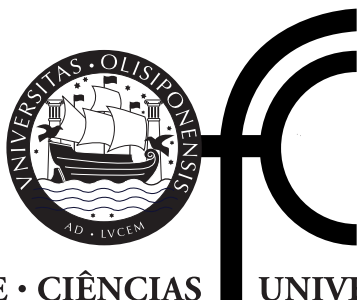


UNIVERSIDADE DE LISBOA

FACULDADE DE CIÊNCIAS

Departamento de Estatística e Investigação Operacional



FACULDADE • DE • CIÊNCIAS

UNIVERSIDADE • DE • LISBOA

Probabilidade *Circa* 1914
e a Construção de
Pacheco d'Amorim

Rui Filipe Vargas de Sousa Santos

Doutoramento em Estatística e Investigação Operacional

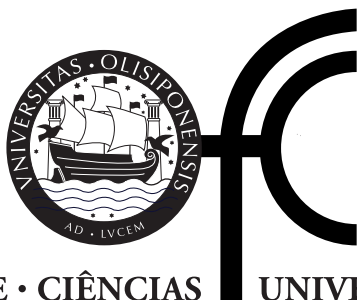
(Especialidade de Probabilidades e Estatística)

2008

UNIVERSIDADE DE LISBOA

FACULDADE DE CIÊNCIAS

Departamento de Estatística e Investigação Operacional



FACULDADE • DE • CIÊNCIAS

UNIVERSIDADE • DE • LISBOA

Probabilidade *Circa* 1914
e a Construção de
Pacheco d'Amorim

Rui Filipe Vargas de Sousa Santos

Tese orientada pelo Professor Doutor Dinis D. F. Pestana

Doutoramento em Estatística e Investigação Operacional

(Especialidade de Probabilidades e Estatística)

2008

Dissertação apresentada à Faculdade de Ciências
da Universidade de Lisboa, para a obtenção do grau
de Doutor em Probabilidades e Estatística.

RESUMO

Diogo Pacheco d'Amorim defendeu a sua tese de doutoramento, intitulada *Elementos de Cálculo das Probabilidades*, em 1914. Nela propõe uma construção rigorosa para a Teoria da Probabilidade baseada no conceito, que considera primitivo, de tiragem ao acaso. Nesta estruturação o autor começa por admitir um modelo padrão, em que o agente da selecção procede a escolhas (ou a lançamentos) em situação de plena aleatoriedade, com total conhecimento do espaço amostra. Assim, pode deduzir a possibilidade de cada elemento sem recorrer ao polémico *princípio da razão insuficiente* de J. Bernoulli e Laplace. O conceito primitivo de que parte é o de probabilidade condicionada, ainda que a sua definição não seja geral. Um outro seu conceito, o de *ponto imagem*, antecipa muitas ideias subjacentes às variáveis aleatórias, ficando perto de alcançar a definição de função de distribuição. Como epílogo, o autor analisa à luz das leis limites, Lei dos Grandes Números e Teorema Limite Central, os casos onde não somos nós os agentes e/ou não temos total informação do espaço amostra, expondo a sua visão sobre as aplicações da Probabilidade, isto é, a sua concepção de Estatística.

Neste trabalho comentaremos as principais ideias apresentadas por Pacheco d'Amorim na sua tese de doutoramento, comparando-a com trabalhos da mesma época, nomeadamente da escola francesa, onde não só salientamos os aspectos mais inovadores na sua conceptualização de Probabilidade, mas também mostramos as limitações de alguns dos conceitos que usa.

Palavras chave: Axiomatização da Probabilidade, Escolha Aleatória, Probabilidade Condicional, Fundamentos da Probabilidade, História da Probabilidade.

AMS (2000) Subject Classification: 60A05, 01A90.

ABSTRACT

Diogo Pacheco d'Amorim presented his thesis *Elements of Probability Calculus* in 1914. His main goal was the axiomatization of Probability. He built up Probability upon the idea of random choice (or random throw); his concept of possibilities, leading to conditional probability, elegantly solves the problem of getting unequal probabilities for elementary events. But his definition of conditional probability is not general. His ideas of *image point* are a predecessor of many interesting developments on functions of random variables, without, unfortunately, inventing the idea of distribution function. His reconstruction of Fubinni's theorem clearly shows that he is aware of the richness brought in, in dealing with random vectors, by the concept of dependence. His construction, distinguishing several layers of incomplete knowledge, begins by a thorough investigation of the standard model (random choice performed by ourselves with perfect knowledge of the sample space); then, using Bernoulli's laws and their consequences, he devises objective ways of deciding whether a random choice performed by someone else, or even by a mechanical device, is undistinguishable from random choice performed by the subject, and thence can be reduced to the standard model.

The main goal of this work is to analyse the contribution to the foundations of Probability Theory, and the bridge between probability and observed data, contained in Pacheco d'Amorim proposal. We also review other previous and contemporary contributions to point out the meaning and complexity of the problem of the foundations of the notion of probability, which is part of Hilbert's sixth problem, and the deep difficulties previous to the definitive axiomatization by Kolmogoroff in 1933.

Keywords: Axiomatization of Probability, Random Choice, Conditional Probability, Foundations of Probability, History of Probability.

AMS (2000) Subject Classification: 60A05, 01A90.

Conteúdo

Lista de Figuras	xv
Lista de Tabelas	xviii
Agradecimentos	xix
 I INTRODUÇÃO	 1
1 Probabilidade <i>Circa</i> 1914 e a Construção de Pacheco d'Amorim	3
 II TRADUÇÃO	 17
2 <i>Elements of Probability Calculus</i>	21
Preface	21
Introduction	23
2.1 CHAPTER I — Finite sets	28
2.2 CHAPTER II — Continuous Probability	47
2.3 CHAPTER III — Random Figures	61

2.3.1	First Part — Random rigid figures	61
2.3.2	Second Part — Random variable figures	71
2.4	CHAPTER IV — Image Point	76
2.4.1	Application	77
2.4.2	Law of possibilities and law of probability	86
2.4.3	<i>A priori</i> and <i>a posteriori</i> laws	87
2.5	CHAPTER V — Jacob Bernoulli's Theorems and the Error Law	93
2.5.1	First Part — Jacob Bernoulli's theorems	93
2.5.2	Second Part — Law of deviations (Error law)	113
2.6	CHAPTER VI — Mathematical Expectation and Mean Value	124
2.7	CHAPTER VII — Conclusion	139
 III A CONSTRUÇÃO DE PACHECO D'AMORIM		153
 O Autor		157
 Prefácio		161
 Introdução		165
 3 Classes Finitas		169
3.1	Elementos e classes possíveis	170
3.2	Possibilidade	172
3.3	Probabilidade de $\mathbf{A}''$ em relação a $\mathbf{A}'$	178
3.4	Teorema da Probabilidade Total e Composta	189

3.5	Teorema de Bayes	196
3.6	Fórmulas inversas da de Bayes	204
3.7	<i>Regra da Sucessão</i> de Laplace	206
3.8	Comentário geral ao capítulo	218
4	Probabilidade Contínua	221
4.1	Pontos e regiões possíveis	222
4.2	Possibilidade	228
4.3	Probabilidade de $\mathbf{X}''$ em relação a $\mathbf{X}'$	239
4.4	Problema do triângulo	245
4.5	Simetria	254
4.6	Paradoxo de Borel-Kolmogoroff	255
4.7	Lançamentos em regiões ilimitadas	262
4.8	Comentário geral ao capítulo	263
5	Lançamento, à Sorte, de Figuras	265
5.1	Lançamento, à sorte, de figuras rígidas	266
5.1.1	Problema do encontro dos dois amigos	271
5.1.2	Paradoxo de Bertrand	274
5.1.3	Problema da Agulha de Buffon	289
5.2	Lançamento, à sorte, de figuras variáveis	296
5.2.1	Figuras poligonais abertas	297
5.2.2	Figuras poligonais fechadas	298
5.2.3	Curvas flexíveis e inextensíveis	313
5.3	Comentário geral ao capítulo	314

6	Ponto Imagem	319
6.1	Definição de Ponto Imagem	320
6.2	Aplicação do Ponto Imagem	330
6.2.1	Princípio de Borel	333
6.2.2	Primeiro caso	341
6.2.3	Segundo caso	342
6.2.4	Terceiro caso	351
6.3	Lei da Possibilidade e Lei da Probabilidade	353
6.4	Leis <i>a priori</i> e leis <i>a posteriori</i>	369
6.4.1	Leis <i>a priori</i>	369
6.4.2	Leis <i>a posteriori</i>	371
6.5	Probabilidade composta e Teorema de Bayes	372
6.5.1	Fórmulas inversas da de Bayes	374
6.6	Comentário geral ao capítulo	375
7	Teorema de JACOB BERNOULLI e Lei dos Desvios	379
7.1	Resultados preliminares	381
7.2	Teoremas de Jacob Bernoulli	395
7.3	Ordem de convergência do número de experiências em relação aos afastamentos	414
7.4	Probabilidade de um número racional	425
7.4.1	Os números Normais de Borel	429
7.5	Lei dos desvios	444
7.6	Comentário geral ao capítulo	458

8	Esperança Matemática e Valor Médio	461
8.1	Esperança Matemática e valor médio em Classes	463
8.1.1	Definição	464
8.1.2	Propriedades	469
8.2	Esperança matemática e valor médio em Regiões	473
8.2.1	Definição	473
8.2.2	Propriedades	474
8.3	Desigualdade	483
8.4	Problema da curva flexível e inextensível	485
8.4.1	Abandonando a hipótese de independência	497
8.4.2	Problema de escala	512
8.4.3	Passeios aleatórios em reticulados	525
8.4.4	O movimento browniano	534
8.5	Problema da agulha de Buffon tratado como valores esperados	545
8.6	Lei dos Grandes Números	550
8.7	Propriedades da Lei de Gauss	559
8.8	Comentário geral ao capítulo	569
9	Conclusão	573
9.1	Redução ao fenómeno padrão	575
9.2	Leis de Bernoulli e análogas	577
9.3	Probabilidade <i>versus</i> certeza	581
9.4	Informação incompleta	586
9.4.1	Primeiro sub-grupo — Caso discreto	586

9.4.2	Segundo sub-grupo — Caso contínuo	589
9.5	Aleatoriedade	596
9.6	Estatística <i>circa</i> 1914	601
9.7	Comentário geral ao capítulo	610

IV CONCLUSÃO 613

10 Comentário Geral à Construção de Pacheco d’Amorim 615

10.1	Formalização da concepção geral de Pacheco d’Amorim	617
10.2	<i>Grundbegriffe</i> de Kolmogoroff (1933)	625
10.2.1	Contextualização	626
10.2.2	Axiomática	633
10.2.3	Variáveis Aleatórias e Função de Distribuição	649
10.2.4	Esperança Matemática e Probabilidade Condicionada .	660
10.2.5	Espaços de dimensão superior e independência	664
10.2.6	Aplicabilidade	669
10.2.7	Comentário aos fundamentos de Kolmogoroff e a versão condicional de Rényi	672
10.3	Os <i>Colectivos</i> de von Mises e a visão frequencista	679
10.3.1	A axiomática de Richard von Mises	681
10.3.2	O Cálculo das Probabilidades nos <i>colectivos</i>	682
10.3.3	Os Teoremas Fundamentais	689
10.3.4	Independência <i>versus</i> Estatística clássica	691
10.4	A <i>Permutabilidade</i> de Bruno de Finetti e a visão bayesiana . .	692

CONTEÚDO

xi

10.4.1	Axiomática de Bruno de Finetti	695
10.4.2	Definição de probabilidade e princípio da coerência . .	696
10.4.3	Aditividade <i>versus</i> σ -aditividade	701
10.4.4	Probabilidade Condicionada	707
10.4.5	Permutabilidade e Teoremas de Representação	710
10.4.6	A visão <i>lógica indutiva</i>	716
10.5	Comentário final à construção de Diogo Pacheco d'Amorim . .	722

BIBLIOGRAFIA

727

Lista de Figuras

2.1	Figure 1	49
2.2	Figure 2	50
2.3	Figure 3	54
2.4	Figure 4	55
2.5	Figure 5	64
2.6	Figure 6	65
2.7	Figure 7	67
2.8	Figure 8	69
2.9	Figure 9	72
2.10	Figure 10	74
2.11	Figure 11	75
2.12	Figure 12	89
4.1	Lançar, à sorte, um ponto numa região composta	226
4.2	Lançar, à sorte, um ponto num complexo de regiões — 1 . . .	226
4.3	Lançar, à sorte, um ponto num complexo de regiões — 2 . . .	227
4.4	Lançar, à sorte, um ponto num círculo	231
4.5	Lançar, à sorte, um ponto num complexo de regiões	232

4.6	Inversão da ordem de lançamento de um complexo de regiões	237
4.7	Lançar, à sorte, um ponto num quadrado	246
4.8	Problemas dos Triângulos	250
4.9	Problemas dos Triângulos 2	250
5.1	Lançar, à sorte, a região plana A na região plana B	271
5.2	Problema do encontro de dois amigos	272
5.3	Problema de Bertrand – Solução de Pacheco d'Amorim	274
5.4	Problema de Bertrand – 1ª Solução de Bertrand	276
5.5	Problema de Bertrand – 2ª Solução de Bertrand	277
5.6	Problema de Bertrand – 3ª Solução de Bertrand	278
5.7	Erro de Pacheco d'Amorim no Problema de Bertrand	280
5.8	Lançamento, à sorte, de pontos num círculo	281
5.9	Cordas escolhidas aleatoriamente	282
5.10	Simulação do Problema de Bertrand	283
5.11	Duas cordas com possibilidades distintas	286
5.12	Problema da Agulha de Buffon	289
5.13	Região de lançamento da agulha	290
5.14	Campo de variação da agulha	292
5.15	O jogo do <i>franc-carreau</i>	293
5.16	Tirar, à sorte, a forma de um vértice	297
5.17	Tirar, à sorte, a forma de um triângulo	299
5.18	Tirar, à sorte, a forma de um quadrilátero	300
5.19	Quando um quadrilátero à sorte sai triângulo	303

5.20	Quadriláteros tirados à sorte	305
5.21	Gráficos dos ângulos de um quadrilátero	306
5.22	Tirar, à sorte, a forma de um pentágono (original)	308
5.23	Tirar, à sorte, a forma de um pentágono (alterado)	309
6.1	Leis à priori	370
6.2	Leis à posteriori	371
8.1	Simulação do Problema curva flexível	490
8.2	Simulação do Problema curva flexível com dependência	498
8.3	Passeios aleatórios utilizando distribuição Beta	501
8.4	Gráfico de $f(x) = \frac{\sin(x)}{x}$	508
8.5	Gráfico de $g(x) = 3 \left(\frac{\sin(x)}{x^3} - \frac{\cos(x)}{x^2} \right)$	510
8.6	Passeios aleatórios com n distinto	514
8.7	Passeios com normalização	518
8.8	Passeios com distância esperada fixa	522
8.9	Passeios aleatórios em reticulados	527

Lista de Tabelas

5.1	Simulação do Problema de Bertrand	283
5.2	Variação dos ângulos de um quadrilátero	307
7.1	Valores de α e $\left \frac{n_p}{n_q} - \frac{p}{q} \right $ em 50 provas com $p = 0.5$	397
8.1	Simulação do Problema curva flexível	491
8.2	Simulação do Problema curva flexível com dependência	499
8.3	Simulação do Problema curva flexível com dependência	500
8.4	Simulação do Problema curva flexível com dependência Beta .	502
8.5	Quadrado das distâncias entre os pontos extremos	513
8.6	Quadrado das distâncias entre os pontos extremos	519
8.7	Simulação com processo com limite esperado fixo	520
8.8	Evolução de $\mathbb{E}(d^2)$ em função de n e ρ	521
8.9	Distâncias entre os pontos extremos	523
8.10	Simulação com distância esperada fixa	524
8.11	Simulação de π pela Lei de Gauss	564
8.12	Simulação de π pela agulha de Buffon com $\delta = 1$ e $\gamma = 0.5$. .	565
8.13	Simulação de π pela agulha de Buffon com $\delta = \gamma = 1$	566

8.14	Simulação de π pelo lançamento de um ponto	567
8.15	Erros de estimação pela fórmula de Wallis	568

Agradecimentos

Para que fosse possível a realização deste trabalho muitos foram os que, directa ou indirectamente, deram um contributo significativo para a sua concretização. Por este motivo, gostaria de expressar aqui a minha enorme gratidão a todos aqueles que, de alguma forma, contribuíram para este fim, especialmente:

Ao Professor Dinis Pestana, pelo tema sugerido, pela contínua orientação e disponibilidade, pela precisão e pertinência dos comentários, pela confiança e apoio desmedido, pelo entusiasmo contagiante e renovado em cada conversa, pela compreensão e amizade patenteada ao longo deste projecto, bem como pelas sugestões e críticas imprescindíveis à concretização deste trabalho;

À Professora Sandra Mendonça, pela troca de ideias, sempre proveitosas e enriquecedoras, que me ajudaram a compreender melhor algumas das teorias concebidas por Pacheco d'Amorim;

Ao Professor José Bayolo Pacheco de Amorim, pela sua autorização na divulgação da edição diplomática da tradução da tese de doutoramento do seu pai, bem como pelo manifesto apoio nesta investigação e na disponibilidade em partilhar connosco histórias sobre os feitos de Diogo Pacheco d'Amorim;

Ao Instituto Politécnico de Leiria e ao Centro de Estatística e Aplicações da Universidade de Lisboa pelas facilidades com que, ao longo dos últimos

anos, me criaram condições favoráveis para a concretização desta dissertação;

Ao Nuno Dias, colega de pesquisa em História da Matemática em Portugal, pelas diversas informações que me facultou e pela sua contínua preocupação com a minha investigação;

Ao Miguel Felgueiras, colega de gabinete e de pesquisa em Probabilidades, pela permanente troca de ideias bem como pelo ambiente que me propiciou;

A todos os meus colegas do Instituto Politécnico de Leiria, nomeadamente do Departamento de Matemática, pela forma como me encorajaram e confiaram nas minhas capacidades, em todas as circunstâncias, tornando-me apto a completar este ciclo;

A todos os meus amigos pelas oportunas manifestações de companheirismo e de encorajamento que me permitiram manter o ânimo sempre elevado;

Ao Nuno, à Dalila e ao David pela sua enorme amizade e por estarem sempre presentes;

Aos meus Pais, Jorge e Antonieta, por toda a dedicação com que me criaram, desde o seu amor e carinho constantes até ao investimento que efectuaram na minha educação, cuja contínua orientação tornou possível atingir esta meta;

À Susana e ao Pedro, minhas fontes inspiradoras, por tudo o que abdicaram em prol das minhas pesquisas; pelo seu amor infinito, a sua paciência ilimitada e a sua confiança inesgotável, que foram factores determinantes ao longo deste trabalho, cujo apoio incondicional tornou esta tarefa exequível.

Aos meus Pais,

à Susana

e ao Pedro.

Parte I

INTRODUÇÃO

Capítulo 1

Probabilidade *Circa* 1914 e a Construção de Pacheco d’Amorim

“Le nom seul de calcul des probabilités est un paradoxe: la probabilité, opposée à la certitude, c’est ce qu’on ne connaît pas? Cependant, beaucoup de savants éminents se sont occupés de ce calcul, et l’on ne saurait nier que la science n’en ait tiré quelque profit. Comment expliquer cette apparente contradiction?”

[Poincaré, 1902, p. 210]

No início do século XX, a Teoria da Probabilidade (ou Cálculo de Probabilidades como nessa altura é apelidada, nomeadamente pela escola francesa) atravessa um período conturbado na sua evolução.

O Cálculo das Probabilidades, para utilizar a mesma nomenclatura que Hacking (1975), emergiu no século XVII com a análise dos resultados de jogos de *azar*⁽¹⁾, sendo habitualmente identificado o desponar desta ciência com a correspondência entre Fermat (1601–1665) e Pascal (1623–1662) em 1654

⁽¹⁾ Segundo o Dicionário Etimológico da Língua Portuguesa, *azar* advém do árabe “*az-zahar*” que significa “*felicidade, acto, caso feliz, dado*”; segundo o Dicionário Aurélio *azar* tem origem no árabe em “*az-zahar*” ou “*az-zahr*” e significa “*má sorte; fortuna adversa;*

para resolver os problemas propostos por Chevalier de Méré (1610–1685) sobre a fracção da aposta que cabe a cada adversário quando uma partida de jogos é interrompida antes da conclusão.

Os resultados dos jogos de *azar* são caracterizados por terem duas facetas à primeira vista antagónicas. Por um lado, são descritos por uma desordem individual, isto é, pela impossibilidade de previsão de um resultado específico, consequência da incerteza associada a cada um dos resultados particulares de um jogo de *azar* (caso contrário não seria realmente um jogo de *azar*); por outro lado, são também caracterizados pela sua ordem colectiva, ou seja, pela regularidade existente quando visualizamos um conjunto (com um número razoável) de resultados. Os fenómenos caracterizados por esta dupla faceta, denominados fenómenos aleatórios, são o alvo de análise da Teoria da Probabilidade. Todavia, os fenómenos aleatórios não se restringem aos resultados de jogos de *azar*.

Ao longo do século XVII verifica-se uma crescente observação e colecção de dados, nomeadamente de tabelas de mortalidade⁽²⁾, sendo cada vez mais notória a regularidade das frequências relativas quando temos um conjunto com um número elevado de observações, tornando-se evidente que os resultados

revés; fatalidade; desgraça; infortúnio; casualidade; acaso”. Tiago de Oliveira (1991a) refere que a palavra *azar* deriva do árabe “*al azhar*” que apresenta não só a conotação que habitualmente lhe damos de má-sorte ou má-fortuna, mas também designa *acaso*, sendo “*az-zahar*” a patrona da fortuna, correspondente à deusa Tykhe (ou Tyché) da mitologia grega e à deusa Fortuna presente na mitologia romana. Refira-se que estes deuses simbolizam o acaso dos destinos humanos que, segundo os seus caprichos, será benévolo ou malévol.

⁽²⁾ Por exemplo, os trabalhos realizados por John Graunt (1620–1674), John de Witt (1625–1672) e Edmond Halley (1656–1742), para citar algumas das mais importantes referências desta época.

dos jogos de *azar* não são os únicos fenómenos aleatórios⁽³⁾.

Por este motivo aparecem, nesta época, as primeiras tentativas de inferência estatística, verificando-se um aumento da importância da Teoria da Probabilidade na Estatística, surgindo novos resultados teóricos para justificar esta aproximação, destacando-se a Lei [fraca] dos Grandes Números de Bernoulli (1713)⁽⁴⁾ e os primeiros enunciados do Teorema Limite Central, inicialmente restritos às provas de Bernoulli, em de Moivre (1738) e Laplace (1812). Assim, a Teoria da Probabilidade conseguiu, lentamente, libertar-se da sua inicial dependência dos jogos de *azar*, o que leva, por exemplo, Laplace a afirmar

“en les appliquant aux questions les plus importantes de la vie, qui ne sont en effet, pour la plupart, que des problèmes de probabilité.”

[Laplace 1812, Introduction, p. i]

No entanto, apesar da crescente progressão de interesse da Teoria da Probabilidade, como será possível desenvolver esta teoria sem ter os seus conceitos básicos, tais com *probabilidade*, *acaso* ou *aleatório*, rigorosamente definidos? Desta forma, sendo essencial obter-se uma definição clara destas

⁽³⁾ Hacking (1975), Stigler (1986), Bernstein (1998) e Hald (2003) fazem uma boa descrição das origens da Teoria da Probabilidade, das suas primeiras obras, não só da sua inicial dependência na análise de resultados de jogos de *azar* e do cálculo combinatório, como ainda do seu aumento de importância na Estatística com a constatação de regularidade quando se possui uma grande quantidade de dados.

⁽⁴⁾ Esta Lei foi apresentada pela primeira vez por Jakob (ou Jacob, que foi traduzido para Jacques pelos francófonos, para James pelos anglo-saxónicos e para Giacomo pelos italianos) Bernoulli (1654–1705) na quarta parte da sua obra inacabada e póstuma, publicada pelo seu sobrinho Nicolau Bernoulli oito anos após a sua morte, sob o título de *Ars Conjectandi* (A arte de conjecturar). Este resultado foi depois designado por Lei dos Grandes Números por Poisson (1837, p. 7), sendo ainda hoje um dos resultados mais importantes da Teoria da Probabilidade e da sua ligação à Estatística.

noções, ao longo do século XIX diversas obras debatem quer as possíveis definições destes conceitos quer os limites de aplicabilidade desta ciência. Porém, a obtenção de uma definição rigorosa de alguns destes conceitos é uma tarefa delicada, como podemos deduzir das seguintes palavras de Bertrand (1822–1900).

“Comment oser parler des lois du hasard? Le hasard n’est-il pas l’antithèse de toute loi? En repoussant cette définition, je n’en proposerai aucune autre.”

[Bertrand, 1888, p. vi]

Assim, neste período é sentida uma insatisfação geral na definição de probabilidade baseada na equiprobabilidade, principal definição adoptada desde a publicação da obra prima de Laplace (1749–1827), *Théorie Analytique des Probabilités* em 1812⁽⁵⁾. Nesta definição, actualmente denominada clássica ou laplaceana, a probabilidade de um dado acontecimento é igual ao quociente entre o número de casos favoráveis a esse acontecimento e o número total de casos possíveis, sob a hipótese de os casos serem igualmente prováveis (equiprobabilidade) e serem em número finito. As hipóteses, excessivamente restritas, subjacentes a esta definição tornam necessária uma nova definição de probabilidade, de forma a incluir as situações em que não existe equiprobabilidade dos acontecimentos ou situações onde o universo não é finito, quer este seja numerável ou não (contínuo). Todavia, parece haver um certo cepticismo na possibilidade de existência de uma definição mais geral de probabilidade, bem patente nas seguintes palavras de Poincaré (1854–1912).

“On ne peut guère donner une définition satisfaisant de la Probabilité. On dit ordinairement: la probabilité d’un événement est le

⁽⁵⁾ Ian Hacking (1975) identifica a origem da utilização da equipossibilidade (ou equiprobabilidade) num *memorandum* intitulado *De incerti aestimatione* de Leibniz que data de 1678.

rapport du nombre des cas favorables à cet événement au nombre total des cas possibles.”

[Poincaré, 1896, pág. 24]

Há ainda a acrescentar, para salientar as dificuldades sentidas nesta procura de uma definição mais geral de probabilidade, um descrédito na Teoria da Probabilidade devido à existência de inúmeros paradoxos, como os apresentados por Bertrand em 1888. Um dos paradoxos (Bertrand 1888, p. 4–5), actualmente conhecido por paradoxo de Bertrand, consiste em apresentar três possibilidades distintas de resolução de um problema, qualquer uma delas rigorosamente coerente e em consonância com uma abordagem intuitiva, mas que fornecem três valores distintos para a probabilidade pretendida. Outro paradoxo apresentado por Bertrand (1988, p. 4) coloca em causa a extensão, muitas vezes utilizada, da definição clássica à probabilidade contínua, isto é, utilizando uma medida representativa da sua proporção geométrica (comprimento, área, volume) e definindo a probabilidade de uma região (região favorável) como sendo o quociente entre a sua medida e a medida da região total (universo) sob a hipótese de a probabilidade ser proporcional à medida e a medida do universo ser finita (interpretação geométrica de probabilidade).

Deste modo, a Teoria de Probabilidade é, nesta altura, assombrada por diversos paradoxos⁽⁶⁾ que ilustram a ambiguidade existente em diversas noções básicas tais como a *escolha ao acaso* (*au hasard*), sendo necessária a construção de uma teoria que clarifique estes conceitos.

Apesar desta depreciação na Teoria da Probabilidade, existe uma enorme motivação na procura de uma definição mais geral de probabilidade. Um dos factores determinantes é a crescente importância da Teoria da Probabilidade na Física, pois é nesta altura que surgem as primeiras ideias que irão dar ori-

⁽⁶⁾ Székely (1986) explora bem diversos paradoxos na evolução da Teoria da Probabilidade e da Estatística.

gem, na terceira década do século XX, à Mecânica Quântica e ao paradigma da incerteza aplicado ao mundo microscópico. Provavelmente por este motivo David Hilbert (1862–1943), um dos matemático mais notáveis da época, na sua famosa alocução no Congresso Internacional de Matemática de Paris de 1900⁽⁷⁾, expõe um conjunto de 23 problemas por resolver que deveriam orientar a investigação em Matemática durante o século XX, consistindo um desses problemas, o sexto, na axiomatização das áreas da Física onde a Matemática desempenha um papel predominante, figurando o Cálculo das Probabilidades e a Mecânica como as primeiras da lista. Desta forma, David Hilbert desafia a *comunidade matemática* a procurar uma fundamentação rigorosa para a Teoria da Probabilidade a fim de que esta possa ser tratada de uma forma axiomática, como as outras áreas da Matemática, de modo a ser utilizada sem ambiguidade. Plato (1994) refere diversas tentativas frustradas de axiomatização da probabilidade anteriores ao tratado de Kolmogoroff, que surge em 1933, onde se apresenta a primeira axiomatização de probabilidade que definitivamente resolve os paradoxos patentes nesta teoria, intitulada *Grundbegriffe der Wahrscheinlichkeitstheorie* e traduzida para inglês em 1950 por *Foundations of the Theory of Probability*. Esta axiomática baseia-se nas características das funções de conjuntos, nomeadamente na área actualmente denominada por Teoria da Medida, definindo probabilidade como uma medida em que a probabilidade do universo é unitária. Contudo, as bases para que a probabilidade fosse tratada como uma medida foram criadas, essencialmente, por Émile Borel (1871–1956), Henri Lebesgue (1875–1941), Maurice Fréchet (1878–1973), Constantin Carathéodory (1873–1950), Johann Radon (1887–1956), Otto Nikodym (1889 – 1974), entre outros, algumas das quais são posteriores à tese de doutoramento de Diogo Pacheco d’Amorim.

É neste panorama, antecedente aos fundamentos de Kolmogoroff e a resul-

⁽⁷⁾ Estes problemas podem ser consultados em Hilbert (1902).

tados que permitiram uma maior abstracção na Teoria da Medida, que surgem os *Elementos de Cálculo das Probabilidades* de Diogo Pacheco d'Amorim, em 1914, sendo objectivo do autor, conforme enuncia no PREFÁCIO da sua obra, responder ao desafio de David Hilbert, isto é, fornecer resposta aos problemas que na época assombravam a Teoria da Probabilidade, generalizando a definição de probabilidade e clarificando os conceitos básicos desta Teoria, de forma que esta consiga desenvolver-se sem ambiguidade ou paradoxos.

Assim, Pacheco d'Amorim procura dar uma definição mais geral de probabilidade, assentando a construção da sua teoria na proposição, que considera primitiva, de *tirar, à sorte, um elemento de uma classe finita* (ou, o correspondente para a probabilidade contínua, *lançar, à sorte, um ponto numa região limitada*), considerando que esta proposição é perfeitamente clara quando somos nós os agentes da tiragem e desfrutamos de total conhecimento do espaço amostra. Pacheco d'Amorim constrói toda a teoria sob estas hipóteses, generalizando na CONCLUSÃO da sua obra quer para o caso em que a tiragem de um elemento (ou lançamento de um ponto) não é efectuada por nós, mas por um ser semelhante a nós ou por um agente de outra natureza, quer para a situação em que nós não possuímos total informação do espaço amostra.

Com base no seu conceito primitivo, Pacheco d'Amorim define possibilidade de uma classe (região) para posteriormente definir probabilidade de uma classe (região) $\mathbf{A}'$ em relação a outra classe (região) $\mathbf{A}$ que contém a primeira. Desta forma, para Pacheco d'Amorim, a probabilidade de uma classe (região) só está definida em relação a outra classe (região), pelo que a definição de probabilidade de Pacheco d'Amorim corresponde ao conceito de probabilidade condicionada actualmente utilizado, restrito ao caso em que $\mathbf{A}' \subset \mathbf{A}$ ⁽⁸⁾. De facto, com a sua construção, Pacheco d'Amorim consegue ge-

⁽⁸⁾ Refira-se que Pacheco d'Amorim não é o único que define probabilidade somente na

neralizar o conceito de probabilidade para acontecimentos não equiprováveis através de uma elaborada construção de complexos de classes (regiões).

Um outro aspecto que consideramos notável na tese de doutoramento de Pacheco d'Amorim é a sua constante tentativa de aproximar Probabilidade e Medida, apresentando, no capítulo PONTO IMAGEM, uma admirável construção que nos permite passar de um ponto lançado à sorte directamente numa região (distribuição uniforme na região) para outro que é sua imagem. Desta forma o autor obtém a função densidade de uma variável aleatória, ou lei de possibilidade de uma região, como ele a denomina, que pode ser distinta da uniforme sem ser necessário recorrer aos complexos de regiões. Assim, parece-nos que Pacheco d'Amorim fica próximo de obter a definição moderna de variável aleatória, faltando-lhe o conceito de função de distribuição, noção que é actualmente fundamental para a construção da Teoria da Probabilidade. Ainda no mesmo capítulo, Pacheco d'Amorim apresenta uma notável reconstrução do Teorema de Fubinni (ou Teorema de Fubinni-Tonelli), usando a noção de projecção e de condicionamento, permitindo abordar o cálculo de integrais múltiplos de forma simples, com integrais iterados.

Parece-nos que, nesta parte da sua obra, Pacheco d'Amorim tem um mérito excepcional no trabalho que desenvolve, demonstrando ter a intuição de que a Teoria da Medida é o caminho correcto para a construção rigorosa da Teoria da Probabilidade, antecipando, de certa forma, as ideias notáveis de Fréchet de que probabilidade é uma faceta de medida, e que os instrumentos para tratar probabilidade e integral são os mesmos.

Temos que admitir, contudo, que a sua obra é, por vezes, pouco clara.

situação de condicionamento, pois, por exemplo, Keynes (1921), de Finetti (1937), Jeffreys (1939) ou Rényi (1955) também o fazem, apesar de a sua fundamentação ser bem distinta da apresentada por Pacheco d'Amorim em 1914.

Uma das razões assenta na notação adoptada, que é pouco diversificada, sendo utilizado o mesmo símbolo para diversos objectos, como, por exemplo, quando o autor utiliza (A) , que entre outras coisas pode significar um elemento da classe (A) , a própria classe (A) , o número de elementos da classe (A) , um ponto de uma região (A) , a própria região (A) , a medida da região (A) . Outra razão que contribui para aumentar a dificuldade de interpretação de algumas afirmações de Pacheco d'Amorim é a escassa, quase inexistente, bibliografia por ele usada. O autor, ao longo de toda a sua obra, limita-se a fazer, em notas de rodapé, sete referências a cinco obras: aos clássicos J. Bernoulli [*Ars Conjectandi*] e Laplace [*Essai Philosophique sur les Probabilités*], e aos seus contemporâneos H. Poincaré [*La Science et l'Hypothèse*], E. Borel [*Éléments de la Théorie des Probabilités*] e Bertrand [*Calcul des Probabilités*] que pertencem à escola francesa de probabilidades. Refira-se também que, destas sete referências, uma é feita no PREFÁCIO e cinco na CONCLUSÃO da tese, restando apenas uma referência ao longo do desenvolvimento da sua teoria. Desta forma, torna-se difícil distinguir o que é criado por Pacheco d'Amorim do que é reconstruído com base em outras obras por ele consultadas.

Torna-se, também, delicado interpretar algumas afirmações do autor, de entre as quais salientamos, pela estranheza que nos causaram, particularmente três. A primeira encontra-se numa referência ao Teorema de Bayes, ao afirmar que:

“é errónea a dedução que desta fórmula se faz nos livros de probabilidade”

[Pacheco d'Amorim, 1914, p. 27]

quando, por exemplo, Laplace (1774, 1812) parece dominar bem o assunto. Refira-se, contudo, que Pacheco d'Amorim não é o único insatisfeito com as demonstrações deste teorema existentes na época, pois Bertrand (1888, p.

25–26) e, alguns anos mais tarde, Keynes (1921, p. 176) também mostram um certo descontentamento com as fundamentações normalmente apresentadas na época acerca da fórmula de Bayes. Todavia, parece-nos que as dificuldades sentidas por Pacheco d’Amorim se devem ao facto de a sua definição de probabilidade condicionada $\mathbb{P}(\mathbf{A} \mid \mathbf{B}) = \frac{\mathbb{P}(\mathbf{A})}{\mathbb{P}(\mathbf{B})}$ ser restrita ao caso $\mathbf{A} \subset \mathbf{B}$, pois, se esta definição fosse estendida para o caso geral, por exemplo através de $\mathbb{P}(\mathbf{A} \mid \mathbf{B}) = \mathbb{P}(\mathbf{A} \cap \mathbf{B} \mid \mathbf{B})$, teria obtido a definição de probabilidade condicionada actualmente utilizada, o que levaria a uma demonstração do Teorema de Bayes menos penosa.

A segunda afirmação enigmática a que nos referimos resulta da distinção que o autor faz entre esperança matemática e valor médio, mencionando em nota de rodapé:

“Embora esta distinção entre esperança matemática e valor médio não costume vir explicitamente feita nos livro de probabilidades, todos os autores dão a estes termos a significação que acabamos de atribuir-lhes.”

[Pacheco d’Amorim, 1914, p. 137]

Ora, nas obras da época por nós consultadas, não identificamos, nem implicitamente nem explicitamente, esta distinção (pelo menos de forma semelhante à utilizada por Pacheco d’Amorim). Há, ainda no capítulo dedicado à esperança matemática e valor médio, outra afirmação bastante misteriosa:

“O valor provável do quadrado da distância que separa os pontos extremos duma curva flexível, lançada à sorte sobre um plano, é nulo, qualquer que seja o comprimento da curva, logo que seja finito.”

[Pacheco d’Amorim, 1914, p. 147]

Se o quadrado da distância, que naturalmente não assume valores negativos, tem valor esperado nulo, então Pacheco d’Amorim está a afirmar que,

se lançarmos à sorte uma curva flexível e inextensível sobre um plano (com comprimento finito e fixo) vamos obter quase certamente uma curva fechada. Não nos parece tal ideia aceitável.

O capítulo do LANÇAMENTO, À SORTE, DE FIGURAS é decepcionante, não sendo decifráveis as intenções do autor. Com as suas definições de lançamentos de figuras consegue, de facto, obter uma única resposta para o problema (paradoxo) de Bertrand. Contudo, acrescente-se ainda que, ao contrário das três resoluções apresentadas por Bertrand (1888, p. 4–5), a solução de Pacheco d'Amorim não é satisfatória, pois nela não existe uma correspondência biunívoca entre o ponto *lançado à sorte* e as cordas, isto é, a cada ponto não corresponde uma única corda e vice-versa, enquanto nas três propostas de resolução de Bertrand esta relação é verificada (com excepção de, numa das soluções, o centro da circunferência que é um conjunto singular, de medida nula). Todavia, neste capítulo, o autor resolve o problema da agulha de Buffon utilizando um argumento interessante e porventura inovador, que para a discussão da probabilidade não é necessário considerar o plano $\mathbb{R}^2$, sendo suficiente considerar um paralelograma com uns lados paralelos às rectas da folha e os outros paralelos à direcção achada para a agulha (utilizando a Proposição IX do Capítulo II, pois todos os paralelogramas assim definidos possuem a mesma probabilidade e a Definição IV pelo facto de ser possível obter uma região plana tão grande quanto se queira com a união de paralelogramas assim definidos).

Há, também, a salientar, pela negativa, alguns erros incomprensíveis. Referimo-nos, por exemplo, às fórmulas inversas do Teorema de Bayes aquando da análise das classes finitas (p. 27 e 28), onde através de $P_i = \frac{\omega_i p_i}{\sum \omega_k p_k}$ deduz (correctamente) a fórmula de $\omega_i = \frac{P_i}{p_i} \left(\sum \frac{P_k}{p_k} \right)^{-1}$ e induz (erradamente), justificando-se pela simetria da primeira fórmula em relação a ω_i e a p_i , que $p_i = \frac{P_i}{\omega_i} \left(\sum \frac{P_k}{\omega_k} \right)^{-1}$, esquecendo-se que na dedução da primeira

utiliza o facto de $\sum \omega_i = 1$ que, obviamente, não pode utilizar na segunda, pois $\sum p_i \neq 1$ (provavelmente uma das razões deste erro é a sua notação não ser clara, não especificando que p_i depende quer das causas quer dos efeitos, pois trata-se de uma probabilidade condicionada).

Em relação ao tratamento dos teoremas limite (resultados assintóticos), onde é dada maior importância à Lei dos Grandes Números que ao Teorema Limite Central, consideramos que Pacheco d'Amorim vacila entre os dois extremos qualitativos opostos. Por um lado, a sua análise é decepcionante pelo facto de ser restrita a uma linha clássica de aproximar probabilidades (em provas de Bernoulli com probabilidade fixa — p) através da regra de Stirling, não se ocupando do contexto mais geral de considerar uma sucessão de variáveis aleatórias convergentes para uma variável aleatória gaussiana, problema deveras mais delicado, cuja análise exige outro tipo de ferramentas mais sofisticadas, como as transformadas de Laplace. Desta forma, parece-nos que o autor desconhece os resultados obtidos pela escola russa, como, por exemplo, as demonstrações de Markoff (1913) dos resultados obtidos por Lyapounov em 1901. Contudo, mesmo tendo em consideração o alheamento do autor em relação a estes resultados seus contemporâneos, Pacheco d'Amorim poderia ter tentado ir mais longe, pois já Laplace (1812) e Poisson (1837) analisam esta situação, apesar de não fornecerem demonstrações rigorosas, e este teorema assume um papel preponderante na Teoria da Probabilidade, razão pela qual Pólya (1920) o denominou por Teorema Limite Central. Por outro lado, existe uma originalidade na justificação de que a convergência para a distribuição gaussiana (*lei dos desvios*) só se verifica se a ordem do número de provas relativas ao afastamento é a segunda, justificando este resultado pela demonstração de que, considerando ε um valor qualquer positivo e α o número de sucessos a mais relativamente ao número de sucessos esperado, a probabilidade de $\alpha > \varepsilon m^{\frac{n}{n+1}}$ para $n > 1$ converge para zero e a probabilidade

de $\alpha < \varepsilon m^{\frac{1}{n}}$ quando $n > 2$ também converge para zero⁽⁹⁾.

Apesar dos erros contidos na sua obra e de algumas afirmações obscuras, as ideias apresentadas por Pacheco d'Amorim, quer a sua abordagem filosófica na construção da probabilidade e na sua aplicação, quer na aproximação constante entre o tratamento integral e probabilidade, quer a sua demonstração do Teorema Limite Central, são de uma originalidade e de uma riqueza invulgar, razão pela qual consideramos urgente uma análise cuidada desta obra. Vamos, então, apresentar uma análise detalhada dos principais assuntos dissecados por Pacheco d'Amorim na sua tese de doutoramento intitulada *Elementos do Cálculo das Probabilidades*. Desta forma, começaremos por expor, na segunda parte deste trabalho, a tradução para Inglês da tese de doutoramento de Pacheco d'Amorim efectuada em colaboração com o Professor Doutor Dinis Pestana e a Professora Doutora Sandra Mendonça. Na Terceira parte, que intitulamos A CONSTRUÇÃO DE DIOGO PACHECO D'AMORIM, comentaremos os *Elementos do Cálculo das Probabilidades* capítulo a capítulo, isto é, vamos expor a construção de Pacheco d'Amorim tal e qual como o autor a concebeu, ilustrando a sua validade e as suas limitações e comparando os seus resultados com os patentes nas principais obras internacionais disponíveis na época (onde nos centramos na escola francesa pelo facto desta constituir a principal fonte bibliográfica do autor). Finalizaremos este trabalho com uma apreciação geral à tese de doutoramento de Pacheco d'Amorim onde apresentaremos, com o objectivo de salientar a riqueza da proposta de Pacheco d'Amorim, outras construções posteriores à sua obra, tais como as axiomáticas de Kolmogoroff e de Rényi, fundamenta-

⁽⁹⁾ Na verdade, não encontramos este tratamento em mais nenhuma obra na época e Manuel dos Reis (1929) atribui, na sua tese de Doutorado, a primazia destes resultados a Pacheco d'Amorim (apesar de considerar que estes poderiam ser obtidos como corolários do Teorema de Moivre-Laplace, como ilustraremos no capítulo 7).

das na Teoria da Medida, a visão frequencista de Richard von Mises, baseada em *colectivos*, e a interpretação bayesiana personalista de Bruno de Finetti, fundada na *permutabilidade*.

Parte II

TRADUÇÃO

Tradução

Esta parte tem como finalidade apresentar a tradução, para Inglês, da tese de doutoramento de Pacheco d'Amorim que foi efectuada em colaboração com o Professor Doutor Dinis Pestana e a Professora Doutora Sandra Mendonça, estando disponível on-line⁽¹⁰⁾ com um formato diferente. Na versão disponível na Internet houve o cuidado de paginar de acordo com a obra original, correspondendo cada página da tradução à página com igual numeração do original, pelo que o ficheiro disponível contém, lado a lado, o *fac simile* da obra e a respectiva versão em inglês. Na tradução que de seguida apresentamos o aspecto foi alterado com o objectivo de melhorar a apresentação e esta não se tornar demasiado longa, mas mantendo o conteúdo. Saliente-se que, nesta tradução, o principal objectivo foi obter um texto que espelhasse as ideias concebidas por Diogo Pacheco d'Amorim. Neste sentido, da parte dos tradutores, houve a tentativa de apresentar, com o maior rigor, a fundamentação de Pacheco d'Amorim, tentando não proceder a qualquer interpretação ou aperfeiçoamento, por maior que fosse a inclinação natural para fazê-lo. Todavia, com o objectivo de tornar mais clara a sua leitura, situações houve onde sentimos a necessidade de alterar ligeiramente a notação e, em alguns casos, efectuar algumas correcções (que assinalámos em forma de nota de rodapé).

Deste modo expomos, de seguida, a tradução dos *Elementos de Cálculo das Probabilidades* de Diogo Pacheco d'Amorim.

⁽¹⁰⁾ Consultar, por exemplo, em

www.estg.ipleiria.pt/~rsantos/Elements_of_Probability_Calculus.pdf.

Capítulo 2

Elements of Probability Calculus

Preface

This volume, for which the title — *An Essay Towards Rationalizing Probability Calculus* — would perhaps be more appropriate, gives an outstanding role to a concept that, until now, never got the relevance it deserves — the concept of *extracting, at random, an element from a set* or of *throwing, at random, a point in a region*.

Henri Poincaré⁽¹⁾ goes as far as saying that such a statement has, by itself, no meaning. But the truth is that this proposition has a very clear and precise meaning for the agent of the random extraction or of the random throw, and this allows us to construct the theory of probability with clarity and rigor. Starting from this primitive concept, the theory of probability can be reduced to a systematic sequence of propositions and definitions, as any other branch of pure mathematics. In this approach, discontinuous and continuous probability are identical in all aspects, and paradoxes have no place in the ensuing theory.

⁽¹⁾ H. POINCARÉ, *La Science et l'Hypothèse*, p. 226.

Once the theory of probability of random extractions and of random throws done by ourselves has been built, its extension to phenomena whose outcomes are similar to extractions or throws performed by agents similar to us is rather easy, in case the extractions are done under some rigid circumstances.

The theory thus constructed can be applied to the study of natural phenomena, insofar as we reject, *a priori*, the determinist hypothesis, that, in fact, is incompatible with probability theory; under this proviso, the application is easily done.

The perspective we have adopted led us to change the form and the essence of Probability. We had to generalize the definition of probability, a generalization needed to prove Bayes formula, and absolutely unavoidable in the study of continuous probability, as we can see in problem 3, page 48.

We had to distinguish the probability of one point from the probability of another point which is the *image* of the first one, and from this the concept of *probability law* emerged, etc.

The order of presentation couldn't, therefore, conform to the classical one.

In this book, continuous probability is presented in parallel with discontinuous probability, and with the development it deserves. Bernoulli's theorems are a natural follow up, since they can be applied to both discontinuous and continuous probability. After Bernoulli's 3rd theorem, we present some variants and extensions, necessary to establish the error law with the rigorous demonstration its usual presentation lacks. We next develop the theory of *Mathematical Expectation*, since the importance of this concept is more evident with the application of Bernoulli's 3rd theorem than with the definition of mathematical expectation itself. Finally, we broaden the scope

of applications of Probability, dealing with phenomena of which we are not the agents. We also postpone until the end a classification of the phenomena that are the object of this science, since we believe that the classification is clear and rational after a deep understanding of how Probability deals with the standard phenomenon, discussed in the Introduction, and developed in the core of this thesis. I had in mind to finish with a justification of our concept of probability, and to add an Appendix developing the study of probability in denumerable sets; but the unusual extension of the present dissertation persuaded me to postpone the publication of these matters.

Introduction

The aim of Probability Calculus, as of any other science, is to find associations relating known facts to other facts that, although being unknown, can be related to the former ones.

We begin by an example, illustrative of what we consider our [degree of] knowledge of the facts.

Suppose that one urn contains balls, identical in all aspects save, eventually, in their color.

There are three possible situations:

1. we do not know the colors of the balls in the urn, and therefore we do not know the percentage of each color, as well;
2. we know the colors of the balls [for instance, there are white balls and there are black balls], but we ignore the percentage of the balls of each color;
3. we know the colors, and the percentage of balls of each color in the urn.

A ball will be randomly extracted from the urn, and we have to bet on the color of the ball.

What bet should we choose?

In the first case, the question doesn't make sense. As we do not know anything about the colors present in the urn, there is no reason whatsoever to prefer any color to bet in.

In the second case, our ignorance has been moderated, since we know that the ball that will be extracted can be either white or black.

But as we still ignore the percentage of balls of each color, there are no grounds to decide which bet to take.

On the other hand, in the third case, assuming for instance that we know that 90% of the balls are white, we would surely decide to bet that a random extraction would produce white ball.

Obviously, we do not know for sure the color of the ball that will be extracted, it can be black or white, but we do not hesitate in choosing white as the sensible bet.

This distinguishes the third case from the former ones. It can serve as an example on how to take rational decisions with incomplete information.

For this reason we shall say that the third case describes a *known* urn.

The third case deals with random extractions from one urn whose composition is qualitatively and quantitatively known.

We shall assume that any phenomenon whose outcomes can be identified with random extractions of balls from one urn of qualitatively and quantitatively known composition is explained once that identification has been made. More generally, we consider explained any phenomenon which can be identified with a random selection of elements in a finite set, qualitatively

and quantitatively known.

As we have seen, we have distinguished the third case as known, because it can serve as a *standard model* on how to take decisions under uncertainty, i.e., under circumstances that we synthesized in the form of taking a bet.

Let us analyze in more detail the reasons that led us, in that example, to bet in white color.

The first reason was, indeed, the fact that we knew that more white balls than black balls existed in the urn, or, as stated in the example, the percentage of white balls was larger than the percentage of black balls.

The second one was the knowledge that the extraction was performed at random.

If one of these assumptions is withdrawn, there is no rationale for choosing to bet “white ball”.

The reason why the first condition is an argument in favor of betting in white ball comes from Arithmetic; the explanation why the second condition is needed can be found only in the emerging science of Probability.

In that science we will therefore take the statement

“to extract an element, at random, from a finite set”

as a primitive concept in this branch of Mathematics.

We shall build Probability Calculus starting from this primitive concept.

It is worth observing that we didn’t choose the concept “randomness” or “uncertainty” as the primitive concept upon which the theory of Probability would be constructed, since these concepts are vague, and as such inadequate to serve as the foundation for any science; our choice has been quite different, the concept of “extracting an element, at random, from a finite set”.

Some could accuse us of using a foundation as vague as the concept of “randomness”, since this concept is used in our primitive statement.

However, in the statement we choose as primitive, it is immaterial whether the formulation “*at random*” is or isn’t vague, insofar as the proposition using it can be understood and expresses an idea that can guide our choices and decisions under precise circumstances.

Whatever we say about this proposition is irrelevant either from the mathematical viewpoint or in the perspective of applications.

The same could be said about the concepts of *space* in Geometry, or of *time* in Mechanics.

The discussion of these concepts is irrelevant in Mathematics, they are from the scope of Philosophy. Mathematics *would* be the same theoretical construct if these concepts didn’t exist. The knowledge of what we consider Geometry and Mechanics would be latent in the symbolism of Mathematical Analysis, and no more, but this knowledge would still be valid, although less visible.

The usefulness of the concepts “space” and “time” can be compared to the usefulness of coloring reagents in Chemistry: they enhance the visibility of the phenomena, but these exist independently of being or not being enhanced by the coloring reagent.

An important question must be raised at once: how can we distinguish between random and non-random extractions?

It is obvious that there are extractions that are non-random, and therefore we need a criterion to distinguish random from non-random extractions.

To construct such criterion, we shall assume that any individual knows whether an extraction has been made at random if *the extraction has been made by him*.

Under this assumption, we shall build up a theory of probability, which is a subjective science, as all pure science is. This theory will allow us to construct a criterion to distinguish between random and non-random extractions, when we are not, ourselves, the agent performing the extraction.

This area of Probability Calculus is at the onset of applications.

The usefulness of science is its general ability to forecast events with an approximation considered good enough in practical applications.

This pragmatism seems unfeasible in Probability Calculus.

In effect, how could we predict the color of the ball that will be extracted from one urn containing two white balls and one black ball?

It is obvious that Probability Calculus is unable to make an useful prediction, in this situation.

If instead of two white balls and one black ball, the urn composition was one thousand white balls and one black ball, prediction of the outcome of a random extraction would still be impossible, but to our intuition it would seem more plausible to forecast that a white ball would be extracted.

The practical usefulness of Probability Calculus lies in this evaluation of the degree of probability of a future event, and in the ensuing confidence that our intuition attaches to the plausibility of events whose probability approximates certitude.

Confidence based in probability will, in its essence, be different from certitude, no matter how nearly the percentage of white balls in the urn approximates 1. But this doesn't deface the real practical usefulness of Probability in decision making under incomplete or unreliable information.

What we have said about random extractions of elements from a finite set can also be said about randomly throwing *points* in a bounded region of space, in any number of dimensions.

2.1 CHAPTER I — Finite sets

We shall denote $A, B, \dots$ sets with a [finite] number of elements.

The symbol $A \times B$ will denote the set of ordered pairs (a, b) , obtained from the sets A and B , by associating each $a \in A$ with each $b \in B$.

The symbol $\#A$ denotes the cardinal of the set A .

With these notations, it is obvious that

$$\#A \times B = \#A \times \#B.$$

Each ordered pair $(a, b) \in A \times B$ is said to be *compound* of a and b .

The set $A \times B$ is *compound* from the sets A and B . A set composed of compound elements (a, b) is not, necessarily, a compound set.

Primitive concept

a)

We consider the statement *to extract, at random [or to select], an element from the set A* as having a self evident meaning, and henceforth needing no further explanation; in other words, *to select, at random*, an element from a finite set is considered a primitive concept.

b)

The statement *a is a randomly chosen element from the set A* has the same meaning; *b)* is better suited to the formal symbolism of mathematical logic, while *a)* is more appropriate for the natural language.

From the above assertions, the propositions “randomly extracting a card from a card deck”, “randomly throwing a die” (randomly selecting of one die

face), “randomly extracting a ball from an urn”, etc., do not need further explanation.

DEFINITION 1

Randomly extracting an element from A , or B , or $C, \dots$, is the same as randomly taking an element from $A \cup B \cup C \cup \dots$, the set having all the elements from the sets $A, B, C, \dots$

DEFINITION 2

a)

Randomly extracting an element from A and, [independently,] another from B is, by definition, the same as randomly extracting an element from $A \times B$.

b)

Randomly extracting an element from A , another from B and another from C , [the extractions being mutually independent] is, by definition, the same as randomly extracting an element from $A \times B$ and another from C , etc.

According to this definition, *randomly choosing a suit and then randomly choosing a number⁽²⁾, [independently,] is the same as randomly choosing a card from the card deck.*

⁽²⁾ In this context the *numbers* are 1 or ace, 2, 3, 4, 5, 6, 7, 8, 9, 10, knave, queen, king, i.e. the card value, whichever the suit.

DEFINITION 3

a)

Let us associate to each $a \in A$ a set B_a , and denote $\{a\} \times B_a$ the set of ordered pairs $\{(a, b) : b \in B_a\}$.

Randomly extracting an element from A and another element from the corresponding set B_a is, by definition, the same as randomly extracting an element (a, b) from $A \times B_a$.

b)

If to each $b \in B_a$ we associate a set C_b , *randomly extracting an element from A , another element from the corresponding set B_a and another element from the corresponding set C_b is, by definition, the same as randomly extracting an element (a, b, c) from $A \times B_a \times C_b$.*

Possibility**1 — Possible elements**

According to the above definitions, random extractions have a meaning either in a single set (primitive concept, definitions 1 and 2) or in a complex of sets [definition 3, a) and b)].

All depends on the extracting system, and on the sets from where the extractions are performed.

When the extractions are performed from a single set, or performed in such a way that they are equivalent to extractions from a single set (primitive concept, definitions 1 and 2), we say that all the elements from that set are *possible*.

When the extractions are sequentially performed from a complex of sets, as explained in definition 3 a), we say that the *possible* elements are those in $A \odot B = \bigcup_{a \in A} \{a\} \times B_a$.

On the other hand, in what concerns definition 3 b), the *possible* elements are those that can be sequentially extracted randomly choosing $a \in A$, and then randomly choosing one element $b \in B_a$, and next randomly choosing an element $c \in C_b$, i.e., the elements from the complex of sets $A \odot B \odot C = \left\{ (a, b, c) \in \bigcup_{(a,b) \in A \odot B} \{(a, b)\} \times C_b \right\}$, etc.

2 — Possible sets

The *total possible set* A [resp. B , $A \times B$, $A \odot B$, etc.] is the set with all possible elements.

Any $A' \subset A$ is a *possible set*, i.e. is a set whose elements are possible.

DEFINITION 4

The *possibility* of a randomly chosen element $a \in A$ (or in any of its possible subsets), or *unit possibility*, is

$$\pi_a = \frac{1}{\#A}.$$

Thus, all elements randomly chosen in the same set (or randomly chosen using an extracting system which is equivalent to random extraction from the same set) are equally possible.

Proposition I

The possibility of a compound element $(a, b) \in A \times B$ is the product of the possibilities of its components.

This is an obvious consequence of $\#A \times B = \#A \times \#B$:

$$\frac{1}{\#A \times B} = \frac{1}{\#A} \times \frac{1}{\#B},$$

and thus

$$\pi_{(a,b)} = \pi_a \times \pi_b.$$

□

DEFINITION 5

The possibility $\varpi_{A'}$ of a possible set A' is the sum of the possibilities of its elements,

$$\varpi_{A'} = \sum_{a \in A'} \pi_a.$$

Proposition II

If A' is a possible set which may be partitioned into pairwise disjoint sets

$$A' = A'_1 \cup A'_2 \cup \cdots \cup A'_n$$

then

$$\varpi_{A'} = \varpi_{A'_1} + \varpi_{A'_2} + \cdots + \varpi_{A'_n}.$$

This is an immediate consequence of Definition 5.

Proposition III

The possibility of the total possible set is 1.

a)

If all the possible elements result from random extractions performed in the same set A , the proposition is obvious, since

$$\varpi_A = \sum_{a \in A} \frac{1}{\#A} = \frac{\#A}{\#A} = 1.$$

b)

Let us consider now sequential extractions from a complex of sets. Without loss of generality, consider the extraction system in definition 3 a).

Let $A = \{a_1, a_2, \dots, a_n\}$, and denote B_{a_k} , $k = 1, 2, \dots, n$ the set associated with each element $a_k \in A$. From Proposition I, the possibility of any element resulting from pairing a_k with $b_j \in B_{a_k}$ is

$$\pi_{(a_k, b_j)} = \frac{1}{\#A \times \#B_{a_k}},$$

and therefore the possibility of the set $a_k \times B_{a_k}$ is

$$\sum_{b_j \in B_{a_k}} \pi_{(a_k, b_j)} = \frac{1}{\#A} \sum_{b_j \in B_{a_k}} \frac{1}{\#B_{a_k}} = \frac{1}{\#A}.$$

Thus, in view of Proposition II, the possibility of the total possible set is

$$\sum_{a \in A} \frac{1}{\#A} = 1.$$

□

The above proof is easily extended for any complex extracting system.

Proposition IV

If the set $A \times B$ is compound from the sets A and B , then

$$\varpi_{A \times B} = \varpi_A \times \varpi_B,$$

since the possibility of each element (a, b) is the product of the possibility of an element of A by the possibility of an element of B .

Probability

DEFINITION 6

Let A' be a possible set and $A'' \subset A'$ another possible set⁽³⁾. We shall call *probability of A'' relative to A'* the number

$$\mathbb{P}_{A'}(A'') = \frac{\varpi_{A''}}{\varpi_{A'}} ,$$

$\varpi_{A''}$ and $\varpi_{A'}$ denoting, as above, the possibilities of A'' and of A' , respectively.

In the above context, the set A'' is said to be the *favorable* set, and $A' - A''$ is said to be the *unfavorable* or *contrary* set.

Sometimes we shall use the word *case* meaning *element*.

If the elements in A' are equally possible, it follows that

$$\varpi_{A'} = \#A' \times \pi_a, \quad \varpi_{A''} = \#A'' \times \pi_a$$

and therefore

$$\mathbb{P}_{A'}(A'') = \frac{\#A''}{\#A'} .$$

In other words: When the elements in the possible set are equally possible, the probability is the number of favorable cases divided by the number of possible cases.

When the possible set A' is the total possible set A , from

$$\varpi_A = 1$$

it follows that

$$\mathbb{P}_A(A'') = \varpi_{A''} \text{ }^{(4)}.$$

⁽³⁾ We shall use, as a rule, $A'' \subset A' \subset A$.

⁽⁴⁾ The most general definition of probability that can be found in Laplace is coincident with this particular case, of the reference set being the total possibility set A , with $\varpi_A = 1$.

If the favorable set A'' is the possible class A' ,

$$\mathbb{P}_{A'}(A') = 1$$

and in this case probability is *certitude*.

If the favorable set is empty, $A'' = \emptyset$,

$$\varpi_{\emptyset} = 0$$

and therefore

$$\mathbb{P}_{A'}(\emptyset) = 0.$$

In this case, probability is renamed *impossibility*.

Therefore, probability takes values between 0 and 1.

Postulate

Let S and S' be two extracting systems, originating qualitatively equal elements. We say that those two systems are equivalent if qualitatively equal sets have the same probability under S and S' .

The term *equivalent* in the above postulate means that similar extractions performed under S and under S' imply similar decisions.

This postulate reduces all extracting systems to extractions from a single set.

Proposition V

Total probability

If the possible set A'' is partitioned pairwise disjoint partial sets $A''_1, A''_2, \dots, A''_n$,

$$A'' = A''_1 \cup A''_2 \cup \dots \cup A''_n,$$

we have (Prop. II)

$$\varpi_{A''} = \varpi_{A''_1} + \varpi_{A''_2} + \cdots + \varpi_{A''_n}$$

and henceforth

$$\mathbb{P}_{A'}(A'') = \mathbb{P}_{A'}(A''_1) + \mathbb{P}_{A'}(A''_2) + \cdots + \mathbb{P}_{A'}(A''_n)$$

i.e., *the probability of the union of pairwise disjoint sets is the sum of the probabilities of the partial sets.*

Proposition VI

Compound probability

a)

If $A' \times B'$ is a possible set compound from A' and B' , and $A'' \times B''$ is a possible subset of $A' \times B'$, we have (Prop. IV)

$$\varpi_{A' \times B'} = \varpi_{A'} \times \varpi_{B'}$$

and

$$\varpi_{A'' \times B''} = \varpi_{A''} \times \varpi_{B''}$$

and therefore

$$\mathbb{P}_{A' \times B'}(A'' \times B'') = \mathbb{P}_{A'}(A'') \times \mathbb{P}_{B'}(B'').$$

In case the sets A' and B' are independent, this proposition may be stated as: *the probability of a compound set is the product of the probabilities of its components.*

b)

Proposition VI has been proved under the hypothesis that both the favorable and the possible sets are compound. It can, however, be generalized in the following ways:

1st

If $A' \times B' = A \times B$ is the total possible class, and thus

$$\varpi_{A' \times B'} = 1$$

we have

$$\varpi_{A'} = \varpi_{B'} = \varpi_{A' \times B'} = 1$$

and from this it follows that

$$\mathbb{P}_{A' \times B'}(A'' \times B'') = \mathbb{P}_{A'}(A'') \times \mathbb{P}_{B'}(B'').$$

2nd

If

$$\varpi_{A' \times B'} = \varpi_{A'},$$

i.e., if the possible set is obtained from the total possible set by excluding some elements $a \in A$ together with all the elements from the corresponding sets B_a , from the fact that

$$\varpi_{B'} = 1$$

it follows that

$$\mathbb{P}_{A' \times B'}(A'' \times B'') = \mathbb{P}_{A'}(A'') \times \mathbb{P}_{B'}(B'').$$

Proposition VII

Let $A'' \subset A' \subset A$ be possible sets. As

$$\frac{\varpi_{A''}}{\varpi_A} = \frac{\varpi_{A'}}{\varpi_A} \times \frac{\varpi_{A''}}{\omega_{A'}}$$

it follows that

$$\mathbb{P}_A(A'') = \mathbb{P}_A(A') \times \mathbb{P}_{A'}(A''). \quad (1.1)$$

Corollary

From (1.1) it follows that

$$\mathbb{P}_{A'}(A'') = \frac{\mathbb{P}_A(A'')}{\mathbb{P}_A(A')}$$

Proposition VIII**On the probability of causes**

When the random extractions are performed as described in Definition 3, the set A is the set of causes, and the sets $B = \bigcup_{a \in A} B_a$ is the set of effects.

The problem of the probability of possible causes may be typified as follows:

Let us consider a set of N urns, n_1 of which have a fraction p_1 of white balls, n_2 of which have a fraction p_2 of white balls, etc.

Randomly choose one among the N urns, and from that urn randomly extract a ball; let's investigate the consequences of assuming that the extracted ball is white.

What is the probability that this ball has been extracted from an urn with percentage p_i of white balls?

The solution may be constructed as follows:

Under the hypothesis that the extracted ball is white, the elements of the possible set A are all the compound elements of the form

(any urn, white ball).

Denoting ϖ_A the possibility of this set, from Prop. II and IV we get that

$$\varpi_A = \frac{n_1}{N} p_1 + \frac{n_2}{N} p_2 + \cdots$$

or, denoting

$$\frac{n_k}{N} = \omega_k,$$

$$\varpi_A = \sum \omega_k p_k.$$

The elements of the favorable set A' are all the compound elements of the form

(urn with $p_i \times 100\%$ white balls, white ball);

and therefore (Prop. IV)

$$\varpi_{A'} = \omega_i p_i.$$

Thus (Def. 6)

$$\mathbb{P}_A(A') = \frac{\omega_i p_i}{\sum \omega_k p_k}, \quad (1.2)$$

an expression known as Bayes formula.

In the above expression, ω_i is the probability of extracting, among the N urns, one with percentage p_i of white balls, and it is known as *a priori probability* of the urns with $p_i \times 100\%$ white balls.

The probability (1.2), $P_i = \mathbb{P}_A(A') = \frac{\omega_i p_i}{\sum \omega_k p_k}$, is the probability of extracting, among the N urns, one with percentage p_i of white balls, after performing the first extraction, resulting in white ball; for that reason, it is known as *a posteriori probability* of the urns with $p_i \times 100\%$ white balls.

It is obvious that the *causes* we are investigating may arise in any random extraction system, and that we cannot limit ourselves with extractions in a single set.

We now generalize formula (1.2) for sequential extractions from a complex of sets:

Denote

$$\omega_1, \omega_2, \dots, \omega_n$$

the *a priori* probabilities of the n causes which may originate the extraction of white ball, and denote

$$p_1, p_2, \dots, p_n$$

the probabilities that each of these causes confers to the event extraction of white ball.

Let us denote A the set that we obtain by associating each of the causes with each of the balls whose extraction it can originate. On the other hand, let us denote A' the set that we obtain by associating each of the causes with each of the *white balls* whose extraction it can originate. And let us denote A''_k , $k = 1, 2, \dots, n$ the set that we obtain by associating each k -th cause with each of the *white balls* whose extraction each of these causes can originate.

From Prop. VI, *b*), 1st, we know that

$$\mathbb{P}_A(A''_i) = \omega_i p_i; \quad (1.3)$$

on the other hand (Prop. VII)

$$\mathbb{P}_A(A''_i) = \mathbb{P}_A(A') \times \mathbb{P}_{A'}(A''_i) \quad (1.4)$$

and (Prop. V)

$$\mathbb{P}_A(A') = \sum_{k=1}^n \omega_k p_k$$

since

$$A' = \bigcup_{k=1}^n A''_k.$$

Thus

$$\mathbb{P}_{A'}(A''_i) = \frac{\omega_i p_i}{\sum_{k=1}^n \omega_k p_k}.$$

The above proof clearly shows that the usual argumentation that appears in other probability books is erroneous. In fact, when the urns do not have the same number of balls, the usual demonstration uses the formulas (1.3) and (1.4), justifying their use with the compound probability principle. But Proposition VI cannot be reduced to Proposition VII, since Prop. VII cannot be applied to compound elements.

This error was not evident due to lack of clarification of the meaning of *compound event* [and of *complex event*].

In fact, without this error it would have been impossible to establish Bayes formula with the definition of probability adopted in those books, since Bayes formula refers to a situation unforeseen in their definition: unequal probability of elementary events in a *possible set* which is a proper subset of the *total probability set*.

Proposition IX

Inverse formulas to Bayes' formula

Denoting the *a posteriori* probability of the *i*-th cause P_i , we have established that

$$P_i = \frac{\omega_i p_i}{\sum \omega_k p_k}. \quad (1.5)$$

This formula is symmetrical in what concerns the use of the ω_k and of p_k , which are given⁽⁵⁾.

Let us now assume that the p_k and of P_k are given, and that our aim is to compute the *a priori* probabilities ω_i . We now prove that

$$\omega_i = \frac{\frac{P_i}{p_i}}{\sum \frac{P_k}{p_k}}.$$

In fact, from (1.5) we get that

$$\frac{P_i}{p_i} = \frac{\omega_i}{\sum \omega_k p_k};$$

therefore,

$$\sum \frac{P_i}{p_i} = \frac{\sum \omega_i}{\sum \omega_k p_k} = \frac{1}{\sum \omega_k p_k}$$

and thus $\omega_i = \frac{\frac{P_i}{p_i}}{\sum \frac{P_k}{p_k}}.$

□

Due to the symmetry of (1.5), we also have the inversion formula

$$p_i = \frac{\frac{P_i}{\omega_i}}{\sum \frac{P_k}{\omega_k}}. \quad (6)$$

⁽⁵⁾ *Editors' note:* this is not true: $\sum \omega_i = 1$, but $\sum p_i$ can be different from 1. For instance, in the classical Laplace's urn problem $\sum p_i = \sum_{k=0}^N \frac{k}{N} = \frac{N+1}{2}$.

⁽⁶⁾ *Editors' note:* this is not true, unless $\sum p_i = 1$. The usefulness of the correct expression $\frac{p_i}{\sum p_k} = \frac{\frac{P_i}{\omega_i}}{\sum \frac{P_k}{\omega_k}}$ seems rather limited. On the other hand, $\omega_i = \frac{\frac{P_i}{p_i}}{\sum \frac{P_k}{p_k}}$ is true.

Proposition X

Let us now solve the problem that follows, where we assume the conditions stated for the problem of the probability of causes.

“From a randomly chosen urn, extract one ball; this ball is white, and after observation it is returned to the urn. What is the probability that a second extraction from this urn will result in white ball?”

1st solution

We shall solve this problem directly using the definition of probability.

The possible set is the set of all compound events of the form

(any urn, white ball, any ball)

and therefore

$$\omega_A = \sum \omega_k p_k \cdot 1 = \sum \omega_k p_k.$$

The favorable set is the set of all compound events of the form

(any urn, white ball, white ball)

and thus

$$\omega_{A'} = \sum \omega_k p_k p_k = \sum \omega_k p_k^2.$$

From the above, we get

$$\mathbb{P}_A(A') = \frac{\sum \omega_k p_k^2}{\sum \omega_k p_k}.$$

2nd solution

We may alternatively solve the problem in the following way: the effect of the observation of white ball in the first extraction is to change the *a priori* probabilities ω_k by the *a posteriori* probabilities P_k formerly computed. This problem is therefore equivalent to the following one:

If n causes with probabilities

$$P_1, P_2, \dots, P_n$$

may result in a given effect with probabilities

$$p_1, p_2, \dots, p_n,$$

respectively, what is the probability of that effect?

The desired probability is (Prop. V)

$$\mathbb{P} = \frac{\sum \omega_k p_k^2}{\sum \omega_k p_k}.$$

More generally:

If we perform $m + n$ extractions from a randomly chosen urn (returning each extracted ball to the urn after observation, before proceeding to the next extraction), resulting in m white balls and n black balls, the probability of getting white ball in the $(m + n + 1)$ -th extraction is

$$\mathbb{P} = \frac{\sum \omega_k p_k^{m+1} q_k^n}{\sum \omega_k p_k^m q_k^n}. \quad [q = 1 - p]$$

This result may be established by any of the two methods used in solving the former problem, which was the particular case of two extractions.

Corollary

If the urns have the same *a priori* probability, i.e., if

$$\omega_k = \text{constant}$$

then

$$\mathbb{P} = \frac{\sum p_k^{m+1} q_k^n}{\sum p_k^m q_k^n}.$$

Problem

One urn contains N balls, either white or black, in unknown proportions. Assuming that all the possible proportions $(0, \frac{1}{N}, \dots, 1)$ of white balls are equiprobable, what is the probability of extracting white ball in the $(m + n + 1)$ -th extraction, if we know that the previous $m + n$ extractions resulted m times in white ball and n times in black ball?

The above problem is equivalent to the following one:

There are $N + 1$ urns, one of them with N black balls, another one with 1 white and $N - 1$ black balls, another one with 2 white and $N - 2$ black balls, etc., until the last urn, containing N white balls.

Performing $m + n$ extractions of one ball from a randomly chosen urn (always returning the extracted ball to the urn before proceeding to the next extraction), white ball is observed in m occasions, and black ball in n occasions. What is the probability of extracting white ball in the $(m + n + 1)$ -th extraction?

The solution is given in the corollary above, where we may use

$$p_k = \frac{k}{N} \quad \text{and} \quad q_k = \frac{N - k}{N},$$

obtaining

$$\mathbb{P} = \frac{\sum_{k=0}^N \left(\frac{k}{N}\right)^{m+1} \left(\frac{N-k}{N}\right)^n}{\sum_{k=0}^N \left(\frac{k}{N}\right)^m \left(\frac{N-k}{N}\right)^n},$$

which may be approximated by

$$\mathbb{P} \approx \frac{\int_0^N \left(\frac{\alpha}{N}\right)^{m+1} \left(\frac{N-\alpha}{N}\right)^n d\alpha}{\int_0^N \left(\frac{\alpha}{N}\right)^m \left(\frac{N-\alpha}{N}\right)^n d\alpha}.$$

Using the substitution

$$\alpha = Nx,$$

$$\mathbb{P} \approx \frac{\int_0^1 x^{m+1} (1-x)^n dx}{\int_0^1 x^m (1-x)^n dx}.$$

As

$$\int_0^1 x^m (1-x)^n dx = \frac{\Gamma(m+1) \Gamma(n+1)}{\Gamma(m+n+2)}$$

and, for natural n ,

$$\Gamma(n) = (n-1)!$$

it follows that

$$\mathbb{P} \approx \frac{\Gamma(m+2) \Gamma(n+1)}{\Gamma(m+n+3)} \frac{\Gamma(m+n+2)}{\Gamma(m+1) \Gamma(n+1)}$$

or

$$\mathbb{P} \approx \frac{m+1}{m+n+2},$$

where the closeness of the approximation improves with the increase of N .

2.2 CHAPTER II — Continuous Probability

If $A, B, \dots$ denote bounded regions in a space with any number of dimensions, $A \cup B \cup \dots$ denotes the region with all the points from $A, B, \dots$

If $A, B, \dots$ denote regions whatever, $A \times B$ will denote the set of ordered pairs (a, b) , obtained from the sets A and B , by associating each point $a \in A$ with each point $b \in B$.

In the above definitions the use of geometric terminology is merely metaphoric, the word *point* meaning no more than any n -uple of numbers.

It has been proved in Pangeometry that if A and B are regions and $\mu(A), \mu(B)$ the corresponding *measures*, then

$$\mu(A \times B) = \mu(A) \times \mu(B);$$

more precisely, Pangeometry has generalized the concept of *measure* in hyperspace in such a way that the relation

$$\mu(A \times B) = \mu(A) \times \mu(B);$$

is valid.

We shall say that $A \times B$ is a *compound region* from A and B , and its points (a, b) are referred to as *compound points* of a and b , similarly to the conventions we have adopted in the former chapter, dealing with discontinuous probability.

Primitive concept

a)

As in the case of probability of finite discontinuous sets, we consider as primitive the concept of *throwing* [or *selecting*, or *choosing*, or *extracting*] *a point, at random, in the bounded region A in any number of dimensions.*

b)

The statement *X is a point thrown, at random, in A* has the same meaning as *a), b)* being better suited to the formal symbolism of mathematical logic.

DEFINITION 1

Randomly extracting one point from A, or B, or C, ..., is the same as randomly choosing one point in the region $A \cup B \cup C \cup \dots$

DEFINITION 2

a)

Randomly throwing one point from A and another[, independently,] from B is, by definition, the same as randomly throwing one point from $A \times B$.

b)

Randomly throwing one point from A, another from B and another from C is, by definition, the same as randomly throwing one point from $A \times B$ and another from C, [independently,] etc.

Thus, randomly choosing one point X in the segment $\overline{ab}$ and one point Y in the segment $\overline{ac}$ is the same as randomly choosing one point Z in the parallelogram $[abcd]$ (Fig. 1).

Randomly choosing one point in one arc and one point in a non coplanar line segment is the same as randomly choosing a point in the cylindrical surface generated by them, etc.

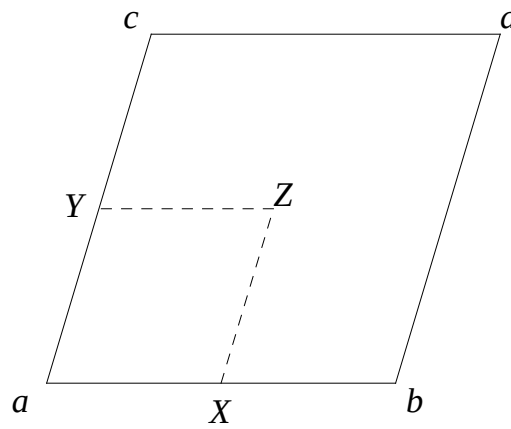


Figura 2.1: Figure 1

DEFINITION 3

a)

We now consider the case of constrained random selection, made in regions subject to some sort of mutual dependence.

Let us associate with each $a \in A$ a region B_a , and denote $\{a\} \times B_a$ the set of ordered pairs $\{(a, b) : b \in B_a\}$.

Randomly selecting (or throwing) one point in A and another point in the corresponding region B_a is, by definition, the same as randomly choosing one point (a, b) from $A \times B_a$.

b)

If to each $b \in B_a$ we associate a region C_b , *randomly throwing one point in A , another point in the corresponding region B_a and another point in the corresponding region C_b is, by definition, the same as randomly throwing one point (a, b, c) in $A \times B_a \times C_b$, etc.*

As an example, let A be (Fig. 2) the line segment $\overline{ab}$ on the $\overrightarrow{OX}$ axis

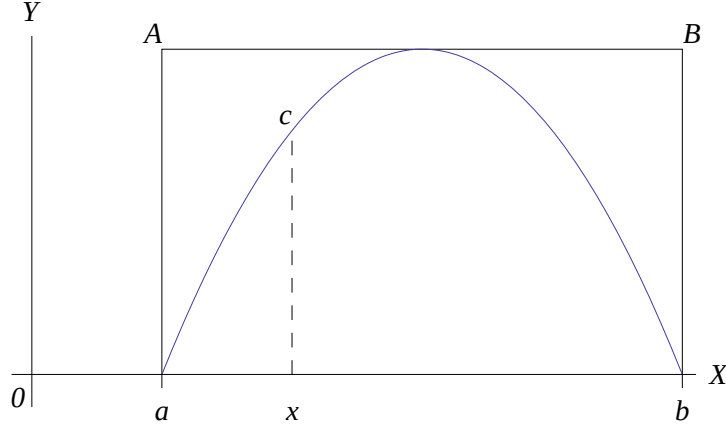


Figura 2.2: Figure 2

and that for each x in $\overline{ab}$ the corresponding B_x is the vertical segment with endpoints in the $\overrightarrow{OX}$ axis and on the curve $\widehat{acb}$.

If in the first point randomly thrown in $\overline{ab}$ results x , the second point will be randomly thrown in $\overline{xc}$ which is equivalent, according to the definition, of making only one random throw in the parallelogram with $\overline{ab}$ as base and the upper side passing from the point c , which is a subset of the parallelogram $[abAB]$.⁽⁷⁾

Possibility

According to the above definitions, we consider either randomly throwing one point in one region, single or compound (primitive concept. and Def. 1

⁽⁷⁾ *Editors' note:* In fact this is not true, and the observation where Pacheco d'Amorim says that all sampling schemes can be reduced to a single selection (or throw) are contradictory to the very detailed construction he builds to overpass the question of dependence; this is never explicitly stated, but it is evident that Pacheco d'Amorim tries to elegantly solve how to deal with joint probabilities. Observe also that his "reconstruction" of Fubinni's theorem in chapter IV clearly shows that this bold statement that hierarchical sampling can be reduced to single sampling cannot hold in dependent settings.

and 2) or constrained randomly throwing one point in a complex region (Def. 3 a and b)

In the first situation, we shall consider *possible* all the points in the region where the random throws are done. For instance, in the example illustrated in Fig. 1, all the points in the parallelogram $[abcd]$ are possible points.

In the case of Def. 3 a) we shall say that the possible points are those that result from associating each point from region A with each point from the corresponding region B_a , i.e., the complex of regions $A \odot B = \bigcup_{a \in A} \{a\} \times B_a$. For instance, in the example illustrated in Fig. 2, the possible points are those lying in the region limited by the segment line $\overline{ab}$ and the curve $\widehat{acb}$, etc.

Point possibility or unit possibility

Let $\mu(A)$ be the measure of the region A where we are throwing points at random.

The number

$$\pi_a = \frac{1}{\mu(A)}$$

will be called *possibility at point a* or *unit possibility*.

Similarly with what happens in the case of discontinuous probability, we may say that, in case all the random throwing of points is performed in the same [single or compound] region, all the sample points are *equally possible*, in the sense that the possibility is the same in each of those points.

According to this definition, all the points from regions as described in the primitive concept and in Def. 1 and 2 are equally possible. But in the cases addressed in Def. 3 the possibility will not be, in general, the same for all points.

However, the possibility is always well defined, since in all those definitions one point belongs to some uniquely defined region, where random throws are performed, its *possibility* resulting from the random throwing system adopted.

The possibility of each point is then a function of its coordinates.

Proposition I

Similarly to what happens in the case of probability in finite sets, *the possibility of a compound point $(a, b) \in A \times B$ is the product of the possibilities of its components*. The proof is in all points similar to the proof for the finite sets case.

Possibility of a region

One region is said to be possible if all its points are possible.

The possibility of a given possible region A' is the integral of the unit possibility over that region, in case this integral exists.

We shall denote $\varpi_{A'}$ the possibility of a possible region A' .

Proposition II

If A' is a possible region which may be partitioned into pairwise disjoint regions

$$A' = A'_1 \cup A'_2 \cup \cdots \cup A'_n$$

then

$$\varpi_{A'} = \varpi_{A'_1} + \varpi_{A'_2} + \cdots + \varpi_{A'_n}.$$

The possible regions A' are subsets of the *total possible region* A , the region of all possible points in the random throwing system considered.

Proposition III

The possibility of the total possible region is 1. (The proof has exactly the same steps detailed in the proof of the similar property in the case of finite sets.)

Proposition IV

If the region $A \times B$ is compound from the regions A and B , then

$$\varpi_{A \times B} = \varpi_A \times \varpi_B.$$

In fact, as the function $\frac{1}{\mu(B)}$ is independent of the coordinates of points from the region A , recalling that

$$\mu(A \times B) = \mu(A) \times \mu(B)$$

we have

$$\varpi_{A \times B} = \int_{A \times B} \frac{d(a, b)}{\mu(A \times B)} = \int_A \frac{d(a)}{\mu(A)} \times \int_B \frac{d(b)}{\mu(B)} = \omega_A \times \omega_B.$$

□

Probability

Let A' be a possible region in what concerns a given random throwing system, and $A'' \subset A'$ another possible region.

We shall call *probability of the region A'' relative to the region A'* the number

$$\mathbb{P}_{A'}(A'') = \frac{\varpi_{A''}}{\varpi_{A'}},$$

$\varpi_{A''}$ and $\varpi_{A'}$ denoting, as above, the possibilities of the regions A'' and of A' , respectively.

If the elements in A' are equally possible, it follows that

$$\mathbb{P}_{A'}(A'') = \frac{\mu(A'')}{\mu(A')}.$$

When the possible set A' is the total possible set A , from

$$\varpi_A = 1$$

it follows that

$$\mathbb{P}_A(A'') = \varpi_{A''}$$

Examples:

1st

A line segment is randomly broken into three parts. What is the probability that the three resulting segments can be taken for sides of a triangle?

To break a segment into three parts, randomly, is the same as to throw randomly two points X and Y on it. By Def. 2, this is the same as to throw, randomly, one point in the square having the segment as one of its sides.



Figura 2.3: Figure 3

Let $\overline{ab}$ be the segment (Fig. 3) and $[abcd]$ be the associated square (Fig. 4).

Let (X, Y) be the coordinates of the point Z corresponding to the positions X and Y of the two points randomly marked in the segment $\overline{ab}$.

Assuming that the segments mentioned in the problem are additive, we shall need to determine the probability that the segments $\overline{aX}$, $\overline{XY}$ and $\overline{Yb}$

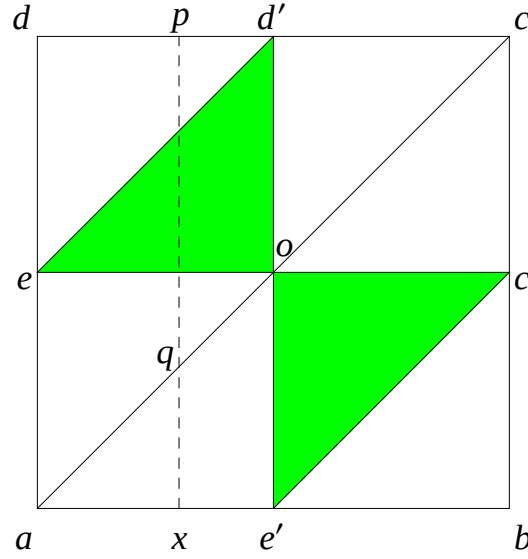


Figura 2.4: Figure 4

can be taken as the sides of a triangle (case $X < Y$); or else, that the segments $\overline{aY}$, $\overline{YX}$ and $\overline{Xb}$ can be taken as the sides of a triangle (case $X > Y$).

Let us analyze first the case $X < Y$.

Denoting α the length of the segment $\overline{ab}$ [and assuming $a = 0$, for simplicity], the three segments can be the sides of a triangle if and only if

$$0 < X < (Y - X) + (\alpha - Y),$$

$$0 < Y - X < X + (\alpha - Y),$$

$$0 < \alpha - Y < X + (Y - X).$$

These conditions are equivalent to

$$\begin{aligned} 0 < X &< \frac{\alpha}{2}, \\ 0 < Y - X &< \frac{\alpha}{2}, \\ \frac{\alpha}{2} < Y &< \alpha. \end{aligned} \tag{2.1}$$

The totality of points from the square $[abcd]$ whose coordinates verify condition (2.1) is the favorable region.

From the analysis of Fig. 4, it is obvious that this region is $[od'e]$.

In the case $X > Y$, a similar analysis shows that the favorable region is the triangle $[oc'e']$, symmetrical to $[od'e]$ in reference to the line $\overline{ac}$.

As in this randomly throwing system all points are equally possible, the probability is given by the quotient of the area of the favorable region by the area of the possible region, i.e.

$$\mathbb{P} = \frac{1}{4}.$$

2nd

Let us now assume that the segment is randomly broken into two segments, and then that the *bigger* subsegment is randomly broken into two. What is the probability that the three resulting segments can be the sides of a triangle?

The favorable region is obviously the same that we have constructed in the previous problem; let us now find the possible region.

When $X < \frac{\alpha}{2}$, we shall have $\overline{aX} < \overline{Xb}$ and therefore $X < Y < \alpha$, i.e., conditionally on the first point being $X < \frac{\alpha}{2}$ the second point is in $\overline{Xb}$, and thus (cf. Fig. 4) all the points in $B_x = \overline{qp}$ are possible. Hence, all the points in region $[aod'd]$ are possible.

For identical reasons, in the case $\frac{\alpha}{2} < X < \alpha$ the favorable region is $[oc'e']$ and the possible region is $[e'ocb]$; as all is symmetrical in reference to $\overline{ac}$, we shall make our computation for the case $X < \frac{\alpha}{2}$, the other one having the same numerical solution.

In the previous problem, all the possible points were equally possible, since each of the two points was randomly thrown into the segment $\overline{ab}$,

without any restriction. In the present problem, this is not so⁽⁸⁾. Denoting F the favorable region and P the possible region, we have

$$\varpi_P = 1$$

and

$$\varpi_F = \iint_{[od'e]} \frac{2 \, dx}{\alpha} \cdot \frac{dy}{\alpha - x} = \frac{2}{\alpha} \int_0^{\frac{\alpha}{2}} \left(\int_{\frac{\alpha}{2}}^{x+\frac{\alpha}{2}} dy \right) \frac{dx}{\alpha - x},$$

and so

$$\begin{aligned} \varpi_F &= \frac{2}{\alpha} \int_0^{\frac{\alpha}{2}} \frac{x}{\alpha - x} \, dx = \\ &= \frac{2}{\alpha} \int_0^{\frac{\alpha}{2}} \left[-1 + \frac{\alpha}{\alpha - x} \right] \, dx = -1 + 2 \left(-\log(\alpha - x) \right)_0^{\frac{\alpha}{2}} = \\ &= 2 \log 2 - 1 \approx 0.386 \end{aligned}$$

and therefore the probability we wanted to compute is

$$\mathbb{P} \approx 0.386$$

i.e., we get in this problem a bigger probability, as it should be expected by the extra conditions, which have increase the possibility of the three segments forming a triangle⁽⁹⁾.

3rd

⁽⁸⁾ *Editors' note:* Observe that to any $X \in (0, \frac{\alpha}{2})$ we associate $B_x = (X, \alpha)$, and hence the possibility of any $(X, Y) \in [aod'd]$ is $\frac{2}{\alpha} \times \frac{1}{\alpha - X}$.

⁽⁹⁾ *Editors' note:* We have corrected the final result given by the author who presents 0.44.

Now we consider a follow up of the above problem.

One point X is randomly thrown in $\overline{ae'}$ (Fig. 4), and another point X' is randomly thrown in the segment $\overline{Xb}$, and we further assume the condition that $X' \in \overline{e'b}$; what is the probability that the three segments $\overline{aX}$, $\overline{XX'}$ and $\overline{X'b}$ can be the sides of a triangle?

The *total possible region*, the possible region and the favorable region are $[aod'd]$, $[eod'd]$ and $[eod']$, respectively.

Therefore

$$\begin{aligned}\varpi_P &= \iint_{[eod'd]} \frac{2 \, dx}{\alpha} \cdot \frac{dy}{\alpha - x} = \\ &= \int_0^{\frac{\alpha}{2}} \frac{dx}{\alpha - x}\end{aligned}$$

or

$$\varpi_P = (-\log(\alpha - x))\Big|_0^{\frac{\alpha}{2}} =$$

$$= \log \alpha - \log \frac{\alpha}{2} =$$

$$= \log 2,$$

and (probl. 2nd)

$$\varpi_F = 2 \log 2 - 1;$$

therefore

$$\mathbb{P} = 2 - \frac{1}{\log 2} \approx 0,557.^{(10)}$$

⁽¹⁰⁾ *Editors' note:* We have detailed the final result given by the author who presents 0.6.

Observation

All that has been said about discontinuous probability, is also valid for continuous probability. Therefore Propositions V, VI and VII from Chapter I may be established for continuous probability using the same arguments that have been used in the case of discontinuous probability, and we take them as Propositions V, VI and VII in this Chapter II, without explicitly rewriting them. The “problem of the probabilities of causes” could be dealt with here as we did in Chapter I.

But we postpone the investigation of that problem to Chapter IV, using a different and more general methodology.

Proposition VIII

If a variable point X in some given region can be decomposed in two components X_1 and X_2 in such a way that whatever the position of X_1 the corresponding X_2 has always the same favorable and possible regions, X_1 and X_2 are *independent*, and henceforth their probabilities can be computed separately; this, as a rule, simplifies considerably the solution of problems. If, in particular, the probability of X_1 is 1, the probability of X is independent of the parameters defining the position of X_1 , which as a consequence we may assume *fixed*.

When the region where points are randomly thrown has a symmetry element, this proposition can in general be used.

Example:

Two points are randomly thrown on a spherical surface. What is the probability that the smaller arc of the maximum circle defined by the two points is smaller than α ?

Whatever the position of one of the points, say X , the favorable and possible regions for the other point Y are always the same.

In fact, given X , the favorable region is the spherical cap having vertex X and an angle 2α , and the possible region is the entire spherical surface. Therefore, the problem can be reformulated as follows: What is the probability that one point Y randomly thrown on a spherical surface lies in a given spherical cap? This problem has immediate solution.

Proposition IX

If the region of variation of X can be partitioned into subregions in such a way that the probability of X in each of them is always the same, the probability in the total region is the same as the probability in any of those regions.

In fact, let $A_1, A_2, \dots, A_L$ be the pairwise disjoint subregions of the possible region A , and $A'_1, A'_2, \dots, A'_L$ the corresponding favorable subregions of the favorable region A' ; from the hypothesis

$$\mathbb{P} = \frac{\varpi_{A'_1}}{\varpi_{A_1}} = \frac{\varpi_{A'_2}}{\varpi_{A_2}} = \dots = \frac{\varpi_{A'_L}}{\varpi_{A_L}}$$

we get that

$$\mathbb{P} = \frac{\sum_{k=1}^L \varpi_{A'_k}}{\sum_{k=1}^L \varpi_{A_k}},$$

which is the probability of X . □

DEFINITION 4

The proposition *randomly throw one point in the region A* , A being unbounded, has the same meaning as *randomly throw one point in the region A'* , where $A' \subset A$ is an arbitrarily large bounded region.

If X is a randomly thrown point in A and A is unbounded, the probability of X in region A is the limit of the probability of X in A' when A' increases indefinitely, i.e. the probability of X is the number $\mathbb{P}$ such that for all $\delta > 0$ there corresponds a region C such that

$$|\mathbb{P} - \mathbb{P}_B| < \delta$$

for all regions B such that

$$C \subset B.$$

This case will be dealt with in detail in an appendix, where we investigate the probability of denumerable sets⁽¹¹⁾.

2.3 CHAPTER III — Random Figures

All the propositions where the terms *randomly extracting* or *randomly throwing*, or equivalents are used in the context of the construction of random figures (either rigid or variable) will be defined through the use of the concepts of random extractions from a finite set or random throws of points in a continuous region, as indicated in definitions 1, 2 and 3 of Chapters I and II.

2.3.1 First Part — Random rigid figures

DEFINITION 1

The random choice of an orientation in a space of dimension n is, by definition, the same as randomly throwing a point $(x_1, \dots, x_n)$ in the set

⁽¹¹⁾ *Editors' note:* In the last sentence of the Preface, Pacheco d'Amorim says that he had conceived the intention of including an appendix on this subject, but that finally he has decided otherwise.

defined by the equation

$$(x_1 - x'_1)^2 + (x_2 - x'_2)^2 + \cdots + (x_n - x'_n)^2 = 1.$$

The orientation will be that of the vector having $(x'_1, \dots, x'_n)$ as origin and the randomly thrown point $(x_1, \dots, x_n)$ as extremity.

Any point in the space may be taken as origin $(x'_1, \dots, x'_n)$.

In the particular case $n = 2$, the random throw is done in a circumference, and in the case $n = 3$ the random throw is done in a spherical surface.

The definition of a random direction is done by analogy.

DEFINITION 2

Let $\overline{ab}$ and $\overline{a'b'}$ be two line segments. Superimpose the two segments in such a way that a and a' coincide, and then let the smaller one slide over the bigger one until points b and b' coincide. In other words, the smaller segment goes through all positions it may have upon the bigger one, and at the end of this procedure each point of the smaller segment will have defined segment trajectories of the same length.

Randomly throwing a smaller segment on a bigger one is the same as randomly throwing any given point of the smaller segment on the segment it defines when the smaller segment slides over the bigger one, as described above.

Randomly throwing the bigger segment on the smaller one is the same as randomly throwing the smaller segment on the bigger one.

The validity of the definition lies in the fact that all the segments defined by each point of the smaller segment when it slides over the bigger one are of equal length. Hence, it doesn't depend on a particular choice of the point, it has the same meaning whatever the point chosen in the segment.

As, in view of definitions 1, 2 and 3 of Chapters I and II, every random choice can be viewed either as a random choice in a single or compound region or as a choice in a complex region, the choice of a random figure corresponds, as a rule, to randomly choosing a point which determines the figure in this random choice. We shall call such a point the *equivalent point* to the figure, in what concerns the random choice at hand. The equivalent point must be independent of all points from the figure, whichever the figure in question.

Problem

Each of two friends goes for a half hour walk to a public garden open from 2 p.m. till 4 p.m., separately. What is the probability that in a given day they meet during their walk in the public garden?

We assume that the time each of them starts his walk is random. Then, as time is continuous in one dimension, the problem may be reformulated as follows:

We randomly throw two segments of lengths b and c , respectively, over a given segment of length a . What is the probability that the two random segments have a nonempty intersection?

If $b + c > a$, the two segments always overlap, and therefore

$$\mathbb{P} = 1.$$

Now we analyze the case $b + c < a$.

When the segment of length b slides over the segment of length a , each of its points describes a segment of length $a - b$ and, similarly, each point of the segment of length c generates segments of length $a - c$. Randomly throwing the two segments over the segment of length a is the same as randomly throwing one point of the segment of length b in $a - b$ and one point of the

segment of length c in $a - c$; and this is the same as randomly throwing one point in the rectangle $(a - b) \times (a - c)$.

Let's take as equivalent points of those segments, in what concerns the random throw described, their right extremities, and let x and y denote the distances of those right extremities to the origin 0 of the segment of length a (Fig. 5).

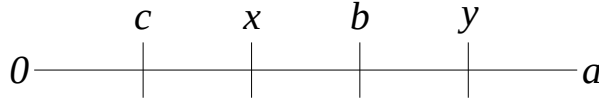


Figura 2.5: Figure 5

The two random segments don't overlap if and only if

$$y - x > b \quad \text{or} \quad x - y > c.$$

The lines with equations

$$y - x = b \quad \text{and} \quad x - y = c$$

determine on the rectangle $(a - b) \times (a - c)$ two half-squares which are, in what concerns this problem, the *contrary* region. It is then easy to compute

$$1 - \mathbb{P} = \frac{\frac{1}{2}(a - b - c)^2 + \frac{1}{2}(a - c - b)^2}{(a - b)(a - c)} = \frac{(a - b - c)^2}{(a - b)(a - c)}.$$

In the special case

$$a = 2, b = c = \frac{1}{2},$$

we get

$$\mathbb{P} = 1 - \frac{4}{9} = \frac{5}{9}.$$

DEFINITION 3

To throw a straight line at random in a given region A means, by definition, to throw a point, at random, in A , and to select at random one direction in the region A , which determine the straight line.

Example⁽¹²⁾:

A straight line is randomly thrown in a circle. What is the probability that its intersection with the circle is a chord smaller than a given chord of length c ?

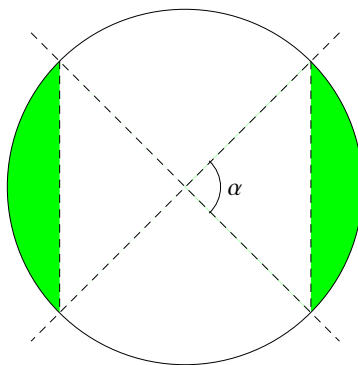


Figura 2.6: Figure 6

Without loss of generality, we shall solve the problem in the unit circle S , i.e. with area π .

Any chord of length c defines a smaller arc of amplitude α , say, and the area of the corresponding smaller circular segment is $\frac{\alpha - \sin \alpha}{2}$. Therefore (Prop. VIII of Chapter II) the probability wanted is independent of the direction D of the random straight line.

Whatever the direction of the random straight line, its intersection with the circle is a chord of length smaller than c if and only if the associated

⁽¹²⁾ *Editors' note:* We have corrected the misprints in the formulation and drawn a figure more suited to follow the arguments in the solution given by Pacheco d'Amorim.

smaller circular segment has area less than $\frac{\alpha - \sin \alpha}{2}$.

Therefore when throwing a random chord in the circle, for any randomly chosen direction, any point in the circle is a possible point so that the straight line is thrown in the circle (Def. 3), and the favorable points M are those in the [two] circular segments S'_1 [and S'_2], with area $\frac{\alpha - \sin \alpha}{2}$ each, defined by the straight line[s] with the given direction whose intersection with the circle is a chord of length c .

Thus, the probability in question is

$$\mathbb{P} = \frac{2 \cdot \text{area of the circular segment } S'}{\text{area of the circle } S} = \frac{\alpha - \sin \alpha}{\pi},$$

since the chord which is the intersection of the random straight line with the circle will have length less or equal to c if and only if the point M lies on a segment in a region $S' = S'_1 \cup S'_2$ with area $\alpha - \sin \alpha$.⁽¹³⁾

The solution has immediate generalization for the case of a straight line thrown in a sphere.

The same solution applies, with the necessary modifications, for a similar problem, where the straight line is replaced by a plane and the chord by the area of a plane section.

⁽¹³⁾ *Editors' note:* Although in the explanation Pacheco d'Amorim seems to overlook the fact that there are two circular segments, symmetrical in respect to the diameter with the chosen direction, whose points are favorable, the final expression he presents is correct. We have introduced the necessary corrections in his arguments.

Note however that, in our opinion, this ingenious solution he gives to the problem, of the class of the famous Bertrand's paradoxes, has a flaw. In fact, all the points that lie in a given chord with the given direction will correspond to the **same** randomly thrown straight line of the given direction, in Pacheco d'Amorim's definition, and it is obviously untrue that the two sets of points that lie in two chords of different lengths carry equal probability.

DEFINITION 4

To throw, at random, one straight line segment in a region A means, by definition, to throw a straight line at random in A (Def. 3) and to throw the segment in question (Def. 2) in the segment which is the intersection of A with the random straight line.

Example:

The needle problem

A needle (straight line segment) is randomly thrown over a sheet of paper (unlimited plane) where parallel and equidistant straight lines have been drawn. What is the probability that the needle intersects one of those straight lines?

Let l denote the needle length, and δ denote the distance between the parallel straight lines $\overline{ab}$, $\overline{a'b'}$, $\dots$ (Fig. 7).

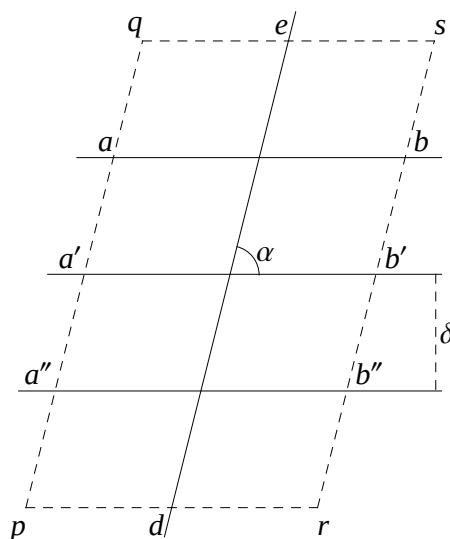


Figura 2.7: Figure 7

We first randomly select the direction α of the straight line that contains the segment; next we throw a random point X in a portion arbitrarily large

of the plane (Chapt. II, Def. 4). However big this portion of the plane, we may always define in it a parallelogram $[pqsr]$ whose sides are parallel to the direction selected, and whose bases are parallel to the straight lines $\overline{ab}$, $\overline{a'b'}$, containing that portion of the plane, and throw the point X inside it. Assume that these two random throws have determined the straight line $\overline{de}$. Once this random straight line has been thrown, the next step is to throw the segment of length l over the segment $\overline{de}$ (Def. 4). Or, whatever the position of point X , the probability that the segment of length l intersects one of the parallel straight lines is always the same. Therefore (Chapt. II, Prop. 8) we may take X as fixed.

Let the segment of length l slides over the straight line $\overline{de}$; when its origin goes from one parallel to the next one, the segment it generates while the segment of length l intersects the next parallel straight line has always the same length.

Thus the parallel straight lines divide the possible region where the segment of length l is randomly thrown over $\overline{de}$ in subregions with equal probability (with the exception of the first and of the last ones, which may be discarded, in view of the arbitrary size of the parallelogram). Therefore, in view of Prop. IX of Chapt. II, it is enough to compute the probability in one of them:

$$\mathbb{P} = \frac{l}{\overline{bb'}} = \frac{l \sin \alpha}{\delta},$$

(where α is the angle $\langle[edr]\rangle$, in case $\delta \geq l$; on the other hand, the unit possibility of α is $\frac{1}{\pi}$; therefore

$$\mathbb{P} = \int_0^\pi \frac{1}{\pi} \cdot \frac{l \sin \alpha}{\delta} d\alpha = \frac{l}{\delta \pi} (-\cos \alpha)_0^\pi = \frac{2l}{\delta \pi}.$$

In the case $\delta \leq l$, let us partition the field of variation of α in two parts: the first one the totality of values of α for which the intersection is void; the

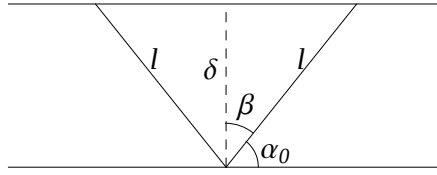


Figure 2.8: Figure 8

second one with all the other values. Using the theorems of total probability and of compound probability, and writing

$$\delta = l \cos \beta = l \sin \alpha_0,$$

we get

$$\begin{aligned} P &= \frac{\pi - 2\alpha_0}{\pi} \cdot 1 + 2 \frac{\alpha_0}{\pi} \int_0^{\alpha_0} \frac{l \sin \alpha}{\delta} \cdot \frac{d\alpha}{\alpha_0} = \\ &= \frac{\pi - 2\alpha_0}{\pi} + \frac{2l}{\delta \pi} [1 - \cos \alpha_0] = \\ &= \frac{2\beta}{\pi} + \frac{2l}{\delta \pi} (1 - \sin \beta). \end{aligned}$$

Note

It is worth noting that when $l < \delta$ the probability that the needle intersects one of the parallel straight lines is directly proportional to the length of the needle.

DEFINITION 5

To throw a plane⁽¹⁴⁾ at random in a plane region A means, by definition, to choose randomly an orientation in A.

⁽¹⁴⁾ The orientation of this plane is determined by a half-line.

DEFINITION 6

To throw, at random, a plane region in another plane region A means, by definition, to throw a random plane in A (Def. 5) and then to throw, at random, a point of the mobile region (or a point invariably tied to that region) in the region that this point defines when the mobile plane region occupies inside A all the positions that are compatible with the orientation randomly chosen in the first step.

It is obvious that an equivalent point will always have the same possibility, whichever the randomly chosen point in the plane of the mobile figure, since all the points of the mobile figure describe identical regions when the plane containing it moves taking on positions which are parallel to each other.

DEFINITION 7

To throw a plane at random in an n -dimensional space, $n \geq 3$ means, by definition, to choose a random point and two directions emerging from it in that space.

DEFINITION 8

To throw randomly a plane figure in an n -dimensional region means, by definition, to throw a random plane in that region and then to throw at random that figure in the portion of the random plane inside that region.

DEFINITION 9

To throw at random a three dimensional space in a three dimensional region A means, by definition, to throw a point at random in A and to choose two random directions⁽¹⁵⁾ emerging from it.

⁽¹⁵⁾ Two directions is sufficient to direct three axes.

The above definitions are easily extended for higher dimensional spaces.

Note

It is worth observing that, according to the above definitions, to throw at random a finite region B on another finite region A is equivalent to randomly throwing a point in a region that depends both from A and from B . From that dependence we may immediately conclude (it is enough to analyze one particular case, for instance throwing a straight line segment on a rectangular plane region) that from the random throw of all B we cannot conclude the random throw of any of its parts, because in the case of finite regions, the equivalent point of a part of B will vary in a different region of the equivalent point of all B , that is, if we throw together all region B the field of variation of its equivalent point (field of the equivalent point of the all) will be different from its field of variation if we throw the parts of B separately.

In the special case of random throws in an unbounded region, it may happen that the global random throw determines the partial random throws.

For instance, in the [Buffon's] needle problem (Def. 4), the random throw of the needle determines the random throw of any of its parts, since we get the same result conceptualizing the random throw of part of the needle either in isolation or as part of the needle, because its equivalent point would have the same field of variation in both cases.

2.3.2 Second Part — Random variable figures

DEFINITION 10

Randomly throwing a variable figure in a given region A is, by definition, to select, at random, the form of the figure and, then, to throw it at random in A , as if it were a rigid figure.

The discussion of the second statement in this definition has been done in the first part of the present Chapter. So, our present task is to discuss the meaning of randomly choosing the form of a variable figure.

Obviously this question cannot have an exhaustive treatment. We shall limit ourselves to articulated polygonal figures (either open or closed) and, as a limit case, of flexible inextensible curves.

DEFINITION 11

Open polygonal lines

An articulated polygonal figure is a polygonal line whose consecutive segments form variable angles.

To choose, at random, its form is, by definition, to choose, randomly, the form of each of its vertices.

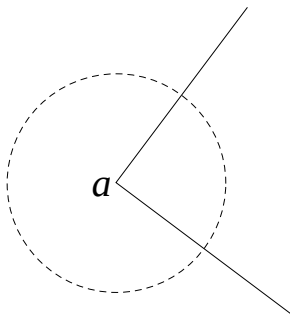


Figura 2.9: Figure 9

To explain the meaning of this last statement, let us consider one vertex or articulation a (Fig. 9), which we may assume to be an element of the n -dimensional space, with coordinates $(x'_1, x'_2, \dots, x'_n)$. Let us consider the hyperspherical surface defined by

$$(x_1 - x'_1)^2 + \dots + (x_n - x'_n)^2 = 1. \quad (3.1)$$

We assume that one of the sides of the articulation a is fixed, and that the other one can occupy any of the possible positions; therefore, at distance 1 from the vertex, it intersects the hyperspherical surface (3.1). Denote B the set of such intersection points ⁽¹⁶⁾. Thus, *to choose at random the form of the articulation a* is, by definition, *to choose a random point in the subset B of the hyperspherical surface*.

DEFINITION 12

Closed polygonal lines

a)

In the plane

We begin with plane figures, and next we shall discuss articulated figures in the higher dimensional spaces.

1st

Let's discuss, to start with, how to choose randomly the form of a four sided plane polygon $[abcd]$ (Fig. 10). When this polygon assumes all possible forms, its angle a can take values of one of two kinds: those corresponding to the position c of its non-adjacent vertex, and those corresponding to the position c' of its non-adjacent vertex. In other words, drawing a circumference

⁽¹⁶⁾ It is possible that B does not correspond to the total hyperspherical surface.

Editors' note: This is a mysterious footnote. How can the region B being constrained when we are choosing, at random, the form of one vertex of an open polygonal line? Perhaps this footnote is in the wrong place and it's relate about the random choose of the form of a closed polygonal line where it makes sense.

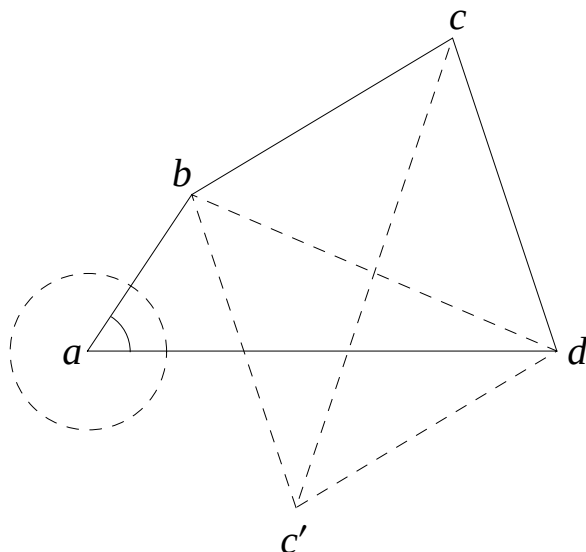


Figura 2.10: Figure 10

with unit radius with center a , and considering the side $\overline{ad}$ fixed, the intersection point of the moving side with the circumference defines two regions, that can have non-empty intersection ⁽¹⁷⁾ when the polygon assumes all its possible forms. But we shall, in all cases, consider the two regions, which we denote A and A_1 , as distinct. Once one point from one of those regions is given, the form of the figure has been determined. Consider similar regions with vertices b , c and d , and denote them B and B_1 , C and C_1 , D and D_1 , respectively.

Randomly choosing the form of the plane four sided polygon is, by definition, to choose randomly one point (Chapter II, Def. 1) from A or A_1 , or from B or B_1 , or from C or C_1 , or from D or D_1 .

With this definition, the equivalent point doesn't depend on any element of the figure.

⁽¹⁷⁾ The articulations can be subjected to restrictions such that the position c' is inadmissible.

Let us now consider a pentagon, and investigate how to progress from the previous case to the random choice of a pentagon.

These forms correspond to groups B and B' , C and C' , D and D' , E and E' , for the angles of $[bcde]$, and similar groups, but in general different ones, for the angles of $[bc'd'e]$. Let's associate the possible values of a with each of the elements of the groups $[bcde]$, and denote A the set thus obtained. Let's

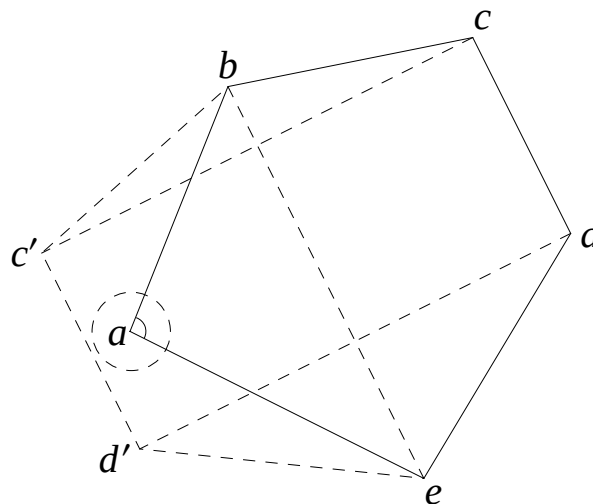


Figura 2.11: Figure 11

Randomly choosing a pentagon is, by definition, to choose, randomly, an element from A or A_1 , or from B or B_1 , ...

The definition of random choice of a hexagon, or of a heptagon, etc., is similar in all points.

b)

In the space

Randomly throwing a closed polygonal line in an n -dimensional space is in all points similar to what we have seen about randomly throwing a closed polygonal line in the plane, we only need to substitute, in the preceding definitions, the points varying in circumferences by points varying in hyperspherical surfaces.

Randomly throwing flexible inextensible curves

By definition, *randomly throwing a flexible inextensible curve*, open or closed *in the space A* is to throw in that space, at random, a polygon with the same length and an arbitrary large number of sides.

Any problem referring to an articulated polygon with an arbitrary number of sides will have a solution which depends on the number and length of those sides. If that solution converges for some limit when the supremum length of the polygons sides decreases to zero, we shall say that this limit is the solution of the same problem in the case of a flexible and inextensible curve.

2.4 CHAPTER IV — Image Point

Proposition I

Let A and B be two regions such that it is possible to define a bijective, complete and continuous correspondence between their points. When a point

M is randomly chosen in A , the corresponding point N in B is said to be the *image point* of M . Generally, each element in B is the *image* of a corresponding randomly chosen element in A .

The image point N is random, insofar as it depends on the original point M randomly chosen in A . But the random status of M in B is clearly different from the random status of M in A , since its random choice in A has been *direct*, while N randomly varies in B , but as an *image* of M .

For that reason we say that M is a *free point*, while N is an image or *dependent* point.

The possibility of a dependent point is, by definition, the possibility of the corresponding free point. The possibility of a region B' which is the image of a region A' is the possibility of A' . In a general way, all that can be said, in what concerns probability, about an image point N , varying in B , is defined *via* the corresponding free point varying in A . It is easily seen that the properties established for the free point [in the previous chapters] are also valid for the image point.

2.4.1 Application

Let f be a continuous increasing function defined in an interval $[\alpha, \beta]$. A number X is randomly chosen in that interval, and we want to know the probability that the corresponding Y has the digit d as its a -th decimal.

Let ω and ω' be the integers which most closely satisfy the inequalities

$$f(\alpha) < \frac{10\omega + d}{10^a} \quad f(\beta) > \frac{10\omega' + d + 1}{10^a},$$

and let us represent f^{-1} the inverse function of f ; the values $f(x)$ which verify the conditions in our problem lie in the intervals

$$\left(\frac{10\omega + d}{10^a}, \frac{10\omega + d + 1}{10^a} \right),$$

$$\begin{aligned} & \left(\frac{10(\omega + 1) + d}{10^a}, \frac{10(\omega + 1) + d + 1}{10^a} \right), \dots \\ & \left(\frac{10(\omega + i) + d}{10^a}, \frac{10(\omega + i) + d + 1}{10^a} \right), \dots \\ & \left(\frac{10\omega' + d}{10^a}, \frac{10\omega' + d + 1}{10^a} \right). \end{aligned}$$

Taking into account the foregoing definitions and the theorem of total probability, the probability we wish to compute is therefore

$$\mathbb{P}_{(\alpha, \beta)}(d, a) = \frac{\sum \left\{ f^{-1} \left(\frac{10(\omega + i) + d + 1}{10^a} \right) - f^{-1} \left(\frac{10(\omega + i) + d}{10^a} \right) \right\}}{\beta - \alpha},$$

1st

Let us apply the above formula to the function

$$y = \log_{\alpha} x$$

in the variation interval for y

$$\left(0, \frac{10\omega + 10}{10^a} \right).$$

In this case,

$$\begin{aligned} \omega = 0, \quad \omega' = \omega, \quad \text{and} \quad \mathbb{P}_{(\alpha, \beta)}(d, a) &= \frac{\sum_{i=0}^{\omega} \left[\alpha^{\frac{10i+d+1}{10^a}} - \alpha^{\frac{10i+d}{10^a}} \right]}{\alpha^{\frac{10\omega+10}{10^a}} - 1} = \\ &= \alpha^{\frac{d}{10^a}} \frac{\left[\alpha^{\frac{1}{10^a}} - 1 \right] \sum_{i=0}^{\omega} \alpha^{\frac{i}{10^a-1}}}{\alpha^{\frac{\omega+1}{10^a-1}} - 1} = \alpha^{\frac{d}{10^a}} \frac{\alpha^{\frac{1}{10^a}} - 1}{\alpha^{\frac{1}{10^a-1}} - 1}. \end{aligned} \quad (4.1)$$

Expression (4.1) shows that this probability doesn't depend on ω , and thus doesn't depend on the interval of variation of y ; therefore, it can be represented $\mathbb{P}(d, a)$.

From (4.1) it is clear that

$$\frac{\mathbb{P}(d+1, a)}{\mathbb{P}(d, a)} = \alpha^{\frac{1}{10^a}},$$

independent of d .

It is easily seen that $\mathbb{P}(d, a)$ goes quickly to $\frac{1}{10}$ when a increases.

2nd

Let us now consider the function

$$y = \alpha^x, \quad \alpha > 1$$

in the interval corresponding to (ω, ω') ; we get

$$\begin{aligned} \mathbb{P}_{(\omega, \omega')}(d, a) &= \frac{\sum_{n=\omega}^{\omega'} \left(\log_{\alpha} \frac{10n+d+1}{10^a} - \log_{\alpha} \frac{10n+d}{10^a} \right)}{\log_{\alpha} \frac{10\omega'+10}{10^a} - \log_{\alpha} \frac{10\omega}{10^a}} = \\ &= \frac{\sum_{n=\omega}^{\omega'} \log_{\alpha} \left(1 + \frac{1}{10n+d} \right)}{\log_{\alpha} \frac{\omega'+1}{\omega}}, \end{aligned}$$

showing that this probability is independent of a .

It is easily seen that

$$\lim_{\omega' \rightarrow \infty} \mathbb{P}_{(\omega, \omega')}(d, a) = \frac{1}{10}.$$

In order to establish this result, we shall first show that the sequence with general term

$$U_n(d) = \frac{\sum_{k=\omega}^n \log_{\alpha} \left(1 + \frac{1}{10k+d} \right)}{\sum_{k=\omega}^n \log_{\alpha} \left(1 + \frac{1}{k} \right)}$$

is increasing; then, we shall prove that it is upper bounded by 1.

a)

The sequence $U_n(d)$ is increasing when $d > 9$ (in the expression of $U_n(d)$, d may be any [integer] number).

In effect

$$f(k) = \frac{\log_{\alpha} \left(1 + \frac{1}{10k+d} \right)}{\log_{\alpha} \left(1 + \frac{1}{k} \right)}$$

is an increasing function, since its derivative

$$f'(k) = \frac{-\frac{10}{(10k+d)(10k+d+1)} \log_{\alpha} \left(1 + \frac{1}{k} \right) + \frac{1}{k(k+1)} \log_{\alpha} \left(1 + \frac{1}{10k+d} \right)}{\log_{\alpha} \left[\log_{\alpha} \left(1 + \frac{1}{k} \right) \right]^2}$$

is such that $f'(k) > 0$. In effect, as

$$\left(1 + \frac{1}{x} \right)^x$$

is an increasing function, we get

$$\left(1 + \frac{1}{10k+d} \right)^{10k+d} > \left(1 + \frac{1}{k} \right)^k$$

and, on the other hand, if $d \geq 9$,

$$\left[\left(1 + \frac{1}{10k+d} \right)^{10k+d} \right]^{10k+d+1} > \left[\left(1 + \frac{1}{k} \right)^k \right]^{10k+10};$$

taking basis α logarithms on both sides of the above inequality, it is immediate that

$$f'(k) > 0.$$

As the general term of the sequence U_n is a fraction whose numerator is the sum of the numerators of

$$f(1), f(2), \dots, f(n)$$

and whose denominator is the sum of their denominators, and as $f(k)$ is increasing, $U_n(d)$ is also increasing.

b)

$$U_n(d) < 1$$

for any n .

In effect

$$\log_{\alpha} \left(1 + \frac{1}{10k+d} \right) < \log_{\alpha} \left(1 + \frac{1}{k} \right)$$

and therefore

$$\sum \log_{\alpha} \left(1 + \frac{1}{10k+d} \right) < \sum \log_{\alpha} \left(1 + \frac{1}{k} \right)$$

from which we get

$$U_n(d) < 1.$$

As a consequence, we may state that $U_n(d)$ has a limit when $n \rightarrow \infty$ and $d \geq 9$.

From the expression of $U_n(d)$ we have that

$$0 \leq d \leq 10 \implies U_n(0) \geq U_n(d) \geq U_n(10);$$

and therefore, for n large enough,

$$U_n(d) - U_n(10) \leq U_n(0) - U_n(10) =$$

$$\begin{aligned}
&= \frac{\sum_{k=\omega}^n \left[\log_{\alpha} \left(1 + \frac{1}{10k} \right) - \log_{\alpha} \left(1 + \frac{1}{10(k+1)} \right) \right]}{\log_{\alpha}(n+1) - \log_{\alpha} \omega} = \\
&= \frac{\log_{\alpha} \left(1 + \frac{1}{10\omega} \right) - \log_{\alpha} \left(1 + \frac{1}{10(n+1)} \right)}{\log_{\alpha}(n+1) - \log_{\alpha} \omega} < \\
&< \frac{\log_{\alpha} \left(1 + \frac{1}{10\omega} \right)}{\log_{\alpha}(n+1)} < \delta,
\end{aligned}$$

for any $\delta > 0$. Thus the limit

$$\lim_{n \rightarrow \infty} U_n(d)$$

exists for any d , and that limit doesn't depend on d ; but as

$$\lim_{n \rightarrow \infty} \sum_{d=0}^9 U_n(d) = 1 = \sum_{d=0}^9 \lim_{n \rightarrow \infty} U_n(d)$$

it follows that

$$\lim_{n \rightarrow \infty} U_n(d) = \frac{1}{10}.$$

It is easily seen that U_n converges towards $\frac{1}{10}$ very quickly.

Observation

The general problem that we have just solved gives the distribution of the digits in an ideal table containing all values of a regular function in an interval (α, β) . In any real table with independent values in arithmetic progression, equal subsets of (α, β) contain approximately the same number of values of x written down in the table, with relative error decreasing with the step of the arithmetic progression of x values. From that, the probability that a value of x randomly chosen in (α, β) lies in a given subinterval is approximately

proportional to the size of that subinterval, exactly as it happens in the ideal table.

This ideal table may be regarded as the limit of a sequence of real tables as described when the step of the arithmetic progression of the x 's decreases towards 0. Thus, the smaller is the step of the arithmetic progression of the x 's, the closer general formula (4.1) will be to the distribution of digits in a table of $f(x)$. Therefore, in a table of basis 10 logarithms, as the mantissa doesn't change when dividing x by an (integer) power of 10, we expect the formula

$$\frac{\mathbb{P}(d+1, a)}{\mathbb{P}(d, a)} = 10^{\frac{1}{10^a}}$$

to give much closer results at the end of the table than at its beginning.

This is in fact so. For instance, counting the number of digits 1 and 2 in the second decimal place in a table of basis 10 logarithms between 1289 and 1319, or between 1319 and 1349, we find 30 of each of those digits, getting

$$\frac{\mathbb{P}(2, 2)}{\mathbb{P}(1, 2)} = \frac{30}{30} = 1;$$

on the other hand, the number of digits 1 and 2 in the second decimal place in a tables of basis 10 logarithms between 10232 and 10471, or between 10471 and 19715, are respectively 239 and 244, and thus

$$\frac{\mathbb{P}(2, 2)}{\mathbb{P}(1, 2)} = \frac{244}{239} \approx 1.0209,$$

much closer to the theoretical value

$$\sqrt[100]{10} \approx 1.0233$$

for the ideal table.

3rd

It has some independent interest to compute the ratio

$$\frac{\mathbb{P}(d+1)}{\mathbb{P}(d)}$$

for the tabular differences of logarithms. Those differences may be regarded as values of the function

$$\begin{aligned} y &= \log(1+x) - \log x = \\ &= \log \left(1 + \frac{1}{x} \right) \end{aligned}$$

corresponding to x values in arithmetic progression. Rewriting

$$\begin{aligned} \log \left(1 + \frac{1}{x} \right) &= \frac{10\omega + d}{10^a}, \\ \log \left(1 + \frac{1}{x'} \right) &= \frac{10\omega + d + 1}{10^a}, \\ \log \left(1 + \frac{1}{x''} \right) &= \frac{10\omega + d + 2}{10^a}, \end{aligned}$$

we get

$$\begin{aligned} \frac{\mathbb{P}(d+1)}{\mathbb{P}(d)} &= \frac{x'' - x'}{x' - x} = \\ &= \frac{\frac{1}{10^{\frac{10\omega+d+2}{10^a}} - 1} - \frac{1}{10^{\frac{10\omega+d+1}{10^a}} - 1}}{\frac{1}{10^{\frac{10\omega+d+1}{10^a}} - 1} - \frac{1}{10^{\frac{10\omega+d}{10^a}} - 1}} \end{aligned}$$

very approximately

$$\begin{aligned} &\approx \frac{\frac{\frac{1}{10^{\frac{10\omega+d+2}{10^a}} - 1} - \frac{1}{10^{\frac{10\omega+d+1}{10^a}} - 1}}{\frac{1}{10^{\frac{10\omega+d+1}{10^a}} - 1} - \frac{1}{10^{\frac{10\omega+d}{10^a}} - 1}}}{\frac{\frac{10\omega+d+2}{10^a} \log 10 - \frac{10\omega+d+1}{10^a} \log 10}{\frac{10\omega+d+1}{10^a} \log 10 - \frac{10\omega+d}{10^a} \log 10}} = \end{aligned}$$

$$= \frac{(10\omega + d)(10\omega + d + 1)}{(10\omega + d + 1)(10\omega + d + 2)} = \frac{10\omega + d}{10\omega + d + 2};$$

therefore

$$\frac{\mathbb{P}(d)}{\mathbb{P}(d+1)} = 1 + \frac{2}{10N + d} = 1 + \frac{2}{10^a D}, \quad (4.2)$$

where

$$D = \log \left(1 + \frac{1}{x} \right) = \frac{10N + d}{10^a}.$$

As $10^a D$ is the integer part of the product of 10^a by any tabular difference, with the digit d in the a -th decimal place, we conclude that, given a tabular difference D , we can get from (4.2) the ratio

$$\frac{\mathbb{P}(d)}{\mathbb{P}(d+1)},$$

where d is the digit in the a -th decimal place of D .

As an example, given the tabular difference 0.0000524, we get for $a = 5$,

$$\frac{\mathbb{P}(5)}{\mathbb{P}(6)} = 1 + \frac{2}{5} = 1.400.$$

Searching in the tables we find that the tabular difference 0.0000500 has a corresponding maximum at 8694; that the tabular difference 0.0000600 has a corresponding maximum at 7243; that the tabular difference 0.0000700 has a corresponding maximum at 6208; we therefore have, for those tables

$$\frac{\mathbb{P}(5)}{\mathbb{P}(6)} = \frac{1451}{1035} = 1.401.$$

Observation

For the integer positions, i.e. for $a = 0, -1, -2, \dots$, the formula

$$\frac{\mathbb{P}(d+1, a)}{\mathbb{P}(d, a)} = 10^{\frac{1}{10^a}}$$

is exact, since the integers are logarithms of values of x written in the tables.

2.4.2 Law of possibilities and law of probability

Proposition II

Law of possibilities

Let $A' \subset A$, and let B and B' be the images of, respectively, A and A' . Let M denote a free point in A , ΔS a neighborhood of M , and let N be its image in B , its neighborhood $\Delta S'$ the image of ΔS . Denoting $\Delta\omega$ the possibility of ΔS , it will also be the possibility of $\Delta S'$. Let

$$\lim_{\mu(\Delta S') \rightarrow 0} \frac{\Delta\omega}{\mu(\Delta S')},$$

where we assume that the limit is taken with the supremum of the projection of $\Delta S'$ on the coordinate axes goes to zero with $\Delta S'$. *The set of points N for which this limit exists is the domain of a function of the coordinates of N , whose value in each point is the above limit. That function is the law of possibility.*

Corollary

The possibility of B' is

$$\omega_{B'} = \int_{B'} \lim_{\mu(\Delta S') \rightarrow 0} \frac{\Delta\omega}{\mu(\Delta S')} dS'.$$

Thus, once the possibility law is known, we can compute the possibility of any region B' without any reference to the region A' of which B' is the image.

Proposition III

Law of probability

In analogy with the above definition, we define the probability law as the function whose value at each point N is given by

$$\lim_{\mu(\Delta S') \rightarrow 0} \frac{\Delta \mathbb{P}}{\mu(\Delta S')}.$$

Proposition IV

If A is the possible region in what regards the probability $\Delta \mathbb{P}$ and $\omega(N)$ denotes the possibility of N , then

$$\lim_{\mu(\Delta S') \rightarrow 0} \frac{\Delta \mathbb{P}}{\mu(\Delta S')} = \lim_{\mu(\Delta S') \rightarrow 0} \frac{\frac{\Delta \omega}{\int_A \omega(N) d\omega}}{\mu(\Delta S')}$$

or

$$\lim_{\mu(\Delta S') \rightarrow 0} \frac{\frac{\Delta \omega}{\mu(\Delta S')}}{\int_A \omega(N) d\omega} = \frac{\omega(N)}{\int_A \omega(N) d\omega},$$

showing that in each possible region the probability law is proportional to the possibility law.

2.4.3 *A priori* and *a posteriori* laws

In what follows, we assume, without loss of generality, that the random point varies in a plane region, so that the arguments can be presented in an easy way.

A priori law

Let $M(x, y)$ denote a random point varying in a plane region (Fig. 12), and let $m(x)$ denote its projection in the OX axis.

The *a priori* law of the point $M(x, y)$ is the probability law of its projection $m(x)$.

Proposition V

If $\varphi(x, y)$ denotes the probability law of $M(x, y)$ in the region A , then its *a priori* probability law is

$$a(x) = \int \varphi(x, y) \, dy.$$

In effect, let us consider a vertical band with width ΔS , containing the points with abscissa x .

The probability $\Delta\mathbb{P}$ that the point m lies in a neighborhood ΔS of x is the probability that the point $M(x, y)$ lies in the region $[abcd]$; therefore

$$\Delta\mathbb{P} = \iint_{[abcd]} \varphi(x, y) \, dx \, dy = \mu(\Delta S) \int_{y'_1}^{y'_2} \varphi(x_1, y) \, dy,$$

where y'_1 and y'_2 stand for the smaller and the greater ordinates of the points in A with abscissa x_1 . Or, according to the definition of *a priori* law,

$$a(x) = \lim_{\mu(\Delta S) \rightarrow 0} \frac{\Delta\mathbb{P}}{\mu(\Delta S)} = \int_{y'_1}^{y'_2} \varphi(x, y) \, dy,$$

□

Similarly,

$$a(y) = \int \varphi(x, y) \, dx.$$

We shall use a to denote an *a priori* law, and p to denote an *a posteriori* law.

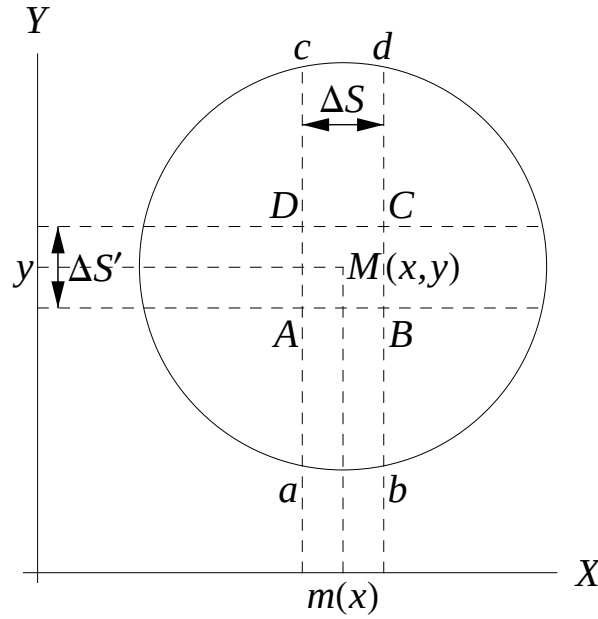


Figura 2.12: Figure 12

A posteriori law

Let us consider an horizontal and a vertical band in region A (Fig. 12), containing respectively the points with ordinate y and the points with abscissa x , and assume that the bandwidths are respectively $\Delta S'$ and ΔS . The probability of the region $[ABCD]$ in relation to the region $[abcd]$ (Chapter II, Prop. VII, Observation) is

$$\begin{aligned} \Delta \mathbb{P} &= \frac{\iint_{[ABCD]} \varphi(x, y) \, dx \, dy}{\iint_{[abcd]} \varphi(x, y) \, dx \, dy} = \frac{\mu(\Delta S) \cdot \mu(\Delta S') \cdot \varphi(x', y')}{\mu(\Delta S) \int_{y'_1}^{y'_2} \varphi(x'_1, y) \, dy} = \\ &= \frac{\mu(\Delta S') \cdot \varphi(x', y')}{\int_{y'_1}^{y'_2} \varphi(x'_1, y) \, dy} \end{aligned}$$

where x' and x'_1 are functions of ΔS that converge to x when ΔS goes to zero. We shall say that

$$\lim_{\mu(\Delta S) \rightarrow 0} \Delta \mathbb{P} = \frac{\mu(\Delta S') \cdot \varphi(x, y')}{\int_{y'_1}^{y'_2} \varphi(x'_1, y) \, dy}$$

is the *a posteriori* probability of $\Delta S'$.

We may interpret that a posteriori probability as the probability that y lies in the interval $\Delta S'$, given that x has taken the particular value x .

$$p(y) = \lim_{\mu(\Delta S') \rightarrow 0} \frac{\Delta \mathbb{P}}{\mu(\Delta S')} = \frac{\varphi(x, y)}{\int \varphi(x, y) \, dy} \quad (4.3)$$

is the *a posteriori* law of y .

Proposition VI

From Prop. V and (4.3) we get

$$\varphi(x, y) = a(x) \cdot p(y) = a(y) \cdot p(x).$$

Prop. V and VI are similar to the propositions about total probability and compound probability.

From those two propositions it is easy to infer a formula similar to

Bayes' formula

and this justifies the terminology *a priori* and *a posteriori* laws that we have been using.

In effect, from Prop. VI

$$\varphi(x, y) = a(x) \cdot p(y) = a(y) \cdot p(x).$$

and

$$a(y) = \int \varphi(x, y) \, dx = \int a(x) p(y) \, dx$$

it follows that

$$p(x) = \frac{\varphi(x, y)}{a(y)} = \frac{a(x) p(y)}{\int a(x) p(y) \, dx},$$

and, similarly,

$$p(y) = \frac{a(y) p(x)}{\int a(y) p(x) \, dy}.$$

Those are the formulas for the *a posteriori* law for each of the variables. Hence the *a posteriori* law of one variable can be computed from its *a priori* law, if the *a posteriori* law of the other variable and the corresponding domain of variation are known.

From the above formulas we can infer other formulas, which we shall call

Inverses to Bayes' formula

From

$$p(x) = \frac{a(x) p(y)}{\int a(x) p(y) \, dx},$$

taking partial derivative in order to x and remembering that

$$\int a(x) p(y) \, dx$$

doesn't depend on x , we get

$$\frac{\frac{\partial p(x)}{\partial x}}{p(x)} = \frac{a'(x)}{a(x)} + \frac{\frac{\partial p(y)}{\partial x}}{p(y)};$$

from this,

$$\frac{a'(x)}{a(x)} = \frac{\frac{\partial p(x)}{\partial x}}{p(x)} - \frac{\frac{\partial p(y)}{\partial x}}{p(y)};$$

thus

$$a(x) = k(y) \frac{p(x)}{p(y)},$$

where $k(y)$ is an arbitrary function of y which we determine using the condition

$$\int a(x) \, dx = k(y) \int \frac{p(x)}{p(y)} \, dx = 1;$$

therefore

$$k(y) = \frac{1}{\int \frac{p(x)}{p(y)} \, dx}$$

and

$$a(x) = \frac{\frac{p(x)}{p(y)}}{\int \frac{p(x)}{p(y)} \, dx}.$$

Similarly,

$$a(y) = \frac{\frac{p(y)}{p(x)}}{\int \frac{p(y)}{p(x)} \, dy}.$$

These formulas are analogous to those deduced in Chapter I.

The definitions and demonstrations we have presented have immediate generalization for higher dimensions. What we have established about a point varying in a plane region can be extended immediately for a point in any region, the only exception being these last formulas inverse to Bayes' formula, since the arguments are not usable in higher dimensions. But the extension to higher dimensions can also be made very easily.

In effect, if $\mathbf{x}$ and $\mathbf{y}$ are vectors, we still have

$$a(\mathbf{x}) \cdot p(\mathbf{y}) = a(\mathbf{y}) \cdot p(\mathbf{x}). \quad (4.4)$$

where $a(\mathbf{x})$ and $a(\mathbf{y})$ are functions of only $\mathbf{x}$ and of $\mathbf{y}$, respectively, and $p(\mathbf{y})$ and $p(\mathbf{x})$ are functions of, simultaneously, $\mathbf{x}$ and $\mathbf{y}$. From (4.4) we get

$$a(\mathbf{x}) = a(\mathbf{y}) \cdot \frac{p(\mathbf{x})}{p(\mathbf{y})}$$

and from this it follows that

$$\int_{(\mathbf{x})} a(\mathbf{x}) \, d\mathbf{x} = a(\mathbf{y}) \int_{(\mathbf{x})} \frac{p(\mathbf{x})}{p(\mathbf{y})} \, d\mathbf{x} = 1.$$

Therefore

$$a(\mathbf{x}) = \frac{\frac{p(\mathbf{x})}{p(\mathbf{y})}}{\int_{(\mathbf{x})} \frac{p(\mathbf{x})}{p(\mathbf{y})} \, d\mathbf{x}}$$

in all cases.

2.5 CHAPTER V — Jacob Bernoulli's Theorems and the Error Law

2.5.1 First Part — Jacob Bernoulli's theorems

As usual, we shall say that the *result* of the random selection of an element from a finite set, or of randomly throwing one point in a bounded region, is a *case*, or an *event*. A result, case or event is said to be *favorable* [or a *success*] if it is an element of the favorable set, and *contrary* [or a *failure*] if it is an element of the contrary set.

Let us denote p the probability of success, and q the probability of failure. Obviously,

$$p + q = 1.$$

Performing once a random experiment [trial] as described, two cases can occur, with probabilities

$$p \quad \text{or} \quad q$$

Performing this random experiment twice, there are four possible outcomes, the corresponding probabilities being

$$pp \quad \text{or} \quad pq \quad \text{or} \quad qp \quad \text{or} \quad qq$$

Three trials would produce eight possible outcomes, with probabilities given by arrangements with repetitions of the two elementary probabilities p and q , three by three, etc. In the above reasoning, we are assuming that any outcome doesn't change the probability of the outcomes in the following experiment.

We may therefore conclude the following:

Proposition I

Performing m times an experiment whose possible results are success, with probability p , or failure, with probability q , the probability of getting $m - i$ successes and i failures, in a given order, is

$$\mathbb{P} = p^{m-i} q^i;$$

this is a direct consequence of the propositions concerning compound probability.

Corollary

The probability of any given sequence of outcomes in pre-arranged order decreases to zero, when the number of trials increase.

As a matter of fact, and assuming, without loss of generality, that $q \leq p$, from

$$\mathbb{P} = p^{m-i} q^i \leq p^m$$

with $p < 1$, the result follows, and we see that p^m goes to zero when m increases to infinity.

When performing m trials as described, the probability of getting as outcome $m - i$ successes and i failures in a pre-arrange order is always the same, $\mathbb{P} = p^{m-i} q^i$; therefore, the probability of getting $m - i$ successes and i failures, whichever their order, depends on the number of possible sequences of m trials whose outcome consists exactly of $m - i$ successes and i failures. Our immediate goal will be to establish the appropriate formula.

For the sake of clarity, we solve the question in the context of the random extraction of white and black balls from an urn whose composition is such that the probability of extracting white ball is p and the probability of extracting black ball is q . Clearly this identification of the two problems is legitimate only when p and q are rational numbers; but, as a metaphor, we shall use this language in all cases.

Proposition II

In one urn there are white and black balls, the probability of getting white ball in a random extraction being p ; performing m extractions, [with replacement of the extracted ball in the urn after each of them,] the probability of getting white ball in n of those extractions, and black ball in the remaining $m - n$ extractions, is

$$\mathbb{P}_{m,n} = \frac{m!}{(m-n)! n!} p^n q^{m-n}.$$

In effect, the probability of any of the sequences of n white and $m - n$

black balls is (Prop. I)

$$p^n q^{m-n}.$$

On the other hand, the number of possible different sequences composed of n white and $m - n$ black balls is the number of combinations of m objects of two types, having n of the first and $m - n$ of the second type, i.e.,

$$\frac{m!}{(m-n)!n!}.$$

Therefore,

$$\mathbb{P}_{m,n} = \frac{m!}{(m-n)!n!} p^n q^{m-n}.$$

□

Corollary

The probability of getting n white and $m - n$ black balls is given by the corresponding term in the expansion of

$$(p + q)^m.$$

Proposition III

Assuming m fixed, we have:

1. *The probability $\mathbb{P}_{m,n}$ increases with n while n goes from zero until the biggest integer not greater than*

$$p(m+1).$$

- 2 *The probability $\mathbb{P}_{m,n}$ decreases from the least integer greater than*

$$p(m+1) - 1$$

until $n = m$.

3 The probability $\mathbb{P}_{m,n}$ assumes its maximum in the integers in the interval $[p(m+1)-1, p(m+1)]$.

In effect,

1st

If

$$n \leq p(m+1) \quad (5.1)$$

we have

$$\frac{1}{p} \leq \frac{m+1}{n},$$

$$\frac{q}{p} \leq \frac{m+1}{n} - 1 = \frac{m-n+1}{n}$$

and

$$1 \leq \frac{m-n+1}{n} \cdot \frac{p}{q}.$$

But

$$\frac{m-n+1}{n} \cdot \frac{p}{q} = \frac{\mathbb{P}_{m,n}}{\mathbb{P}_{m,n-1}},$$

and therefore

$$\mathbb{P}_{m,n-1} \leq \mathbb{P}_{m,n}.$$

□

2nd

If

$$n \geq p(m+1)-1 \quad (5.2)$$

we have

$$\frac{1}{p} \geq \frac{m+1}{1+n},$$

$$1 + \frac{q}{p} \geq \frac{m+1}{1+n},$$

$$\frac{q}{p} \geq \frac{m-n}{n+1},$$

and

$$1 \geq \frac{m-n}{n+1} \cdot \frac{p}{q}.$$

But

$$\frac{m-n}{n+1} \cdot \frac{p}{q} = \frac{\mathbb{P}_{m,n+1}}{\mathbb{P}_{m,n}},$$

and therefore

$$\mathbb{P}_{m,n} \geq \mathbb{P}_{m,n+1}.$$

□

3rd

From the first and the second part of the present proposition, we can deduce that $\mathbb{P}_{m,n}$ will assume its maximum value when n verifies both (5.1) and (5.2), i.e., when it is an integer of the form

$$p(m+1) - r, \quad (0 \leq r \leq 1).$$

When $p(m+1)$ is non integer, there is a unique value of n in $[p(m+1) - 1, p(m+1)]$ for which $\mathbb{P}_{m,n}$ assumes its maximum value. When $p(m+1)$ is an integer, $\mathbb{P}_{m,n}$ assumes its maximum for any of the terms of order

$$p(m+1) \quad \text{or} \quad p(m+1) - 1.$$

□

Proposition IV

Let us consider all the sequences that may be obtained by repeatedly extracting two objects, with replacement after each extraction.

For clarity, let the two objects be white ball and black ball [extracted from one urn such that the probability of extracting white ball is p and that of extracting black ball is $q = 1 - p$]. We shall call a combination of outcomes the totality of sequences with the same number k of white and $m - k$ of black balls. With these assumptions, we have:

The probability of the most probable combination decreases to zero when the number of trials increases to infinity.

In effect, using Prop. III, the probability of the most probable combination is

$$\mathbb{P}_{m, p(m+1)-r} = \frac{m!}{[p(m+1)-r]! [q(m+1)-1+r]!} p^{p(m+1)-r} q^{q(m+1)-1+r},$$

since when the number of white balls is $(m+1)p - r$ (the most probable outcome in m extractions), the corresponding number of black balls is

$$m - (m+1)p + r = (m+1)q + r - 1.$$

However,

$$m! = m^m e^{-m} \sqrt{2\pi m} (1 + \varepsilon_m),$$

where ε_m is a function going to zero when m goes to infinity. Therefore we may write

$$\begin{aligned} \mathbb{P}_{m, p(m+1)-r} &= \\ &= \frac{m^m e^{-m} \sqrt{2\pi m} (1 + \varepsilon_m)}{(mp)^{mp} e^{-mp} \sqrt{2\pi mp} (1 + \varepsilon_{mp}) (mq)^{mq} e^{-mq} \sqrt{2\pi mq} (1 + \varepsilon_{mq})} p^{mp} q^{mq}, \end{aligned}$$

an expression obtained using Stirling's approximation for the factorials in the previous expression, and cutting out the vanishingly small terms in $p - r$ and $q + r - 1$ (a legitimate approximation when m becomes as large as we want). This may be rewritten

$$\mathbb{P}_{m,p(m+1)-r} = \frac{1 + \alpha_m}{\sqrt{2\pi m p q}}$$

where α_m denotes a function going to zero when m goes to infinity. Therefore, $\mathbb{P}_{m,p(m+1)-r}$ goes to zero when m goes to infinity, as stated. $\square$

Proposition V

(Jacob Bernoulli's 1st Theorem)

Let p denote the probability of the favorable event or success, and q the probability of the contrary event or failure. Performing a certain number of trials, let us denote by (p) and by (q) the number of successes and the number of failures that occur [in the $m = (p) + (q)$ trials], respectively.

The ratio $\frac{(p)}{(q)}$ can take different values; but the most probable among them is the one nearer to $\frac{p}{q}$; the more $\frac{(p)}{(q)}$ differs from $\frac{p}{q}$, the less probable it is.

The probability of $\frac{(p)}{(q)}$ in the above proposition is the combination of (p) successes and (q) failures in any order.

The proposition may be established as follows:

As seen in Prop. III (3rd part), the most probable number of successes in m trials is the greatest integer that can be written in the form

$$p(m+1) - r, \quad (0 \leq r \leq 1);$$

the corresponding number of failures is

$$q(m+1) - 1 + r;$$

the ratio of those numbers is

$$\frac{(p)}{(q)} = \frac{p(m+1) - r}{q(m+1) - 1 + r} = \frac{p}{q} + \frac{(1-r)p - qr}{(m+1)q - 1 + r} \cdot \frac{1}{q}.$$

Assuming that the number of successes increases, the ratio with immediately lower probability (Prop. III, 2nd part) is

$$\frac{(p)}{(q)} = \frac{p(m+1) - r + 1}{q(m+1) - 1 + r - 1} = \frac{p}{q} + \frac{(1-r)p - qr + 1}{(m+1)q - 1 + r - 1} \cdot \frac{1}{q};$$

and the immediate one is

$$\frac{(p)}{(q)} = \frac{p(m+1) - r + 2}{q(m+1) - 1 + r - 2} = \frac{p}{q} + \frac{(1-r)p - qr + 2}{(m+1)q - 1 + r - 2} \cdot \frac{1}{q};$$

when the most probable number of successes is exceeded by α units, the ratio in question will be

$$\frac{(p)}{(q)} = \frac{p(m+1) - r + \alpha}{q(m+1) - 1 + r - \alpha} = \frac{p}{q} + \frac{(1-r)p - qr + \alpha}{(m+1)q - 1 + r - \alpha} \cdot \frac{1}{q}.$$

As those expressions show, the difference

$$\frac{(p)}{(q)} - \frac{p}{q}$$

increases with α ; therefore (Prop. III, 2nd part), its probability decreases.

If the number of successes would decrease, we would work out similarly with the ratio $\frac{(q)}{(p)}$, and so this case could be dealt with as the above one.

□

Proposition VI

(Jacob Bernoulli's 2nd Theorem)

As the number of trials increases, the probability of each ratio $\frac{(p)}{(q)}$ decreases, and the greater is the absolute value of the difference

$$\left| \frac{p}{q} - \frac{(p)}{(q)} \right|$$

the greater will be the rate of decrease.

The number α used in the proof of the previous proposition will be called deviation; the probability of $\frac{(p)}{(q)}$ attains its maximum when that deviation is zero (Prop. V), and since in that case it decreases to zero when m goes to infinity (Prop. IV), it will also decrease to zero in all the other less probable cases. On the other hand,

$$\frac{\mathbb{P}_{\alpha-1}}{\mathbb{P}_{\alpha}} = \frac{q(m+1) + r - \alpha}{p(m+1) - r - \alpha + 1} \cdot \frac{p}{q} = \frac{A + p\alpha}{B - q\alpha},$$

an expression that shows that the ratio

$$\frac{\mathbb{P}_{\alpha-1}}{\mathbb{P}_{\alpha}}$$

decreases with α . In other words, $\frac{\mathbb{P}_{\alpha-1}}{\mathbb{P}_{\alpha}}$ decreases with

$$\left| \frac{p}{q} - \frac{(p)}{(q)} \right|,$$

since what is true for the deviation from p is also true for the deviation from q .

Proposition VII
(Vallée-Poussin's Lemma)

We now denote T_i the probability of a combination of outcomes with i successes, and by T_n the probability of the most probable combination of outcomes.

Denoting

$$S = T_{n-\alpha} + T_{n-\alpha+1} + \cdots + T_n + \cdots + T_{n+\alpha}$$

we have

$$1 - S < \frac{m}{\left(1 + \frac{\alpha}{(m+1)pq}\right)^{\frac{\alpha}{2}}}.$$

In effect, we have

$$\frac{T_{n+1}}{T_n} = \frac{m-n}{n+1} \cdot \frac{p}{q},$$

where

$$n = (m+1)p - r$$

and

$$n+1 = (m+1)p - r + 1$$

and

$$m-n = (m+1)q - 1 + r;$$

with

$$1 - r = \varepsilon, \quad (0 \leq \varepsilon \leq 1),$$

we get

$$\frac{T_{n+1}}{T_n} = \frac{(m+1)q - \varepsilon}{(m+1)p + \varepsilon} \cdot \frac{p}{q} = \frac{1 - \frac{\varepsilon}{q(m+1)}}{1 + \frac{\varepsilon}{p(m+1)}}.$$

It is easily established that for any $\varepsilon > 0$

$$\frac{1 - \frac{\varepsilon}{q(m+1)}}{1 + \frac{\varepsilon}{p(m+1)}} < \frac{1}{1 + \frac{\varepsilon}{pq(m+1)}};$$

therefore

$$\frac{T_{n+1}}{T_n} < \frac{1}{1 + \frac{\varepsilon}{pq(m+1)}}.$$

As ε can be any positive number, using similar arguments we get

$$\frac{T_{n+2}}{T_{n+1}} = \frac{m-n-1}{n+1+1} \cdot \frac{p}{q} = \frac{(m+1)q - \varepsilon - 1}{(m+1)p + \varepsilon - 1} \cdot \frac{p}{q} =$$

$$= \frac{(m+1)q - (1 + \varepsilon)}{(m+1)p + (1 + \varepsilon)} \cdot \frac{p}{q} <$$

$$< \frac{1}{1 + \frac{\varepsilon + 1}{pq(m+1)}}$$

and, in general

$$\frac{T_{n+\alpha+1}}{T_{n+\alpha}} < \frac{1}{1 + \frac{\varepsilon + \alpha}{pq(m+1)}}.$$

Multiplying term by term these inequalities, we get

$$\begin{aligned} \frac{T_{n+\alpha+1}}{T_n} &< \frac{1}{1 + \frac{\varepsilon}{pq(m+1)}} \cdot \frac{1}{1 + \frac{\varepsilon + 1}{pq(m+1)}} \dots \frac{1}{1 + \frac{\varepsilon + \alpha}{pq(m+1)}} < \\ &< \frac{1}{1 + \frac{1}{pq(m+1)}} \dots \frac{1}{1 + \frac{\alpha}{pq(m+1)}} = \end{aligned}$$

$$= \prod_{k=1}^{\alpha} \frac{1}{1 + \frac{k}{pq(m+1)}};$$

reversing the order of all the above factors

$$\frac{T_{n+\alpha+1}}{T_n} < \prod_{k=1}^{\alpha} \frac{1}{1 + \frac{\alpha+1-k}{pq(m+1)}};$$

multiplying term by term those two inequalities, we get

$$\left(\frac{T_{n+\alpha+1}}{T_n}\right)^2 < \prod_{k=1}^{\alpha} \left[\frac{1}{1 + \frac{k}{pq(m+1)}} \cdot \frac{1}{1 + \frac{\alpha+1-k}{pq(m+1)}} \right];$$

but

$$\begin{aligned} & \left[1 + \frac{k}{pq(m+1)} \right] \left[1 + \frac{\alpha+1-k}{pq(m+1)} \right] = \\ & = 1 + \frac{\alpha+1}{pq(m+1)} + \frac{k(\alpha+1-k)}{[pq(m+1)]^2} > \\ & > 1 + \frac{\alpha+1}{pq(m+1)}; \end{aligned}$$

therefore

$$\begin{aligned} \left(\frac{T_{n+\alpha+1}}{T_n}\right)^2 & < \prod_{k=1}^{\alpha} \left[1 + \frac{\alpha+1}{pq(m+1)} \right]^{-1} = \\ & = \left[1 + \frac{\alpha+1}{pq(m+1)} \right]^{-\alpha} < \\ & < \left[1 + \frac{\alpha}{pq(m+1)} \right]^{-\alpha}; \end{aligned}$$

from the above inequality, it follows that

$$T_{n+\alpha+1} < T_n \left[1 + \frac{\alpha}{pq(m+1)} \right]^{-\frac{\alpha}{2}}$$

and, as

$$T_n < 1,$$

we get

$$T_{n+\alpha+1} < \left[1 + \frac{\alpha}{pq(m+1)} \right]^{-\frac{\alpha}{2}}$$

On the other hand, the probabilities of the combinations of outcomes which are not in S are all smaller than $T_{n+\alpha+1}$ (Prop. III); in consequence,

$$1 - S < [m - (2\alpha + 1)] T_{n+\alpha+1} < m T_{n+\alpha+1} <$$

$$< \frac{m}{\left[1 + \frac{\alpha}{pq(m+1)} \right]^{\frac{\alpha}{2}}}.$$

□

Proposition VIII

(Jacob Bernoulli's 3rd Theorem)

The probability that the deviation of the ratio $\frac{(p)}{(q)}$ between the number of successes and the number of failures from the odds ratio $\frac{p}{q}$ of the corresponding probabilities falls within given bounds is always increasing to 1, when the number of trials is large enough, however tight these bounds may be.

In effect, in Prop. V we have seen that, assuming the deviation to be positive,

$$\frac{(p)}{(q)} = \frac{p}{q} + \frac{p - r + \alpha}{mq - p + r - \alpha} \cdot \frac{1}{q},$$

and therefore

$$\left| \frac{\binom{p}{q}}{\binom{q}{q}} - \frac{p}{q} \right| = \frac{p - r + \alpha}{mq - p + r - \alpha} \cdot \frac{1}{q}.$$

For

$$\left| \frac{\binom{p}{q}}{\binom{q}{q}} - \frac{p}{q} \right| > \varepsilon$$

it is necessary that

$$\frac{p - r + \alpha}{mq - p + r - \alpha} \cdot \frac{1}{q} > \varepsilon$$

or

$$p - r + \alpha > mq^2\varepsilon - pq\varepsilon + rq\varepsilon - q\alpha\varepsilon$$

i.e.

$$\alpha(1 + q\varepsilon) > mq^2\varepsilon + \dots$$

or

$$\alpha > A(m + 1) + B, \tag{5.3}$$

where $A > 0$ and B are constants.

But, in Prop. VII, S denotes the probability that the deviation is less than or equal to α ; in other words, $1 - S$ is the probability that the deviation is greater than α .

In consequence, the probability that

$$\left| \frac{\binom{p}{q}}{\binom{q}{q}} - \frac{p}{q} \right| > \varepsilon$$

is

$$\begin{aligned} 1 - S &< \frac{m}{\left[1 + \frac{\alpha}{(m + 1)pq} \right]^{\frac{\alpha}{2}}} < \\ &< \frac{m}{\left(1 + \frac{A(m + 1) + B}{pq(m + 1)} \right)^{\frac{A(m + 1) + B}{2}}} \end{aligned}$$

an expression that decreases to zero when m goes to infinity, since $A > 0$.

This theorem is also known as *the law of large numbers*.

Observation

We do not explicit the case $\alpha < 0$ since in that case an inequality similar to (5.3) holds, and therefore it reduces to the former situation $\alpha > 0$.

The 3rd Bernoulli's theorem may be rephrased as follows:

The probability that the deviation [of $\frac{(p)}{(q)}$ from the odds ratio $\frac{p}{q}$] is of the order of the number of trials decreases to zero when the number of trials goes to infinity.

Proposition IX

The probability that the deviation α is such that

$$\frac{\alpha^{n+1}}{m^n} > \varepsilon, \quad (5.4)$$

decreases to zero when m increases, if $n > 1$.

In effect, from (5.4) we get

$$\frac{m}{\left[1 + \frac{\alpha}{pq(m+1)}\right]^{\frac{\alpha}{2}}} < \frac{m}{\left[1 + \frac{\varepsilon' m^{\frac{n}{n+1}}}{pq(m+1)}\right]^{\frac{\varepsilon' m^{\frac{n}{n+1}}}{2}}}.$$

Disregarding vanishing terms, the second member of the above inequality can be approximated by

$$\frac{m}{\left[1 + Am^{-\frac{1}{n+1}}\right]^{B m^{\frac{n}{n+1}}}},$$

which in turn may be expanded as

$$\begin{aligned}
 & \frac{m}{1 + A_1 m^{\frac{n}{n+1}} m^{-\frac{1}{n+1}} + A_2 m^{2\frac{n}{n+1}} m^{-\frac{2}{n+1}} + \dots} = \\
 & = \frac{m}{1 + A_1 m^{\frac{n-1}{n+1}} + A_2 m^{2\frac{n-1}{n+1}} + \dots} = \\
 & = \frac{1}{\frac{1}{m} + A_1 m^{\frac{n-1}{n+1}-1} + A_2 m^{2\frac{n-1}{n+1}-1} + \dots}
 \end{aligned}$$

an expression that decreases to zero when the number of trials goes to infinity, provided $n > 1$, since there exists some integer i for which

$$i \frac{n-1}{n+1} - 1 > 0.$$

This proposition may be rephrased as follows:

The probability that the number of trials is of order less than two of the deviation α is zero.

Proposition X

The probability that

$$\frac{\alpha}{\sqrt[n]{m}} < \varepsilon$$

decreases to zero when the number of trials goes to infinity, if $n > 2$.

In effect, as

$$T_{n-\alpha} < T_n = \frac{1 + \alpha_m}{\sqrt{2\pi p q m}},$$

it follows that

$$S < (2\alpha + 1)T_n = \frac{2\alpha + 1}{\sqrt{2\pi p q m}} (1 + \alpha_m),$$

or

$$S < \frac{\alpha}{\sqrt{m}} C_m,$$

where C_m converges to a constant when m goes to infinity. Therefore, if

$$\frac{\alpha}{\sqrt[n]{m}} < \varepsilon$$

we get

$$\begin{aligned} S &< \frac{\sqrt[n]{m}}{\sqrt{m}} \varepsilon C_m = \\ &= m^{\frac{1}{n}-\frac{1}{2}} \cdot \varepsilon C_m \end{aligned}$$

and thus

$$\lim_{m \rightarrow \infty} S = 0$$

if

$$\frac{1}{n} - \frac{1}{2} < 0, \quad \text{or} \quad n > 2.$$

□

From this, it follows that *the probability that the number of trials is of order greater than two of the deviation α is zero.*

Henceforth,

Proposition XI

The number of trials is of order 2 in what regards the deviation α , an immediate consequence of Prop. IX and X.

Any real number can be written in decimal form with an infinite number of decimal places. For instance, the number $\frac{1}{2}$ can be written 0.50000...

Assuming that we are dealing with numbers in that form of representation, we shall examine the following

Problem

A number is randomly chosen in the interval $(0, 1)$; what is the probability that the sequence of digits in its decimal expansion satisfies Bernoulli's law?

The problem as stated has no solution with the definition of probability that we have adopted.

In effect, we have a problem of continuous probability, since the random extraction is performed in the interval $(0, 1)$, and in Chapter II we have solely defined the probability of *regions* in reference to other possible *regions*. Therefore, in the context we have adopted, the problem would have a solution if the favorable class, i.e. the sequence of digits of a number, would be a complete interval contained in $(0, 1)$, which is not true, or at least *a priori* cannot be taken for granted.

In view of that, we shall instead consider a simpler problem, namely: *what is the probability that the N first digits in the decimal expression of a number randomly chosen in $(0, 1)$ satisfy Bernoulli's law?*

This problem has a solution, since the set of numbers which share the first N digits is an interval.

Some of these intervals are favorable, others aren't.

The measure of the union of the favorable intervals is the solution we search.

As all the intervals are identical, they are equally possible, and as there are 10^N of those intervals, the possibility of each of them is

$$\frac{1}{10^N}.$$

Let us consider one urn with 10 balls, numbered $0, 1, \dots, 9$. The first N digits of any number in $(0, 1)$ can be identified with the sequence of digits

obtained in N random extractions, with replacement, of balls from that urn. In effect, the possibility of each sequence is

$$\frac{1}{10} \cdot \frac{1}{10} \cdots \frac{1}{10} = \frac{1}{10^N},$$

the same that we had found previously for each of the partial intervals.

Therefore, in the problem at hand, randomly selecting a number in the interval $(0, 1)$ is the same as performing N random extractions, with replacement, of balls of the urn described.

The same conclusion could be reached, also, using the expression obtained in the problem solved in page 74 of these *Elements*, assuming $f(x) = x$, that would immediately imply that

$$\mathbb{P} = \frac{1}{10}$$

whatever a and d .

From the identification of the modified problem with random extractions, it is immediate that the ratio between the number of combinations of outcomes that satisfy Bernoulli's law and the total number of combinations of outcomes increases towards 1, as N increases. This is the core of Bernoulli's theorem.

Thus, the probability asked for in the restated problem, increases to 1 when N increases. We may conclude that the probability asked for in the original problem, which corresponds to the limit when $N \rightarrow \infty$, is equal to 1.

The argument above also shows that the probability that the sequence of digits of a randomly chosen number in $(0, 1)$ satisfies all laws similar to that of Bernoulli is 1.

From this we may deduce that *the probability that a randomly chosen number in $(0, 1)$ is rational is zero.*

In effect, a rational number has periodic decimal representation.

And the two possible cases are: either all digits from 0 to 9 appear, in the same proportion, in its period, or this is not so. In the second instance, the sequence of digits doesn't satisfy Bernoulli's law. On the other hand, in the first case, the distribution of the digits in the sequence satisfies Bernoulli's law, but none of the others, since the absolute deviation will periodically take on the same values, therefore it will never remain lower than a given bound.

Therefore, the set of rational numbers corresponds to a combination of outcomes that either doesn't satisfy Bernoulli's law or doesn't satisfy the analogous laws. Its probability is, therefore, zero.

2.5.2 Second Part — Law of deviations (Error law)

After having presented a rigorous proof of Jacob Bernoulli's 3rd theorem and others similar theorems on the order of magnitude of deviations [from the most probable combination of outcomes], we shall now establish an approximate relation between the deviations and their probabilities.

Proposition XII

Denoting T_n the probability of the combination of outcomes of maximum probability, which we may call *normal* combination, the probability that the deviation has absolute value less than or equal to k is

$$\mathbb{P}(k) = T_{n-k} + \cdots + T_n + \cdots + T_{n+k} = \sum_{i=-k}^k T_{n+i},$$

where

$$T_{n+i} = \frac{m!}{(n+i)!(m-n-i)!} p^{n+i} q^{m-n-i},$$

where n denotes the number of white balls in the normal combination, and therefore is of the form

$$n = (m+1)p - r, \quad (0 \leq r \leq 1).$$

Assuming that m is large enough, we can use

$$mp$$

as an approximation for

$$(m+1)p - r;$$

in Stirling's approximation for

$$(m-n-i)! \quad \text{and} \quad (n+i)!.$$

T_{n+i} may then be approximated by

$$\begin{aligned} T_{n+i} &= \frac{m^m e^{-m} \sqrt{2\pi} mp^{mp+i} q^{mq-i} (1 + \alpha_m)}{(mp+i)^{mp+i} e^{-mp-i} \sqrt{2\pi} (mp+i) (mq-i)^{mq-i} e^{-mq+i} \sqrt{2\pi} (mq-i)} = \\ &= \frac{\sqrt{m} p^{mp+i} q^{mq-i}}{\left(p + \frac{i}{m}\right)^{mp+i} \left(q - \frac{i}{m}\right)^{mq-i} \sqrt{2\pi} (mp+i) (mq-i)} (1 + \alpha_m) = \\ &= \frac{p^{mp+i} q^{mq-i}}{\left(p + \frac{i}{m}\right)^{mp+i+\frac{1}{2}} \left(q - \frac{i}{m}\right)^{mq-i+\frac{1}{2}} \sqrt{2\pi} m} (1 + \alpha_m) = \\ &= \frac{1}{\sqrt{2\pi} mpq} \cdot \frac{1}{\left(1 + \frac{i}{mp}\right)^{mp+i+\frac{1}{2}} \left(1 - \frac{i}{mq}\right)^{mq-i+\frac{1}{2}}} (1 + \alpha_m) = \\ &= \frac{1}{\sqrt{2\pi} mpq} \cdot H \end{aligned}$$

where

$$H = \left(1 + \frac{i}{mp}\right)^{-mp-i-\frac{1}{2}} \left(1 - \frac{i}{mq}\right)^{-mq+i-\frac{1}{2}} (1 + \alpha_m)$$

and

$$\begin{aligned} \log H = & - \left(mp + i + \frac{1}{2} \right) \log \left(1 + \frac{i}{mp} \right) - \\ & - \left(mq - i + \frac{1}{2} \right) \log \left(1 - \frac{i}{mq} \right) + \log(1 + \alpha_m) \end{aligned}$$

But (Bernoulli's 3rd theorem) the probability that

$$\left| \frac{i}{m} \right| > \varepsilon$$

decreases to zero when m increases; therefore the probability that

$$\begin{aligned} \log H = & - \left(mp + i + \frac{1}{2} \right) \left[\frac{i}{mp} - \frac{i^2}{2m^2p^2} + \frac{i^3}{3m^3p^3} - \cdots \right] - \\ & - \left(mq - i + \frac{1}{2} \right) \left[-\frac{i}{mq} - \frac{i^2}{2m^2q^2} - \frac{i^3}{3m^3q^3} - \cdots \right] + \\ & + \log(1 + \alpha_m) = \\ = & -\frac{i^2}{m} \left(\frac{1}{p} + \frac{1}{q} \right) - \frac{1}{2} \cdot \frac{i}{m} \left(\frac{1}{p} - \frac{1}{q} \right) + \\ & + \frac{i^2}{2m} \left(\frac{1}{p} + \frac{1}{q} \right) + \frac{i^3}{2m^2} \left(\frac{1}{p^2} - \frac{1}{q^2} \right) + \\ & + \frac{1}{2} \frac{i^2}{2m^2} \left(\frac{1}{p^2} + \frac{1}{q^2} \right) + \cdots + \log(1 + \alpha_m) = \\ = & -\frac{i^2}{2mpq} - \frac{i(q-p)}{2mpq} + \frac{i^3(q^2-p^2)}{2m^2p^2q^2} + \cdots \end{aligned}$$

is always increasing.

On one hand, from what we know about Stirling's approximation, the term in α_m goes to zero; on the other hand, the probability that the summands

$$\frac{i}{m}, \frac{i^3}{m^2}, \dots, \frac{i^n}{m^n}, \frac{i^{n+1}}{m^n}, \dots$$

remain greater than ε , however small, also goes to zero when m increases (Prop. IX and X); therefore, the probability that

$$\log H \approx -\frac{i^2}{2mpq}$$

is always increasing, and thus

$$H \approx e^{-\frac{i^2}{2mpq}};$$

from this it follows that

$$T_{n+i} \approx \frac{1}{\sqrt{2\pi mpq}} e^{-\frac{i^2}{2mpq}}$$

and

$$\mathbb{P}(k) = \sum_{i=-k}^k T_{n+i} \approx \frac{2}{\sqrt{2\pi mpq}} \sum_{i=0}^k e^{-\frac{i^2}{2mpq}} :$$

substituting $\sum$ by $\int$ to be computed between the same limits, we have

$$\mathbb{P}(k) \approx \frac{2}{\sqrt{2\pi mpq}} \int_0^k e^{-\frac{x^2}{2mpq}} dx,$$

and with the substitution

$$x = \lambda \sqrt{2mpq},$$

$$\begin{aligned} \mathbb{P}(k) &= \frac{2}{\sqrt{2\pi mpq}} \int_0^{\lambda_1} e^{-\lambda^2} d\lambda \sqrt{2mpq} = \\ &= \frac{2}{\sqrt{\pi}} \int_0^{\lambda_1} e^{-\lambda^2} d\lambda, \end{aligned}$$

where

$$\lambda_1 = \frac{k}{\sqrt{2mpq}}.$$

We shall say that λ_1 is the *relative deviation*, to distinguish it from the *absolute deviation* k . The value

$$\sqrt{2mpq}$$

is the *deviation unit*.

The probability that the absolute value of the relative deviation is smaller than λ_1 is the *probability of λ_1* .

Proposition XIII

When m increases, the probability that the probability of the absolute deviation is less than λ_1 is given by

$$\theta(\lambda_1) = \frac{2}{\sqrt{\pi}} \int_0^{\lambda_1} e^{-\lambda^2} d\lambda$$

also increases.

We shall refer to $\theta(\lambda_1)$ as the *error law*, the *deviations law*, or *Gauss' law*.

The law we have established is only a *probable* law and, in addition to that, an approximate result. Its probability, however, rapidly converges to 1 when m increases, and the errors incurred in the approximations used to deduce it vanish very quickly. The rate of convergence is so high that the approximation it gives is, in many practical applications, equal to the true result. It is always used in problems about deviations.

The probability that the variable $|\Lambda|$, the absolute value of the relative deviation, is between 0 and ∞ is

$$\mathbb{P} = \frac{2}{\sqrt{\pi}} \int_0^{\infty} e^{-\lambda^2} d\lambda = \frac{2}{\sqrt{\pi}} \frac{\sqrt{\pi}}{2} = 1,$$

this being a rigorous result, as should be expected; when $\lambda_1 \rightarrow \infty$, $m \rightarrow \infty$ also, and under that condition the law is *exact*.

The table that follows⁽¹⁸⁾ shows the value of $\theta(\lambda_1)$ for centesimal increases in the argument. They show how quickly $\theta(\lambda_1) \rightarrow 1$ when λ_1 increases.

λ_1	$\theta(\lambda_1)$	λ_1	$\theta(\lambda_1)$	λ_1	$\theta(\lambda_1)$
0.00	0.0000000	0.24	0.2657001	0.48	0.5027497
0.01	0.0112834	0.25	0.2763264	0.49	0.5116683
0.02	0.0225646	0.26	0.2868997	0.50	0.5204999
0.03	0.0338412	0.27	0.2974182	0.51	0.5292436
0.04	0.0451111	0.28	0.3078801	0.52	0.5378986
0.05	0.0563720	0.29	0.3182835	0.53	0.5464641
0.06	0.0676216	0.30	0.3286268	0.54	0.5549393
0.07	0.0788577	0.31	0.3389082	0.55	0.5633234
0.08	0.0900781	0.32	0.3491260	0.56	0.5716158
0.09	0.1012806	0.33	0.3592787	0.57	0.5798158
0.10	0.1124629	0.34	0.3693645	0.58	0.5879229
0.11	0.1236229	0.35	0.3793821	0.59	0.5959365
0.12	0.1347584	0.36	0.3893297	0.60	0.6038561
0.13	0.1458671	0.37	0.3992060	0.61	0.6116812
0.14	0.1569470	0.38	0.4090095	0.62	0.6194115
0.15	0.1679960	0.39	0.4187387	0.63	0.6270464
0.16	0.1790118	0.40	0.4283924	0.64	0.6345858
0.17	0.1899925	0.41	0.4379691	0.65	0.6420293
0.18	0.2009358	0.42	0.4474676	0.66	0.6493767
0.19	0.2118399	0.43	0.4568867	0.67	0.6566277
0.20	0.2227026	0.44	0.4662251	0.68	0.6637822
0.21	0.2335219	0.45	0.4754817	0.69	0.6708401
0.22	0.2442959	0.46	0.4846554	0.70	0.6778012
0.23	0.2550226	0.47	0.4937451	0.71	0.6846656

⁽¹⁸⁾ This table has been recalculated using Mathematica 5.1. Observe the accuracy of the computations in the original.

λ_1	$\theta(\lambda_1)$		λ_1	$\theta(\lambda_1)$		λ_1	$\theta(\lambda_1)$
0.72	0.6914331		1.08	0.8733262		1.44	0.9582966
0.73	0.6981039		1.09	0.8768031		1.45	0.9596950
0.74	0.7046781		1.10	0.8802051		1.46	0.9610535
0.75	0.7111556		1.11	0.8835330		1.47	0.9623729
0.76	0.7175368		1.12	0.8867879		1.48	0.9636541
0.77	0.7238216		1.13	0.8899707		1.49	0.9648979
0.78	0.7300104		1.14	0.8930823		1.50	0.9661051
0.79	0.7361035		1.15	0.8961238		1.51	0.9672767
0.80	0.7421010		1.16	0.8990962		1.52	0.9684135
0.81	0.7480033		1.17	0.9020004		1.53	0.9695162
0.82	0.7538108		1.18	0.9048374		1.54	0.9705857
0.83	0.7595238		1.19	0.9076083		1.55	0.9716227
0.84	0.7651427		1.20	0.9103140		1.56	0.9726281
0.85	0.7706681		1.21	0.9129555		1.57	0.9736026
0.86	0.7761003		1.22	0.9155339		1.58	0.9745470
0.87	0.7814398		1.23	0.9180501		1.59	0.9754620
0.88	0.7866873		1.24	0.9205052		1.60	0.9763484
0.89	0.7918432		1.25	0.9229001		1.61	0.9772068
0.90	0.7969082		1.26	0.9252359		1.62	0.9780381
0.91	0.8018828		1.27	0.9275136		1.63	0.9788428
0.92	0.8067677		1.28	0.9297342		1.64	0.9796218
0.93	0.8115636		1.29	0.9318986		1.65	0.9803756
0.94	0.8162710		1.30	0.9340079		1.66	0.9811049
0.95	0.8208908		1.31	0.9360631		1.67	0.9818104
0.96	0.8254236		1.32	0.9380652		1.68	0.9824928
0.97	0.8298703		1.33	0.9400150		1.69	0.9831526
0.98	0.8342315		1.34	0.9419137		1.70	0.9837905
0.99	0.8385081		1.35	0.9437622		1.71	0.9844070
1.00	0.8427008		1.36	0.9455614		1.72	0.9850028
1.01	0.8468105		1.37	0.9473124		1.73	0.9855785
1.02	0.8508380		1.38	0.9490160		1.74	0.9861346
1.03	0.8547842		1.39	0.9506733		1.75	0.9866717
1.04	0.8586499		1.40	0.9522851		1.76	0.9871903
1.05	0.8624361		1.41	0.9538524		1.77	0.9876909
1.06	0.8661436		1.42	0.9553762		1.78	0.9881742
1.07	0.8697733		1.43	0.9568573		1.79	0.9886405

λ_1	$\theta(\lambda_1)$	λ_1	$\theta(\lambda_1)$	λ_1	$\theta(\lambda_1)$
1.80	0.9890905	2.16	0.9977472	2.52	0.9996345
1.81	0.9895245	2.17	0.9978511	2.53	0.9996537
1.82	0.9899432	2.18	0.9979506	2.54	0.9996720
1.83	0.9903468	2.19	0.9980459	2.55	0.9996893
1.84	0.9907359	2.20	0.9981372	2.56	0.9997058
1.85	0.9911110	2.21	0.9982244	2.57	0.9997215
1.86	0.9914725	2.22	0.9983079	2.58	0.9997364
1.87	0.9918207	2.23	0.9983878	2.59	0.9997505
1.88	0.9921562	2.24	0.9984642	2.60	0.9997640
1.89	0.9924793	2.25	0.9985373	2.61	0.9997767
1.90	0.9927904	2.26	0.9986071	2.62	0.9997888
1.91	0.9930899	2.27	0.9986739	2.63	0.9998003
1.92	0.9933782	2.28	0.9987377	2.64	0.9998112
1.93	0.9936557	2.29	0.9987986	2.65	0.9998215
1.94	0.9939226	2.30	0.9988568	2.66	0.9998313
1.95	0.9941793	2.31	0.9989124	2.67	0.9998406
1.96	0.9944263	2.32	0.9989655	2.68	0.9998494
1.97	0.9946637	2.33	0.9990162	2.69	0.9998578
1.98	0.9948920	2.34	0.9990646	2.70	0.9998657
1.99	0.9951114	2.35	0.9991107	2.71	0.9998732
2.00	0.9953223	2.36	0.9991548	2.72	0.9998803
2.01	0.9955248	2.37	0.9991968	2.73	0.9998870
2.02	0.9957195	2.38	0.9992369	2.74	0.9998934
2.03	0.9959063	2.39	0.9992751	2.75	0.9998994
2.04	0.9960858	2.40	0.9993115	2.76	0.9999051
2.05	0.9962581	2.41	0.9993462	2.77	0.9999105
2.06	0.9964235	2.42	0.9993793	2.78	0.9999156
2.07	0.9965822	2.43	0.9994108	2.79	0.9999204
2.08	0.9967344	2.44	0.9994408	2.80	0.9999250
2.09	0.9968805	2.45	0.9994694	2.81	0.9999293
2.10	0.9970205	2.46	0.9994966	2.82	0.9999334
2.11	0.9971548	2.47	0.9995226	2.83	0.9999373
2.12	0.9972836	2.48	0.9995472	2.84	0.9999409
2.13	0.9974070	2.49	0.9995707	2.85	0.9999443
2.14	0.9975253	2.50	0.9995930	2.86	0.9999476
2.15	0.9976386	2.51	0.9996143	2.87	0.9999507

λ_1	$\theta(\lambda_1)$	λ_1	$\theta(\lambda_1)$	λ_1	$\theta(\lambda_1)$
2.88	0.9999536	3.24	0.9999954	3.60	0.99999964414
2.89	0.9999563	3.25	0.9999957	3.61	0.99999966975
2.90	0.9999589	3.26	0.9999960	3.62	0.99999969358
2.91	0.9999613	3.27	0.9999962	3.63	0.99999971574
2.92	0.9999636	3.28	0.9999965	3.64	0.99999973635
2.93	0.9999658	3.29	0.9999967	3.65	0.99999975552
2.94	0.9999679	3.30	0.9999969	3.66	0.99999977333
2.95	0.9999698	3.31	0.9999971	3.67	0.99999978989
2.96	0.9999716	3.32	0.9999973	3.68	0.99999980528
2.97	0.9999733	3.33	0.9999975	3.69	0.99999981957
2.98	0.9999750	3.34	0.9999977	3.70	0.99999983285
2.99	0.9999765	3.35	0.9999978	3.71	0.99999984518
3.00	0.9999779	3.36	0.9999980	3.72	0.99999985663
3.01	0.9999793	3.37	0.9999981	3.73	0.99999986726
3.02	0.9999805	3.38	0.9999982	3.74	0.99999987712
3.03	0.9999817	3.39	0.9999984	3.75	0.99999988627
3.04	0.9999829	3.40	0.9999985	3.76	0.99999989476
3.05	0.9999839	3.41	0.9999986	3.77	0.99999990264
3.06	0.9999849	3.42	0.9999987	3.78	0.99999990995
3.07	0.9999859	3.43	0.9999988	3.79	0.99999991672
3.08	0.9999867	3.44	0.9999989	3.80	0.99999992300
3.09	0.9999876	3.45	0.9999989	3.81	0.99999992881
3.10	0.9999884	3.46	0.99999900780	3.82	0.99999993421
3.11	0.9999891	3.47	0.99999907671	3.83	0.99999993920
3.12	0.9999898	3.48	0.99999914100	3.84	0.99999994383
3.13	0.9999904	3.49	0.99999920097	3.85	0.99999994811
3.14	0.9999910	3.50	0.99999925690	3.86	0.99999995208
3.15	0.9999916	3.51	0.99999930905	3.87	0.99999995575
3.16	0.9999921	3.52	0.99999935766	3.88	0.99999995915
3.17	0.9999926	3.53	0.99999940297	3.89	0.99999996230
3.18	0.9999931	3.54	0.99999944518	3.90	0.99999996521
3.19	0.9999936	3.55	0.99999948452	3.91	0.99999996790
3.20	0.9999940	3.56	0.99999952115	3.92	0.99999997039
3.21	0.9999944	3.57	0.99999955527	3.93	0.99999997269
3.22	0.9999947	3.58	0.99999958704	3.94	0.99999997482
3.23	0.9999951	3.59	0.99999961661	3.95	0.99999997678

λ_1	$\theta(\lambda_1)$	λ_1	$\theta(\lambda_1)$	λ_1	$\theta(\lambda_1)$
3.96	0.99999997860	4.10	0.99999999330	4.60	0.99999999992
3.97	0.99999998028	4.20	0.99999999714	4.70	0.99999999997
3.98	0.99999998183	4.30	0.99999999881	4.80	0.99999999999
3.99	0.99999998326	4.40	0.99999999951	4.90	1.00000000000
4.00	0.99999998458	4.50	0.99999999980	5.00	1.00000000000

Proposition XIV

The expression of $\frac{(p)}{(q)}$ that corresponds to an absolute deviation

$$k = \lambda_1 \sqrt{2mpq}$$

is

$$\frac{(p)}{(q)} = \frac{mp + \lambda_1 \sqrt{2mpq}}{mq - \lambda_1 \sqrt{2mpq}}.$$

where we assume, without loss of generality, that the deviation is the consequence of an excess of successes; therefore

$$\frac{(p)}{(q)} - \frac{p}{q} = \frac{mp + \lambda_1 \sqrt{2mpq}}{mq - \lambda_1 \sqrt{2mpq}} - \frac{p}{q} =$$

$$\frac{\lambda_1 \sqrt{2mpq}}{mq^2 - \lambda_1 q \sqrt{2mpq}} = \frac{\lambda_1 \sqrt{2pq}}{q^2 \sqrt{m} - \lambda_1 q \sqrt{2pq}}.$$

In case the deviation would be a consequence of an excess of failures, the only alteration would be to change the sign of λ_1 [in the numerator]. In any case, we have that

$$\left| \frac{(p)}{(q)} - \frac{p}{q} \right| = \left| \frac{\lambda_1 \sqrt{2pq}}{q^2 \sqrt{m} - \lambda_1 q \sqrt{2pq}} \right| > \varepsilon$$

if

$$|\lambda_1| > \frac{\varepsilon q^2 \sqrt{m}}{\sqrt{2pq} (1 \pm \varepsilon q)} \quad (5.5)$$

The probability that (5.5) holds is (Prop. XII)

$$\mathbb{P} = 1 - \theta \left(\frac{\varepsilon q^2 \sqrt{m}}{\sqrt{2pq} (1 \pm \varepsilon q)} \right).$$

Even for very small values of ε , $\mathbb{P}$ decreases very quickly towards 0, because of the factor $\sqrt{m}$.

Example:

What is the probability of winning or loosing more than 10 cents in a sequence of 200 bets on the result of coin throwing, when the money at stake in each trial is 1 cent?

Assuming that

$$p = q = \frac{1}{2}; \quad m = 200; \quad k > 10;$$

therefore, if

$$k = \lambda_1 \sqrt{2 \cdot 200 \cdot \frac{1}{2} \cdot \frac{1}{2}} = 10 \cdot \lambda_1,$$

we get

$$\lambda_1 > 1 \quad \text{and} \quad 1 - \theta(1) = \frac{16}{100} = \frac{4}{25},$$

approximately.

Proposition XV

Borel⁽¹⁹⁾ proved a generalization of the law of deviations

$$\theta(\lambda_1) = \frac{2}{\sqrt{\pi}} \int_0^{\lambda_1} e^{-\lambda^2} d\lambda$$

assuming that the extractions are done from urns with different compositions.

⁽¹⁹⁾ E. BOREL, *Eléments de la Théorie des Probabilités*, deuxième édition, p. 77.

Let

$$p_1 \text{ and } q_1, \quad p_2 \text{ and } q_2, \quad \dots \quad p_n \text{ and } q_n$$

be the compositions of n urns with white and black balls.

Assume that we extract m_1 balls from the first urn, m_2 balls from the second urn, $\dots$, m_n balls from the last urn. The most probable number of white balls in the lot will be

$$m_1 p_1 + m_2 p_2 + \dots + m_n p_n.$$

But in general this number of white balls, which we could call *normal*, is not the number of white balls that we fetch in a real experiment. We shall in general get an absolute deviation h , which is the addition of deviations $h_1, h_2, \dots, h_n$ of the white balls extracted from each urn when compared to the corresponding normal number.

Borel has proved, and using very simple arguments, that the law of deviations still holds true, provided we use as unit deviation h the square root of the sum of squares of the deviations corresponding to the different urns.

2.6 CHAPTER VI — Mathematical Expectation and Mean Value

DEFINITION 1

Let A' be a subset of the possible class A . We associate to each element $a \in A'$ one value, obtaining therefore one function that we shall denote f . The sum of products of the probability of each $a \in A'$ by $f(a)$ is *the mathematical expectation of the class A' , in what concerns the function f* .

Assuming that the probability of the elements of A' is relative to A , the mathematical expectation of the class $A' = A$ is called the *mean value* or *probable value* of the function f . ⁽²⁰⁾

We denote the mathematical expectation of A by

$$\mathbb{E}_A [f]$$

and the mean or probable value

$$\mathbb{M}_A [f].$$

As an example, in dice throwing, consider the function that associates to each face the number of dots in it. The mean value of that function is

$$\mathbb{M} = \frac{1}{6} \cdot 1 + \frac{1}{6} \cdot 2 + \frac{1}{6} \cdot 3 + \frac{1}{6} \cdot 4 + \frac{1}{6} \cdot 5 + \frac{1}{6} \cdot 6 = 3.5.$$

The mathematical expectation relative to the faces 1 and 2 is

$$\mathbb{E} = \frac{1}{6} \cdot 1 + \frac{1}{6} \cdot 2 = 0.5.$$

Proposition I

The mathematical expectation of one class is the sum of the mathematical expectations of its parts, an immediate consequence of Def. 1.

Proposition II

It is also obvious that *the mathematical expectation of a sum of functions is the sum of the mathematical expectations of the summands*.

⁽²⁰⁾ Although this distinction between mathematical expectation and mean value isn't explicitly stated in most Probability books, all authors attach to these terms the concepts we state.

Proposition III

Let A and B be two possible classes, and $A \times B$ the compound class of the two. Define

$$f(a, b) = f_1(a) \cdot f_2(b).$$

Under these assumptions, we get

$$\mathbb{E}_{A \times B} [f] = \mathbb{E}_A [f_1] \cdot \mathbb{E}_B [f_2],$$

i.e., *the mathematical expectation of the compound class is the product of the expectations of its components.*

In effect, denoting $\mathbb{P}_x$ the probability of x , we have

$$\mathbb{E}_A [f_1] = \sum_{a \in A} f_1(a) \mathbb{P}_a, \quad \mathbb{E}_B [f_2] = \sum_{b \in B} f_2(b) \mathbb{P}_b$$

$$\mathbb{E}_{A \times B} [f] = \sum_{(a,b) \in A \times B} f(a, b) \mathbb{P}_{(a,b)}$$

As

$$\begin{aligned} \sum_{(a,b) \in A \times B} f(a, b) \mathbb{P}_{(a,b)} &= \sum_{(a,b) \in A \times B} f_1(a) \cdot f_2(b) \mathbb{P}_a \mathbb{P}_b = \\ &= \sum_{a \in A} f_1(a) \mathbb{P}_a \cdot \sum_{b \in B} f_2(b) \mathbb{P}_b \end{aligned}$$

we conclude that

$$\mathbb{E}_{A \times B} [f] = \mathbb{E}_A [f_1] \cdot \mathbb{E}_B [f_2].$$

□

DEFINITION 2

Let X denote a free point or an image point varying in some region containing the possible region A ; and $\mathbb{P}_x$ denote its probability law relative to A . Consider a function $\varphi(X)$, defined in A , of the coordinates of point X . If $A' \subset A$, the *mathematical expectation of the region A' relative to the function $\varphi(X)$* is

$$\mathbb{E}_{A'} [\varphi(X)] = \int_{A'} \mathbb{P}_x \varphi(X) \, da.$$

In the particular case $A' = A$, that number is the *mean value* or *probable value* of $\varphi(X)$.

Proposition IV

It is evident that if A' can be partitioned in pairwise disjoint sets

$$A' = A_1 \cup A_2 \cup \cdots \cup A_n$$

then

$$\mathbb{E}_{A'} [X] = \mathbb{E}_{A_1} [X] + \mathbb{E}_{A_2} [X] + \cdots + \mathbb{E}_{A_n} [X].$$

Proposition V

It is also obvious that

$$\mathbb{E} [\varphi_1(X) + \varphi_2(X) + \cdots] = \mathbb{E} [\varphi_1(X)] + \mathbb{E} [\varphi_2(X)] + \cdots.$$

Proposition VI

Let X denote a varying point in A , Y a varying point in B . Similarly to what we stated in Prop. III ⁽²¹⁾,

$$\mathbb{E}_{A \times B} [\varphi(X) \cdot \psi(Y)] = \int_{A \times B} \mathbb{P}_x \mathbb{P}_y \varphi(X) \psi(Y) \, d(a, b) =$$

⁽²¹⁾ *Editors' note:* Perhaps this reference to Proposition III is to include the independence hypotheses in this Proposition, because otherwise it would not be correct.

$$\begin{aligned}
&= \int_A \mathbb{P}_X \varphi(X) \, da \cdot \int_B \mathbb{P}_Y \psi(Y) \, db = \\
&= \mathbb{E}_A [\varphi(X)] \cdot \mathbb{E}_B [\psi(Y)].
\end{aligned}$$

Proposition VII

Let X denote a varying point in some region containing A , φ denote some function of its coordinates, and $\mathbb{P}_X$ the probability law of X relative to A .

We have defined

$$\mathbb{E}_A [\varphi(X)] = \int_A \mathbb{P}_X \varphi(X) \, da.$$

Writing $\varphi(X) = Z$, we get

$$\mathbb{E} = \int_A Z \mathbb{P}_X \, da = \int_{z_0}^{z_1} Z \int \mathbb{P}_X \, da$$

where the second integral is to be computed for the values of A for which Z is between z and $z + dz$. But this integral is, by definition, $\mathbb{P}_Z$, the probability that Z lies between z and $z + dz$, and its value may be denoted

$$\mathbb{P}_Z(z) \, dz,$$

where $\mathbb{P}_Z(z)$ is the probability law of Z . Therefore,

$$\mathbb{E}_A [\varphi(X)] = \int_{z_0}^{z_1} Z \mathbb{P}_Z(z) \, dz = \mathbb{E}_Z(Z).$$

Later on, we shall present an example showing this proposition usefulness in the computation of mathematical expectations.

Proposition VIII

Given the probability law of the variable Z , *the mathematical expectation of any function $\varphi(Z)$ of Z is*

$$\mathbb{E}_Z [\varphi(Z)] = \int_{z_0}^{z_1} \varphi(z) \mathbb{P}_Z(z) dz.$$

To prove it let

$$f(X) = \varphi(Z)$$

in the previous proposition, computing the second integral in the region where Z takes values between z and $z + dz$.

Proposition IX

The mathematical expectation of a region A' , in what concerns a constant function c , is the product of that constant by the probability of the region, i.e.,

$$\mathbb{E}_{A'}(c) = c \cdot \mathbb{P}(A'),$$

a result that follows immediately from the definition.

In particular, the mean value of a constant is that constant, since

$$\mathbb{P}(A) = 1 \quad \implies \quad \mathbb{M}(c) = c.$$

Proposition X

If $f(X)$ is a positive function and its mean value is smaller than a number δ , however small this δ may be, the probability, that $f(X)$ is greater than a given m , is lower than $\frac{\delta}{m}$.

In effect, if A' is the region where

$$f(X) \geq m,$$

it follows that

$$\int_A f(X) \mathbb{P}_x da = \int_{A-A'} f(X) \mathbb{P}_x da + \int_{A'} f(X) \mathbb{P}_x da \geq$$

$$\geq \int_{A'} f(X) \mathbb{P}_x da \geq$$

$$\geq m \cdot \mathbb{P}(A');$$

therefore

$$m \cdot \mathbb{P}(A') \leq \delta$$

and

$$\mathbb{P}(A') \leq \frac{\delta}{m}.$$

□

Problem

Consider a random plane open polygonal line with sides of lengths

$$l_1, l_2, \dots, l_n,$$

denote A and B its endpoints, and d the length of $\overline{AB}$.

Compute

$$\mathbb{M}_n(d^2).$$

Solution

To compute $\mathbb{M}_n(d^2)$ we can use the probability law of d (Prop. VIII). But, by Prop. VII, instead of the probability law of d we may use the probability law of any convenient point X tied to it, for instance the equivalent point to the polygonal line. Consider first the simplest case of a polygonal line with only one side of length l_1 ; from Prop. IX, it follows that

$$\mathbb{M}_1(d^2) = l_1^2.$$

Let us now consider the case of a random polygonal line with two sides; from Prop. VIII it follows that

$$\begin{aligned} \mathbb{M}_2(d^2) &= \int_0^{l_1+l_2} \mathbb{P}(d) d^2 \, dd = \\ &= \int_0^\pi \frac{1}{\pi} (l_1^2 + l_2^2 - 2l_1 l_2 \cos \alpha) \, d\alpha = \\ &= l_1^2 + l_2^2, \end{aligned}$$

where in the above computation α denotes the angle between the two sides of the polygonal line.

Let us now assume the induction hypothesis that in the case of a polygonal line with i sides we have

$$\mathbb{M}_i(d^2) = l_1^2 + l_2^2 + \cdots + l_i^2$$

in order to prove that we also get

$$\mathbb{M}_{i+1}(d^2) = l_1^2 + l_2^2 + \cdots + l_i^2 + l_{i+1}^2.$$

Denoting δ the length of the segment from the origin of l_1 with the end-point of l_i , from the induction hypothesis

$$\mathbb{M}_i(\delta^2) = l_1^2 + l_2^2 + \cdots + l_i^2.$$

Denoting $\alpha_1, \alpha_2, \dots, \alpha_i$ the angles in the articulations of the polygonal line, d^2 is a function $f(\alpha_1, \alpha_2, \dots, \alpha_i)$ of those angles, and the mean value we want to compute is of the form

$$\begin{aligned} \mathbb{M}_{i+1}(d^2) &= \int f(\alpha_1, \alpha_2, \dots, \alpha_i) \frac{d\alpha_1}{2\pi} \cdot \frac{d\alpha_2}{2\pi} \cdots \frac{d\alpha_i}{2\pi} = \\ &= \int \frac{d\alpha_1}{2\pi} \cdot \frac{d\alpha_2}{2\pi} \cdots \frac{d\alpha_{i-1}}{2\pi} \int_0^{2\pi} f(\alpha_1, \alpha_2, \dots, \alpha_i) \frac{d\alpha_i}{2\pi} = \\ &= \int \frac{d\alpha_1}{2\pi} \cdot \frac{d\alpha_2}{2\pi} \cdots \frac{d\alpha_{i-1}}{2\pi} \int_0^{2\pi} \frac{1}{2\pi} (\delta^2 + l_{i+1}^2 - 2\delta l_{i+1} \cos \alpha_i) d\alpha_i = \\ &= \int \frac{d\alpha_1}{2\pi} \cdot \frac{d\alpha_2}{2\pi} \cdots \frac{d\alpha_{i-1}}{2\pi} (\delta^2 + l_{i+1}^2) = \\ &= \mathbb{M}_i(\delta^2 + l_{i+1}^2) = \\ &= \mathbb{M}_i(\delta^2) + \mathbb{M}_i(l_{i+1}^2) = \\ &= l_1^2 + l_2^2 + \cdots + l_i^2 + l_{i+1}^2. \end{aligned}$$

□

From this it follows that

$$\mathbb{M}_n(d^2) = l_1^2 + l_2^2 + \cdots + l_n^2.$$

If the sides of the random polygonal line are all of equal length, denoting L its perimeter we have

$$\mathbb{M}_n(d^2) = n l_1^2 = \frac{L^2}{n}.$$

From this, letting $n \rightarrow \infty$, we get that:

The probable value of the square of the distance between the endpoints of a random plane flexible curve is zero, whatever the length of the curve, provided this is finite.

Proposition XI

The mathematical expectation can be computed, in some cases, without previous computation of the summands involved in its definition, or of the probability law. We exemplify using a curious example, an alternative way of solving Buffon's needle problem. As remarked in the observation in page 64, dividing the needle into equal parts, each of those would have equal probability of intersection one of the separation lines. If those parts, instead of being collinear, have different relative positions, forming a polygonal line, each part will still have the same probability, an immediate consequence of the definition of randomly throwing a variable form figure.

Let us assume, now, that we associate the same number, one, to each intersection of one side of the polygonal line with one of the parallel lines. The mathematical expectation of each part will be the probability that it hits one parallel. The sum of all these expectations is proportional to the number of parts, and thus proportional to the perimeter of the polygonal line. And this is so, whatever the polygonal line, rigid or articulated, and whatever the length of its sides. In the limit, we can still say that the integral of the elementary expectations of a rigid or flexible curve is proportional to

its length:

$$\mathbb{E}(l) = K l,$$

where K is independent of the form, nature and perimeter of the figure.

To determine K , let us consider a needle which is a circumference whose diameter is the distance between two consecutive parallels.

Randomly throwing this circumference in the plane of the parallels, it will always have two points in common with the system of parallels, [either] because it intersects one parallel in two points, [or because it is tangent to two consecutive parallels]⁽²²⁾; therefore

$$\mathbb{E}(\pi a) = K \pi a = 2$$

and

$$K = \frac{2}{\pi a}.$$

From that we get

$$\mathbb{E}(l) = \frac{2l}{\pi a}.$$

This is the value we had obtained for the probability that the needle intersects one of the parallels, when solving Buffon's needle problem, when $l \leq a$. In fact, when $l \leq a$, the needle either intersects one of the parallels in one point, or it doesn't, and therefore the mathematical expectation is the probability of the event that it intersects one of the parallels. So, another way of solving Buffon's needle problem is via the exploitation of the concept of mathematical expectation.

If the parallel lines are substituted by equidistant circumferences with the same center, the mathematical expectation would still be the same, but we couldn't state anything about the probability of intersection of a linear segment with one of the circumferences.

⁽²²⁾ *Editors' note:* Pacheco d'Amorim forgets the second possibility.

Proposition XII

Let us consider an experiment with two possible outcomes, [success or failure,] and denote p and q the corresponding probabilities. Consider the function $f(\text{success}) = a$, $f(\text{failure}) = b$.

The mean value of this function is

$$\mathbb{M} = ap + bq.$$

Repeat the experiment a large number of times, and let's assume that (p) times we get success and (q) times we get failure.

Denote

$$\bar{x} = \frac{(p) a + (q) b}{(p) + (q)}$$

the arithmetic mean of the observed values of the function we have defined.

The expression

$$|\mathbb{M} - \bar{x}| = \left| ap + bq - \frac{(p) a + (q) b}{(p) + (q)} \right| \leq$$

$$\leq |a| \cdot \left| p - \frac{(p)}{(p) + (q)} \right| + |b| \cdot \left| q - \frac{(q)}{(p) + (q)} \right|;$$

goes to zero when $(p) + (q)$ increases; more precisely, the probability that $|\mathbb{M} - \bar{x}|$ is smaller than ε , however small ε is, goes to one when $(p) + (q) \rightarrow \infty$ (3rd Bernoulli's theorem).

What we just established for a random experiment with two possible outcomes is valid for an experiment with any number of possible outcomes.

What we have proved about mean values is valid, with the necessary adaptations, for the mathematical expectation of any class. Therefore:

The mathematical expectation of a finite class of numerical elements i.e., of elements to which we associated a number, is the limit, when the number of

experiments goes to infinity, of the sum of the observed numbers in that class when we perform repeated experiments, divided by the number of experiments.

The practical importance of the mathematical expectation comes from the above statement.

If the class considered in the above statement is the *total possible class*, the *mathematical expectation* is the *mean value*, and the above proposition becomes:

The mean value of a function that can assume a finite number of values is the limit of the averages of the observed values of that function, in repeated experiments, when the number of experiments goes to infinity.

This is the reason why averages are so important in the applications of Probability.

All those propositions can be generalized for functions continuously varying in some region.

For instance, considering the mean value, that only formally differs from the mathematical expectation:

Let A be the region, $f(X)$ the function, $\mathbb{P}_x$ the probability law.

The mean value of the function is

$$\mathbb{M} = \int_A f(X) \mathbb{P}_x da.$$

Let's partition A in n parts A_i , $i = 1, \dots, n$. The mathematical expectation of each of the parts is

$$\begin{aligned} \mathbb{E}_i &= \int_{A_i} f(X) \mathbb{P}_x da = \\ &= f(X_i) \int_{A_i} \mathbb{P}_x da = \end{aligned}$$

$$= f(X_i) \cdot \mathbb{P}(A_i), \quad (6.1)$$

since $\mathbb{P}_x$ is always a positive function, and thus we can use the mean value 1st theorem. In (6.1), X_i denotes the value of the function f in a point $X_i \in A_i$, and $\mathbb{P}(A_i)$ the probability of the region A_i .

On the other hand, we have

$$\mathbb{M} = \sum \mathbb{E}_i = \sum f(X_i) \mathbb{P}(A_i).$$

Let us now assume that we group the observed values of the function f in classes corresponding to the partial regions A_i and that, inside each of those regions, we substitute $f(X)$ by $f(X_i) + \varepsilon_i$, where $\varepsilon_i = o(f(X_i))$, because of the assumed continuity of f . Thus $\varepsilon_i \rightarrow 0$ when A_i decreases to $\emptyset$.

Let us consider the average of the observed values $f(x_k)$ of $f(X)$ decomposed into two summands, the first one corresponding to the values to the values $f(X_i)$ and the second one corresponding to the values ε_i . When the number of experiments goes to infinity, the first summand goes to $\sum f(X_i) \mathbb{P}(A_i)$, and therefore to $\mathbb{M}$, whatever the partition of A (Prop. XII). The second summand, as we can partition A in subsets whose measure is as small as we want, converges to zero.

Therefore the limit of the averages of the observed values $f(x_k)$ of $f(X)$ exists, and it is the *mean value* or *probable value* of $f(X)$. $\square$

As we have seen in the previous chapter, the probability that the absolute value of the relative deviation, denoted $|\Lambda|$, is less than λ_1 is

$$\theta(\lambda_1) = \frac{2}{\sqrt{\pi}} \int_0^{\lambda_1} e^{-\lambda^2} d\lambda.$$

Therefore the probability law of the relative deviation Λ is

$$\mathbb{P}_{\Lambda}(\lambda) = \frac{1}{\sqrt{\pi}} e^{-\lambda^2}$$

for all real λ .

The mean value of Λ is

$$\mathbb{M}(\Lambda) = \frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} e^{-\lambda^2} \lambda \, d\lambda = 0.$$

The mean value of $|\Lambda|$ is

$$\begin{aligned} \mathbb{M}(|\Lambda|) &= \frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} e^{-\lambda^2} |\lambda| \, d\lambda = \\ &= \frac{1}{\sqrt{\pi}} \int_0^{\infty} 2 e^{-\lambda^2} \lambda \, d\lambda = \\ &= \frac{1}{\sqrt{\pi}} \left(-e^{-\lambda^2} \right) \Big|_0^{\infty} = \\ &= \frac{1}{\sqrt{\pi}}. \end{aligned}$$

The mean value of Λ^2 is

$$\begin{aligned} \mathbb{M}[\Lambda^2] &= \frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} e^{-\lambda^2} \lambda^2 \, d\lambda = \\ &= \frac{1}{\sqrt{\pi}} \int_0^{\infty} \lambda \cdot e^{-\lambda^2} 2\lambda \, d\lambda = \\ &= \frac{1}{\sqrt{\pi}} \left(-\lambda e^{-\lambda^2} + \int e^{-\lambda^2} d\lambda \right) \Big|_0^{\infty} = \end{aligned}$$

$$= \frac{1}{\sqrt{\pi}} \cdot \frac{\sqrt{\pi}}{2} = \frac{1}{2}.$$

Therefore

$$\frac{\mathbb{M}[\Lambda^2]}{[\mathbb{M}(|\Lambda|)]^2} = \frac{\frac{1}{2}}{\frac{1}{\pi}} = \frac{\pi}{2}.$$

Proposition XII from this chapter confers a remarkable interpretation to this result: it is possible, using random throws, to rectify the circumference.

2.7 CHAPTER VII — Conclusion

In the preceding chapters, we have investigated the probability of events which can be thought of as random extractions from a finite set, or as random throws in a region, on the assumption that we are, ourselves, the agents of the random selection, and that the set or the region are qualitatively and quantitatively specified. We also described how a sequence of random extractions or of random throws can be reduced to a single extraction from a finite set or to a single throw in a region.

We now describe how the scope of Probability can be broadened, using the principles formerly established.

For clarity, we start with a classification of the facts we want to investigate.

To do so, we shall admit the possibility that someone like us, or even essentially diverse agents can, *in some circumstances*, perform random extractions (or random throws) with analogous outcomes to those performed by us.

Once this has been accepted as admissible, we shall consider three groups of phenomena whose probability can be investigated.

In the first group, we enclose the phenomena that can be assimilated either to random extractions or to random throws performed by us; in the second group, phenomena that can be viewed as random extractions or as random throws, but performed by someone similar to us; in the third group, phenomena that can be assimilated either to random extractions or to random throws, done by some essentially different agent.

Each of those three groups will be further divided into two subgroups. In the first one, we consider the phenomena that can have a finite number of possible outcomes. In the second one, the phenomena whose possible outcomes conceptually form a continuous region.

In each of those subgroups, we consider three possible situations. In the first subgroup of each of the three groups, the three possible situations are: the set of phenomenon on outcomes is qualitatively and quantitatively known (first case); it is qualitatively known, but quantitatively unknown (second case); or it is unknown, both qualitatively and quantitatively (third case).

In the second subgroup of each group, we may know the probability law of the phenomenon, and the corresponding support (first case); or we may know the support, but ignore the probability law (second case); or we may ignore both the probability law and its support (third case).

The criterion used in the first classification is the nature of the agent of random extractions or of random throws. The second classification is done on the nature of the phenomenon; the third classification is based on our degree of knowledge of the phenomenon.

As we have seen in the *Introduction* to these *Elements of Probability Cal-*

culus, we consider a phenomenon which can be identified to a random extraction (or to a random throw) done by us, ourselves, from qualitatively and quantitatively known finite set (or in a qualitatively and quantitatively known bounded region) as known, in the sense that everything is well specified. This is the description of our *standard phenomenon*, *standard model*, or *elementary fact*.

Only the phenomena amenable to the standard model can be the object of Probability.

We shall start our program of standardization of phenomena with the

1st Group

which, as we have seen, is the one in which random extractions or random throws are performed by us. In this group, as we said before, we must consider two subgroups, of which the

1st SUBGROUP

contains those phenomena having a finite number of outcomes. Such phenomena may be conceptualized as random extractions from finite discrete sets, or to random throws in bounded regions divided into a finite number of parts.

Its

1st case

has been studied in Chapters I, II and III of these *Elements of Probability Calculus*. It includes the *standard model*; thus, the starting point of our standardization program must investigate how to reduce to it the

2nd case

which deals with phenomena amenable to random extractions from sets qualitatively known but quantitatively unknown.

In this situation, the aim of the reduction procedure is, therefore, the quantitative determination of the set from which the random extractions are done.

This can be done with high probability and precision, i.e. with the accuracy we wish in the approximation, and with as probability as large ⁽²³⁾ as we want, insofar as it is feasible to perform as many trials as needed (J. Bernoulli's 3rd theorem).

The determination is therefore *approximate* and *probabilistic*. But the probability that the approximation produces an error whose absolute value is lower than ε , however small ε is, will approach 1 as much as we want. Thus this 2nd case is separated from the 1st case by the same hiatus that separates *probability* from *certitude*. ⁽²⁴⁾

⁽²³⁾ I.e., $1 - \varepsilon$, with ε as small as desired.

⁽²⁴⁾ Certitude is the probability of extracting one white ball from an urn containing only white balls. Laplace "Quand tous les cas sont favorables à un évènement, sa probabilité se change en certitude, et son expression devient égale à l'unité. Sous ce rapport, la certitude et la probabilité sont comparables, quoiqu'il y ait une différence essentielle entre les deux états de l'esprit, lorsqu'une vérité lui est rigoureusement démontrée, ou lorsqu'il aperçoit encore une petite source d'erreur." (LAPLACE, *Essai Philosophique sur les Probabilités*). In Jacob Bernoulli's view, there is no *essential difference* between probability and certitude: "Certitudo rerum, spectata in ordine ad nos, non omnium eadem est, sed multipliciter variat secundum majis et minus. Illa de quibus revelatione, ratione, sensu, experientia, ἀνισοψία aut aliter ita constat, ut de eorum existentia vel futuritione nullo modo dubitare possimus, summa et absoluta certitudine gaudent. Caetera omnia imperfectiorem ejus mensuram in mentibus nostris obtinent, majorem minoremve, prout plures vel pauciores sunt probabilitates, quae suadent rem aliquam esse, fore aut fuisse.

The identification cannot be but *probable*, but, having no better choice we accept it, since probabilistic knowledge is a useful degree of knowledge.

“All the science of philosophers and all the prudence of politicians deals with” the evaluation of the probability of events ([...] *in quo solo omnis Philosophi sapientia et Politici prudentia versatur*) ⁽²⁵⁾

The identification of the

3rd case

with the 1st one is done in a similar fashion, with the extra task of identifying qualitatively the set of possible outcomes. But the identification procedure is, in all steps, similar to the former one.

Let us now consider the

2nd SUBGROUP

Probabilitas enim est gradus certitudinis, et ab hac differt ut pars a toto.” (J. BERNOULLI, *Ars Conjectandi*, Pars Quarta, Chap. I)

(Laplace’s text: “When all cases are favorable to an event, its probability becomes certitude, and its value is unity. In this perspective, probability and certitude are comparable, although there is an essential difference between the two states of mind, resulting from the rigorous proof of a true statement, or from an argument where a possible source of error is still perceived.”)

Bernoulli’s text: “Our view on the certitude of things is not always the same, it varies, being high in what concerns some, low in respect to others. We have complete and absolute certitude on those things that we know by revelation, by the exercise of the intellect or of the senses, by experience, by direct observation, or otherwise constated, and in no way doubt that they will exist or occur in the future. Under other circumstances, our mind assigns to things some lower degree of belief, higher or lower according to whether we judge large or small the probability that they exist, existed or will exist.

Probability is, thus, a degree of certitude, and differs from it as a part differs from the whole.”)

⁽²⁵⁾ J. BERNOULLI, *Ars Conjectandi*, Pars Quarta, Chap. II.

which, as we have seen, contains the phenomena whose set of possible outcomes is infinite, building up a *continuous* that we shall assume to be of the *second species*, in the terminology of H. Poincaré ⁽²⁶⁾. We shall therefore assume that to each possible outcome of the phenomenon we associate one point in a space with the convenient number of dimensions. The

1st case

from this subgroup, characterized by the fact that its probability law and the corresponding support are known, has been dealt with in Chapters II, III and IV. We may therefore proceed to the

2nd case

in which the support is known, but the probability law itself is not known. The reduction of this case to the previous one consists, therefore, in the determination of the probability law.

When we execute a sequence of random throws in some region and we observe directly the points that result from each trial, the ratio of the number of points that lie in a given subregion to the number of points that lie in another region of the same size will converge to one as the number of trials increases (Bernoulli's 3rd theorem). In other words, as the number of trials increases, the distribution of the points approaches the uniform distribution.

But if, instead of directly observing the points randomly chosen, our goal is to study the law of the point's projections or of some other image point, the probability law is no longer uniform. But, according to the above mentioned Bernoulli's 3rd theorem, the distribution of these image points will be governed by the corresponding probability law. The observed points will

⁽²⁶⁾ H. POINCARÉ, *La Science et l'Hypothèse*, Chap. II.

be concentrate in the neighborhood of the maxima of such probability law. When the probability law is known, we can *forecast* the distribution of the points projections or of other image points of the randomly thrown points, and the probability that this forecast agrees with the reality is increasing with the number of points.

The other way round, observing a large number of points, we can determine the corresponding probability law, with a probability as large as we wish; more precisely, we may compute the value of the integral, in any given interval from its support, of that unknown probability function (Bernoulli's 3rd theorem).

From this fact we get two methods to determine the unknown law:

1st method

It may happen that some reasons which are inherent to the nature of the phenomenon we are studying point towards the adoption, *a priori*, of some specified probability law — as it happens, for instance, when we are dealing with observation errors. In order to decide whether this is so, we perform a large number of sequences of trials, each sequence with a number large enough of trials, so that the probability that its distribution doesn't agree with the *a priori* law, provided this one is the true one, is negligible.

The ratio of the number of sequences whose empirical law matches with the hypothesized probability law to the total number of series approaches a number that (as we have already seen in the 2nd case of the 1st subgroup) we may call the probability of that law. If in our view that probability is large enough, the *a priori* law is maintained; otherwise, it is rejected.

The

2nd method

is the following: partition the region — which is assumed to be known — where the observed point varies into a large enough ⁽²⁷⁾ number of subregions.

Once this has been done, we perform a large enough sequence of random throws in that region. The ratio of the number of points we observe in each subregion to the total number of random throws gives us, with the degree of probability and approximation we want, the integral of the unknown function in each of the partial subregions.

Dividing each number computed as described by the size of the corresponding subregion, we obtain an equal number of points of the function we wish. This, however, isn't sufficient to determine the probability law.

In fact, there is an infinite number of functions that, when integrated in the considered subregions, furnish the same results, namely the ones computed as described above. How to select one among this infinite number of possibilities?

All the functions whose integrals in the subregions match those numbers are equally plausible, since they have the same degree of agreement with Bernoulli's 3rd theorem. Among them, we choose the simpler one, the one which is more adequate for our goals, if no deeper reasons can guide our judgement.

⁽²⁷⁾ The number of subregions and the number of trials can be tuned so that we can achieve the accuracy we wish in the final results.

As in the preceding subgroup, the

3rd case

reduces to the 2nd one. The observed points are distributed in a region of arbitrary boundary.

We can even assume that the support of the probability law is unbounded in all directions; the specification of the law will indicate, afterwards, which subregions do have null probability; in other words, the law itself will limit its domains of existence.

We might as well assume, as in the 3rd case of the 1st subgroup, that the phenomenon was qualitatively unknown; the quantitative determination of its probability law would then specify the events of null probability.

In that perspective, these two cases are identical. We insist however in distinguishing the two cases, for the sake of clarity in the exposition.

Note

In all that has been said it was implicitly assumed that the sets and regions in which the random operations are performed remain invariable during the experiments, both qualitatively and quantitatively. Otherwise, no sound conclusions could be reached, unless the law of such variation was known, thus allowing the necessary *corrections*.

Let's now investigate the

2nd Group

of phenomena, starting with an explanation of the hypothesis assumed for its construction, and of its meaning.

The proposition *to extract, at random, one element from a set* has for us a precise meaning, when we are the agents of such random selection.

But when the agent of the selection is someone else, this proposition is ambiguous, in the sense that it has no different meaning from the proposition *to extract one element from a set*. In some situations, however, it is legitimate to retain the expression *random*, essential for our study.

What are the conditions needed to accept that the selection *has been done at random*?

First of all, the distribution of the elements in the class where the selection is done must be ignored.

Further, the extraction has to be done in such a way that it is impossible to *forecast* either which element will be selected, or its quality; this *impossibility of forecasting* must hold in what concerns not only *us, ourselves*, as *anyone similar to us*.

Hence, if the extractions are done by someone similar to us, according to the above requirements, nothing opposes, *a priori*, the hypothesis that the outcome of such extractions has the same informative value that would have the outcome of a random selection done by us.

For instance, assuming that 90% of the balls in one urn are white and the remaining 10% are black, we would favor a bet that the outcome of a random extraction would be white ball, in case we would be the agent of the random extraction. But we would surely stick to this bet, the extraction being performed by someone else, in case we would be satisfied that the circumstances of the extraction were as described: namely, the agent of the extraction was unable to *predict* the element he would extract, and ignorant of how the balls are mixed in the urn. All games of chance take those assumptions for granted; shuffling the deck before dealing the cards, and the fact that the back

of all cards in a deck is the same, fulfills the above requirements.

Whenever the above requirements are not fulfilled, the phenomenon is out of the scope of the Science of Probability.

We must reject the hypothesis that another agent is doing random extractions whenever the outcome doesn't agree with Bernoulli's and similar laws. In fact, as in any other Science, external circumstances can cheat us.

This is clearly demonstrated in an anecdote that Bertrand reports about Galiani, philosopher and economist of the XVIII century:

“Un jour, à Naples, un homme de Basilicate, en présence de l'abbé Galiani, agita trois dés dans un cornet et paria d'amener raffle de 6; il l'amena sur-le-champ. Cette chance est possible, dit-on; l'homme réunit une seconde fois, et l'on répéta la même chose; il réunit les dés dans le cornet trois, quatre, cinq fois, et toujours raffle de 6. *Sangue di Bacco*, s'écria l'abbé, *les dés sont pipés*”.⁽²⁸⁾

As we have just discussed, the phenomena in the second group can be reduced to phenomena in the first group, and in that case we can view them as studied, or they are not amenable to phenomena in the first group, and they do not fall in the scope of Probability.

In what we said above about phenomena in the second group, we skipped the phenomena in the second subgroup. In fact, this is a reasonable option, in view of the analogy between the qualitative composition of a finite set and the support of the probability law, and between the quantitative composition

⁽²⁸⁾ J. BERTRAND, *Calcul des Probabilités*, Préface.

“One day, at Naples, a man from Basilicate, the abbot Galiani being present, shacked a cornet with three dice and bet he would throw a 6; he did, in fact, throw a 6. This is possible, no doubt; but he made a second throw, obtaining once again 6, and the same happened in a third throw, a fourth throw, a fifth throw — always he got a 6. *Sangue di Bacco*, said the abbot, *these dice are loaded*.”

in the former case and the form of the probability law.

Let us now consider the

3rd Group

of phenomena.

When we execute a sequence of random extractions (or of random throws), the elements we get are unordered, with runs that apparently do not follow any law. Otherwise, we could *predict* them, and this is incompatible with what intuition tells us about random extractions. Henceforth, when the outcomes of some phenomenon do occur in an unordered fashion, this hazardous character gives us a vague feeling that it is governed by *chance*.

Can we identify that vague feeling of chance intervention with the random character of sequences of random extractions we have studied? In other words, can we quantitatively determine the set of the qualitative outcomes of the said phenomenon, so that we may *assume* that the outcomes produced by natural causes have the same random character possessed by random extractions done by us in that set?

Or, if only part of the possible outcomes of the phenomenon is known (and this may always be *assumed*), can we in all cases determine qualitatively and quantitatively the corresponding set?

We shall hypothesize that this is so.

How can we justify such a hypothesis?

There is one way out: we must verify whether the conclusions that we can deduce from it conform to the observed facts.

Bernoulli's theorems, and those analogous to them, are the appropriate tools to judge whether this is so.

The first step is to analyze what we get in the light of Bernoulli's 3rd theorem, since it is useful in the qualitative and quantitative determination of the law.

In fact, Bernoulli's 3rd theorem tells us that the probability that the relative frequency of each possible outcome approaches, as much as desired, the probability of that outcome is *always increasing*. In other words: if after experiment we divide the number of times we got each of the outcomes by the total number of experiments, we obtain relative frequencies that converge to the probabilities of the outcomes. The probability that those numbers get closer and closer to the corresponding limits increases with the number of trials.

So, while the number of trials is moderate, those numbers will fluctuate showing some irregularity (since the probability that they are close to their limits is small); but their fluctuation will be smoother and smoother when the number of trials increases.

If the observed facts do not agree with this pattern of behavior, this must be interpreted as an indication that the hypothesis we assumed is wrong, either these phenomena have a pattern that is not identifiable with random extractions, or the phenomenon is varying in time.

If the facts are in agreement with Bernoulli's 3rd theorem and with the analogous theorems we discussed in Chapter V, we may maintain the hypothesis that the observed outcomes of the phenomenon behaved like random extractions, during the period taken in the verification.

So, while there is no reason to question the stability of the process producing this phenomenon, there is no substantial reason to doubt that hypothesis. Even if the process changes, the hypothesis may still be acceptable, but further investigation has to be carried out, to make the necessary adaptations

and amendments, and as often as required to determine the eventual pattern of variation.

Parte III

A CONSTRUÇÃO DE PACHECO D'AMORIM

A Construção de Diogo Pacheco d'Amorim

Na terceira parte deste trabalho vamos apresentar uma análise detalhada de todos os assuntos apresentados por Pacheco d'Amorim na sua tese *Elementos de Cálculo das Probabilidades*, com a qual, em 4 de Abril de 1914, obteve o grau de Doutor.

Principiaremos com uma pequena descrição da obra criada por Diogo Pacheco d'Amorim ao longo da sua vida, para posteriormente concentrarmos a nossa atenção no principal objectivo deste trabalho, a análise da tese de doutoramento de Pacheco d'Amorim. Com o intuito de facilitar a exposição desta crítica, tornando mais facilmente perceptível os nossos comentários, seguiremos a mesma ordem de conteúdos que Pacheco d'Amorim usou na sua tese. Assim, à medida que expusermos a sua construção de Probabilidade, exibiremos os nossos comentários e compararemos as suas concepções com as de outros autores seus contemporâneos, nomeadamente da escola francesa que é a sua principal fonte (a avaliar pelas poucas referências que efectua).

Vamos, desta forma, percorrer e dissecar as principais concepções consagradas por Diogo Pacheco d'Amorim nesta obra ímpar da História da Matemática em Portugal.

O Autor

Diogo Pacheco d'Amorim, filho de Manuel Pacheco e Juliana de Amorim, nasceu em Troviscoso, concelho de Monção e distrito de Viana do Castelo, a sete de Novembro de 1888, tendo falecido em nove de Fevereiro de 1976 em Coimbra com 87 anos⁽¹⁾.

Pacheco d'Amorim fez o ensino primário em Monção e o ensino secundário em Braga, tendo-se licenciado em Coimbra com vinte valores e pouco depois, em quatro de Abril de 1914, defendido a sua Tese de Doutoramento. Em 1919, com pouco mais de 30 anos, já era Professor Catedrático da Faculdade de Ciências da Universidade de Coimbra, tendo proferido a *Oração da sapientia* na abertura solene da Universidade de Coimbra do ano lectivo 1923-1924⁽²⁾, onde leccionou diversas disciplinas, tais como Análise, Geometria, Cálculo das Probabilidades, Mecânica e Astronomia, tendo-se jubulado em 1958 por limite de idade. A sua influência na Universidade de Coimbra não se restringiu à leccionação de disciplinas na área das Matemáticas puras

⁽¹⁾ Podemos encontrar diversas informações sobre a sua vida e obra em Rodrigues (1992), Carvalho e Gomes (1994), Pestana (1994) e Pestana e Velosa (2001).

⁽²⁾ Esta oração foi publicada recentemente (2002) na *Antologia de Textos essenciais sobre a História da Matemática em Portugal* editada pela Sociedade Portuguesa da Matemática. É estimulante ler as palavras de Pacheco d'Amorim, quer o seu enorme elogio à Matemática, quer a sua visão sobre o ensino da Matemática da época, que bem se poderia aplicar à actualidade.

ou aplicadas, como ilustra o facto de ter regido o curso de Economia Política da Faculdade de Direito da Universidade de Coimbra (1936–1938).

Ao longo da sua vida publicou diversas obras e artigos em variadas áreas, tais como a Matemática, a Economia, a Política, a Religião, a História, a Filosofia ou até a culinária. De entre os seus escritos, para ilustrar a diversidade da sua obra, refiram-se alguns títulos: *Elementos de Cálculo de Probabilidades* (1914), *Compêndio de Geometria* (1920)⁽³⁾, *Sobre o determinante de Ronsky* (1923), *Compêndio de Aritmética Racional* (1931), *A obra matemática de Newton* (1932), *O teorema da probabilidade composta na teoria da probabilidade dos conjuntos* (1933), *A Matemática e a Economia Política* (1934), *Relações de Portugal com a Inglaterra* (1937), *Finanças e economia: comentários, 1936* (1937), *O poderio económico e financeiro da Inglaterra* (1939), *Do erro e da sua eliminação* (1940), *Capitalismo* (1941), *Sur la loi binomiale* (1942), *Princípios fundamentais da pensamento marxista* (1942), *Princípios fundamentais da sociologia católica* (1944), *Da cultura geral do universitário* (1951), *Esquemas indefinidos de Poisson* (1951), *Direito e deveres da família na educação* (1952), *Projecção nacional do Centro Académico de Democracia Cristã* (1952), *O jovem católico perante o casamento* (1953), *O cálculo das probabilidades e a classificação das ciências* (1956), *Lições de sociologia geral* (1956), *Arte culinária* (1957)⁽⁴⁾, *O fenómeno solar de 13 de Outubro de 1917* (1961), *Algarismo* (1964), *Aritmética das classes* (1966), *Política monetária* (1967), *Lições de mecânica racional* (1971), *Cál-*

⁽³⁾ Esta obra foi recentemente publicada pela Sociedade Portuguesa de Matemática num projecto, intitulado *Biblioteca Básica de Textos Didácticos de Matemática*, que consiste na publicação *fac simile* de um conjunto de textos didácticos de qualidade escritos em língua portuguesa. A nona edição do *Compêndio de Geometria* de Pacheco d'Amorim foi a primeira obra a ser patenteada com esta colecção.

⁽⁴⁾ Publicada sob o pseudónimo Blandimar.

culo simbólico (1973), entre outras. Grande parte da sua obra está dispersa, sendo em muitos casos de difícil acesso.

Pacheco d'Amorim também se destacou na política, tendo sido Deputado pela Covilhã em 1919, deputado na Assembleia Nacional em 1935–1938 e 1945–1949, vogal do Conselho de Normalização do Ministério da Economia (1955), vogal do Conselho Superior da Indústria (1957), vogal do Conselho de Normalização da Secretaria de Estado da Indústria (1958), entre outras funções.

Foi um dos fundadores do Centro Católico de Coimbra em 1915, presidente do Instituto de Coimbra, membro da Comissão Portuguesa da União Internacional de Matemática, do Centro de Estudos Económicos, do Conselho Superior da Indústria, sócio efectivo da Academia de Ciências desde 1947. Possuía as condecorações de *S. Gregório Magno* e a *King's Medal* por *service in the Cause of Freedom*.

Estes dados demonstram que são muitos os motivos que justificam uma investigação da vasta obra deixada por Diogo Pacheco d'Amorim. Contudo, neste trabalho delimitaremos a nossa análise à sua tese de doutoramento, da qual iremos focar o enquadramento na época e comentar as principais ideias, sublinhando, desde já, que esta é a segunda tese de doutoramento na área das Probabilidades e Estatística realizada em Portugal⁽⁵⁾.

⁽⁵⁾ Tiago de Oliveira (1991) enumera as teses de doutoramento realizadas nas universidades Portuguesas até ao ano de 1989, onde apenas aparecem três teses antes de 1950, a de Sidónio Paes (1898), a de Diogo Pacheco d'Amorim (1914) e a de Manuel dos Reis (1929), todas na Universidade de Coimbra.

Prefácio

Pacheco d'Amorim, no seu prefácio, propõe claramente construir uma base sólida para a Teoria da Probabilidade.

“O presente volume que mais se poderia chamar — Uma tentativa de racionalização do Cálculo das Probabilidades — põe em especial relevo uma proposição a que, até hoje, ninguém deu a importância devida — a proposição tirar, à sorte, um elemento de uma classe, ou lançar, à sorte, um ponto numa região.

Henrique Poincaré diz mesmo que ela, por si só, não tem significação nenhuma⁽¹⁾. Ora a verdade é que ela tem um sentido muito preciso e claro quando nós mesmos somos os agentes das tiragens ou lançamentos e isso permite-nos construir a teoria das probabilidades com toda a clareza e rigor. Partindo dela, a teoria das probabilidades pode reduzir-se a uma sucessão de proposições e definições, como qualquer outro ramo das Matemáticas Puras; a probabilidade contínua e a probabilidade descontínua aparecem com feições em tudo idênticas; os paradoxos desaparecem.”

[Pacheco d'Amorim, 1914, Prefácio, p. ix]

Salientamos, imediatamente, algumas semelhanças neste prefácio, em relação aos objectivos propostos, com o apresentado quase duas décadas depois

⁽¹⁾ H. Poincaré, *La Science et l'Hypothèse*, pág. 226

por Kolmogoroff (1933) nos seus fundamentos da Teoria da Probabilidade — a primeira axiomática aceite pela generalidade.

“The theory of probability, as a mathematical discipline, can and should be developed from axioms in exactly the same way as Geometry and Algebra. This means that after we have defined the elements to be studied and their basic relations, and have stated the axioms by which these relations are to be governed, all further exposition must be based exclusively on these axioms, independent of the usual concrete meaning of these elements and their relations.”

[Kolmogoroff 1933, pág. 1]

Pacheco d’Amorim considera que o conceito de *tirar, à sorte, um elemento de uma classe* (finita) é primitivo, isto é, não precisa de ser esclarecido, desde que sejamos nós os agentes das tiragens e conheçamos a classe, salientando que esta proposição é completamente distinta da noção de *sorte* ou *acaso*, demasiado obscura para que nela assentem as bases de qualquer ciência. Deste modo, considera que, com base neste conceito, se pode construir toda a teoria, como em qualquer outra área da matemática, sem paradoxos.

O autor indica, desta forma, que na sua tese pretende construir a Teoria da Probabilidade com base no seu conceito primitivo, para que posteriormente, após construída a teoria, esta possa ser generalizada para as restantes situações onde não somos nós os agentes das tiragens (ou lançamentos) ou não tenhamos total informação acerca do espaço amostra.

Deste modo, o autor no PREFÁCIO propõe generalizar a definição de probabilidade, quer para a probabilidade discreta, onde considera necessária para a demonstração do Teorema de Bayes, quer para a probabilidade contínua, onde considera fundamental, além de distinguir entre a probabilidade de um ponto (*livre*) e a da sua imagem (*ponto imagem*) com o conceito de

lei de probabilidade. O autor pretende expor os Teoremas de Bernoulli e as suas generalizações para a dedução rigorosa da *lei dos desvios* (Teorema Limite Central restrito à aproximação da binomial à gaussiana), e apresentar a teoria da Esperança Matemática de forma a clarificar a importância dos Teoremas de Bernoulli que nos fornecem a convergência da média empírica para a Esperança Matemática. Por fim, Pacheco d’Amorim propõe generalizar, na CONCLUSÃO da sua tese, a Teoria da Probabilidade que desenvolve utilizando o *fenómeno padrão* que define na INTRODUÇÃO e que utiliza como base para a construção da sua Teoria da Probabilidade.

Pacheco d’Amorim refere que era ainda objectivo seu apresentar, num apêndice, “*a probabilidade dos conjuntos numeráveis, depois de numerados*”, justificando a ausência deste pela “*já demasiada extensão*” da sua Tese de Doutorado⁽²⁾. Lamentamos esta situação, pois gostaríamos de descobrir como Pacheco d’Amorim se desembaraçaria deste problema, pois a definição da probabilidade em conjuntos (infinitos) numeráveis foi um obstáculo delicado de transpor, como ilustra o paradoxo da escolha aleatória de um número natural onde, supondo σ -aditividade⁽³⁾, todos os números terão probabilidade nula e, no entanto, a probabilidade de a escolha recair sobre um $n \in \mathbb{N}$ é 1.

Pacheco d’Amorim mostra claramente no prefácio que a sua principal preocupação nesta tese é obter uma fundamentação da Teoria da Probabilidade utilizando o seu *fenómeno tipo*, sem contradições ou paradoxos, e generalizar os resultados obtidos para estes fenómenos debatendo as limitações da sua

⁽²⁾ Há ainda alusão a este apêndice (sem mencionar a sua inexistência) no final do segundo capítulo PROBABILIDADE CONTÍNUA onde, após definir, de forma pouco clara, a possibilidade numa região ilimitada, anuncia que desenvolverá esse assunto num apêndice em que estudará a “*probabilidade dos conjuntos numeráveis, depois de numerados*”.

⁽³⁾ Não é necessário supor σ -aditividade, é suficiente considerar aditividade finita para tornar-se impossível definir a distribuição uniforme aos números naturais.

aplicabilidade na CONCLUSÃO. Saliente-se que, em monografias da mesma época da tese de Pacheco d'Amorim, não existe uma preocupação explícita com a axiomatização da Teoria da Probabilidade, como ilustram os tratados de Bertrand (1888), Poincaré (1896), Borel (1909, 1914) e Bachelier (1912). Poincaré, por exemplo, publicou sobre a epistemologia da ciência, e decerto o tema o preocupou, mas os avanços teóricos que permitiram a Kolmogoroff construir finalmente uma axiomática que colheu aceitação dos seus pares devem-se antes a Borel e a Fréchet, cujas publicações são mais tardias que a de Pacheco d'Amorim.

Introdução

Na INTRODUÇÃO o autor apresenta o seu fenómeno padrão e justifica o valor (utilidade) da Teoria da Probabilidade. Como referido, o principal pilar de sustentação da fundamentação de Pacheco d'Amorim é o seu conceito primitivo de *tirar, à sorte, um elemento de uma classe*⁽¹⁾. Pacheco d'Amorim começa por explicar o que entende por este conceito, principiando por analisar o nosso grau de conhecimento da classe, situação que divide em três níveis. Para facilitar a exposição, utilizaremos a mesma idealização do autor entre uma classe e uma urna que contém bolas. Deste modo os três níveis de conhecimento são:

- desconhecimento qualitativo e quantitativo — não conhecemos as cores das bolas nem o número de bolas que a urna contém;

⁽¹⁾ De referir que Diogo Pacheco d'Amorim utiliza classe como sinónimo de um conjunto de elementos. Recordemos que a teoria dos conjuntos foi criada por Cantor nos finais do século XIX, sendo um dos trabalhos mais importantes a axiomática da teoria dos conjuntos apresentada por Zermelo em 1908. A noção de classe, distinta da de conjunto, deve-se a von Neumann e Bernays com o objectivo de evitar o paradoxo de Russell, e é posterior à obra analisada (paradoxo que surge ao definir o conjunto constituído por todos os conjuntos que não são elemento de si próprio, isto é, $A = \{x : x \notin x\}$, pois com esta definição concluímos que $A \in A \Leftrightarrow A \notin A$). Consequentemente, a diferença entre classes e conjuntos não é utilizada por Pacheco de Amorim nesta obra. Nós também não faremos qualquer distinção entre estes dois conceitos.

- conhecimento qualitativo e desconhecimento quantitativo — conhecemos as cores das bolas mas desconhecemos o seu número;
- conhecimento qualitativo e quantitativo — conhecemos as cores das bolas e a quantidade de bolas de cada cor.

Se, em cada uma destas situações, retirássemos, à sorte, uma bola da urna, em que cor deveríamos apostar?

No primeiro caso nem sabemos quais as cores possíveis, pois desconhecemos por completo o conteúdo da urna; no segundo caso sabemos quais as cores, no entanto não sabemos as suas percentagens e, por conseguinte, não temos nenhuma razão para apostar em qualquer uma dessas cores em detrimento da(s) outra(s); no terceiro caso sabemos quais as cores e quais as percentagens, logo não hesitaríamos em apostar na cor com maior percentagem (não sabemos qual é a cor que irá sair, mas não teríamos dúvidas em qual apostar). Desta forma, Pacheco d'Amorim afirma que unicamente o terceiro caso nos pode servir de “*guia de conduta em certas circunstâncias*” e, como tal, supõe que todo o fenómeno pode ser identificado com uma tiragem, feita à sorte, numa classe onde possuímos o conhecimento qualitativo e quantitativo (*fenómeno padrão*).

Contudo, o seu conceito primitivo não depende unicamente do conhecimento da composição da urna, há também a ter em consideração a forma como a bola é retirada. Então, como poderemos distinguir uma tiragem à sorte de uma que não o é? Para o autor é vital esta distinção, considerando que qualquer indivíduo, desde que seja o agente da tiragem, sabe se esta é feita, ou não, à sorte, pois só nesta situação é clara a noção do seu conceito primitivo. Assim, completando a definição do seu *fenómeno padrão*, podemos dizer que o conceito base de que Pacheco d'Amorim parte para a fundamentação da Teoria da Probabilidade é *tirar, à sorte, um elemento de*

uma classe finita, onde possuímos o conhecimento qualitativo e quantitativo, quando somos nós os agentes das tiragens.

Daqui se infere que, para Pacheco d'Amorim, o seu conceito primitivo só não apresenta qualquer ambiguidade na situação em que somos nós os agentes da tiragem. Assim, o conceito a partir do qual Pacheco d'Amorim vai definir Probabilidade depende do sujeito que pratica a tiragem (ou lançamento em relação às probabilidades contínuas), demonstrando, desta forma, uma visão subjectiva ou personalista de probabilidade⁽²⁾.

Pacheco d'Amorim dedica o último capítulo da sua tese, a CONCLUSÃO, a generalizar as deduções, obtidas com base no *fenómeno padrão*, para as situações onde alguma condição do seu conceito primitivo não se verifique, decompondo a análise em relação ao agente (nós, semelhante a nós ou de outra natureza), em relação à natureza do fenómeno (tiragens ou lançamentos) e em relação ao nosso grau de conhecimento (qualitativo e quantitativo), reduzindo todas as situações ao *fenómeno padrão* acima definido e que será a base de toda a sua construção.

No entanto, mesmo conhecendo totalmente uma urna, quer as cores (consideremos, como exemplo, brancas e pretas) quer a quantidade de bolas de cada cor (suponhamos que há pelo menos uma de cada), se nós retirarmos uma bola, à sorte, poderemos prever, com a Ciência construída com base no fenómeno padrão anteriormente mencionado, qual a bola que vai sair? A resposta é claramente negativa, mesmo que na urna haja um número muito superior de bolas de uma cor em relação à outra. Então, se não é possível prever os acontecimentos como em qualquer outra Ciência, qual é a utilidade desta?

⁽²⁾ Refira-se que esta visão personalista antecedeu as interpretações subjectivas que mais se destacaram no desenvolvimento da Teoria da Probabilidade, de que são exemplo os estudos de Ramsey (1926), de Finetti (1937), de Jeffreys (1939) ou de Savage (1954).

Pacheco d'Amorim considera que, apesar da imprevisibilidade dos acontecimentos que esta Ciência retrata, ela serve de “*guia de conduta em certas circunstâncias*”, referindo que

“É nessa previsão que a intuição nos sugere quando o número de bolas brancas é muito maior que o das pretas que reside o valor prático do Cálculo de Probabilidades.

Essa intuição ficará irremediavelmente separada da certeza, por mais que da unidade se aproxime a percentagem das bolas brancas.

Mas nem por isso ela deixa de ter para nós um valor prático real.”

[Pacheco d'Amorim, 1914, p. 8]

Com esta analogia com uma urna, Pacheco d'Amorim claramente visualiza a Teoria da Probabilidade como a Ciência que nos deve guiar na tomada de decisão sob incerteza, considerando que é nesta Teoria que nos devemos fundamentar tendo em vista a tomada de decisão quando estamos perante factores não deterministas.

Capítulo 3

Classes Finitas

Pacheco d'Amorim no primeiro capítulo da sua Tese de Doutorado apresenta uma proposta de fundamentação da probabilidade discreta utilizando como modelo standard a tiragem, à sorte, de um elemento de uma classe finita cujos elementos conhecemos (conhecimento qualitativo e quantitativo). Assim, começa por distinguir três sistemas de escolha aleatória de elementos: a tiragem de um só elemento a de uma classe $\mathbf{A}$, a tiragem de um elemento a de uma classe $\mathbf{A}$ e de outro elemento b de outra classe $\mathbf{B}$ independente da primeira e a tiragem de um elemento a de uma classe $\mathbf{A}$ e de um outro elemento b de outra classe $\mathbf{B}_a$ que depende do elemento a escolhido na primeira tiragem. Com estes três sistemas de tiragens define a possibilidade de cada elemento de uma classe e a possibilidade de cada classe, que servirá de base para a sua definição de probabilidade de uma classe $\mathbf{A}''$ em relação a outra classe $\mathbf{A}'$. Desta forma Pacheco d'Amorim faz uma clara distinção entre possibilidade e probabilidade que, para o autor, é sempre uma probabilidade condicionada.

Com esta construção o autor deduz os tradicionais teoremas de probabilidade, tais como o da Probabilidade Total, o da Probabilidade Composta

e o de Bayes. Em relação a este último teorema Pacheco d'Amorim afirma que apresenta a primeira demonstração rigorosa deste resultado, uma vez que considera que sem o recurso a complexos de classes, que criou, não seria possível demonstrar tais fórmulas. Deduz ainda, de forma errônea, as fórmulas inversas da de Bayes. Por fim aplica a fórmula de Bayes para resolver o problema da Sucessão de Laplace.

3.1 Elementos e classes possíveis

Pacheco d'Amorim começa por apresentar o seu conceito primitivo *tirar à sorte, um elemento de uma classe $\mathbf{A}$* , para, com base neste conceito, definir a tiragem, à sorte, de uma união de classes [disjuntas], de uma classe composta ($\mathbf{A} \times \mathbf{B}$) e de um complexo de classes ($\mathbf{A} \odot \mathbf{B}$), fazendo uma clara distinção entre uma classe composta e um complexo de classes. Enquanto na classe composta $\mathbf{A} \times \mathbf{B}$, produto cartesiano entre classes, o autor considera haver independência entre as classes, num complexo de classes existe uma dependência da classe $\mathbf{B}_{a_i}$ em relação ao elemento a_i escolhido da classe $\mathbf{A}^{(1)}$.

⁽¹⁾ Apesar de o autor não usar os termos independência *versus* dependência ao longo deste capítulo, com uma exceção na página 22, parece-nos que é este o seu intuito. Saliente-se que não há uma definição rigorosa de independência de acontecimentos antes do tratado de Kolmogoroff (1933), embora já de Moivre (1718, p. 6) defina acontecimentos independentes: “*Two Events are independent, when they have no connexion one with the other, and that the happening of one neither forwards nor obstructs the happening of the other. Two Events are dependent, when they are so connected together as that the Probability of either’s happening is altered by the happening of the other.*”. Laplace (1812, p. *viii*) também emprega estes conceitos na apresentação do seu terceiro princípio geral do Cálculo das Probabilidades: “*Si les événements sont indépendans les uns des autres, la probabilité de l’existence de leur ensemble, est le produit de leurs probabilités particulières.*”. Refira-se que para Kolmogoroff (1933, p. 8) a noção de independência

Desta forma Pacheco d'Amorim define três sistemas de tiragens de elementos:

1. tirar, à sorte, um elemento de uma classe $\mathbf{A}$;
2. tirar, à sorte, um elemento de uma classe $\mathbf{A}$ e seguidamente tirar, à sorte, um elemento de uma classe $\mathbf{B}$ independente de $\mathbf{A}$ (a constituição da classe $\mathbf{B}$ não depende do resultado do primeiro elemento a tirado em $\mathbf{A}$). Esta situação é equivalente a tirar, à sorte, somente um elemento na classe $\mathbf{A} \times \mathbf{B}$, podendo, por este motivo, este sistema de tiragens ser reduzido à primeira situação que retrata uma única tiragem numa classe;
3. tirar, à sorte, um elemento de uma classe $\mathbf{A}$ e posteriormente tirar, à sorte, um elemento de um classe $\mathbf{B}_{a_i}$ que depende do resultado do primeiro elemento a_i tirado em $\mathbf{A}$, isto é, a tiragem é efectuada num complexo de classes $\mathbf{A} \odot \mathbf{B}$. Pacheco d'Amorim considera que, se o elemento tirado à sorte em $\mathbf{A}$ coincidir com a_i , a segunda tiragem será efectuada na classe $\mathbf{B}_{a_i}$ e tudo se passará como se apenas fosse realizada uma única tiragem numa classe composta $\mathbf{A} \times \mathbf{B}_{a_i}$. Notemos que é inexequível reduzir uma tiragem num complexo de classes a uma única tiragem numa classe (composta ou não), pois a definição da classe onde é efectuada a segunda tiragem depende do elemento resultante da primeira tiragem, o que torna impossível traduzir *a priori* estas situações numa única tiragem. Realce-se que o autor apresenta esta definição com o objectivo de posteriormente servir de base à sua definição de possibilidade, sendo a possibilidade de cada elemento $(a_i, b_i^{a_i})$ do complexo de

é fundamental na Teoria da Probabilidade, referindo “*Historically, the independence of experiments and random variables represents the very mathematical concept that has given the theory of probability its peculiar stamp.*”

classes $\mathbf{A} \odot \mathbf{B}$ determinada sob esta hipótese. Assim, como analisaremos mais adiante, a possibilidade de um elemento $(a_i, b_i^{a_i}) \in \mathbf{A} \odot \mathbf{B}$ é determinada considerando que o resultado da primeira tiragem é já conhecido (condicionado a que o elemento escolhido em $\mathbf{A}$ foi a_i).

Pacheco d'Amorim utiliza os três sistemas de tiragens acima referidos para definir quais são os elementos possíveis de uma *tiragem à sorte*. Deste modo, numa classe $\mathbf{A} = \{a_1, \dots, a_{n_{\mathbf{A}}}\}$ serão possíveis os $n_{\mathbf{A}}$ elementos a_i , $i = 1, \dots, n_{\mathbf{A}}$ da classe $\mathbf{A}$; numa classe composta $\mathbf{A} \times \mathbf{B}$, com $\mathbf{A} = \{a_1, \dots, a_{n_{\mathbf{A}}}\}$ e $\mathbf{B} = \{b_1, \dots, b_{n_{\mathbf{B}}}\}$, serão possíveis os $n_{\mathbf{A}} \times n_{\mathbf{B}}$ elementos (a_i, b_j) , $i = 1, \dots, n_{\mathbf{A}}$ e $j = 1, \dots, n_{\mathbf{B}}$, da classe $\mathbf{A} \times \mathbf{B}$ e num complexo de classes $\mathbf{A} \odot \mathbf{B}$, onde a cada elemento a_i é associada uma classe $\mathbf{B}_{a_i} = \{b_1^{a_i}, \dots, b_{n_{\mathbf{B}_{a_i}}}^{a_i}\}$, serão possíveis os $\sum_{i=1}^{n_{\mathbf{A}}} n_{\mathbf{B}_{a_i}}$ elementos $(a_i, b_j^{a_i})$, $i = 1, \dots, n_{\mathbf{A}}$ e $j = 1, \dots, n_{\mathbf{B}_{a_i}}$, da classe $\mathbf{A} \odot \mathbf{B}$.

Uma classe possível, em relação a determinada tiragem, é definida como qualquer classe constituída apenas por elementos possíveis a essa mesma tiragem. A classe total possível é a classe constituída por todos os elementos possíveis, que corresponde à nossa noção actual de espaço amostra⁽²⁾.

3.2 Possibilidade

Pacheco d'Amorim define a possibilidade de cada elemento a tirado à sorte da classe $\mathbf{A}$, representado por π_a , através de

$$\pi_a = \frac{1}{\#\mathbf{A}} = \frac{1}{n_{\mathbf{A}}}. \quad (3.1)$$

Por conseguinte, a possibilidade de um elemento composto (a, b) será o

⁽²⁾ Um conceito que von Mises criou no primeiro quartel do século XX, posterior à Tese de Doutoramento de Pacheco d'Amorim

produto das possibilidades dos elementos componentes, isto é, $\pi_{(a,b)} = \pi_a \cdot \pi_b$ (Prop. I) que é consequência imediata de, em classes compostas, verificar-se $\#(\mathbf{A} \times \mathbf{B}) = \#\mathbf{A} \cdot \#\mathbf{B}$. Saliente-se que, por definição, um elemento de um complexo de classes pode ser visto como um elemento composto; por exemplo o elemento $(a_i, b_j^{a_i})$ da classe $\mathbf{A} \odot \mathbf{B}$ é um elemento composto da classe composta $a_i \times \mathbf{B}_{a_i}$ sendo a sua possibilidade dada por

$$\pi_{(a_i, b_j^{a_i})} = \pi_{a_i} \cdot \pi_{b_j^{a_i}} = \frac{1}{n_{\mathbf{A}} n_{\mathbf{B}_{a_i}}}. \quad (3.2)$$

Desta forma, todos os elementos de uma classe (composta ou não) têm a mesma possibilidade, enquanto num complexo de classes isto não é necessariamente verdadeiro (pois só se verificará se todas as classes $\mathbf{B}_{a_i}$ tiverem o mesmo número de elementos, isto é, se $n_{\mathbf{B}_{a_i}} = n_{\mathbf{B}}, \forall i$).

A possibilidade de uma classe é definida pela soma das possibilidades dos elementos que a compõem, isto é, representando a possibilidade da classe $\mathbf{A}'$ por $\varpi_{\mathbf{A}'}$ esta é definida por

$$\varpi_{\mathbf{A}'} = \sum_{a \in \mathbf{A}'} \pi_a. \quad (3.3)$$

Naturalmente, com esta definição, a possibilidade de uma classe resultante da união de classes disjuntas será obtida pela soma das possibilidades de cada uma das classes que a compõem (Prop. II). Desta forma é possível determinar a possibilidade de qualquer classe, mesmo sendo um complexo de classes, pois qualquer complexo de classes pode ser decomposto na união de classes compostas (disjuntas), bastando fazer

$$\mathbf{A}' \odot \mathbf{B}' = \bigcup_{a_i \in \mathbf{A}'} a_i \times \mathbf{B}'_{a_i}, \quad (3.4)$$

que recorrendo à fórmula (3.2), obtemos

$$\begin{aligned}\varpi_{\mathbf{A}' \odot \rightarrow \mathbf{B}'} &= \sum_{a_i \in \mathbf{A}'} \sum_{b_j^i \in \mathbf{B}'_{a_i}} \pi_{(a_i, b_j^i)} = \sum_{a_i \in \mathbf{A}'} \sum_{b_j^i \in \mathbf{B}'_{a_i}} \frac{1}{n_{\mathbf{A}} n_{\mathbf{B}_{a_i}}} = \\ &= \frac{1}{n_{\mathbf{A}}} \sum_{a_i \in \mathbf{A}'} \frac{\#\mathbf{B}'_{a_i}}{n_{\mathbf{B}_{a_i}}} = \frac{1}{n_{\mathbf{A}}} \sum_{a_i \in \mathbf{A}'} \varpi_{\mathbf{B}'_{a_i}}.\end{aligned}\quad (3.5)$$

Exemplo 3.1 (Complexo de Classes). Seja $\mathbf{A} \odot \rightarrow \mathbf{B} \odot \rightarrow \mathbf{C}$ um complexo de classes, sendo $\mathbf{A} = \{a_1, \dots, a_n\}$, onde a cada elemento $a_i \in \mathbf{A}$ se associa um conjunto $\mathbf{B}_{a_i} = \mathbf{B}_i = \{b_1^i, \dots, b_{n_i}^i\}$ onde $\#\mathbf{B}_i = n_i$ e a cada elemento b_j^i se associa um novo conjunto $\mathbf{C}_{ij} = \{c_1^{ij}, \dots, c_{n_{ij}}^{ij}\}$ onde $\#\mathbf{C}_{ij} = n_{ij}$.

Facilmente se obtém, para qualquer elemento da classe $\mathbf{A}$, a sua possibilidade através de

$$\pi_{a_i} = \frac{1}{n}, \quad (3.6)$$

onde se nota que, nesta classe, os elementos têm todos a mesma possibilidade. Para qualquer subconjunto $\mathbf{A}'$ de $\mathbf{A}$ tem-se que

$$\varpi_{\mathbf{A}'} = \frac{\#\mathbf{A}'}{n}, \quad \mathbf{A}' \subset \mathbf{A}. \quad (3.7)$$

Em relação aos conjuntos $\mathbf{B}_i$ temos que a possibilidade de qualquer elemento é dada por

$$\pi_{b_j^i} = \frac{1}{n_i}, \quad (3.8)$$

de onde se conclui que a possibilidade de cada elemento depende da classe $\mathbf{B}_i$ a que pertence, isto é, dentro de cada classe todos os elementos possuem a mesma possibilidade, mas para classes $\mathbf{B}_i$ distintas os elementos possuem geralmente possibilidades distintas (excepto se estas classes possuírem o mesmo número de elementos). A possibilidade de qualquer subconjunto $\mathbf{B}'_i$ de $\mathbf{B}_i$ é determinada através de

$$\varpi_{\mathbf{B}'_i} = \frac{\#\mathbf{B}'_i}{n_i}, \quad \mathbf{B}'_i \subset \mathbf{B}_i. \quad (3.9)$$

No que se refere ao complexo de classes $\mathbf{A} \odot \rightarrow \mathbf{B}$, temos que para cada elemento (a_i, b_j^i) (que é um elemento composto) a sua possibilidade é determinada por (Prop. I)

$$\pi_{(a_i, b_j^i)} = \pi_{a_i} \cdot \pi_{b_j^i} = \frac{1}{n \ n_i}, \quad (3.10)$$

para $i = 1, \dots, n$ e $j = 1, \dots, n_i$. Podemos constatar uma vez mais, com base nesta fórmula, que num complexo de classes os elementos só são equipossíveis no caso em que todas as classes $\mathbf{B}_i$ possuem o mesmo número de elementos. O mesmo raciocínio poderia facilmente ser aplicado às classes $\mathbf{C}_{ij}$, obtendo-se

$$\pi_{c_k^{ij}} = \frac{1}{n_{ij}}, \quad \varpi_{\mathbf{C}'_{ij}} = \frac{\#\mathbf{C}'_{ij}}{n_{ij}}, \quad \mathbf{C}'_{ij} \subset \mathbf{C}_{ij}, \quad (3.11)$$

e para as possibilidades dos elementos (a_i, b_j^i, c_k^{ij}) do complexo de classes $\mathbf{A} \odot \rightarrow \mathbf{B} \odot \rightarrow \mathbf{C}$ teremos

$$\pi_{(a_i, b_j^i, c_k^{ij})} = \pi_{a_i} \cdot \pi_{b_j^i} \cdot \pi_{c_k^{ij}} = \frac{1}{n \ n_i \ n_{ij}}, \quad (3.12)$$

para $i = 1, \dots, n$, $j = 1, \dots, n_i$ e $k = 1, \dots, n_{ij}$.

Desta forma, se pretendermos a possibilidade de um complexo de classes $\mathbf{A}' \odot \rightarrow \mathbf{B}'$ contido em $\mathbf{A} \odot \rightarrow \mathbf{B}$ apenas temos de somar as possibilidades dos seus elementos como determina a fórmula (3.3) (página 173). Assim, por exemplo, se desejarmos determinar a possibilidade da classe $\mathbf{A}' \odot \rightarrow \mathbf{B}'$ teremos

$$\varpi_{\mathbf{A}' \odot \rightarrow \mathbf{B}'} = \sum_{a_i \in \mathbf{A}'} \sum_{b_j^i \in \mathbf{B}'} \pi_{(a_i, b_j^i)} = \sum_{a_i \in \mathbf{A}'} \sum_{b_j^i \in \mathbf{B}'} \frac{1}{n \ n_i} = \frac{1}{n} \sum_{a_i \in \mathbf{A}'} \frac{\#\mathbf{B}'_i}{n_i}. \quad (3.13)$$

Para a possibilidade de um complexo de classes $\mathbf{A}' \odot \rightarrow \mathbf{B}' \odot \rightarrow \mathbf{C}'$ contido em $\mathbf{A} \odot \rightarrow \mathbf{B} \odot \rightarrow \mathbf{C}$, uma vez que

$$\mathbf{A}' \odot \rightarrow \mathbf{B}' \odot \rightarrow \mathbf{C}' = \bigcup_{a_i \in \mathbf{A}'} \bigcup_{b_j^i \in \mathbf{B}'_i} \bigcup_{c_k^{ij} \in \mathbf{C}'_{ij}} (a_i, b_j^i, c_k^{ij}), \quad (3.14)$$

recorrendo à fórmula (3.12) obtemos

$$\begin{aligned}\varpi_{\mathbf{A}' \odot \rightarrow \mathbf{B}' \odot \rightarrow \mathbf{C}'} &= \sum_{a_i \in \mathbf{A}'} \sum_{b_j^i \in \mathbf{B}'_i} \sum_{c_k^{ij} \in \mathbf{C}'_{ij}} \pi_{(a_i, b_j^i, c_k^{ij})} = \\ &= \frac{1}{n} \sum_{a_i \in \mathbf{A}'} \frac{1}{n_i} \sum_{b_j^i \in \mathbf{B}'_i} \frac{\#\mathbf{C}'_{ij}}{n_{ij}}.\end{aligned}\quad (3.15)$$

◆

A Proposição III de Pacheco d'Amorim garante-nos que a possibilidade da classe total possível é igual à unidade. De facto, se considerarmos o complexo de classes $\mathbf{A} \odot \rightarrow \mathbf{B}$ como a classe total possível em relação a uma determinada tiragem, temos que

$$\varpi_{\mathbf{A} \odot \rightarrow \mathbf{B}} = \sum_{a_i \in \mathbf{A}} \sum_{b_j^i \in \mathbf{B}_i} \pi_{(a_i, b_j^i)} = \sum_{a_i \in \mathbf{A}} \frac{1}{n_{\mathbf{A}}} \cdot \frac{\#\mathbf{B}_i}{n_i} = \sum_{a_i \in \mathbf{A}} \frac{1}{n_{\mathbf{A}}} = \frac{\#\mathbf{A}}{n_{\mathbf{A}}} = 1, \quad (3.16)$$

que pode ser demonstrado para um qualquer complexo de classes.

A Proposição IV decompõe a possibilidade de uma classe composta na multiplicação das possibilidades das suas classes constituintes, isto é,

$$\varpi_{\mathbf{A} \times \mathbf{B}} = \varpi_{\mathbf{A}} \cdot \varpi_{\mathbf{B}}, \quad (3.17)$$

consequência imediata de, em classes compostas, verificar-se para cada elemento $\pi_{(a,b)} = \pi_a \cdot \pi_b$ e por conseguinte

$$\varpi_{\mathbf{A} \times \mathbf{B}} = \sum_{a \in \mathbf{A}} \sum_{b \in \mathbf{B}} \pi_{(a,b)} = \sum_{a \in \mathbf{A}} \sum_{b \in \mathbf{B}} \pi_a \cdot \pi_b = \sum_{a \in \mathbf{A}} \pi_a \sum_{b \in \mathbf{B}} \pi_b = \varpi_{\mathbf{A}} \cdot \varpi_{\mathbf{B}}.$$

Esta proposição permite-nos determinar a possibilidade de uma classe composta através da multiplicação das possibilidades das classes que a compõem. Consequentemente, também nos permite determinar a possibilidade de um complexo de classes pelos valores das possibilidades das classes que o compõem, pois como

$$\mathbf{A}' \odot \rightarrow \mathbf{B}' = \bigcup_{a_i \in \mathbf{A}'} a_i \times \mathbf{B}'_{a_i}, \quad (3.18)$$

então, para qualquer complexo de classes, teremos

$$\varpi_{\mathbf{A}' \curvearrowright \mathbf{B}'} = \sum_{a_i \in \mathbf{A}'} \pi_{a_i} \cdot \varpi_{\mathbf{B}'_i},^{(3)} \quad (3.19)$$

visto que

$$\begin{aligned} \varpi_{\mathbf{A}' \curvearrowright \mathbf{B}'} &= \sum_{a_i \in \mathbf{A}'} \sum_{b_j^i \in \mathbf{B}'_i} \pi_{(a_i, b_j^i)} = \sum_{a_i \in \mathbf{A}'} \sum_{b_j^i \in \mathbf{B}'_i} \pi_{a_i} \cdot \pi_{b_j^i} = \\ &= \sum_{a_i \in \mathbf{A}'} \pi_{a_i} \sum_{b_j^i \in \mathbf{B}'_i} \pi_{b_j^i} = \sum_{a_i \in \mathbf{A}'} \pi_{a_i} \cdot \varpi_{\mathbf{B}'_i}. \end{aligned}$$

Exemplo 3.2 (Distribuição hipergeométrica). Consideremos uma urna com m bolas das quais m_1 são brancas e m_2 pretas ($m_1 + m_2 = m$). Retirem-se n bolas, com $n \leq m$, desta urna sem repor as bolas retiradas. Qual a possibilidade de, das n bolas retiradas, n_1 serem brancas e n_2 pretas? (com $n_1 \leq m_1$, $n_2 \leq m_2$ e $n_1 + n_2 = n$)

Notemos que a possibilidade de uma qualquer sequência em que, das n bolas retiradas, n_1 sejam brancas, é dada por (Prop. I)⁽⁴⁾

$$\begin{aligned} \varpi' &= \frac{m_1}{m} \cdot \frac{m_1 - 1}{m - 1} \cdots \frac{m_1 - n_1 + 1}{m - n_1 + 1} \cdot \frac{m_2}{m - n_1} \cdot \frac{m_2 - 1}{m - (n_1 + 1)} \cdots \\ &\cdots \frac{m_2 - n_2 + 1}{m - n + 1} = \frac{m_1! \, m_2! \, (m - n)!}{(m_1 - n_1)! \, (m_2 - n_2)! \, m!} \end{aligned}$$

e, como existem $\binom{n}{n_1}$ sequências favoráveis à ocorrência de n_1 bolas brancas

⁽³⁾ Num complexo de classe não é obrigatório serem associadas classes $\mathbf{B}_i$ distintas a cada elemento $a_i \in \mathbf{A}$. Podemos associar classes $\mathbf{B}_i$ a grupos de elementos de $\mathbf{A}$, isto é, fazendo $\mathbf{A} = \mathbf{A}_1 \cup \cdots \cup \mathbf{A}_n$ com $\mathbf{A}_i$ disjuntas, onde a cada classe $\mathbf{A}_i$ se associa uma classe $\mathbf{B}_i$. Nestas situações teríamos $\varpi_{\mathbf{A}' \curvearrowright \mathbf{B}'} = \sum_{\mathbf{A}'_i \in \mathbf{A}} \varpi_{\mathbf{A}'_i} \cdot \varpi_{\mathbf{B}'_i}$ onde $\mathbf{A}'_i \subset \mathbf{A}_i$ e $\mathbf{B}'_i \subset \mathbf{B}_i$. Contudo, este caso está incluído na fórmula apresentada.

⁽⁴⁾ Para simplificar a apresentação suporemos que as primeiras n_1 bolas retiradas são as brancas. Contudo o resultado é o mesmo, pois, se considerarmos outra ordem, iremos obter uma permutação destas parcelas, sendo o resultado final o mesmo.

das n bolas retiradas, teremos

$$\begin{aligned}\varpi &= \binom{n}{n_1} \cdot \frac{m_1! m_2! (m-n)!}{(m_1-n_1)! (m_2-n_2)! m!} = \\ &= \frac{n!}{n_1! (n-n_1)!} \cdot \frac{m_1! m_2! (m-n)!}{(m_1-n_1)! (m_2-n_2)! m!} = \frac{\binom{m_1}{n_1} \binom{m_2}{n_2}}{\binom{m}{n}}\end{aligned}$$

uma vez que $n_2 = n - n_1$. Desta forma obtemos a função de probabilidade da distribuição hipergeométrica. $\blacklozenge$

3.3 Probabilidade de $\mathbf{A}''$ em relação a $\mathbf{A}'$

Considerando $\mathbf{A}$ a classe total possível em relação a uma tiragem, Pacheco d'Amorim define probabilidade de uma classe $\mathbf{A}''$ (classe favorável) em relação a outra classe $\mathbf{A}'$ (classe possível), com $\mathbf{A}'' \subset \mathbf{A}' \subset \mathbf{A}$, pelo rácio das possibilidades destas classes, ou seja,

$$\mathbb{P}_{\mathbf{A}'}(\mathbf{A}'') = \frac{\varpi_{\mathbf{A}''}}{\varpi_{\mathbf{A}'}}. \quad (3.20)$$

Pacheco d'Amorim, após apresentar a sua definição de Probabilidade, acrescenta alguns comentários. Em primeiro lugar salienta que, se os elementos de $\mathbf{A}'$ possuírem todos a mesma possibilidade, então a probabilidade será igual ao número de casos favoráveis a dividir pelo número de casos possíveis, isto é, teremos

$$\mathbb{P}_{\mathbf{A}'}(\mathbf{A}'') = \frac{\#\mathbf{A}''}{\#\mathbf{A}'} . \quad (3.21)$$

Outra situação particular analisada pelo autor é aquela em que a classe possível corresponde à classe total possível ($\mathbf{A}' = \mathbf{A}$), onde teremos, como anteriormente vimos, $\varpi_{\mathbf{A}} = 1$ e a probabilidade de $\mathbf{A}''$ em relação a $\mathbf{A}$ resume-se a

$$\mathbb{P}_{\mathbf{A}}(\mathbf{A}'') = \varpi_{\mathbf{A}''}. \quad (3.22)$$

Na opinião do autor este último caso particular, onde temos presentes as hipóteses de os elementos serem igualmente possíveis e a classe possível identificar-se com a classe total possível, equivale à definição mais geral das definições que Laplace apresentou de probabilidade.

De facto, Laplace na sua obra capital *Théorie Analytique des Probabilités* define probabilidade nos seguintes termos:

“la probabilité d’un événement, est le rapport du nombre des cas qui lui sont favorables, au nombre de tous les cas possibles; lorsque rien ne porte à croire que l’un de ces cas doit arriver plutôt que les autres, ce qui les rend pour nous, également possibles. La juste appréciation de ces cas divers, est un des points les plus délicats de l’analyse des hasards.

Si tous les cas ne sont pas également possibles, on déterminera leurs possibilités respectives; et alors la probabilité de l’événement sera la somme des probabilités de chaque cas favorables.”

[Laplace, 1812, Livre II, p. 179]

Laplace apresenta, na Introdução da obra acima citada⁽⁵⁾, definições semelhantes de probabilidade, quer na página *iv* quer na página *vii*, como sendo o primeiro dos seus dez princípios gerais do Cálculo de Probabilidade. Das definições presentes na obra de Laplace é notória a dependência destas da condição de equiprobabilidade. No entanto, para Laplace só existirá equiprobabilidade? Cremos que não, dado que Laplace aceita situações onde não há equiprobabilidade, apesar de não as explicar, referindo apenas que, quando tal acontece, devemos aumentar o número de casos possíveis de forma a permitir a redução de todas as probabilidades a um quociente entre um determi-

⁽⁵⁾ A *Introduction* da segunda edição da *Théorie Analytique des Probabilités* de 1814 foi publicada isoladamente com o nome *Essai Philosophique sur les Probabilités*.

nado número de casos favoráveis e o número (aumentado) de casos possíveis, como refere Hald (1998).

“For Laplace equipossibility of simple events is the fundamental notion in probability theory. He remarks that if the cases considered are not equally possible, they should be made so by subdividing the more possible cases. If $p_i = \frac{n_i}{N}$ is the probability of case number i , then this case should be considered as n_i equally possible cases (Laplace, 1982, OC 7, p. 181).

It will be seen from these statements that Laplace “solved” the problem of measuring probability by introducing the undefined notion “equally possible cases” in relation to “simple events” or “events of the same kind”.”

[Hald, 1998, pág. 159]

Poder-se-á questionar se, nos complexos de classes criados por Pacheco d'Amorim, não será possível actuar de forma análoga. Começemos por analisar um exemplo simples para ilustrar esta ideia.

Exemplo 3.3 (Complexo de Classes reduzido à equipossibilidade). Consideremos as classes $\mathbf{A} = \{a_1, a_2, a_3\}$, $\mathbf{B}_1 = \{b_1^1\}$, $\mathbf{B}_2 = \{b_1^2, b_2^2\}$ e $\mathbf{B}_3 = \{b_1^3, b_2^3, b_3^3\}$. Se determinarmos as possibilidades de cada elemento (a_i, b_j^i) do complexo de classes $\mathbf{A} \odot \mathbf{B}$ obteremos

$$\begin{array}{ll} \pi_{(a_1, b_1^1)} = \frac{1}{3}; & \pi_{(a_3, b_1^3)} = \frac{1}{9}; \\ \pi_{(a_2, b_1^2)} = \frac{1}{6}; & \pi_{(a_3, b_2^3)} = \frac{1}{9}; \\ \pi_{(a_2, b_2^2)} = \frac{1}{6}; & \pi_{(a_3, b_3^3)} = \frac{1}{9}. \end{array}$$

Para que possamos reduzir este complexo de classes a uma situação de equipossibilidade, temos de reduzir ao mesmo denominador estes valores. Assim,

em virtude de cada um destes denominadores ter sido obtido através de $n \cdot n_i$, uma forma de determinar um denominador comum é utilizar

$$n_T = n \prod_{k=1}^3 n_k = 3 \times 1 \times 2 \times 3 = 18. \quad (3.23)$$

Deste modo, poderíamos considerar que esta experiência era formada por 18 elementos com a mesma possibilidade, sendo o novo número de casos favoráveis à realização de cada um dos seis acontecimentos possíveis determinado através de

$$\begin{aligned} n_{(a_1, b_1^1)} &= 2 \times 3 = 6; & n_{(a_3, b_1^3)} &= 1 \times 2 = 2; \\ n_{(a_2, b_1^2)} &= 1 \times 3 = 3; & n_{(a_3, b_2^3)} &= 1 \times 2 = 2; \\ n_{(a_2, b_2^2)} &= 1 \times 3 = 3; & n_{(a_3, b_3^3)} &= 1 \times 2 = 2. \end{aligned}$$

Assim sendo, a possibilidade de cada acontecimento (anteriormente indicada) pode ser determinada pelo quociente entre o novo número de casos que lhe são favoráveis e o novo número de casos possíveis, através de

$$n_{(a_i, b_j^i)} = \frac{n_{(a_i, b_j^i)}}{n_T}. \quad (3.24)$$

◆

De facto, utilizando o mesmo raciocínio do exemplo 3.3, é sempre possível reduzir um complexo de classes à situação de equipossibilidade da forma como Laplace refere na sua obra, pois, considerando um complexo de classes $\mathbf{A} \odot \mathbf{B}$ como definido no exemplo 3.1 (página 174), temos

$$\pi_{(a_i, b_j^i)} = \frac{1}{n} \frac{\prod_{\substack{k=1 \\ k \neq i}}^n n_k}{n \prod_{k=1}^n n_k}, \quad (3.25)$$

para $i = 1, \dots, n$ e $j = 1, \dots, n_i$. Desta forma, num complexo de classes, podemos pensar que temos uma experiência com $n \times \prod_{k=1}^n n_k$ resultados

possíveis e equiprováveis, dos quais $\prod_{\substack{k=1 \\ k \neq i}}^n n_k$ são favoráveis à realização do elemento (a_i, b_j^i) .

Pensamos, contudo, que Laplace, na sua obra, não explica, exemplifica ou define como é que situações com probabilidades distintas podem ocorrer, enquanto Pacheco d'Amorim justifica a existência de probabilidades distintas com o recurso a complexos de classes. Desta forma, Pacheco d'Amorim fornece, de facto, uma definição mais geral que a de Laplace.

Mesmo se considerarmos obras mais recentes que a de Laplace, não encontramos uma definição (objectiva) mais geral de probabilidade. Referimo-nos, por exemplo, a Lacroix (1816), Poisson (1837), Cournot (1843), Bertrand (1888), Poincaré (1896), Guimarães (1904), Montessus de Ballore (1908), Borel (1909, 1914), Bachelier (1912) e Pinto (1913).

Outro aspecto importante a salientar na definição de probabilidade de Pacheco d'Amorim é a sua definição restringir-se à probabilidade condicionada, isto é, o autor denomina por possibilidades o conceito, usualmente utilizado, de probabilidade incondicional e apelida de probabilidade o actual conceito de probabilidade condicionada (restrito à situação em que a classe condicionante contém a classe condicionada). Naturalmente, se determinarmos a probabilidade de uma classe $\mathbf{A}''$ em relação à classe total possível, temos $\mathbb{P}_{\mathbf{A}}(\mathbf{A}'') = \varpi_{\mathbf{A}''}$ (na notação actual correspondente a $\mathbb{P}(\mathbf{A}'' \mid \Omega) = \mathbb{P}(\mathbf{A}'')$) que equivale à probabilidade absoluta (ou incondicionada) da classe $\mathbf{A}''$, ou seja, nesta construção podemos visualizar a probabilidade absoluta como um caso particular da probabilidade condicionada e não a habitual estrutura onde a probabilidade condicionada é definida através da probabilidade absoluta. Porém, Pacheco d'Amorim não é único nesta característica de apenas definir a probabilidade condicionada, pois alguns anos mais tarde, por exemplo, Keynes (1921), de Finetti (1937), Jeffreys (1939) e Rényi (1955, 1970) fazem o mesmo utilizando abordagens bastante distintas. Para Keynes

e Jeffreys, defensores de uma definição lógica indutiva da probabilidade⁽⁶⁾, habitualmente denominada por visão bayesiana objectiva de probabilidade, a probabilidade representa uma relação lógica entre uma proposição e as hipóteses consideradas (ou nível de conhecimento), isto é, uma probabilidade $\mathbb{P}(\mathbf{P}|\mathbf{H})$ representa o grau de credibilidade (*degree of belief*) sobre uma relação lógica entre uma hipótese (premissa) $\mathbf{H}$ e uma proposição $\mathbf{P}$. Deste modo, consideram que qualquer probabilidade depende das hipóteses consideradas (sendo reavaliada caso as hipóteses sejam alteradas) e, como tal, a probabilidade de qualquer proposição só faz sentido em relação ao nível de conhecimento considerado. Para Bruno de Finetti a probabilidade de uma proposição é vista como o grau de credibilidade de um indivíduo (*personal degrees of believe*) relativamente à proposição sendo, por esta razão, denominada por visão bayesiana personalista (subjectiva) de probabilidade. Nesta visão, ao contrário da de Keynes e Jeffreys, não é obrigatório que haja apenas uma probabilidade para determinada proposição, podendo esta variar de indivíduo para indivíduo, o que significa que, para a mesma quantidade de informação indivíduos diferentes podem associar diferentes graus de credibilidade à mesma proposição⁽⁷⁾, pois é possível associar qualquer grau a

⁽⁶⁾ As visões deste dois autores são semelhantes, mas não coincidem na totalidade, por exemplo Jeffreys não concorda com Keynes por este considerar que as probabilidades só são parcialmente ordenadas. Para Keynes (1921, p. 38–39) as probabilidades das relações lógicas entre uma hipótese e uma proposição nem sempre são mensuráveis e por vezes nem sequer comparáveis, enquanto para Jeffreys (1939, p. 17) quaisquer duas probabilidades podem ser comparadas, através do seu primeiro axioma (axioma da comparação de probabilidades), no que se refere às probabilidades condicionadas aos mesmos dados, considerando que não é preciso incluir (no axioma) o caso em que não se referem ao mesmo conjunto de dados pois este pode sempre ser reduzido ao anterior.

⁽⁷⁾ Apesar de, em rigor, a informação disponível ser sempre diferente de indivíduo para indivíduo.

uma proposição, desde que os graus de credibilidade obedeçam a determinados critérios de consistência, tais como o princípio da coerência (ausência do *dutch book*⁽⁸⁾). Rényi, em 1955, numa visão baseada na Teoria da Medida, apresenta uma axiomática semelhante à de Kolmogoroff, mas o seu ponto de partida, o seu conceito primitivo, é a probabilidade condicionada. Assim, considerando Ω o espaço dos resultados, $\mathcal{A}$ uma σ -álgebra sobre Ω e $\mathcal{B}$ um sistema de conjuntos não vazio tal que $\mathcal{B} \subset \mathcal{A}$, então $\mathbb{P}(\mathbf{A} \mid \mathbf{B})$, com $\mathbf{A} \in \mathcal{A}$ e $\mathbf{B} \in \mathcal{B}$, é denominada a *probabilidade condicionada do acontecimento A em relação ao condicionante B* se verificar os seguintes axiomas:

1. $\mathbb{P}(\mathbf{A} \mid \mathbf{B}) \geq 0$ e $\mathbb{P}(\mathbf{B} \mid \mathbf{B}) = 1$,
2. Seja $\{\mathbf{A}_i\}_{i=1}^{\infty}$ uma sucessão de acontecimentos de $\mathcal{A}$ incompatíveis dois a dois, $\mathbf{A}_i \cap \mathbf{A}_j = \emptyset, \forall i \neq j$, então, para qualquer $\mathbf{B} \in \mathcal{B}$ tem-se

$$\mathbb{P}\left(\bigcup_{i=1}^{+\infty} \mathbf{A}_i \mid \mathbf{B}\right) = \sum_{i=1}^{+\infty} \mathbb{P}(\mathbf{A}_i \mid \mathbf{B}). \quad (3.26)$$

Este axioma significa que, para qualquer acontecimento $\mathbf{B} \in \mathcal{B}$ fixo, $\mathbb{P}(\mathbf{A} \mid \mathbf{B})$ é uma medida em $\mathcal{A}$.

3. Sejam $\mathbf{A} \in \mathcal{A}$, $\mathbf{B} \in \mathcal{B}$ e $\mathbf{C} \in \mathcal{B}$ três acontecimentos, tais que $\mathbf{B} \subset \mathbf{C}$ e $\mathbb{P}(\mathbf{B} \mid \mathbf{C}) > 0$, então

$$\mathbb{P}(\mathbf{A} \mid \mathbf{B}) = \frac{\mathbb{P}(\mathbf{A} \cap \mathbf{B} \mid \mathbf{C})}{\mathbb{P}(\mathbf{B} \mid \mathbf{C})}. \quad (3.27)$$

Deste modo, o sistema $[\Omega, \mathcal{A}, \mathcal{B}, \mathbb{P}(\mathbf{A} \mid \mathbf{B})]$ forma um espaço de probabilidade condicionado. Rényi (1970) expõe a sua proposta de axiomatização da probabilidade baseada nestes axiomas.

⁽⁸⁾ O *dutch book* (ou banca holandesa) consiste em, numa aposta, não haver perda garantida, isto é, a aposta ser feita de tal forma que um dos apostadores, independentemente do resultado que ocorra, perca sempre dinheiro.

A definição de probabilidade [condicionada] de Pacheco d'Amorim é, apesar do referido, inesperadamente restritiva⁽⁹⁾ à situação onde $\mathbf{A}'' \subset \mathbf{A}'$, podendo este conceito ser facilmente extendido para o caso geral, definindo, para a situação em que $\mathbf{A}'' \not\subset \mathbf{A}'$, a probabilidade através de

$$\mathbb{P}_{\mathbf{A}'}(\mathbf{A}'') = \mathbb{P}_{\mathbf{A}'}(\mathbf{A}' \cap \mathbf{A}''). \quad (3.28)$$

Deste modo é possível, e até de uma forma elementar, estender a definição do autor para o caso geral, obtendo-se, desta forma, a definição de probabilidade condicionada que é normalmente adoptada. Outro motivo para a nossa surpresa, nesta restrição de Pacheco d'Amorim, é o facto de diversas obras anteriores à sua tese, que decerto conhecia, utilizarem a definição geral, como por exemplo a de Laplace (1812) aquando da demonstração do seu princípio da probabilidade composta.

“Pour démontrer ce principe d’une manière générale, nommons p le nombre de tous les cas possibles, et supposons que dans ce nombre, il y en ait p' favorables au premier événement. Supposons ensuite que dans le nombre p' , il y en ait q favorables au second événement; il est clair que $\frac{q}{p}$ sera la probabilité de l’événement composé.”

[Laplace, 1812, p. 180]

Notemos que, nesta passagem aqui transcrita da obra de Laplace, o autor não faz nenhuma restrição no sentido de inclusão, isto é, não obriga a que o acontecimento condicionado (segundo acontecimento — $\mathbf{B}$) esteja contido no condicionante (primeiro acontecimento — $\mathbf{A}$), definindo $\mathbb{P}(\mathbf{B} \mid \mathbf{A}) = \frac{\mathbb{P}(\mathbf{A} \cap \mathbf{B})}{\mathbb{P}(\mathbf{A})}$.

Refira-se, no entanto, que o conceito de probabilidade condicionada actualmente utilizado somente foi rigorosamente definido com Kolmogoroff em

⁽⁹⁾ Saliente-se, uma vez mais, que na definição de Pacheco d'Amorim tem-se obrigatoriamente $\mathbf{A}' \subset \mathbf{A}$ e, nestes casos, na probabilidade condicionada actualmente utilizada teríamos $\mathbb{P}(\mathbf{A}' \mid \mathbf{A}) = \frac{\mathbb{P}(\mathbf{A}' \cap \mathbf{A})}{\mathbb{P}(\mathbf{A})} = \frac{\mathbb{P}(\mathbf{A}')}{\mathbb{P}(\mathbf{A})}$.

1933. Antes do aparecimento dos fundamentos de Kolmogoroff, no conceito de probabilidade condicionada está (quase) sempre presente a ideia de uma sequência, organização cronológica dos acontecimentos, sendo tipicamente definida a probabilidade de o segundo acontecimento se realizar sabendo que o primeiro acontecimento se realizou, não havendo a abstracção que a definição de Kolmogoroff nos patenteia, de a probabilidade condicionada estar definida matematicamente não surgindo da observação da realidade. Uma tal dependência poderá ter levado Pacheco d'Amorim a impor esta condição desnecessária. Contudo, apesar da atenuante representada pela dificuldade de abstracção presente no conceito de probabilidade condicionada na altura, pensamos que Pacheco d'Amorim poderia ter ido mais longe se não tivesse restringido escusadamente a sua definição de probabilidade [condicionada].

Pacheco d'Amorim continua a sua obra referindo que qualquer probabilidade se situa entre zero e a unidade, considerando que, no caso de a classe favorável ser a classe possível, teremos $\mathbb{P}_{\mathbf{A}}(\mathbf{A}) = 1$ tendo a probabilidade o nome de *certeza*, caso seja *nula* (conjunto vazio), teremos $\mathbb{P}_{\mathbf{A}}(\emptyset) = 0$ que tem o nome de *impossibilidade* e, conseqüentemente, teremos para quaisquer classes $\mathbf{A}'' \subset \mathbf{A}' \subset \mathbf{A}$

$$0 \leq \mathbb{P}_{\mathbf{A}'}(\mathbf{A}'') \leq 1. \quad (3.29)$$

Em forma de Postulado, o autor considera que dois sistemas de tiragens são equivalentes se derem origem a elementos qualitativamente iguais e a cada classe qualitativamente igual for associada a mesma probabilidade nos dois sistemas de tiragens. Desta forma Pacheco d'Amorim reduz todos os sistemas de tiragens a tiragens realizadas numa só classe. Em relação a tiragens efectuadas em classes compostas, esta afirmação faz todo o sentido, pois, por exemplo, tirar um elemento da classe $\mathbf{A}$ e outro da classe $\mathbf{B}$ é o mesmo que tirar um só elemento da classe $\mathbf{A} \times \mathbf{B}$. Contudo, esta afirmação de Pacheco d'Amorim torna-se algo estranha quando pensamos em tiragens

efectuadas em complexos de classes, pois parece-nos que um complexo de classes não pode ser decomponível, *a priori*, num único lançamento numa só classe. Esta posição permite-nos pensar que o autor pretende dizer com este Postulado que, como cada elemento (a_i, b_k^i) de um complexo de classes pode ser considerado um elemento composto da classe $a_i \times \mathbf{B}_i$, podemos então definir a sua possibilidade como de um só lançamento se tratasse nessa classe composta. Deste modo, para Pacheco d'Amorim, os lançamentos efectuados em complexos de classes podem ser tratados, se condicionados ao resultado do primeiro lançamento em $\mathbf{A}$, como se de um lançamento numa classe composta se tratasse.

Exemplo 3.4 (Urnas de Pólya). Consideremos uma urna com m bolas das quais m_1 são brancas e m_2 pretas ($m_1 + m_2 = m$). Retirem-se n bolas desta urna da seguinte forma: retira-se uma bola, anota-se a sua cor, repõe-se a bola na urna conjuntamente com mais k bolas da mesma cor.

Desta forma, a composição da urna, quando é retirada a segunda bola, depende da bola que já foi retirada, logo, estamos perante um complexo de classes. Representando por $\mathbf{B}_i$ o acontecimento sair bola branca na i -ésima tiragem e por $\mathbf{P}_i$ sair bola preta, obtemos para as possibilidades da primeira bola

$$\varpi_{\mathbf{B}_1} = \frac{m_1}{m}, \quad \varpi_{\mathbf{P}_1} = \frac{m_2}{m},$$

sendo as possibilidades da segunda bola, pela aditividade das possibilidades, uma vez que existem duas sequências favoráveis à saída de cada cor quando é retirada a segunda bola da urna⁽¹⁰⁾, obtidas através de

$$\varpi_{\mathbf{B}_2} = \varpi_{(\mathbf{B}_1, \mathbf{B}_2)} + \varpi_{(\mathbf{P}_1, \mathbf{B}_2)} = \frac{m_1 + k}{m + k} \cdot \frac{m_1}{m} + \frac{m_1}{m + k} \cdot \frac{m_2}{m} = \frac{m_1}{m}$$

⁽¹⁰⁾ Representaremos por $(\mathbf{B}_1, \mathbf{P}_2)$ uma sequência de duas bolas retiradas, neste caso particular, a primeira bola ser branca e a segunda ser preta.

e, de forma semelhante,

$$\varpi_{\mathbf{P}_2} = \varpi_{(\mathbf{P}_1, \mathbf{P}_2)} + \varpi_{(\mathbf{B}_1, \mathbf{P}_2)} = \frac{m_2 + k}{m + k} \cdot \frac{m_2}{m} + \frac{m_2}{m + k} \cdot \frac{m_1}{m} = \frac{m_2}{m}.$$

Estes resultados significam que, se não conhecermos a cor da primeira bola retirada (só temos conhecimento da constituição inicial da urna), a possibilidade de a segunda bola retirada ser branca é igual à possibilidade de a primeira bola ser branca. Assim, tendo em conta que podemos generalizar este resultado por indução, a possibilidade de sair bola branca [ou preta] é sempre a mesma, apesar de a constituição da urna ser variável, desde que não tenhamos qualquer informação sobre quais foram as cores das bolas que saíram previamente.

As probabilidades da segunda bola retirada em relação ao resultado da primeira bola são obtidos, por definição, pelo rácio das respectivas possibilidades obtendo-se:

$$\begin{aligned} \mathbb{P}_{\mathbf{B}_1}(\mathbf{B}_1, \mathbf{B}_2) &= \frac{m_1 + k}{m + k}; & \mathbb{P}_{\mathbf{P}_1}(\mathbf{P}_1, \mathbf{B}_2) &= \frac{m_1}{m + k}; \\ \mathbb{P}_{\mathbf{P}_1}(\mathbf{P}_1, \mathbf{P}_2) &= \frac{m_2 + k}{m + k}; & \mathbb{P}_{\mathbf{B}_1}(\mathbf{B}_1, \mathbf{P}_2) &= \frac{m_2}{m + k}. \end{aligned}$$

Representemos por $\mathbf{B}^{n_1}$ o acontecimento constituído por todas as sequências onde saem n_1 bolas brancas (e n_2 pretas) das n bolas retiradas ($n_1 + n_2 = n$) e por $\mathbf{B}_i^{n_1}$ cada uma das $\binom{n}{n_1}$ sequências que verificam esta condição, verificando-se

$$\mathbf{B}^{n_1} = \bigcup_{i=1}^{\binom{n}{n_1}} \mathbf{B}_i^{n_1}.$$

Facilmente se verifica, pela possibilidade das classes compostas, que a possibilidade de qualquer sequência com n_1 bolas brancas em n bolas retiradas é dada por

$$\begin{aligned} \varpi_{\mathbf{B}_i^{n_1}} &= \frac{m_1}{m} \cdot \frac{m_1 + k}{m + k} \cdots \frac{m_1 + (n_1 - 1)k}{m + (n_1 - 1)k} \cdot \frac{m_2}{m + n_1 k} \cdot \\ &\quad \cdot \frac{m_2 + k}{m + (n_1 + 1)k} \cdots \frac{m_2 + (n_2 - 1)k}{m + (n - 1)k} \end{aligned}$$

e, como existem $\binom{n}{n_1}$ seqüências que verificam esta condição, teremos

$$\begin{aligned} \varpi_{\mathbf{B}^{n_1}} &= \sum_{i=1}^{\binom{n}{n_1}} \varpi_{\mathbf{B}_i^{n_1}} = \binom{n}{n_1} \cdot \varpi_{\mathbf{B}_i^{n_1}} = \\ &= \frac{n! m_1 (m_1 + k) \cdots [m_1 + (n_1 - 1) k] \cdot m_2 (m_2 + k) \cdots [m_2 + (n_2 - 1) k]}{n_1! n_2! m (m + k) \cdots [m + (n - 1) k]} \end{aligned}$$

que corresponde à função de probabilidade de uma variável aleatória com distribuição de Pólya. Como casos particulares, mais conhecidos, teremos as situações $k = 0$, que corresponde à distribuição binomial⁽¹¹⁾, e $k = -1$ que corresponde à distribuição hipergeométrica (a bola retirada não é recolocada da urna, isto é, as bolas são retiradas sem reposição). ♦

3.4 Teorema da Probabilidade Total e Composta

A Proposição V, da Probabilidade Total, diz que, considerando que as classes $\mathbf{A}_1'', \dots, \mathbf{A}_n''$ formam uma partição de uma classe $\mathbf{A}''$, podemos determinar a probabilidade da classe $\mathbf{A}''$ pela adição das probabilidades das classes que constituem a sua partição, todas em relação [condicionadas] à mesma classe possível, ou seja,

$$\mathbb{P}_{\mathbf{A}'}(\mathbf{A}'') = \sum_{i=1}^n \mathbb{P}_{\mathbf{A}'}(\mathbf{A}_i'') \quad (3.30)$$

⁽¹¹⁾ Nesta situação estamos a trabalhar com classes compostas, não sendo necessário recorrer aos complexos de classes propostos por Pacheco d'Amorim, pois corresponde à única situação onde a composição da urna não depende das bolas anteriormente retiradas, isto é, a composição da urna é fixa. Esta situação, pela sua importância fundamental na Teoria da Probabilidade e Estatística nesta construção de Pacheco d'Amorim, será analisada mais detalhadamente no capítulo TEOREMAS DE JACOB BERNOULLI E LEI DOS DESVIOS.

pois, de facto, pela aditividade das possibilidades (Proposição II) temos que

$$\varpi_{\mathbf{A}''} = \sum_{i=1}^n \varpi_{\mathbf{A}_i''} \quad (3.31)$$

e, conseqüentemente, verificar-se-á

$$\mathbb{P}_{\mathbf{A}'}(\mathbf{A}'') = \frac{\varpi_{\mathbf{A}''}}{\varpi_{\mathbf{A}'}} = \sum_{i=1}^n \frac{\varpi_{\mathbf{A}_i''}}{\varpi_{\mathbf{A}'}} = \sum_{i=1}^n \mathbb{P}_{\mathbf{A}'}(\mathbf{A}_i''). \quad (3.32)$$

Pacheco d'Amorim apresenta o Teorema da Probabilidade Composta, Proposição VI, na seguinte forma

$$\mathbb{P}_{\mathbf{A}' \times \mathbf{B}'}(\mathbf{A}'' \times \mathbf{B}'') = \mathbb{P}_{\mathbf{A}'}(\mathbf{A}'') \cdot \mathbb{P}_{\mathbf{B}'}(\mathbf{B}''), \quad (3.33)$$

onde $\mathbf{A}' \times \mathbf{B}'$ e $\mathbf{A}'' \times \mathbf{B}''$ são classes compostas⁽¹²⁾. A demonstração deste resultado é imediata, uma vez que

$$\mathbb{P}_{\mathbf{A}' \times \mathbf{B}'}(\mathbf{A}'' \times \mathbf{B}'') = \frac{\varpi_{\mathbf{A}'' \times \mathbf{B}''}}{\varpi_{\mathbf{A}' \times \mathbf{B}'}} = \frac{\varpi_{\mathbf{A}''}}{\varpi_{\mathbf{A}'}} \cdot \frac{\varpi_{\mathbf{B}''}}{\varpi_{\mathbf{B}'}} = \mathbb{P}_{\mathbf{A}'}(\mathbf{A}'') \cdot \mathbb{P}_{\mathbf{B}'}(\mathbf{B}''). \quad (3.34)$$

O autor considera que esta proposição pode ser extendida para tiragens em complexos de classe em duas situações distintas. A primeira situação corresponde ao caso em que a classe possível é igual à classe total possível, isto é, $\mathbf{A}' \odot \rightarrow \mathbf{B}' = \mathbf{A} \odot \rightarrow \mathbf{B}$, pois neste caso (Prop. III) $\varpi_{\mathbf{A}' \odot \rightarrow \mathbf{B}'} = \varpi_{\mathbf{A} \odot \rightarrow \mathbf{B}} = 1$ e, deste facto, resulta que

$$\mathbb{P}_{\mathbf{A}' \odot \rightarrow \mathbf{B}'}(\mathbf{A}'' \times \mathbf{B}'') = \frac{\varpi_{\mathbf{A}'' \times \mathbf{B}''}}{\varpi_{\mathbf{A}' \odot \rightarrow \mathbf{B}'}} = \frac{\varpi_{\mathbf{A}''} \cdot \varpi_{\mathbf{B}''}}{1} = \varpi_{\mathbf{A}''} \cdot \varpi_{\mathbf{B}''}, \quad (3.35)$$

onde a classe favorável $\mathbf{A}'' \times \mathbf{B}''$ continua a ser uma classe composta. A segunda situação considerada por Pacheco d'Amorim corresponde ao caso onde a classe possível $\mathbf{A}' \odot \rightarrow \mathbf{B}'$ é obtida da classe total possível $\mathbf{A} \odot \rightarrow \mathbf{B}$

⁽¹²⁾ Note-se que nesta Proposição Pacheco d'Amorim não faz referência à classe total possível, se esta é uma classe composta ou se pode ser um complexo de classes. Contudo, consideramos que, para deduzir os resultados apresentados não é necessário impor qualquer restrição à classe total possível.

retirando unicamente alguns elementos de $\mathbf{A}$ (e as correspondentes classes $\mathbf{B}_i$ que lhe estão associadas), isto é,

$$\mathbf{A}' \odot \mathbf{B}' = \bigcup_{a_i \in \mathbf{A}'} a_i \times \mathbf{B}_i, \quad (3.36)$$

onde os elementos a_i de $\mathbf{A}'$ têm associados todos os possíveis elementos da classe $\mathbf{B}_i$. Por este motivo vamos obter

$$\begin{aligned} \varpi_{\mathbf{A}' \odot \mathbf{B}'} &= \sum_{a_i \in \mathbf{A}'} \sum_{b_j^i \in \mathbf{B}'} \pi_{(a_i, b_j^i)} = \sum_{a_i \in \mathbf{A}'} \sum_{b_j^i \in \mathbf{B}_i} \frac{1}{n \cdot n_i} = \\ &= \sum_{a_i \in \mathbf{A}'} \frac{n_i}{n \cdot n_i} = \sum_{a_i \in \mathbf{A}'} \frac{1}{n_i} = \sum_{a_i \in \mathbf{A}'} \pi_{a_i} = \varpi_{\mathbf{A}'}. \end{aligned} \quad (3.37)$$

Assim, podemos concluir que

$$\mathbb{P}_{\mathbf{A}' \odot \mathbf{B}'}(\mathbf{A}'' \times \mathbf{B}'') = \frac{\varpi_{\mathbf{A}'' \times \mathbf{B}''}}{\varpi_{\mathbf{A}' \odot \mathbf{B}'}} = \frac{\varpi_{\mathbf{A}''} \cdot \varpi_{\mathbf{B}''}}{\varpi_{\mathbf{A}'}} = \mathbb{P}_{\mathbf{A}'}(\mathbf{A}'') \cdot \varpi_{\mathbf{B}''}. \quad (3.38)$$

Pacheco d'Amorim apresenta estes resultados de forma distinta, utilizando a fórmula inicial $\mathbb{P}_{\mathbf{A}' \times \mathbf{B}'}(\mathbf{A}'' \times \mathbf{B}'') = \mathbb{P}_{\mathbf{A}'}(\mathbf{A}'') \cdot \mathbb{P}_{\mathbf{B}'}(\mathbf{B}'')$ em ambas as generalizações. Contudo, consideramos que esta fórmula não é adequada para a situação em que a classe possível é um complexo de classes, pois a probabilidade $\mathbb{P}_{\mathbf{B}'}(\mathbf{B}'')$ não está definida ou, dito de forma mais precisa, se $\mathbf{A}' \odot \mathbf{B}'$ é um complexo de classes, a classe $\mathbf{B}'$ não está definida sem a sua ligação à classe $\mathbf{A}'$. Notemos que, por definição de probabilidade, temos

$$\mathbb{P}_{\mathbf{B}'}(\mathbf{B}'') = \frac{\varpi_{\mathbf{B}''}}{\varpi_{\mathbf{B}'}} \quad (3.39)$$

contudo, como poderemos determinar a possibilidade da classe $\mathbf{B}'$ que é um conjunto de conjuntos $\mathbf{B}_i$? Não está definida.

Exemplo 3.5. Seja $\mathbf{A} \odot \mathbf{B}$ onde $\mathbf{A} = \{a_1, \dots, a_n\}$ e a cada elemento $a_i \in \mathbf{A}$ associa-se um conjunto $\mathbf{B}_{a_i} = \mathbf{B}_i = \{b_1^i, \dots, b_{n_i}^i\}$. Seja $\mathbf{A}' \odot \mathbf{B}' \subset \mathbf{A} \odot \mathbf{B}$ um complexo de classes onde $\mathbf{A}' = \{a_1, \dots, a_k\}$ com $k < n$ e $\mathbf{B}'$ é definido de forma a que a cada elemento $a_i \in \mathbf{A}'$ seja associado o conjunto $\mathbf{B}_i$, isto é,

aos elementos $a_i \in \mathbf{A}'$ associa-se a mesma classe que na classe total possível. Neste caso como poderemos determinar $\varpi_{\mathbf{B}'}$?

Para clarificar esta ideia, consideremos que no complexo de classes $\mathbf{A} \odot \mathbf{B}$ as classes $\mathbf{B}_i$ verificam $\#\mathbf{B}_i = n_i = i$ e que pretendemos determinar a possibilidade da classe

$$\mathbf{A}' \odot \mathbf{B}' = \left\{ \left(a_1, b_1^1 \right), \left(a_2, b_1^2 \right), \left(a_2, b_2^2 \right), \left(a_3, b_1^3 \right), \left(a_3, b_2^3 \right), \left(a_3, b_3^3 \right) \right\}.$$

Nesta situação é impossível determinar $\varpi_{\mathbf{B}'}$, pois, como a possibilidade de cada elemento $b_i^k \in \mathbf{B}_i$ é dada por

$$\pi_{b_i^k} = \frac{1}{\#\mathbf{B}_i} = \frac{1}{i}, \quad (3.40)$$

obtemos, recorrendo à definição de possibilidade de uma classe,

$$\varpi_{\mathbf{B}'} = \sum_{i=1}^3 \sum_{j=1}^i \pi_{b_j^i} = 1 + 2 \times \frac{1}{2} + 3 \times \frac{1}{3} = 3, \quad (3.41)$$

o que, naturalmente, não faz sentido. Consequentemente não é possível determinar $\mathbb{P}_{\mathbf{B}'}(\mathbf{B}'')$. ◆

Pacheco d'Amorim considera, em ambas as generalizações, que $\varpi_{\mathbf{B}'} = 1$ pelo facto de a classe $\mathbf{B}'$ incluir todos os elementos possíveis em relação aos elementos $a_i \in \mathbf{A}'$. Porém, pensamos que $\varpi_{\mathbf{B}'}$ não está definido e, como tal, existem duas situações distintas. A primeira verifica-se quando o conjunto $\mathbf{B}'$ é constituído unicamente por elementos, isto é, a todos os elementos a_i da classe $\mathbf{A}'$ é associada a mesma classe $\mathbf{B}'$ e o complexo de classes $\mathbf{A}' \odot \mathbf{B}'$ pode ser encarado como uma classe composta. Nesta situação estamos nas condições iniciais do Teorema da Probabilidade Composta e não será necessário recorrer a uma generalização. Na segunda situação a classe $\mathbf{B}'$ é constituída por vários conjuntos, sendo associado, a cada elemento a_i de $\mathbf{A}'$, um conjunto $\mathbf{B}_i$ distinto, ou seja, o complexo de classes $\mathbf{A}' \odot \mathbf{B}'$ não pode ser reduzido a uma classe composta. Neste caso $\mathbb{P}_{\mathbf{B}'}(\mathbf{B}'')$ não está definido, pois não faz

sentido analisar o conjunto $\mathbf{B}'$ sem o utilizar conjuntamente com a classe $\mathbf{A}'$, por dela depender.

Saliente-se que as generalizações do Teorema da Probabilidade Composta que Pacheco d'Amorim realiza se referem a situações onde a classe possível pode ser um complexo de classes (sob determinadas condições anteriormente enunciadas), mas a classe favorável continua a ser obrigatoriamente uma classe composta. No entanto, denota-se na apresentação do autor uma enorme confusão pelo facto de este usar uma notação onde é impossível diferir as classes compostas dos complexos de classes.

Exemplo 3.6. Consideremos uma classe $\mathbf{A} = \{a_1, \dots, a_6\}$ onde a cada elemento a_i é associada uma classe $\mathbf{B}_i$ com $\#\mathbf{B}_i = i$, sendo os seus elementos representados por b_j^i para $i = 1, \dots, 6$ e $j = 1, \dots, i$. Nesta situação o complexo de classes total possível é

$$\begin{aligned} \mathbf{A} \odot \mathbf{B} = & \left\{ (a_1, b_1^1), (a_2, b_1^2), (a_2, b_2^2), (a_3, b_1^3), (a_3, b_2^3), \right. \\ & (a_3, b_3^3), (a_4, b_1^4), (a_4, b_2^4), (a_4, b_3^4), (a_4, b_4^4), \\ & (a_5, b_1^5), (a_5, b_2^5), (a_5, b_3^5), (a_5, b_4^5), (a_5, b_5^5), \\ & \left. (a_6, b_1^6), (a_6, b_2^6), (a_6, b_3^6), (a_6, b_4^6), (a_6, b_5^6), (a_6, b_6^6) \right\} \end{aligned}$$

que, para facilitar, representaremos por $\mathbb{A}$. Para cada elemento possível a sua possibilidade é determinada por

$$\pi_{(a_i, b_j^i)} = \pi_{a_i} \cdot \pi_{b_j^i} = \frac{1}{\#\mathbf{A}} \cdot \frac{1}{\#\mathbf{B}_i} = \frac{1}{6} \cdot \frac{1}{i},$$

para $i = 1, \dots, 6$ e $j = 1, \dots, i$. Para determinar a possibilidade de qualquer classe contida na classe total possível, pela definição, será somar as possibilidades dos elementos que pertencem a essa classe. Deste modo, se considerarmos o complexo de classes $\mathbf{A}' \subset \mathbb{A}$ definido por

$$\begin{aligned} \mathbf{A}' = \mathbf{A}' \odot \mathbf{B}' = & \left\{ (a_1, b_1^1), (a_3, b_1^3), (a_3, b_2^3), (a_3, b_3^3), \right. \\ & \left. (a_5, b_1^5), (a_5, b_2^5), (a_5, b_3^5), (a_5, b_4^5), (a_5, b_5^5) \right\} \end{aligned}$$

teremos

$$\begin{aligned}\varpi_{\mathbb{A}'} &= \sum_{(a_i, b_j^i) \in \mathbb{A}'} \pi_{(a_i, b_j^i)} = \pi_{(a_1, b_1^1)} + \pi_{(a_3, b_1^3)} + \cdots + \pi_{(a_5, b_5^5)} = \\ &= \frac{1}{6} \cdot \frac{1}{1} + \frac{1}{6} \cdot \frac{1}{3} + \cdots + \frac{1}{6} \cdot \frac{1}{5} = \frac{45}{90} = 0,5.\end{aligned}$$

A probabilidade de $\mathbb{A}'$ em relação a $\mathbb{A}$ é dada por

$$\mathbb{P}_{\mathbb{A}}(\mathbb{A}') = \frac{\varpi_{\mathbb{A}'}}{\varpi_{\mathbb{A}}} = \frac{0,5}{1} = 0,5,$$

pois a probabilidade de uma classe condicionada à classe total possível é igual à sua possibilidade, tal como, utilizando os conceitos actuais, qualquer probabilidade condicionada ao universo é igual à probabilidade incondicionada, isto é, $\mathbb{P}(\mathbf{A} \mid \Omega) = \mathbb{P}(\mathbf{A})$. Seja o complexo de classes

$$\mathbb{A}'' = \left\{ (a_1, b_1^1), (a_3, b_1^3), (a_5, b_1^5) \right\} \subset \mathbb{A}',$$

então a probabilidade de $\mathbb{A}''$ em relação a $\mathbb{A}'$ é dada por

$$\mathbb{P}_{\mathbb{A}'}(\mathbb{A}'') = \frac{\varpi_{\mathbb{A}''}}{\varpi_{\mathbb{A}'}} = \frac{\frac{1}{6} + \frac{1}{18} + \frac{1}{30}}{0,5} = \frac{23}{45}.$$

Determinemos agora a possibilidade dos complexos de classes $\mathbb{C}_i$, com $i = 1, \dots, 6$ definidos por

$$\begin{aligned}\mathbb{C}_1 &= \left\{ (a_1, b_1^1), (a_2, b_1^2), (a_3, b_1^3), (a_4, b_1^4), (a_5, b_1^5), (a_6, b_1^6) \right\}; \\ \mathbb{C}_2 &= \left\{ (a_2, b_2^2), (a_3, b_2^3), (a_4, b_2^4), (a_5, b_2^5) \right\}; \\ &\dots \\ \mathbb{C}_6 &= \left\{ (a_6, b_6^6) \right\}\end{aligned}$$

sendo a possibilidade de cada uma destas classes dada por

$$\begin{aligned}
 \varpi_{c_1} &= \frac{1}{6} + \frac{1}{12} + \frac{1}{18} + \frac{1}{24} + \frac{1}{30} + \frac{1}{36} = \frac{147}{360}; \\
 \varpi_{c_2} &= \frac{1}{12} + \frac{1}{18} + \frac{1}{24} + \frac{1}{30} + \frac{1}{36} = \frac{87}{360}; \\
 \varpi_{c_3} &= \frac{1}{18} + \frac{1}{24} + \frac{1}{30} + \frac{1}{36} = \frac{57}{360}; \\
 \varpi_{c_4} &= \frac{1}{24} + \frac{1}{30} + \frac{1}{36} = \frac{37}{360}; \\
 \varpi_{c_5} &= \frac{1}{30} + \frac{1}{36} = \frac{22}{360}; \\
 \varpi_{c_6} &= \frac{1}{36} = \frac{10}{360}
 \end{aligned}$$

Note-se que estes complexos de classes constituem uma partição da classe total possível, logo, a soma das suas possibilidades é igual à unidade. $\blacklozenge$

Pacheco d'Amorim na sua Proposição VII, considerando $\mathbf{A}'' \subset \mathbf{A}' \subset \mathbf{A}$, afirma que

$$\mathbb{P}_{\mathbf{A}}(\mathbf{A}'') = \mathbb{P}_{\mathbf{A}}(\mathbf{A}') \times \mathbb{P}_{\mathbf{A}'}(\mathbf{A}''). \quad (3.42)$$

De facto, utilizando a sua definição de probabilidade, este resultado é imediato. O autor apresenta, como corolário,

$$\mathbb{P}_{\mathbf{A}'}(\mathbf{A}'') = \frac{\mathbb{P}_{\mathbf{A}}(\mathbf{A}'')}{\mathbb{P}_{\mathbf{A}}(\mathbf{A}')} . \quad (3.43)$$

Notemos que esta propriedade, Proposição VII, não é válida se utilizarmos a actual definição de probabilidade condicionada, isto é, ela só é verdadeira pelo facto de a definição de probabilidade de Pacheco d'Amorim corresponder à actual definição de probabilidade condicionada restrita às situações onde o condicionado é subconjunto do condicionante, pois, caso contrário, esta proposição (geralmente) não se verificaria.

3.5 Teorema de Bayes

Pacheco d'Amorim apresenta, como Proposição VIII, a resolução do problema das causas (Teorema de Bayes), considerando que quando as tiragens, à sorte, se efectuarem num complexo de classes $\mathbf{A} \odot \rightarrow \mathbf{B}$, à classe $\mathbf{A}$ chama-se *causa* e às classes $\mathbf{B}_i$ de $\mathbf{B}$, *efeitos*. O problema chamado das probabilidades das causas pode ser tipificado pela situação a seguir apresentada.

Problema 3.1 (Problema das probabilidades das causas). Considere-se um conjunto de urnas que contêm bolas brancas e pretas. Sejam $\omega_1, \omega_2, \dots, \omega_n$ as probabilidades *a priori* das n urnas (causas) que dão lugar à saída de uma bola branca com probabilidade, respectivamente, $p_1, p_2, \dots, p_n$. Tire-se, à sorte, uma urna e, da urna que sair, tire-se, à sorte, uma bola que, por hipótese, sai branca. Qual a probabilidade $\mathbb{P}_i$ de que a bola tirada pertença a uma urna cuja probabilidade de sair bola branca seja p_i ?

Teorema 3.1 (Teorema de Bayes). *Designa-se por $\mathbf{A} \odot \rightarrow \mathbf{B}$ a classe (complexo de classes) que se obtém associando cada uma das causas a cada uma das bolas a que ela pode dar origem — classe total possível; por $\mathbf{A}' \odot \rightarrow \mathbf{B}'$ a classe (complexo de classes) que se obtém associando cada uma das causas às bolas brancas a que ela pode dar origem; por $a_i \times \mathbf{B}_i''$ a classe (composta) que se obtém associando cada causa de ordem i com as bolas brancas a que pode dar origem, então⁽¹³⁾*

$$\mathbb{P}_{\mathbf{A}' \odot \rightarrow \mathbf{B}'}(a_i \times \mathbf{B}_i'') = \frac{\omega_i p_i}{\sum_{j=1}^n \omega_j p_j}. \quad (3.44)$$

Demonstração. Notemos que $a_i \times \mathbf{B}_i'' \subset \mathbf{A}' \odot \rightarrow \mathbf{B}' \subset \mathbf{A} \odot \rightarrow \mathbf{B}$, logo, podemos recorrer à Proposição VII, isto é, ao facto de $\mathbb{P}_{\mathbf{A}}(\mathbf{A}'') = \mathbb{P}_{\mathbf{A}}(\mathbf{A}') \times \mathbb{P}_{\mathbf{A}'}(\mathbf{A}'')$

⁽¹³⁾ Notemos que, pelo enunciado do problema das causas, $\varpi_{\mathbf{B}_i''} = p_i$ e $\varpi_{\{a_i\}} = \pi_{a_i} = \omega_i$.

quando $\mathbf{A}'' \subset \mathbf{A}' \subset \mathbf{A}$, obtendo-se

$$\mathbb{P}_{\mathbf{A}' \odot \rightarrow \mathbf{B}'} (a_i \times \mathbf{B}_i'') = \frac{\mathbb{P}_{\mathbf{A} \odot \rightarrow \mathbf{B}} (a_i \times \mathbf{B}_i'')}{\mathbb{P}_{\mathbf{A} \odot \rightarrow \mathbf{B}} (\mathbf{A}' \odot \rightarrow \mathbf{B}')}. \quad (3.45)$$

Por outro lado, como

$$\mathbf{A}' \odot \rightarrow \mathbf{B}' = \bigcup_{j=1}^n a_j \times \mathbf{B}_j'', \quad (3.46)$$

pelo Teorema da Probabilidade Total tem-se

$$\mathbb{P}_{\mathbf{A} \odot \rightarrow \mathbf{B}} (\mathbf{A}' \odot \rightarrow \mathbf{B}') = \sum_{j=1}^n \mathbb{P}_{\mathbf{A} \odot \rightarrow \mathbf{B}} (a_j \times \mathbf{B}_j''). \quad (3.47)$$

Finalmente, visto que $\mathbf{A} \odot \rightarrow \mathbf{B}$ é a classe total possível, podemos aplicar a generalização do Teorema da Probabilidade Composta para o caso em que a classe possível é a classe total possível, obtendo-se

$$\mathbb{P}_{\mathbf{A} \odot \rightarrow \mathbf{B}} (a_i \times \mathbf{B}_i'') = \varpi_{\{a_i\}} \cdot \varpi_{\mathbf{B}_i''} = \omega_i p_i. \quad (3.48)$$

Assim, conclui-se que

$$\begin{aligned} \mathbb{P}_{\mathbf{A}' \odot \rightarrow \mathbf{B}'} (a_i \times \mathbf{B}_i'') &= \frac{\mathbb{P}_{\mathbf{A} \odot \rightarrow \mathbf{B}} (a_i \times \mathbf{B}_i'')}{\mathbb{P}_{\mathbf{A} \odot \rightarrow \mathbf{B}} (\mathbf{A}' \odot \rightarrow \mathbf{B}')} = \\ &= \frac{\mathbb{P}_{\mathbf{A} \odot \rightarrow \mathbf{B}} (a_i \times \mathbf{B}_i'')}{\sum_{j=1}^n \mathbb{P}_{\mathbf{A} \odot \rightarrow \mathbf{B}} (a_j \times \mathbf{B}_j'')} = \frac{\omega_i p_i}{\sum_{j=1}^n \omega_j p_j}. \end{aligned} \quad (3.49)$$

□

Pacheco d'Amorim considera que as demonstrações usuais do Teorema de Bayes, suas contemporâneas, estão incorrectas, afirmando que a sua demonstração

“mostra bem como é errónea a dedução que desta fórmula se faz nos livros de probabilidades.”

[Pacheco d'Amorim, 1914, p. 27]

Desta forma, Pacheco d'Amorim reclama apresentar a primeira demonstração completa do Teorema de Bayes, baseada numa extensão do Teorema da Probabilidade Composta a um complexo de classes. No entanto, já Laplace tinha redescoberto⁽¹⁴⁾ o Teorema de Bayes numa publicação de 1774, onde o resultado é deduzido a partir do seu princípio (axioma) da probabilidade inversa. Em 1781 demonstra este resultado sem recurso ao referido axioma, simplificando a sua demonstração em 1786. Em todas estas demonstrações Laplace considera unicamente o caso de causas com igual probabilidade [Hald (1998, p. 164–165)]. Finalmente, em 1812, na sua sublime obra *Théorie Analytique des Probabilités*, Laplace generaliza o resultado para causas com diferentes probabilidades (assumindo para a causa C_i uma probabilidade genérica $\mathbb{P}(C_i)$ sem a explicar — Laplace (1812, p. 180–182)) obtendo, neste caso, a forma do Teorema de Bayes actualmente usada (ideia também presente em Hald (1998, p. 161,165)). Laplace (1812, p. 182) faz a demonstração considerando que as causas têm probabilidades iguais ($\mathbb{P}(C_i) = \frac{1}{n}$) e depois generaliza referindo que, tendo as causas probabilidades distintas, a fórmula obtém-se utilizando o mesmo raciocínio⁽¹⁵⁾.

Bertrand (1888), apesar de a definição de probabilidade presente na sua obra também se restringir às situações de equiprobabilidade e de a sua demonstração ser bastante distinta da de Pacheco d'Amorim, também apresenta o problema de forma geral, aliás, de uma forma muito semelhante à de Pacheco d'Amorim.

⁽¹⁴⁾ Provavelmente Laplace, na altura, não teria conhecimento do artigo de Bayes (1764), onde este resultado foi pela primeira vez apresentado, como referem Hald (1998, p. 161) e Stigler na tradução do artigo de Laplace (1774, p. 359).

⁽¹⁵⁾ Notemos que, nesta demonstração, Laplace fornece mais um exemplo do que foi referido anteriormente em relação à sua definição de probabilidade, que aceita e usa probabilidades distintas (ou não baseadas em equiprobabilidade), mas nunca explicita o seu fundamento.

“Diverses causes $E_1, E_2, \dots, E_n$ ont pu produire un événement observé. Les probabilités de ces causes, lorsque le résultat n’était pas encore connu, étaient $\varpi_1, \varpi_2, \dots, \varpi_n$. L’événement se produit; la cause E_i , lorsqu’on est certain que c’est elle qui agit, donne à l’événement la probabilité p_i . Quelle est la probabilité de chacune des causes qui sont, on l’admet, les seules possibles?”

[Bertrand, 1888, p. 140]

Deste modo, em diversas obras contemporâneas a Pacheco d’Amorim o Teorema de Bayes é apresentado e demonstrado na sua forma geral. Assim, consideramos que Bertrand (1888, p. 140–142), Poincaré (1896, p. 154–156), Bachelier (1912, p. 485), Borel (1909, p. 145; 1914, p. 99), apesar de apresentarem o Teorema de Bayes de uma forma menos elaborada que Pacheco d’Amorim, demonstram-no, sem dificuldade, nas suas obras. Mesmo restringindo-nos à literatura nacional, também Rodolpho Guimarães (1904, p. 29–30) apresenta o Teorema de Bayes e correspondente demonstração na forma geral.

Em suma, consideramos que estas demonstrações não estão erradas, nem as realizadas pelos seus contemporâneos nem a apresentada por Laplace um século antes. Parece-nos que a abordagem de Pacheco d’Amorim apenas aparentemente vai mais longe, pois com uma definição mais geral de probabilidade condicional o caso geral decorre facilmente (a discussão é mesmo irrelevante no contexto operacional do conceito de probabilidade na axiomática de Kolmogoroff). De facto, para Pacheco d’Amorim, a definição de probabilidade condicional foi restringida, como anteriormente indicado, mas bastaria generalizá-la para a situação em que $\mathbf{A} \not\subset \mathbf{B}$ através de $\mathbb{P}(\mathbf{A} \mid \mathbf{B}) = \mathbb{P}(\mathbf{A} \cap \mathbf{B} \mid \mathbf{B})$ para obter a definição de probabilidade condicionada que é normalmente adoptada. Como o conceito de probabilidade condicional utilizado por Pacheco d’Amorim se restringe à situação $\mathbf{A} \subset \mathbf{B}$,

isso obriga-o a uma construção morosa de uma extensão à situação de classes complexas.

Refira-se, contudo, que uma das justificações do autor para esta afirmação é a definição de probabilidade usada ser baseada na equiprobabilidade, considerando que

“seria impossível deduzir a fórmula de Bayes com a definição de probabilidade dada nesses livros, pela simples razão de que a fórmula de Bayes se refere a um caso de probabilidade que não está previsto nessa definição, a classe possível ser menor do que a classe total possível”

[Pacheco d'Amorim, 1914, p. 27]

Será que Pacheco d'Amorim considera que na concepção de probabilidade condicionada de Laplace a *classe possível* não é menor que a *classe total possível*? A afirmação afigura-se-nos deveras estranha.

Todavia, Pacheco d'Amorim não está isolado na sua insatisfação em relação às demonstrações deste teorema realizadas com base na definição *Laplaciana* de probabilidade. Por exemplo, Bertrand (1888) partilha da mesma opinião, referindo em relação aos Teoremas da Probabilidade Total e da Probabilidade Composta (que posteriormente utiliza na demonstração do Teorema de Bayes)

“Les deux théorèmes précédents sont et doivent être incomplètement démontrés. La probabilité, en effet, n'a été définie que pour une classe très restreinte d'événements. Il en existe d'autres, incertains comme eux, dans lesquels l'énumération des cas ne peut rien apprendre. Les principes leur sont-ils applicables? Sont-ils dès à présent démontrés pour eux? Les principes sont applicables. Ils ne sont pas encore démontrés. Comment le seraient-ils? Les

probabilités dont ils donnent la mesure n'ont même été définies."

[Bertrand, 1888, p. 25–26]

Desta forma, Bertrand considera que estes teoremas são válidos (aplicáveis) para qualquer probabilidade, no entanto, considera que só estão verdadeiramente demonstrados para o caso de equiprobabilidade, pois a probabilidade só está definida rigorosamente nestas situações. Possivelmente, foi nesta ideia de Bertrand que Pacheco d'Amorim se apoiou, embora, como anteriormente referimos, consideremos que, apesar de os autores referidos não justificarem a existência de probabilidades distintas, isto é, de eles não fornecerem definições que fundamentem o aparecimento de acontecimentos com probabilidades distintas, eles utilizam-nas. Deste modo, não julgamos necessário para a demonstração do Teorema de Bayes a fundamentação de causas com probabilidades distintas, parece-nos suficiente partir, como muitos o fizeram, de que seja p_i a probabilidade da causa i , sem justificar como este valor é obtido, pois, abstraindo-nos desta falta de justificação da origem de acontecimentos de probabilidades não baseadas em equiprobabilidade, não nos parece haver qualquer problema no raciocínio utilizado nas obras supracitadas. Refira-se, para clarificar a nossa opinião, que a axiomática de Kolmogoroff não justifica o valor de nenhuma probabilidade⁽¹⁶⁾, isto é, ela *per se* não nos permite determinar probabilidades, sejam baseadas em equiprobabilidade ou não, mas, naturalmente, quando com base nela é demonstrado, de forma simples, o Teorema de Bayes, ninguém duvida da sua validade.

Keynes (1921) também partilha da ideia de que as demonstrações do Teorema de Bayes não estão bem fundamentadas, afirmando

"The direct and indeed fundamental dependence of the inverse principle on the rule for compound probabilities was not appre-

⁽¹⁶⁾ Com excepção dos dois extremos, a probabilidade do universo e a do acontecimento vazio ou impossível.

ciated by Laplace. A number of proofs of the theorem have been attempted since Laplace's time, but most of them are not very satisfactory, and generally couched in such a form that they do no more than recommend the plausibility of their thesis. Mr. McColl ("Sixth Paper on the Calculus of Equivalent Statements," Proc. Lond. Math. Soc., 1897, vol. xxviii. p. 567) gave a symbolic proof, (...); and a very similar proof has also been given by A. A. Markoff (Wahrscheinlichkeitsrechnung, p. 178). I am not acquainted with any other rigorous discussion of it."

[Keynes, 1921, p. 176]

Keynes parece salientar o facto de Laplace não se ter apercebido da dependência da fórmula de Bayes em relação ao Teorema da Probabilidade Composta. De facto, actualmente o Teorema de Bayes é uma consequência imediata dos Teoremas da Probabilidade Composta e da Probabilidade Total, ainda que consideremos que esta ideia já está presente nas habituais demonstrações da época (como foi já referido em relação a Bertrand).

Exemplo 3.7. Consideremos as condições enunciadas no exemplo 3.6 (página 193). Suponhamos ainda que os elementos b_j^i só dependem de j , isto é, que representam o mesmo objecto para os diferentes valores de i , isto significa que $b_j^i = b_j, \forall i$. Desta forma o elemento b_1 existe nas seis urnas, enquanto o elemento b_6 só existe na sexta urna. Se retirarmos, à sorte, um elemento da classe **A**, onde cada elemento representa uma urna, e de dentro da urna escolhida retirarmos, à sorte, um elemento (que corresponde a escolher, à sorte, um elemento da classe **B_i** associada ao elemento retirado de **A**) que verificamos ser o elemento b_j , qual a probabilidade de ter sido escolhida a urna a_i ?

Representemos por **A** a classe total possível (que é um complexo de classes

que contém todos os resultados possíveis desta experiência); por

$$\mathbb{A}_j = \bigcup_{k=j}^6 a_k \times b_j \quad (3.50)$$

a classe possível (complexo de classes que contém todas as possibilidades de o elemento b_j sair) e por $\mathbf{A}_i^j$ a classe favorável (classe composta que corresponde à escolha da urna a_i e posterior saída do elemento b_j). Desta forma, as probabilidades pretendidas são $\mathbb{P}_{\mathbb{A}_j}(\mathbf{A}_i^j)$ que são nulas no caso de $i < j$. Para os casos em que $i \geq j$ o resultado é obtido utilizando o Teorema de Bayes, obtendo-se

$$\mathbb{P}_{\mathbb{A}_j}(\mathbf{A}_i^j) = \frac{\mathbb{P}_{\mathbb{A}}(\mathbf{A}_i^j)}{\mathbb{P}_{\mathbb{A}}(\mathbb{A}_j)} = \frac{\frac{1}{6} \cdot \frac{1}{i}}{\sum_{k=j}^6 \frac{1}{6} \cdot \frac{1}{k}} = \frac{\frac{1}{i}}{\sum_{k=j}^6 \frac{1}{k}}. \quad (3.51)$$

Assim, se, por exemplo, sair o elemento b_1 , teremos que a probabilidade de cada urna será dada por

$$\mathbb{P}_{\mathbb{A}_1}(\mathbf{A}_i^1) = \frac{\frac{1}{i}}{\sum_{k=1}^6 \frac{1}{k}} = \frac{\frac{1}{i}}{\frac{1}{1} + \dots + \frac{1}{6}} = \frac{20}{49 i}, \quad (3.52)$$

que obrigatoriamente a sua soma para todos os valores de i é igual à unidade.

Urnas	Urna 1	Urna 2	Urna 3	Urna 4	Urna 5	Urna 6
$\mathbb{P}_{\mathbb{A}_1}(\mathbf{A}_i^1)$	$\frac{20}{49}$	$\frac{20}{98}$	$\frac{20}{147}$	$\frac{20}{196}$	$\frac{20}{245}$	$\frac{20}{294}$

No caso de sair o elemento b_5 , a probabilidade das primeiras quatro urnas é nula (estas urnas não têm o elemento b_5) sendo a probabilidade da quinta e sexta urnas dada por

$$\mathbb{P}_{\mathbb{A}_1}(\mathbf{A}_i^5) = \frac{\frac{1}{i}}{\sum_{k=5}^6 \frac{1}{k}} = \frac{\frac{1}{i}}{\frac{1}{5} + \frac{1}{6}} = \frac{30}{11 i}. \quad (3.53)$$

Assim, a probabilidade de ter sido escolhida a quinta urna é $\frac{30}{55} = \frac{6}{11}$ e a probabilidade de ter sido escolhida a sexta urna é $\frac{30}{66} = \frac{5}{11}$. $\blacklozenge$

3.6 Fórmulas inversas da de Bayes

Do exposto anteriormente, o Teorema de Bayes permite deduzir as probabilidades *a posteriori* de cada uma das n causas ($\mathbb{P}_i$) através da probabilidade *a priori* de cada causa (ω_i) e da probabilidade de cada causa dar origem ao efeito (p_i). De seguida, como Proposição IX, Pacheco d'Amorim tenta determinar as fórmulas inversas, começando por determinar as probabilidades *a priori* de cada causa ω_i em função de $\mathbb{P}_i$ e de p_i .

Ora, na fórmula de Bayes, temos

$$\mathbb{P}_i = \frac{\omega_i p_i}{\sum_{j=1}^n \omega_j p_j}, \quad (3.54)$$

de onde podemos deduzir que

$$\begin{aligned} \frac{\mathbb{P}_i}{p_i} = \frac{\omega_i}{\sum_{j=1}^n \omega_j p_j} &\Leftrightarrow \sum_{i=1}^n \frac{\mathbb{P}_i}{p_i} = \frac{\sum_{i=1}^n \omega_i}{\sum_{j=1}^n \omega_j p_j} = \frac{1}{\sum_{j=1}^n \omega_j p_j} \Leftrightarrow \\ &\Leftrightarrow \sum_{j=1}^n \omega_j p_j = \frac{1}{\sum_{i=1}^n \frac{\mathbb{P}_i}{p_i}} \end{aligned} \quad (3.55)$$

e, por isso,

$$\frac{\mathbb{P}_i}{p_i} = \frac{\omega_i}{\sum_{j=1}^n \omega_j p_j} \Leftrightarrow \omega_i = \frac{\mathbb{P}_i}{p_i} \times \sum_{j=1}^n \omega_j p_j = \frac{\frac{\mathbb{P}_i}{p_i}}{\sum_{i=1}^n \frac{\mathbb{P}_i}{p_i}}. \quad (3.56)$$

Esta fórmula corresponde, na notação actual, a

$$\mathbb{P}(\mathbf{A}_i) = \frac{\frac{\mathbb{P}(\mathbf{A}_i|\mathbf{B})}{\mathbb{P}(\mathbf{B}|\mathbf{A}_i)}}{\sum_j \frac{\mathbb{P}(\mathbf{A}_j|\mathbf{B})}{\mathbb{P}(\mathbf{B}|\mathbf{A}_j)}}, \quad (3.57)$$

que facilmente se verifica. Mas, se efectuarmos a mesma operação para p_i temos

$$\frac{\mathbb{P}_i}{\omega_i} = \frac{p_i}{\sum_{j=1}^n \omega_j p_j} \Leftrightarrow \sum_{i=1}^n \frac{\mathbb{P}_i}{\omega_i} = \frac{\sum_{i=1}^n p_i}{\sum_{j=1}^n \omega_j p_j} \quad (3.58)$$

que não simplifica da mesma forma que a dedução anterior, uma vez que $\sum_{i=1}^n \omega_i = 1$ (uma das urnas será necessariamente escolhida) mas $\sum_{i=1}^n p_i \neq 1$ (em princípio). Por esta razão o resultado de Pacheco d'Amorim não é correcto, pois a sua fórmula

$$p_i = \frac{\frac{\mathbb{P}_i}{\omega_i}}{\sum_{j=1}^n \frac{\mathbb{P}_j}{\omega_j}}. \quad (3.59)$$

corresponde, na notação usual mais detalhada, a

$$\mathbb{P}(\mathbf{B}|\mathbf{A}_i) = \frac{\frac{\mathbb{P}(\mathbf{A}_i|\mathbf{B})}{\mathbb{P}(\mathbf{A}_i)}}{\sum_j \frac{\mathbb{P}(\mathbf{A}_j|\mathbf{B})}{\mathbb{P}(\mathbf{A}_j)}} \quad (3.60)$$

que só se verifica se, por acaso, $\sum_{i=1}^n \mathbb{P}(\mathbf{B}|\mathbf{A}_i) = 1$, o que normalmente não acontece. Porventura este erro passou despercebido pela notação ambígua usada por Pacheco d'Amorim, não se tendo apercebido o autor de que p_i corresponde à probabilidade de cada urna dar origem a uma bola branca, pelo que a sua soma não é normalmente igual à unidade. Por exemplo, se considerarmos cem urnas onde cada uma tem metade das bolas brancas, teremos $p_i = 0.5$ para $i = 1, \dots, 100$ e, neste caso, resultará para a soma destas probabilidades $\sum_{i=1}^{100} p_i = 50 \neq 1$.

3.7 Regra da Sucessão de Laplace

Para finalizar este capítulo das CLASSES FINITAS o autor resolve o problema conhecido por regra da sucessão de Laplace (*rule of succession*⁽¹⁷⁾), começando por resolver problemas mais simples. Nestes problemas vamos recorrer à mesma notação simplificada que temos utilizado desde o problema das probabilidades das causas (Problema 3.1), isto é, usaremos ω_i para a probabilidade *a priori* de cada uma das urnas, p_i para a probabilidade de na urna i sair bola branca e $\mathbb{P}_i$ para a probabilidade *a posteriori* (de, quando sai bola branca, a urna escolhida ter sido a número i).

Problema 3.2. Tire-se, à sorte, uma urna e da urna tirada, uma bola que se verifica ser branca. Metida a bola na urna, pergunta-se: qual a probabilidade de que, feita outra tiragem da mesma urna, se obtenha uma bola branca?

Resolveremos este problema das duas formas que Pacheco d'Amorim apresentou na sua obra.

SOLUÇÃO: 1 — Começemos por determinar as possibilidades das classes intervenientes. A classe possível, que é um complexo de classes, será representada por $\mathbb{B}'$ de modo a simplificar a notação, podendo ser considerada como a totalidade dos elementos compostos de

$$\text{urna qualquer} \rightarrow \text{bola branca} \rightarrow \text{bola qualquer};$$

cuja possibilidade é dada por

$$\varpi_{\mathbb{B}'} = \sum_{j=1}^n \omega_j p_j \cdot 1 = \sum_{j=1}^n \omega_j p_j. \quad (3.61)$$

A classe favorável, que também é um complexo de classes, será representada por $\mathbb{B}''$, sendo constituída pela totalidade dos elementos compostos de

⁽¹⁷⁾ Este nome foi utilizado pela primeira vez por Venn (1866) correspondendo ao título do *chapter VII* desta obra.

urna qualquer $\rightarrow$ bola branca $\rightarrow$ bola branca;

logo a sua possibilidade é dada por

$$\varpi_{\mathbb{B}''} = \sum_{j=1}^n \omega_j p_j p_j = \sum_{j=1}^n \omega_j p_j^2 \quad (3.62)$$

e consequentemente concluímos que

$$\mathbb{P}_{\mathbb{B}'}(\mathbb{B}'') = \frac{\varpi_{\mathbb{B}''}}{\varpi_{\mathbb{B}'}} = \frac{\sum_{i=1}^n \omega_i p_i^2}{\sum_{j=1}^n \omega_j p_j}. \quad (3.63)$$

◇

SOLUÇÃO: 2 — Pacheco d'Amorim considera que este problema é equivalente a substituir, no problema das probabilidades das causas (Problema 3.1), que deu origem ao Teorema de Bayes, a probabilidade de cada urna ser escolhida (ω_i) por $\mathbb{P}_i$, sendo ω_i as probabilidades *a priori* e $\mathbb{P}_i$ as probabilidades *a posteriori*. Deste modo o autor considera que este problema é equivalente ao seguinte.

Problema 3.3. Se n causas de probabilidades $\mathbb{P}_1, \mathbb{P}_2, \dots, \mathbb{P}_n$, dão origem a um determinado efeito com probabilidades, respectivamente, $p_1, p_2, \dots, p_n$, qual a probabilidade desse efeito?

A probabilidade pretendida é (pela Proposição V - Teorema das Probabilidades Totais) dada por

$$\mathbb{P} = \sum_{i=1}^n \mathbb{P}_i p_i = \frac{\sum_{i=1}^n \omega_i p_i^2}{\sum_{j=1}^n \omega_j p_j}, \quad (3.64)$$

obtendo-se, deste modo, o mesmo resultado anteriormente determinado. ◇

Utilizando raciocínios análogos, o autor generaliza este resultado servindo-se do problema a seguir enunciado.

Problema 3.4. Se numa urna, tirada à sorte, fizermos m tiragens (metendo na urna cada bola tirada, antes de feita a tiragem imediata) e dessas tiragens nos resultarem m_1 bolas brancas e m_2 pretas ($m_1 + m_2 = m$), qual a probabilidade de sair na tiragem de ordem $m + 1$ uma bola branca?

SOLUÇÃO: A solução deste problema, que pode ser obtido pelos mesmos dois processos que o problema anterior (caso particular de duas tiragens), é dada por

$$\mathbb{P} = \frac{\sum_{i=1}^n \omega_i p_i^{m_1+1} q_i^{m_2}}{\sum_{j=1}^n \omega_j p_j^{m_1} q_j^{m_2}}. \quad (3.65)$$

◇

Tal resultado, referente ao Problema 3.4, corresponde à Proposição X da tese de Pacheco d'Amorim. Como corolário o autor apresenta uma fórmula simplificada para as situações em que as urnas têm igual probabilidade *a priori* de serem escolhidas, pois nesta situação teremos $\omega_i = \frac{1}{n}$ e facilmente obtemos

$$\mathbb{P} = \frac{\sum_{i=1}^n p_i^{m_1+1} q_i^{m_2}}{\sum_{j=1}^n p_j^{m_1} q_j^{m_2}}. \quad (3.66)$$

Problema 3.5. Dá-se uma urna contendo N bolas, brancas e pretas, de percentagens desconhecidas. Supondo que todas as percentagens são igualmente prováveis, qual a probabilidade de tirar uma bola branca na tiragem de ordem $m + 1$, sabendo-se que nas primeiras m tiragens se obtiveram m_1 bolas brancas e m_2 bolas pretas?

Ou, de forma equivalente:

Problema 3.6. Dão-se $N + 1$ urnas, contendo a primeira N bolas pretas, a segunda uma bola branca e $N - 1$ pretas, a terceira 2 bolas brancas e $N - 2$ pretas e assim sucessivamente até à última urna que contém N bolas

brancas⁽¹⁸⁾. Tira-se uma urna, à sorte, e fazem-se m tiragens (tornando a pôr na urna cada bola, antes de tirar a seguinte) que dão m_1 bolas brancas e m_2 pretas. Pergunta-se: qual a probabilidade de que a tiragem de ordem $m + 1$ dê uma bola branca?

SOLUÇÃO: Pacheco d'Amorim recorre ao corolário da sua Proposição X, fórmula (3.66), para resolver este problema, utilizando para a probabilidade de bola branca em cada urna

$$p_\alpha = \frac{N - \alpha}{N}, \quad \alpha = 0, 1, \dots, N, \quad (3.67)$$

logo teremos $q_\alpha = \frac{\alpha}{N}$, obtendo-se

$$\mathbb{P} = \frac{\sum_{\alpha=0}^N \left(\frac{N - \alpha}{N} \right)^{m_1+1} \left(\frac{\alpha}{N} \right)^{m_2}}{\sum_{\alpha=0}^N \left(\frac{N - \alpha}{N} \right)^{m_1} \left(\frac{\alpha}{N} \right)^{m_2}}, \quad (3.68)$$

que considera poder ser aproximado, para valores de N grandes, por

$$\mathbb{P} = \frac{\int_0^N \left(\frac{N - \alpha}{N} \right)^{m_1+1} \left(\frac{\alpha}{N} \right)^{m_2} d\alpha}{\int_0^N \left(\frac{N - \alpha}{N} \right)^{m_1} \left(\frac{\alpha}{N} \right)^{m_2} d\alpha} \quad (3.69)$$

que, aplicando a transformação $\alpha = Nx$, obtemos

$$\mathbb{P} = \frac{\int_0^1 (1 - x)^{m_1+1} x^{m_2} dx}{\int_0^1 (1 - x)^{m_1} x^{m_2} dx}. \quad (3.70)$$

Esta probabilidade corresponde ao quociente entre duas funções Beta (**B**), pois

$$\mathbf{B}(a, b) = \int_0^1 x^{a-1} (1 - x)^{b-1} dx = \frac{\Gamma(a) \cdot \Gamma(b)}{\Gamma(a + b)} \quad (3.71)$$

⁽¹⁸⁾ Desta forma obtém-se todas as possibilidades de composição de urnas com N bolas brancas e pretas.

onde a função Gamma (Γ) é definida por

$$\Gamma(\alpha) = \int_0^{\infty} x^{\alpha-1} e^{-x} dx \quad (3.72)$$

que, para valores α inteiros, verifica $\Gamma(\alpha) = (\alpha - 1)!$. Desta forma a probabilidade pretendida é dada por

$$\begin{aligned} \mathbb{P} &= \frac{\Gamma(m_1 + 2) \Gamma(m_2 + 1)}{\Gamma(m_1 + m_2 + 3)} \cdot \frac{\Gamma(m_1 + m_2 + 2)}{\Gamma(m_1 + 1) \Gamma(m_2 + 1)} = \\ &= \frac{(m_1 + 1)!}{(m_1 + m_2 + 2)!} \cdot \frac{(m_1 + m_2 + 1)!}{m_1!} = \\ &= \frac{m_1 + 1}{m_1 + m_2 + 2} = \frac{m_1 + 1}{m + 2}, \end{aligned} \quad (3.73)$$

onde esta fórmula é tanto mais aproximada quanto maior for o valor de N , isto é,

$$\frac{\sum_{\alpha=0}^N \left(\frac{N-\alpha}{N}\right)^{m_1+1} \left(\frac{\alpha}{N}\right)^{m_2}}{\sum_{\alpha=0}^N \left(\frac{N-\alpha}{N}\right)^{m_1} \left(\frac{\alpha}{N}\right)^{m_2}} \xrightarrow{N \rightarrow \infty} \frac{m_1 + 1}{m + 2}. \quad (3.74)$$

◇

Este resultado foi apresentado pela primeira vez por Laplace (1774), sendo analisado frequentemente nas obras da época, do século XIX e início do século XX, tais como em Poisson (1837, p. 124–126), Bertrand (1888, p. 167–168), Poincaré (1896, p. 163), Borel (1909, p. 170–171), Bachelier (1912, p. 488), entre outras. Contudo, o problema proposto por Laplace no seu artigo de 1774 não é exactamente igual ao de Pacheco d'Amorim, pois Laplace parte de uma urna com uma infinidade de bolas brancas e pretas com proporções desconhecidas, sendo naturalmente a sua resolução distinta da apresentada por Pacheco d'Amorim. Pela importância que este resultado tem nas discussões dos fundamentos da Teoria da Probabilidade, apresentamos de seguida o *Problem I* de Laplace (1774), que é considerado por alguns o primeiro artigo bayesiano.

Problema 3.7 (*Rule of succession*). De uma urna contendo uma infinidade⁽¹⁹⁾ de bolas brancas e pretas com percentagem desconhecida retiramos $n = p + q$ bolas, das quais p são brancas e q são pretas. Qual a probabilidade de a próxima bola a ser retirada seja branca?

SOLUÇÃO: Consideremos, por hipótese e pelo facto de desconhecermos o valor da proporção de bolas, que a proporção de bolas brancas é caracterizada por uma variável aleatória θ com distribuição uniforme no intervalo $[0, 1]$ e que as bolas retiradas são representadas por uma sucessão de variáveis aleatórias $X_1, X_2, \dots, X_n, \dots$, onde X_i representa a i -ésima bola retirada, que assume o valor 1 no caso de a bola ser branca e assume o valor 0 caso seja preta, isto é, as variáveis aleatórias X_i têm distribuição de Bernoulli com probabilidade de sucesso desconhecida θ . Suponhamos ainda que as variáveis aleatórias X_i são independentes condicionalmente a um valor de θ fixo. Neste contexto, podemos obter

$$\begin{aligned} \mathbb{P}(X_{n+1} = 1 \mid X_1 + \dots + X_n = p) &= \mathbb{E}(\theta \mid X_1 + \dots + X_n = p) = \\ &= \frac{p+1}{n+2}, \end{aligned} \quad (3.75)$$

pois

$$\begin{aligned} \mathbb{P}(X_1 = x_1, \dots, X_n = x_n \mid \theta) &= \prod_{i=1}^n \theta^{x_i} (1-\theta)^{1-x_i} = \\ &= \theta^{\sum_{i=1}^n x_i} (1-\theta)^{\sum_{i=1}^n (1-x_i)} = \theta^p (1-\theta)^{n-p} \end{aligned} \quad (3.76)$$

onde cada x_i , $i = 1, \dots, n$, assume o valor zero (insucesso) ou o valor unitário (sucesso) de forma que a sua soma seja igual a p (número de sucessos). Como

⁽¹⁹⁾ Se considerarmos um número finito de bolas onde as bolas são retiradas com reposição, o valor obtido será o mesmo; contudo, nessa situação, a proporção de bolas brancas apenas pode assumir valores da forma $\frac{i}{n}$ para $i = 1, \dots, n$ onde n é o número de bolas, enquanto com um número infinito de bolas a proporção pode assumir qualquer valor (racional) no intervalo $[0, 1]$.

existem $\binom{n}{p}$ formas distintas de obter p sucessos em n provas (permutações dos valores de x_i tal que a soma seja igual a p) teremos

$$\begin{aligned} f(X_1 + \cdots + X_n = p \mid \theta) &= \mathbb{P}(X_1 + \cdots + X_n = p \mid \theta) = \\ &= \binom{n}{p} \theta^p (1 - \theta)^{n-p}. \end{aligned} \quad (3.77)$$

Como, por hipótese, a variável aleatória θ tem distribuição uniforme, a sua função densidade h_θ é igual à unidade no intervalo $[0, 1]$ sendo nula nos restantes valores, isto é,

$$h_\theta(x) = \begin{cases} 1 & 0 \leq x \leq 1 \\ 0 & x < 0 \vee x > 1 \end{cases}, \quad (3.78)$$

obtemos⁽²⁰⁾

$$\begin{aligned} h(\theta \mid X_1 + \cdots + X_n = p) &= \frac{f(X_1 + \cdots + X_n = p \mid \theta) h_\theta(x)}{\int_0^1 f(X_1 + \cdots + X_n = p \mid \theta) h_\theta(x) d\theta} = \\ &= \frac{\binom{n}{p} \theta^p (1 - \theta)^{n-p}}{\int_0^1 \binom{n}{p} \theta^p (1 - \theta)^{n-p} d\theta} = \frac{\theta^p (1 - \theta)^{n-p}}{\mathbf{B}(p+1, n-p+1)}, \end{aligned} \quad (3.79)$$

que representa a função densidade da distribuição Beta(a,b) com parâmetros $a = p+1$ e $b = n-p+1$, sendo o seu valor esperado $\frac{a}{a+b}$, que, concretizando na situação analisada, alcançamos o resultado $\frac{p+1}{n+2}$ previamente apresentado.

◇

Uma das razões do seu aparecimento em diversas obras deve-se ao facto de este resultado ser controverso. Não é a autenticidade desde resultado, sob as hipótese enunciadas, que é questionada. O que é problemático são as duas hipóteses, quer a dependência entre as variáveis aleatórias X_i que

⁽²⁰⁾ Utilizando o Teorema de Bayes na sua forma contínua $h(\theta \mid x) = \frac{f(x|\theta) h(\theta)}{\int_{\Theta} f(x|\theta) h(\theta) d\theta}$.

só são consideradas independentes quando condicionadas a um valor fixo de θ , quer a utilização da distribuição uniforme para caracterizar o nosso desconhecimento da proporção de bolas brancas – θ .

A primeira hipótese assume enorme importância na Estatística bayesiana. Esta escola considera que não pode haver independência entre as variáveis aleatórias com probabilidade constante e desconhecida⁽²¹⁾, uma vez que, se fossem independentes, não se aprenderia nada com a experiência, pois, no caso de X_i serem independentes, teríamos

$$\mathbb{P}(X_{n+1} = 1 \mid X_1 + \cdots + X_n = p) = \mathbb{P}(X_{n+1} = 1),$$

não havendo qualquer processo de aprendizagem, isto é, não se concluindo nada sobre o valor da proporção de bolas brancas ao serem observadas tiragens de bolas da urna. Por conseguinte, nesta situação, com o facto de termos realizado extracções de bolas da urna, fosse qual fosse o resultado, nada aprenderíamos sobre a sua constituição. A visão bayesiana refuta desta forma a visão frequencista de Richard von Mises para quem as variáveis, na situação descrita no problema, seriam independentes. Assim, na visão frequencista, se retirarmos $p + q$ bolas, das quais p são brancas, a nossa visão sobre a probabilidade não é alterada e, consequentemente, é impossível efectuar inferência estatística nestes casos. Esta situação é fortemente criticada pelos defensores da interpretação bayesiana de probabilidade. Se, no início, não sabemos qual a probabilidade de cada cor e se retirarmos 1000 bolas e todas elas forem brancas, será que continuamos sem saber nada sobre a probabilidade de uma bola ser branca? Para os bayesianos, como a probabilidade de sucesso é desconhecida, o facto de saírem bolas brancas ou

⁽²¹⁾ Condições que foram utilizadas na resolução do Problema 3.7 (*Rule of succession*), pois considerámos que as variáveis X_i eram dependentes, mas independentes condicionalmente a um valor fixo para a variável aleatória θ .

pretas altera o nosso grau de credibilidade sobre a cor que vai sair a seguir. Se, em 1000 bolas retiradas todas forem brancas, teremos um forte grau de credibilidade de que a próxima bola a sair também será branca.

Deste modo, na escola bayesiana de Probabilidade o Teorema de Bayes desempenha um papel fundamental na actualização das probabilidades após a obtenção de novas informações, sendo essa a razão da sua denominação bayesiana⁽²²⁾. Nesta perspectiva, para que seja possível a existência de um processo de aprendizagem, as variáveis aleatórias X_i não podem ser consideradas independentes⁽²³⁾. Bruno de Finetti (1937), o principal criador da visão bayesiana (subjectiva), desenvolve um conceito de dependência que, por um lado, permite realizar inferência (haver aprendizagem com as experiências) e, por outro lado, deverá ser tão fraco quanto possível de forma a incluir os outros tipos de dependência, sendo denominado por permutabilidade (*exchangeability*). Uma sequência de variáveis aleatórias diz-se permutável se a sua distribuição conjunta for invariante para permutações das variáveis. Deste modo, um conjunto finito de variáveis (quantidades)⁽²⁴⁾ ale-

(22) Para uma análise mais detalhada, e actual, sobre a Estatística bayesiana, podem ser consultados Paulino (2003) ou Bernardo e Smith (1994).

(23) Isto não significa que os bayesianos não aceitem o conceito de independência das variáveis aleatórias, pois, se estivermos perante provas de Bernoulli com probabilidades conhecidas, estas provas são consideradas independentes mesmo pelos defensores desta teoria. No entanto, não aceitam que, no caso de a probabilidade de sucesso ser desconhecida, as provas sejam independentes, pois consideram que, à medida que vão efectuar provas e conhecendo os seus resultados, estes fornecem informação sobre a probabilidade desconhecida e, consequentemente, os nossos graus de credibilidade sobre o seu valor serão alterados. Assim, nesta visão, quando a probabilidade não é conhecida as provas não podem ser consideradas independentes.

(24) Bruno de Finetti não utiliza a nomenclatura variáveis aleatórias, preferindo utilizar quantidades aleatórias.

atórias $(X_1, X_2, \dots, X_n)$ diz-se permutável quando

$$f(x_1, x_2, \dots, x_n) = f(x_{\pi(1)}, x_{\pi(2)}, \dots, x_{\pi(n)}) \quad (3.80)$$

para qualquer permutação $\{\pi(1), \pi(2), \dots, \pi(n)\}$ do conjunto $\{1, 2, \dots, n\}$. Um conjunto infinito de variáveis (quantidades) aleatórias X_i , $i = 1, 2, \dots$ diz-se permutável se qualquer subsucessão finita for permutável. Notemos que as variáveis aleatórias independentes são permutáveis, mas as variáveis aleatórias permutáveis não são necessariamente independentes. Com base no conceito de permutabilidade Bruno de Finetti deduz o seu famoso Teorema de Representação que para uma sucessão infinita de quantidades aleatórias binárias permutáveis X_i , $i = 1, 2, \dots$ garante que qualquer que seja o inteiro n existe sempre uma função de distribuição F_θ tal que

$$f(x_1, x_2, \dots, x_n) = \int_0^1 \prod_{i=1}^n \theta^{x_i} (1 - \theta)^{1-x_i} dF_\theta(\theta), \quad (3.81)$$

onde

$$F_\theta(\theta) = \lim_{n \rightarrow \infty} \mathbb{P} \left(\frac{\sum_{i=1}^n X_i}{n} \leq \theta \right), \quad (3.82)$$

e $\theta = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n X_i$. Desta forma, a amostra $X_1, \dots, X_n$ comporta-se como se tratasse de um conjunto de variáveis aleatórias independentes condicionalmente a θ , pois

$$f(x_1, x_2, \dots, x_n | \theta) = \prod_{i=1}^n f(x_i | \theta) = \prod_{i=1}^n \theta^{x_i} (1 - \theta)^{1-x_i}. \quad (3.83)$$

Podemos encontrar este assunto desenvolvido em Bernardo e Smith (1994) e Bernardo (1996). Na secção 10.4 analisaremos com mais detalhe a visão de Bruno de Finetti onde salientaremos a importância do conceito de permutabilidade e dos Teoremas de Representação.

Mesmo aceitando as duas hipóteses consideradas na resolução do Problema 3.7, são ainda discutíveis os limites de aplicação da fórmula deduzida,

sendo extremamente polémica a utilização em algumas situações famosas, como ilustram as célebres discussões filosóficas sobre a probabilidade de o sol nascer amanhã. Esta situação é comparada com uma urna com composição desconhecida, onde cada dia é retirada uma bola, representando as bolas brancas o sol nascer nesse dia e as bolas pretas a situação de o sol não nascer. Após seis mil anos em que a urna foi diariamente consultada, observamos 2191500 vezes bola branca, então a probabilidade de o sol nascer amanhã será, pela forma deduzida⁽²⁵⁾

$$\mathbb{P} = \frac{2\ 191\ 501}{2\ 191\ 502} \approx 0.9999995437.$$

Esta análise é efectuada em diversas obras, uns defendendo a sua aplicação, outros criticando fortemente a sua utilização. Por exemplo, Cournot (1843) refere que não faz qualquer sentido apostar 2 para 1 que vai sair caras num segundo lançamento apenas pelo facto de no primeiro lançamento ter saído caras.

Pacheco d'Amorim, neste capítulo, limitou-se a deduzir o resultado, não entrando na discussão desta questão sobre os limites da aplicabilidade da fórmula deduzida, ou da Teoria da Probabilidade, a estas situações. Contudo, na resolução do Problema 3.5 utilizou as mesmas duas hipótese que Laplace. Em primeiro lugar considerou (tendo o cuidado de incluir esta hipótese no enunciado do problema evitando, desta forma, qualquer polémica), pelo seu desconhecimento em relação à proporção de bolas brancas na urna, que as diferentes proporções possíveis eram equiprováveis (utilizando a distribuição uniforme, neste caso, discreta entre os possíveis proporções de bolas brancas quando uma urna contém N bolas, aplicando o princípio que Laplace (1812,

⁽²⁵⁾ Os valores aqui apresentados foram retirados de Bertrand (1888, p. 169), mas por exemplo Laplace (1814, p. *xiii*) utiliza cinco mil anos, obtendo uma probabilidade de $\frac{1\ 826\ 213}{1\ 826\ 214} \approx 0,9999994524$ de o sol nascer no dia seguinte.

p. *ii*) denominava de *princípio da razão insuficiente*). O autor considerou também a dependência entre as variáveis (bolas retiradas), sendo esta dependência estabelecida pelo facto de Pacheco d'Amorim considerar o Problema 3.5 equivalente ao Problema 3.6 (página 208), pois no segundo problema a dependência é visível dado que a probabilidade de sair bola branca depende da urna que é previamente escolhida à sorte (havendo uma ligação entre as tiragens por estas depois serem todas efectuadas na mesma urna), enquanto no primeiro problema esta dependência não é clara por estarmos a retirar bolas, com reposição, de uma única urna (com probabilidade de sair bola branca desconhecida). Deste modo, como a base de resolução do problema é o enunciado do Problema 3.6, Pacheco d'Amorim considerou que as bolas retiradas só são independentes quando consideradas condicionalmente a um valor fixo da proporção de bolas brancas, isto é, quando consideramos uma urna fixa.

Contudo, é unicamente na CONCLUSÃO da sua Tese, que analisaremos mais adiante, que o autor debate os limites de aplicação desta Ciência.

Refira-se que a definição de Probabilidade [condicionada] de Pacheco d'Amorim é, como usual nas obras da sua época, ainda dependente das origens desta teoria nos jogos de azar. Desta forma, o autor limitou-se a conceptualizar a probabilidade para tiragens de elementos de classes finitas, e não conseguiu apresentar uma definição mais vasta, porventura mais abstracta, que englobasse outras situações. Por exemplo, poderia ter recorrido a uma ideia semelhante à que utilizou na probabilidade contínua, onde as regiões são sub-conjuntos de $\mathbb{R}^n$, usando as classes como subconjuntos de $\mathbb{N}^n$ ou $\mathbb{Z}^n$. Porém, convém salientar que qualquer definição de probabilidade presente em livros seus contemporâneos não insere a abstracção necessária para alcançar uma definição mais geral de probabilidade.

3.8 Comentário geral ao capítulo

A construção apresentada por Pacheco d'Amorim para fundamentar a probabilidade discreta é, de facto, bastante original e de qualidade inegável.

A sua formalização de três sistemas de tiragens de elementos à sorte permite-lhe conceber os complexos de classes com os quais justifica a inexistência de equiprobabilidade em todos os sistemas de tiragens aleatórias de elementos. Com este sistema define uma tiragem hierarquizada de elementos, isto é, uma tiragem sequencial de elementos de classes onde a definição da classe de onde é retirado o elemento de ordem n depende dos $n - 1$ elementos retirados previamente. Nestas tiragens a possibilidade condicionada aos elementos já retirados continua a ser uniforme, contudo a possibilidade conjunta já não o é. Desta forma o autor justifica a existência de elementos com possibilidades distintas sem tem que recorrer ao princípio da razão insuficiente de Bernoulli e Laplace⁽²⁶⁾.

Pacheco d'Amorim define probabilidade de uma classe $\mathbf{A}''$ em relação a outra $\mathbf{A}'$, com $\mathbf{A}'' \subset \mathbf{A}'$, pelo rácio das possibilidades das duas classes. Desta definição concluímos que, para o autor, qualquer probabilidade é uma probabilidade condicionada. Assim sendo, a probabilidade incondicional é interpretada como sendo um caso particular da condicionada, ao condicionar em relação à classe total possível (universo), não utilizando a construção habitual onde a probabilidade condicionada é determinada pelo rácio de duas probabilidade incondicionais. O conceito primitivo de que parte é a probabi-

⁽²⁶⁾ O princípio da razão insuficiente, enunciado por Bernoulli (1713), diz que se não existir fundamento para suspeitar que hajam resultados mais ou menos prováveis do que outros, ou seja, se não houver qualquer razão para considerar que um acontecimento é mais provável do que outro, então devemos considerar que todos os resultados têm a mesma probabilidade. Este princípio polémico criou diversos paradoxos na Teoria da Probabilidade, tais como os apresentados por Bertrand (1888).

lidade condicionada, antecipando, apesar de utilizar argumentos distintos, as visões de Keynes (1921), de Finetti (1937), Jeffreys (1939) ou Rényi (1955).

A sua definição de probabilidade condicionada $\mathbb{P}_{\mathbf{A}'}(\mathbf{A}'')$ é restrita à situação $\mathbf{A}'' \subset \mathbf{A}'$ e esta restrição faz com que a sua construção se torne bastante mais complicada, pois bastaria definir $\mathbb{P}_{\mathbf{A}'}(\mathbf{A}'') = \mathbb{P}_{\mathbf{A}'}(\mathbf{A}'' \cap \mathbf{A}')$ para as situações onde $\mathbf{A}'' \not\subset \mathbf{A}'$ para alcançar a usual definição de probabilidade condicionada. Provavelmente esta estranha restrição, que outros autores anteriores a Pacheco d'Amorim não utilizam na sua definição de probabilidade condicionada⁽²⁷⁾, está na origem da sua afirmação de apresentar a primeira demonstração rigorosa do Teorema de Bayes. Todavia, na nossa opinião, as demonstrações previamente apresentadas por outros autores não contêm qualquer erro. O autor apresenta ainda alguns erros evidentes na determinação das fórmulas inversas da de Bayes.

Apesar da ineficiência da sua definição de probabilidade condicional, do *estardalhaço* sobre a demonstração do Teorema de Bayes e dos erros ingênuos na determinação das fórmulas inversas da de Bayes, o autor apresenta uma abordagem filosófica à construção da probabilidade discreta extremamente bem concebida. Saliente-se que, nas obras suas contemporâneas por nós consultadas, apesar de todos os autores destacarem os problemas de fundamentação da probabilidade e os paradoxos daí resultantes, não encontramos nenhuma tentativa de *racionalização do Cálculo das Probabilidades* do mesmo género, o que é, por si só, demonstrativo da dificuldade existente na época em fundamentar a Teoria da Probabilidade sem o recurso a princípios polémicos como o princípio da razão insuficiente.

⁽²⁷⁾ Shafer e Vovk (2006) referem que Felix Hausdorff em 1901 utiliza $\mathbb{P}_{\mathbf{A}}(\mathbf{B})$ para representar a probabilidade relativa de $\mathbf{B}$ dado $\mathbf{A}$, explicando que $\mathbf{A}$ e $\mathbf{B}$ podem ser quaisquer acontecimentos independentemente da sua relação temporal ou lógica.

Capítulo 4

Probabilidade Contínua

No segundo capítulo da sua tese Pacheco d’Amorim formaliza o conceito de probabilidade em regiões, isto é, de probabilidade contínua, de uma forma análoga à que efectuou com a probabilidade em classes.

Neste contexto Pacheco d’Amorim principia a construção da fundamentação da probabilidade contínua com o seu modelo standard (conceito primitivo) correspondente ao lançamento, à sorte, de um ponto numa região limitada da qual temos todo o conhecimento e somos nós os agentes do lançamento. O autor define três sistemas de lançamentos à sorte de pontos, concebendo os complexos de regiões para definir lançamentos hierarquizados onde o lançamento do n -ésimo ponto é efectuado numa região que depende dos pontos que saíram nos primeiros $n - 1$ lançamentos. Neste sistema de lançamentos os pontos pertencentes à região total possível (espaço amostra) não possuem todos a mesma possibilidade. Como tal, neste tipo de lançamento criado por Pacheco d’Amorim não é utilizada a usual definição geométrica de probabilidade, onde a probabilidade de uma região é proporcional à sua medida, consequência da existência de pontos com possibilidades distintas.

Tal como tinha feito na probabilidade discreta, Pacheco d’Amorim só

define probabilidade de uma região em relação a outra que contenha a primeira, ou seja, o seu conceito primitivo é a probabilidade condicionada. Para ilustrar a utilização da sua definição de probabilidade, Pacheco d'Amorim resolve alguns problemas, tais como o do triângulo e o de Borel-Kolmogoroff. Em relação a este último problema, só definitivamente resolvido com a axiomática proposta por Kolmogoroff (1933), a sua resolução não é satisfatória, pois recorre a algumas simplificações baseadas na simetria do problema, mas que não são válidas em todas as situações, nomeadamente no que se refere a aplicações em conjuntos não numeráveis como retrata este problema.

Por fim o autor propõe uma generalização da definição de possibilidade (probabilidade) para as situações em que estamos a lidar com regiões não limitadas, que será utilizada na sua elegante resolução do problema de Buffon no capítulo que dedicou a problemas geométricos.

Destacamos que a construção da probabilidade contínua efectuada por Pacheco d'Amorim não termina neste capítulo, pois algumas das ideias mais originais e mais bem concebidas do autor são apresentadas apenas no capítulo PONTO IMAGEM, que será analisado posteriormente.

4.1 Pontos e regiões possíveis

Tal como nas CLASSES FINITAS, Pacheco d'Amorim começa por definir o domínio de aplicação dos lançamentos, identificando os pontos e as regiões que considera possíveis em relação a cada sistema de lançamentos à sorte. Primeiramente define a união de regiões [disjuntas] com a mesma dimensão e a região composta $\mathbf{X} \times \mathbf{Y}$ (produto cartesiano entre duas regiões), demonstrando que a medida desta região composta é igual à multiplicação das medidas das regiões $\mathbf{X}$ e $\mathbf{Y}$ que a compõem, utilizando, como referência, a

Pangeometria⁽¹⁾. Deste modo, para Pacheco d'Amorim, lançar, à sorte, um ponto na região $\mathbf{X} \subset \mathbb{R}^n$ e outro na região $\mathbf{Y} \subset \mathbb{R}^m$ é o mesmo que lançar à sorte um ponto na região $\mathbf{X} \times \mathbf{Y} \subset \mathbb{R}^{n+m}$ considerando que, nestas situações, as regiões são independentes. Notemos que esta definição não nos restringe a rectângulos ou paralelepípedos, pois se a região $\mathbf{X} \subset \mathbb{R}^2$ for uma circunferência e a região $\mathbf{Y} \subset \mathbb{R}$ for um segmento de recta tudo decorrerá como se de um único lançamento se tratasse na superfície cilíndrica $\mathbf{X} \times \mathbf{Y} \subset \mathbb{R}^3$.

De forma semelhante à construção de complexos de classes nas CLASSES FINITAS, o autor considera que situações há onde as regiões $\mathbf{X}$ e $\mathbf{Y}$ são dependentes, ou seja, onde a região $\mathbf{Y}$ na qual é efectuado o segundo lançamento depende do ponto x lançado previamente em $\mathbf{X}$. Deste modo, Pacheco d'Amorim define um complexo de regiões $\mathbf{X} \odot \mathbf{Y} \subset \mathbb{R}^{m+n}$, com $\mathbf{X} \subset \mathbb{R}^n$ e $\mathbf{Y} \subset \mathbb{R}^m$, onde a região $\mathbf{Y}$ é dependente do ponto $x \in \mathbf{X}$, pelo que, para cada ponto $x \in \mathbf{X}$, teremos uma região $\mathbf{Y}|_{\mathbf{x}=x}$ associada que é, em geral, distinta das outras regiões associadas aos outros pontos de $\mathbf{X}$ (caso contrário seria uma região composta e não se tornaria necessário recorrer a este novo conceito). Este conceito de complexo de regiões pode ser formalizado através

⁽¹⁾ Pangeometria é o nome do último livro, publicado em 1855, do russo Nicolai Lobachewski (1792–1856). Nesta obra Lobachewski apresenta a sua geometria não euclidiana que inicialmente, quando a apresentou em 1829, apelidou de imaginária por contrariar o senso comum, e que é actualmente conhecida por Geometria Hiperbólica. Lobachewski criou esta geometria na tentativa de provar o quinto Postulado de Euclides que é equivalente a considerar que por cada ponto externo a uma determinada recta passa uma única recta paralela. Assim, ao tentar provar este postulado de Euclides, Lobachewski admitiu que este seria impossível, surgindo assim uma nova geometria ao considerar que por um ponto exterior a uma recta passam duas (e consequentemente um número infinito) rectas paralelas à dada. O húngaro János Bolyai (1802-1860), trabalhando independentemente, constrói uma geometria semelhante. Riemann (1826-1866) em 1854 apresenta outra geometria não Euclidiana, considerando que por um ponto exterior a uma recta não passa nenhuma recta paralela à dada, criando assim a Geometria Esférica.

de

$$\mathbf{X} \odot \mathbf{Y} = \bigcup_{x \in \mathbf{X}} x \times \mathbf{Y}|_{\mathbf{X}=x}, \quad (4.1)$$

onde $\mathbf{X} \subset \mathbb{R}^n$, $\mathbf{Y} \subset \mathbb{R}^m$ (isto significa que $\mathbf{Y}|_{\mathbf{X}=x} \subset \mathbb{R}^m, \forall x \in \mathbf{X}$) e, consequentemente, $\mathbf{X} \odot \mathbf{Y} \subset \mathbb{R}^{n+m}$.

Desta forma Pacheco d'Amorim considera três sistemas de lançamentos.

1. Lançar, à sorte, um ponto numa região limitada $\mathbf{X} \subset \mathbb{R}^n$, onde todos os pontos dessa região são possíveis;
2. Lançar, à sorte, um ponto na região $\mathbf{X} \subset \mathbb{R}^n$ e outro na região $\mathbf{Y} \subset \mathbb{R}^m$ (independente da região $\mathbf{X}$), onde todos os pontos da região $\mathbf{X} \times \mathbf{Y} \subset \mathbb{R}^{n+m}$ (região composta) são possíveis. Assim sendo, estes lançamentos são equivalentes a lançar um único ponto na região $\mathbf{X} \times \mathbf{Y}$ e este sistema de lançamentos pode ser reduzido à primeira situação que retrata um único lançamento numa região;
3. Lançar, à sorte, um ponto na região $\mathbf{X} \subset \mathbb{R}^n$ e outro na região $\mathbf{Y}|_{\mathbf{X}=x} \subset \mathbb{R}^m$, onde as regiões $\mathbf{X}$ e $\mathbf{Y}|_{\mathbf{X}=x}$ estão “*sujeitas a uma certa dependência*”, isto é, o lançamento é efectuado num complexo de regiões $\mathbf{X} \odot \mathbf{Y} \subset \mathbb{R}^{n+m}$ onde são possíveis todos os pontos que resultam de associar a cada ponto possível do primeiro lançamento (pontos da região $\mathbf{X}$) todos os pontos da região $\mathbf{Y}|_{\mathbf{X}=x}$ que lhe corresponde. Pacheco d'Amorim considera que, se o ponto lançado à sorte em $\mathbf{X}$ coincidir com x , o segundo lançamento far-se-á na região $\mathbf{Y}|_{\mathbf{X}=x}$ e tudo se passará como se apenas fosse efectuado um único lançamento na região $\mathbf{X} \times \mathbf{Y}|_{\mathbf{X}=x}$. Esta afirmação de Pacheco d'Amorim causa alguma estranheza, pois é notória a impossibilidade de reduzir *a priori* um lançamento num complexo de regiões num outro que seja efectuado numa região composta. Contudo, pensamos que o autor apresenta esta

definição apenas com o objectivo de posteriormente servir de base à sua definição de possibilidade, sendo a possibilidade de cada elemento (x, y) , do complexo de regiões $\mathbf{X} \odot \mathbf{Y}$, determinada sob esta hipótese. Desta forma, como veremos mais adiante, a possibilidade de um elemento $(x, y) \in \mathbf{X} \odot \mathbf{Y}$ é determinada considerando que o resultado do primeiro lançamento é já conhecido (condicionada a $\mathbf{X} = x$).

Uma região é possível em relação a determinado sistema de lançamentos se todos os seus pontos forem pontos possíveis. A região total possível (espaço amostra) é a região que contém todos os pontos possíveis em relação a um determinado lançamento. Anotemos que, com estas definições, num complexo de regiões a classe total possível será dada por

$$\mathbf{X} \odot \mathbf{Y} = \bigcup_{x \in \mathbf{X}} x \times \mathbf{Y}|_{\mathbf{X}=x}. \quad (4.2)$$

Exemplo 4.1 (Lançamentos em regiões compostas). Considere-se que é efectuado um lançamento no intervalo $\mathbf{X} = [x_{\min}, x_{\max}]$ e, posteriormente, um outro lançamento na região $\mathbf{Y} = [y_{\min}, y_{\max}]$. Os pontos possíveis de serem obtidos em relação a este lançamento serão todos os pontos contidos no rectângulo com vértices nos pontos $(x_{\min}, y_{\min})$, $(x_{\min}, y_{\max})$, $(x_{\max}, y_{\min})$ e $(x_{\max}, y_{\max})$. Assim, lançar à sorte um ponto em $[x_{\min}, x_{\max}]$ e outro em $[y_{\min}, y_{\max}]$ é o mesmo que lançar, à sorte, um único ponto na região composta $\mathbf{X} \times \mathbf{Y} = [x_{\min}, x_{\max}] \times [y_{\min}, y_{\max}]$.

Por exemplo, se $\mathbf{X} = [1, 5]$ e $\mathbf{Y} = [2, 5]$ a região $\mathbf{X} \times \mathbf{Y}$ será a região (rectângulo) a sombreado na Figura 4.1. ♦

Exemplo 4.2 (Lançamentos em complexos de regiões). Considere-se que é efectuado um lançamento no intervalo $\mathbf{X} = [x_{\min}, y_{\min}]$ e, posteriormente, um outro lançamento na região $\mathbf{Y}|_{\mathbf{X}=x} = [f_{\min}(x), f_{\max}(x)]$, região esta que depende do ponto x escolhido no primeiro lançamento em $\mathbf{X}$.

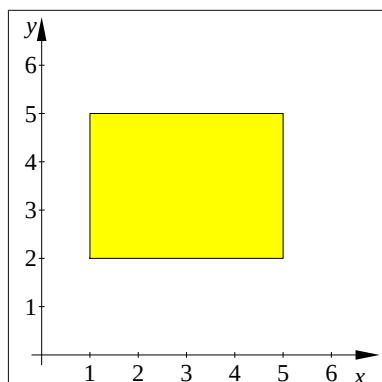


Figura 4.1: Lançar, à sorte, um ponto numa região composta

A região possível será a região em $\mathbb{R}^2$ limitada na abcissa por $x = x_{\min}$ e $x = x_{\max}$ e nas ordenadas pelas curvas $y = f_{\min}(x)$ e $y = f_{\max}(x)$.

Assim, se o ponto lançado à sorte em $\mathbf{X}$ coincidir com x_0 , o segundo lançamento far-se-á no segmento entre os pontos de coordenadas $(x_0, f_{\min}(x_0))$ e $(x_0, f_{\max}(x_0))$ e tudo se passará, segundo a definição, como se fosse efectuado um só lançamento num paralelograma $[x_{\min}, x_{\max}] \times [f_{\min}(x_0), f_{\max}(x_0)]$, isto no que se refere ao lançamento no complexo de regiões $\mathbf{X} \odot \mathbf{Y}$ quando no primeiro lançamento, em $\mathbf{X}$, é escolhido o ponto x_0 .

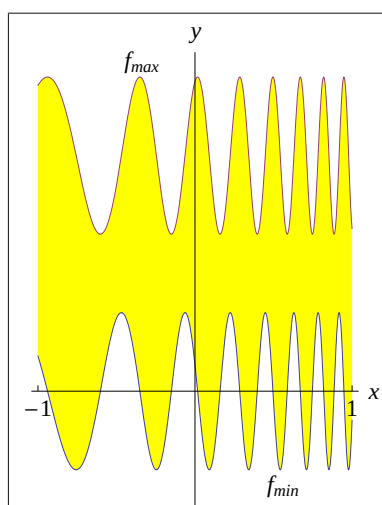


Figura 4.2: Lançar, à sorte, um ponto num complexo de regiões — 1

Suponhamos que é primeiro lançado um ponto em $\mathbf{X} = [-1, 1]$ e depois é lançado outro ponto em $\mathbf{Y}|\mathbf{X}=x = [f_{\min}(x), f_{\max}(x)]$, sendo a região total possível destes dois lançamentos representada na Figura 4.2.

Para Pacheco d'Amorim se, ao efectuar os lançamentos no complexo de regiões $\mathbf{X} \odot \mathbf{Y}$, o primeiro lançamento, na região $\mathbf{X}$, resultar o ponto $\mathbf{X} = -0.5$ estes lançamentos são equivalentes a lançar um só ponto na região $[-1, 1] \times [f_{\min}(-0.5), f_{\max}(-0.5)]$, rectângulo representado a sombreado na Figura 4.3.

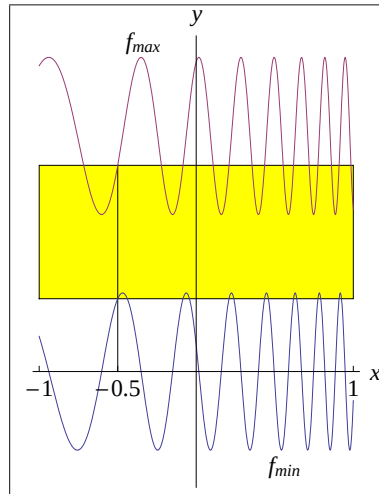


Figura 4.3: Lançar, à sorte, um ponto num complexo de regiões — 2

Salientemos, desde já, que é a área deste rectângulo que irá determinar a possibilidade dos pontos (x, y) em que $x = -0.5$, contendo este rectângulo pontos que não são admissíveis no sistema de lançamentos, isto é, pontos que não pertencem à região total possível dos dois lançamentos. Por outro lado, notemos também que o rectângulo não contém alguns pontos admissíveis no sistema de lançamentos, isto é, pontos que pertencem à região total possível. Conforme podemos observar no gráfico da Figura 4.2, se o primeiro lançamento originar outro ponto para x que não o -0.5 retratado no gráfico da Figura 4.3, o rectângulo obtido será distinto, pois o comprimento será sempre

o mesmo, correspondente à amplitude do intervalo $[-1, 1]$, mas a sua altura variará consoante o ponto considerado em $\mathbf{X}$. $\blacklozenge$

Deverá referir-se que não é a forma da região total possível que distingue se estamos a trabalhar numa região (composta ou não) ou num complexo de regiões, pois com as definições fornecidas por Pacheco d'Amorim poder-se-ia efectuar um único lançamento na região sombreada no gráfico da Figura 4.2, que utilizamos para retratar os lançamentos num complexo de regiões. Assim, podemos efectuar um único lançamento em qualquer região (limitada) de $\mathbb{R}^n$ independentemente da sua forma. O que diferencia um lançamento numa região de lançamentos em complexos de regiões é o sistema de lançamentos utilizado. Como vamos analisar de seguida, nos casos em que é efectuado um só lançamento (ou se pode reduzir a tal) os pontos são todos equipossíveis, nos casos de lançamentos num complexo de regiões não é possível reduzir a um só lançamento pelo facto de a região onde é efectuado o segundo lançamento só ser definida pelo resultado do primeiro lançamento e, conseqüentemente, os pontos não terem geralmente a mesma possibilidade.

4.2 Possibilidade

Pacheco d'Amorim define a possibilidade de cada ponto x lançado à sorte em $\mathbf{X}$, ou possibilidade por unidade, por

$$\pi_x = \frac{1}{\mu_{\mathbf{X}}}, \quad (4.3)$$

onde $\mu_{\mathbf{X}}$ representa a medida da região $\mathbf{X}$, limitada e de qualquer dimensão, em que se faz o lançamento à sorte, isto é,

$$\mu_{\mathbf{X}} = \int_{\mathbf{X}} dx. \quad (4.4)$$

Com esta definição a possibilidade de um ponto composto é igual ao produto das possibilidades dos pontos componentes (Proposição I), ou seja, se (x, y) é um ponto da região $\mathbf{X} \times \mathbf{Y}$, então

$$\pi_{(x,y)} = \pi_x \cdot \pi_y, \quad (4.5)$$

com $x \in \mathbf{X}$ e $y \in \mathbf{Y}$.

Demonstração. Visto que, como foi anteriormente referido, a medida de uma região composta é igual ao produto das medidas das regiões que a compõem ($\mu_{\mathbf{X} \times \mathbf{Y}} = \mu_{\mathbf{X}} \cdot \mu_{\mathbf{Y}}$), então

$$\pi_{(x,y)} = \frac{1}{\mu_{\mathbf{X} \times \mathbf{Y}}} = \frac{1}{\mu_{\mathbf{X}}} \cdot \frac{1}{\mu_{\mathbf{Y}}} = \pi_x \cdot \pi_y.$$

□

Numa região, composta ou não, a possibilidade é igual para todos os pontos; por isso podemos afirmar que, nestes casos, existe equipossibilidade (tal como acontecia anteriormente nas classes). Analisemos agora o que acontece nos complexos de regiões. Começemos por notar que este resultado também é válido para complexos de regiões se condicionarmos a um ponto fixo no primeiro lançamento. De facto Pacheco d'Amorim considera, como foi previamente referido, que lançar um ponto à sorte em $\mathbf{X}$ onde ocorre x e depois lançar outro ponto à sorte em $\mathbf{Y}|_{\mathbf{X}=x}$ é, por definição, o mesmo que lançar um único ponto na região composta $\mathbf{X} \times \mathbf{Y}|_{\mathbf{X}=x}$. Assim, sob estas considerações, podemos generalizar a fórmula (4.5) para lançamentos em complexos de regiões. Considerando então o complexo de regiões $\mathbf{X} \odot \mathbf{Y}$ teremos, para um determinado ponto (x, y) desse complexo de regiões,

$$\pi_{(x,y)} = \frac{1}{\mu_{\mathbf{X} \times \mathbf{Y}|_{\mathbf{X}=x}}} = \frac{1}{\mu_{\mathbf{X}}} \cdot \frac{1}{\mu_{\mathbf{Y}|_{\mathbf{X}=x}}} = \pi_x \cdot \pi_{y|\mathbf{X}=x}. \quad (4.6)$$

Assim, num complexo de regiões os pontos não são (geralmente) equipossíveis, pois a possibilidade de cada ponto composto depende do resultado

do primeiro lançamento, isto é, do valor de x lançado em $\mathbf{X}$, pois o valor de $\pi_{y|\mathbf{X}=x}$ presente na fórmula (4.6) é uma função de x e consequentemente $\pi_{(x,y)}$ também o é ($\pi_{(x,y)} = \pi(x)$).

Analisemos agora a possibilidade de uma região. Seja $\mathbf{X}$ a região total possível do sistema de lançamentos com que estamos a trabalhar. A possibilidade de uma região possível $\mathbf{X}' \subset \mathbf{X}$ é definida pelo integral da possibilidade por unidade estendido a essa região, caso esse integral exista, isto é, se representarmos por $\varpi_{\mathbf{X}'}$ a possibilidade da região $\mathbf{X}'$, teremos

$$\varpi_{\mathbf{X}'} = \int_{\mathbf{X}'} \pi_x dx. \quad (4.7)$$

Exemplo 4.3 (Possibilidade de uma região 1). Seja efectuado um lançamento, à sorte, na região $\mathbf{X} = [a, b]$. A possibilidade por unidade é

$$\pi_x = \frac{1}{\int_{\mathbf{X}} dx} = \frac{1}{b-a}, \quad (4.8)$$

obtendo-se, desta forma, a função densidade de uma distribuição uniforme. A possibilidade do intervalo $\mathbf{X}' = [c, d] \subset [a, b]$ será dada por

$$\varpi_{\mathbf{X}'} = \int_{\mathbf{X}'} \pi_x dx = \int_c^d \frac{1}{b-a} dx = \frac{d-c}{b-a}, \quad (4.9)$$

onde a probabilidade é proporcional à amplitude do intervalo. $\blacklozenge$

Como foi previamente referido, os lançamentos definidos no primeiro sistema não têm que ser efectuados unicamente em segmentos de recta, pois estes podem ser efectuados em qualquer região limitada de $\mathbb{R}^n$.

Exemplo 4.4 (Possibilidade de uma região 2). Lancemos, à sorte, um ponto num círculo centrado na origem e de raio 1 (região $\mathbf{C}$ representada a sombreado na Figura 4.4). A possibilidade por unidade de qualquer ponto do círculo é

$$\pi_{(x,y)} = \frac{1}{\int_{\mathbf{C}} d(x,y)} = \frac{1}{\pi}. \quad (4.10)$$

Qual é a possibilidade de o ponto lançado se situar no círculo centrado na origem e de raio $\frac{1}{2}$ (região $\mathbf{C}'$)? A possibilidade é dada por

$$\varpi_{\mathbf{C}'} = \int_{\mathbf{C}'} \pi_{(x,y)} d(x,y) = \int_{\mathbf{C}'} \frac{1}{\pi} d(x,y) = \frac{1}{4}. \quad (4.11)$$

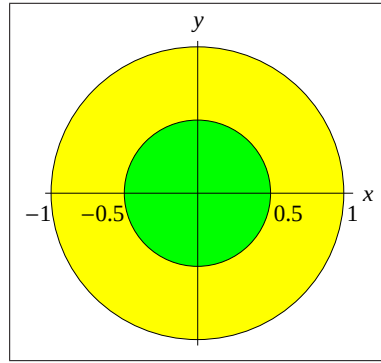


Figura 4.4: Lançar, à sorte, um ponto num círculo

◆

Exemplo 4.5 (Possibilidade de uma região composta). Consideremos que lançamos, à sorte, um ponto num círculo (região $\mathbf{C}$ do exemplo anterior) e depois lançamos outro ponto num segmento de recta de comprimento igual à dois (região $\mathbf{R}$ que é independente da região $\mathbf{C}$, isto é, o segundo lançamento é efectuado numa região fixa, que não depende do primeiro ponto lançado em $\mathbf{C}$).

Por definição, estes lançamentos são equivalentes a lançar um só ponto num cilindro (região $\mathbf{C} \times \mathbf{R}$) com raio um e comprimento dois que tem volume igual a 2π . Desta forma a possibilidade de qualquer ponto do cilindro será dada por

$$\pi_{(x,y,z)} = \frac{1}{\mu_{\mathbf{C} \times \mathbf{R}}} = \frac{1}{2\pi}. \quad (4.12)$$

◆

Os exemplos anteriores ilustram bem que, quando se efectua um único lançamento de um ponto numa região $\mathbf{A} \subset \mathbb{R}^n$, ou k lançamentos em regiões

independentes $\mathbf{A}_i \subset \mathbb{R}^{n_i}$, com $i = 1, \dots, k^{(2)}$, os pontos são equipossíveis, isto é, todos os pontos têm a mesma possibilidade, sendo a possibilidade de cada região $\mathbf{R}$ proporcional à sua medida $\mu_{\mathbf{R}}$.

Exemplo 4.6 (Possibilidade de um complexo de regiões). Seja lançado um ponto à sorte em $\mathbf{X} = [1, 3]$ e um outro lançado em $\mathbf{Y}|_{\mathbf{X}=x} = [f_{\min}(x), f_{\max}(x)]$ onde

$$f_{\min}(x) = \begin{cases} 3 - x & 1 \leq x < 2 \\ x - 1 & 2 \leq x \leq 3 \end{cases} \quad (4.13)$$

e

$$f_{\max}(x) = \begin{cases} x + 1 & 1 \leq x < 2 \\ 5 - x & 2 \leq x \leq 3 \end{cases}, \quad (4.14)$$

conforme região a sombreado na Figura 4.5.

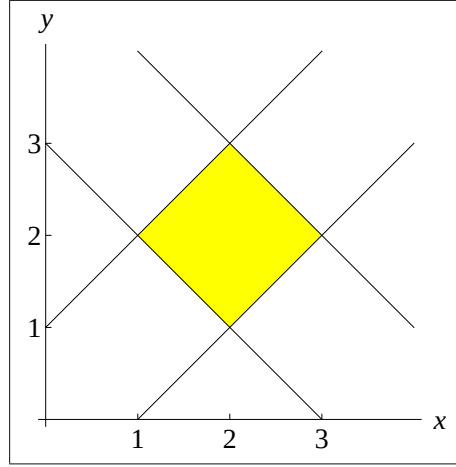


Figura 4.5: Lançar, à sorte, um ponto num complexo de regiões

Em relação ao ponto x , lançado à sorte em $\mathbf{X}$, a sua possibilidade é dada por

$$\pi_x = \frac{1}{\int_{\mathbf{x}} dx} = \frac{1}{\int_1^3 dx} = \frac{1}{2}. \quad (4.15)$$

⁽²⁾ Por definição, efectuar, à sorte, estes k lançamentos é equivalente a lançar, à sorte, um único ponto na região $\mathbf{A} = \mathbf{A}_1 \times \mathbf{A}_2 \times \dots \times \mathbf{A}_k \subset \mathbb{R}^n$, com $n = \sum_{i=1}^k n_i$.

Em relação aos pontos (x, y) do complexo de regiões $\mathbf{X} \odot \rightarrow \mathbf{Y}$ teremos, recorrendo à fórmula (4.6), para as situações onde o lançamento de x se situa em $1 \leq x < 2$

$$\pi_{(x,y)} = \pi_x \cdot \pi_{y_{\mathbf{X}=x}} = \frac{1}{2} \cdot \frac{1}{\int_{\mathbf{Y}|\mathbf{X}=x} dy} = \frac{1}{2} \cdot \frac{1}{\int_{3-x}^{x+1} dy} = \frac{1}{2(2x-2)} = \frac{1}{4x-4},$$

e, de forma idêntica, para o caso de o primeiro ponto lançado se situar em $2 \leq x \leq 3$, teremos

$$\pi_{(x,y)} = \pi_x \cdot \pi_{y_{\mathbf{X}=x}} = \frac{1}{2} \cdot \frac{1}{\int_{\mathbf{Y}|\mathbf{X}=x} dy} = \frac{1}{2} \cdot \frac{1}{\int_{x-1}^{5-x} dy} = \frac{1}{2(6-2x)} = \frac{1}{12-4x}.$$

Assim, a possibilidade de qualquer ponto $(x, y) \in \mathbf{X} \odot \rightarrow \mathbf{Y}$ é dada pela função

$$\pi_{(x,y)} = \begin{cases} \frac{1}{4x-4} & 1 \leq x < 2 \\ \frac{1}{12-4x} & 2 \leq x \leq 3 \end{cases} \quad (4.16)$$

que, como foi já referido, é uma função de x , não sendo nesta situação os pontos (x, y) equipossíveis.

Se, por exemplo, pretendermos determinar a possibilidade da região composta $\mathbf{A} = \left[\frac{3}{2}, \frac{5}{2}\right] \times \left[\frac{3}{2}, \frac{5}{2}\right] \subset \mathbf{X} \odot \rightarrow \mathbf{Y}$ teremos

$$\begin{aligned} \varpi_{\mathbf{A}} &= \int_{\mathbf{A}} \pi_{(x,y)} d(x, y) = \int_{\frac{3}{2}}^2 \int_{\frac{3}{2}}^{\frac{5}{2}} \frac{1}{4x-4} dy dx + \int_2^{\frac{5}{2}} \int_{\frac{3}{2}}^{\frac{5}{2}} \frac{1}{12-4x} dy dx = \\ &= \int_{\frac{3}{2}}^2 \frac{1}{4x-4} dx + \int_2^{\frac{5}{2}} \frac{1}{12-4x} dx = \frac{1}{2} \ln(2). \end{aligned}$$

Consideremos o complexo de regiões $\mathbf{X}' \odot \rightarrow \mathbf{Y}' \subset \mathbf{X} \odot \rightarrow \mathbf{Y}$ onde $\mathbf{X}' = \left[1, \frac{3}{2}\right]$ e $\mathbf{Y}'|_{\mathbf{X}=x} = \mathbf{Y}|_{\mathbf{X}=x} = [3-x, x+1]$. A possibilidade deste complexo de

regiões é dada por

$$\begin{aligned}\varpi_{\mathbf{X}' \odot \rightarrow \mathbf{Y}'} &= \int_{\mathbf{X}' \odot \rightarrow \mathbf{Y}'} \pi_{(x,y)} \, d(x,y) = \int_1^{\frac{3}{2}} \int_{3-x}^{x+1} \frac{1}{4x-4} \, dy \, dx = \\ &= \int_1^{\frac{3}{2}} \frac{2x-2}{4x-4} \, dx = \int_1^{\frac{3}{2}} \frac{1}{2} \, dx = \frac{1}{4}.\end{aligned}$$

Analiseemos agora o complexo de regiões $\mathbf{X}'' \odot \rightarrow \mathbf{Y}'' \subset \mathbf{X} \odot \rightarrow \mathbf{Y}$ onde $\mathbf{X}'' = [\frac{3}{2}, \frac{5}{2}]$ e $\mathbf{Y}''|_{\mathbf{X}=x} = [\frac{5}{2}, f_{\max}(x)]$. A possibilidade deste complexo de regiões é dada por

$$\begin{aligned}\varpi_{\mathbf{X}'' \odot \rightarrow \mathbf{Y}''} &= \int_{\mathbf{X}'' \odot \rightarrow \mathbf{Y}''} \pi_{(x,y)} \, d(x,y) = \\ &= \int_{\frac{3}{2}}^2 \int_{\frac{5}{2}}^{x+1} \frac{1}{4x-4} \, dy \, dx + \int_2^{\frac{5}{2}} \int_{\frac{5}{2}}^{5-x} \frac{1}{12-4x} \, dy \, dx = \\ &= \int_{\frac{3}{2}}^2 \frac{x-\frac{3}{2}}{4x-4} \, dx + \int_2^{\frac{5}{2}} \frac{\frac{5}{2}-x}{12-4x} \, dx = \frac{1-\ln(2)}{4}.\end{aligned}$$

Notemos que os complexos de regiões $\mathbf{X}' \odot \rightarrow \mathbf{Y}'$ e $\mathbf{X}'' \odot \rightarrow \mathbf{Y}''$ têm a mesma medida (área), mas não têm a mesma possibilidade. Isto significa que, nos complexos de regiões, a possibilidade não é proporcional à medida, como nas definições habituais de probabilidade contínua, consequência de os pontos dos complexos de regiões não serem equipossíveis.

Contudo, esclareçamos uma vez mais que poderíamos aplicar outro sistema de lançamentos à mesma região analisada ao longo deste exemplo e, em vez de lançarmos um ponto em $\mathbf{X}$ e outro em $\mathbf{Y}|_{\mathbf{X}=x}$, efectuarmos um único lançamento, à sorte, de um ponto no losango (Região $\mathbf{L}$) representado na Figura 4.5. Neste caso todos os pontos do losango teriam a mesma

possibilidade, pois

$$\pi_{(x,y)} = \frac{1}{\int_{\mathbf{L}} d(x,y)} = \frac{1}{\int_1^2 \int_{3-x}^{x+1} dy \, dx + \int_2^3 \int_{x-1}^{5-x} dy \, dx} = \frac{1}{2}. \quad (4.17)$$



Neste último exemplo é clara a distinção entre lançamentos efectuados num complexo de regiões e o lançamento numa única região (que não se restringe ao lançamento efectuado num paralelepípedo n -dimensional). A região total possível em relação a estes dois sistemas de lançamentos pode ser a mesma, mas a possibilidade de cada ponto, e, consequentemente, de cada região, serão geralmente distintos nos dois sistemas de lançamentos. Por conseguinte, não é a região total possível do lançamento que define a função de possibilidade de um ponto, mas sim o sistema de lançamentos adoptado.

Após expor a sua definição de possibilidade de uma região, Pacheco d'Amorim deduz as mesmas propriedades que nas possibilidades de classes. Assim, como Proposição II, apresenta a aditividade das possibilidades, isto é, se considerarmos uma região possível $\mathbf{X}'$ tal que possa ser decomposta por $\mathbf{X}' = \mathbf{X}'_1 \cup \mathbf{X}'_2 \cup \dots \cup \mathbf{X}'_n$, onde as regiões $\mathbf{X}'_i$ são incompatíveis duas a duas, então teremos

$$\varpi_{\mathbf{X}'} = \varpi_{\mathbf{X}'_1} + \varpi_{\mathbf{X}'_2} + \dots + \varpi_{\mathbf{X}'_n}. \quad (4.18)$$

Demonstração. De facto, facilmente se verifica que

$$\begin{aligned} & \varpi_{\mathbf{X}'_1} + \varpi_{\mathbf{X}'_2} + \dots + \varpi_{\mathbf{X}'_n} = \\ &= \int_{\mathbf{X}'_1} \pi_x \, dx + \int_{\mathbf{X}'_2} \pi_x \, dx + \dots + \int_{\mathbf{X}'_n} \pi_x \, dx = \int_{\mathbf{X}'} \pi_x \, dx = \varpi_{\mathbf{X}'}. \end{aligned}$$

□

A proposição III assegura que a possibilidade da região total possível $\mathbf{X}$, região constituída pela totalidade dos pontos possíveis em relação a determinado sistema de lançamentos, é igual à unidade, isto é,

$$\varpi_{\mathbf{X}} = 1. \quad (4.19)$$

Demonstração. Se os pontos são todos provenientes de lançamentos feitos numa só região $\mathbf{X}$ (composta ou não) o resultado é trivial, pois

$$\varpi_{\mathbf{X}} = \int_{\mathbf{X}} \frac{1}{\mu_{\mathbf{X}}} dx = \frac{1}{\mu_{\mathbf{X}}} \int_{\mathbf{X}} dx = \frac{\mu_{\mathbf{X}}}{\mu_{\mathbf{X}}} = 1.$$

Nos casos onde os lançamentos são efectuados num complexo de regiões $\mathbf{X} \odot \rightarrow \mathbf{Y}$ teremos

$$\begin{aligned} \varpi_{\mathbf{X} \odot \rightarrow \mathbf{Y}} &= \int_{\mathbf{X} \odot \rightarrow \mathbf{Y}} \pi_{(x,y)} d(x,y) = \\ &= \int_{\mathbf{X} \odot \rightarrow \mathbf{Y}} \frac{1}{\mu_{\mathbf{X}}} \cdot \frac{1}{\mu_{\mathbf{Y}|\mathbf{X}=x}} d(x,y) = \int_{\mathbf{X}} \frac{1}{\mu_{\mathbf{X}}} \int_{\mathbf{Y}|\mathbf{X}=x} \frac{1}{\mu_{\mathbf{Y}|\mathbf{X}=x}} dy dx = \\ &= \int_{\mathbf{X}} \frac{1}{\mu_{\mathbf{X}}} \frac{\mu_{\mathbf{Y}|\mathbf{X}=x}}{\mu_{\mathbf{Y}|\mathbf{X}=x}} dx = \frac{\mu_{\mathbf{X}}}{\mu_{\mathbf{X}}} = 1. \end{aligned}$$

□

A Proposição IV garante que a possibilidade de uma região composta é igual ao produto das regiões que a compõem. Deste modo teremos

$$\varpi_{(\mathbf{X} \times \mathbf{Y})} = \varpi_{\mathbf{X}} \cdot \varpi_{\mathbf{Y}}. \quad (4.20)$$

Demonstração. Uma vez que a região $\mathbf{X} \times \mathbf{Y}$ é composta das regiões $\mathbf{X}$ e $\mathbf{Y}$ (regiões independentes), teremos

$$\begin{aligned} \varpi_{\mathbf{X} \times \mathbf{Y}} &= \int_{\mathbf{X} \times \mathbf{Y}} \pi_{(x,y)} d(x,y) = \int_{\mathbf{X}} \int_{\mathbf{Y}} \pi_x \cdot \pi_y dy dx = \\ &= \int_{\mathbf{X}} \pi_x dx \cdot \int_{\mathbf{Y}} \pi_y dy = \varpi_{\mathbf{X}} \cdot \varpi_{\mathbf{Y}}. \end{aligned}$$

□

Saliente-se que é impossível permutar a ordem de lançamento dos pontos num complexo de regiões, isto é, se lançamos à sorte um ponto no complexo de regiões $\mathbf{X} \odot \rightarrow \mathbf{Y}$ é impossível traduzir este lançamento noutro lançamento no complexo de regiões $\mathbf{Y}^* \odot \rightarrow \mathbf{X}^{*(3)}$ que seja equivalente, por outras palavras, em que a ordem dos lançamentos seja inversa mas a função de possibilidade de cada ponto (x, y) da região total possível assuma o mesmo valor nos dois sistemas de lançamentos e, conseqüentemente, as regiões idênticas sejam associadas possibilidades idênticas. Para percebermos melhor esta ideia observemos os dois gráficos patentes na Figura 4.6 (que corresponde ao mesmo complexo de regiões retratado no exemplo 4.6 da página 232).

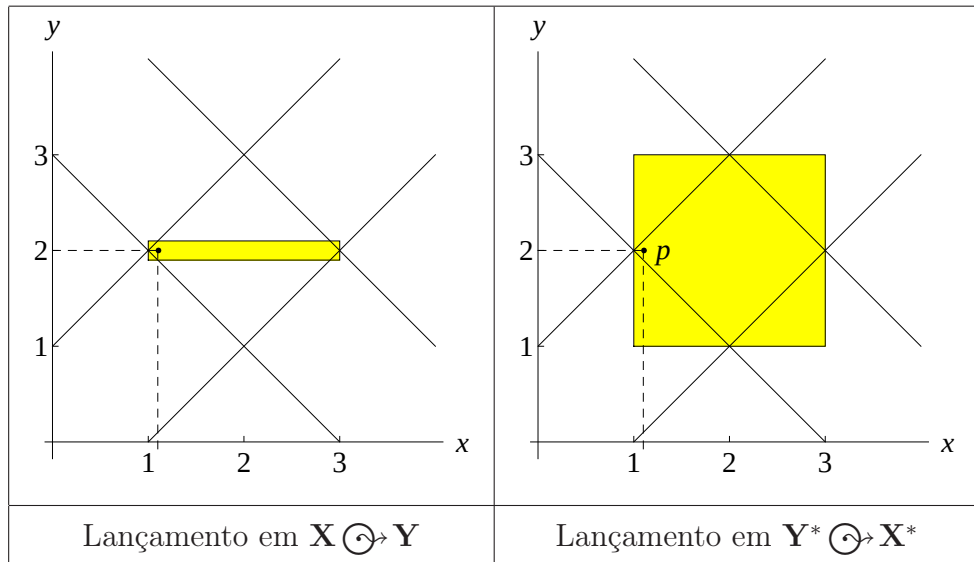


Figura 4.6: Inversão da ordem de lançamento de um complexo de regiões

No primeiro lançamento, correspondente ao complexo de regiões $\mathbf{X} \odot \rightarrow \mathbf{Y}$ que é retratado no primeiro gráfico da Figura 4.6, a possibilidade do ponto

⁽³⁾ As regiões $\mathbf{Y}^*$ e $\mathbf{X}^*$ são definidas de forma que o complexo de regiões $\mathbf{Y}^* \odot \rightarrow \mathbf{X}^*$ tenha a mesma região total possível que o complexo de regiões $\mathbf{X} \odot \rightarrow \mathbf{Y}$, isto é, que

$$\bigcup_{x \in \mathbf{X}} x \times \mathbf{Y}|_{\mathbf{X}=x} = \bigcup_{y \in \mathbf{Y}^*} y \times \mathbf{X}^*|_{\mathbf{Y}^*=y}.$$

$p = (x, y) = (1.1, 2)$ é obtida através da área do rectângulo a sombreado $[1, 3] \times [1.9, 2.1]$ uma vez que, se no primeiro lançamento em $[1, 3]$ sair o ponto $x = 1.1$, o segundo lançamento será efectuado na região $[1.9, 2.1]$ (segmento que liga os pontos $(1.1, 1.9)$ e $(1.1, 2.1)$), sendo a possibilidade do ponto $(1.1, 2)$ igual a $\pi_{(1.1, 2)} = \frac{1}{(3-1) \cdot (2.1-1.9)} = \frac{1}{0.4} = 2.5$. Em relação ao lançamento retratado no segundo gráfico da Figura 4.6, correspondente ao complexo de regiões $\mathbf{Y}^* \odot \mathbf{X}^*$, a possibilidade do ponto $(1.1, 2)$ será obtida usando a área do rectângulo a sombreado, uma vez que o primeiro ponto y é escolhido no intervalo $[1, 3]$ e, supondo que saiu $y = 2$, o segundo ponto x será escolhido no intervalo $[1, 3]$ (pois para $y = 2$ o valor de x pode ser qualquer um pertencente a este intervalo). Assim, a possibilidade do ponto $(1.1, 2)$, quando lançado no complexo $\mathbf{Y}^* \odot \mathbf{X}^*$, será igual a $\pi_{(1.1, 2)}^* = \frac{1}{(3-1) \cdot (3-1)} = \frac{1}{4} = 0.25$, o que é dez vezes inferior ao valor determinado para o mesmo ponto em relação ao lançamento no complexo de regiões $\mathbf{X} \odot \mathbf{Y}$. Desta forma, num lançamento de um ponto numa região, temos de distinguir o lançamento directo na região onde a possibilidade é igual para todos os pontos (no exemplo retratado seria $\pi = 0.5$ para qualquer ponto (x, y) pertencente à região total possível), dos lançamentos em complexos de regiões e, de entre estes, a ordem com que são efectuados os lançamentos, uma vez que não é possível converter uma ordem na outra. Notemos que a possibilidade de qualquer ponto num lançamento no complexo de regiões $\mathbf{X} \odot \mathbf{Y}$ é sempre uma função de x (uma vez que este valor determina a região $\mathbf{Y}|_{\mathbf{X}=x}$ onde vai ser lançado o segundo ponto y), enquanto a possibilidade de qualquer ponto num lançamento no complexo de regiões $\mathbf{Y}^* \odot \mathbf{X}^*$ será sempre uma função de y , pois este valor determinará a região $\mathbf{X}^*|_{\mathbf{Y}^*=y}$ na qual será lançado o segundo ponto x . Por esta razão, a única situação em que é possível inverter a ordem dos lançamentos e obter a mesma possibilidade para todos os pontos da região total possível corresponde aos lançamentos em regiões compostas onde a possibilidade é

constante, isto é, igual para todos os pontos.

Apesar de nesta construção hierarquizada de complexos de regiões não ser possível a inversão dos lançamentos, no capítulo PONTO IMAGEM, que analisaremos na secção 6 desta tese, Pacheco d'Amorim tenta lidar com a função de probabilidade conjunta, projectando a probabilidade conjunta num eixo, correspondente a um dos pontos lançados, e tomando a probabilidade do outro ponto de forma condicional ao valor obtido na projecção do primeiro, onde, a partir desta transformação, pode determinar o integral da função conjunta de forma iterada, reinventando, desta forma, o Teorema de Fubinni.

4.3 Probabilidade de $\mathbf{X}''$ em relação a $\mathbf{X}'$

Pacheco d'Amorim define probabilidade em regiões de forma análoga à que tinha efectuado em classes. Consideremos a região total possível $\mathbf{X}$ em relação a um sistema de lançamentos, a probabilidade de uma região $\mathbf{X}''$ (região favorável) em relação a outra região $\mathbf{X}'$ (região possível), com $\mathbf{X}'' \subset \mathbf{X}' \subset \mathbf{X}$, é definida como sendo o quociente entre as possibilidades destas regiões, ou seja, através de

$$\mathbb{P}_{\mathbf{X}'}(\mathbf{X}'') = \frac{\varpi_{\mathbf{X}''}}{\varpi_{\mathbf{X}'}}. \quad (4.21)$$

Esta fórmula, para as situações em que os pontos das regiões são equipossíveis, pode ser simplificada para

$$\mathbb{P}_{\mathbf{X}'}(\mathbf{X}'') = \frac{\mu_{\mathbf{X}''}}{\mu_{\mathbf{X}'}}. \quad (4.22)$$

Outra simplificação possível verifica-se quando a região possível coincide com a região total possível ($\mathbf{X}' = \mathbf{X}$), nestas situações teremos

$$\mathbb{P}_{\mathbf{X}'}(\mathbf{X}'') = \varpi_{\mathbf{X}''}, \quad (4.23)$$

que identifica a probabilidade condicionada à região total possível (probabilidade absoluta ou incondicionada) com a possibilidade de uma região.

O raciocínio que Pacheco d'Amorim utiliza para chegar à definição de probabilidade contínua (probabilidade em regiões) é análogo ao que previamente empregou para a definição de probabilidade discreta (probabilidade em classes). Desta forma, também nestas situações, o conceito de que parte é o de probabilidade condicionada, definindo unicamente a probabilidade de uma região em relação a outra região.

Nas obras suas contemporâneas por nós consultadas não é frequente encontrar uma definição rigorosa de probabilidade contínua, apesar de todos os autores determinarem probabilidades em regiões usando uma definição análoga à definição clássica de probabilidade baseada em equiprobabilidade, utilizando uma medida representativa da sua proporção geométrica (consoante a sua dimensão teremos o comprimento, a área, o volume, etc.). Desta forma a probabilidade, de uma dada região $\mathbf{A}$, é determinada pelo quociente entre a medida da região favorável e a medida da região possível, sob a hipótese de a probabilidade ser proporcional à medida (equipossibilidade) e a medida do universo ser finita. Esta definição é habitualmente designada por interpretação geométrica de probabilidade.

Contudo, Bertrand (1888, p. 1–5), após apresentar a definição de probabilidade (discreta) baseada em equiprobabilidade onde considera haver uma só condição, a equipossibilidade, salienta o facto de esta fórmula não ser aplicável nas situações onde o número de casos possíveis é infinito, afirmando

“Une condition est sous-entendue: tous les cas doivent être également possibles. (...) Une remarque encore est nécessaire: l’infini n’est pas un nombre; on ne doit pas, sans explication, l’introduire dans les raisonnements. La précision illusoire des mots pourrait faire naître des contradictions. Choisir au hasard, entre un nombre infini de cas possibles, n’est pas une indication suffisante.”

[Bertrand 1888, p. 2]

Bertrand fundamenta esta opinião com a apresentação dos seus famosos paradoxos. O mais célebre destes paradoxos, e por tal denominado por paradoxo de Bertrand, será analisado detalhadamente no capítulo LANÇAMENTO, À SORTE, DE FIGURAS, pelo facto de Pacheco d'Amorim, nesse capítulo, propor uma resolução para este problema⁽⁴⁾. Atente-se ainda em outro paradoxo apresentado por Bertrand, para salientar o seu cepticismo em relação à existência de uma definição rigorosa de probabilidade contínua:

“On demande, par exemple, la probabilité pour qu’un nombre, entier ou fractionnaire, commensurable ou incommensurable, choisi au hasard entre 0 et 100, soit plus grand que 50. La réponse semble évidente: le nombre des cas favorables est la moitié de celui des cas possibles. La probabilité est $\frac{1}{2}$.

Au lieu du nombre, cependant, on peut choisir son carré. Si le nombre est compris entre 50 et 100, le carré le sera entre 2500 et 10000.

La probabilité pour qu’un nombre choisi au hasard entre 0 et 10000 surpasse 2500 semble évidente: le nombre des cas favorables est le trois quarts du nombre des cas possibles. La probabilité est $\frac{3}{4}$.

Les deux problèmes sont identiques. D’où vient la différence des réponses? Les énoncés manquent de précision.

Les contradictions de ce genre peuvent être multipliées à l’infini.”

[Bertrand 1888, p. 4]

De facto, este paradoxo surge pelo facto de a expressão “*au hasard*” não ter um significado claro quando aplicado a um universo contínuo, mostrando a necessidade de clarificar o espaço-amostra⁽⁵⁾ associado à experiência aleatória

⁽⁴⁾ Consultar página 274.

⁽⁵⁾ Este conceito, que é basilar nos fundamentos apresentados por Kolmogoroff em 1933,

que está sendo realizada. Com as ferramentas de que actualmente dispomos na Teoria da Probabilidade é fácil demonstrarmos que, se uma variável aleatória X é caracterizada por uma distribuição uniforme no intervalo $[0, 100]$, então o seu quadrado, variável aleatória $Y = X^2$, não segue uma distribuição uniforme no intervalo $[0, 10000]$, resultando, deste facto, as probabilidades distintas obtidas por Bertrand no texto supracitado. Deste modo, para Bertrand, quando os resultados possíveis de um experiência aleatória são em número infinito, situação presente em qualquer problema da probabilidade contínua, a escolha à sorte (*au hasard*) numa região não é um enunciado suficiente para a sua resolução.

Borel (1909) não tem uma postura tão céptica, dando uma importância elevada a este tipo de probabilidade, como se nota pelo facto de lhe dedicar uma das três partes em que divide os seus *Éléments de la Théorie des Probabilités*: Probabilidades descontínuas, Probabilidades contínuas e Probabilidades das causas. Na parte dedicada às probabilidades contínuas começa por referir que, como a definição da probabilidade descontínua depende do número de caso possíveis e, por exemplo, num segmento de recta, o número de casos possíveis é indefinido, é urgente uma nova definição de probabilidade para a probabilidade contínua, sendo a definição adoptada

“La probabilité pour que le point M se trouve sur un certain segment PQ de AB est proportionnelle à la longueur de ce segment.”

[Borel 1910, p. 84]

Nas páginas seguintes Borel generaliza esta definição para superfícies planas e de dimensão superior através do quociente dos integrais ao longo da região favorável e da região possível. Borel, ciente dos paradoxos que a utilização

foi unicamente criado por Richard von Mises na tentativa de formalização da interpretação frequentista de Probabilidade com o recurso ao seu conceito de *colectivo* e é posterior à tese de Pacheco d'Amorim.

desta definição origina nos problemas de probabilidade geométrica, julga-a convencional e crê que deve ser confirmada *a posteriori* com os resultados de experiências, afirmando, no capítulo que dedica à resolução de problemas geométricos, que apenas considera os problemas geométricos *concretos*, que define como sendo aqueles em que é possível determinar um método de “*vérification expérimentale des résultats*”.

Borel (1909, p. 84–85), contudo, após definir a probabilidade contínua, apresenta, de seguida, uma forte objecção a esta definição, baseada no quociente das medidas favorável e possível, expondo um problema semelhante ao paradoxo apresentado por Bertrand previamente transcrito (na página 241) e generalizando este problema do modo que passamos a expor. Seja x uma variável aleatória contínua definida no intervalo $[a, b]$, sendo a probabilidade, obtida utilizando a definição geométrica de probabilidade, de esta variável se situar no intervalo $[c, d] \subset [a, b]$ dada por $\frac{d-c}{b-a}$. Se considerarmos uma função crescente f aplicada à variável x , obteremos a variável aleatória contínua $f(x)$ definida no intervalo $[f(a), f(b)]$, sendo a probabilidade de pertencer ao intervalo $[f(c), f(d)] \subset [f(a), f(b)]$, utilizando o mesmo raciocínio, igual a $\frac{f(d)-f(c)}{f(b)-f(a)}$. Todavia, como, sob a hipótese da monotonia previamente referida, temos $a \leq x \leq b \Leftrightarrow f(a) \leq f(x) \leq f(b)$ e $c \leq x \leq d \Leftrightarrow f(c) \leq f(x) \leq f(d)$, as probabilidades referidas deveriam ser iguais para qualquer função f , mas geralmente verifica-se

$$\frac{d-c}{b-a} \neq \frac{f(d)-f(c)}{f(b)-f(a)}. \quad (4.24)$$

Deste modo podemos até escolher a forma da função f de forma conveniente, para que a probabilidade obtida assumo o valor, seja ele qual for, que pretendamos.

Poincaré (1896, p. 118 e seguintes) começa por ilustrar os problemas que podem surgir quando trabalhamos com um número de casos possíveis infinito, recorrendo ao paradoxo de Bertrand e às suas três possibilidades

de resolução, para depois considerar que a probabilidade de uma variável aleatória contínua X se situar entre x_0 e x_1 , representada por $\mathbb{P}(x_0, x_1)$, é dada por

$$\mathbb{P}(x_0, x_1) = \int_{x_0}^{x_1} \varphi(x) \, dx, \quad (4.25)$$

onde considera que φ é uma função que nós ignoramos e, como tal, é arbitrária. Por tais razões, Poincaré considera que é através do problema em análise que devemos definir, por uma qualquer convenção que forneça a essa função um sentido no contexto do problema em resolução, a forma da função φ , referindo

“Le mathématicien n’a plus aucune prise sur le choix de cette hypothèse; mais il doit, une fois qu’elle est choisie, porter son attention à ne pas en faire une autre qui la contredise.”

[Poincaré 1896, p. 121]

Keynes (1921, p. 47–48) utiliza o mesmo argumento (o de ser possível obter probabilidades distintas a acontecimentos equivalentes obtidos através da aplicação de uma transformação à variável original) enquanto debate as limitações da utilização do princípio da razão insuficiente de Laplace que apelida de princípio da indiferença.

Pacheco d’Amorim não se refere explicitamente a este problema, mas certamente teve-o em consideração, pois a sua construção, nomeadamente a efectuada no capítulo do PONTO IMAGEM, permite-lhe resolver este paradoxo, ao fazer uma distinção clara entre a probabilidade dum ponto lançado à sorte, onde a probabilidade é obrigatoriamente uniforme, e a probabilidade dum outro ponto que seja sua imagem, onde define a lei de probabilidade. Deste modo, para Pacheco d’Amorim, está justificada a diferença de resultados apresentados em (4.24), pois, se o ponto x é lançado à sorte no intervalo $[a, b]$, este terá distribuição uniforme, mas, como $f(x)$ é o *ponto imagem*

de x , então não terá (necessariamente) distribuição uniforme, sendo a probabilidade do *ponto imagem* $f(x)$ definida em função da probabilidade do acontecimento equivalente do ponto directamente lançado (x), isto é, através de $\mathbb{P}(c \leq x \leq d) = \mathbb{P}[f(c) \leq f(x) \leq f(d)]$ (no que se refere à situação anteriormente apontada de f ser uma função crescente). Assim, para o autor, unicamente utilizamos a definição geométrica de probabilidade, isto é, o quociente das medidas das regiões favorável e possível, quando lançamos pontos, à sorte, directamente numa região (pois neste caso existe equipossibilidade e, conseqüentemente, a probabilidade numa região é proporcional à sua medida). Quando analisarmos o capítulo PONTO IMAGEM da tese de Pacheco d'Amorim comentaremos mais detalhadamente esta ideia.

Em forma de observação, após a apresentação do problema do triângulo, Pacheco d'Amorim considera válidas para a probabilidade contínua as Propriedades V, VI e VII apresentadas e demonstradas na probabilidade em classes. O autor refere ainda que na probabilidade contínua o Problema das Causas pode ser analisado de forma análoga à efectuada na probabilidade discreta, contudo essa análise será efectuada, de uma forma mais geral, no capítulo PONTO IMAGEM da sua tese.

4.4 Problema do triângulo

O primeiro exemplo que Pacheco d'Amorim apresenta de aplicação da probabilidade contínua é o problema do triângulo que passamos a enunciar.

Problema 4.1. Consideremos um segmento de recta $[ab]$ que se *parte*, à sorte, em três partes: qual a probabilidade de que essas três partes (segmentos) possam formar um triângulo?

SOLUÇÃO: Em primeiro lugar, partir um segmento de recta, à sorte, em três partes é o mesmo que lançar, à sorte, dois pontos no mesmo segmento de

recta; e lançar, à sorte, dois pontos no segmento $[ab]$ é o mesmo que lançar, à sorte, um único ponto num quadrado com lados iguais a esse segmento. Seja,

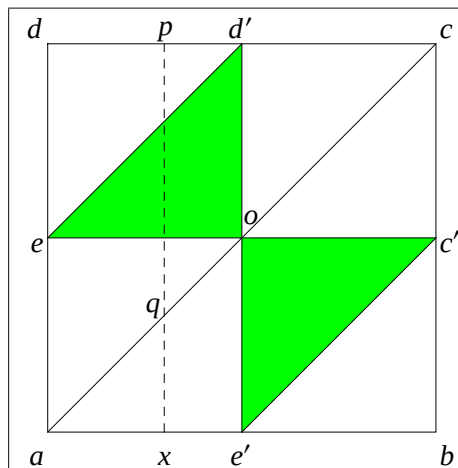


Figura 4.7: Lançar, à sorte, um ponto num quadrado

então, lançado um ponto no quadrado $[abcd]$ da Figura 4.7 e sejam (x, y) as coordenadas do ponto determinado pelo lançamento, correspondendo a coordenada x ao primeiro ponto e a coordenada y ao segundo ponto lançado sobre o segmento $[ab]$. Deste modo, temos duas situações distintas onde os três segmentos obtidos possam formar um triângulo. O caso em que $x < y$ e os segmentos $[ax]$, $[xy]$ e $[yb]$ formem um triângulo; e o caso $x > y$ onde os segmentos a considerar serão $[ay]$, $[yx]$ e $[xb]$.

Consideremos a primeira situação $x < y$ que corresponde na figura à situação em que o ponto (x, y) se situa no triângulo $[adc]$. Para que os três segmentos possam formar um triângulo é necessário e suficiente que o comprimento de cada lado seja inferior à soma dos comprimentos dos outros dois, situação que é representada na figura pelo triângulo $[od'e]$ (região favorável $\mathbf{X''}$). Utilizando raciocínio análogo, conclui-se que, no caso de $x > y$ (triângulo $[abc]$), a região favorável seria formada pelo triângulo $[e'c'o]$. E como todos os pontos são igualmente possíveis (região composta), a probabilidade pedida será dada pelo quociente das áreas das regiões favorável e possível

que é igual $\mathbb{P} = \frac{1}{4}$. $\diamond$

Este problema foi proposto por Lemoine (1872) no ano de fundação da *Société Mathématique de France*, tendo sido publicado no seu primeiro volume.

“Une tige se brise en trois morceaux; quelle est la probabilité pour que, avec ces trois morceaux, on puisse former un triangle?”

[Lemoine 1872, p. 39]

Neste artigo, para a resolução do problema, Lemoine considera que o segmento inicial está dividido em $2m$ segmentos de igual comprimento, sendo cada um dos três segmentos resultantes da partição (à sorte) constituído por um determinado número destes segmentos que representaremos por x , y e z ($x + y + z = 2m$). Utilizando este raciocínio, recorre à definição clássica de Probabilidade, determinando a probabilidade pretendida, em função de m , através do quociente entre o número de casos favoráveis e o número de casos possíveis. Para o número de casos possíveis, considerando que $x + y + z = 2m$, teremos

$$\begin{array}{ll} x = 0 & y = 0, 1, \dots, 2m; \\ x = 1 & y = 0, 1, \dots, 2m - 1; \\ \dots & \dots; \\ x = 2m & y = 0. \end{array}$$

O número de casos possíveis é então obtido através de

$$1 + 2 + 3 + \dots + (2m + 1) = \frac{(2m + 1)(2m + 2)}{2}. \quad (4.26)$$

Para determinar o número de casos favoráveis teremos que impor que o comprimento de cada lado seja inferior à soma dos outros dois lados que, utilizando $x + y + z = 2m$, se obtém $x \leq m$, $y \leq m$ e $x + y \geq m$. Deste modo teremos como resultados favoráveis

$$\begin{array}{ll}
x = 0 & y = m; \\
x = 1 & y = m, m - 1; \\
\cdots & \cdots; \\
x = m & y = m, m - 1, \cdots, 0.
\end{array}$$

O número de casos favoráveis será determinado por

$$1 + 2 + 3 + \cdots + (m + 1) = \frac{(m + 1)(m + 2)}{2}. \quad (4.27)$$

Pela definição clássica de Probabilidade, teremos

$$\mathbb{P} = \frac{m + 2}{2(2m + 1)}. \quad (4.28)$$

Lemoine determina depois o limite desta expressão quando m tende para infinito, obtendo o valor $\frac{1}{4}$.

Uns anos mais tarde, Lemoine (1883) afirma que este problema, desde a sua proposta, captou a atenção de numerosos matemáticos, referindo Halphen, Jordan e Lalanne. Neste artigo Lemoine recorre à definição geométrica de Probabilidade, considerando que a probabilidade de cada região é proporcional à sua área. Lemoine começa por considerar que o problema é equivalente a escolher à sorte um ponto m no interior de um qualquer triângulo equilátero $[abc]$, pois a soma das distâncias de qualquer ponto (interior ao triângulo) aos lados do triângulo é sempre igual à altura desse triângulo (ou ao comprimento de qualquer bissectriz do triângulo). Se representarmos por a' , b' e c' os pontos médios de cada um dos lados do triângulo $[abc]$, o triângulo $[a'b'c']$ corresponde à região favorável, sendo a probabilidade pretendida o quociente entre a área deste triângulo e a do triângulo $[abc]$. Lemoine apresenta, ainda no mesmo artigo, outras variantes deste problema, como, por exemplo, qual a probabilidade de os três segmentos obtidos formarem um triângulo acutângulo (triângulo com os três ângulos agudos).

Czuber (1884, p. 13–14) também identifica a origem do problema com Lemoine e apresenta duas formas de resolvê-lo. Na primeira considera-o como

um caso particular de um outro problema por ele resolvido, problema esse que consiste em determinar a probabilidade de, ao partir um segmento de recta de comprimento α em três partes, nenhum dos três segmentos resultantes ter um comprimento superior a um determinado valor β . Assim, para resolver o problema do triângulo bastará utilizar $\beta = \frac{\alpha}{2}$. Na segunda resolução Czuber considera x , y e z as coordenadas (x, y, z) de um ponto no espaço onde cada uma destas variáveis representa o comprimento de cada um dos três segmentos resultantes. No plano representamos as restrições do problema: $x + y + z = 1$ (Czuber considerou o comprimento do segmento igual à unidade) e $x, y, z \geq 0$. Obtemos desta forma a região possível — triângulo com vértices em $(0, 0, 1)$ e $(0, 1, 0)$ e $(1, 0, 0)$, que é equilátero (correspondente à ideia apresentada por Lemoine em 1883). A região favorável é obtida ao impor, dentro da região possível, as condições $x + y \geq z$, $x + z \geq y$ e $y + z \geq x$, obtendo-se o triângulo com vértices $(0.5, 0.5, 0)$ e $(0, 0.5, 0.5)$ e $(0.5, 0, 0.5)$. O quociente das áreas dos triângulo dá-nos a probabilidade pretendida.

Poincaré (1896, p. 123–126), denominando o problema por *problème du bâton brisé*, resolve-o de uma forma semelhante à de Lemoine e Czuber. Assim, considerou um segmento de comprimento unitário que é partido em três segmentos com comprimentos x , y e z ($x + y + z = 1$). O comprimento destes três segmentos pode ser representado por um ponto m interno a um triângulo equilátero com altura igual à unidade, pois a soma das distâncias deste ponto a cada um dos lados do triângulo assume este valor (conforme Figura 4.8), representando então cada uma destas distâncias o comprimento de cada um dos segmentos. Para que os três segmentos possam formar um triângulo, o comprimento de cada lado tem de ser inferior à soma dos outros dois, sendo, então, a área favorável a apresentada a sombreado.

Poincaré considera que a probabilidade de cada um dos pontos é proporcional ao comprimento, isto é, utilizando a linguagem e notação deste autor,

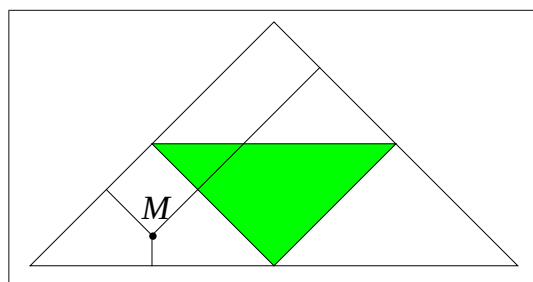


Figura 4.8: Problemas dos Triângulos

a probabilidade de a variável X estar compreendida entre x e $x + dx$ é proporcional a dx e a probabilidade de Y estar compreendida entre y e $y + dy$ é proporcional a dy . Por este motivo a probabilidade de se verificarem ambas as condições é proporcional a $dx dy$.

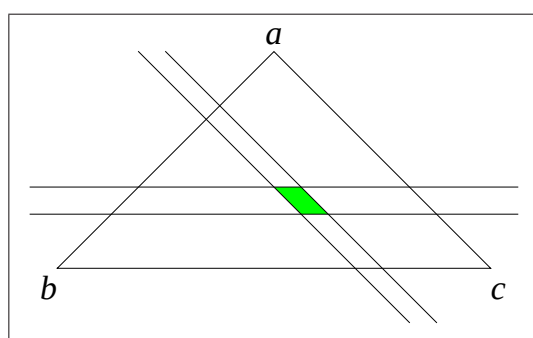


Figura 4.9: Problemas dos Triângulos 2

Assim, se representarmos as rectas $X = x$, $X = x + dx$ (horizontais), $Y = y$ e $Y = y + dy$ (paralelas ao lado $[ac]$ do triângulo), obteremos um paralelograma (a sombreado na Figura 4.9) cuja área é $\frac{2dx dy}{\sqrt{3}}$. A probabilidade será então proporcional à área do paralelograma e, de forma geral, proporcional à área da região (contida no triângulo $[abc]$) da qual pretendemos determinar a probabilidade. Desta forma, como facilmente se verifica na Figura 4.8, a probabilidade pretendida é igual a $\frac{1}{4}$.

Bachelier (1912, p. 405) utiliza um raciocínio análogo à primeira resolução de Lemoine, dividindo o segmento original em um número elevado de

segmentos de igual comprimento e utilizando o quociente entre o número de casos favoráveis e o número de casos possíveis, determinando posteriormente o seu limite.

Do conjunto de soluções aqui apresentadas, contemporâneas à tese de Pacheco d'Amorim, ressalta-nos que, apesar de não questionarmos a validade destas resoluções e de algumas serem, na nossa opinião, extremamente bem concebidas, nenhuma delas usufrui da simplicidade e da elegância presentes na de Pacheco d'Amorim. De facto este autor, ao traduzir o *partir* à sorte um segmento de recta em três partes num lançamento, à sorte, de um ponto num quadrado, consegue transformar o problema do triângulo num outro de resolução bastante acessível, não sendo, desta forma, necessário recorrer a processos tão elaborados como os de Lemoine, Czuber ou Poincaré previamente apresentados. Mais tarde, Rényi (1970, p. 66–67) resolve este problema de forma semelhante à de Pacheco d'Amorim, considerando que escolher dois pontos num segmento unitário é o mesmo que escolher um ponto no quadrado $(0, 1) \times (0, 1)$ e que a probabilidade de o ponto cair numa determinada região é proporcional (neste caso será mesmo igual) à sua área. Esta ideia é em tudo semelhante à utilizada por Pacheco d'Amorim. Refira-se ainda que a figura apresentada por Rényi (1970, Fig. 12, p. 67) para fundamentar a sua resolução é, curiosamente, igual à apresentada por Pacheco d'Amorim (1914, Fig. 4, p. 45).

Problema 4.2. Consideremos agora que existe uma certa dependência entre os pontos escolhidos. Assim, comecemos por *partir*, à sorte, o segmento $[ab]$ em dois. De seguida escolhemos, de entre os dois segmentos obtidos, o segmento maior e *partimos*, à sorte, esse segmento em dois. Nesta situação, qual a probabilidade de que os três segmentos achados possam formar um triângulo?

SOLUÇÃO: Neste problema o campo favorável é o mesmo do problema antecedente, mas o campo possível foi alterado, pois deixou de ser possível obter pontos na região $[aoe']$ (consultar Figura 4.7 presente na página 246), situação em que se verifica $\overline{ax} < \overline{xb}$ e é o segmento $[xb]$ que é *partido* para determinar o valor de y , e na região $[d'oc]$ onde $\overline{ax} > \overline{xb}$ e é utilizado o segmento $[ax]$ para determinar o valor de y . Deste modo, temos como regiões possíveis as regiões $[aod'd]$ e $[e'ocb]$. Já não estamos, assim, a trabalhar numa região composta, mas num complexo de regiões (as regiões já não são independentes, logo, os pontos não têm a mesma possibilidade). Pacheco d'Amorim, pelo facto de a questão ser simétrica em relação a $[ac]$, considerou nos seus cálculos só $[aod'd]$ como região possível e $[od'e]$ como região favorável. Representando por $\mathbf{F}$ a região favorável e por $\mathbf{T}$ a região total possível, será $\varpi_{\mathbf{T}} = 1$ e, representando o comprimento do segmento $[ab]$ por α , teremos

$$\begin{aligned}
 \varpi_{\mathbf{F}} &= \int_{\mathbf{F}} \frac{dx}{\frac{\alpha}{2}} \cdot \frac{dy}{\alpha - x} = \frac{2}{\alpha} \int_0^{\frac{\alpha}{2}} \frac{dx}{\alpha - x} \int_{\frac{\alpha}{2}}^{x+\frac{\alpha}{2}} dy = \\
 &= \frac{2}{\alpha} \int_0^{\frac{\alpha}{2}} \frac{x}{\alpha - x} dx = \frac{2}{\alpha} \int_0^{\frac{\alpha}{2}} \left(-1 + \frac{\alpha}{\alpha - x} \right) dx = \\
 &= -1 + 2 \left[-\ln(\alpha - x) \right]_0^{\frac{\alpha}{2}} = -1 + 2 \ln 2 \simeq 0,38629 \quad (4.29)
 \end{aligned}$$

e, portanto,

$$\mathbb{P}_{\mathbf{P}}(\mathbf{F}) \simeq 0,38629. \quad (4.30)$$

Refira-se que o resultado final difere do apresentado pelo autor, obtido pela mesma expressão, que exhibe 0.44. $\diamond$

Pacheco d'Amorim apresenta ainda uma outra variante do problema do triângulo.

Problema 4.3. Começemos por lançar, à sorte, o ponto x em $[ae']$ (Figura 4.7 apresentada na página 246) e depois lança-se o ponto y no segmento $[xb]$. Supondo que o ponto y caiu no segmento $[e'b]$, qual é a probabilidade de que os três segmentos $([ax], [xy]$ e $[yb])$ possam formar um triângulo?

SOLUÇÃO: As regiões total possível, possível e favorável são respectivamente $\mathbf{T} = [aod'd]$, $\mathbf{P} = [eod'd]$ e $\mathbf{F} = [eod']$. Teremos pois:

$$\begin{aligned}\varpi_{\mathbf{P}} &= \int_{\mathbf{P}} \frac{dx}{\frac{\alpha}{2}} \cdot \frac{dy}{\alpha - x} = \int_0^{\frac{\alpha}{2}} \int_{\frac{\alpha}{2}}^{\alpha} \frac{dx}{\frac{\alpha}{2}} \cdot \frac{dy}{\alpha - x} = \\ &= \int_0^{\frac{\alpha}{2}} \frac{dx}{\alpha - x} = [-\ln(\alpha - x)]_0^{\frac{\alpha}{2}} = \ln \alpha - \ln \frac{\alpha}{2} = \ln 2\end{aligned}\quad (4.31)$$

e (ver solução do problema anterior – Problema 4.2)

$$\varpi_{\mathbf{F}} = 2 \ln 2 - 1;\quad (4.32)$$

logo,

$$\mathbb{P}_{\mathbf{P}}(\mathbf{F}) = 2 - \frac{1}{\ln 2} \simeq 0.5573,\quad (4.33)$$

tendo, para a mesma expressão, Pacheco d'Amorim obtido 0.6. $\diamond$

Não encontramos estas duas últimas variantes do problema de Lemoine em mais obra nenhuma. Pensamos que, porventura, Pacheco d'Amorim tenha criado estas variantes para justificar a necessidade de generalizar a definição de probabilidade e de recorrer à criação de complexos de regiões. Recordamos que no PREFÁCIO da sua tese, quando se propõe generalizar a definição de probabilidade, Pacheco d'Amorim justifica a importância desta pelo facto de ser necessária para a demonstração do Teorema de Bayes, assunto analisado no capítulo anterior, e “*absolutamente indispensável na probabilidade contínua, como claramente mostra*” a resolução desta última variante do problema do triângulo.

4.5 Simetria

Com o intuito de facilitar a determinação de probabilidades em alguns problemas que manifestem algum elemento de simetria, Pacheco d'Amorim apresenta duas proposições.

Na Proposição VIII o autor analisa os problemas em que a região $\mathbf{X}$, onde é efectuado o lançamento do ponto, pode ser decomposta em duas componentes $\mathbf{X}_1$ e $\mathbf{X}_2$, de tal forma que para cada *coordenada* de $\mathbf{X}_1$ se tem sempre o mesmo campo favorável e o mesmo campo possível em $\mathbf{X}_2$, isto é, as regiões favoráveis e possíveis assumem a mesma possibilidade independentemente do valor das coordenadas de $\mathbf{X}_1$. Nestas situações as componentes $\mathbf{X}_1$ e $\mathbf{X}_2$ da região $\mathbf{X}$ são *independentes*, podendo ser determinadas de forma separada. É o que acontece quando analisamos a possibilidade ou probabilidade numa região composta $\mathbf{X} = \mathbf{X}_1 \times \mathbf{X}_2$. Se, além do referido, a possibilidade de uma das componentes, seja $\mathbf{X}_1$, é igual à unidade, então a possibilidade da região $\mathbf{X}$ não depende das *coordenadas* de $\mathbf{X}_1$ e podemos determinar essa possibilidade considerando uma coordenada qualquer de $\mathbf{X}_1$ fixa. Este resultado será aplicado na resolução do problema na superfície esférica (paradoxo de Borel-Bertrand) retratado na página 255⁽⁶⁾.

Na Proposição IX Pacheco d'Amorim considera os problemas em que a região analisada pode ser decomposta em n regiões disjuntas onde as probabilidades são iguais. Isto significa que se pode decompor a região possível $\mathbf{X}$

⁽⁶⁾ Destaquemos imediatamente que esta propriedade não é sempre válida, nomeadamente no que respeita a situações onde dividimos o universo numa infinidade (não numerável) de conjuntos de medida nula, como ilustra o paradoxo de Borel-Kolmogoroff. Contudo, o alheamento de Pacheco d'Amorim em relação a este problema pode ser justificado pelo facto de, nesta altura, as noções elementares da Teoria de Medida serem ainda muito incipientes.

através de

$$\mathbf{X} = \bigcup_{i=1}^n \mathbf{X}_i, \quad \mathbf{X}_i \cap \mathbf{X}_j = \emptyset, \forall i \neq j, \quad (4.34)$$

onde as regiões $\mathbf{X}_i$ formam uma partição da região $\mathbf{X}$ e, consequentemente, a região favorável $\mathbf{X}'$ será decomposta através de

$$\mathbf{X}' = \bigcup_{i=1}^n \mathbf{X}'_i, \quad \mathbf{X}'_i = \mathbf{X}' \cap \mathbf{X}_i, \quad i = 1, \dots, n, \quad (4.35)$$

verificando-se, em cada uma destas regiões,

$$\mathbb{P}_{\mathbf{x}_1}(\mathbf{X}'_1) = \mathbb{P}_{\mathbf{x}_2}(\mathbf{X}'_2) = \dots = \mathbb{P}_{\mathbf{x}_n}(\mathbf{X}'_n). \quad (4.36)$$

Então, nestas condições, é suficiente determinar a probabilidade numa das regiões que constituem a partição da região possível, pois temos

$$\mathbb{P}_{\mathbf{x}}(\mathbf{X}') = \frac{\varpi_{\mathbf{x}'}}{\varpi_{\mathbf{x}}} = \frac{\sum_{i=1}^n \varpi_{\mathbf{x}'_i}}{\sum_{i=1}^n \varpi_{\mathbf{x}_i}} = \frac{\varpi_{\mathbf{x}'_i}}{\varpi_{\mathbf{x}_i}} = \mathbb{P}_{\mathbf{x}_i}(\mathbf{X}'_i) \quad (4.37)$$

para qualquer $i = 1, \dots, n$. Este argumento será utilizado por Pacheco d'Amorim na resolução do problema de Buffon⁽⁷⁾ no capítulo LANÇAMENTO, À SORTE, DE FIGURAS.

4.6 Paradoxo de Borel-Kolmogoroff

Problema 4.4 (Problema da superfície esférica). Lançam-se, à sorte, dois pontos sobre a superfície duma esfera de raio r . Pergunta-se: qual a probabilidade de que o menor arco do círculo máximo que liga os dois pontos seja inferior a α ?

SOLUÇÃO: Sejam m_1 e m_2 os pontos escolhidos. A resolução de Pacheco d'Amorim baseia-se na sua Proposição VIII (presente na página 254), pois

⁽⁷⁾ Consultar página 289.

o autor considera que qualquer que seja a posição de um dos pontos, por exemplo m_1 , para o outro corresponderá sempre a mesma região possível e uma região favorável com a mesma possibilidade. Desta forma podemos considerar para m_1 um qualquer ponto fixo. Supondo, então, m_1 fixo, a região favorável será dada pela calote tendo o ângulo 2α de abertura e m_1 por vértice, sendo a região possível toda a superfície da esfera. Desta forma, para Pacheco d'Amorim, o problema é equivalente a este outro mais elementar: qual a probabilidade de que um ponto lançado à sorte na superfície duma esfera caía sobre um segmento dessa superfície? Este problema é considerado de “*solução imediata*” por Pacheco d'Amorim.

Deste modo, para Pacheco d'Amorim, a resolução deste problema deverá ser efectuada considerando que todos os pontos da superfície esférica são equipossíveis, sendo suficiente para o cálculo da probabilidade determinar o quociente das áreas das regiões favorável e possível. Assim, a área da região favorável, superfície dada pela calote com ângulo 2α de abertura, é igual a $2\pi r^2 [1 - \cos(\alpha)]$, enquanto a região possível, toda a superfície da esfera, tem área igual a $4\pi r^2$. Deste modo a probabilidade pretendida é dada por

$$\mathbb{P} = \frac{2\pi r^2 [1 - \cos(\alpha)]}{4\pi r^2} = \frac{1 - \cos(\alpha)}{2} = \sin^2 \left(\frac{\alpha}{2} \right). \quad (4.38)$$

◇

O Problema 4.4 aparece, pela primeira vez, na obra de Bertrand (1888, p. 6) restrito à situação $\alpha = 10'$, isto é, o autor francês utiliza uma distância igual a $10'$ e não um valor qualquer arbitrário α . Contudo, Bertrand propõe para este problema duas resoluções com as quais obtém duas probabilidades distintas. Nas duas resoluções Bertrand considera, tal como Pacheco d'Amorim, que o primeiro ponto pode ser considerado conhecido, pois, seja qual for a sua posição, a probabilidade obtida será a mesma. Na primeira situação Bertrand considera ainda que o arco de círculo máximo que une os dois pontos também pode ser considerado conhecido, pois as medidas das

regiões favorável e total são as mesmas qualquer que seja o arco considerado, isto é, a probabilidade obtida é independente da direcção utilizada para definir o arco. Assim, Bertrand considerou que o arco de círculo máximo estava dividido em 2160 arcos, cada um com comprimento $10'$. Desta forma a probabilidade pretendida, tendo em conta que só existem dois arcos favoráveis, é igual a $\frac{2}{2160} = \frac{1}{1080}$. A segunda resolução de Bertrand é análoga à descrita por Pacheco d'Amorim, obtendo-se

$$\mathbb{P} = \sin^2 \left(\frac{10'}{2} \right) = \sin^2 \left(\frac{\pi}{2160} \right) \approx 0.0000021 \neq \frac{1}{1080} \approx 0,0009259! \quad (4.39)$$

Obtemos assim duas probabilidades distintas, uma considerando um arco de círculo máximo (pois as regiões favorável e total são as mesmas, qualquer que seja o arco considerado) e a outra utilizando as áreas das regiões favorável e possível da superfície esférica. Bertrand exhibe, assim, para o mesmo problema duas respostas distintas, utilizando duas formas de resolução que intuitivamente deveriam fornecer o mesmo resultado, pois parecem ser aplicados os mesmos princípios, apresentando desta forma mais um paradoxo na Teoria da Probabilidade, paradoxo este que é actualmente usual denominar por paradoxo de Borel-Kolmogoroff⁽⁸⁾. Para Bertrand as duas respostas estão correctas, utilizando este facto para salientar, uma vez mais, a sua ideia de que os problemas em probabilidade contínua têm de ser bem definidos *a priori*, o que considera não acontecer com este problema particular.

Borel (1909, p. 100–101) e (1914, p. 86–87) resolve o problema considerando o caso geral, como Pacheco d'Amorim, da distância ser α e utilizando as duas formas de resolução de Bertrand. Contudo, Borel não aceita que as análises de Bertrand sejam ambas correctas, defendendo que somente a segunda resolução de Bertrand, que corresponde à que Pacheco d'Amorim adoptou, é correcta, pois é a única que verifica o postulado da definição de

⁽⁸⁾ Kolmogoroff (1933) explica este paradoxo denominando-o de paradoxo de Borel.

probabilidade contínua, da probabilidade de uma região ser proporcional à sua medida. Assim, neste problema, se dividirmos a superfície esférica em regiões de igual área, então a probabilidade do ponto m_1 (ou m_2) pertencer a uma dessas regiões deve ser igual em todas essas regiões. Borel considera que isto acontece na segunda solução de Bertrand, mas não na primeira, justificando esta asserção com o raciocínio a seguir reconstituído. Se considerarmos um ponto m_1 fixo e um dos arcos de círculo máximo que passa pelo ponto m_1 igualmente fixo, então a probabilidade de m_2 se situar em determinada região do arco será proporcional ao seu comprimento. No entanto, Borel afirma que, se considerarmos que os arcos não têm largura, então a probabilidade dos pontos m_1 e m_2 se situarem sobre o arco é nula, como tal, para tornar este cálculo possível, devemos considerar dois arcos (próximos) de círculo máximo que passem pelo ponto fixo m_1 . Neste raciocínio é notório que os pontos não têm todos a mesma probabilidade, pois, se o ponto m_1 se situar no pólo norte (fazendo comparação da esfera com o nosso planeta), os pontos situados no equador serão mais prováveis que os situados no pólo sul.

Notemos que, apesar das soluções díspares apresentadas quer por Bertrand quer por Borel⁽⁹⁾, Pacheco d'Amorim considera tal problema de “*solução imediata*”. Contudo, salientemos que a primeira resolução apresentada por Bertrand, distinta da de Pacheco d'Amorim, também é válida segundo os fundamentos do Cálculo das Probabilidades defendidas na tese de doutoramento de Pacheco d'Amorim. Para conferir esta ideia, basta verificar que, seja qual for a direcção do arco, temos sempre a mesma probabilidade de o segundo ponto se situar a uma distância menor que α do primeiro, logo,

⁽⁹⁾ Relembremos que os livros de Bertrand (1888) e Borel (1909) pertencem à pequena lista de cinco obras citadas ao longo da tese de Pacheco d'Amorim, conjuntamente com Bernoulli (*Ars Conjectandi*), Poincaré (*La Science et l'Hypothèse*) e Laplace (*Essai Philosophique sur les Probabilités*).

segundo a Proposição VIII apresentada por Pacheco d'Amorim (que também é utilizada para justificar o seu resultado), é suficiente utilizar um dos círculos máximos para determinar a probabilidade. Pensamos que, por este motivo, Pacheco d'Amorim deveria ter notado que, mesmo com a sua construção, as duas respostas apresentadas previamente por Bertrand e Borel são válidas e, conseqüentemente, fica demonstrado que este não é um problema de resolução fácil. Este facto deveria ter alertado Pacheco d'Amorim para a possibilidade de a Proposição VIII não poder ser aplicada em todas as situações. Com efeito, tal como foi previamente referido, esta proposição não é (sempre) válida, nomeadamente nestas situações em que dividimos o universo em um número infinito não numerável de conjuntos de medida nula, como bem ilustra este paradoxo.

Relembramos, no entanto, que este paradoxo só foi definitivamente resolvido com os fundamentos de Kolmogoroff (1933, p. 50–51), utilizando a sua definição de probabilidade condicionada com recurso ao Teorema de Radon-Nikodym que garante a existência da medida condicionada. Apresentamos, de seguida, uma explicação deste problema semelhante à proposta por Kolmogoroff.

Para simplificar a apresentação vamos recorrer à imagem do planeta como representação da esfera e considerar o primeiro ponto m_1 fixo correspondendo ao pólo norte. Naturalmente, pelo facto de estarmos a actuar numa superfície esférica, vamos trabalhar com coordenadas esféricas que simplificam a exposição. Deste modo teremos as coordenadas esféricas (ρ, θ, φ) onde ρ é o raio da esfera (fixo *a priori* — $\rho = r$), $0 \leq \varphi < \pi$ a longitude que define univocamente o meridiano (arcos de círculo máximo que passam pelos dois pólos), e $-\pi \leq \theta < \pi$ a latitude que define o paralelo (onde o equador corresponde a $\theta = 0$).

Consideremos ainda que a probabilidade de o segundo ponto pertencer

a determinada região $\mathbf{R}$ da superfície da esfera é proporcional à área da região. Será esta hipótese congruente com a probabilidade ser proporcional ao comprimento do arco de círculo máximo que passa pelos dois pontos? Analisemos a função densidade conjunta e a função densidade condicionada a um meridiano fixo ($\varphi = \varphi_0$), isto é, a medida induzida num meridiano sob a hipótese de a probabilidade ser homogênea na superfície esférica. Será essa distribuição também homogênea?

Se a resposta a esta questão for afirmativa, então a função densidade em cada ponto do meridiano (em coordenadas cartesianas) será $g(\theta) = \frac{1}{2\pi r}$ e a probabilidade de o segundo ponto m_2 se situar entre dois paralelos θ_0 e θ_1 (seja região $\mathbf{C}$ onde $-\pi \leq \theta_0 \leq \theta_1 < \pi$) será dada por

$$\mathbb{P}(\theta_0 \leq \theta \leq \theta_1) = \int_{\mathbf{C}} \frac{1}{2\pi r} d\vec{v} = \int_{\theta_0}^{\theta_1} \frac{r}{2\pi r} d\theta = \frac{\theta_1 - \theta_0}{2\pi}. \quad (4.40)$$

Analisemos então se a hipótese de a probabilidade de cada região da superfície da esfera ser proporcional à sua área é compatível com este resultado. Sob esta hipótese teremos, uma vez que a área da superfície esférica é $4\pi r^2$, para a probabilidade de qualquer região $\mathbf{R}$ contida na superfície esférica

$$\mathbb{P}(m_2 \in \mathbf{R}) = \int_{\mathbf{R}} \frac{1}{4\pi r^2} d\mathbf{S} = \frac{\mu_{\mathbf{R}}}{4\pi r^2}, \quad (4.41)$$

onde $\mu_{\mathbf{R}}$ representa a área da região $\mathbf{R}$. Representando por $\mathbf{R}'$ a região $\mathbf{R}$ em coordenadas esféricas teremos, devido a $d\mathbf{S} = \rho^2 |\sin \theta| d\varphi d\theta$,

$$\mathbb{P}(m_2 \in \mathbf{R}) = \int_{\mathbf{R}} \frac{1}{4\pi r^2} d\mathbf{S} = \int_{\mathbf{R}'} \frac{|\sin(\theta)|}{4\pi} d\varphi d\theta \quad (4.42)$$

onde a probabilidade é proporcional à área. Em relação à probabilidade de θ se situar entre θ_0 e θ_1 teremos

$$\mathbb{P}(\theta_0 \leq \theta \leq \theta_1) = \int_{\theta_0}^{\theta_1} \int_0^{\pi} \frac{|\sin \theta|}{4\pi} d\varphi d\theta = \int_{\theta_0}^{\theta_1} \frac{|\sin \theta|}{4} d\theta. \quad (4.43)$$

Para que as duas resoluções sejam congruentes teria que acontecer

$$\frac{\theta_1 - \theta_0}{2\pi} = \int_{\theta_0}^{\theta_1} \frac{|\sin \theta|}{4} d\theta, \quad (4.44)$$

o que (normalmente) não acontece.

Kolmogoroff, após mostrar que a distribuição do segundo ponto, supondo a probabilidade de cada região proporcional à sua área, condicionada a um meridiano (círculo máximo) fixo não é uniforme, escreve o seguinte:

“This shows that the concept of a conditional probability with regard to an isolated given hypothesis whose probability equals 0 is inadmissible. For we can obtain a probability distribution for θ on the meridian circle only if we regard this circle as an element of the decomposition of the entire spherical surface into meridian circles with the given poles.”

[Kolmogoroff 1933, p. 51]

Kolmogoroff afirma que este exemplo demonstra que o conceito de probabilidade condicionada não é aplicável quando o condicionante é um acontecimento de medida nula. Assim, este paradoxo resulta do facto de, para a determinação de probabilidades (ou funções de densidade) condicionadas a acontecimentos de medida nula (tal como na superfície esférica considerar apenas um círculo máximo), estas não deverem ser determinadas substituindo as funções de densidade na fórmula usual de probabilidade condicional, isto é, utilizando

$$f_{X|Y}(x|y) = \frac{f_{(X,Y)}(x,y)}{f_Y(y)}, \quad (4.45)$$

onde $f_{(X,Y)}(x,y)$ é a função densidade conjunta das variáveis aleatórias X e Y , $f_Y(y)$ a função densidade (marginal) da variável aleatória Y e $f_{X|Y}(x|y)$ a função densidade da variável X condicionada a Y . Nesta análise deve ser considerado todo o problema (toda a esfera) e só depois determinar o limite,

dependendo o resultado final da forma como se determina este limite. Daí que seja necessário definir no problema a forma de convergência pretendida para tornar possível a existência de uma só medida condicionada. Por este motivo Bertrand estava correcto quando afirmava que este problema não estava bem definido, isto é, que o enunciado não era suficientemente claro pelo facto de não conter todos os dados do problema.

4.7 Lançamentos em regiões ilimitadas

Pacheco d'Amorim finaliza o seu capítulo da probabilidade contínua definindo o lançamento de um ponto numa região ilimitada. Para o autor lançar um ponto, à sorte, numa região ilimitada $\mathbf{X}$ é o mesmo que lançar à sorte um ponto numa região $\mathbf{X}'$, contida em $\mathbf{X}$, que seja limitada e arbitrariamente grande, isto é, o lançamento de um ponto numa região ilimitada será determinado pelo lançamento de um ponto numa região limitada (contida na primeira), quando esta converge para a região ilimitada, isto é, quando $\mathbf{X}' \rightarrow \mathbf{X}$.

Representemos, ao longo desta secção, por $\varpi_{\mathbf{A}|\mathbf{R}}$ a possibilidade da região $\mathbf{A}$ quando é efectuado um lançamento, à sorte, de um ponto na região $\mathbf{R}$. Deste modo, o autor define a possibilidade⁽¹⁰⁾ de uma região $\mathbf{A}$, quando é lançado um ponto à sorte na região ilimitada $\mathbf{X}$, por

$$\varpi_{\mathbf{A}|\mathbf{X}} = \lim_{\mathbf{X}' \rightarrow \mathbf{X}} \varpi_{\mathbf{A}|\mathbf{X}'} \quad (4.46)$$

⁽¹⁰⁾ Pacheco d'Amorim escreve «probabilidade», contudo, como para o autor, a avaliar pelas definições já apresentadas, a probabilidade de uma região só está definida quando em relação a outra região, consideramos que nesta secção Pacheco d'Amorim está a definir «possibilidade» e não «probabilidade», como escreve. Saliente-se, nesta correcção, que a apresentação desta secção de Pacheco d'Amorim é das menos cuidadas. Vamos expressar aquilo que consideramos estar na mente do autor quando incluiu esta definição.

se este limite existir, isto é, se existir uma região $\mathbf{B}$, verificando $\mathbf{A} \subset \mathbf{B} \subset \mathbf{X}$, tal que

$$\left| \varpi_{\mathbf{A}|\mathbf{X}} - \varpi_{\mathbf{A}|\mathbf{X}'} \right| < \delta, \quad (4.47)$$

para todo o $\delta > 0$ e para todas as regiões $\mathbf{X}'$ tais que $\mathbf{B} \subset \mathbf{X}' \subset \mathbf{X}$.

Refira-se que a apresentação de Pacheco d'Amorim desta definição é pouco clara e, apesar de o autor referir que desenvolverá num *apêndice* esta situação juntamente com a probabilidade dos conjuntos numeráveis, tal não é efectuado⁽¹¹⁾.

Todavia, Pacheco d'Amorim recorre a esta definição para obter a solução do problema da agulha de Buffon, problema que o autor resolve de uma forma bastante apelativa e que será comentado em detalhe na secção 5.1.3.

4.8 Comentário geral ao capítulo

Pacheco d'Amorim fundamenta a probabilidade contínua de forma semelhante à efectuada com a probabilidade discreta, baseando-se em três sistemas de lançamentos. No terceiro sistema de lançamentos concebe os complexos de regiões, definindo lançamentos sequenciais onde o lançamento do n -ésimo ponto é efectuado numa região cuja definição depende dos primeiros $n - 1$ pontos anteriormente lançados, sendo a sua distribuição uniforme quando condicionada aos pontos previamente definidos. Contudo, nestas situações, os pontos pertencentes ao espaço amostra não possuem todos a mesma possibilidade, contrariando a definição geométrica de probabilidade. Desta forma o autor concebe uma definição de probabilidade mais genérica que a usual, pois em tal definição é possível a existência de pontos com possibilidades

⁽¹¹⁾ Recordemos que Pacheco d'Amorim justifica, no PREFÁCIO da sua tese, a ausência deste *apêndice* pela dimensão, já demasiado elevada, da sua tese.

distintas. Saliente-se que uma análise à construção da probabilidade contínua efectuada por Pacheco d'Amorim estará incompleta caso não façamos qualquer referência a uma das suas ideias mais importantes, que é a sua construção do *ponto imagem*, pois um dos paradoxos criados por Bertrand em relação à escolha aleatória de um número em regiões é claramente resolvido com a distinção entre a probabilidade de um ponto lançado directamente à sorte numa região, onde a probabilidade é obrigatoriamente uniforme, e a probabilidade dum outro ponto que seja sua imagem, para a qual Pacheco d'Amorim define a lei de possibilidade.

O autor apresenta duas propriedades que permitem simplificar o cálculo de probabilidades em situações em que exista alguma simetria, ainda que as suas proposições não sejam sempre válidas, pois, em situações onde o espaço amostra é dividido numa infinidade não numerável de conjuntos de medida nula, as propriedades apresentadas não serão, em geral, válidas. Por este motivo o autor, na tentativa de resolução do paradoxo de Borel-Kolmogoroff com o recurso a estas propriedades, não se apercebe de que este paradoxo surge da inexactidão do enunciado e não da existência de qualquer tipo de erro nas distintas formas de resolução apresentadas por Bertrand.

A tentativa de generalização da definição de possibilidade e/ou probabilidade para as situações em que a região é ilimitada parece-nos muito pouco precisa, não nos parecendo que o autor lhe tenha dedicado o relevo devido. De facto, a definição de probabilidade em regiões ilimitadas e em classes infinitas não é de fácil resolução, como ilustram os diversos paradoxos existentes quando trabalhamos em classes ou regiões ilimitadas.

Capítulo 5

Lançamento, à Sorte, de Figuras

Pacheco d'Amorim dedica este capítulo à análise de problemas de Probabilidade Geométrica, criando uma sequência de definições de escolha aleatória de alguns conceitos geométricos, tais como segmentos, rectas, polígonos, entre outros, reduzindo estes novos conceitos a lançamentos aleatórios de pontos em regiões, analisados no capítulo antecedente.

De uma forma geral Pacheco d'Amorim define a escolha aleatória de uma figura em função de lançamentos de pontos em regiões ou complexos de regiões, sendo a cada figura lançada à sorte associado um ponto que lhe serve de definição no lançamento. Este ponto é denominado por Pacheco d'Amorim por *ponto equivalente* da figura em relação ao lançamento considerado e, segundo o autor, não pode depender de nenhum dos pontos da figura lançada. Por este motivo a região na qual é lançado o *ponto equivalente* é fixa, isto é, independente do ponto da figura que determinamos para *ponto equivalente* (que pode ser qualquer ponto da figura).

Pacheco d'Amorim divide este capítulo em duas partes, o lançamento de figuras rígidas e o lançamento de figuras variáveis. Na primeira parte teremos uma figura com a sua forma já definida e iremos lançar essa figura

numa região. Desta forma o autor define uma região, onde será lançado o *ponto equivalente*, na qual cada ponto corresponderá a uma localização possível para a figura ocupar na região onde é lançada. Com estas definições Pacheco d'Amorim propõe resoluções para alguns problemas célebres nesta área, tais como o paradoxo de Bertrand e o problema da agulha de Buffon.

Na segunda parte, referente ao lançamento de figuras variáveis, é definida previamente a forma da figura que depois é lançada numa região como se tratasse de uma figura rígida. Assim, Pacheco d'Amorim define como tirar, à sorte, a forma de polígonos abertos e fechados com um determinado número de lados fixo. Por fim concebe também o lançamento, à sorte, de um polígono com comprimento total fixo e um número de lados a crescer indefinidamente, isto é, um polígono com o número de lados a tender para infinito e, consequentemente, o comprimento de cada lado a tender para zero, que designa por curvas flexíveis e inextensíveis.

5.1 Lançamento, à sorte, de figuras rígidas

Nesta secção vamos expor a sequência de definições apresentada por Pacheco d'Amorim no que se refere ao lançamento aleatório de figuras rígidas, isto é, figuras com a sua forma fixa. Neste contexto iremos inicialmente restringir a exposição às definições com o objectivo de salientar a tentativa do autor em reduzir todos os lançamentos aleatórios de figuras a lançamentos à sorte de pontos em regiões que foram analisados no capítulo antecedente. De seguida comentaremos os exemplos utilizados pelo autor para ilustrar estes novos conceitos referentes a lançamentos, à sorte, de figuras rígidas.

Definição 5.1. Lançar, à sorte, **um sentido num espaço a n dimensões** significa, por definição, o mesmo que lançar, à sorte, um ponto $(x_1, x_2, \dots, x_n)$ na região $\mathbf{E} \subset \mathbb{R}^n$ (lançamento de um ponto numa região

analisado no capítulo 4) determinada pelos pontos que satisfazem

$$(x_1 - x'_1)^2 + (x_2 - x'_2)^2 + \cdots + (x_n - x'_n)^2 = 1, \quad (5.1)$$

sendo o sentido o do vector tendo o ponto $(x'_1, x'_2, \dots, x'_n)$ como origem e o ponto lançado como extremidade.

Definição 5.2. Lançar, à sorte, **um segmento de recta $[ab]$ num outro segmento de recta $[a'b']$** com comprimento maior significa, por definição, lançar à sorte o ponto equivalente na região que este descreve quando $[ab]$ percorre $[a'b']$. Isto é, se fizermos coincidir os pontos a e a' e depois o segmento $[ab]$ percorrer $[a'b']$ até os pontos b e b' coincidirem (de forma a passar por todas as possibilidades que pode ocupar no lançamento), durante este percurso qualquer ponto de $[ab]$ define o mesmo segmento, seja $[a''b'']$. Por este motivo, lançar à sorte $[ab]$ sobre $[a'b']$ (ou $[a'b']$ sobre $[ab]$ que tem o mesmo significado) é equivalente a lançar à sorte um ponto no segmento $[a''b'']$, pois esse ponto (*ponto equivalente*) definirá (de forma unívoca) o local que o segmento $[ab]$ ocupará no segmento $[a'b']$. Pacheco d'Amorim exemplifica a utilização desta definição na resolução do problema do encontro dos dois amigos, que será analisado na secção 5.1.1.

Definição 5.3. Lançar, à sorte, **uma recta numa região $\mathbf{A} \subset \mathbb{R}^n$, com $n \geq 2$** , significa, por definição, lançar, à sorte, um ponto em $\mathbf{A}$ (capítulo 4) e à volta desse ponto tirar, à sorte, uma direcção (através da DEFINIÇÃO 5.1, pois o sentido define a direcção). A recta fica definida por estes dois elementos, o ponto e a direcção. Com base nesta definição o autor irá apresentar uma solução para o paradoxo de Bertrand, que comentaremos na secção 5.1.2.

Definição 5.4. Lançar, à sorte, **um segmento de recta $[ab]$ numa região $\mathbf{A} \subset \mathbb{R}^n$, com $n \geq 2$** , significa, por definição, lançar, à sorte, uma recta na região $\mathbf{A}$ (DEFINIÇÃO 5.3) e, na parte contida em $\mathbf{A}$ dessa recta (que forma um segmento de recta), lançar, à sorte, o segmento $[ab]$ (DEFINIÇÃO 5.2).

Com base nesta definição o autor resolve outro problema célebre na história da Probabilidade, o problema da agulha de Buffon, que será analisado na secção 5.1.3.

Definição 5.5. Lançar, à sorte, **um plano dentro de uma região plana A** significa, por definição, tirar, à sorte, um sentido (DEFINIÇÃO 5.1). Desta forma a região plana A fica orientada em relação ao plano.

Definição 5.6. Lançar, à sorte, **uma região plana A dentro de outra região também plana B** significa, por definição, o mesmo que lançar, à sorte, um plano em B (DEFINIÇÃO 5.5) e, com a orientação obtida, lançar, à sorte, o *ponto equivalente* na região plana C (capítulo 4), constituída pelos pontos que este pode assumir. Isto é, com o lançamento do plano na região B , a região A fica orientada em relação à região B ⁽¹⁾. Escolhendo um qualquer ponto de A (como *ponto equivalente*), quando a região A percorre todas as posições que pode ocupar na região B obedecendo à direcção definida no primeiro lançamento, o ponto percorre uma região que será a região possível para o seu lançamento à sorte (esta região é igual — tem a mesma forma e as mesmas medidas — seja qual for o ponto de A que seja escolhido para *ponto equivalente*). O lançamento à sorte deste ponto definirá uma única posição para a região A ocupar dentro da região B .

Definição 5.7. Lançar, à sorte, **um plano P num espaço E de dimensão n , sendo $n \geq 3$** , significa, por definição, lançar, à sorte, um ponto nesse espaço (capítulo 4) e, à volta desse ponto, dois sentidos (DEFINIÇÃO 5.1).

⁽¹⁾ Notemos que, em alguns exemplos, poderia acontecer que o sentido encontrado originasse uma orientação da região A em relação à região B que não é possível. Por exemplo, se tivermos como região A um rectângulo com 10 centímetros de comprimento e um de largura e como região B um rectângulo com 20 centímetros de comprimento e 3 de largura, existirão sentidos (orientações da região A em relação à região B) que tornam este lançamento impossível.

Estes dois sentidos e o ponto definem o plano dentro do espaço.

Definição 5.8. Lançar, à sorte, **uma figura plana F numa região A definida num espaço E a n dimensões ($n \geq 3$)** significa, por definição, lançar, à sorte, um plano P no espaço E (DEFINIÇÃO 5.7) e na região que resulta da intersecção desse plano P com a região A ⁽²⁾ lançar, à sorte, a figura F (DEFINIÇÃO 5.6).

Definição 5.9. Lançar, à sorte, **um espaço E a três dimensões, numa região A também a três dimensões** significa, por definição, lançar à sorte um ponto em A (conforme capítulo 4) e à volta dele dois sentidos (DEFINIÇÃO 5.1)⁽³⁾.

Pacheco d'Amorim considera que estas definições podem ser facilmente generalizadas para qualquer espaço de dimensão superior.

O autor salienta o facto de a região, onde é lançado o *ponto equivalente*, não poder depender do ponto escolhido. Analisemos esta ideia com um caso concreto, o do lançamento de uma região plana A numa outra região plana B . Primeiro, segundo a definição, lançamos, à sorte, um plano em B que definirá a orientação da região A em relação à região B . Escolhe-se em A o *ponto equivalente*, que pode ser qualquer ponto da região A pois o resultado final será o mesmo, contudo poderá haver pontos que tornem a análise mais acessível. A região onde é efectuado o lançamento do *ponto equivalente* é constituída por todos os pontos que o *ponto equivalente* pode tomar quando a região A passa por todas as localizações possíveis em B com a orientação definida no primeiro lançamento. Se escolhermos outro ponto de A para

⁽²⁾ No caso de existir, pois nada nos garante que o plano escolhido intersecte a região A . A apresentação de Pacheco d'Amorim omite, no entanto, esta possibilidade.

⁽³⁾ Os dois sentidos definidos são suficientes para orientar três eixos.

ponto equivalente, a região obtida será igual (tendo a mesma forma, apenas sofrendo uma translação para outra posição).

Por exemplo, consideremos que a região **A** é um triângulo equilátero cujos lados têm comprimento igual a 4 centímetros e que a região **B** é um rectângulo com comprimento 10 centímetros e altura 5 centímetros (conforme primeiros dois gráficos da Figura 5.1). Suponhamos que no lançamento de um plano na região **B** sai o correspondente a um ângulo de 45 graus, isto é, a figura **A** ficará orientada em relação a **B** através de uma rotação de 45 graus (consultar terceiro gráfico da Figura 5.1). Escolhamos para *ponto equivalente* da região **A** o vértice inferior (ponto $(-\sqrt{2}, -\sqrt{2})$) e façamos a região **A** percorrer todas as possibilidades que pode ocupar dentro da região **B**. Ao longo deste processo os pontos por onde o *ponto equivalente* passou formam uma região **C** (rectângulo representado dentro do rectângulo **B** no quarto gráfico da Figura 5.1) onde será lançado, à sorte, o *ponto equivalente*. Notemos que a cada ponto da região **C** corresponde uma única posição do triângulo **A**, bem como a cada posição possível do triângulo (condicionada à direcção determinada pela sorte) corresponde um único ponto da região **C**. Se optássemos por outro ponto para *ponto equivalente*, a região onde esse ponto seria escolhido à sorte teria a mesma forma que a região **C**, ainda que deslocada para outra posição (através da mesma translação que permite passar do *ponto equivalente* previamente analisado para esse novo *ponto equivalente*).

Pacheco d'Amorim refere que, se dividirmos a região **A** em partes **A**₁ e **A**₂, tais que $\mathbf{A} = \mathbf{A}_1 \cup \mathbf{A}_2$ e $\mathbf{A}_1 \cap \mathbf{A}_2 = \emptyset$, o campo de variação (região possível) do ponto equivalente do lançamento à sorte de **A** em **B** é (em geral) distinto do campo de variação do *ponto equivalente* do lançamento à sorte de **A**₁ em **B**, pois esta última região terá normalmente uma medida superior, porque o seu *ponto equivalente* tem uma região de lançamento que inclui a

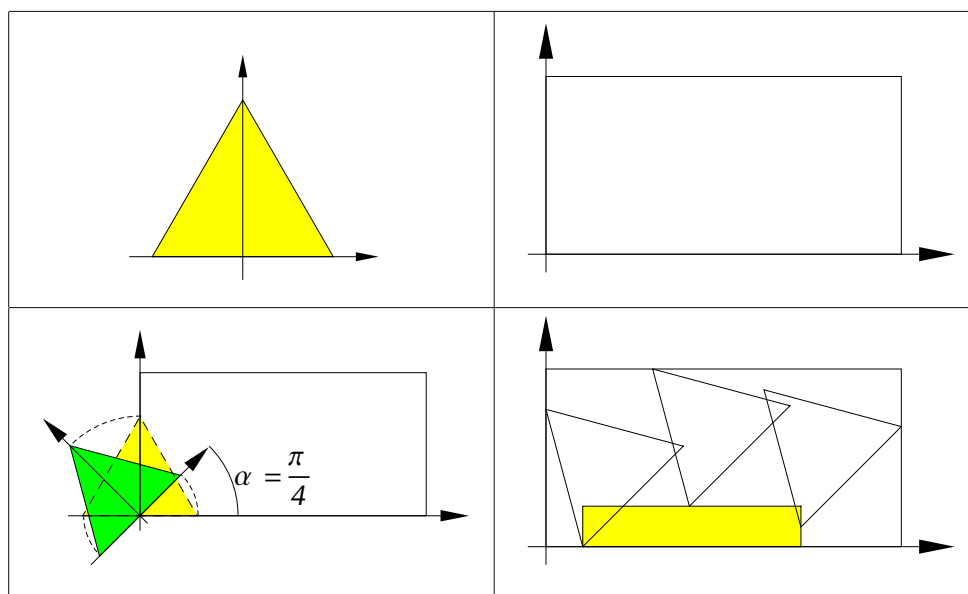


Figura 5.1: Lançar, à sorte, a região plana **A** na região plana **B**

região da primeira situação onde é lançada toda a região **A**.

No entanto, se o campo de variação for ilimitado, como no exemplo da agulha de Buffon que analisaremos na secção 5.1.3, quer seja lançada toda a agulha quer seja lançada apenas uma parte desta, o campo de variação do *ponto equivalente* será sempre o mesmo pelo facto de não haver restrições na definição da região onde é lançado o *ponto equivalente* (nem derivadas da forma da agulha, que pode ser qualquer uma, nem do sentido que esta tem em relação à região onde é lançada) pelo facto de a região, onde é lançada a figura, ser ilimitada.

5.1.1 Problema do encontro dos dois amigos

Problema 5.1 (Problema do encontro dos dois amigos). Dois amigos passeiam todas as tardes, durante meia hora, num jardim público que está aberto das duas horas até às quatro. Qual a probabilidade de que se encontrem em certo dia?

SOLUÇÃO: Supondo casual (e independente) a hora a que qualquer um dos dois amigos chega ao jardim (e que este é um local tão pequeno que é impossível os dois amigos estarem a passear ao mesmo tempo e não se encontrarem), o problema pode ser generalizado através do enunciado que a seguir se expõe: Considere-se um segmento de recta **A** de comprimento α (correspondente ao tempo em que o jardim está aberto) e mais dois outros (**B** e **C**) de comprimentos β e γ , respectivamente (que representam o tempo que cada um dos amigos passeia no jardim); lançam-se os segmentos **B** e **C**, à sorte, sobre o segmento **A**: qual a probabilidade de que estes segmentos se sobreponham?

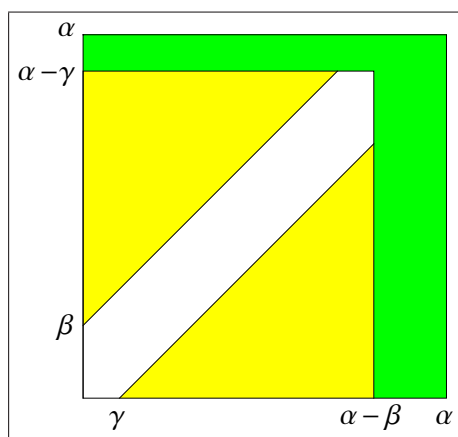


Figura 5.2: Problema do encontro de dois amigos

No caso de $\beta + \gamma \geq \alpha$ os segmentos sobrepõem-se sempre e por isso $\mathbb{P} = 1$.

Analisemos, então, o caso em que $\beta + \gamma < \alpha$. Quando o segmento **B** percorre o segmento **A**, qualquer dos seus pontos descreve um segmento com comprimento $\alpha - \beta$ (seja segmento **B'**) e, de forma análoga, qualquer ponto de **C** descreve um segmento de comprimento $\alpha - \gamma$ (seja segmento **C'**). Lançar, à sorte, os dois segmentos sobre o segmento **A** (DEFINIÇÃO 5.2), é o mesmo que lançar à sorte um ponto em **B'** e um ponto em **C'** que, pela definições apresentadas no capítulo referente à probabilidade contínua, é o mesmo que

lançar, à sorte, um ponto (x, y) no rectângulo $\mathbf{B}' \times \mathbf{C}'$ (região composta). Escolhamos então para pontos equivalentes de $\mathbf{B}$ e $\mathbf{C}$ as suas extremidades que representam o momento em que chegam ao jardim⁽⁴⁾ e sejam x e y as distâncias dessas extremidades à origem de $\mathbf{A}$.

Para que $\mathbf{B}$ e $\mathbf{C}$ se não sobreponham terá de ser $y - x > \beta$ (no caso de $y > x$) ou $x - y > \gamma$ (no caso de $y < x$). As rectas cujas equações são $y - x = \beta$ e $x - y = \gamma$ determinam sobre o rectângulo $\mathbf{B}' \times \mathbf{C}'$ dois semiquadrados que são, para o problema proposto, a região contrária. Assim, a possibilidade pretendida é dada por

$$\begin{aligned} 1 - \mathbb{P} &= \frac{0.5(\alpha - \beta - \gamma)^2 + 0.5(\alpha - \gamma - \beta)^2}{(\alpha - \beta)(\alpha - \gamma)} = \\ &= \frac{(\alpha - \beta - \gamma)^2}{(\alpha - \beta)(\alpha - \gamma)}. \end{aligned}$$

No problema particular enunciado temos $\alpha = 2$ e $\beta = \gamma = 0.5$, sendo consequentemente a probabilidade igual a $\mathbb{P} = 1 - \frac{4}{9} = \frac{5}{9}$. $\diamond$

Refira-se que Czuber (1884, p. 45), numa obra integralmente dedicada à análise de problemas em probabilidade geométrica, resolve um problema semelhante, se bem que, na sua resolução, os amigos podem chegar a qualquer momento, isto é, utilizando os dados do enunciado apresentado por Pacheco d'Amorim, os amigos possam chegar ao jardim quando faltam apenas 5 minutos para este fechar não cumprindo os 30 minutos de passeio. Desta forma, para Czuber, a região possível é todo o quadrado $\mathbf{A} \times \mathbf{A}$ não impondo, tal como Pacheco d'Amorim, que a região possível se restrinja ao rectângulo $\mathbf{B}' \times \mathbf{C}' \subset \mathbf{A} \times \mathbf{A}$.

⁽⁴⁾ Como foi previamente referido, o ponto equivalente pode ser qualquer um. Por este motivo poderíamos ter optado por quaisquer outros pontos, mas pensamos que estes tornam a análise do problema mais perceptível.

5.1.2 Paradoxo de Bertrand

Problema 5.2 (Problema de Bertrand). Lança-se, à sorte, uma recta dentro de um círculo de raio r . Qual a probabilidade de que ela tenha um comprimento inferior a uma determinada corda de comprimento c ?

SOLUÇÃO: Lançar, à sorte, uma recta numa região é, segundo a DEFINIÇÃO 5.3 de Pacheco d'Amorim, o mesmo que lançar, à sorte, um ponto na região e tirar, à sorte, uma direcção. O autor começa por justificar que, devido à simetria do círculo, a possibilidade da região favorável e da região possível é sempre a mesma, seja qual for a direcção utilizada, isto é, para qualquer direcção a probabilidade pretendida será a mesma. Por esta razão a probabilidade pretendida é independente da direcção e, pela Proposição VIII (consultar página 254), pode-se considerar uma qualquer direcção fixa. Seja então considerada a direcção do eixo das ordenadas – OY .

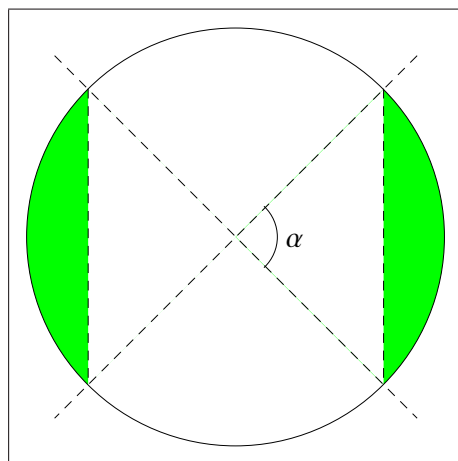


Figura 5.3: Problema de Bertrand – Solução de Pacheco d'Amorim

Desta forma o problema é reduzido a um lançamento de um ponto numa região. Como tal, todos os pontos têm a mesma possibilidade, sendo suficiente, para determinar a probabilidade, calcular o rácio entre a área favorável e possível. Assim, a região possível $\mathbf{P}$ será toda a área do círculo, sendo a

sua área πr^2 . A região favorável **F** será a região onde as cordas, com a direção estabelecida, possuem um comprimento menor que c , que corresponde à região representada a sombreado na Figura 5.3. Representando por α o ângulo subentendido a uma qualquer corda de comprimento c (definido pelos segmentos que vão dos extremos da corda ao centro do círculo), temos que a área da região **F** é dada por

$$2 \left[\frac{\alpha}{2\pi} \pi r^2 - r^2 \sin \frac{\alpha}{2} \cos \frac{\alpha}{2} \right] = 2r^2 \left[\frac{\alpha}{2} - \sin \frac{\alpha}{2} \cos \frac{\alpha}{2} \right] \quad (5.2)$$

e, conseqüentemente, a probabilidade pretendida é

$$\mathbb{P} = 2r^2 \frac{\frac{\alpha}{2} - \sin \frac{\alpha}{2} \cos \frac{\alpha}{2}}{r^2 \pi} = \frac{\alpha - \sin \alpha}{\pi}. \quad (5.3)$$

◇

Este problema corresponde a uma versão mais geral do problema proposto por Bertrand (1888, p. 4–5), que como segmento considera os lados do triângulo equilátero inscrito no círculo, logo, $c = \sqrt{3}r$ e $\alpha = \frac{2}{3}\pi$, obtendo-se, desta forma, uma probabilidade aproximadamente igual a $\mathbb{P} \approx 0.391002$. Refira-se que Bertrand, no enunciado do problema, questiona a probabilidade de o comprimento da corda ser inferior ao dos lados do triângulo, mas nas três resoluções que apresenta fornece como resultado a probabilidade complementar. Por este motivo enunciaremos o problema de Bertrand de acordo com as resoluções que ele apresenta e, desta forma, as probabilidades por ele obtidas devem ser confrontadas com o valor complementar da probabilidade obtida por Pacheco d'Amorim, isto é, com o valor aproximado $1 - 0.391 = 0.609$.

Problema 5.3 (Problema original de Bertrand). Qual a probabilidade de uma corda, escolhida ao acaso num círculo de raio r , ter um comprimento superior ao comprimento dos lados do triângulo equilátero inscrito nesse círculo?

SOLUÇÃO: Bertrand apresenta três resoluções distintas deste problema.

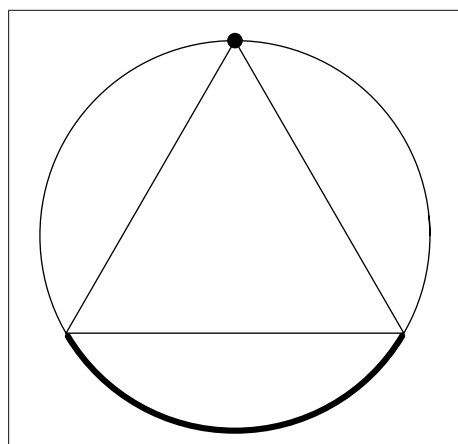


Figura 5.4: Problema de Bertrand – 1ª Solução de Bertrand

Em primeiro lugar, Bertrand considera que uma das extremidades da corda é fixa, pois, devido à simetria do círculo, qualquer que seja este ponto (da circunferência) obteremos sempre a mesma probabilidade. Deste modo, para determinarmos a corda, basta-nos considerar este ponto como sendo a sua origem e estabelecermos o sentido da corda. Se utilizarmos este ponto como um dos vértices do triângulo equilátero (por exemplo, como sendo o vértice superior do triângulo presente na Figura 5.4), então os lados do triângulo e a recta tangente à circunferência nesse ponto dividem o campo possível em três ângulos de 60 graus (os restantes 180 graus não são possíveis⁽⁵⁾), sendo apenas uma destas possibilidades favorável ao acontecimento analisado. Desta forma obtém-se a probabilidade $\mathbb{P} = \frac{1}{3}$. Saliente-se que, nesta análise, existe uma relação biunívoca entre o ângulo e cada corda, podendo, por este motivo, a escolha da corda ser considerada aleatória (à sorte). Esta solução, supondo fixa uma das extremidades da corda, é, pelo menos na actualidade, mais frequentemente descrita utilizando a determinação ale-

⁽⁵⁾ Poderíamos pensar em termos de direcção em vez de sentido, uma vez que numa corda normalmente não se define a sua origem e o seu fim. Neste caso todas as direcções seriam possíveis, contudo o raciocínio seria análogo e o resultado final o mesmo.

atória da outra extremidade. Assim, nesta descrição, tendo em conta que uma extremidade é conhecida, escolhemos aleatoriamente na circunferência a outra extremidade. Para facilitar, podemos fazer corresponder a extremidade fixa a um dos vértices do triângulo equilátero inscrito (podemos utilizar uma vez mais o vértice superior do triângulo representado na Figura 5.4), verificando-se facilmente que o triângulo divide a circunferência em três partes iguais, das quais apenas uma é favorável à realização do acontecimento pretendido, obtendo-se a mesma probabilidade $\mathbb{P} = \frac{1}{3}$. Notemos que, também nesta forma de apresentação, a cada ponto da circunferência corresponde uma só corda e a cada corda um só ponto da circunferência, logo, todas as cordas são *equipossíveis*.

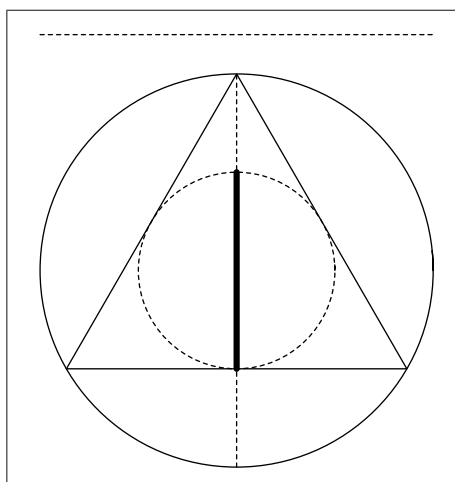


Figura 5.5: Problema de Bertrand – 2ª Solução de Bertrand

Na segunda resolução, utilizando o mesmo raciocínio da simetria da circunferência, Bertrand considera que a probabilidade é independente da direcção da corda, isto é, seja qual for a direcção da corda que utilizemos para determinar a probabilidade vamos sempre obter o mesmo valor. Assim, nesta resolução, supondo que a direcção da corda é fixa (conforme representação na Figura 5.5 onde consideramos uma corda horizontal), como todas estas cordas obrigatoriamente se irão intersectar com o diâmetro do círculo que

é perpendicular à sua direcção, bastará determinar, nesse diâmetro, o rácio entre os comprimentos da região favorável (que corresponde à metade central do diâmetro) e da região possível obtendo-se, desta forma, o valor $\mathbb{P} = \frac{1}{2}$. Saliente-se que, neste raciocínio utilizado por Bertrand, é aceitável a escolha aleatória de uma corda uma vez que a cada ponto do diâmetro perpendicular está associada uma e uma só corda e a cada corda (com a direcção preestabelecida) está associado um e um só ponto do diâmetro. Logo, utilizando a nomenclatura de Pacheco d'Amorim, todas as cordas possuem a mesma possibilidade de serem escolhidas.

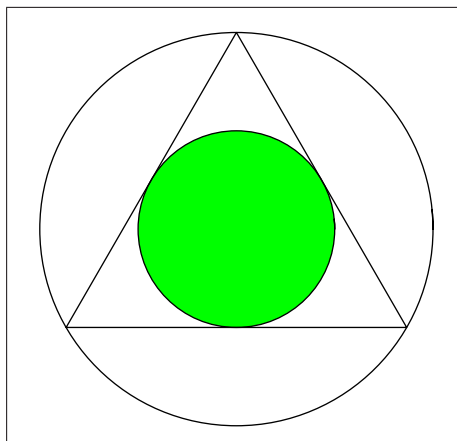


Figura 5.6: Problema de Bertrand – 3ª Solução de Bertrand

Na terceira resolução Bertrand considera que escolher a corda à sorte é o mesmo que escolher aleatoriamente o seu ponto central, isto é, escolher um ponto do círculo que lhe corresponderá ao seu ponto central, pois, desta forma, a cada ponto está associada uma única corda⁽⁶⁾ e a cada corda está associado um único ponto. Deste modo, para determinarmos a probabilidade devemos calcular o quociente entre as áreas da região favorável e da possível.

⁽⁶⁾ Com excepção do centro do círculo que tem associadas infinitas cordas, que correspondem aos diâmetros. Contudo, nesta situação, não só todos esses diâmetros têm o mesmo comprimento como o ponto, centro da circunferência, é um conjunto singular e, como tal, de medida nula.

Uma vez que a região favorável corresponde ao círculo inscrito no triângulo, conforme Figura 5.6, que tem raio $\frac{r}{2}$, a probabilidade pretendida é dada por $\mathbb{P} = \frac{\pi(0.5r)^2}{\pi r^2} = \frac{1}{4}$. $\diamond$

Realcemos, uma vez mais, que nas três soluções de Bertrand a cada ponto escolhido temos associada uma e uma só corda, com uma única situação excepcional na terceira resolução, onde num ponto (centro do círculo) temos associadas infinitas cordas (neste caso diâmetros), mas todas elas com o mesmo comprimento. Na solução de Pacheco d'Amorim isto não acontece. Uma vez que a direcção da corda é fixa *a priori* e depois é escolhido um ponto no círculo, a cada corda é associada uma região favorável que é um segmento de recta (com a direcção definida), não havendo uma relação biunívoca entre o ponto escolhido e a corda. Deste modo, a cada corda estão associados infinitos pontos (que constituem um segmento de recta que é a própria corda) e, consequentemente, as regiões favoráveis das cordas não são igualmente possíveis, tendo em conta que a cada corda é associado um segmento de recta e estes não têm o mesmo comprimento. Assim, nesta resolução consideram-se cordas mais prováveis (no sentido de possuírem uma maior densidade, pois na realidade todas elas têm probabilidade nula) do que outras, sendo a densidade de cada corda proporcional ao seu comprimento. Por exemplo, a corda com a direcção definida que passa pelo centro do círculo (corda $\mathbf{C}_1$ na Figura 5.7) terá um segmento de recta associado com comprimento igual ao diâmetro do círculo, enquanto qualquer outra corda (por exemplo $\mathbf{C}_2$) terá associada uma região favorável (segmento) com menor comprimento. Desta forma as cordas que passam perto do centro do círculo terão maior possibilidade que aquelas que se afastam do centro, fazendo, naturalmente, com que a probabilidade de que a corda escolhida tenha um comprimento grande seja mais elevada. Recordamos que, para o enunciado do problema apresentado por Bertrand, a resolução de Pacheco d'Amorim fornece uma probabilidade

Bertrand na segunda solução, quando fixa um diâmetro, também parte do mesmo princípio — seja qual for a direcção da recta a probabilidade será a mesma, logo, podemos considerar uma direcção fixa — mas escolhe um ponto no diâmetro perpendicular à direcção da corda, fazendo com que todas as rectas (cordas) tenham igual possibilidade (na terminologia de Pacheco d’Amorim). Assim, nas soluções de Bertrand pode-se considerar uma escolha aleatória de uma corda, pois em qualquer uma as cordas têm a mesma possibilidade de saírem, enquanto na solução de Pacheco d’Amorim isto não acontece, pois a possibilidade de cada corda é proporcional ao seu comprimento, sendo mais provável serem escolhidas cordas de comprimento mais elevado. Desta forma não podemos considerar que as cordas sejam escolhidas à sorte na resolução de Pacheco d’Amorim. Eis porque razão consideramos a sua proposta de resolução incorrecta.

Uma vez que Pacheco d’Amorim na resolução deste problema recorre à Proposição VIII do capítulo 4 que, como destacamos na análise efectuada.

do paradoxo de Borel, não é sempre válida, decidimos realizar uma pequena simulação para analisar qual seria o resultado obtido para o mesmo problema, utilizando o enunciado original apresentado por Bertrand restrito a $c = \sqrt{3}r$ (comprimento dos lados do triângulo equilátero inscrito no círculo) com $r = 1$ (para simplificar a apresentação, pois a probabilidade pretendida é independente do raio do círculo considerado) e sem recorrer a esta Proposição. Daqui se depreende que utilizamos directamente a DEFINIÇÃO 5.3 referente ao lançamento, à sorte, de uma recta numa região **A**. Para este fim recorreremos à simulação de Monte Carlo utilizando o software *Mathematica* 5.1. Assim, começamos por simular o lançamento à sorte de um ponto no círculo (com raio unitário) através da simulação de dois números (independentes) com distribuição uniforme no intervalo $[-1, 1]$ e utilizando apenas os pontos que *caíam* no círculo, isto é, usando uma distribuição uniforme bivariada no quadrado $[-1, 1] \times [-1, 1]$ e truncando a distribuição para os pontos que não nos interessam, de forma que os pontos $(x, y) \in \mathbb{R}^2$ tais que $x^2 + y^2 > 1$ sejam omitidos e obtenhamos uma distribuição uniforme para os pontos $(x, y) \in \mathbb{R}^2$ dentro do círculo unitário ($x^2 + y^2 \leq 1$). A título de exemplo apresentamos na Figura 5.8 o resultados de três simulações de, respectivamente, 1000, 5000 e 10000 pontos nestas condições.

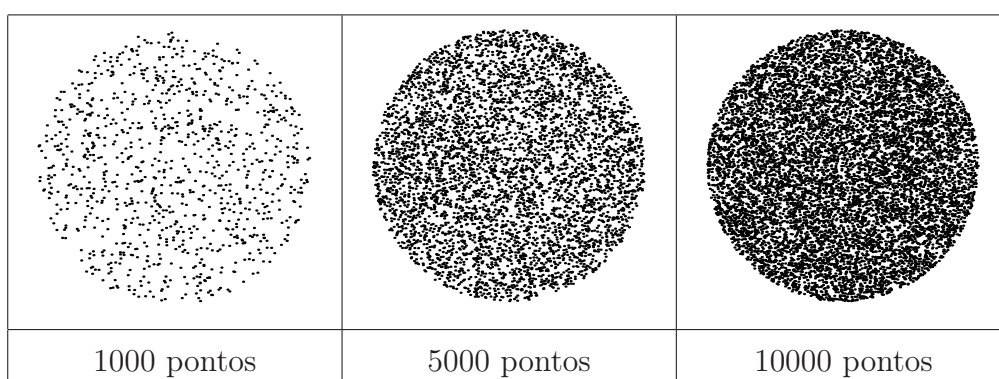


Figura 5.8: Lançamento, à sorte, de pontos num círculo

Após esta operação, conforme a DEFINIÇÃO 5.3, foi escolhido à sorte

um sentido, isto é, foi simulado um número com distribuição uniforme no intervalo $[0, 2\pi)$ que representa o ângulo α com o eixo das abcissas. Notemos que este processo é idêntico ao presente na DEFINIÇÃO 5.1, referente à escolha aleatória de um sentido, pois, se representarmos uma circunferência unitária à volta do ponto, esta terá perímetro 2π e, ao escolhermos um ponto à sorte na circunferência, estaremos de igual forma a lançar um ponto à sorte no intervalo $[0, 2\pi)$, correspondendo o ponto $\alpha \in [0, 2\pi)$ da circunferência a um ângulo α entre o segmento escolhido e o eixo horizontal. Desta forma, com o ponto (x, y) e o sentido α determinados, temos definida uma única corda; bastará então determinar o seu comprimento e averiguar se este é superior a $\sqrt{3}$ (comprimento dos lados do triângulo equilátero inscrito no círculo de raio unitário). Apresentamos na Figura 5.9 dois exemplos, um com 30 cordas e outro com 100 cordas escolhidas aleatoriamente, onde estão também representados os pontos que deram origem às cordas.

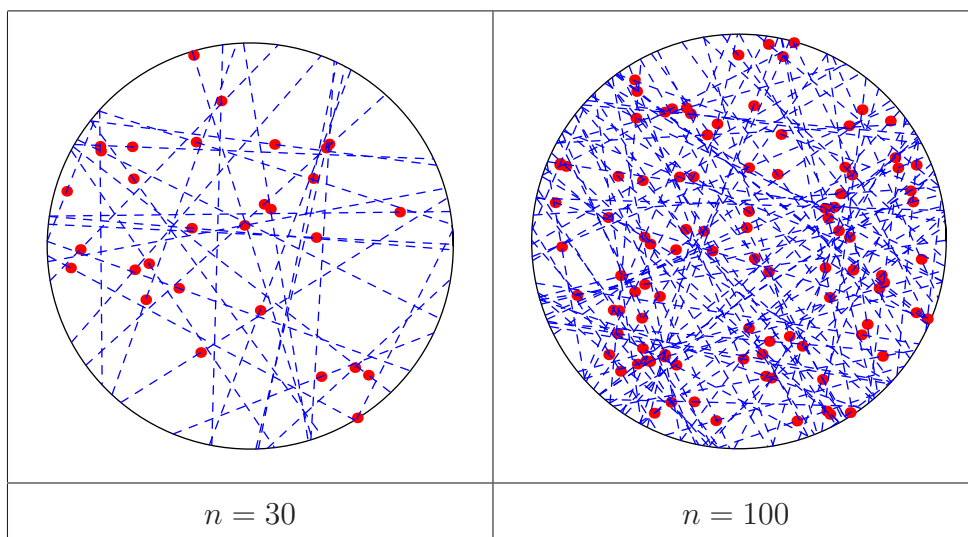


Figura 5.9: Cordas escolhidas aleatoriamente

Cada corda simulada por este método corresponderá a uma prova de Bernoulli com probabilidade de sucesso p que corresponde à probabilidade de uma corda, escolhida aleatoriamente pelo processo descrito, ter um com-

primento superior ao dos lados do triângulo equilátero inscrito no círculo, probabilidade esta que pretendemos averiguar se assume o valor obtido por Pacheco d'Amorim analiticamente. Nas três séries de simulações efectua- das foram utilizadas r réplicas de amostras de dimensão n , onde registamos o proporção amostral observada ($\hat{p}_{\text{obs.}}$), isto é, a proporção de cordas com comprimento superior a $\sqrt{3}$ em cada amostra.

Dimensão n	1000	10000	5000
Réplicas r	10000	1000	5000
Média	0.607899	0.608079	0.608124
Desvio Padrão	0.0152866	0.00481569	0.00686065
Mínimo	0.55	0.5917	0.5818
Quartil 1	0.598	0.6048	0.6034
Quartil 2	0.608	0.6082	0.6082
Quartil 3	0.618	0.611	0.6128
Máximo	0.666	0.6221	0.6310
I. C. (1)	95.25	95.2	95.14
I. C. (2)	94.93	95.3	95.32

Tabela 5.1: Simulação do Problema de Bertrand

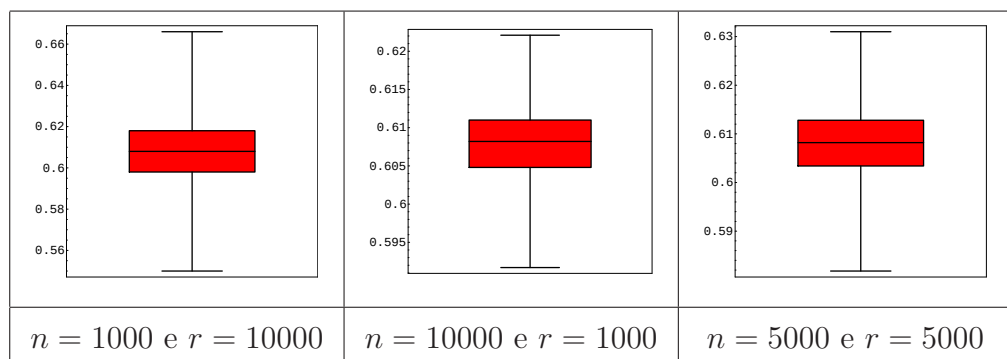


Figura 5.10: Simulação do Problema de Bertrand

Apresentamos na Tabela 5.1 as principais características dos valores obti-

dos pela simulação, no conjunto das r amostras observadas, para a proporção amostral, tais como a média, o desvio padrão, os extremos e os quartis e representamos o diagrama de extremos e quartis na Figura 5.10. Construímos, com base no valor da proporção amostral observado em cada amostra, dois intervalos com noventa e cinco por cento de confiança para determinarmos em cada série a percentagem de intervalos que contêm o valor obtido por Pacheco d'Amorim. Uma vez que, como a dimensão utilizada para as amostras é elevada (a amostra mínima tem dimensão 1000) e a probabilidade não está próxima dos extremos (nem perto de zero nem da unidade), podemos considerar, como consequência do Teorema Limite Central, uma distribuição aproximadamente gaussiana para cada proporção amostral

$$\hat{p} \overset{\circ}{\sim} \text{Gaussiana} \left(p, \sqrt{\frac{p(1-p)}{n}} \right). \quad (5.4)$$

Com esta estatística podemos construir dois intervalos de confiança. O primeiro é obtido pelo método da variável fulcral, tendo o intervalo de confiança a seguinte forma

$$\left(\frac{\hat{p} + \frac{z^2}{2n} - z\sqrt{\frac{\hat{p}(1-\hat{p}) + \frac{z^2}{4n}}{n}}}{1 + \frac{z^2}{n}}, \frac{\hat{p} + \frac{z^2}{2n} + z\sqrt{\frac{\hat{p}(1-\hat{p}) + \frac{z^2}{4n}}{n}}}{1 + \frac{z^2}{n}} \right) \quad (5.5)$$

onde z representa o percentil 97,5 da distribuição gaussiana standard. O segundo intervalo de confiança utilizado é o mais usual, sendo obtido também pelo método da variável fulcral, mas recorrendo a algumas simplificações (substituindo o parâmetro p , que aparece na fórmula (5.4) no desvio padrão da distribuição gaussiana de $\hat{p}$, por um seu estimador $\hat{p}$). É denominado por intervalo de confiança de Wald e tem a seguinte forma

$$\left(\hat{p} - z\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, \hat{p} + z\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \right). \quad (5.6)$$

Ainda que não efectuemos uma análise exaustiva aos resultados da simulação, poderemos salientar que estes são congruentes com o valor obtido

analiticamente por Pacheco d'Amorim, pois, apesar de a média (e também a mediana) das proporções amostrais obtidas em qualquer uma das três séries de simulações ser inferior ao valor obtido analiticamente por Pacheco d'Amorim (aproximadamente 0.608998), denota-se que as percentagens de intervalos de confiança que contêm este valor são muito próximas das previstas, sendo os valores obtidos pelo intervalo descrito na fórmula (5.5) sempre superiores a noventa e cinco por cento e os obtidos pelos intervalos de confiança de Wald também muito próximos do valor teórico.

Consideramos, pelas razões previamente descritas, que os valores resultantes destas três séries de simulações são compatíveis com o resultado obtido por Pacheco d'Amorim. Assim, na determinação da solução deste problema, não nos parece haver qualquer incongruência na utilização da Proposição VIII, como, aliás, tínhamos verificado na sua aplicação na resolução do paradoxo de Borel. Como tal, o problema da solução apresentada por Pacheco d'Amorim não está no recurso a esta proposição, mas, antes, na própria definição de lançamentos de rectas em regiões — DEFINIÇÃO 5.3. Utilizando unicamente esta definição, as rectas possíveis não possuem todas a mesma possibilidade, como facilmente se verifica na simulação deste problema. Por exemplo, as cordas C_1 e C_2 presentes na Figura 5.11, escolhidas à sorte no círculo unitário, não têm a mesma possibilidade, pois, apesar de as direcções das rectas serem equipossíveis, para que seja escolhida a corda C_1 teremos os pontos do segmento $[ab]$ como possíveis, segmento este que tem comprimento igual a 0.71, enquanto para ser escolhida a corda C_2 teremos como possíveis os pontos do segmento $[cd]$, que tem um comprimento igual a 2. Por consequência, o segmento C_2 terá maior possibilidade de ser escolhido que o segmento C_1 .

Desta forma, a DEFINIÇÃO 5.3 fornece uma maior possibilidade às rectas cuja intersecção com a região A , região onde as rectas são lançadas, resulta

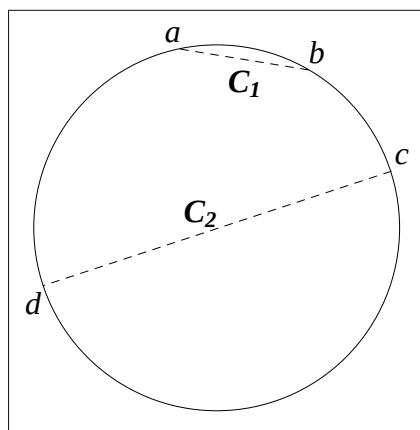


Figura 5.11: Duas cordas com possibilidades distintas

um segmento com maior comprimento. Assim, não podemos estar de acordo com esta definição (e consequentemente com a DEFINIÇÃO 5.4 que desta depende), pois as rectas obtidas não são equipossíveis, contrariando, deste modo, a ideia intuitiva de escolha à sorte de uma recta. No que se refere ao paradoxo de Bertrand, aqui analisado, a definição favorece as cordas com maior comprimento, sobrevalorizando a probabilidade pretendida.

Borel (1909, p. 110–111) e (1914, p. 83–85) expõe este problema apresentando as três resoluções de Bertrand e, como foi anteriormente referido na análise do paradoxo de Borel-Kolmogoroff⁽⁷⁾, fundamenta a sua análise à validade de cada uma das soluções com a verificação experimental do valor obtido. Borel, seguindo este raciocínio, não considera que exista qualquer probabilidade, das obtidas pelas resoluções de Bertrand, que seja errada, mas que todas elas são válidas de acordo com um determinado modo de escolha à sorte da corda, isto é, se vamos escolher “*au hasard*” as suas extremidades, se vamos considerar uma direcção qualquer fixa e escolher “*au hasard*” o seu

⁽⁷⁾ Para Borel só são válidos os problemas geométricos concretos, isto é, problemas cujo enunciado é suficientemente preciso para que seja possível deduzir um método de verificação experimental dos resultados.

ponto intermédio ou se vamos escolher “*au hasard*” um ponto do círculo que será o seu ponto médio. Borel, tal como Bertrand, considera que o problema, para ter uma única solução, deveria conter o modo de acordo com o qual será efectuada a verificação experimental, isto é, o método pelo qual se vai traçar a corda à sorte no círculo. Contudo, Borel, nomeadamente em *Le Hasard* (apesar de grande parte desta obra ser uma reprodução dos *Éléments de la Théorie des Probabilités*), defende a segunda solução (a que dá resultado $\mathbb{P} = \frac{1}{2}$), pelo facto de considerar que a maior parte dos métodos experimentais fornecer esta probabilidade. Outra razão apontada por Borel para defender esta solução é o facto de considerar que os três métodos de resolução do problema de Bertrand também podem ser aplicados para a obtenção da solução do problema da Agulha de Buffon (Borel, 1909, p. 111–113). Borel considera, assim, que poderíamos igualmente obter três soluções distintas para o problema da agulha de Buffon, mas que, por a solução deste problema não ser posta em causa e ser confirmada pela experiência⁽⁸⁾, devemos utilizar a mesma metodologia para o problema de Bertrand, pela qual se obtém $\mathbb{P} = \frac{1}{2}$.

Bachelier (1912, p. 403–411), num pequeno capítulo que dedica aos problemas geométricos, utiliza a definição clássica de probabilidade, rácio entre o número de casos favoráveis e o número de casos possíveis, entendendo todos os casos como igualmente possíveis, e salienta que é um dado essencial no enunciado de qualquer problema o modo como os casos poderão ser divididos em casos igualmente possíveis. No que se refere a este problema, Bachelier considera que o enunciado não é suficientemente claro para podermos dividir em casos igualmente possíveis, isto é, para que só possa ser considerada uma

⁽⁸⁾ Borel não menciona nenhuma experiência particular, mas provavelmente refere-se a experiências semelhantes à efectuada pelo matemático italiano Lazzerini em 1901, que realizou 34080 lançamentos da agulha para *estimar* o valor de π em $\frac{355}{113} = 3.1415929$, que está correcto até à sexta casa decimal.

forma de divisão.

De facto, o enunciado deste problema, tal como muitos outros em probabilidade contínua, não é suficientemente preciso, permitindo várias soluções correspondentes a diferentes formas de escolha aleatória. Por este motivo o “*paradoxo*” surge da inexactidão do enunciado e não de qualquer incorrecção nas resoluções dispare apresentadas por Bertrand. Mesmo que recorramos à axiomática de Kolmogoroff (1933) para a resolução do problema, o enunciado do problema não é suficientemente claro, pois com ele não é possível definir de uma só forma o espaço de probabilidade. Deste modo, mesmo que tenhamos igual espaço amostra e/ou espaço de acontecimentos, o enunciado permite associar várias medidas de probabilidade a estes espaços e, conseqüentemente, poderemos associar diferentes probabilidades aos mesmos acontecimentos. Assim, o enunciado não define a medida de probabilidade que deve ser utilizada na sua resolução, correspondendo a cada uma das três soluções apresentadas por Bertrand um espaço de probabilidade distinto, compatível com todos os dados do problema.

Rényi (1970, p. 64–65) salienta que cada resultado corresponde a uma interpretação distinta da escolha aleatória da corda, isto é, a cada resultado corresponde uma experiência distinta, sendo, por este motivo, associada uma medida de probabilidade diferente aos mesmos acontecimentos. Refira-se, uma vez mais, que é o enunciado que é omissivo em relação à forma que a escolha, à sorte, da corda deve ser efectuada.

Parece-nos que Pacheco d’Amorim, ao contrário de diversos autores seus contemporâneos, não se apercebe que o equívoco surge da imprecisão do enunciado. Por este motivo, apesar de não referir explicitamente às resoluções de Bertrand, Pacheco d’Amorim claramente considera que a única solução válida deste problema é a sua. Pensamos, por isso, que Pacheco d’Amorim concebe esta sequência de definições, apresentadas ao longo deste

capítulo, com o intuito de torneiar os diversos paradoxos existentes neste campo da Teoria da Probabilidade, de forma a que cada problema permita uma única solução. Contudo, como mencionámos, consideramos que a solução encontrada para o paradoxo de Bertrand por Pacheco d'Amorim não é aceitável, dado não haver uma relação biunívoca entre cada ponto determinado pela sorte e cada corda, isto é, por não ser razoável que as cordas estejam a ser escolhidas à sorte quando possuem possibilidades distintas.

5.1.3 Problema da Agulha de Buffon

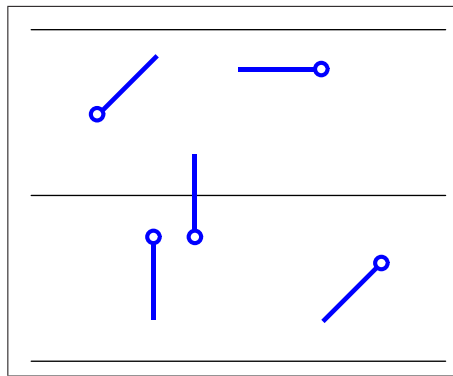


Figura 5.12: Problema da Agulha de Buffon

Problema 5.4 (Problema da Agulha de Buffon). Lança-se, à sorte, uma agulha (segmento de recta **A** de comprimento γ) sobre uma folha de papel ilimitada (plano) dividida por linhas rectas paralelas e equidistantes (com as distâncias entre as rectas igual a δ). Qual a probabilidade de que a agulha intersecte uma dessas rectas?

SOLUÇÃO: Neste problema pretende-se lançar, à sorte, um segmento de recta num plano que, segundo a DEFINIÇÃO 5.4, corresponde a lançar à sorte uma recta na região (isto é lançar, à sorte, um ponto e, à volta desse ponto, uma direcção conforme a DEFINIÇÃO 5.3) e lançar, à sorte, o segmento de recta no segmento que resulta da intersecção da recta lançada com a região.

Contudo, uma vez que estamos a lançar um segmento de recta num plano, logo numa região ilimitada, vamos ter que recorrer à definição de lançamento de pontos em regiões ilimitadas apresentada na secção 4.7, através da qual lançar, à sorte, um ponto numa região ilimitada $\mathbf{X}$ significa, por definição, lançar, à sorte, um ponto numa região $\mathbf{X}' \subset \mathbf{X}$ limitada e arbitrariamente grande.

Pacheco d'Amorim começa por tirar, à sorte, a direcção α da recta que contém o segmento (α será o ângulo que a direcção escolhida para a recta faz com as linhas rectas paralelas e equidistantes da folha); em seguida, para definir a recta, o autor lança, à sorte, um ponto m dentro dum segmento do plano arbitrariamente grande (conforme a definição de lançamento de um ponto numa região ilimitada). Porém, por maior que seja esse segmento do plano, poderemos sempre determinar um paralelograma $[pqsr]$ com os lados paralelos à direcção achada para a recta e às linhas rectas da folha (sejam $[ab]$, $[a'b']$, etc.), paralelograma esse que contém o segmento do plano considerado e fazer dentro dele o lançamento do ponto m (Figura 5.13).

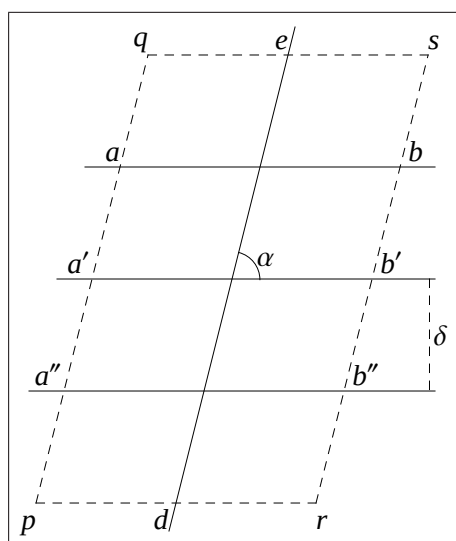


Figura 5.13: Região de lançamento da agulha

Suponhamos, então, que estes dois lançamentos determinaram a recta

que contém o segmento $[ed]$. Determinada a recta, resta lançar o segmento **A** sobre o segmento $[ed]$ (intersecção da recta obtida com a região — paralelograma $[pqsr]$) recorrendo à DEFINIÇÃO 5.4.

No entanto, qualquer que seja a posição do ponto m , a probabilidade de **A** encontrar uma das rectas paralelas é sempre a mesma (com uma direcção fixa, o ponto só vai definir a localização da recta, mas, seja ela qual for, teremos sempre o mesmo rácio entre o comprimento da região favorável e o comprimento da região possível). Assim, pela Proposição VIII (página 254) podemos supor o ponto m fixo. Sob esta hipótese devemos fazer o segmento de recta **A** percorrer o segmento de recta $[ed]$ (intersecção da recta com o paralelograma $[pqsr]$). Quando a origem do segmento **A** percorrer um segmento de recta compreendido entre duas paralelas consecutivas, o espaço durante o qual esse segmento encontra uma das paralelas é sempre o mesmo, isto é, a região favorável é igual. Por este motivo as paralelas dividem o campo possível do lançamento do segmento **A** sobre o segmento $[ed]$ em partes com igual probabilidade. Desta forma, pela Proposição IX (página 254), será suficiente determinar a probabilidade numa delas. Deste modo, para determinada direcção α fixa (sendo α o ângulo $\angle [rde]$ da Figura 5.13), vamos obter para as situações em que a distância entre as paralelas é pelo menos igual ao comprimento da agulha ($\delta \geq \gamma$)

$$\mathbb{P} = \frac{\gamma}{\frac{\delta}{\sin \alpha}} = \frac{\gamma \sin \alpha}{\delta}. \quad (5.7)$$

Por outro lado, a possibilidade por unidade de α é $\frac{1}{\pi}$, logo

$$\mathbb{P} = \int_0^\pi \frac{1}{\pi} \cdot \frac{\gamma \sin \alpha}{\delta} d\alpha = \frac{\gamma}{\delta\pi} (-\cos \alpha)|_0^\pi = \frac{2\gamma}{\delta\pi}. \quad (5.8)$$

Deste modo concluímos que, caso a agulha tenha um comprimento inferior à distância entre as rectas ($\gamma < \delta$), a probabilidade obtida é proporcional ao comprimento da agulha.

No caso de a distância entre as paralelas ser inferior ao comprimento da agulha ($\delta < \gamma$), teremos de decompor o campo da variação de α em duas partes: a primeira constituída pela totalidade de valores de α para os quais pode ou não haver encontro; a segunda constituída pelos outros valores onde obrigatoriamente a agulha vai cruzar uma das paralelas. Aplicando os Teoremas da Probabilidade Total e Composta e notando que, se representarmos por α_0 o ângulo que a agulha faz com as linhas paralelas quando toca nas duas linhas (conforme Figura 5.14), teremos

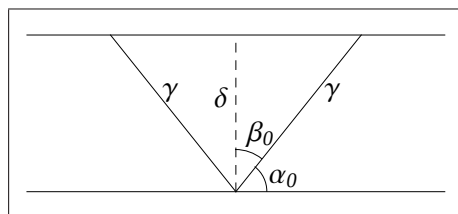


Figura 5.14: Campo de variação da agulha

$$\delta = \gamma \sin \alpha_0, \quad (5.9)$$

sendo a probabilidade obtida através de

$$\begin{aligned} \mathbb{P} &= \frac{\pi - 2\alpha_0}{\pi} \cdot 1 + 2 \int_0^{\alpha_0} \frac{1}{\pi} \cdot \frac{\gamma \sin \alpha}{\delta} d\alpha = \\ &= \frac{\pi - 2\alpha_0}{\pi} + \frac{2\gamma}{\delta\pi} (1 - \cos \alpha_0) = \\ &= \frac{2\beta_0}{\pi} + \frac{2\gamma}{\delta\pi} (1 - \sin \beta_0), \end{aligned} \quad (5.10)$$

onde $\beta_0 = \frac{\pi}{2} - \alpha_0$. ◇

O problema da agulha, um dos problemas mais famosos de probabilidade geométrica, foi proposto pela primeira vez por Buffon numa conferência na *Académie Royale des Sciences de Paris* em 1733⁽⁹⁾, onde apresentou a solução

⁽⁹⁾ Contudo os resultados só foram publicados após mais de 40 anos, em 1777, no

de diversos problemas em Probabilidade geométrica, conforme nota do editor da *Histoire de L'Académie Royale des Sciences* de 1733 páginas 43–45. Um dos problemas debatidos refere-se ao jogo designado como *le jeu du franc-carreau* que consiste em considerar um pavimento dividido em quadrados iguais, onde é lançada uma moeda; um dos jogadores aposta que a moeda cairá num só quadrado (“*franc-carreau*”), o outro aposta que a peça tocará em pelo menos dois quadrados. Nesta situação pretende-se determinar a probabilidade de cada jogador ganhar, considerando o raio da moeda (seja r) e o comprimento dos lados do quadrado (seja l) fixos ou, numa outra variante do problema, determinar o comprimento dos lados do quadrado de forma a tornar o jogo equitativo (probabilidades iguais para a vitória de cada jogador). Buffon resolve o problema do jogo à *franc-carreau* notando que, para determinar a probabilidade de vitória de um jogador, é suficiente utilizar o local do pavimento onde cai o centro da moeda.

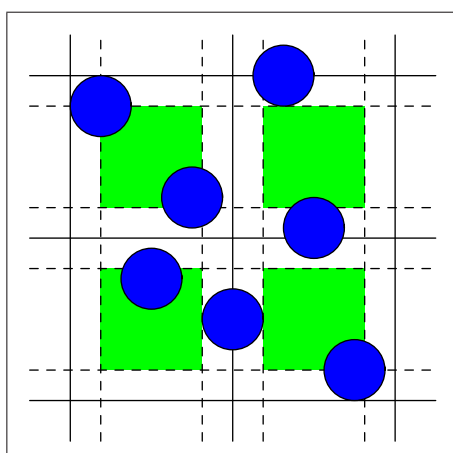


Figura 5.15: O jogo do *franc-carreau*

Essai d'arithmétique moral, p. 95–105. Segundo Holgate (1981), o texto patente no *Essai d'arithmétique moral* não é muito distinto de um artigo submetido em 1733 por Buffon à *Académie Royale des Sciences de Paris*. Todhunter (1865) apresenta uma boa descrição da análise do problema da agulha efectuada no *Essai d'arithmétique moral* por Buffon.

Neste problema a região favorável será constituída pelos quadrados inscritos nos quadrados do pavimento com o mesmo ponto central, mas com os lados com um comprimento igual a $l - 2r$ (inferiores no valor do diâmetro da moeda). Assim, se o centro da moeda se situar nestes quadrados menores, então toda a moeda estará dentro de um dos quadrados do pavimento; caso contrário, significará que a moeda se situará em pelo menos dois quadrados do pavimento. Desta forma, a probabilidade não será determinada dividindo o espaço amostra em acontecimento de igual probabilidade, e calculando o quociente entre o número de casos favoráveis e o número de casos possíveis, mas utilizando o quociente entre as medidas (áreas) favorável e possível, sendo, por este motivo, considerado o primeiro problema de probabilidade geométrica.

Buffon generaliza a questão para outras formas do pavimento, onde, em vez de quadrados, teremos triângulos, losangos, hexágonos, entre outras formas; e determina não só a probabilidade de sair um *franc-carreau*, como a probabilidade de a moeda intersectar um determinado número de ladrilhos. Outra generalização efectuada refere-se a alterações da forma do objecto lançado, onde, em vez de uma moeda circular, utiliza objectos tais como um quadrado, uma vara (*baguette*) ou uma agulha. No problema da agulha, bem mais complicado que o jogo *franc-carreau* e que requer obrigatoriamente um recurso ao cálculo integral para a sua resolução, Buffon começa por considerar um pavimento dividido por linhas rectas paralelas e uma agulha com um comprimento ($2b$) inferior à distância das rectas do pavimento ($2a$). Buffon resolve correctamente esta versão do problema, embora enuncie uma versão ainda mais complicada do problema da agulha, onde o pavimento, além de estar dividido pelas rectas paralelas, está também dividido por outras rectas paralelas, que são perpendiculares às primeiras. Desta forma temos o problema onde o pavimento está dividido em rectângulos. Todavia, acrescente-se

que a sua resolução desta versão está errada.

Laplace (1812, p. 359–362) resolve correctamente a segunda versão do problema da agulha de Buffon. Na obra de Laplace não existe nenhuma referência à origem deste problema, sendo por este motivo apelidado, por alguns autores, por problema de Laplace. Actualmente, a versão em que o pavimento está dividido em rectângulos é usualmente denominado por problema da agulha de Buffon-Laplace.

Barbier (1860), que refere erradamente Laplace como autor deste problema, resolve-o utilizando o valor esperado do ganho de um jogador que deverá receber um franco por cada ponto de intersecção da agulha e substitui a agulha (segmento de recta) por um círculo após, de forma brilhante, observar que a forma do objecto não altera o valor esperado. Contudo, a resolução deste problema com o recurso ao valor esperado será comentado unicamente na secção 8.5 presente na página 545.

Este problema é analisado, com mais ou menos detalhe, em diversas obras contemporâneas à tese de doutoramento de Pacheco d'Amorim, tais como Czuber (1884, p. 84–91, 116–117), Bertrand (1888, p. 52–53), Poincaré (1896, p. 126–130), Borel (1909, p. 104–109) e Borel (1914, p. 80), entre outras.

Pensamos que a resolução do problema da agulha apresentada por Pacheco d'Amorim está extremamente bem concebida, pois o autor conseguiu simplificar de uma forma bastante criativa o problema. Assim, recorre à definição de lançamento de um ponto num plano (região ilimitada) para transpor o problema para um segmento limitado do plano e, uma vez que podemos obter esse segmento plano, por maior que seja, através da união de paralelogramas, o autor recorre à Proposição IX para limitar a análise a um destes paralelogramas. Finalmente, como o campo possível e favorável é o mesmo entre cada duas paralelas do pavimento, o problema é resumido à análise entre duas paralelas do pavimento. Desta forma, para determinar a probabilidade

pretendida é suficiente analisar essa probabilidade numa região restrita a um segmento de um paralelograma, não sendo necessário considerar toda a região $\mathbb{R}^2$. Por este motivo, se compararmos a resolução de Pacheco d'Amorim com a apresentada por outros autores seus contemporâneos, concluimos que é elementar e evidente.

Refira-se que Pacheco d'Amorim, na resolução deste problema, se baseia na DEFINIÇÃO 5.4 que, por sua vez, depende da DEFINIÇÃO 5.3 e esta, como salientamos na crítica à resolução do paradoxo de Bertrand, não é aceitável pelo facto de as rectas (e, por consequência, as cordas) escolhidas não serem equiprováveis. Contudo, como no problema da agulha de Buffon a agulha é lançada num plano (região ilimitada) a sua resolução não é afectada por esta questão.

5.2 Lançamento, à sorte, de figuras variáveis

Pacheco d'Amorim refere que é impossível retratar o lançamento de todas as figuras variáveis e que, por este motivo, se vai restringir às figuras poligonais (abertas e fechadas) articuladas e, como caso limite, às curvas flexíveis e inextensíveis, onde uma figura poligonal articulada é uma linha poligonal de ângulos variáveis.

Definição 5.10. Lançar, à sorte, **uma figura variável numa dada região $\mathbf{A}$** significa, por definição, tirar, à sorte, a forma da figura e, em seguida, lançar a figura, à sorte, na região $\mathbf{A}$ como se fosse uma figura rígida (conforme as definições apresentadas na secção 5.1).

5.2.1 Figuras poligonais abertas

Definição 5.11. Fixar, à sorte, **a forma de uma figura poligonal articulada aberta** é, por definição, tirar, à sorte, a forma de cada um dos seus vértices.

Seja **a** um dos vértices ou articulações de uma dada figura poligonal definida num espaço a n dimensões e sejam $(x_1, x_2, \dots, x_n)$ as coordenadas desse vértice. Consideremos, tal como na DEFINIÇÃO 5.1, o espaço **E** formado pelos pontos que satisfazem

$$(x_1 - x'_1) + (x_2 - x'_2) + \dots + (x_n - x'_n) = 1, \quad (5.11)$$

e suponhamos um dos ramos do vértice fixo e o outro movendo-se de modo a ocupar, em relação ao ramo fixo, todas as posições possíveis. Este processo definirá uma subregião $\mathbf{B} \subset \mathbf{E}$ que corresponde às possíveis direcções do segundo ramo em relação ao primeiro.

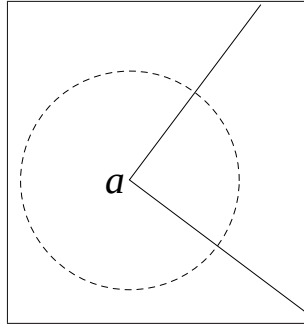


Figura 5.16: Tirar, à sorte, a forma de um vértice

Desta forma tirar, à sorte, **a forma da articulação a** é, por definição, lançar, à sorte, um ponto na região **B**.

5.2.2 Figuras poligonais fechadas

As definições a seguir apresentadas, DEFINIÇÃO 5.12 e 5.13, referem-se ao lançamento, à sorte, de figuras poligonais fechadas. Saliente-se que, tal como nas figuras poligonais abertas, Pacheco d'Amorim nunca menciona a determinação do comprimento dos lados. Desta forma, a análise efectuada por Pacheco d'Amorim, no que se refere ao lançamento de figuras, não só se restringe a figuras poligonais abertas e fechadas⁽¹⁰⁾, como o autor explicitamente afirma, mas também se limita a figuras com os lados com comprimentos e ordem previamente fixados (seja de que forma for, pois a exposição de Pacheco d'Amorim é omissa em relação a este aspecto). Assim, quer na tiragem, à sorte, da forma das figuras poligonais abertas, quer das figuras poligonais fechadas, a definição da forma destas figuras apenas acarreta a escolha, à sorte, dos seus vértices, isto é, dos ângulos que a figura possui, pois os restantes factores são dados do problema.

Salientemos, uma vez mais, que as definições que vamos apresentar de tiragem, à sorte, da forma de um polígono fechado serão sempre restritas a polígonos com os lados com o comprimento fixo e a ordem dos lados também estabelecida *a priori*.

Pacheco d'Amorim começa por analisar o tirar, à sorte, a forma de um quadrilátero sem fazer qualquer referência à forma de um triângulo. Contudo notemos que um triângulo com os lados com o comprimento fixo tem duas formas possíveis. Por exemplo, consideremos um triângulo com lados com comprimentos iguais a, respectivamente, 10, 3 e 9 centímetros. Neste caso teremos, como formas possíveis, os triângulos $[acb]$ e $[ac'b]$ da Figura 5.17 onde foi considerado fixo o lado $[ab]$.

⁽¹⁰⁾ O autor analisa ainda o limite destas figuras ao qual apelida de lançamento de curvas flexíveis e inextensíveis, que será comentado na secção 5.2.3.

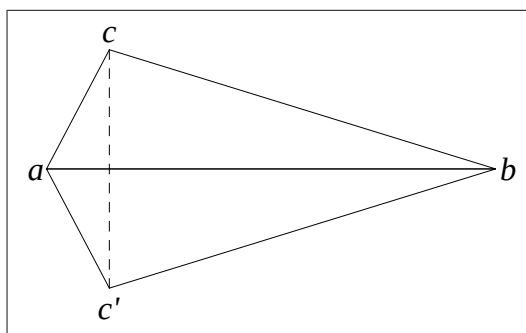
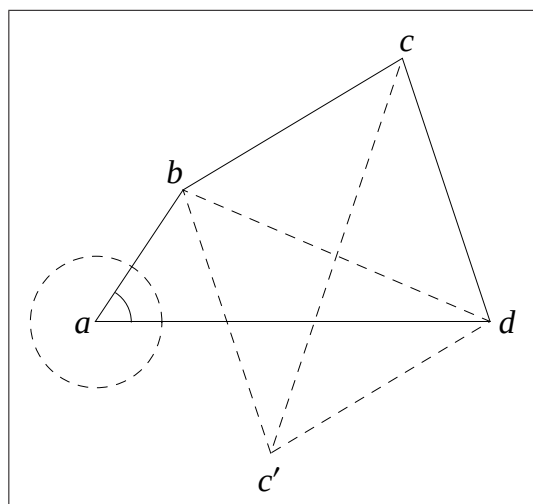


Figura 5.17: Tirar, à sorte, a forma de um triângulo

Se o objectivo for lançar um triângulo (com a forma variável) numa região plana **A** (conforme DEFINIÇÃO 5.10), o próprio lançamento define uma orientação para o triângulo em relação à região **A**. Assinalemos, todavia, que o triângulo $[ac'b]$ não pode ser obtido através do triângulo $[acb]$, isto é, um triângulo não é obtido através do outro usando unicamente uma rotação, pois, para obter um através do outro, seria necessário recorrer a uma reflexão. Por este motivo, no lançamento referido, os dois triângulos iriam obter posições (possíveis) distintas, devendo, então, ser considerados duas formas diferentes que o triângulo pode assumir. Desta forma, apesar de os dois triângulos serem considerados equivalentes na geometria Euclidiana, no lançamento destes dois triângulos à sorte num plano não podemos obter os mesmos resultados, razão pela qual não podemos considerar as duas formas mencionadas para o triângulo como equivalentes no que respeita aos lançamentos definidos por Pacheco d'Amorim. Refira-se, porém, que a apresentação de Pacheco d'Amorim é omissa em relação a esta situação.

Analisemos, então, o que Pacheco d'Amorim define em relação à tiragem, à sorte, da forma de um quadrilátero. Consideremos o quadrilátero $[abcd]$ representado na Figura 5.18 e suponhamos o lado $[ad]$ fixo.

Quando analisamos todas as formas possíveis que o quadrilátero $[abcd]$ pode assumir, o lado $[ab]$ intersecta um conjunto de pontos na circunferência



à volta do vértice **a** (pontos estes que definem o ângulo $\angle [bad]$). Notemos que, para cada valor do ângulo $\angle [bad]$, teremos três possibilidades: ou o valor do ângulo $\angle [bad]$ não permite a junção dos outros dois lados (segmentos $[bc]$ e $[cd]$) pelo facto de os pontos b e d estarem demasiado distantes (a distância entre estes pontos ser superior à soma dos comprimentos dos lados $[bc]$ e $[cd]$) e, conseqüentemente, não é possível formar nenhum quadrilátero com esse ângulo $\angle [bad]$; ou (como ilustra a Figura 5.18) o ângulo $\angle [bad]$ permite a junção dos outros dois lados ($[bc]$ e $[cd]$) de duas formas distintas, possibilitando assim a formação de dois quadriláteros, que são $[abcd]$ e $[abc'd]$ (apesar de a forma destes quadriláteros poder ser a mais variada, podendo-se obter quadriláteros convexos, côncavos, estrelados, entre outros⁽¹¹⁾); ou estamos numa situação intermédia das duas anteriores e só é possível, com esse ângulo $\angle [bad]$, formar um quadrilátero utilizando os lados $[bc]$ e $[cd]$ com o mesmo sentido (a distância entre os pontos b e d é igual à soma dos comprimentos dos lados $[bc]$ e $[cd]$ e o quadrilátero terá a forma de um triângulo).

⁽¹¹⁾ Pacheco d'Amorim (2004, Vol. 1, capítulo III) faz uma boa descrição dos tipos de polígonos existentes e das suas propriedades.

Quando analisarmos a Figura 5.21 (página 306), onde são analisadas as diferentes possibilidades associadas às distintas formas do ângulo $\angle [bad]$ para um exemplo particular, estas três possibilidades serão mais perceptíveis.

Os ângulos, em relação à forma do vértice **a** (ângulo $\angle [bad]$), que tornam o quadrilátero $[abcd]$ possível constituem o conjunto **A** e os ângulos que tornam o quadrilátero $[abc'd]$ possível constituem o conjunto **A**₁. Os conjuntos **A** e **A**₁ podem ou não conter os mesmo pontos, contudo consideramos sempre que os pontos pertencentes a **A** são distintos dos pertencentes a **A**₁ (mesmo que assumam o mesmo ângulo, consideramos os valores como distintos). Façamos o mesmo com os vértices **b**, **c** e **d** de forma a obter os conjuntos **B**, **B**₁, **C**, **C**₁, **D** e **D**₁⁽¹²⁾.

Definição 5.12 (No plano). Lançar, à sorte, **a forma de um quadrilátero** $[abcd]$ é, por definição, lançar à sorte um ponto em **A** ou **A**₁ ou **B** ou **B**₁ ou **C** ou **C**₁ ou **D** ou **D**₁, isto é, escolher um ponto em

$$\mathbf{T} = \mathbf{A} \sqcup \mathbf{A}_1 \sqcup \mathbf{B} \sqcup \mathbf{B}_1 \sqcup \mathbf{C} \sqcup \mathbf{C}_1 \sqcup \mathbf{D} \sqcup \mathbf{D}_1.$$

Em relação a esta definição salientamos que, para o autor, a região **A** é (mesmo que contenha os mesmo pontos) disjunta da região **A**₁, isto é, apesar de os pontos poderem ser os mesmos (correspondendo ao mesmo ângulo $\angle [bad]$) “*nem por isso essas regiões deixam de ser distintas*” (razão pela qual utilizamos o símbolo $\sqcup$ na definição). Assim, o ponto escolhido através desta definição corresponde a uma só forma para o quadrilátero, se bem que, cada possível forma do quadrilátero não tenha apenas um ponto associado na região **T** pois, por exemplo, qualquer quadrilátero associado a um ponto da região **A** pode ser obtido através da rotação de um quadrilátero associado a

⁽¹²⁾ Se fosse aplicado o mesmo raciocínio no triângulo representado na Figura 5.17 (analisando as possíveis formas dos vértices **b** e **c**), teriam resultado triângulos que podem ser obtidos por rotação através de um dos dois triângulos apresentados na figura.

um outro ponto da região **B** e a um outro associado à região **C** e ainda a um outro ponto associado à região **D**. Desta forma, existem quatro pontos em **T** em que se obtêm quadriláteros semelhantes, isto é, em que qualquer um destes quadriláteros pode ser obtido por qualquer um dos outros através da aplicação de uma rotação (recordemos que, quando uma figura é lançada numa região, o primeiro factor a ser determinado é a sua orientação, pelo que é indiferente lançar qualquer um destes quatro quadriláteros).

Exemplifiquemos como poderíamos simular este mecanismo de tirar à sorte a forma de um quadrilátero $[abcd]$ cujos lados têm comprimento respectivamente $\overline{ab} = 4$, $\overline{bc} = 3$, $\overline{cd} = 2$ e $\overline{da} = 1$. Com o objectivo de simplificar a apresentação deste exemplo, iremos representar as regiões em termos da variação dos ângulos em análise em vez dos pontos que resultam da intersecção do lado móvel com o circunferência unitária, uma vez que estes métodos são análogos. Para começar consideremos o lado $[da]$ como fixo e analisemos os valores que o ângulo $\angle [dab]$ pode assumir. Em primeiro lugar atentemos em que, para que seja possível formar um quadrilátero (ligar os outros lados — $[bc]$ e $[cd]$, que juntos têm um comprimento igual a 5), os vértices **d** e **b** têm que estar, no máximo, a uma distância igual a 5. Como os segmentos $[da]$ e $[ab]$ juntos também têm um comprimento igual a 5, o ângulo $\angle [dab]$ pode assumir qualquer valor (notemos que no caso de o ângulo $\angle [dab]$ assumir o valor π teremos os segmentos sobrepostos e a forma resultante será apenas um segmento de recta). Saliente-se que também não existe qualquer restrição quando analisamos o valor mínimo pois, seja qual for o valor do ângulo $\angle [dab]$ os vértices **d** e **b** terão sempre uma distância situada no intervalo $[3, 5]$ sendo, para qualquer destes valores, sempre possível formar um quadrilátero com os outros lados $[bc]$ e $[cd]$. Constatemos também que, para cada valor do ângulo $\angle [dab]$, é possível construir dois quadriláteros, um representado na Figura 5.18 através de $[abcd]$ e o outro através de $[abc'd]$. Desta forma

teremos as regiões $\mathbf{A} = [0, 2\pi)$ e $\mathbf{A}_1 = [0, 2\pi)$ onde, apesar de assumirem o mesmo intervalo, serão consideradas como duas regiões distintas quando formos escolher um ponto à sorte em $\mathbf{T}$. Pacheco d’Amorim refere que a região $\mathbf{A}_1$ pode não ser igual à região $\mathbf{A}$, pois “*as articulações podem estar sujeitas a ligações que não tornem possível a posição c*”. Todavia, o autor não explicita esta questão. Pensamos que unicamente em situações em que não pretendamos, por exemplo, polígonos estrelados (onde os lados se intersectam) estas duas regiões podem ser distintas, pois, caso não façamos nenhuma restrição deste género, as duas regiões serão sempre iguais, uma vez que para cada ângulo de um vértice, por exemplo $\mathbf{a}$, há sempre duas possibilidades de ligar os lados opostos ($[bc]$ e $[cd]$), havendo desta forma dois possíveis quadriláteros ou, caso contrário, não é possível formar qualquer quadrilátero com esse ângulo no vértice $\mathbf{a}$.

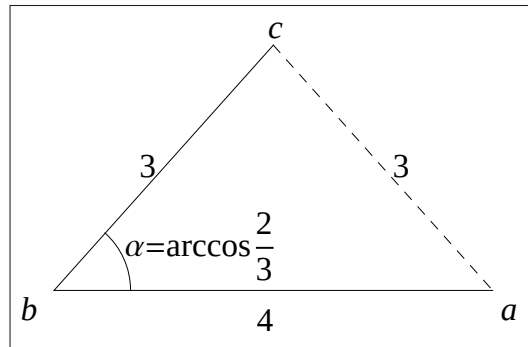


Figura 5.19: Quando um quadrilátero à sorte sai triângulo

Analisemos agora a situação em que o lado $[ab]$ é o fixo e investiguemos os valores que o ângulo $\angle[abc]$ pode assumir. Uma vez que os restantes lados ($[cd]$ e $[da]$) somam comprimento igual a 3, a distância entre os vértices $\mathbf{a}$ e $\mathbf{c}$ terá que assumir no máximo este valor. Por conseguinte o ângulo $\angle[abc]$ só pode assumir valores no intervalo $\mathbf{B} = \mathbf{B}_1 = [0, \arccos(\frac{2}{3})] \cup [2\pi - \arccos(\frac{2}{3}), 2\pi]$. Notemos que nos casos em que $\angle[abc] = \arccos(\frac{2}{3})$ ou $\angle[abc] = 2\pi - \arccos(\frac{2}{3})$ os vértices $\mathbf{a}$ e $\mathbf{c}$ têm uma distância igual a

3, logo os lados $[cd]$ e $[da]$ terão obrigatoriamente de fazer um ângulo raso entre si ($\angle[cd] = \pi$) para formar um polígono fechado, resultando, nestas duas situações, um triângulo. Os vértices **a** e **c** no mínimo, para $\angle[abc] = 0$, terão uma distância igual a 1, situação que torna possível a formação de um polígono fechado com os lados $[cd]$ e $[da]$, não havendo qualquer restrição neste caso apesar de, neste limite, o polígono obtido assumir a forma de um segmento.

Considerando agora o lado $[bc]$ fixo, o ângulo $\angle[bcd]$ também não pode assumir qualquer valor em $[0, 2\pi)$, pois para estes valores a distância entre os vértices **b** e **d** situar-se-á no intervalo $[1, 5]$ e com os restantes dois lados, $[da]$ e $[ab]$, só é possível formar um quadrilátero caso a distância entre **b** e **d** esteja compreendida no intervalo $[3, 5]$. Desta forma só serão possíveis para o ângulo $\angle[bcd]$ os valores que tornem os vértices **b** e **d** com uma distância superior a três, obtendo-se $\mathbf{C} = \mathbf{C}_1 = [\arccos(\frac{1}{3}), 2\pi - \arccos(\frac{1}{3})]$.

Utilizando raciocínio semelhante ao efectuado para o vértice **a** obtém-se facilmente $\mathbf{D} = \mathbf{D}_1 = [0, 2\pi)$. Vamos, então, tirar à sorte um ponto da região $\mathbf{T} = \mathbf{A} \sqcup \mathbf{A}_1 \sqcup \mathbf{B} \sqcup \mathbf{B}_1 \sqcup \mathbf{C} \sqcup \mathbf{C}_1 \sqcup \mathbf{D} \sqcup \mathbf{D}_1$ que tem amplitude total igual a $12\pi + 4\arccos(\frac{2}{3}) - 4\arccos(\frac{1}{3})$. Assim, podemos escolher um ponto aleatoriamente no intervalo $[0, 12\pi + 4\arccos(\frac{2}{3}) - 4\arccos(\frac{1}{3})]$ que pode ser decomposto, representando por k_1 o valor $\arccos(\frac{1}{3})$ e por k_2 o valor $\arccos(\frac{2}{3})$, em

$$\begin{aligned} & [0, 2\pi) \cup [2\pi, 4\pi) \cup [4\pi, 4\pi + 2k_2) \cup [4\pi + 2k_2, 4\pi + 4k_2) \cup \\ & \cup [4\pi + 4k_2, 6\pi + 4k_2 - 2k_1) \cup [6\pi + 4k_2 - 2k_1, 8\pi + 4k_2 - 4k_1) \cup \\ & \cup [8\pi + 4k_2 - 4k_1, 10\pi + 4k_2 - 4k_1) \cup [10\pi + 4k_2 - 4k_1, 12\pi + 4k_2 - 4k_1) \end{aligned}$$

onde cada um destes conjuntos é associado, respectivamente, a **A**, **A**₁, **B**, **B**₁, **C**, **C**₁, **D** e **D**₁.

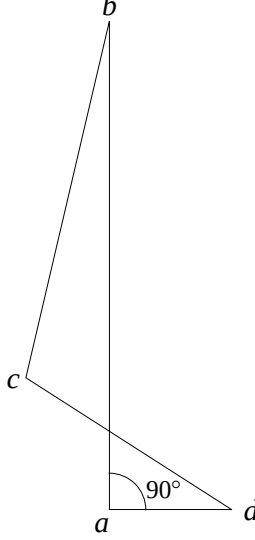
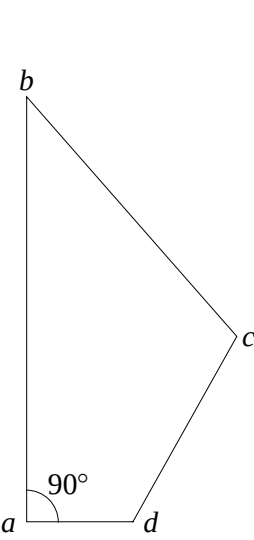
		
$\mathbf{T} = \frac{5\pi}{2}$	$\mathbf{T} = \frac{\pi}{2}$	$\mathbf{T} = 10\pi - 4 \arccos\left(\frac{1}{3}\right)$

Figura 5.20: Quadriláteros tirados à sorte

Assim, se o ponto escolhido à sorte em $\mathbf{T}$ for $\frac{5\pi}{2}$, por exemplo, este valor está associado à região $\mathbf{A}_1 = [0, 2\pi)$, mais concretamente ao valor $\frac{\pi}{2}$, isto é, o ângulo $\angle[dab] = \frac{\pi}{2}$, e, como pertence a $\mathbf{A}_1$ (e não a $\mathbf{A}$), o quadrilátero não será convexo (consultar Figura 5.20 onde apresentamos também o quadrilátero associado ao valor $\mathbf{T} = \frac{\pi}{2}$ que corresponde à região $\mathbf{A} = [0, 2\pi)$ e consequentemente ao quadrilátero convexo com $\angle[dab] = \frac{\pi}{2}$). Se o ponto escolhido à sorte na região em $\mathbf{T}$ for $10\pi - 4 \arccos\left(\frac{1}{3}\right)$, este valor está associado ao valor $2\pi - 4 \arccos\left(\frac{2}{3}\right)$ da região $\mathbf{D} = [0, 2\pi)$, correspondendo neste caso a um quadrilátero convexo com o ângulo $\angle[cda] = 2\pi - 4 \arccos\left(\frac{2}{3}\right) \approx 167.24^\circ$.

Quando apresentamos a definição do lançamento de um quadrilátero referimos que na região $\mathbf{T} = \mathbf{A} \sqcup \mathbf{A}_1 \sqcup \dots \sqcup \mathbf{D}_1$, onde é lançado à sorte o *ponto equivalente* do quadrilátero, existem pontos que definem o mesmo quadrilátero, isto é, temos um ponto em $\mathbf{A}$, outro em $\mathbf{B}$, outro em $\mathbf{C}$ e outro em $\mathbf{D}$ que têm associados quadriláteros equivalentes, que podem ser obtidos uns através dos outros com a aplicação de uma rotação. Contudo, caso tivéssemos definido a região $\mathbf{T}$ utilizando unicamente as regiões $\mathbf{A}$ e $\mathbf{A}_1$ o resultado

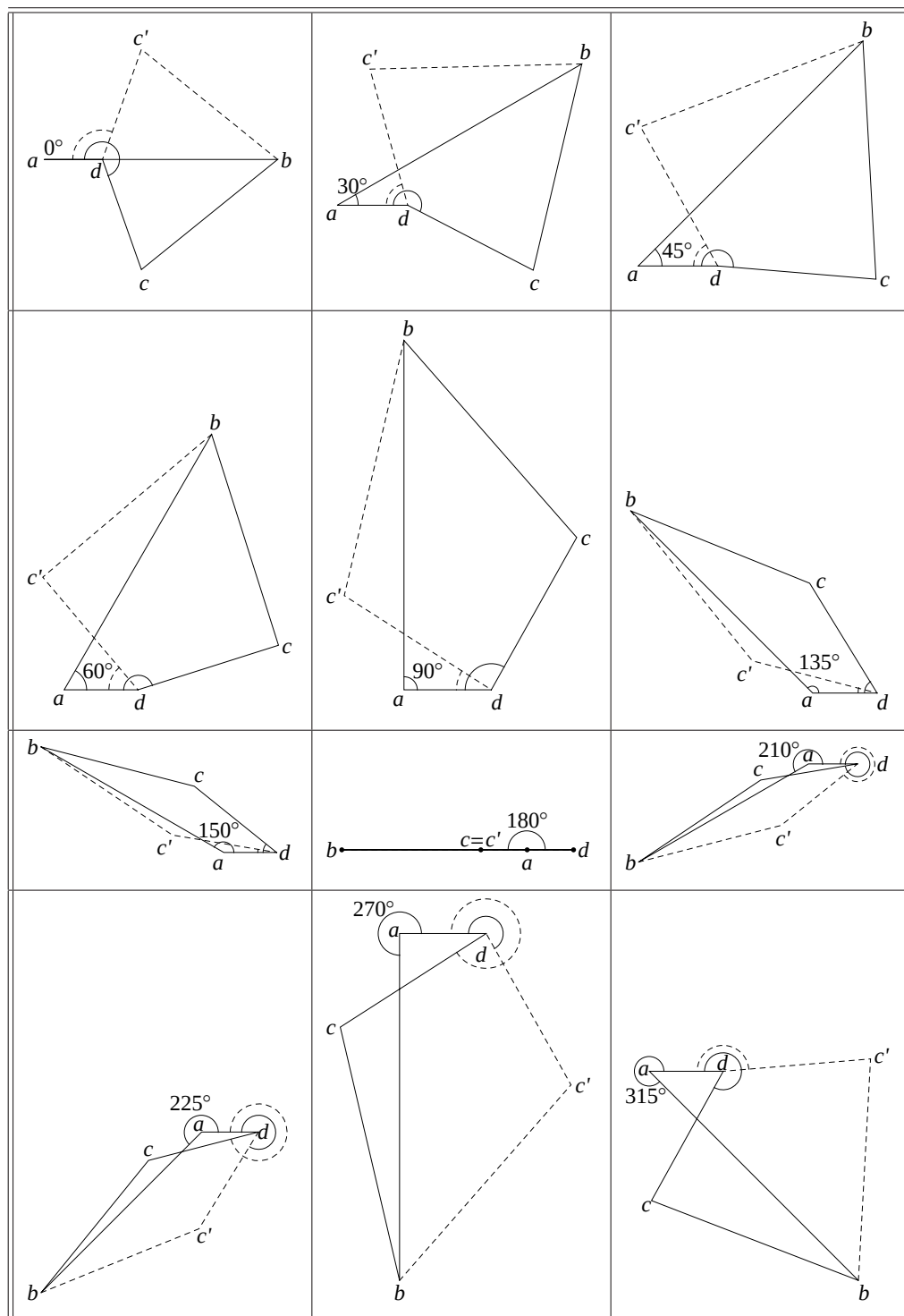


Figura 5.21: Gráficos dos ângulos de um quadrilátero

seria diferente, pois a função densidade associada a cada forma do quadrilátero seria distinta. Para ilustrarmos esta ideia analisemos os valores patentes na Tabela 5.2, onde são analisados os possíveis valores para o ângulo $\angle [dab]$ (que definem as regiões $\mathbf{A}$ e $\mathbf{A}_1$), aos quais estão associados, em geral, dois possíveis quadriláteros, um correspondente ao ângulo $\angle [adc]$ (quadriláteros associados à região $\mathbf{A}$), e outro correspondente ao ângulo $\angle [adc']$ (quadriláteros associados à região $\mathbf{A}_1$). Assim, na tabela são analisadas as variações dos ângulos referentes aos vértices $\mathbf{a}$ e $\mathbf{d}$.

$\angle [dab]$	$\angle [adc]$	$\angle [adc']$	$\angle [dab]$	$\angle [adc]$	$\angle [adc']$
0°	250.529°	109.471°	150°	38.809°	9.457°
30°	207.387°	74.484°	180°	0°	0°
45°	184.791°	60.97°	210°	350.543°	308.809°
60°	162.412°	49.792°	225°	345.515°	301.572°
90°	119.278°	32.65°	270°	327.25°	240.722°
135°	58.428°	14.486°	315°	299.03°	175.209°

Tabela 5.2: Variação dos ângulos de um quadrilátero

Por outro lado, os distintos valores que o ângulo correspondente ao vértice $\mathbf{d}$ do quadrilátero pode assumir estarão descritos nos valores $\angle [adc]$ e $\angle [adc']$ que percorrem o intervalo $[0, 2\pi)$. Se as formas dos quadriláteros tivessem a mesma densidade nas regiões $\mathbf{A} \sqcup \mathbf{A}_1$, $\mathbf{B} \sqcup \mathbf{B}_1$, $\mathbf{C} \sqcup \mathbf{C}_1$ e $\mathbf{D} \sqcup \mathbf{D}_1$, a possibilidade de cada forma possível para o quadrilátero teria obrigatoriamente de ser igual, independentemente do vértice que escolhêssemos para tirar à sorte o seu ponto equivalente. Todavia, os valores patentes na Tabela 5.2 claramente demonstram que esta proporcionalidade (ou igualdade, pois ambos os ângulos possuem o mesmo campo de variação — $[0, 2\pi)$) não se verifica. Por exemplo, quando o vértice $\mathbf{a}$ varia entre os 0° e os 30° , o vértice $\mathbf{d}$ tem dois tipos de variação: um assumirá valores entre 207.387° e

250.529° e o outro entre 74.484° e 109.471° . Neste caso o ângulo $\angle [dab]$ variou 30° , que representa 8.33 por cento do seu campo de variação, enquanto a correspondente variação do vértice **d** foi superior, uma vez que o ângulo $\angle [adc]$ variou 43.142° (11.98 por cento do seu campo de variação) e o ângulo $\angle [adc']$ 34.987° (9.72 por cento do seu campo de variação). Mesmo que efectuemos uma média ponderada entre as duas possíveis variações do vértice **d**, independentemente dos pesos atribuídos na determinação dessa média, a probabilidade das possíveis formas do quadrilátero, quando o vértice **a** assume valores no intervalo $[0, 30^\circ]$, será superior a da hipótese de ser lançado em $\mathbf{D} \sqcup \mathbf{D}_1$ do que em $\mathbf{A} \sqcup \mathbf{A}_1$.

Desta forma, caso a definição da região **T** fosse restrita à forma de apenas um dos vértices do quadrilátero, a própria escolha desse vértice iria implicar uma distinção entre as possíveis formas do quadrilátero, isto é, iria associar uma possibilidade superior a algumas formas do quadrilátero, pois a escolha de outro vértice iria associar outra possibilidade a cada possível forma do quadrilátero.

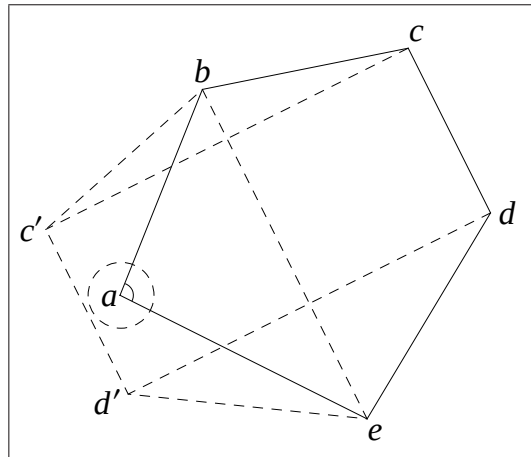


Figura 5.22: Tirar, à sorte, a forma de um pentágono (original)

Se pretendermos tirar, à sorte, a forma de um pentágono $[abcde]$ (Fi-

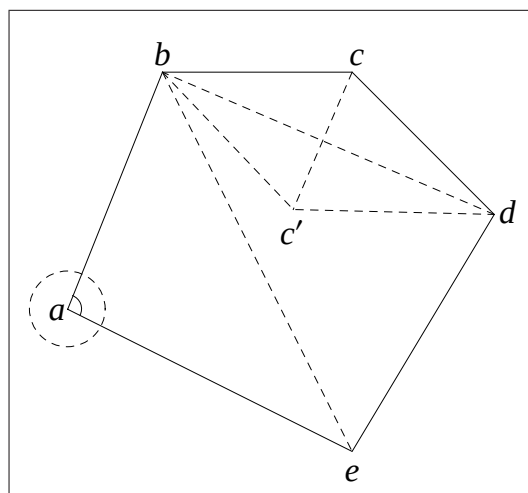


Figura 5.23: Tirar, à sorte, a forma de um pentágono (alterado)

gura 5.23⁽¹³⁾) o raciocínio é análogo. Neste caso, para cada valor do ângulo $\angle[eab]$ o quadrilátero $[bcde]$ pode assumir uma infinidade de formas (que foram descritas pelo processo de tiragem à sorte da forma de um quadrilátero). Deste modo, se fixarmos a forma do vértice **a** (considerarmos o ângulo $\angle[eab]$ fixo), estaremos consequentemente a fixar o comprimento do lado $[eb]$ e, partindo deste pressuposto, podemos analisar as possíveis formas do quadrilátero $[bcde]$. Assim, neste problema de tirar a forma de um pentágono ao acaso, teremos uma região para o lançamento do *ponto equivalente* que será constituída por duas coordenadas. A primeira coordenada definirá a forma de um dos cinco ângulos do pentágono e, consequentemente, os dois lados adjacentes a esse vértice. Se, por exemplo, o valor da primeira coordenada fixar o ângulo $\angle[eab]$, os lados $[ea]$ e $[ab]$ estarão fixos e, naturalmente, teremos definido o comprimento do lado $[eb]$, com o qual deveremos tirar, à sorte, as possibilidades do quadrilátero $[ebcd]$. A região na qual vamos lançar o *ponto*

⁽¹³⁾ A Figura 5.23 é distinta da utilizada por Pacheco d'Amorim (1914, Figura 11, p. 69) (cf. Figura 5.22) para explicar o lançamento de um pentágono. Contudo, pensamos que a figura que aqui utilizamos descreve melhor a nossa interpretação do método concebido por Pacheco d'Amorim para a escolha à sorte da forma de um pentágono.

equivalente para definir a forma do quadrilátero $[ebcd]$, conforme previamente analisado, corresponderá à segunda coordenada. Portanto, a região onde é lançado o *ponto equivalente* do pentágono terá uma primeira coordenada que determinará a forma de um dos vértices (logo, de dois lados do pentágono) e uma outra coordenada que determinará a forma do quadrilátero formado pelo lado definido pelo vértice fixo pela primeira coordenada (que une os extremos opostos ao vértice dos dois lados que lhe são adjacentes) e os restantes três lados do pentágono que não têm ligação a esse vértice (logo, dos restantes três lados do pentágono).

Desta forma, supondo o lado $[eb]$ fixo (através de um qualquer valor da primeira coordenada) e fazendo variar o lado $[bc]$, construímos dois conjuntos de possibilidades para o vértice **b** (que fixará um valor para o ângulo $\angle [ebc]$ do quadrilátero e naturalmente para o ângulo $\angle [abc]$ do pentágono), correspondentes à região **B**, constituída pelos valores que tornam possível a formação do quadrilátero $[ebcd]$, e à região **B**₁, constituída pelos valores que tornam possível a formação do quadrilátero $[ebc'd]$. Podemos ainda considerar que o lado fixo do quadrilátero $[bcde]$ não é o lado $[eb]$, mas o lado $[bc]$, construindo as regiões **C** e **C**₁. Assim, para cada forma do vértice **a** teremos as regiões **B**, **B**₁, **C**, **C**₁, **D**, **D**₁, **E** e **E**₁, que dependem do valor do ângulo **a** considerado e que em conjunto correspondem às possibilidades da forma do quadrilátero $[bcde]$.

Associando a cada valor possível para o vértice **a** (primeira coordenada definida na região **A**) as regiões **B**, **C**, **D** e **E**, vamos obter uma região⁽¹⁴⁾ que representaremos por $\mathbb{A} = \{\mathbf{A}, \mathbf{B} \sqcup \mathbf{C} \sqcup \mathbf{D} \sqcup \mathbf{E}\}$. De igual forma, para cada valor possível do vértice **a** associamos as regiões **B**₁, **C**₁, **D**₁ e **E**₁ formando

⁽¹⁴⁾ Notemos que a forma de lançar à sorte um ponto desta região será distinta caso consideremos a região como uma região composta ou como um complexo de regiões. Contudo, Pacheco d'Amorim nada diz sobre este assunto.

a região $A_1 = \{A, B_1 \sqcup C_1 \sqcup D_1 \sqcup E_1\}$. Resta-nos fazer o mesmo com os vértices **b**, **c**, **d** e **e**, de forma a obter as regiões B , B_1 , C , C_1 , D , D_1 , E e E_1 .

Tirar, à sorte, **a forma de um pentágono** $[abcde]$ é, por definição, escolher um ponto em A ou A_1 ou B ou B_1 ou C ou C_1 ou D ou D_1 ou E ou E_1 , isto é, escolher um ponto em

$$T = A \sqcup A_1 \sqcup B \sqcup B_1 \sqcup C \sqcup C_1 \sqcup D \sqcup D_1 \sqcup E \sqcup E_1,$$

onde, tal como na definição do lançamento à sorte da forma de um quadrilátero, a cada ponto da região T é associada uma só forma para o pentágono, mas cada forma possível do polígono terá associados vários pontos da região T .

Pacheco d'Amorim considera que pode ser definida de forma análoga a tiragem, à sorte, da forma de um hexágono, heptágono, etc. De facto, seguindo o raciocínio proposto é possível fazê-lo, mas tal tarefa será indubitavelmente árdua. Notemos que a região onde é lançado o *ponto equivalente* que define a forma de um pentágono apresenta já uma complexidade elevada e, se pensarmos em tirar a forma de um hexágono à sorte, a região terá três coordenadas: a primeira coordenada fixaria a forma de um dos vértices do hexágono, a segunda fixaria a forma de um dos vértices do pentágono (onde um dos lados é definido pelo valor da primeira coordenada) e a terceira coordenada definiria a forma do quadrilátero (cujos lados dependem dos valores das primeiras duas coordenadas). Felizmente que não foi efectuada a análise de um decágono, pois teríamos então de definir uma região de sete coordenadas para o lançamento do *ponto equivalente*. Se continuássemos, para definir um icoságono teríamos 17 coordenadas, para um hectágono 97 coordenadas e para um quilógono teríamos uma região labiríntica com 997 coordenadas!

De facto não nos parece que Pacheco d'Amorim tenha obtido uma forma eficaz de tirar ao acaso a forma de um polígono fechado, pois não só a defini-

ção da região onde é lançado o *ponto equivalente* é bastante delicada, como ainda a complexidade dessa região, quando aumentamos o número de lados do polígono, aumenta significativamente. Por outro lado, a sua apresentação é demasiado obscura para ser possível afirmar, com certeza absoluta, que a descrição aqui efectuada corresponde às ideias expostas pelo autor, pelo menos no que se refere ao lançamento de um pentágono e à generalização. A Figura 5.22 parece-nos carecida de sentido, pois, se o vértice **a** está fixo (logo o lado $[be]$ também), quando analisamos as (infinitas) formas que o quadrilátero $[ebcd]$ pode assumir devemos analisar a forma de cada um dos seus vértices, tal como foi efectuada na definição de tirar, à sorte, a forma de um quadrilátero. Assim, se fixarmos o vértice **a**, teremos duas formas para o quadrilátero que correspondem ao quadrilátero convexo $[ebcd]$ e ao quadrilátero côncavo $[ebc'd]$ representados na Figura 5.23. Na figura apresentada por Pacheco d'Amorim, Figura 5.22, os quadriláteros $[ebcd]$ e $[ebc'd']$ correspondem a dois valores distintos para o vértice **b** e não a dois quadriláteros distintos para um determinado valor do vértice **b**. Por esta razão alteramos a definição de lançamento de um pentágono, pois consideramos que a definição que aqui apresentamos corresponde melhor à ideia inicialmente apresentada por Diogo Pacheco d'Amorim para o lançamento de um quadrilátero.

Definição 5.13 (No espaço). Tirar, à sorte, a forma de uma figura poligonal fechada, com qualquer número finito de vértices, num espaço a n dimensões, com $n > 2$, é efectuada de forma análoga a tirar, à sorte, a forma da mesma figura poligonal fechada num plano (DEFINIÇÃO 5.12) substituindo as circunferências que definem os ângulos dos vértices por hiper-esferas de dimensão n .

5.2.3 Curvas flexíveis e inextensíveis

Para finalizar este capítulo dedicado à probabilidade geométrica, Pacheco d'Amorim define o lançamento, à sorte, de curvas com comprimento fixo, mas cujo número de vértices aumenta indefinidamente.

Definição 5.14. Lançar, à sorte, **uma curva flexível e inextensível (fechada ou aberta) num espaço A** é, por definição, o mesmo que lançar à sorte um polígono do mesmo comprimento e com um número de lados arbitrariamente grande, isto é, um polígono com o mesmo comprimento e com o número de lados a tender para o infinito.

A solução de qualquer problema relacionado com o lançamento de polígonos articulados depende do número de lados considerado e do comprimento de cada lado. O autor refere que se, nessa solução encontrada, determinarmos (se existir) o limite quando o número de lados aumentar indefinidamente (e consequentemente o comprimento de cada lado tender para zero em virtude de o comprimento total ser fixo), esse limite será então a solução do mesmo problema relativo a uma curva flexível e inextensível.

Esta definição será utilizada por Pacheco d'Amorim, no capítulo ESPERANÇA MATEMÁTICA E VALOR MÉDIO⁽¹⁵⁾, para demonstrar que o valor esperado do quadrado da distância que separa os pontos extremos, quando é lançada uma linha curva flexível e inextensível, é nulo.

Apesar de este resultado nos surpreender, notemos que pelo processo descrito nesta definição não é possível obter todas as curvas. Por exemplo, o resultado deste processo será sempre uma curva que não é diferenciável em qualquer ponto. Assim sendo, apesar de Pacheco d'Amorim denominar curva flexível e inextensível, o que poderia levar-nos a pensar que este resultado

⁽¹⁵⁾ Consultar secção 8.4 na página 485.

seria válido para uma qualquer curva lançada, este restringe-se a um tipo de curvas muito específico, àquelas que podem ser obtidas pelo processo transcrito na definição, ou seja, o de considerar um segmento e *dobrá-lo* indefinidamente.

Saliente-se ainda que uma das áreas em que Pacheco d'Amorim mais se notabilizou foi a geometria. Neste campo editou diversas sebtas que, devido à sua excelente qualidade, foram recentemente editadas pela Sociedade Portuguesa da Matemática na colecção *Biblioteca Básica de Textos Didácticos de Matemática*. Refiram-se, então, algumas ideias presentes na primeira edição do *Compêndio de Geometria* de Pacheco d'Amorim que nos permitem perceber melhor as ideias do autor. Para Pacheco d'Amorim os segmentos de recta são a linha mais simples, sendo depois a *linha quebrada*, constituída no mínimo por dois segmentos e, de seguida, a curva,

“porque a linha curva se compõe de segmentos de recta infinitamente pequenos. A linha curva não passa duma linha quebrada com um número infinitamente grande de lados.”

[Pacheco d'Amorim 1920, p. 4]

Somos assim levados a crer que Pacheco d'Amorim considera que todas as curvas podem ser obtidas por este processo. Contudo, analisaremos mais detalhadamente este problema no capítulo 8.

5.3 Comentário geral ao capítulo

O capítulo dos LANÇAMENTO, À SORTE, DE FIGURAS VARIÁVEIS é, na nossa opinião, um dos pontos mais fracos da Tese de Doutoramento de Pacheco d'Amorim. De facto, parece-nos que o autor crê que, com a sequência de definições de lançamentos aleatórios que criou, pode resolver os diversos paradoxos existentes na época em Teoria da Probabilidade, nomeadamente os

referentes a problemas de probabilidade geométrica. Esta crença baseia-se no facto de considerar que através das suas definições só existe uma única solução válida para cada problema e, consequentemente, os paradoxos estão resolvidos. Esta não é, no entanto, a nossa opinião.

Por um lado, não concordamos que as definições fornecidas por Pacheco d'Amorim permitam uma única solução para cada problema, como previamente verificámos no capítulo anterior aquando da análise ao paradoxo de Borel-Kolmogoroff (que também pode ser considerado um problema de lançamento de uma figura, um segmento de recta que corresponde a uma corda, numa região em $\mathbb{R}^3$ que é uma superfície esférica).

Por outro lado, não podemos concordar com algumas definições de lançamento, à sorte, de figuras criadas por Pacheco d'Amorim neste capítulo. Por exemplo, na definição de escolha à sorte de uma recta ou de um segmento de recta numa região (DEFINIÇÃO 5.3 e 5.4), os possíveis resultados não são equipossíveis e, como tal, não podemos considerar que sejam o produto de uma escolha à sorte. O resultado proposto pelo autor para o paradoxo de Bertrand recorre a estas definições e, por este motivo, as cordas escolhidas à sorte no círculo não são igualmente possíveis, razão pela qual consideramos que a sua resolução deste problema não está correcta.

Saliente-se que diversos autores da sua época, tais como Borel (que, sublinhamos, corresponde a uma das poucas referências bibliográficas que Pacheco d'Amorim faz ao longo da sua tese), já possuíam a intuição de que o paradoxo surge da falta de clareza do enunciado pelo facto de este não definir a forma como as cordas são escolhidas à sorte. Pensamos, por conseguinte, que Pacheco d'Amorim se deveria ter apercebido de que o paradoxo surge pelo facto de o enunciado permitir várias interpretações de escolha aleatória das cordas e não por qualquer erro nas soluções apresentadas por Bertrand. Assim, para resolver este paradoxo não seria necessário criar novas definições

de lançamentos de figuras, mas tão somente complementar o enunciado com a forma como a corda seria escolhida à sorte.

Apesar de discordarmos de algumas definições de escolha à sorte criadas por Pacheco d'Amorim, entendemos como positiva a sua resolução do problema da agulha de Buffon, onde salientamos a forma como o autor justifica a utilização de apenas um “*pedaço*” do plano (um paralelograma com uns lados paralelos às linhas da folha de papel e os outros lados paralelos à direcção da agulha) para a sua resolução, não considerando necessário analisar toda a região $\mathbb{R}^2$.

As definições de tiragem, à sorte, da forma de um polígono constituem, na nossa opinião, um emaranhado de ideias de impossível decifração integral. Em primeiro lugar não se compreende o objectivo desta secção, mesmo porque para alcançar o seu objectivo de *racionalização do Cálculo das Probabilidades* esta secção poderia ter sido evitada. Refira-se que o autor não apresenta nenhum exemplo de polígonos com a forma escolhida aleatoriamente, utilizando unicamente no capítulo da ESPERANÇA MATEMÁTICA E VALOR MÉDIO uma destas definições, a do lançamento de uma curva flexível e inextensível num exemplo que analisaremos na secção 8.4. Apesar disso, antecipemos já que nem mesmo esta aplicação poderá ser declarada como um argumento positivo para a existência deste capítulo, pois os resultados deduzidos, nessa aplicação, por Pacheco d'Amorim sobre valores médios são também eles de valor dubitável. Com efeito, considerar que o valor esperado do quadrado da distância que separa os pontos extremos duma curva flexível e inextensível (de comprimento finito), lançada, á sorte, sobre um plano, é nulo, corresponde sem dúvida a um resultado deveras estranho, o que pode reforçar os argumentos favoráveis à inexistência desta secção obscura de lançamentos aleatórios de figuras.

Ao longo da sua exposição o autor nunca refere de que modo o número de

lados, a sua ordem ou o comprimento de cada lado dos polígonos são definidos, limitando-se a analisar os ângulos que os vértices dos polígonos podem assumir. Há ainda algumas restrições, tais como o facto de, no lançamento de figuras poligonais abertas, ao fixar ao acaso a forma de um vértice, este não poder assumir qualquer ângulo⁽¹⁶⁾, que não explica nem justifica. Em relação ao lançamento de polígonos fechados a confusão ainda é maior, sendo tarefa árdua decifrar as ideias do autor.

Em suma, consideramos que o capítulo LANÇAMENTO, À SORTE, DE FIGURAS poderia ter sido evitado, pois nada contribui para a fundamentação do Cálculo das Probabilidades que Pacheco d'Amorim propõe na sua tese. Isto equivale a dizer que, na nossa opinião, a tese de Doutoramento de Pacheco d'Amorim não seria, de forma nenhuma, desvalorizada com o desaparecimento deste capítulo.

⁽¹⁶⁾ O autor, em nota de rodapé, afirma que “*pode não ser a totalidade da superfície da hiper-esfera*”. Contudo, não é detectável a razão desta restrição. Porventura esta nota de rodapé está mal localizada e a restrição indicada não se refere à situação da escolha, à sorte, do ângulo de um vértice de um polígono aberto, mas antes de um polígono fechado. Nesta última situação concordamos que, de facto, os vértices (por vezes) não podem assumir todos os ângulos, como visualizamos nos lançamentos à sorte de quadriláteros.

Capítulo 6

Ponto Imagem

Pacheco d'Amorim dedica o capítulo intitulado PONTO IMAGEM à distinção entre a probabilidade de um ponto lançado directamente, à sorte, numa região (ponto livre), onde a probabilidade é uniforme⁽¹⁾, e a probabilidade dum outro ponto que é sua imagem. Assim, o autor considera que podemos aplicar uma função f , contínua e bijectiva, a um qualquer ponto x lançado directamente à sorte na região $\mathbf{X}$, formando, desta forma, o ponto imagem $y = f(x)$ lançado (indirectamente) à sorte na região $\mathbf{Y} = f(\mathbf{X})$ ⁽²⁾. As probabilidades inerentes ao lançamento à sorte do ponto imagem $y = f(x)$ serão definidas em virtude das probabilidades correspondentes no lançamento à sorte do ponto

⁽¹⁾ Com excepção dos lançamentos em complexos de regiões $\mathbf{X} \odot \mathbf{Y}$, onde a densidade não é uniforme pelo facto de a região na qual o segundo ponto é lançado estar definida condicionalmente ao primeiro ponto escolhido, conforme analisamos no capítulo 4. Contudo, ao longo do presente capítulo, o ponto livre será derivado de um único lançamento numa região. Consequentemente, todos os pontos dessa região terão igual possibilidade, isto é, o ponto livre será necessariamente caracterizado por uma distribuição uniforme definida na região em que o ponto é lançado.

⁽²⁾ Representamos por $\mathbf{Y} = f(\mathbf{X})$ a região constituída pelos pontos y que são imagem de um qualquer ponto x da região $\mathbf{X}$, isto é, $\mathbf{Y} = f(\mathbf{X}) = \{y : y = f(x), x \in \mathbf{X}\}$.

livre x através de $\mathbb{P}[f(\theta_1) \leq f(x) \leq f(\theta_2)] = \mathbb{P}(\theta_1 \leq x \leq \theta_2)$ (no caso de a aplicação f ser crescente). O autor exemplifica esta relação analisando a probabilidade de obter o algarismo d na casa decimal número a num qualquer ponto imagem, apresentando três aplicações, onde investiga, por exemplo, a tábua do logaritmo de base decimal.

Com esta construção a distribuição do ponto imagem depende da função f aplicada, não sendo esta restrita à situação uniforme tal como a distribuição do ponto livre. Por este motivo, para caracterizar o lançamento do ponto imagem, o autor define a lei de possibilidade (função densidade) e a lei de probabilidade (função densidade condicionada) do ponto imagem. Utilizando como base a noção de projecção, define, em relação a um ponto lançado numa região plana, a lei de probabilidade *a priori* e a lei de probabilidade *a posteriori*, que utiliza para deduzir o Teorema da Probabilidade Composta, o Teorema de Bayes e as suas fórmulas inversas para o contexto de lançamentos de pontos em regiões (probabilidade contínua). Por fim, com base nas leis de probabilidade *a priori* e *a posteriori* e no Teorema da Probabilidade Composta, determina integrais definidos numa região plana através de integrais iterados. Desta forma projecta a probabilidade, de uma região plana, num dos eixos (lei *a priori*) e toma a probabilidade da outra variável (correspondente ao outro eixo) condicionalmente a esse valor obtido na projecção (lei *a posteriori*), reinventando deste modo o Teorema de Fubini.

6.1 Definição de Ponto Imagem

Pacheco d'Amorim começa por considerar duas regiões $\mathbf{X}$ e $\mathbf{Y}$ tais que entre os seus pontos se estabeleça uma correspondência biunívoca e contínua através de uma aplicação f . Desta forma, quando se lança à sorte um ponto x na região $\mathbf{X}$, o ponto $y = f(x)$ designa-se por ponto imagem e assume valores

na região $\mathbf{Y} = f(\mathbf{X})$.

Definição 6.1 (Ponto Imagem). Seja f uma aplicação biunívoca e contínua da região $\mathbf{X}$ para a região $\mathbf{Y}$,

$$\begin{aligned} f: \mathbf{X} &\longrightarrow \mathbf{Y} \\ x &\longmapsto y = f(x) \end{aligned} .$$

O ponto x lançado (directamente) à sorte em $\mathbf{X}$ denomina-se por **ponto livre** e o ponto y , definido por $y = f(x)$, que é lançado (de forma indirecta) à sorte em $\mathbf{Y}$, apelida-se por **ponto imagem** ou **ponto sujeito**.

Com esta definição o lançamento à sorte do ponto imagem será definido pelo lançamento à sorte do ponto livre correspondente, lançamento este cujas características foram analisadas detalhadamente no capítulo 4 referente à probabilidade contínua. As propriedades do ponto imagem $y = f(x)$ serão estabelecidas em função das características do correspondente ponto livre. Assim, se no lançamento à sorte do ponto livre x em $\mathbf{X}$ for escolhido o ponto x_0 , significará que no lançamento do ponto imagem y lançado em $\mathbf{Y} = f(\mathbf{X})$ foi escolhido o ponto $y_0 = f(x_0)$ e, conseqüentemente, a possibilidade do ponto y_0 será definida em função da possibilidade do ponto x_0 (como analisaremos em pormenor na secção 6.3, em que comentaremos a definição, concebida por Pacheco d'Amorim, para a lei de possibilidade do ponto imagem y). A possibilidade de uma região $\mathbf{Y}' = f(\mathbf{X}')$, contida em $\mathbf{Y}$, será igual à possibilidade da correspondente região $\mathbf{X}'$, isto é⁽³⁾,

⁽³⁾ Ao longo deste capítulo vamos acrescentar um índice à representação que temos adoptado para a possibilidade de uma região. Assim, para evitar qualquer dúvida sobre a origem da possibilidade de uma região, se esta se refere ao ponto livre x ou ao ponto imagem y , iremos representar por $\varpi_y(\mathbf{Y})$ a possibilidade da região $\mathbf{Y}$ quando o ponto imagem y é lançado e por $\varpi_x(\mathbf{X})$ a possibilidade da região $\mathbf{X}$ quando o ponto livre x é lançado.

$$\varpi_y(\mathbf{Y}') = \varpi_x(f^{-1}(\mathbf{Y}')) = \varpi_x(\mathbf{X}') = \int_{\mathbf{X}'} \pi_x dx \quad (6.1)$$

que, uma vez que no lançamento do ponto livre todos os pontos do espaço amostra possuem a mesma possibilidade, pode ser simplificado para

$$\varpi_y(\mathbf{Y}') = \varpi_x(\mathbf{X}') = \frac{\mu_{\mathbf{X}'}}{\mu_{\mathbf{X}}}. \quad (6.2)$$

Daqui resulta que qualquer possibilidade referente ao ponto imagem y será definida em função da correspondente possibilidade no ponto livre x . Deste modo, a probabilidade, em relação ao ponto imagem y lançado em $\mathbf{Y} = f(\mathbf{X})$, da região $\mathbf{Y}'' = f(\mathbf{X}'')$ em relação à região $\mathbf{Y}' = f(\mathbf{X}')$, será igual à probabilidade correspondente do ponto livre x , isto é,

$$\mathbb{P}_{\mathbf{Y}'}(\mathbf{Y}'') = \frac{\varpi_y(\mathbf{Y}'')}{\varpi_y(\mathbf{Y}')} = \frac{\varpi_x(\mathbf{X}'')}{\varpi_x(\mathbf{X}')} = \mathbb{P}_{\mathbf{X}'}(\mathbf{X}''), \quad (6.3)$$

e, naturalmente, em consequência de (6.2) e (6.3), teremos

$$\mathbb{P}_{\mathbf{Y}'}(\mathbf{Y}'') = \frac{\mu_{\mathbf{X}''}}{\mu_{\mathbf{X}'}}. \quad (6.4)$$

Por este motivo Pacheco d'Amorim considera que as propriedades válidas para o ponto livre x são, de um modo geral, válidas também para o ponto imagem $y = f(x)$.

Deste modo o ponto imagem y , lançado indirectamente, à sorte, na região $\mathbf{Y}$, dependerá também da sorte, dado que depende do ponto x lançado, à sorte, em $\mathbf{X}$. Contudo, o autor destaca as diferenças entre estes dois tipos de lançamentos, considerando que o lançamento de um ponto imagem é muito distinto do lançamento de um ponto livre. Enquanto um ponto livre x tem distribuição uniforme na região $\mathbf{X}$ onde é lançado, a distribuição do ponto imagem y lançado na região $\mathbf{Y}$ dependerá da função f utilizada. Para aclararmos esta ideia, analisemos alguns exemplos.

Exemplo 6.1. Seja x um ponto livre lançado, à sorte, no intervalo $(0, 1)$ e seja $y = -\frac{1}{\alpha} \ln(1 - x)$, com $\alpha > 0$, o seu ponto imagem com suporte em $\mathbb{R}^+$. Qual a probabilidade de o ponto imagem y assumir um valor no intervalo $(y_0, y_1) \subset \mathbb{R}^+$?

A função inversa da função f é $x = f^{-1}(y) = 1 - e^{-\alpha y}$ com domínio $\mathbb{R}^+$. A probabilidade⁽⁴⁾ pretendida será então

$$\begin{aligned} \mathbb{P}_{\mathbb{R}^+}(y_0 < y < y_1) &= \mathbb{P}_{(0,1)}(f^{-1}(y_0) < x < f^{-1}(y_1)) = \\ &= \frac{f^{-1}(y_1) - f^{-1}(y_0)}{1 - 0} = \frac{1 - e^{-\alpha y_1} - (1 - e^{-\alpha y_0})}{1 - 0} = \\ &= e^{-\alpha y_0} - e^{-\alpha y_1}, \end{aligned} \quad (6.5)$$

que corresponde à probabilidade de uma variável aleatória Y com distribuição exponencial de parâmetro α situar-se no intervalo $(y_0, y_1) \subset \mathbb{R}^+$. ♦

Exemplo 6.2. Seja x um ponto livre lançado, à sorte, no intervalo $(0, 1)$ e seja $y = f(x)$ o seu ponto imagem, sendo a função f definida por

$$f(x) = \begin{cases} (b - a) \sqrt{\frac{x}{2}} + a & 0 \leq x \leq 0.5 \\ (a - b) \sqrt{\frac{1-x}{2}} + b & 0.5 < x \leq 1 \end{cases}. \quad (6.6)$$

Qual a probabilidade de o ponto imagem y assumir um valor no intervalo $(y_0, y_1) \subset [a, b]$?

A função inversa de f , função f^{-1} , é dada por

$$f^{-1}(y) = \begin{cases} 2 \frac{(y - a)^2}{(b - a)^2} & a \leq y \leq \frac{a + b}{2} \\ 1 - 2 \frac{(y - b)^2}{(b - a)^2} & \frac{a + b}{2} < y \leq b \end{cases}, \quad (6.7)$$

⁽⁴⁾ Pretendemos avaliar o valor de uma probabilidade incondicionada, como tal, utilizando a nomenclatura adoptada por Pacheco d'Amorim, o que estamos a determinar é a possibilidade de uma região, pois, uma vez que o domínio do ponto imagem é $\mathbb{R}^+$, $\mathbb{P}_{\mathbb{R}^+}(y_0 < y < y_1) = \varpi_y((y_0, y_1))$.

sendo a probabilidade pretendida

$$\begin{aligned}
 \mathbb{P}_{[a,b]}(y_0 < y < y_1) &= \mathbb{P}_{(0,1)}\left(f^{-1}(y_0) < y < f^{-1}(y_1)\right) = \\
 &= \frac{f^{-1}(y_1) - f^{-1}(y_0)}{1 - 0} = f^{-1}(y_1) - f^{-1}(y_0) = \\
 &= \begin{cases} 2 \frac{(y_1 - a)^2 - (y_0 - a)^2}{(b - a)^2} & a \leq y_0 \leq y_1 \leq \frac{a+b}{2} \\
 1 - 2 \frac{(y_1 - b)^2 + (y_0 - a)^2}{(b - a)^2} & a \leq y_0 \leq \frac{a+b}{2} \leq y_1 \leq b \\
 2 \frac{(y_0 - b)^2 - (y_1 - b)^2}{(b - a)^2} & \frac{a+b}{2} \leq y_0 \leq y_1 \leq b \end{cases}, \quad (6.8)
 \end{aligned}$$

que corresponde à probabilidade de uma variável aleatória com distribuição triangular no intervalo $[a, b]$ assumir um valor pertencente ao intervalo $(y_0, y_1) \subset [a, b]$. $\blacklozenge$

A construção apresentada por Pacheco d'Amorim não é restrita a pontos livres lançados no intervalo unitário $(0, 1)$, nem às situações em que a função f é crescente, condições presentes nos exemplos 6.1 e 6.2 por nós descritos e nas três aplicações expostas por Pacheco d'Amorim. Para ilustrar esta ideia procedamos à análise de mais alguns exemplos.

Exemplo 6.3. Seja x um ponto livre lançado, à sorte, no intervalo $(e^{-1}, 1)$ e seja $y = -\log(x)$ o seu ponto imagem com suporte em $(0, 1)$. Qual a probabilidade de o ponto imagem y assumir um valor no intervalo $(y_0, y_1) \subset (0, 1)$?

A função inversa de f é $f^{-1}(y) = e^{-y}$ e a probabilidade pretendida será, uma vez que neste caso a função f é decrescente, dada por

$$\begin{aligned}
 \mathbb{P}_{(0,1)}(y_0 < y < y_1) &= \mathbb{P}_{(e^{-1}, 1)}\left(f^{-1}(y_1) < x < f^{-1}(y_0)\right) = \\
 &= \frac{f^{-1}(y_0) - f^{-1}(y_1)}{1 - e^{-1}} = \frac{e^{-y_0} - e^{-y_1}}{1 - e^{-1}}, \quad (6.9)
 \end{aligned}$$

que corresponde à probabilidade de uma variável aleatória com distribuição exponencial, com parâmetro igual à unidade e condicionada (truncada) de

forma a que só assuma valores inferiores a um, se situar no intervalo $(y_0, y_1) \subset (0, 1)$. $\blacklozenge$

Exemplo 6.4. Seja x um ponto livre lançado, à sorte, no intervalo $(-\frac{\pi}{2}, \frac{\pi}{2})$ e seja $y = \sigma \tan(x)$ o seu ponto imagem. Qual a probabilidade de o ponto imagem y assumir um valor no intervalo $(y_0, y_1) \subset \mathbb{R}$?

A função f tem como inversa a função $x = f^{-1}(y) = \arctan\left(\frac{y}{\sigma}\right)$, sendo a probabilidade pretendida

$$\begin{aligned} \mathbb{P}_{\mathbb{R}}(y_0 < y < y_1) &= \mathbb{P}_{(-\frac{\pi}{2}, \frac{\pi}{2})}\left(f^{-1}(y_0) < x < f^{-1}(y_1)\right) = \\ &= \frac{f^{-1}(y_1) - f^{-1}(y_0)}{\frac{\pi}{2} - \left(-\frac{\pi}{2}\right)} = \frac{\arctan\left(\frac{y_1}{\sigma}\right) - \arctan\left(\frac{y_0}{\sigma}\right)}{\pi}, \end{aligned} \quad (6.10)$$

que corresponde à probabilidade de uma variável aleatória com distribuição de Cauchy, com parâmetro σ , se situar no intervalo $(y_0, y_1) \subset \mathbb{R}$. $\blacklozenge$

Exemplo 6.5. Sejam x um ponto livre lançado, à sorte, no intervalo $(0, 1)$ e $y = \frac{b}{x^{\frac{1}{a}}}$, com a e b valores reais positivos, um seu ponto imagem. Qual a probabilidade de o ponto imagem y assumir um valor no intervalo $(y_0, y_1) \subset (b, +\infty)$?

A função inversa de f é $f^{-1}(x) = \left(\frac{b}{x}\right)^a$ sendo a probabilidade pretendida, uma vez que a função f é decrescente, dada por

$$\begin{aligned} \mathbb{P}_{\mathbb{R}}(y_0 < y < y_1) &= \mathbb{P}_{(0,1)}\left(f^{-1}(y_1) < x < f^{-1}(y_0)\right) = \\ &= \frac{f^{-1}(y_0) - f^{-1}(y_1)}{1 - 0} = \left(\frac{b}{y_0}\right)^a - \left(\frac{b}{y_1}\right)^a, \end{aligned} \quad (6.11)$$

que corresponde à probabilidade de uma variável aleatória com distribuição de Pareto, com parâmetros a e b , se situar no intervalo $(y_0, y_1) \subset (b, +\infty)$. $\blacklozenge$

Conforme foi analisado no capítulo referente à probabilidade contínua (capítulo 4), a definição usualmente utilizada neste contexto é a definição

geométrica de probabilidade, onde é aplicado um raciocínio análogo ao da definição clássica de probabilidade para a situação discreta, sendo o número de elementos substituído por uma medida representativa da sua proporção geométrica (comprimento, área, volume, ...). Desta forma a probabilidade de uma região $\mathbf{A}$ é determinada pelo quociente entre a medida da região favorável $\mathbf{A}$ e a medida da região total possível $\mathbf{\Omega}$, sob a hipótese de a probabilidade ser proporcional à medida e a medida do universo ser finita. Contudo, este raciocínio originou diversos paradoxos, nomeadamente os apresentados por Bertrand (1888). Um destes paradoxos, referido na secção 4.2 (página 241), consiste em atribuir probabilidades distintas a acontecimentos que são equivalentes. Por exemplo, se compararmos a probabilidade de um número x , escolhido à sorte entre 0 e 100, ter valor superior a 50 com a probabilidade de o seu quadrado $y = x^2$, escolhido à sorte entre 0 e 10000, ter valor superior a 2500, concluímos, utilizando sempre a definição geométrica de probabilidade, que

$$\mathbb{P}(50 < x < 100) = \frac{1}{2} \neq \mathbb{P}(2500 < y < 10000) = \frac{3}{4}, \quad (6.12)$$

apesar de $50 < x < 100 \Leftrightarrow 2500 < y < 10000$.⁽⁵⁾ Comentamos, na secção 4.2, as diversas propostas de interpretação, apresentadas por distintos autores contemporâneos a Pacheco d'Amorim, deste *misterioso* resultado. De facto, apesar de Pacheco d'Amorim não efectuar qualquer referência a este paradoxo, a sua construção do ponto imagem resolve-o ao distinguir claramente um ponto livre, lançado directamente numa região, de um outro que é sua imagem. Com esta sua construção a possibilidade dos pontos pertencentes à região $\mathbf{Y}$, onde é lançado (indirectamente) à sorte o ponto imagem y , não são geralmente caracterizados por uma distribuição uniforme⁽⁶⁾. Sa-

⁽⁵⁾ Os valores utilizados são os mencionados por Bertrand (1888, p. 4).

⁽⁶⁾ Só serão caracterizados por uma distribuição uniforme se o ponto imagem y for uma transformação linear de x , isto é, se a função f assumir a forma $f(x) = a + bx$ com a e b

lientamos, desta forma, que o ponto imagem vai ter uma distribuição que depende da função f considerada, não se podendo recorrer à definição geométrica de probabilidade para o cálculo de probabilidades relativas ao ponto imagem, pois esta definição pressupõe igual densidade na região onde o ponto é lançado. Refira-se que, com este método proposto por Pacheco d'Amorim, o ponto imagem pode ser caracterizado por qualquer distribuição⁽⁷⁾, dependendo da função f utilizada. No exemplo 6.1 (página 323) o ponto imagem é caracterizado por uma distribuição exponencial com parâmetro α , isto é, a função f utilizada permite passar de uma distribuição uniforme no intervalo $(0, 1)$ para uma distribuição exponencial com parâmetro α e suporte $\mathbb{R}^+$. O ponto imagem caracterizado no exemplo 6.2 (página 323) segue uma distribuição triangular no intervalo $[a, b]$ obtido também através de uma distribuição uniforme no intervalo $(0, 1)$. No exemplo 6.3 (página 324) o ponto livre é caracterizado por uma distribuição uniforme no intervalo $(e^{-1}, 1)$ obtendo-se um ponto imagem com distribuição exponencial truncada. De forma semelhante no exemplo 6.4 (página 325) resulta uma distribuição de Cauchy com parâmetro σ e no exemplo 6.5 (página 325) uma distribuição de Pareto com parâmetros a e b .

No paradoxo de Bertrand temos $y = x^2$ logo $x = \sqrt{y} = f^{-1}(y)$ (como o ponto x está restrito ao intervalo $(0, 100)$, de valores positivos, a função $f(x) = x^2$ é bijectiva nesta restrição) e a probabilidade retratada, referente à região $(2500, 10000)$, é dada por

$$\begin{aligned} \mathbb{P}_{(0,10000)}(2500 < y < 10000) &= \frac{f^{-1}(10000) - f^{-1}(2500)}{f^{-1}(10000) - f^{-1}(0)} = \\ &= \frac{100 - 50}{100 - 0} = \frac{1}{2} \end{aligned} \quad (6.13)$$

constantes reais.

⁽⁷⁾ Desenvolveremos melhor esta ideia quando analisarmos a definição de lei de possibilidade na secção 6.3.

que, naturalmente, é igual à probabilidade equivalente em relação ao ponto livre x , isto é, a $\mathbb{P}_{(0,100)}(50 < x < 100)$. A construção do ponto imagem efectuada por Pacheco d'Amorim é, de facto, uma ideia extremamente bem concebida e, caso a sua tese tivesse tido impacto internacional, poderia ter motivado uma investigação na época, em relação à ideia de que é possível obter qualquer distribuição através de uma transformação de uma variável uniforme, e constituiria hoje uma referência na história da Teoria da Probabilidade. Saliente-se que a ideia base da concepção do ponto imagem é análoga à do Teorema da Transformação Uniformizante.

Teorema 6.1 (Teorema da Transformação Uniformizante). *Para qualquer variável aleatória contínua X , com função de distribuição $F_X(x)$ estritamente crescente (existindo portanto F_X^{-1}), a variável aleatória $Y = F_X(X)$ tem distribuição uniforme no intervalo $[0, 1]$, i.e.,*

$$Y = F_X(X) \sim \text{Uniforme}(0, 1). \quad (6.14)$$

O resultado do Teorema da Transformação Uniformizante pode ser utilizado na sua forma inversa⁽⁸⁾, isto é, considerando uma variável aleatória com distribuição uniforme no intervalo $(0, 1)$ e definindo a variável aleatória X através de $X = F_X^{-1}(U)$, então X terá função de distribuição F_X . Desta forma podemos obter qualquer distribuição contínua através de uma distribuição uniforme, bastando para tal utilizar o ponto imagem $y = f(x) = F^{-1}(x)$ onde F é a função de distribuição pretendida para o ponto imagem que, em relação às variáveis (absolutamente) contínuas, é sempre invertível (se restrita ao suporte da variável). Saliente-se que a ideia patente neste teorema é o fundamento da Simulação, uma secção que actualmente assume uma importância capital na Teoria da Probabilidade e que, tal como o Teorema da

⁽⁸⁾ Resultado que é habitualmente denominado por método da transformação inversa.

Transformação Uniformizante, emergiu numa fase posterior à tese de doutoramento de Pacheco d'Amorim. Refira-se que existem generalizações do teorema apresentado, que podem ser aplicadas a variáveis cuja função de distribuição não pode ser escrita de forma explícita ou não possui inversa, tal como acontece com as variáveis aleatórias discretas, recorrendo, por exemplo, a funções inversas generalizadas. Assim, na nossa opinião, esta definição de Pacheco d'Amorim, caso tivesse tido a divulgação merecida, poderia ter sido o embrião da investigação desta ideia central de que, através de uma distribuição uniforme, é possível obter qualquer outra distribuição através de uma transformação.

Apesar de, sob um ponto de vista teórico, ser possível obter resultados que nos permitam gerar uma variável com uma qualquer distribuição (absolutamente) contínua a partir de outra variável com distribuição uniforme (no intervalo $[0, 1]$ ou noutra região qualquer), na prática esta tarefa não é fácil, pois nem sempre conhecemos de forma explícita a expressão da função de distribuição F_X e conseqüentemente da sua inversa. Deste modo, este resultado não é *per se* suficiente para conseguirmos simular qualquer variável, pois muitas distribuições de probabilidade, mesmo que nos restrinjamos às variáveis contínuas, têm uma função de distribuição que não é facilmente manejável de forma a obtermos a sua inversa para determinar a função $f = F^{-1}$ a ser utilizada. Para nos apercebermos disto é suficiente pensarmos na distribuição gaussiana, provavelmente a mais usada de entre as distribuições, para logo deduzirmos que esta tarefa pode ser bastante complicada, ou mesmo impossível. Nestes casos, para obtermos valores aproximados dos pretendidos, teríamos que recorrer a métodos numéricos que, sendo na actualidade por vezes de fácil aplicação, uma vez que possuímos computadores capazes de efectuarem cálculos a uma velocidade vertiginosa, na época de Pacheco d'Amorim eram praticamente impossíveis. Apesar de tudo, por vezes, foi

possível tornear o problema com alguma imaginação, como fizeram Box e Müller (1958) que, por ser tarefa complicada obter uma gaussiana através de uma uniforme, conceberam uma forma de obter duas gaussianas independentes com base em duas uniformes independentes, o que actualmente é denominado por método de Box-Muller⁽⁹⁾.

Desta forma, pensamos que, mesmo tendo em consideração as limitações práticas inerentes à inexistência de computadores na altura da apresentação da tese de Pacheco d'Amorim, caso este resultado tivesse usufruído de divulgação internacional deveria ter causado impacto ao nível da investigação efectuada na época, pois este resultado, na suas versões actuais, que são bem mais gerais, assume uma importância vital, sendo o seu aparecimento posterior à tese de doutoramento de Diogo Pacheco d'Amorim.

6.2 Aplicação do Ponto Imagem

Problema 6.1. Seja f uma função contínua e crescente definida no intervalo $\mathbf{X} = (\alpha, \beta)$. Tire-se, à sorte, um número x do intervalo $\mathbf{X} = (\alpha, \beta)$. Qual a probabilidade de que o valor de $y = f(x)$ correspondente tenha o número

⁽⁹⁾ Utilizando a linguagem criada por Pacheco d'Amorim, o método de Box-Muller consiste em lançar, à sorte, um ponto livre $x = (x_1, x_2)$ no intervalo $(0, 1) \times (0, 1)$ e definir o ponto imagem $y = (y_1, y_2)$ através de

$$\begin{cases} y_1 &= \sqrt{-2 \ln(x_1)} \cos(2\pi x_2) \\ y_2 &= \sqrt{-2 \ln(x_1)} \sin(2\pi x_2) \end{cases},$$

onde as suas coordenadas y_1 e y_2 serão caracterizados por uma distribuição gaussiana standard (podendo-se facilmente transformar em duas gaussianas com valor esperado e desvio padrão quaisquer). Há actualmente métodos mais eficientes de obtenção de gaussianas através de uniformes, sendo inclusive possível simular as gaussianas com uma qualquer correlação pretendida.

dígito d na casa decimal de ordem a ?

SOLUÇÃO: Representemos por ω e ω' os números inteiros definidos por

$$\begin{aligned}\omega &= \min \left\{ n \in \mathbb{N} : f(\alpha) \leq \frac{10n + d}{10^a} \right\} \quad \text{e} \\ \omega' &= \max \left\{ n \in \mathbb{N} : f(\beta) \geq \frac{10n + d + 1}{10^a} \right\},\end{aligned}\quad (6.15)$$

e representemos por f^{-1} a função inversa de f (que existe uma vez que a função f é contínua e crescente).

Pacheco d'Amorim define a região favorável, que representamos por $\mathbf{F}$, contida em $\mathbf{Y} = f(\mathbf{X})$, por

$$\mathbf{F} = \bigcup_{i=\omega}^{\omega'} \left[\frac{10i + d}{10^a}, \frac{10i + d + 1}{10^a} \right). \quad (6.16)$$

Notemos que a região favorável $\mathbf{F}^{(10)}$ aqui apresentada pode estar incompleta nos casos em que $f(\alpha)$ ou $f(\beta)$ são números com o dígito d na casa decimal de ordem a . Para visualizarmos esta questão analisemos um exemplo.

Exemplo 6.6. Suponhamos que nos encontramos nas condições do Problema 6.1 com $d = 5$ e $a = 1$, isto é, que pretendemos a probabilidade de que o valor $y = f(x)$ tenha o dígito 5 na primeira casa decimal. Consideremos ainda que, ao aplicarmos a função f , obtemos, como contradomínio, o intervalo $\mathbf{Y} = f(\mathbf{X}) = (f(\alpha), f(\beta)) = (10.55, 100.59)$. Utilizando a expressão (6.16) obtemos como região favorável

$$\mathbf{F} = \bigcup_{i=11}^{99} \left[\frac{10i + 5}{10}, \frac{10i + 6}{10} \right) = [11.5, 11.6) \cup [12.5, 12.6) \cup \dots \cup [99.5, 99.6).$$

No entanto, a região favorável assim obtida não inclui todos os valores pretendidos, faltando incluir nela os intervalos $(10.55, 10.6)$ e $[100.5, 100.59)$.



⁽¹⁰⁾ Esta região só difere da apresentada pelo autor pelo facto de os intervalos apresentados na fórmula (6.16) serem fechados à esquerda, enquanto os utilizados pelo autor são abertos. Contudo, a diferença entre os dois conjuntos é um conjunto de medida nula.

Salientamos que se, por acaso, o contradomínio resultante da aplicação f utilizada no exemplo 6.6 fosse $\mathbf{Y} = f(\mathbf{X}) = (f(\alpha), f(\beta)) = (0.51, 1.59)$, o erro seria enorme, pois a região favorável obtida pela fórmula (6.16) seria igual ao conjunto vazio (logo de probabilidade nula), enquanto na realidade a região favorável deveria ser $(0.51, 0.6) \cup [1.5, 1.59)$, o que representa $\frac{1}{6}$ da amplitude do suporte do ponto imagem y (podendo este valor ter associada uma probabilidade grande ou não consoante a função f em análise).

Todavia, vamos considerar doravante que a região $\mathbf{F}$, definida através da fórmula (6.16), fornece uma boa aproximação para os nossos propósitos e continuemos o desenvolvimento com o intervalo sugerido por Pacheco d'Amorim. Refira-se que esta ideia nos parece razoável para os exemplos analisados pelo autor. Além disso, a correcção da região favorável apenas iria complicar a exposição e nenhuma alteração traria às conclusões desta análise.

Nestas condições a probabilidade pretendida, uma vez que o ponto livre é caracterizado por uma distribuição uniforme em $\mathbf{X}$, será uma consequência do Teorema das Probabilidades Totais, pois

$$\mathbb{P}_{(d,a)} = \varpi_y(\mathbf{F}) = \sum_{i=\omega}^{\omega'} \varpi_y \left(\left[\frac{10i+d}{10^a}, \frac{10i+d+1}{10^a} \right] \right), \quad (6.17)$$

expressão esta que, pela definição de ponto imagem, é equivalente a

$$\mathbb{P}_{(d,a)} = \sum_{i=\omega}^{\omega'} \varpi_x \left(\left[f^{-1} \left(\frac{10i+d}{10^a} \right), f^{-1} \left(\frac{10i+d+1}{10^a} \right) \right] \right) \quad (6.18)$$

e, uma vez que o ponto livre x é caracterizado por uma distribuição uniforme, teremos

$$\mathbb{P}_{(d,a)} = \sum_{i=\omega}^{\omega'} \frac{[f^{-1}(\frac{10i+d+1}{10^a}) - f^{-1}(\frac{10i+d}{10^a})]}{\beta - \alpha}. \quad (6.19)$$

◇

Pacheco d'Amorim ilustra a aplicação da fórmula (6.19), solução do Problema 6.1, com três casos distintos que de seguida comentaremos, após apresentarmos o princípio de Borel (1909a) para a atribuição de probabilidades

através da medida de Lebesgue no intervalo $(0, 1]$ e a sua associação com uma sequência infinita de provas de Bernoulli independentes, pois consideramos que este princípio de Borel possui algumas semelhanças com a aplicabilidade do conceito ponto imagem apresentado por Pacheco d'Amorim.

6.2.1 Princípio de Borel

O princípio de Borel foi introduzido num artigo com extrema importância na história da Teoria da Probabilidade, publicado em 1909, onde, além do princípio de Borel que nesta secção apresentaremos, surge pela primeira vez a aditividade numerável aplicada no cálculo de probabilidades e as primeiras versões dos actualmente denominados Lema de Borel-Cantelli e Lei Forte dos Grandes Números (algumas destas ideias serão analisadas na secção 7.4.1 que principia na página 429). A ideia central deste princípio é criar uma associação biunívoca entre a sequência de dígitos que constituem a representação de um número, do intervalo $[0, 1]$, na base q , com uma sequência infinita de provas de Bernoulli independentes com probabilidade de sucesso constante (igual a $\frac{1}{q}$). Com esta bijecção poderemos resolver problemas de probabilidade contínua através de resultados sobre sequências de provas de Bernoulli independentes e vice-versa, isto é, resolver problemas relativos a provas de Bernoulli através do cálculo de comprimentos de regiões.

Começemos por recordar que é possível representar, utilizando uma base $q \in \mathbb{N}$, qualquer número ω pertencente ao intervalo $[0, 1]$ através de uma série

$$\omega = \sum_{i=1}^{\infty} \frac{b_i}{q^i}, \quad (6.20)$$

onde $b_n \in \{0, 1, \dots, q-1\}$. A sequência $b_1, \dots, b_n, \dots$ corresponde à representação de base q do número ω . Habitualmente utilizamos a base decimal,

isto é, representamos os números através de

$$\omega = \sum_{i=1}^{\infty} \frac{b_i}{10^i}. \quad (6.21)$$

Borel (1909a) pretende associar a escolha aleatória de um número no intervalo $[0, 1]$ com o resultado de uma sequência infinita de provas de Bernoulli independentes (cujos resultados definem os valores $b_i, i = 1, 2, \dots$), tendo subjacentes as hipóteses de que

1. as provas são independentes (os dígitos b_i que formam a representação do número do intervalo $[0, 1]$ são independentes);
2. cada dígito possível $(0, 1, \dots, q-1)$ tem probabilidade $\frac{1}{q}$ de ocorrer em cada prova.

Desta forma, a cada sequência (ou conjunto de sequências) de dígitos está associado um intervalo de $[0, 1]$ e, sob as hipóteses enunciadas, Borel considera que a probabilidade associada a esta sequência é igual ao comprimento do intervalo formado.

O princípio de Borel é habitualmente apresentado recorrendo à representação diádica (binária) do número $\omega^{(11)}$, isto é, utilizando $q = 2$ nas ideias previamente expostas. Assim, Borel faz equivaler a escolha à sorte de um número no intervalo $[0, 1]$ com uma sequência de provas de Bernoulli independentes com probabilidade de sucesso $\frac{1}{2}$. Concentremos a nossa atenção neste caso particular. Consideremos uma experiência aleatória $\mathcal{E}$ que consiste numa sequência infinita de lançamentos de uma moeda (infinitas provas independentes de Bernoulli com probabilidade de sucesso $p = 0.5$) representados pelas variáveis

$$X_1, X_2, \dots, X_n, \dots, \quad (6.22)$$

⁽¹¹⁾ Consultar, por exemplo, Adams e Guillemin (1996) ou Billingsley (1986).

onde X_i representa o i -ésimo lançamento da moeda. A cada lançamento X_i associamos um de dois possíveis resultados: cara, ao qual associamos o número um à variável; e coroa ao qual associamos o número zero. O conjunto de todos os possíveis resultados (possíveis sequências) da experiência aleatória $\mathcal{E}$ representamos por $\mathcal{B}$. Desta forma, uma concretização da experiência aleatória $\mathcal{E}$ constituirá uma sequência infinita de zeros e uns,

$$0101101001011110101000101010\dots$$

a que podemos sempre associar um número real do intervalo $[0, 1]$ fazendo

$$\omega = \sum_{i=1}^{\infty} \frac{x_i(\omega)}{2^i}, \quad (6.23)$$

onde $x_i(\omega) \in \{0, 1\}$ representa a concretização da variável aleatória $X_i(\omega)$. Para facilitar a exposição, por vezes representaremos esta sequência através de

$$\omega = .x_1(\omega) x_2(\omega) x_3(\omega) \cdots x_n(\omega) \cdots. \quad (6.24)$$

Desta forma a cada concretização da experiência aleatória $\mathcal{E}$ temos associado um único número do intervalo $[0, 1]$. Contudo, nem todos os número do intervalo $[0, 1]$ assumem uma só representação diádica, pois os números que têm representação binária degenerada (isto é, que a partir de determinado valor de $i \in \mathbb{N}$ verificam $x_i = 0$) também podem ser representados de forma não degenerada, pois

$$.x_1(\omega) \dots x_{n-1}(\omega) 100000 \cdots = .x_1(\omega) \dots x_{n-1}(\omega) 011111 \cdots,$$

uma vez que

$$\frac{1}{2^n} = \sum_{i=n+1}^{\infty} \frac{1}{2^i}. \quad (6.25)$$

Para que seja possível criar uma bijecção entre estes dois conjuntos, vamos restringir-nos à representação não degenerada, pois desta forma todos os nú-

mero do intervalo $(0, 1]$ terão uma única representação binária⁽¹²⁾. Seja então $\mathcal{B}'$ o conjunto de todas as sequências de Bernoulli não degeneradas, isto é, $\mathcal{B}'$ é igual a $\mathcal{B}$ retirando as sequências que, a partir de determinada ordem, só assumem zero. Com esta restrição podemos então definir uma aplicação bijectiva $\mathcal{B}' \leftrightarrow (0, 1]$, isto é, a cada possível sequência da experiência aleatória $\mathcal{E}$ associar um único número real no intervalo $(0, 1]$ e a cada número pertencente a este intervalo ter associada somente uma sequência possível para a experiência $\mathcal{E}$, essencial para enunciarmos o princípio de Borel. Notemos que, com esta restrição a expressão (6.23) fica bem definida, uma vez que para cada $\omega \in (0, 1]$ existe uma única sequência

$$x_1(\omega) x_2(\omega) \dots x_n(\omega) \dots \quad (6.26)$$

correspondente à representação binária (não degenerada) do número ω .

Definição 6.2 (Princípio de Borel). Seja $\mathbf{A}$ um acontecimento associado a um determinado conjunto de sequências de Bernoulli $\mathcal{B}_{\mathbf{A}} \in \mathcal{B}'$. Seja $\mathbf{I}_{\mathbf{A}} \subset (0, 1]$ o intervalo associado a $\mathcal{B}_{\mathbf{A}}$, então

$$\mathbb{P}(\mathbf{A}) = \lambda(\mathbf{I}_{\mathbf{A}}), \quad (6.27)$$

onde λ representa a medida de Lebesgue.

Recordemos que a medida de Lebesgue⁽¹³⁾, restringindo a nossa análise a uma dimensão, corresponde ao comprimento do intervalo, ou seja,

$$\lambda((a, b]) = b - a, \quad (6.28)$$

⁽¹²⁾ Desta forma retiramos o número zero da nossa análise e trabalharemos com o intervalo $(0, 1]$, uma vez que o número zero não assume representação não degenerada.

⁽¹³⁾ Esta medida foi apresentada em 1902 por Henri Lebesgue (1875–1941) na sua tese de doutoramento intitulada *Intégrale, longueur, aire*.

utilizando o conceito de aditividade numerável ou σ -aditividade criada por Borel na sua tese de doutoramento em 1895, isto é, a medida de Lebesgue de uma união numerável de intervalos disjuntos é igual à soma numerável das suas medidas,

$$\lambda \left(\bigcup_{i=1}^{\infty} \mathbf{I}_i \right) = \sum_{i=1}^{\infty} \lambda(\mathbf{I}_i), \quad (6.29)$$

onde $\mathbf{I}_1, \mathbf{I}_2, \dots, \mathbf{I}_n, \dots$ representam intervalos disjuntos dois a dois ($\mathbf{I}_i \cap \mathbf{I}_j = \emptyset$, $\forall i \neq j$). Salientemos então algumas propriedades deste princípio de atribuição de probabilidades criado por Borel no seu famoso artigo de 1909.

Se fixarmos o primeiro dígito na representação diádica, por exemplo $X_1(\omega) = x_1(\omega) = 1$, teremos que desta forma podemos obter todos os números situados (continuando em numeração binária) entre $0.1 = 0.01111 \dots$ (não incluindo este ponto, pois na representação não degenerada o primeiro dígito é zero) e $0.1111 \dots$ que corresponde então ao intervalo $(0.5, 1]$. Assim, a probabilidade de o primeiro dígito ser igual a um é dada por

$$\mathbb{P}(\omega : X_1 = 1) = \lambda((0.5, 1]) = 0.5. \quad (6.30)$$

Aplicando raciocínio análogo podemos deduzir que, de uma forma geral, se fixarmos os primeiros n dígitos do número ω , teremos a correspondência

$$\begin{aligned} & \{\omega : X_i(\omega) = x_i(\omega), i = 1, \dots, n\} = \\ & = \left[\sum_{i=1}^n \frac{x_i(\omega)}{2^i}, \sum_{i=1}^n \frac{x_i(\omega)}{2^i} + \frac{1}{2^n} \right]. \end{aligned} \quad (6.31)$$

Por conseguinte, a probabilidade de qualquer sequência, quando consideramos fixos os primeiros n dígitos, é dada por

$$\begin{aligned} & \mathbb{P}(\{\omega : X_i(\omega) = x_i(\omega), i = 1, \dots, n\}) = \\ & = \lambda \left(\left[\sum_{i=1}^n \frac{x_i(\omega)}{2^i}, \sum_{i=1}^n \frac{x_i(\omega)}{2^i} + \frac{1}{2^n} \right] \right) = \frac{1}{2^n}. \end{aligned} \quad (6.32)$$

Naturalmente, para qualquer valor de n , teremos

$$\mathbb{P}(\omega : X_n = 0) = \mathbb{P}(\omega : X_n = 1) = \frac{1}{2^n} 2^{n-1} = \frac{1}{2}, \quad (6.33)$$

uma vez que, de entre as possíveis 2^n representações utilizando n dígitos existem 2^{n-1} representações que verificam $X_n = 0$ (e outras 2^{n-1} representações em que se verifica $X_n = 1$).

Se representarmos por S_n a soma dos primeiros n dígitos, i.e.,

$$S_n(\omega) = \sum_{i=1}^n X_i(\omega), \quad (6.34)$$

existem $\binom{n}{k}$ sequências distintas que satisfazem a condição $S_n = k$ (com $0 \leq k \leq n$) e, conseqüentemente, uma vez que cada possível sequência tem associado um intervalo disjunto dos intervalos associados às restantes sequências, pela aditividade da medida de Lebesgue obtemos

$$\mathbb{P}(\omega : S_n = k) = \binom{n}{k} \frac{1}{2^n}, \quad (6.35)$$

que corresponde à probabilidade de ocorrerem k sucessos em n provas de Bernoulli independentes com probabilidade de sucesso igual a $\frac{1}{2}$.

É possível, com base neste princípio, demonstrar resultados mais elaborados que os exemplos que acabamos de expor. Na secção 7.4.1, referente aos números normais de Borel, desenvolveremos algumas destas situações, nomeadamente a demonstração da Lei Fraca dos Grandes Números e da Lei Forte dos Grandes Números para esta situação particular de somas de provas de Bernoulli independentes com probabilidade de sucesso igual a $\frac{1}{2}$, o que nos permitirá deduzir que a probabilidade de, ao escolhermos um número à sorte no intervalo $(0, 1]$, este ser Normal é igual à unidade e, por esta razão, a probabilidade de o número ser racional é nula.

Nesta secção gostaríamos de salientar as semelhanças entre este princípio de Borel, provavelmente o autor que mais influenciou Pacheco d'Amorim, e as aplicações que Pacheco d'Amorim efectua do ponto imagem. No princípio de Borel é feita uma correspondência biunívoca (obtida através da imposição de uma restrição à representação diádica do número) entre uma sucessão de

provas, cada uma com q resultados equiprováveis, e um número no intervalo $(0, 1]$, sendo q a base em que representamos esse número. Deste modo, uma vez definido o intervalo $\mathbf{A} \subset (0, 1]$ favorável para o nosso problema, a probabilidade deste acontecimento corresponderá à medida de Lebesgue desse intervalo. Se considerarmos $q = 10$, estaremos a trabalhar com a base decimal, situação que nos permite investigar a probabilidade de cada número, proveniente de uma escolha à sorte no intervalo $(0, 1]$ possuir um determinado dígito d na casa decimal de ordem a . Este problema é semelhante ao aqui apresentado por Pacheco d'Amorim. No Problema 6.1 (página 330) pretendia-se analisar a probabilidade de que o número y , onde $y = f(x)$, escolhido à sorte (indirectamente) em $\mathbf{Y} = f(\mathbf{X})$ ter o número dígito d na casa decimal de ordem a . Assim, se na fórmula (6.19) (página 332) considerarmos a transformação $y = x$ e o ponto livre for lançado no intervalo $\mathbf{X} = (0, 1]$, esta corresponderá à definição do princípio de Borel (fórmula (6.27)), uma vez que nestas condições teremos $\beta - \alpha = 1$ e $f^{-1}(y) = x$. Consequentemente da expressão (6.19) obtemos

$$\mathbb{P}_{(d,a)} = \sum_{i=\omega}^{\omega'} \frac{\left[\left(\frac{10i+d+1}{10^a}\right) - \left(\frac{10i+d}{10^a}\right)\right]}{1} = \sum_{i=\omega}^{\omega'} \frac{1}{10^a}. \quad (6.36)$$

Esta expressão corresponde a somar as amplitudes dos intervalos constituídos pelos números que possuem o dígito d na casa decimal número a . Esta semelhança é mais notória se analisarmos as expressões (6.17) e (6.18) (consultar página 332), pois

$$\mathbb{P}_{(d,a)} = \varpi_y(\mathbf{F}) = \varpi_x \left[f^{-1}(\mathbf{F}) \right] = \varpi_x(\mathbf{F}) \quad (6.37)$$

onde ϖ_x representa a possibilidade no ponto livre que, uma vez que é caracterizado pela equipossibilidade, é determinado pela definição geométrica de probabilidade, isto é, pelo quociente entre a medida da região favorável e a medida da região possível. Se $\mathbf{X}$ corresponder ao intervalo $(0, 1]$ (ou qualquer

outra região com medida unitária), então teremos

$$\varpi_x(\mathbf{F}) = \lambda(\mathbf{F}). \quad (6.38)$$

Caso a região possível não tenha uma medida igual à unidade, então, para normalizar a probabilidade, teremos a solução de Pacheco d'Amorim do Problema 6.1 igual a

$$\varpi_x(\mathbf{A}) = \frac{\lambda(\mathbf{A})}{\lambda(\mathbf{X})}. \quad (6.39)$$

Desta forma podemos considerar que o princípio de atribuição de probabilidade de Borel é semelhante à aplicação do ponto imagem apresentada por Pacheco d'Amorim, sendo esta última mais geral, pois nos permite não só analisar o comportamento dos dígitos (na representação decimal) dos números resultantes de uma escolha aleatória no intervalo $(0, 1]$, mas também investigar o comportamento dos dígitos de uma transformação, obtida através da aplicação de uma função bijectiva f , de um número escolhido à sorte de um intervalo qualquer limitado $\mathbf{X}$.

Contudo, Pacheco d'Amorim não efectua, neste capítulo, qualquer referência de que os dígitos que constituem a sequência que representa o número escolhido aleatoriamente possam ter origem numa sequência infinita de provas de Bernoulli independentes, como acontece no princípio de Borel. No capítulo TEOREMAS DE BERNOULLI E LEI DOS DESVIOS o autor, na resolução de um problema relacionado com a probabilidade de ao ser escolhido um número do intervalo $(0, 1)$ sair um número racional, efectua a correspondência entre um número escolhido à sorte no intervalo $(0, 1)$ e a sua representação em base decimal, onde considera que cada dígito pode ser considerado proveniente de uma prova de Bernoulli com dez resultados equipossíveis. Porém, adiamos esta análise para a secção 7.4.1 (página 429).

6.2.2 Primeiro caso

A primeira situação apresentada por Pacheco d'Amorim refere-se ao ponto imagem definido por

$$y = \log_{\alpha} x, \quad (6.40)$$

onde o ponto livre x é lançado à sorte na região $\mathbf{X}$ e, portanto, o ponto imagem y é lançado na região $\mathbf{Y} = f(\mathbf{X})$, que consideremos assumir a forma $\mathbf{Y} = [0, \frac{10k+10}{10^a}]$. Assim, no que se refere à aplicação da fórmula (6.19), nesta situação temos $\omega = 0$ e $\omega' = k$ e, uma vez que a função inversa da função f é $x = f^{-1}(y) = \alpha^y$, obtemos

$$\begin{aligned} \mathbb{P}_{(d,a)} &= \frac{\sum_{i=0}^k \left[\alpha^{\frac{10i+d+1}{10^a}} - \alpha^{\frac{10i+d}{10^a}} \right]}{\alpha^{\frac{10k+10}{10^a}} - 1} = \frac{\alpha^{\frac{d}{10^a}} \sum_{i=0}^k \left[\alpha^{\frac{10i+1}{10^a}} - \alpha^{\frac{10i}{10^a}} \right]}{\alpha^{\frac{10k+10}{10^a}} - 1} = \\ &= \alpha^{\frac{d}{10^a}} \frac{\sum_{i=0}^k \alpha^{\frac{i}{10^{a-1}}} \left[\alpha^{\frac{1}{10^a}} - 1 \right]}{\alpha^{\frac{10k+10}{10^a}} - 1} = \alpha^{\frac{d}{10^a}} \frac{\left[\alpha^{\frac{1}{10^a}} - 1 \right] \sum_{i=0}^k \alpha^{\frac{i}{10^{a-1}}}}{\alpha^{\frac{k+1}{10^{a-1}}} - 1} = \\ &= \alpha^{\frac{d}{10^a}} \frac{\left[\alpha^{\frac{1}{10^a}} - 1 \right] \times \frac{1 - \alpha^{\frac{k+1}{10^{a-1}}}}{1 - \alpha^{\frac{1}{10^{a-1}}}}}{\alpha^{\frac{k+1}{10^{a-1}}} - 1} = \alpha^{\frac{d}{10^a}} \frac{\alpha^{\frac{1}{10^a}} - 1}{\alpha^{\frac{1}{10^{a-1}}} - 1}. \end{aligned} \quad (6.41)$$

Desta forma o autor conclui que a probabilidade não depende do valor de k e conseqüentemente também não depende da amplitude do intervalo em questão.

Através da fórmula (6.41) podemos determinar o valor do rácio entre as probabilidades de dois valores consecutivos para o dígito d , obtendo-se

$$\frac{\mathbb{P}_{(d+1,a)}}{\mathbb{P}_{(d,a)}} = \alpha^{\frac{1}{10^a}} \quad (6.42)$$

que é independente do valor de d . Podemos também, com base no resultado (6.41), determinar o limite de $\mathbb{P}_{(d,a)}$ quando a aumenta indefinidamente,

verificando-se

$$\begin{aligned} \lim_{a \rightarrow +\infty} \mathbb{P}_{(d,a)} &= \lim_{a \rightarrow +\infty} \alpha^{\frac{d}{10^a}} \frac{\alpha^{\frac{1}{10^a}} - 1}{\alpha^{\frac{1}{10^{a-1}}} - 1} = \lim_{a \rightarrow +\infty} \alpha^{\frac{d}{10^a}} \lim_{a \rightarrow +\infty} \frac{\alpha^{\frac{1}{10^a}} - 1}{\alpha^{\frac{1}{10^{a-1}}} - 1} = \\ &= \lim_{a \rightarrow +\infty} \frac{\alpha^{\frac{1}{10^a}} \frac{1}{10^a} \ln \alpha \ln \frac{1}{10}}{\alpha^{\frac{1}{10^{a-1}}} \frac{1}{10^{a-1}} \ln \alpha \ln \frac{1}{10}} = \lim_{a \rightarrow +\infty} \frac{\alpha^{\frac{1}{10^a}}}{\alpha^{\frac{1}{10^{a-1}}}} \times \frac{1}{10} = \frac{1}{10}, \end{aligned} \quad (6.43)$$

que não depende do valor de d nem da base α considerada no logaritmo. O autor deduz deste resultado que, apesar de haver diferenças entre as probabilidades para valores distintos de d (como deduzimos no cálculo do rácio de duas probabilidades com o mesmo valor de a mas valores de d consecutivos), estas tendem todas para o mesmo valor quando a tende para infinito.

6.2.3 Segundo caso

A segunda função analisada por Pacheco d'Amorim corresponde à função

$$y = \alpha^x, \quad \alpha > 1, \quad (6.44)$$

para o ponto livre x lançado à sorte em $\mathbf{X} = [\log_{\alpha}(\frac{10\omega}{10^a}), \log_{\alpha}(\frac{10\omega'+10}{10^a})]$, correspondendo para o ponto imagem um suporte $\mathbf{Y} = [\frac{10\omega}{10^a}, \frac{10\omega'+10}{10^a}]$. Assim, recorrendo uma vez mais à fórmula (6.19), obtemos

$$\begin{aligned} \mathbb{P}_{(d,a)} &= \frac{\sum_{k=\omega}^{\omega'} [\log_{\alpha}(\frac{10k+d+1}{10^a}) - \log_{\alpha}(\frac{10k+d}{10^a})]}{\log_{\alpha}(\frac{10\omega'+10}{10^a}) - \log_{\alpha}(\frac{10\omega}{10^a})} = \\ &= \frac{\sum_{k=\omega}^{\omega'} \log_{\alpha}[1 + \frac{1}{10k+d}]}{\log_{\alpha} \frac{\omega'+1}{\omega}} \end{aligned} \quad (6.45)$$

de onde se conclui que a probabilidade pretendida não depende da casa decimal utilizada, isto é, do valor de a considerado. Outra conclusão apresentada por Pacheco d'Amorim refere-se ao valor do limite de $\mathbb{P}_{(d,a)}$ quando a amplitude do intervalo de variação do ponto imagem y (suporte de y) aumenta, isto é, quando aumentamos o valor de ω' , obtendo-se

$$\lim_{\omega' \rightarrow +\infty} \mathbb{P}_{(d,a)} = \frac{1}{10}. \quad (6.46)$$

Demonstração. Consideremos a sucessão $U_n(d)$ com termo geral definido através de

$$U_n(d) = \frac{\sum_{k=\omega}^n \log_{\alpha} \left(1 + \frac{1}{10k+d} \right)}{\sum_{k=\omega}^n \log_{\alpha} \left(1 + \frac{1}{k} \right)}. \quad (14) \quad (6.46)$$

- A sucessão U_n é crescente quando $d > 9$ (d pode ser um número qualquer na expressão de U_n).

Se considerarmos a função

$$f(n) = \frac{\log_{\alpha} \left(1 + \frac{1}{10n+d} \right)}{\log_{\alpha} \left(1 + \frac{1}{n} \right)} \quad (6.47)$$

temos que a sua derivada é igual a

$$f'(n) = \frac{\frac{-10}{(10n+d)(10n+d+1)} \log_{\alpha} \left(1 + \frac{1}{n} \right) + \frac{1}{n(n+1)} \log_{\alpha} \left(1 + \frac{1}{10n+d} \right)}{\log_{\alpha} \cdot \log_{\alpha}^2 \left(1 + \frac{1}{n} \right)}. \quad (6.48)$$

Uma vez que

$$\left(1 + \frac{1}{n} \right)^n \quad (6.49)$$

é uma função crescente⁽¹⁵⁾ podemos concluir que

$$\left(1 + \frac{1}{10n+d} \right)^{10n+d} > \left(1 + \frac{1}{n} \right)^n \quad (6.50)$$

⁽¹⁴⁾ Notemos que o denominador desta expressão é

$$\begin{aligned} \sum_{k=\omega}^n \log_{\alpha} \left(1 + \frac{1}{k} \right) &= \sum_{k=\omega}^n [\log_{\alpha} (k+1) - \log_{\alpha} (k)] = \\ &= \log_{\alpha} (n+1) - \log_{\alpha} (\omega) = \log_{\alpha} \left(\frac{n+1}{\omega} \right), \end{aligned}$$

e, como tal, verifica-se $U_n(d) = \mathbb{P}_{(d,a)}$ onde n substitui o valor de ω' .

⁽¹⁵⁾ Este resultado pode ser deduzido pela desigualdade de Bernoulli,

$$(1+x)^n \geq 1+nx, \forall n \in \mathbb{N} \text{ e } x \geq -1$$

onde a igualdade se verifica unicamente nos casos em que $n = 1$ ou $x = 0$. Esta desigualdade pode ser demonstrada por indução, pois para $n = 1$ verifica-se facilmente e, supondo

e também, considerando que $d \geq 9$

$$\begin{aligned}
 & \left\{ \left(1 + \frac{1}{10n+d} \right)^{10n+d} \right\}^{10n+d+1} > \left\{ \left(1 + \frac{1}{n} \right)^n \right\}^{10n+10} \Leftrightarrow \\
 \Leftrightarrow & \log_{\alpha} \left[\left\{ \left(1 + \frac{1}{10n+d} \right)^{10n+d} \right\}^{10n+d+1} \right] > \log_{\alpha} \left[\left\{ \left(1 + \frac{1}{n} \right)^n \right\}^{10n+10} \right] \Leftrightarrow \\
 \Leftrightarrow & (10n+d)(10n+d+1) \log_{\alpha} \left(1 + \frac{1}{10n+d} \right) > \\
 & > 10n(n+1) \log_{\alpha} \left(1 + \frac{1}{n} \right) \Leftrightarrow \\
 \Leftrightarrow & \frac{1}{n(n+1)} \log_{\alpha} \left(1 + \frac{1}{10n+d} \right) > \frac{10}{(10n+d)(10n+d+1)} \log_{\alpha} \left(1 + \frac{1}{n} \right) \Leftrightarrow \\
 \Leftrightarrow & \frac{-10}{(10n+d)(10n+d+1)} \log_{\alpha} \left(1 + \frac{1}{n} \right) + \frac{1}{n(n+1)} \log_{\alpha} \left(1 + \frac{1}{10n+d} \right) > 0.
 \end{aligned}$$

Assim o numerador de $f'(n)$ é positivo, faltando-nos analisar o denominador. Uma vez que $\alpha > 1$, temos que $\log \alpha \cdot \log_{\alpha}^2 \left(1 + \frac{1}{n} \right) > 0$ e conseqüentemente $f'(n) > 0$. Notemos que o numerador de U_n é a soma dos numeradores de $f(1), f(2), \dots, f(n)$ e o denominador de U_n é a soma dos denominadores de $f(1), f(2), \dots, f(n)$, e, como $f(n)$ é crescente (e positivo), U_n também o é. Para verificar esta propriedade representemos o numerador de $f(k)$ por n_k e o denominador por d_k (onde $n_k > 0$ e $d_k > 0$), isto é, $f(k) = \frac{n_k}{d_k}$. Como f é uma função crescente, temos que

$$\frac{n_1}{d_1} \leq \frac{n_2}{d_2} \leq \dots \leq \frac{n_k}{d_k}, \quad (6.51)$$

verdadeira para um qualquer valor n , verificar-se-á também para $n+1$, pois

$$(1+x)^{n+1} = (1+x)(1+x)^n \geq (1+x)(1+nx) = 1 + (n+1)x + nx^2 \geq 1 + (n+1)x.$$

Deste resultado podemos deduzir que $u_k = \left(1 + \frac{1}{k} \right)^k$, para $k \in \mathbb{N}$, é crescente, uma vez que

$$u_{k+1} > u_k \Leftrightarrow \left(\frac{1 + \frac{1}{k+1}}{1 + \frac{1}{k}} \right)^{k+1} > \left(1 + \frac{1}{k} \right)^{-1} \Leftrightarrow \left(1 - \frac{1}{(k+1)^2} \right)^{k+1} > 1 - \frac{1}{k+1}$$

que corresponde à desigualdade de Bernoulli com $x = -\frac{1}{(k+1)^2} \geq -1$ e $n = k+1$ (sendo a desigualdade estrita, dado que $n \neq 1$ e $x \neq 0$).

que é suficiente para garantir que $U_k = \frac{\sum_{i=1}^k n_i}{\sum_{i=1}^k d_i}$ também o é. Podemos facilmente deduzir esta conclusão, pois, dado que n_k e d_k são positivos para qualquer valor de $k \in \mathbb{N}$, teremos

$$\frac{n_1}{d_1} \leq \frac{n_2}{d_2} \Leftrightarrow \frac{n_1}{d_1} \leq \frac{n_1 + n_2}{d_1 + d_2} \leq \frac{n_2}{d_2}, \quad (6.52)$$

pois

$$\begin{aligned} \frac{n_1}{d_1} \leq \frac{n_1 + n_2}{d_1 + d_2} &\Leftrightarrow n_1 (d_1 + d_2) \leq d_1 (n_1 + n_2) \Leftrightarrow \frac{n_1}{d_1} \leq \frac{n_2}{d_2} \\ \frac{n_1 + n_2}{d_1 + d_2} \leq \frac{n_2}{d_2} &\Leftrightarrow d_2 (n_1 + n_2) \leq n_2 (d_1 + d_2) \Leftrightarrow \frac{n_1}{d_1} \leq \frac{n_2}{d_2}. \end{aligned}$$

Este resultado garante que, se tivermos duas fracções com numerador e denominador positivos e construirmos uma terceira fracção cujo numerador é a soma dos numeradores das primeiras duas e o denominador é igual à soma dos denominadores das primeiras duas, então a terceira fracção assumirá um valor intermédio das duas primeiras fracções. Com base neste resultado podemos demonstrar, por indução, que U_k é crescente. Notemos que a expressão (6.52) é equivalente a ter $U_1 \leq U_2 \leq \frac{n_2}{d_2}$, pelo que se verifica $U_2 \leq \frac{n_3}{d_3}$ pois $\frac{n_2}{d_2} \leq \frac{n_3}{d_3}$. Como tal, aplicando o resultado patente na fórmula (6.52), teremos

$$U_2 \leq \frac{n_3}{d_3} \Leftrightarrow \frac{\sum_{i=1}^2 n_i}{\sum_{i=1}^2 d_i} \leq \frac{\sum_{i=1}^2 n_i + n_3}{\sum_{i=1}^2 d_i + d_3} \leq \frac{n_3}{d_3} \Leftrightarrow U_2 \leq U_3 \leq \frac{n_3}{d_3}.$$

Assim, supondo $U_{k-1} \leq U_k \leq \frac{n_k}{d_k}$ e consequentemente $U_k \leq \frac{n_{k+1}}{d_{k+1}}$, podemos recorrer, uma vez mais, ao resultado (6.52) e concluir que

$$U_k \leq \frac{n_{k+1}}{d_{k+1}} \Leftrightarrow \frac{\sum_{i=1}^k n_i}{\sum_{i=1}^k d_i} \leq \frac{\sum_{i=1}^k n_i + n_{k+1}}{\sum_{i=1}^k d_i + d_{k+1}} \leq \frac{n_{k+1}}{d_{k+1}} \Leftrightarrow U_k \leq U_{k+1} \leq \frac{n_{k+1}}{d_{k+1}}.$$

Logo $U_k \leq U_{k+1}, \forall k \in \mathbb{N}$, isto é, U_n é crescente.

- $U_n(d) < 1, \forall n \in \mathbb{N}$. Para demonstrar esta propriedade é suficiente notar que

$$\begin{aligned}
& \log_{\alpha} \left(1 + \frac{1}{10n+d} \right) < \log_{\alpha} \left(1 + \frac{1}{n} \right) \Leftrightarrow \\
& \Leftrightarrow \sum_{k=\omega}^n \log_{\alpha} \left(1 + \frac{1}{10n+d} \right) < \sum_{k=\omega}^n \log_{\alpha} \left(1 + \frac{1}{n} \right) \Leftrightarrow \\
& \Leftrightarrow U_n = \frac{\sum_{k=\omega}^n \log_{\alpha} \left(1 + \frac{1}{10n+d} \right)}{\sum_{k=\omega}^n \log_{\alpha} \left(1 + \frac{1}{n} \right)} < 1.
\end{aligned} \tag{6.53}$$

Deste modo deduzimos que $U_n(d)$ é crescente (considerando $d \geq 9$) e limitado ($U_n(d) < 1$), logo tende para um limite quando n tende para infinito.

Uma vez que U_n verifica, para $0 \leq d \leq 10$,

$$U_n(0) \geq U_n(d) \geq U_n(10),^{(16)}$$

deste facto podemos concluir que, na referida restrição, temos

$$\begin{aligned}
& U_n(d) - U_n(10) \leq U_n(0) - U_n(10) = \\
& = \frac{\sum_{k=\omega}^n \log_{\alpha} \left(1 + \frac{1}{10k} \right)}{\sum_{k=\omega}^n \log_{\alpha} \left(1 + \frac{1}{k} \right)} - \frac{\sum_{k=\omega}^n \log_{\alpha} \left(1 + \frac{1}{10k+10} \right)}{\sum_{k=\omega}^n \log_{\alpha} \left(1 + \frac{1}{k} \right)} = \\
& = \frac{\sum_{k=\omega}^n \left[\log_{\alpha} \left(1 + \frac{1}{10k} \right) - \log_{\alpha} \left(1 + \frac{1}{10(k+1)} \right) \right]}{\log_{\alpha}(n+1) - \log_{\alpha} \omega} = \\
& = \frac{\log_{\alpha} \left(1 + \frac{1}{10\omega} \right) - \log_{\alpha} \left(1 + \frac{1}{10(n+1)} \right)}{\log_{\alpha}(n+1) - \log_{\alpha} \omega} < \delta
\end{aligned} \tag{6.54}$$

para qualquer δ positivo, desde que n seja suficientemente grande⁽¹⁷⁾.

⁽¹⁶⁾ Na expressão de $U_n(d)$ o valor de d só influencia o numerador, verificando-se facilmente que, quanto maior for o valor de d , menor será o valor de $U_n(d)$.

⁽¹⁷⁾ Pacheco d'Amorim nesta dedução apresenta

$$\frac{\log_{\alpha} \left(1 + \frac{1}{10\omega} \right) - \log_{\alpha} \left(1 + \frac{1}{10(n+1)} \right)}{\log_{\alpha}(n+1) - \log_{\alpha} \omega} < \frac{\log_{\alpha} \left(1 + \frac{1}{10\omega} \right)}{\log_{\alpha}(n+1)}$$

Deste resultado podemos concluir que o limite $\lim_{n \rightarrow \infty} U_n(d)$ existe para qualquer valor de d , consequência de o limite existir para $d \geq 9$ (em particular para $d = 10$) e de o limite ser o mesmo para qualquer valor inteiro de d no intervalo $[0, 10]$. Uma vez que o valor desse limite é independente do valor de d , que a soma das probabilidades é igual à unidade e que d em $\mathbb{P}_{(d,a)}$ só assume valores inteiros compreendidos entre zero e nove, teremos

$$\begin{aligned} \sum_{d=0}^9 \mathbb{P}_{(d,a)} = 1 &\Leftrightarrow \lim_{n \rightarrow \infty} \sum_{d=0}^9 U_n(d) = 1 \Leftrightarrow \\ &\Leftrightarrow 10 \lim_{n \rightarrow \infty} U_n(d) = 1 \Leftrightarrow \lim_{n \rightarrow \infty} U_n(d) = \frac{1}{10}. \end{aligned} \quad (6.55)$$

□

Pacheco d'Amorim acrescenta ainda, sem o justificar, que facilmente se constata que U_n converge muito rapidamente para $\frac{1}{10}$. Deste modo, na página 77 da sua tese, depois de afirmar que “*facilmente se vê que*” U_n converge para $\frac{1}{10}$, ocupa quase quatro página para demonstrar esse facto *óbvio*, afirmando posteriormente que “*facilmente se vê que*” a convergência é rápida. Parece-nos que a velocidade de convergência da expressão U_n não é assim tão evidente como o autor afirma. Porém, a análise da referida velocidade de convergência não é apresentada por Pacheco d'Amorim.

Pacheco d'Amorim considera que este problema, na sua fórmula geral (fórmula (6.19) presente na página 332), fornece a distribuição dos algarismos numa tábua ideal que contenha todos os valores duma função num intervalo (α, β) , isto é, uma tábua em que qualquer valor, racional ou não, do intervalo

que não é verdadeiro. Basta verificar que para $\omega = 5$ e $n = 10$ (α pode ser qualquer, pois o valor destes rácios é independente do valor de α) obteríamos aproximadamente $0,0136378 < 0,0082583$. Contudo, esta passagem não é necessária para alcançar o objetivo do autor. Refira-se que, se restringirmos a nossa análise a $\omega = 1$, esta desigualdade, como facilmente se verifica, já será verdadeira.

(α, β) tenha nessa tábua imaginária o valor da sua imagem. Contudo, numa tábua em que os valores da variável independente (ponto livre x) estejam em progressão aritmética, intervalos com igual amplitude contidos no intervalo (α, β) compreendem, aproximadamente, o mesmo número de valores de x escritos na tábua. Esta aproximação é tanto melhor quanto menor for a razão da progressão dos valores de x , pois teremos mais valores em cada sub-região. Consequentemente, a probabilidade de que um valor de x , tirado à sorte na região (α, β) , pertença a um dado intervalo parcial é, aproximadamente, proporcional à amplitude do intervalo, tal qual como acontece na tábua ideal, sendo o desfasamento em relação à situação ideal tanto menor quanto menor for a razão da progressão. Podemos, assim, visualizar a tábua ideal como o limite duma sucessão de tábuas, em que a razão da progressão dos valores de x vá decrescendo até zero. Deste modo, a fórmula (6.41) (patente na página 341), exacta para a tábua ideal, fornecerá valores tanto mais aproximados, para a distribuição dos algarismos numa tábua de um ponto imagem $y = f(x)$, quanto menor for a razão da progressão dos valores de x . Por este motivo o autor considera que numa tábua de logaritmos decimais (primeira aplicação com $\alpha = 10$), visto que a mantissa⁽¹⁸⁾ se não altera com a divisão de x por uma potência inteira de 10, conclui que a fórmula

$$\frac{\mathbb{P}_{(d+1,a)}}{\mathbb{P}_{(d,a)}} = 10^{\frac{1}{10^a}} \quad (6.56)$$

fornece um valor mais próximo nos valores mais altos (fim da tábua) que nos

⁽¹⁸⁾ A mantissa de um logaritmo de base decimal corresponde à parte decimal da sua imagem, isto é, a mantissa de $\log(50) \approx 1,698970004$ é 698970004, sendo o valor das unidades (1) a sua característica. Se o valor que pretendemos calcular o logaritmo for multiplicado por 10^n , a característica aumentará n valores e a mantissa manter-se-á, uma vez que $\log(k10^n) = \log(k) + n \log(10) = \log(k) + n$. Assim, a título de exemplo, com base no resultado previamente apresentado poderíamos concluir que $\log(500) \approx 2,698970004$.

mais baixos, justificando com os valores a seguir apresentados⁽¹⁹⁾. Pacheco d'Amorim recorre aos valores de x situados entre 1289 e 1318 ($\Delta = 30$), onde o ponto imagem $y = \log(x)$ apresenta o algarismo 1 na segunda casa decimal, e aos valores de x situados entre 1319 e 1348 ($\Delta = 30$), cujos pontos imagem apresentam o algarismo 2 na segunda casa decimal, concluindo que

$$\frac{\mathbb{P}_{(2,2)}}{\mathbb{P}_{(1,2)}} = \frac{30}{30} = 1, \quad (6.57)$$

nos referidos intervalos. Utilizando o mesmo raciocínio, temos que, para valores do ponto livre x compreendidos entre 10233 a 10471 ($\Delta = 239$), o correspondente ponto imagem apresenta o algarismo 1 na segunda casa decimal e, para valores do ponto livre compreendidos entre 10472 e 10715 ($\Delta = 244$), o ponto imagem apresenta o algarismo 2 na mesma casa decimal. Assim, obtemos para o valor da probabilidade

$$\frac{\mathbb{P}_{(2,2)}}{\mathbb{P}_{(1,2)}} = \frac{244}{239} = 1,0209 \dots; \quad (6.58)$$

número este mais próximo do número teórico fornecido pela fórmula (6.42) (página 341)

$$\sqrt[100]{10} = 1,0233 \quad (6.59)$$

que o primeiro. Assim, no quociente entre as probabilidades de, na segunda casa decimal, o algarismo ser 2 em relação a o algarismo ser igual a 1, obteremos uma probabilidade mais próxima da teórica no caso de utilizarmos

⁽¹⁹⁾ Para não reproduzirmos uma tabela de logaritmos, a tabela seguinte apresenta os valores para os logaritmos utilizados por Pacheco d'Amorim aproximados à quinta casa decimal, calculados através do *Mathematica* 5.2.

x	$y = \log(x)$	x	$y = \log(x)$	x	$y = \log(x)$
1288	3.10992	1348	3.12969	10471	4.01999
1289	3.11025	1349	3.13001	10472	4.02003
1318	3.11992	10232	4.00996	10715	4.02999
1319	3.12024	10233	4.01000	10716	4.03003

valores mais elevados, pois teremos mais números (no ponto imagem) em cada situação, sendo os valores obtidos cada vez mais próximos do valor teórico.

O conceito de ponto imagem construído por Pacheco d'Amorim é restrito ao contexto contínuo, isto é, ao lançamento de um ponto numa região (ponto livre) e análise de uma sua imagem (ponto imagem) que variará também ela numa região, sendo a possibilidade de uma região referente ao ponto imagem definida pela possibilidade da equivalente região do ponto livre. Contudo, salientemos que, nesta análise às tábuas de logaritmos apresentada por Pacheco d'Amorim, o autor considera que as suas conclusões baseadas na concepção de ponto imagem em regiões, nomeadamente o resultado a que chegou para o Problema 6.1, também podem ser aplicadas, como uma aproximação, para situações discretas. Assim, enquanto numa tábua ideal, onde o ponto imagem está definido numa região e consequentemente estamos a trabalhar num ambiente contínuo, a fórmula (6.42) é exacta, nas tábuas usuais, onde os valores que o ponto livre assume correspondem aos valores de uma progressão aritmética de razão r , a fórmula (6.42) será aproximada, sendo esta aproximação tanto melhor quanto menor for o valor da razão r . Parece, assim, que o autor idealiza que a sua concepção de ponto imagem também pode ser generalizada para os casos discretos, sendo que nestas situações os valores não serão exactos, mas aproximados. Contudo, não há mais nenhuma reflexão sobre esta questão na tese de doutoramento de Pacheco d'Amorim.

De facto, é mesmo até possível construir uma análise semelhante, à efectuada por Pacheco d'Amorim para a situação contínua do ponto imagem, para o caso discreto, ou seja, efectuar uma transformação de uma variável discreta noutra discreta ou uma transformação de uma variável contínua numa discreta e obter resultados exactos. Com tal construção poderemos obter, se nos abstrairmos das dificuldades que podem surgir em algumas distribuições mais complexas, qualquer distribuição discreta de probabilidade

através de uma transformação de uma distribuição uniforme numa região (variável contínua) ou numa classe (variável discreta). Todavia, a análise deste problema não foi efectuada por Pacheco d'Amorim. Assim, uma vez que o autor não abordou esta situação, nesta aplicação faz uma análise onde as fórmulas deduzidas para a transformação de uma variável contínua, exactas no seu contexto, podem ser visualizadas como uma aproximação de uma transformação de variáveis discretas (tal como acontece na tábua onde temos uma distribuição uniforme entre o ponto livre que assume valores numa progressão aritmética), sendo que a aproximação é tanto melhor quanto menor for o afastamento dos pontos livre discretos (no caso analisado, razão da progressão aritmética utilizada).

6.2.4 Terceiro caso

A terceira aplicação apresentada por Pacheco d'Amorim refere-se à análise de

$$\frac{\mathbb{P}_{d+1}}{\mathbb{P}_d} \quad (6.60)$$

para as diferenças tabulares dos logaritmos, isto é, considerando que temos valores de x escritos em progressão aritmética de razão unitária, estaremos a analisar a imagem de dois valores consecutivos do logaritmos,

$$y = \log(1+x) - \log(x) = \log\left(1 + \frac{1}{x}\right). \quad (6.61)$$

Sejam então os valores x , x' e x'' definidos por

$$\begin{aligned} \log\left(1 + \frac{1}{x}\right) &= \frac{10\omega + d}{10^a}, \\ \log\left(1 + \frac{1}{x'}\right) &= \frac{10\omega + d + 1}{10^a}, \\ \log\left(1 + \frac{1}{x''}\right) &= \frac{10\omega + d + 2}{10^a}, \end{aligned} \quad (6.62)$$

a expressão $\frac{\mathbb{P}_{d+1}}{\mathbb{P}_d}$ será então

$$\begin{aligned} \frac{\mathbb{P}_{d+1}}{\mathbb{P}_d} &= \frac{x'' - x'}{x' - x} = \frac{\frac{1}{10^{\frac{10\omega+d+2}{10^a}} - 1} - \frac{1}{10^{\frac{10\omega+d+1}{10^a}} - 1}}{\frac{1}{10^{\frac{10\omega+d+1}{10^a}} - 1} - \frac{1}{10^{\frac{10\omega+d}{10^a}} - 1}} = \\ &= \frac{\frac{10^{\frac{10\omega+d+1}{10^a}} - 1 - 10^{\frac{10\omega+d+2}{10^a}} + 1}{\left(10^{\frac{10\omega+d+2}{10^a}} - 1\right)\left(10^{\frac{10\omega+d+1}{10^a}} - 1\right)}}{\frac{10^{\frac{10\omega+d}{10^a}} - 1 - 10^{\frac{10\omega+d+1}{10^a}} + 1}{\left(10^{\frac{10\omega+d+1}{10^a}} - 1\right)\left(10^{\frac{10\omega+d}{10^a}} - 1\right)}} = \frac{10^{\frac{10\omega+d}{10^a}} - 1}{10^{\frac{10\omega+d+2}{10^a}} - 1} 10^{\frac{1}{10^a}}. \end{aligned} \quad (6.63)$$

O limite quando a aumenta indefinidamente será dado por

$$\lim_{a \rightarrow \infty} \frac{10^{\frac{10\omega+d}{10^a}} - 1}{10^{\frac{10\omega+d+2}{10^a}} - 1} 10^{\frac{1}{10^a}} = \frac{10\omega + d}{10\omega + d + 2}, \quad (6.64)$$

uma vez que

$$\begin{aligned} \lim_{a \rightarrow \infty} \frac{10^{\frac{10\omega+d}{10^a}} - 1}{10^{\frac{10\omega+d+2}{10^a}} - 1} &= \lim_{a \rightarrow \infty} \frac{-(10\omega + d) 10^{\frac{10\omega+d}{10^a} - a} \ln^2(10)}{-(10\omega + d + 2) 10^{\frac{10\omega+d+2}{10^a} - a} \ln^2(10)} = \\ &= \lim_{a \rightarrow \infty} \frac{10\omega + d}{10\omega + d + 2} 10^{\frac{-2}{10^a}} = \frac{10\omega + d}{10\omega + d + 2}. \end{aligned} \quad (6.65)$$

Por este motivo podemos concluir que

$$\frac{\mathbb{P}_d}{\mathbb{P}_{d+1}} = \frac{10\omega + d + 2}{10\omega + d} = 1 + \frac{2}{10\omega + d} = 1 + \frac{2}{10^a D} \quad (6.66)$$

onde D é uma diferença tabular, isto é

$$D = \log\left(1 + \frac{1}{x}\right) = \log(1 + x) - \log(x) = \frac{10\omega + d}{10^a}. \quad (6.67)$$

Desta forma, se conhecermos o valor da diferença tabular D , podemos através da expressão (6.66) determinar o valor da relação pretendida $\frac{\mathbb{P}_d}{\mathbb{P}_{d+1}}$, onde d é o número dígito que D tem na casa decimal de ordem a . Pacheco d'Amorim ilustra esta ideia utilizando a diferença tabular 0,0000524, através da qual deduz, para a quinta casa decimal ($a = 5$), que

$$\frac{\mathbb{P}_5}{\mathbb{P}_6} = 1 + \frac{2}{5} = \frac{7}{5} = 1,4. \quad (6.68)$$

Se analisarmos os valores presentes numa tábua de logaritmos⁽²⁰⁾, verificamos que a diferença tabular 0,0000500 tem por máximo correspondente o número 8686; a diferença tabular 0,0000600 tem por máximo correspondente o número 7238; a diferença tabular 0,0000700 tem por número máximo correspondente o número 6204. Assim, utilizando os valores apresentados concluimos que

$$\frac{\mathbb{P}_5}{\mathbb{P}_6} = \frac{8686 - 7239 + 1}{7238 - 6205 + 1} = \frac{1448}{1034} \approx 1.4004. \quad (6.69)$$

Pacheco d'Amorim acrescenta ainda, em forma de observação, que a fórmula

$$\frac{\mathbb{P}_d}{\mathbb{P}_{d+1}} = 10^{\frac{1}{10^a}} \quad (6.70)$$

é exacta para as casas inteiras, isto é, para $a = 0, -1, -2, \dots$, consequência de todos os números inteiros serem logaritmos de valores de x escritos nas tábuas.

6.3 Lei da Possibilidade e Lei da Probabilidade

Seja x um ponto livre lançado à sorte em $\mathbf{X} \subset \mathbb{R}^n$ e $y = f(x)$ uma sua imagem definida em $\mathbf{Y} = f(\mathbf{X}) \subset \mathbb{R}^n$. Seja a região $\mathbf{V}_{x_0} \subset \mathbf{X}$ uma vizinhança do ponto $x_0 \in \mathbf{X}$ ao qual corresponde o ponto imagem $y_0 = f(x_0)$ com vizinhança $\mathbf{V}_{y_0} = f(\mathbf{V}_{x_0}) \subset \mathbf{Y}$. Seja $\varpi_x(\mathbf{V}_{x_0})$ a possibilidade da região

⁽²⁰⁾ A seguinte tabela apresenta as diferenças tabulares referidas por Pacheco d'Amorim. Contudo, os valores foram ligeiramente corrigidos, pois as tabelas consultadas pelo autor (que não identifica) porventura apresentavam valores ligeiramente distintos destes por nós apresentados, calculados utilizando o *Mathematica* 5.2 e truncados na sétima casa decimal.

x	$y = \log(1 + \frac{1}{x})$	x	$y = \log(1 + \frac{1}{x})$	x	$y = \log(1 + \frac{1}{x})$
6204	0,0000700	7238	0,0000600	8686	0,0000500
6205	0,0000699	7239	0,0000599	8687	0,0000499

$\mathbf{V}_{x_0}$, que, pela definição de ponto imagem, será também a possibilidade da região $\mathbf{V}_{y_0}$, isto é, $\varpi_x(\mathbf{V}_{x_0}) = \varpi_y(\mathbf{V}_{y_0})$. Neste contexto Pacheco d'Amorim define a lei de possibilidade do ponto imagem y para cada ponto $y_0 \in \mathbf{Y}$, através de

$$\pi_y(y_0) = \lim_{\Delta \mathbf{V}_{y_0} \rightarrow 0} \frac{\varpi_y(\mathbf{V}_{y_0})}{\Delta \mathbf{V}_{y_0}}, \quad (6.71)$$

onde $\Delta \mathbf{V}_{y_0}$ representa uma medida da vizinhança $\mathbf{V}_{y_0}$ que Pacheco d'Amorim não define claramente, referindo apenas que o limite $\Delta \mathbf{V}_{y_0} \rightarrow 0$ é efectuado de forma a que a maior dimensão da vizinhança $\mathbf{V}_{y_0}$ convirja também para zero.

Assim, para o autor, os pontos y_0 para os quais existe o limite (6.71) formam o domínio da função π_y . Esta função, cujo valor em cada ponto é dado por esse limite, é denominada **lei da possibilidade** do ponto (imagem) y .

Exemplo 6.7. No exemplo 6.1, apresentado na página 323, deduzimos que, se x for um ponto livre lançado à sorte no intervalo $(0, 1)$ e $y = -\frac{1}{\alpha} \ln(1 - x)$, com $\alpha > 0$, um seu ponto imagem, então para qualquer região $(y_0, y_1) \subset \mathbb{R}^+$ teremos

$$\mathbb{P}_{\mathbb{R}^+}(y_0 < y < y_1) = \varpi_y((y_0, y_1)) = e^{-\alpha y_0} - e^{-\alpha y_1}. \quad (6.72)$$

Consequentemente, a lei da possibilidade do ponto imagem y no ponto y_0 , para qualquer $y_0 \in \mathbb{R}^+$ com vizinhança $\mathbf{V}_{y_0} = (y_0 - \epsilon, y_0 + \epsilon)$, é dada por

$$\begin{aligned} \pi_y(y_0) &= \lim_{\Delta \mathbf{V}_{y_0} \rightarrow 0} \frac{\varpi_y(\mathbf{V}_{y_0})}{\Delta \mathbf{V}_{y_0}} = \lim_{\epsilon \rightarrow 0} \frac{e^{-\alpha(y_0 - \epsilon)} - e^{-\alpha(y_0 + \epsilon)}}{2\epsilon} = \\ &= \lim_{\epsilon \rightarrow 0} \frac{\alpha \left(e^{-\alpha(y_0 - \epsilon)} + e^{-\alpha(y_0 + \epsilon)} \right)}{2} = \alpha e^{-\alpha y_0}, \end{aligned} \quad (6.73)$$

que corresponde à função densidade de uma variável aleatória Y com distribuição exponencial de parâmetro α . ◆

Exemplo 6.8. Apliquemos a definição de lei de possibilidade à situação ilustrada no exemplo 6.2 (página 323), onde deduzimos que o ponto imagem

$$y = f(x) = \begin{cases} (b-a) \sqrt{\frac{x}{2}} + a & 0 \leq x \leq 0.5 \\ (a-b) \sqrt{\frac{1-x}{2}} + b & 0.5 < x \leq 1 \end{cases}, \quad (6.74)$$

do ponto livre x lançado à sorte em $(0,1)$, tem possibilidade da região $(y_0, y_1) \subset (a, b)$ dada pela função $\varpi_y((y_0, y_1))$ definida por

$$\begin{cases} 2 \frac{(y_1 - a)^2 - (y_0 - a)^2}{(b - a)^2} & a \leq y_0 \leq y_1 \leq \frac{a+b}{2} \\ 1 - 2 \frac{(y_1 - b)^2 + (y_0 - a)^2}{(b - a)^2} & a \leq y_0 \leq \frac{a+b}{2} \leq y_1 \leq b \\ 2 \frac{(y_0 - b)^2 - (y_1 - b)^2}{(b - a)^2} & \frac{a+b}{2} \leq y_0 \leq y_1 \leq b \end{cases}. \quad (6.75)$$

A lei de possibilidade do ponto imagem y_0 , quando $y_0 < \frac{a+b}{2}$, será

$$\begin{aligned} \pi_y(y_0) &= \lim_{\Delta \mathbf{V}_{y_0} \rightarrow 0} \frac{\varpi_y(\mathbf{V}_{y_0})}{\Delta \mathbf{V}_{y_0}} = \lim_{\epsilon \rightarrow 0} \frac{2 \frac{(y_0 + \epsilon - a)^2 - (y_0 - \epsilon - a)^2}{(b - a)^2}}{2\epsilon} = \\ &= \lim_{\epsilon \rightarrow 0} \frac{(y_0 - a)^2 + 2\epsilon(y_0 - a) + \epsilon^2 - (y_0 - a)^2 + 2\epsilon(y_0 - a) - \epsilon^2}{\epsilon(b - a)^2} = \\ &= \lim_{\epsilon \rightarrow 0} \frac{4\epsilon(y_0 - a)}{\epsilon(b - a)^2} = \frac{4(y_0 - a)}{(b - a)^2}. \end{aligned} \quad (6.76)$$

No caso em que $y_0 = \frac{a+b}{2}$ teremos

$$\begin{aligned} \pi_y(y_0) &= \lim_{\Delta \mathbf{V}_{y_0} \rightarrow 0} \frac{\varpi_y(\mathbf{V}_{y_0})}{\Delta \mathbf{V}_{y_0}} = \lim_{\epsilon \rightarrow 0} \frac{1 - 2 \frac{(y_0 + \epsilon - b)^2 + (y_0 - \epsilon - a)^2}{(b - a)^2}}{2\epsilon} = \\ &= \lim_{\epsilon \rightarrow 0} \frac{(b - a)^2 - 2 \left(-\left(\frac{b-a}{2} - \epsilon\right) \right)^2 - 2 \left(\frac{b-a}{2} - \epsilon \right)^2}{2\epsilon(b - a)^2} = \\ &= \lim_{\epsilon \rightarrow 0} \frac{(b - a)^2 - 4 \frac{(b-a)^2}{4} + 4\epsilon(b - a) - 4\epsilon^2}{2\epsilon(b - a)^2} = \\ &= \lim_{\epsilon \rightarrow 0} \frac{2}{(b - a)} - 2 \frac{\epsilon}{(b - a)^2} = \frac{2}{(b - a)}. \end{aligned} \quad (6.77)$$

Por fim, na situação $y_0 > \frac{a+b}{2}$, obteremos

$$\begin{aligned}
 \pi_y(y_0) &= \lim_{\Delta \mathbf{V}_{y_0} \rightarrow 0} \frac{\varpi_y(\mathbf{V}_{y_0})}{\Delta \mathbf{V}_{y_0}} = \lim_{\epsilon \rightarrow 0} \frac{2 \frac{(y_0 - \epsilon - b)^2 - (y_0 + \epsilon - b)^2}{(b-a)^2}}{2\epsilon} = \\
 &= \lim_{\epsilon \rightarrow 0} \frac{(y_0 - b)^2 - 2\epsilon(y_0 - b) + \epsilon^2 - (y_0 - b)^2 - 2\epsilon(y_0 - b) - \epsilon^2}{\epsilon(b-a)^2} = \\
 &= \lim_{\epsilon \rightarrow 0} \frac{-4\epsilon(y_0 - b)}{\epsilon(b-a)^2} = \frac{4(b - y_0)}{(b-a)^2}. \tag{6.78}
 \end{aligned}$$

Desta forma concluímos que a lei de possibilidade do ponto imagem y é

$$g(y_0) = \begin{cases} \frac{4}{(b-a)^2}(-a + y_0) & a \leq y_0 \leq \frac{a+b}{2} \\ \frac{4}{(b-a)^2}(b - y_0) & \frac{a+b}{2} < y_0 \leq b \end{cases} \tag{6.79}$$

que corresponde à função densidade de uma variável aleatória com distribuição triangular no intervalo $[a, b]$. $\blacklozenge$

Exemplo 6.9. No lançamento analisado no exemplo 6.3 (página 324), onde x é um ponto livre lançado, à sorte, no intervalo $(e^{-1}, 1)$ e $y = -\log(x)$ um seu ponto imagem, concluímos que a possibilidade no intervalo $(y_0, y_1) \subset (0, 1)$ é determinada por

$$\varpi_y((y_0, y_1)) = \frac{e^{-y_0} - e^{-y_1}}{1 - e^{-1}}. \tag{6.80}$$

Assim, a lei de possibilidade deste ponto imagem, para qualquer ponto y_0 , tal que $y_0 \in (0, 1)$, é dada por

$$\begin{aligned}
 \pi_y(y_0) &= \lim_{\Delta \mathbf{V}_{y_0} \rightarrow 0} \frac{\varpi_y(\mathbf{V}_{y_0})}{\Delta \mathbf{V}_{y_0}} = \lim_{\epsilon \rightarrow 0} \frac{\frac{e^{-(y_0-\epsilon)} - e^{-(y_0+\epsilon)}}{1-e^{-1}}}{2\epsilon} = \\
 &= \lim_{\epsilon \rightarrow 0} \frac{e^{-(y_0-\epsilon)} - e^{-(y_0+\epsilon)}}{2\epsilon(1-e^{-1})} = \lim_{\epsilon \rightarrow 0} \frac{e^{-(y_0-\epsilon)} + e^{-(y_1+\epsilon)}}{2(1-e^{-1})} = \frac{e^{-y_0}}{1-e^{-1}}, \tag{6.81}
 \end{aligned}$$

que corresponde à função densidade de uma variável aleatória com distribuição exponencial, truncada para valores inferiores à unidade, com parâmetro igual à unidade. $\blacklozenge$

Exemplo 6.10. Aplicando a definição de lei de possibilidade aos dados do exemplo 6.4 (página 325), onde deduzimos que o ponto imagem $y = \sigma \tan(x)$ do ponto livre x lançado à sorte em $(-\frac{\pi}{2}, \frac{\pi}{2})$ verifica

$$\varpi_y((y_0, y_1)) = \frac{\arctan\left(\frac{y_1}{\sigma}\right) - \arctan\left(\frac{y_0}{\sigma}\right)}{\pi}, \quad (6.82)$$

podemos concluir que

$$\begin{aligned} \pi_y(y_0) &= \lim_{\Delta \mathbf{V}_{y_0} \rightarrow 0} \frac{\varpi_y(\mathbf{V}_{y_0})}{\Delta \mathbf{V}_{y_0}} = \lim_{\epsilon \rightarrow 0} \frac{\arctan\left(\frac{y_0 + \epsilon}{\sigma}\right) - \arctan\left(\frac{y_0 - \epsilon}{\sigma}\right)}{2\epsilon\pi} = \\ &= \lim_{\epsilon \rightarrow 0} \frac{\frac{\frac{1}{\sigma}}{1 + \left(\frac{y_0 + \epsilon}{\sigma}\right)^2} + \frac{\frac{1}{\sigma}}{1 + \left(\frac{y_0 - \epsilon}{\sigma}\right)^2}}{2\pi} = \frac{1}{\sigma\pi \left(1 + \left(\frac{y_0}{\sigma}\right)^2\right)} = \\ &= \frac{\sigma}{\pi(\sigma^2 + y_0^2)}, \end{aligned} \quad (6.83)$$

que corresponde à função densidade de probabilidade de uma variável aleatória com distribuição de Cauchy com parâmetro σ . $\blacklozenge$

Exemplo 6.11. Recorrendo aos dados patentes no exemplo 6.5 (página 325), onde deduzimos que o ponto imagem $y = \frac{b}{x^a}$ do ponto livre x lançado à sorte na região $(0, 1)$ verifica

$$\varpi_y((y_0, y_1)) = \left(\frac{b}{y_0}\right)^a - \left(\frac{b}{y_1}\right)^a, \quad (6.84)$$

podemos deduzir que, para qualquer ponto $y_0 > b > 0$ teremos

$$\begin{aligned} \pi_y(y_0) &= \lim_{\Delta \mathbf{V}_{y_0} \rightarrow 0} \frac{\varpi_y(\mathbf{V}_{y_0})}{\Delta \mathbf{V}_{y_0}} = \lim_{\epsilon \rightarrow 0} \frac{\left(\frac{b}{y_0 - \epsilon}\right)^a - \left(\frac{b}{y_0 + \epsilon}\right)^a}{2\epsilon} = \\ &= \frac{b^a}{2} \lim_{\epsilon \rightarrow 0} \frac{(y_0 + \epsilon)^a - (y_0 - \epsilon)^a}{\epsilon (y_0 - \epsilon)^a (y_0 + \epsilon)^a} = \\ &= \frac{b^a}{2} \lim_{\epsilon \rightarrow 0} \frac{a(y_0 + \epsilon)^{a-1} + a(y_0 - \epsilon)^{a-1}}{(y_0 - \epsilon)^a (y_0 + \epsilon)^a - 2a\epsilon^2 (y_0 - \epsilon)^{a-1} (y_0 + \epsilon)^{a-1}} = \\ &= \frac{b^a 2a y_0^{a-1}}{2y_0^{2a}} = \frac{ab^a}{y_0^{a+1}}, \end{aligned} \quad (6.85)$$

que corresponde à função densidade de probabilidade de uma variável aleatória com distribuição de Pareto (a, b) . $\blacklozenge$

Apesar de a aplicação da definição de lei de possibilidade de um ponto imagem não ser tarefa muito complicada em $\mathbb{R}$, quando aumentamos o número de dimensões a tarefa deixará de ser fácil, como ilustra a aplicação da definição de lei de possibilidade ao método Box-Muller apresentado em nota de rodapé na página 330. Se considerarmos $x = (x_1, x_2)$ um ponto livre lançado à sorte na região $(0, 1) \times (0, 1)$ (equivalente a lançar de forma independente dois pontos em $(0, 1)$) e definirmos o ponto imagem $y = (y_1, y_2)$ através de

$$\begin{cases} y_1 &= \sqrt{-2 \ln(x_1)} \cos(2\pi x_2) \\ y_2 &= \sqrt{-2 \ln(x_1)} \sin(2\pi x_2) \end{cases}, \quad (6.86)$$

podemos deduzir a função inversa. Uma vez que se verifica

$$\begin{cases} y_1^2 &= -2 \ln(x_1) \cos^2(2\pi x_2) \\ y_2^2 &= -2 \ln(x_1) \sin^2(2\pi x_2) \end{cases}, \quad (6.87)$$

podemos deduzir que

$$y_1^2 + y_2^2 = -2 \ln(x_1) \Leftrightarrow x_1 = e^{\left(-\frac{y_1^2 + y_2^2}{2}\right)} \quad (6.88)$$

e também

$$\begin{cases} y_1 &= \sqrt{y_1^2 + y_2^2} \cos(2\pi x_2) \\ y_2 &= \sqrt{y_1^2 + y_2^2} \sin(2\pi x_2) \end{cases}, \quad (6.89)$$

logo,

$$\frac{y_2}{y_1} = \tan(2\pi x_2) \Leftrightarrow x_2 = \frac{1}{2\pi} \arctan\left(\frac{y_2}{y_1}\right). \quad (6.90)$$

Deste modo a função inversa de f é

$$f^{-1} = \begin{cases} x_1 &= e^{-\frac{y_1^2 + y_2^2}{2}} \\ x_2 &= \frac{1}{2\pi} \arctan\left(\frac{y_2}{y_1}\right) \end{cases}, \quad (6.91)$$

embora, mesmo com o conhecimento da função inversa, não conseguimos determinar a lei de possibilidade. Com a função inversa conseguimos facilmente de um ponto y' de $\mathbf{Y}$ regressar para o correspondente ponto livre x' de $\mathbf{X}$, contudo, uma peça chave na determinação da lei de possibilidade é a passagem de $\mathbf{V}_{y_0} \subset \mathbf{Y}$ para $\mathbf{V}_{x_0} \subset \mathbf{X}$, uma vez que precisamos de recorrer a $\varpi_y(\mathbf{V}_{y_0}) = \varpi_x(\mathbf{V}_{x_0})$ e através de uma região geral $\mathbf{V}_{y_0}$ não podemos obter a correspondente região $\mathbf{V}_{x_0}$. Além disso, Pacheco d'Amorim não define a forma que considera para a vizinhança $\mathbf{V}_{y_0}$, nem a forma como é tomado o limite patente na definição de lei de possibilidade.

Restringindo a nossa análise a pontos lançados à sorte numa região de $\mathbb{R}$, isto é, $\mathbf{X} \subset \mathbb{R}$, temos que

$$\varpi_x(\mathbf{X}') = \frac{\mu(\mathbf{X}')}{\mu(\mathbf{X})}, \quad (6.92)$$

consequência de a distribuição do ponto livre ser uniforme na região onde é lançado o ponto. A vizinhança num ponto y_0 será definida através de $\mathbf{V}_{y_0} = [y_0 - \epsilon, y_0 + \epsilon]$ que, sob a hipótese de que f é uma função crescente, a equivalente vizinhança do ponto livre de x_0 terá a forma

$$\mathbf{V}_{x_0} = \left[f^{-1}(y_0 - \epsilon), f^{-1}(y_0 + \epsilon) \right]. \quad (6.93)$$

Assim, podemos concluir que

$$\varpi_y(\mathbf{V}_{y_0}) = \varpi_x(\mathbf{V}_{x_0}) = \frac{f^{-1}(y_0 + \epsilon) - f^{-1}(y_0 - \epsilon)}{\mu(\mathbf{X})} \quad (6.94)$$

e, consequentemente,

$$\pi_y(y_0) = \lim_{\epsilon \rightarrow 0} \frac{\varpi_y(\mathbf{V}_{y_0})}{\Delta \mathbf{V}_{y_0}} = \lim_{\epsilon \rightarrow 0} \frac{\frac{f^{-1}(y_0 + \epsilon) - f^{-1}(y_0 - \epsilon)}{\mu(\mathbf{X})}}{2\epsilon} = \frac{\left(f^{-1}(y_0) \right)'}{\mu(\mathbf{X})}. \quad (6.95)$$

Uma vez que $\epsilon > 0$, podemos concluir que

$$\begin{aligned} & \lim_{\epsilon \rightarrow 0} \frac{f^{-1}(y_0 + \epsilon) - f^{-1}(y_0 - \epsilon)}{2\epsilon} = \\ &= \frac{1}{2} \lim_{\epsilon \rightarrow 0} \left[\frac{f^{-1}(y_0 + \epsilon) - f^{-1}(y_0)}{\epsilon} + \frac{f^{-1}(y_0 - \epsilon) - f^{-1}(y_0)}{-\epsilon} \right] = \\ &= \left(f^{-1}(y_0) \right)'. \end{aligned} \quad (6.96)$$

A expressão (6.95) corresponde ao rácio entre a derivada da inversa da função da transformação f e a medida (comprimento) da região onde o ponto livre é lançado à sorte. Caso o ponto seja lançado no intervalo $[0, 1]$, distribuição uniforme no intervalo $[0, 1]$ para o ponto livre x , então $\mu(\mathbf{X}) = 1$ e a expressão (6.95) pode ser simplificada para

$$\pi_y(y_0) = \left(f^{-1}(y_0) \right)' . \quad (6.97)$$

Nesta circunstância a lei de possibilidade do ponto imagem $y = f(x)$ é simplesmente a derivada da função inversa da transformação utilizada.

Notemos ainda que, se utilizássemos a função inversa da função de distribuição de uma variável aleatória para a transformação (tal como no método da transformação inversa), isto é, $f = F^{-1} \Leftrightarrow f^{-1} = F$ (que existe pois F^{-1} é sempre crescente) obteríamos

$$\varpi_y(\mathbf{V}_{y_0}) = \frac{F(y_0 + \epsilon) - F(y_0 - \epsilon)}{\mu(\mathbf{X})} \quad (6.98)$$

e, consequentemente,

$$\pi_y(y_0) = \lim_{\epsilon \rightarrow 0} \frac{\frac{F(y_0 + \epsilon) - F(y_0 - \epsilon)}{\mu(\mathbf{X})}}{2\epsilon} = \frac{F'(y_0)}{\mu(\mathbf{X})} . \quad (6.99)$$

Como neste caso o ponto livre x é lançado à sorte na região $[0, 1]$ (a função de distribuição só assume valores neste intervalo), teríamos $\mu(\mathbf{X}) = 1$ e a expressão (6.71), presente na página 354, seria simplificada para

$$\pi_y(y_0) = \lim_{\epsilon \rightarrow 0} \frac{F(y_0 + \epsilon) - F(y_0 - \epsilon)}{2\epsilon} = F'(y_0) , \quad (6.100)$$

isto é, a lei de possibilidade corresponderia à derivada da função de distribuição. Em tal situação, a condição de existência do limite presente na definição de lei de possibilidade corresponderia à existência de derivada da função de distribuição (excepto num conjunto de medida nula), isto é, de esta função corresponder a uma função de distribuição de uma variável absolutamente contínua.

Estas expressões, onde a lei de possibilidade de um ponto imagem é determinada através da derivada, tornam a determinação destas leis bastante mais acessível. Analisemos como através destas expressões poderíamos determinar as mesmas leis de possibilidades retratadas nos exemplos 6.7 a 6.11.

Exemplo 6.12. No exemplo 6.1 (página 323) analisamos o ponto imagem $y = -\frac{1}{\alpha} \ln(1-x)$, com $\alpha > 0$, do ponto livre x lançado, à sorte, no intervalo $(0, 1)$. Uma vez que $x = f^{-1}(y) = 1 - e^{-\alpha y}$ temos $\pi_y(y) = \alpha e^{-\alpha y}$ com suporte $\mathbb{R}^+$. ◆

Exemplo 6.13. No exemplo 6.2 (página 323) analisamos o ponto imagem

$$y = f(x) = \begin{cases} (b-a) \sqrt{\frac{x}{2}} + a & 0 \leq x \leq 0.5 \\ (a-b) \sqrt{\frac{1-x}{2}} + b & 0.5 < x \leq 1 \end{cases} \quad (6.101)$$

do ponto livre x lançado, à sorte, no intervalo $(0, 1)$. Uma vez que a função inversa de f , função f^{-1} , é dada por

$$f^{-1}(y) = \begin{cases} 2 \frac{(y-a)^2}{(b-a)^2} & a \leq y \leq \frac{a+b}{2} \\ 1 - 2 \frac{(y-b)^2}{(b-a)^2} & \frac{a+b}{2} < y \leq b \end{cases}, \quad (6.102)$$

a lei de possibilidade do ponto imagem y será

$$\pi_y(y) = \begin{cases} \frac{4}{(b-a)^2} (-a+y) & a \leq y \leq \frac{a+b}{2} \\ \frac{4}{(b-a)^2} (b-y) & \frac{a+b}{2} < y \leq b \end{cases}. \quad (6.103)$$

◆

Notemos que, nos casos em que a função de transformação utilizada para definir o ponto imagem é decrescente, as expressões apresentadas sofrerão uma ligeira correcção. Assim, se considerarmos uma função f decrescente e a vizinhança do ponto $y_0 \in \mathbf{Y}$ da forma $\mathbf{V}_{y_0} = [y_0 - \epsilon, y_0 + \epsilon]$, a correspondente

vizinhança do ponto x_0 será $\mathbf{V}_{x_0} = [f^{-1}(y_0 + \epsilon), f^{-1}(y_0 - \epsilon)]$, verificando-se

$$\varpi_y(\mathbf{V}_{y_0}) = \varpi_x(\mathbf{V}_{x_0}) = \frac{f^{-1}(y_0 - \epsilon) - f^{-1}(y_0 + \epsilon)}{\mu(\mathbf{X})} \quad (6.104)$$

e, conseqüentemente,

$$\begin{aligned} \pi_y(y_0) &= \lim_{\epsilon \rightarrow 0} \frac{\varpi_y(\mathbf{V}_{y_0})}{\Delta \mathbf{V}_{y_0}} = \\ &= \lim_{\epsilon \rightarrow 0} \frac{\frac{f^{-1}(y_0 - \epsilon) - f^{-1}(y_0 + \epsilon)}{\mu(\mathbf{X})}}{2\epsilon} = -\frac{(f^{-1}(y_0))'}{\mu(\mathbf{X})}. \end{aligned} \quad (6.105)$$

Assim, uma vez que nesta situação $(f^{-1}(y_0))' < 0$ (se f é decrescente então f^{-1} também o é), podemos, de uma forma geral, utilizar

$$\pi_y(y_0) = \left| \frac{(f^{-1}(y_0))'}{\mu(\mathbf{X})} \right| \quad (6.106)$$

que é válido quer a função f seja crescente quer seja decrescente.

Exemplo 6.14. No exemplo 6.3 (página 324) analisamos o ponto imagem $y = -\ln(x)$ do ponto livre x lançado, à sorte, no intervalo $(e^{-1}, 1)$. Dado que a função f é decrescente, sendo a sua inversa $x = f^{-1}(y) = e^{-y}$, e que $\mu(X) = 1 - e^{-1}$, obtemos para a lei de possibilidade do ponto imagem y

$$\pi_y(y) = \frac{e^{-y}}{1 - e^{-1}}. \quad (6.107)$$

◆

Exemplo 6.15. No exemplo 6.4 (página 325) analisamos o ponto imagem $y = \sigma \tan(x)$, com $\sigma > 0$, do ponto livre x lançado, à sorte, no intervalo $(-\frac{\pi}{2}, \frac{\pi}{2})$. A função inversa de f é $x = f^{-1}(y) = \arctan(\frac{y}{\sigma})$ e, como $\mu(X) = \pi$, obtemos

$$\pi_y = \frac{\sigma}{\pi(\sigma^2 + y^2)}. \quad (6.108)$$

◆

Exemplo 6.16. No exemplo 6.5 (página 325) analisamos o ponto imagem $y = \frac{b}{x^{\frac{1}{a}}}$, com $b > 0$, do ponto livre x lançado, à sorte, no intervalo $(0, 1)$. A inversa da função f é $x = \left(\frac{b}{y}\right)^a$ e, por conseguinte, a lei de possibilidade do ponto imagem y será

$$\pi_y(y) = \frac{ab^a}{y^{a+1}}. \quad (6.109)$$



Apesar de Pacheco d'Amorim não ter explorado esta ideia, notemos que, generalizando o resultado patente na fórmula (6.106) para $\mathbb{R}^n$, substituindo $\left| \left(f^{-1}(y_0) \right)' \right|$ pelo determinante da matriz Jacobiana da função f^{-1} no ponto $y_0 = (y_1, \dots, y_n)$, que representaremos por $\mathcal{J}_{f^{-1}}(y_0)$, definida por

$$\mathcal{J}_{f^{-1}}(y_0) = \frac{\partial (f_1^{-1}, \dots, f_n^{-1})}{\partial (y_1, \dots, y_n)} = \begin{vmatrix} \frac{\partial f_1^{-1}}{\partial y_1} & \dots & \frac{\partial f_1^{-1}}{\partial y_n} \\ \dots & \dots & \dots \\ \frac{\partial f_n^{-1}}{\partial y_1} & \dots & \frac{\partial f_n^{-1}}{\partial y_n} \end{vmatrix}, \quad (6.110)$$

obteríamos

$$\pi_y(y_0) = \left| \frac{\mathcal{J}_{f^{-1}}(y_0) \left(f^{-1} \right)}{\mu(\mathbf{X})} \right| \quad (6.111)$$

onde x e y são vectores de $\mathbb{R}^n$. Esta generalização poderia facilmente ser aplicada ao método de Box-Muller para deduzir a lei de possibilidade do seu ponto imagem.

Exemplo 6.17. No método Box-Muller, apresentado em nota de rodapé na página 330, o ponto imagem é $y = (y_1, y_2)$ definido por

$$\begin{cases} y_1 &= \sqrt{-2 \ln(x_1)} \cos(2\pi x_2) \\ y_2 &= \sqrt{-2 \ln(x_1)} \sin(2\pi x_2) \end{cases} \quad (6.112)$$

e o correspondente ponto livre $x = (x_1, x_2)$ é lançado, à sorte, na região

$(0, 1) \times (0, 1)$. A função inversa de f é

$$f^{-1} = \begin{cases} x_1 &= e^{-\frac{y_1^2 + y_2^2}{2}} \\ x_2 &= \frac{1}{2\pi} \arctan\left(\frac{y_2}{y_1}\right) \end{cases}, \quad (6.113)$$

sendo a lei de possibilidade (conjunta) obtida através da Jacobiana

$$\begin{aligned} \pi_y((y_1, y_2)) &= \left| \begin{array}{cc} \frac{\partial f_1^{-1}}{\partial y_1} & \frac{\partial f_1^{-1}}{\partial y_2} \\ \frac{\partial f_2^{-1}}{\partial y_1} & \frac{\partial f_2^{-1}}{\partial y_2} \end{array} \right| = \frac{1}{2\pi} e^{-\frac{y_1^2 + y_2^2}{2}} = \\ &= \frac{1}{\sqrt{2\pi}} e^{-\frac{y_1^2}{2}} \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{y_2^2}{2}}, \end{aligned} \quad (6.114)$$

uma vez que $\mu(\mathbf{X}) = 1$. ◆

Depois de deduzida a lei de possibilidade do ponto imagem $y = f(x)$, podemos determinar a possibilidade de qualquer região $\mathbf{Y}' \subset \mathbf{Y}$, sem ser necessário recorrer ao ponto livre x , nem à sua região correspondente $\mathbf{X}' = f^{-1}(\mathbf{Y}')$, pois a possibilidade de uma região $\mathbf{Y}' \subset \mathbf{Y}$ é dada por

$$\varpi_y(\mathbf{Y}') = \int_{\mathbf{Y}'} \pi_y(y) \, dy. \quad (6.115)$$

Demonstração. Por definição de ponto imagem $y = f(x)$ temos

$$\varpi_y(\mathbf{Y}') = \varpi_x(f^{-1}(\mathbf{Y}')) = \varpi_x(\mathbf{X}') = \int_{\mathbf{X}'} \pi_x(x) \, dx. \quad (6.116)$$

Uma vez que x é um ponto livre lançado em $\mathbf{X}$, verifica-se

$$\pi_x(x) = \frac{1}{\mu(\mathbf{X})}. \quad (6.117)$$

Como $y = f(x) \Leftrightarrow x = f^{-1}(y) \Leftrightarrow dx = |(f^{-1}(y))'| \, dy$, concluímos que

$$\begin{aligned} \int_{\mathbf{X}'} \pi_x(x) \, dx &= \int_{\mathbf{X}'} \frac{1}{\mu(\mathbf{X})} \, dx = \int_{f^{-1}(\mathbf{X}')} \frac{1}{\mu(\mathbf{X})} |(f^{-1}(y))'| \, dy = \\ &= \int_{\mathbf{Y}'} \pi_y(y) \, dy, \end{aligned} \quad (6.118)$$

obtendo-se, desta forma, o resultado pretendido. □

Realcemos que, com este resultado, conhecida a lei de possibilidade de um ponto imagem, já não precisamos de recorrer ao correspondente ponto livre para determinar qualquer probabilidade ou possibilidade do ponto imagem. Desta forma, a lei de possibilidade tem o papel actual de função densidade e a fórmula (6.115) corresponde a

$$\mathbb{P}(Y \in \mathbf{A}) = \int_{\mathbf{A}} f_Y(y) \, dy, \quad (6.119)$$

onde $f_Y(y)$ representa a função densidade de probabilidade da variável aleatória contínua Y .

Pacheco d'Amorim define ainda a lei de probabilidade de forma análoga à lei de possibilidade. Considerando que o ponto livre x é lançado à sorte na região $\mathbf{X}$ (região total possível) e a região $\mathbf{X}'$ (região possível em relação à probabilidade pretendida) é uma região contida na primeira e que contém a vizinhança $\mathbf{V}_{x_0}$ do ponto x_0 , então a função de probabilidade φ do ponto imagem $y = f(x)$, sendo $y_0 = f(x_0)$, é dada por

$$\varphi_{\mathbf{Y}'}(y_0) = \lim_{\Delta \mathbf{V}_{y_0} \rightarrow 0} \frac{\mathbb{P}_{\mathbf{Y}'}(\mathbf{V}_{y_0})}{\Delta \mathbf{V}_{y_0}}, \quad (6.120)$$

onde $\mathbf{Y}' = f(\mathbf{X}')$ e $\mathbf{V}_{y_0} = f(\mathbf{V}_{x_0})$. Da definição de lei de probabilidade podemos deduzir que

$$\begin{aligned} \varphi_{\mathbf{Y}'}(y_0) &= \lim_{\Delta \mathbf{V}_{y_0} \rightarrow 0} \frac{\frac{\varpi_y(\mathbf{V}_{y_0})}{\varpi_y(\mathbf{Y}')}}{\Delta \mathbf{V}_{y_0}} = \lim_{\Delta \mathbf{V}_{y_0} \rightarrow 0} \frac{\frac{\varpi_y(\mathbf{V}_{y_0})}{\Delta \mathbf{V}_{y_0}}}{\int_{\mathbf{Y}'} \pi_y(y) \, dy} = \\ &= \frac{\pi_y(y_0)}{\int_{\mathbf{Y}'} \pi_y(y) \, dy} = \frac{\pi_y(y_0)}{\varpi_y(\mathbf{Y}')} \end{aligned} \quad (6.121)$$

Desta forma concluímos que a lei de probabilidade é proporcional à lei de possibilidade, isto é, utilizando a nomenclatura actualmente usual, a função densidade condicional é proporcional à função densidade (incondicional),

$$f_{X|X \in \mathbf{A}}(x) = \frac{f_X(x)}{\int_{\mathbf{A}} f_X(x) \, dx} = \frac{f_X(x)}{\mathbb{P}(X \in \mathbf{A})}. \quad (6.122)$$

Notemos que uma (grande) vantagem desta construção de Pacheco d'Amorim, apesar de não conter todo o formalismo que deveria, é em primeiro lugar permitir passar de um ponto livre x , caracterizado pela equiprobabilidade, para um outro ponto y , denominado ponto imagem, que é caracterizado pela sua lei de possibilidade π_y . Depois de efectuada a transformação, através da aplicação da função f , não teremos que preocupar-nos com o seu domínio $\mathbf{X}$ mas apenas com o seu contradomínio $\mathbf{Y}$ e a função de possibilidade π_y que caracteriza o lançamento do ponto imagem y . Esta é a mesma ideia que usamos nos nossos dias em Probabilidade quando aplicamos uma variável aleatória X .

Na actualidade modelamos as experiências aleatórias na Teoria da Probabilidade e Estatística tipicamente da seguinte forma. Começamos por caracterizar a experiência aleatória por um espaço de probabilidade definido pelo terno $(\Omega, \mathcal{A}, \mathbb{P})$, onde Ω representa o espaço-amostra ou espaço de resultados (conjunto constituído pelos possíveis resultados da experiência aleatória), $\mathcal{A}$ o espaço dos acontecimentos (σ -álgebra gerada por Ω) e $\mathbb{P}$ a medida de probabilidade associada ao par $(\Omega, \mathcal{A})$ satisfazendo os axiomas de Kolmogoroff (1933). Este espaço de probabilidade tem, assim, toda a estrutura matemática necessária para ser o alicerce da Teoria da Probabilidade. Contudo, para atingirmos o nosso objectivo de modelar experiências aleatórias, é conveniente transpor a medida de probabilidade para outro espaço mais apelativo de operar, existindo para tal as variáveis aleatórias que correspondem a funções mensuráveis em Teoria da Medida. As variáveis aleatórias são uma aplicação $X(\cdot)$ do espaço de probabilidade $(\Omega, \mathcal{A}, \mathbb{P})$ no espaço $(\mathbb{R}, \mathcal{B}, \mathbb{P}_X)$, isto é,

$$(\Omega, \mathcal{A}, \mathbb{P}) \xrightarrow{X(\cdot)} (\mathbb{R}, \mathcal{B}, \mathbb{P}_X).$$

Com esta aplicação deixamos de trabalhar num espaço de resultados abstracto Ω e transpomos a medida de probabilidade para conjuntos de números reais $\mathbb{R}$. O conjunto $\mathcal{B}(\mathbb{R})$ representa a σ -álgebra de Borel sobre $\mathbb{R}$ e $\mathbb{P}_X$ é

a medida associada aos borelianos (a medida $\mathbb{P}$ transposta para $\mathbb{R}$). Deste modo, em $(\Omega, \mathcal{A}, \mathbb{P})$ a equiprobabilidade é uma assunção razoável, e que serve de base para probabilizar (levando a discriminação do espaço amostra até à fase de equiprováveis). Mas em $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathbb{P}_X)$, como X não é bijectiva, em certo sentido há uma aglutinação de ω_i de Ω conducentes a uma imagem probabilizável, e borelianos (acontecimentos de $\mathbb{R}$) distintos têm probabilidades distintas. Assim, um aspecto fundamental na perspectiva que temos sobre as funções $X : \Omega \longrightarrow \mathbb{R}$ é interessar-nos mais o contradomínio que o domínio — o qual, a partir de certa fase, é completamente omissso. O que está em causa é uma hierarquização de borelianos mais ou menos prováveis, essencial para a tomada racional de decisões. Temos assim duas fases essenciais do estudo do que é aleatório:

1. Em Ω , apostando na equiprobabilidade, construir os modelos de base;
2. Em $\mathbb{R}$, usando $X : \Omega \longrightarrow \mathbb{R}$, e, antes e depois disso, usando condicionamentos, modificação dos modelos por forma a obter a hierarquização favorável a tomar decisões.

O ganho parece a princípio modesto, porque passamos de uma medida de probabilidade $\mathbb{P}$ para a medida de probabilidade $\mathbb{P}_X$ — e as medidas, sendo funções de conjunto, são pouco familiares. Mas como $\mathcal{B}(\mathbb{R})$ pode ser gerado pelas semi-rectas $] - \infty, x], \forall x \in \mathbb{R}$, a função de distribuição

$$F_X(x) = \mathbb{P}_X(] - \infty, x]) = \mathbb{P}(\{\omega \in \Omega : X(\omega) \leq x\})^{(21)} \quad (6.123)$$

é uma função usual, cujo argumento x é um ponto, o que permite uma familiaridade de tratamento que é, de facto, um passo de gigante. Deste

⁽²¹⁾ Kolmogoroff (1933) definiu a função de distribuição através de $F_X(x) = \mathbb{P}(X < x)$, embora o raciocínio seja análogo. Decidimos apresentar desta forma por ser a que habitualmente é utilizada na actualidade.

modo, em vez de utilizarmos o espaço de probabilidade, podemos associar a cada experiência aleatória uma variável aleatória X que é caracterizada pela sua função de distribuição — $F_X(x)$.

Assim, pensamos que Pacheco d'Amorim, apesar de a sua formalização ser ainda muito rudimentar, uma vez que a Teoria da Medida era uma área ainda muito incipiente na altura em que defende a sua tese de doutoramento, esteve próximo de obter algumas das ideias actualmente centrais da Teoria da Probabilidade, tais como a variável aleatória e a função de distribuição, conceitos estes que são posteriores a sua tese de doutoramento⁽²²⁾.

Pacheco d'Amorim atinge o objectivo de passar de um espaço caracterizado pela equipossibilidade para um outro que não o é, não sendo necessário o conhecimento do espaço original para a análise do ponto imagem. Esta ideia é clara nas aplicações que Pacheco d'Amorim desenvolve na CONCLUSÃO da sua tese (analisadas no capítulo 9), pois o autor não se preocupa com a origem dos dados (em que região foi lançado o ponto livre), apenas com a lei de possibilidade dos dados que observa (que serão a imagem de um qualquer ponto livre que não precisamos de caracterizar), pois basta-nos conhecer o domínio (suporte) do ponto imagem y (região $\mathbf{Y}$) e a sua lei de possibilidade π_y para caracterizar o fenómeno em análise. Tal e qual como na Estatística actual onde não nos preocupamos com o espaço de probabilidade original dos dados, apenas com a função de distribuição F_x que caracteriza

⁽²²⁾ Segundo David e Edwards (2001), no apêndice intitulado “*First (?) Occurrence of Common Terms in Statistics and Probability*”, o conceito função de distribuição foi pela primeira vez utilizado por Richard von Mises (1883–1953) em 1919 (*Verteilungsfunktion*) e em Inglês (*Distribution function*) por Joseph Leo Doob (1910–2004) em 1935. Em relação ao conceito variável aleatória identificam a sua origem em Francesco Paolo Cantelli (1875–1966) (*variabile casuale*) num artigo publicado em 1916, sendo este conceito utilizado pela primeira vez em Inglês (*random variable*) num artigo de Aurel Wintner (1903–1958) em 1934.

o espaço de probabilidade obtido após a aplicação de uma qualquer variável aleatória X .

6.4 Leis *a priori* e leis *a posteriori*

Nesta secção⁽²³⁾ vamos supor que estamos a trabalhar com um ponto $M(x, y)$ que varia numa região plana $\mathbf{A} \subset \mathbb{R}^2$ sendo $\pi(x, y)$ a sua lei de possibilidade⁽²⁴⁾. Vamos representar por $m_X(\mathbf{A})$ a projecção da região $\mathbf{A}$ sobre o eixo X , $m_Y(\mathbf{A})$ a projecção sobre o eixo Y , $\mathbf{A}_{X=x}$ a região formada pelos pontos de $\mathbf{A}$ cuja abcissa assume o valor x e $\mathbf{A}_{Y=y}$ a região formada pelos pontos de $\mathbf{A}$ cuja ordenada assume o valor y .

6.4.1 Leis *a priori*

Pacheco d'Amorim define a **lei de probabilidade *a priori*** do ponto $M(x, y)$ como sendo a lei da probabilidade das suas projecções m_X e m_Y .

Assim, a probabilidade de que o ponto m_X (projecção do ponto M) caia na vizinhança $\mathbf{V}_x$ de x (com comprimento $\Delta \mathbf{V}_x$ e contida em $m_X(\mathbf{A})$) é dada pela probabilidade de que o ponto M caia dentro da região $\mathbf{B}$ (destacada na Figura 6.1) constituída pelos pontos de $\mathbf{A}$ com abcissa em $\mathbf{V}_x$. Desta forma obtém-se

$$\mathbb{P}_{\mathbf{A}}(\mathbf{V}_x) = \int_{\mathbf{B}} \pi(x, y) d(x, y) = \Delta \mathbf{V}_x \int_{m_Y(\mathbf{A}_{X=x'})} \pi(x', y) dy \quad (6.124)$$

⁽²³⁾ Esta secção foi escrita com base em Santos (2005).

⁽²⁴⁾ $M(x, y)$ pode ser um ponto livre, e neste caso $\pi(x, y)$ será constante, como pode ser um ponto imagem e $\pi(x, y)$ pode assumir a forma de qualquer função resultante da aplicação de (6.71) (página 354).

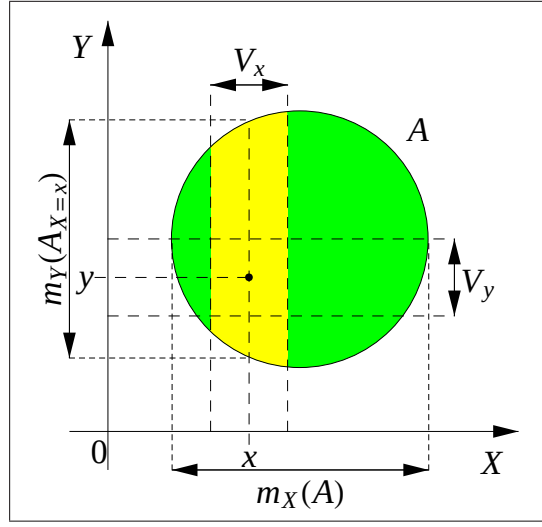


Figura 6.1: Leis à priori

onde $x' \in \mathbf{V}_x$ ⁽²⁵⁾. Deste resultado podemos concluir que

$$a_x(x) = \lim_{\Delta \mathbf{V}_x \rightarrow 0} \frac{\mathbb{P}_{\mathbf{A}}(\mathbf{V}_x)}{\Delta \mathbf{V}_x} = \int_{m_Y(\mathbf{A}_{X=x})} \pi(x, y) \, dy, \quad (6.125)$$

e, de forma análoga, através de

$$\mathbb{P}_{\mathbf{A}}(\mathbf{V}_y) = \int_{\mathbf{B}'} \pi(x, y) \, d(x, y) = \Delta \mathbf{V}_y \int_{m_X(\mathbf{A}_{Y=y'})} \pi(x, y') \, dx \quad (6.126)$$

onde $\mathbf{V}_y$ representa uma vizinhança de y com comprimento $\Delta \mathbf{V}_y$ e $y' \in \mathbf{V}_y$, podemos concluir

$$a_y(y) = \lim_{\Delta \mathbf{V}_y \rightarrow 0} \frac{\mathbb{P}_{\mathbf{A}}(\mathbf{V}_y)}{\Delta \mathbf{V}_y} = \int_{m_X(\mathbf{A}_{Y=y})} \pi(x, y) \, dx, \quad (6.127)$$

que correspondem às leis de probabilidade *a priori* do ponto $M(x, y)$.

⁽²⁵⁾ As deduções aqui apresentadas recorrem ao Teorema do Valor Intermédio, isto é, para qualquer função contínua $f : [a, b] \rightarrow \mathbb{R}$ se escolhermos um valor qualquer γ situado entre $f(a)$ e $f(b)$, então existe um ponto intermédio $c \in [a, b]$ tal que $f(c) = \gamma$.

6.4.2 Leis *a posteriori*

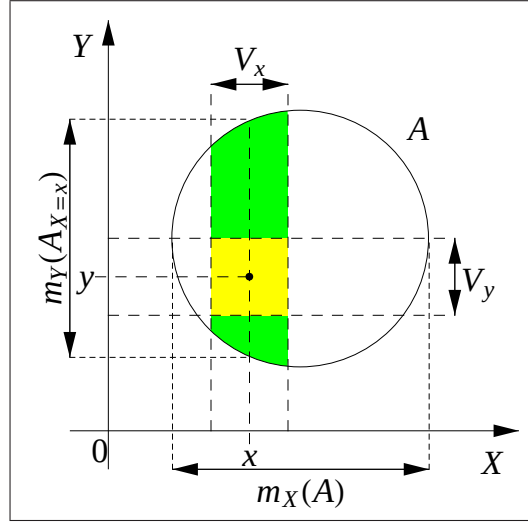


Figura 6.2: Leis à posteriori

A probabilidade de que um ponto M caia na região $\mathbf{C} = \mathbf{V}_x \times \mathbf{V}_y$ em relação à região $\mathbf{B}$, com $\mathbf{C} \subset \mathbf{B}$ (ambas estas regiões estão representadas na Figura 6.2), é determinada através de

$$\begin{aligned} \mathbb{P}_{\mathbf{B}}(\mathbf{C}) &= \frac{\varpi_{\mathbf{C}}}{\varpi_{\mathbf{B}}} = \frac{\int_{\mathbf{C}} \pi(x, y) \, d(x, y)}{\int_{\mathbf{B}} \pi(x, y) \, d(x, y)} = \frac{\Delta \mathbf{V}_x \cdot \Delta \mathbf{V}_y \pi(x'', y'')}{\Delta \mathbf{V}_x \int_{m_Y(\mathbf{A}_{X=x'})} \pi(x', y) \, dy} = \\ &= \frac{\Delta \mathbf{V}_y \pi(x'', y'')}{\int_{m_Y(\mathbf{A}_{X=x'})} \pi(x', y) \, dy} \end{aligned} \quad (6.128)$$

onde $(x'', y'') \in \mathbf{C}$ e $x' \in \mathbf{V}_x$. A probabilidade *a posteriori* de $\mathbf{V}_y$ é definida por

$$\mathbb{P}_x(\mathbf{V}_y) = \lim_{\Delta \mathbf{V}_x \rightarrow 0} \mathbb{P}_{\mathbf{B}}(\mathbf{C}) = \frac{\Delta \mathbf{V}_y \pi(x, y'')}{\int_{m_Y(\mathbf{A}_{X=x})} \pi(x, y) \, dy} \quad (6.129)$$

onde $y'' \in \mathbf{V}_y$, que representa a probabilidade de a ordenada do ponto M pertencer ao intervalo $\mathbf{V}_y$ quando a sua abcissa assume o valor particular x . A correspondente lei de probabilidade *a posteriori*, representada por $p(y | x)$,

é obtida através de

$$p(y | x) = \lim_{\Delta \mathbf{V}_y \rightarrow 0} \frac{\mathbb{P}_x(\mathbf{V}_y)}{\Delta \mathbf{V}_y} = \frac{\pi(x, y)}{\int_{m_Y(\mathbf{A}_{X=x})} \pi(x, y) dy}. \quad (6.130)$$

A lei de probabilidade *a posteriori* $p(x | y)$, obtida de forma análoga, é

$$p(x | y) = \frac{\pi(x, y)}{\int_{m_X(\mathbf{A}_{Y=y})} \pi(x, y) dx}. \quad (6.131)$$

6.5 Probabilidade composta e Teorema de Bayes

Com base nas leis de probabilidade *a priori* e *a posteriori* podemos facilmente deduzir o Teorema da Probabilidade Composta,

$$\pi(x, y) = p(y | x) \cdot a_x(x) = p(x | y) \cdot a_y(y), \quad (6.132)$$

e o Teorema de Bayes,

$$p(x | y) = \frac{p(y | x) \cdot a_x(x)}{\int_{m_X(\mathbf{A}_{Y=y})} p(y | x) \cdot a_x(x) dx} \quad (6.133)$$

e

$$p(y | x) = \frac{p(x | y) \cdot a_y(y)}{\int_{m_Y(\mathbf{A}_{X=x})} p(x | y) \cdot a_y(y) dy}. \quad (6.134)$$

Assim, podemos determinar a possibilidade de qualquer região $\mathbf{A}' \subset \mathbf{A}$ através de

$$\int_{\mathbf{A}'} \pi(x, y) d(x, y) = \begin{cases} \int_{m_X(\mathbf{A}')} a_x(x) \left(\int_{m_Y(\mathbf{A}'_{X=x})} p(y | x) dy \right) dx \\ \int_{m_Y(\mathbf{A}')} a_y(y) \left(\int_{m_X(\mathbf{A}'_{Y=y})} p(x | y) dx \right) dy \end{cases}, \quad (6.135)$$

onde podemos resolver o integral da função $\pi(x, y)$, sobre a região plana $\mathbf{A}'$, por iteração, isto é, utilizando a mesma ideia presente no Teorema de Fubinni-Tonelli

$$\int_{\mathbb{R}^2} dF_{X,Y}(x, y) = \begin{cases} \int_{\mathbb{R}} \left(\int_{\mathbb{R}} dF_{Y|X}(y) \right) dF_X(x) \\ \int_{\mathbb{R}} \left(\int_{\mathbb{R}} dF_{X|Y}(x) \right) dF_Y(y) \end{cases} . \quad (6.136)$$

Esta construção de Pacheco d'Amorim possibilita, de facto, a determinação de integrais da função de possibilidade definidos em regiões planas através de integrais iterados, projectando a possibilidade de uma das variáveis no eixo que lhe corresponde (pela lei *a priori*) e utilizando na outra variável (correspondente ao outro eixo) a função de possibilidade condicionada ao valor da variável projectada (pela lei *a posteriori*). Sublinhamos a importância desta questão. Não é de estranhar que Pacheco d'Amorim não conhecesse a demonstração rigorosa do Teorema de Fubinni, que permite avaliar um integral de uma região n -dimensional através de n integrais iterados bem como trocar a ordem de integração destes⁽²⁶⁾ — quantos resultados dos últimos 5 anos, e que virão a ser reconhecidos como fundamentais no futuro, escapam a cada um de nós. Há, por isso, um mérito excepcional no trabalho que desenvolve, que é, de facto, uma reconstrução notável daquela peça fundamental da Teoria do Integral. É, por outro lado, uma peça abonatória de que Pacheco d'Amorim antecipa as ideias notáveis de Fréchet — de que probabilidade é uma faceta de medida e de que os instrumentos para tratar probabilidade e integral são os mesmos — que Kolmogoroff (1933) reconhece terem sido a

⁽²⁶⁾ Este resultado foi demonstrado, apesar de conter erros, por Fubinni em 1907, sendo a primeira demonstração correcta efectuada por Tonelli em 1909, razão pela qual é por vezes denominado Teorema de Fubinni-Tonelli.

origem da sua meditação sobre os fundamentos da Probabilidade.

6.5.1 Fórmulas inversas da de Bayes

Para a situação de um lançamento de um ponto numa região plana, podemos ainda deduzir as fórmulas inversas da de Bayes, i.e.,

$$a_x(x) = \frac{\frac{p(x|y)}{p(y|x)}}{\int_{m_X(\mathbf{A})} \frac{p(x|y)}{p(y|x)} dx}; \quad (6.137)$$

e

$$a_y(y) = \frac{\frac{p(y|x)}{p(x|y)}}{\int_{m_Y(\mathbf{A})} \frac{p(y|x)}{p(x|y)} dy}. \quad (6.138)$$

Demonstração. Recorrendo à formula (6.133) temos

$$p(x | y) = \frac{p(y | x) \cdot a_x(x)}{\int_{m_X(\mathbf{A}_{Y=y})} p(y | x) \cdot a_x(x) dx} \quad (6.139)$$

e consequentemente

$$\begin{aligned} \ln(p(x | y)) &= \\ &= \ln(p(y | x)) + \ln(a_x(x)) - \ln\left(\int p(y | x) \cdot a_x(x) dx\right) \end{aligned} \quad (6.140)$$

que derivando em ordem a x resulta

$$\frac{\frac{\partial p(x|y)}{\partial x}}{p(x | y)} = \frac{\frac{\partial p(y|x)}{\partial x}}{p(y | x)} + \frac{a'_x(x)}{a_x(x)}, \quad (6.141)$$

uma vez que $\int p(y | x) \cdot a_x(x) dx$ não depende de x . Daqui podemos deduzir que

$$\frac{a'_x(x)}{a_x(x)} = \frac{\frac{\partial p(x|y)}{\partial x}}{p(x | y)} - \frac{\frac{\partial p(y|x)}{\partial x}}{p(y | x)} \Leftrightarrow a_x(x) = k(y) \frac{p(x | y)}{p(y | x)}, \quad (6.142)$$

onde $k(y)$ é uma qualquer função de y . Esta função pode ser determinada através de

$$\begin{aligned} \int_{m_X(\mathbf{A})} a_x(x) \, dx &= k(y) \int_{m_X(\mathbf{A})} \frac{p(x|y)}{p(y|x)} \, dx = 1 \Leftrightarrow \\ \Leftrightarrow k(y) &= \frac{1}{\int_{m_X(\mathbf{A})} \frac{p(x|y)}{p(y|x)} \, dx} \end{aligned} \quad (6.143)$$

e, conseqüentemente, teremos

$$a_x(x) = \frac{\frac{p(x|y)}{p(y|x)}}{\int_{m_X(\mathbf{A})} \frac{p(x|y)}{p(y|x)} \, dx}. \quad (6.144)$$

De forma análoga também podemos deduzir

$$a_y(y) = \frac{\frac{p(y|x)}{p(x|y)}}{\int_{m_Y(\mathbf{A})} \frac{p(y|x)}{p(x|y)} \, dy}. \quad (6.145)$$

□

Pacheco d'Amorim, nesta dedução das fórmulas inversas de Bayes, já não utilizou raciocínios errôneos, tal como tinha feito para determinar as correspondentes fórmulas no caso discreto.

6.6 Comentário geral ao capítulo

Este capítulo é um dos pontos mais inovadores apresentados por Pacheco d'Amorim na sua tese de doutoramento. A ideia de distinguir claramente um ponto lançado à sorte numa região de um outro que é sua imagem é verdadeiramente genial e permite resolver o paradoxo de Bertrand da atribuição de probabilidades distintas a acontecimentos equivalentes. Além disso, toda a sua construção está extremamente bem concebida, apesar de ainda não

estar munida dos conceitos da Teoria da Medida, que na altura constituíam uma área muito incipiente, necessários para a correcta afinação das suas conclusões.

A concepção do ponto imagem permite-nos transformar uma distribuição uniforme, onde é aplicável a definição geométrica de probabilidade, numa qualquer distribuição contínua. Pensamos que, caso Pacheco d'Amorim tivesse utilizado a actual noção de função de distribuição⁽²⁷⁾, teria chegado ainda mais longe, simplificando significativamente a sua análise. Contudo, a sua concepção insere a ideia de que podemos obter qualquer distribuição através da transformação de um lançamento numa região, ideia essencial na actualidade e fundamento da simulação.

Por outro lado, este conceito permite nas aplicações apenas nos preocuparmos com o ponto imagem y que caracteriza o fenómeno que estamos a observar, não sendo necessário investigarmos o ponto livre que lhe dá origem. Desta forma, se nada conhecermos acerca do fenómeno que estamos a analisar, apenas teremos que estimar a região $\mathbf{Y}$ (suporte do ponto imagem) e a sua lei de possibilidade π_y . Esta ideia é semelhante à de variável aleatória e de função de distribuição, dois conceitos centrais nos fundamentos da Teoria da Probabilidade e da Estatística, pois permite-nos abstrair do espaço de probabilidade original associado ao fenómeno em análise, sendo suficiente concentrarmo-nos unicamente na distribuição que caracteriza a variável aleatória associada a esse espaço.

⁽²⁷⁾ Guimarães (1904, p. 34) apresenta uma ideia muito semelhante à de função de distribuição ao definir, no contexto da Teoria dos Erros, a função $F(x)$ como a probabilidade de o erro situar-se entre 0 e x , deduzindo que $p_x = F(x + dx) - F(x) = \varphi(x) dx$ onde $\varphi(x)$ representa a derivada da função $F(x)$ e p_x a probabilidade do erro situar-se entre x e $x + dx$. Recordemos que, considerando $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mu)$ uma medida finita definida sobre $\mathbb{R}$, denominamos F por função de distribuição da medida μ se e só se $\mu([a, b]) = F(b) - F(a)$ para quaisquer valores $a \leq b \in \mathbb{R}$.

Restringindo-se à análise de variáveis definidas numa região em $\mathbb{R}^2$, Pacheco d'Amorim demonstra o Teorema da Probabilidade Composta através da definição de lei *a priori*, que corresponde à projecção da lei de possibilidade (conjunta) num dos eixos, e da lei *a posteriori*, que corresponde a condicionar a lei de possibilidade ao valor obtido numa das variáveis e projectar na outra variável essa possibilidade condicionada. Parece-nos importante salientar que o autor, ao longo deste desenvolvimento, considera que os integrais definidos numa região plana podem ser determinados através de integrais iterados, utilizando como base a noção de projecção numa construção notável do Teorema de Fubinni-Tonelli. Deste modo projectando a probabilidade num dos eixos e tomando a probabilidade (da outra variável) condicionalmente a esse valor obtido na projecção — donde a necessidade de uma classe de eventos $\mathbb{Y}_x$ associada a cada $x \in \mathbb{X}$, ou da classe de eventos $\mathbb{X}_y$ associados a cada $y \in \mathbb{Y}$ — podemos obter o integral em $\mathbb{R}^2$ por iteração

$$\int_{\mathbb{R}^2} dF_{X,Y}(x,y) = \int_{\mathbb{R}} \left(\int_{\mathbb{R}} dF_{Y|X}(y) \right) dF_X(x).$$

Por fim, o autor deduz o Teorema de Bayes e suas fórmulas inversas para os lançamentos, isto é, para o caso contínuo.

Reiteramos a nossa opinião de que, caso a tese de Pacheco d'Amorim tivesse usufruído de divulgação internacional, a ideia do ponto imagem concebida no quarto capítulo da sua tese de doutoramento poderia ter sido a origem de uma maior investigação e, consequente, afinamento desta concepção. A sua tese de doutoramento poderia agora ser, sem dúvida, uma obra de referência na história da Teoria da Probabilidade.

Capítulo 7

Teorema de JACOB BERNOULLI e Lei dos Desvios

No quinto capítulo da sua tese de doutoramento, Pacheco d'Amorim apresenta detalhadamente a dedução da Lei (Frac) dos Grande Números (Teorema de Jacob Bernoulli) e demonstra o Teorema Limite Central (Lei dos Desvios). Contudo, a sua análise destes resultados fundamentais na Teoria da Probabilidade é restrita à situação de somas de provas de Bernoulli independentes com probabilidade de sucesso constante.

Neste sentido, Pacheco d'Amorim começa por analisar a probabilidade dos possíveis resultados quando efectuamos m provas de Bernoulli independentes, deduzindo a sua fórmula (função de probabilidade da distribuição binomial), determinando a sua monotonia, identificando a combinação modal e concluindo, com recurso à fórmula de Stirling, que a probabilidade desta (logo de todas as outras combinações) converge para zero quando aumentamos indefinidamente o número de provas de Bernoulli. Com base nestes resultados e no Lema de Vallée-Poussin, deduz os três Teoremas de Jacob Bernoulli. Começa por analisar a evolução da probabilidade, em m provas

de Bernoulli, em função da distância entre o rácio do número de sucessos n_p e do número de insucessos n_q e o rácio entre a probabilidade de sucesso p e a de insucesso q , i.e. $\left| \frac{n_p}{n_q} - \frac{p}{q} \right|$, onde, em vez de utilizar a usual distância entre a proporção de sucessos observada e a respectiva probabilidade, o autor investiga a probabilidade dos possíveis valores para a distância entre as vantagens (*odds*) observadas e as teóricas. Conclui, desta análise, que a probabilidade varia inversamente com esta distância (Primeiro Teorema de Bernoulli), que, quando aumentamos o número de provas, a probabilidade de qualquer valor, de entre os possíveis valores que esta distância pode assumir, diminui e tanto mais rapidamente quanto maior for esta distância (Segundo Teorema de Bernoulli) e, finalmente, conclui que, seja qual for a distância ε que fixemos, por mais pequeno que seja o valor de ε , a probabilidade de que a distância verificada seja inferior a ε em m provas de Bernoulli, quando aumentamos indefinidamente o número de provas, tenderá para a unidade (Terceiro Teorema de Bernoulli), que é equivalente à Lei Fraca dos Grandes Números.

Posteriormente, fundamentando-se no Lema de Vallée-Poussin, Pacheco d'Amorim infere que o número de experiências em relação aos afastamentos α (distância entre o número de sucessos observado em m provas e o valor modal) é de segunda ordem, começando por deduzir que existe uma probabilidade nula de que a ordem do número de experiência em relação ao afastamento α seja inferior à segunda, através de $\lim_{m \rightarrow \infty} \mathbb{P}(\alpha^n < k m) = 1$, para qualquer valor k positivo e $n < 2$. Depois o autor demonstra que também existe uma probabilidade nula de que a ordem do número de experiências em relação ao afastamento α seja superior à segunda, através de $\lim_{m \rightarrow \infty} \mathbb{P}(\alpha^n < k m) = 0$, para quaisquer $k > 0$ e $n > 2$. Desta forma, Pacheco d'Amorim justifica que a ordem do número de provas tem de ser de segunda em relação aos afastamentos, deduzindo de seguida o Teorema Limite Central de Moivre-

-Laplace ao demonstrar que $\lim_{m \rightarrow \infty} \mathbb{P}(\alpha^2 < k/m) = \Theta(k, \alpha)$, onde Θ é uma função de distribuição de uma variável aleatória não degenerada, isto é, que a expressão $\frac{\alpha^2}{m}$ converge em distribuição para uma lei, que corresponde à lei de Gauss, resultado este que apelida por Lei dos Desvios⁽¹⁾.

Pacheco d'Amorim apresenta ainda uma análise à probabilidade de um número escolhido, à sorte, no intervalo $[0, 1]$ ser racional, um problema típico da época, concluindo que esta probabilidade é nula. O autor refere ainda que Borel apresenta uma generalização da Lei dos Desvios para as situações em que a probabilidade de sucesso das provas de Bernoulli não é constante (igual em todas as provas).

7.1 Resultados preliminares

Pacheco d'Amorim começa esta secção por descrever uma prova de Bernoulli para posteriormente deduzir as características de m provas e, por fim, analisar o que acontece quando o número de provas aumenta indefinidamente.

Assim, consideremos que numa tiragem, à sorte, de um elemento numa classe finita de elementos (conforme analisado no capítulo 3 das CLASSES FINITAS) ou num lançamento, à sorte, de um ponto numa região limitada (conforme analisado no capítulo 4 da PROBABILIDADE CONTÍNUA), podemos

⁽¹⁾ A Lei dos Desvios corresponde ao Teorema Limite Central, neste caso restrito à convergência da distribuição binomial para a gaussiana, quando aumentamos indefinidamente o número de provas de Bernoulli, resultado usualmente denominado por Teorema de Moivre-Laplace. A denominação Teorema Limite Central só surgiu em 1920 através de George Pólya (1887–1985), por considerar que este teorema assume um papel central entre os resultados sobre convergência, sendo por isso fundamental na Teoria da Probabilidade. Por vezes esta denominação é interpretada de forma errónea, ao considerar-se que o nome deriva de o resultado apresentar a convergência do centro (média) dos dados.

dividir os possíveis resultados por dois acontecimentos (classes ou regiões) que denominaremos por favorável (ou sucesso) e contrário (ou insucesso). Representaremos estes acontecimentos respectivamente por $\mathbb{F}$ e $\overline{\mathbb{F}}$. Consideremos então que p representa a probabilidade do acontecimento favorável e q a do acontecimento contrário que, naturalmente, verificam $p + q = 1$.

Notemos que esta aplicação abarca ideia semelhante à concebida por Pacheco d'Amorim no capítulo PONTO IMAGEM, ainda que num contexto diferente, pois o conceito *ponto imagem* concebido pelo autor só se aplica a transformações obtidas através da aplicação de uma função f que estabeleça uma correspondência biunívoca e contínua de uma região $\mathbf{X}$ para outra região $\mathbf{Y}$. Na situação aqui apresentada, o autor transforma uma tiragem à sorte de um elemento de uma classe com um número qualquer (finito) de elementos (que será caracterizada pela equipossibilidade) numa outra tiragem à sorte de um elemento de uma classe constituída por apenas dois elementos (sucesso e insucesso) que poderão ou não ser equipossíveis (dependendo da transformação aplicada). Por exemplo, se a classe inicial for constituída por n elementos, dos quais n_B são associados a bola branca e os restantes n_P a bola preta (com $n_B + n_P = n$), obteríamos como probabilidades $p = \frac{n_B}{n}$ e $q = \frac{n_P}{n}$. Desta forma apenas conseguimos atribuir valores racionais para a probabilidade de sucesso p .

Na segunda ideia apresentada por Pacheco d'Amorim, o autor estabelece uma correspondência entre o lançamento à sorte de um ponto numa região limitada (caracterizado pela equipossibilidade) e uma tiragem à sorte de um elemento de uma classe constituída por apenas dois elementos (sucesso e insucesso). Desta forma teremos uma transformação do tipo

$$f(x) = \begin{cases} \mathbb{F} & x \in \mathbf{X}_1 \\ \overline{\mathbb{F}} & x \in \mathbf{X}_2 \end{cases} \quad (7.1)$$

onde $\mathbf{X}_1 \cup \mathbf{X}_2 = \mathbf{X}$ e $\mathbf{X}_1 \cap \mathbf{X}_2 = \emptyset$ (i.e., $\mathbf{X}_1$ e $\mathbf{X}_2$ definem uma partição de

$\mathbf{X}$), sendo $p = \frac{\mu_{\mathbf{x}_1}}{\mu_{\mathbf{x}}}$ e $q = \frac{\mu_{\mathbf{x}_2}}{\mu_{\mathbf{x}}}$. Desta forma é possível obter qualquer valor para a probabilidade p (racional ou não). Assim, nesta situação, a aplicação f tem como domínio a região $\mathbf{X}$ e como contradomínio dois elementos: $\mathbb{F}$ e $\overline{\mathbb{F}}$ (logo, não é uma aplicação bijectiva e, conseqüentemente, não se enquadra nas definições de ponto imagem fornecidas pelo autor).

Como referimos na análise ao capítulo PONTO IMAGEM da tese de doutoramento de Pacheco d'Amorim, o autor poderia ter explorado outras situações que não as transformações de regiões em regiões, tais como estes dois exemplos aqui retratados, onde temos uma transformação de uma classe noutra classe (onde os elementos da segunda classe não são necessariamente caracterizados pela equipossibilidade) e de uma região numa classe (onde a probabilidade de cada elemento pode assumir qualquer valor real no intervalo $[0, 1]$ e não apenas valores racionais). Contudo, para este tipo de aplicações ser definido seriam necessários outros argumentos além dos apresentados na definição de lei de possibilidade (secção 6.3 presente na página 353), que têm como base a derivada da função f utilizada na transformação. Recordemos, em relação à transformação de uma tiragem de um elemento à sorte numa classe numa outra tiragem noutra classe, que a função de probabilidade de uma variável aleatória discreta é a derivada de Radon-Nikodym da medida de contagem⁽²⁾.

⁽²⁾ Consideremos que a variável aleatória X transforma o espaço de probabilidade $(\Omega, \mathcal{A}, \mathbb{P})$ no espaço $(\mathbb{R}, \mathcal{B}, \mathbb{P}_X)$, que é caracterizado pela função de distribuição $F_X(x)$ para qualquer $x \in \mathbb{R}$. Pelo Teorema de Radon-Nikodym existe uma única função $f \in L_1$ (excepto num conjunto de medida nula) tal que, para qualquer $A \subset \mathcal{B}$, se verifica $\mathbb{P}_X(A) = \int_A f \, d\mu$ onde μ é uma medida absolutamente contínua em relação a $\mathbb{P}_X$, relação que é habitualmente representada através de $\mu \ll \mathbb{P}_X$, e significa que $\mu(A) = 0$ para todo o conjunto $A \in \mathcal{B}$ tal que $\mathbb{P}_X(A) = 0$. Então, se X for uma variável aleatória contínua, a sua função densidade de probabilidade f será a derivada de Radon-Nikodym da medida $\mathbb{P}_X$ em relação à medida de Lebesgue μ (pois teremos $f = \frac{d\mathbb{P}_X}{d\mu}$); no caso de X ser uma

Deste modo, se efectuarmos uma prova com estas características (prova de Bernoulli), podem ocorrer duas situações: $\mathbb{F}$ ou $\overline{\mathbb{F}}$. Se realizarmos duas provas nestas condições teremos então quatro distintas possibilidades de resultado: $\mathbb{F} \mathbb{F}$, $\mathbb{F} \overline{\mathbb{F}}$, $\overline{\mathbb{F}} \mathbb{F}$ ou $\overline{\mathbb{F}} \overline{\mathbb{F}}$. Se repetirmos o fenómeno três vezes, o número de resultados distintos possíveis seria dado pelos arranjos com repetição dos dois objectos, $\mathbb{F}$ e $\overline{\mathbb{F}}$, três a três. Desta forma podemos obter o número de resultados distintos para qualquer número de provas de Bernoulli. Supondo que a produção de um fenómeno não altera as probabilidades associadas às restantes provas⁽³⁾, podemos concluir o teorema a seguir apresentado.

Teorema 7.1. *A probabilidade de que, em m provas, ocorram $m - i$ acontecimentos $\mathbb{F}$ e i acontecimentos $\overline{\mathbb{F}}$ segundo uma ordem previamente fixada é dada por*

$$\mathbb{P} = p^{m-i} q^i. \quad (7.2)$$

Este resultado corresponde à probabilidade de, em m provas de Bernoulli, obtermos $m - i$ sucessos e i insucessos numa determinada ordem fixada *a priori*, resultado este que é consequência imediata dos teoremas relativos à probabilidade composta. Assim, quando aumentamos indefinidamente o nú-

variável aleatória discreta, a sua função de probabilidade será dada pela derivada de Radon-Nikodym da medida $\mathbb{P}_X$ em ordem à medida de contagem μ e teremos $f = \frac{d\mathbb{P}_X}{d\mu}$. Refira-se que os resultados necessários para estabelecer estas relações ainda não estavam disponíveis na época de Pacheco d'Amorim. O Teorema de Radon-Nikodym, segundo Dudley (2002), tem a sua primeira versão, restrita ao espaço Euclidiano, apresentada por Johann Radon (1887–1956) em 1913, sendo posteriormente, em 1920, o resultado extendido por Percy Daniell (1889–1946), e a sua forma mais abstracta, central nos fundamentos apresentados por Kolmogoroff em 1933, foi deduzida por Otto Nikodym (1889–1974) unicamente em 1930.

⁽³⁾ Desta forma, tal como acontecera na sua análise à probabilidade em classes e em regiões, Pacheco d'Amorim considera a hipótese de independência, embora não utilize ou defina explicitamente este conceito.

mero de provas de Bernoulli, a probabilidade de qualquer sequência específica irá tender para zero.

Corolário 7.1.1. *A probabilidade de que os fenómenos $\mathbb{F}$ e $\overline{\mathbb{F}}$ se sucedam segundo uma ordem previamente fixada tende para zero quando o número de experiências m aumenta.*

Demonstração. Considerando $r = \max\{p, q\}$, concluímos que

$$0 \leq \mathbb{P} = p^{m-i} q^i \leq r^m, \quad i = 0, \dots, m, \quad (7.3)$$

e, como $r < 1$ (supondo naturalmente que $p \neq 1$ e $p \neq 0$), deduzimos que r^m tende para zero quando m aumenta indefinidamente e consequentemente $\mathbb{P}$ também. $\square$

No Teorema 7.1 deduzimos a probabilidade $\mathbb{P}$ de uma sequência de m provas com $m - i$ sucessos, probabilidade esta que não depende da ordem pela qual se sucedem os seus elementos componentes, dependendo unicamente do número de casos que nela entram de uma e de outra categoria (favorável ou contrária). É o modo como esta variação se comporta que o autor investiga nesta secção, onde, com o objectivo de tornar mais clara a sua exposição, utiliza como imagem figurativa associada a estas experiências a tiragem de bolas de uma urna, que tenha probabilidade de sair bola branca igual a p e a probabilidade de sair bola preta q . O autor está ciente de que nesta experiência específica p e q só podem assumir valores racionais, enquanto da forma que definiu estas provas no início desta secção p e q podem assumir qualquer valor do intervalo $(0, 1)$, conforme referimos previamente. Contudo, apesar de os resultados que deduz serem igualmente válidos para estas situações, Pacheco d'Amorim recorre a esta imagem para tornar mais explícita a sua apresentação, situação que manteremos nesta exposição crítica das ideias de Pacheco d'Amorim.

Teorema 7.2. *Efectuando m tiragens (com reposição) de uma urna que contém bolas brancas e pretas, a probabilidade de obtermos n (com $n \leq m$) bolas brancas e $m - n$ pretas é dada por*

$$\mathbb{P}_{m,n} = \frac{m!}{(m-n)! n!} p^n q^{m-n}, \quad (7.4)$$

onde p e q representam respectivamente as probabilidades de bola branca e bola preta em cada tiragem.

Demonstração. A probabilidade de se verificar uma qualquer sucessão de n bolas brancas e $m - n$ pretas é (pelo Teorema 7.1) dada por

$$\mathbb{P} = p^n q^{m-n}. \quad (7.5)$$

O número de sucessões distintas que se podem obter com n bolas brancas e $m - n$ pretas é dado pelo número de permutações de m objectos, dos quais $m - n$ e n são repetidos, ou seja, por

$$\frac{m!}{(m-n)! n!}. \quad (7.6)$$

Consequentemente teremos

$$\mathbb{P}_{m,n} = \frac{m!}{(m-n)! n!} p^n q^{m-n} = \binom{m}{n} p^n q^{m-n}. \quad (7.7)$$

□

Desta forma Pacheco d'Amorim determina a expressão que nos permite calcular probabilidades nas situações de aplicabilidade da distribuição binomial com m provas de Bernoulli. O autor deduz um corolário deste teorema, ao comparar a fórmula (7.4) com o desenvolvimento do binómio de Newton, deduzido por Isaac Newton (1643–1727) em 1663, de $(p + q)^m$, isto é,

$$(p + q)^m = \sum_{i=0}^m \binom{m}{i} p^i q^{m-i}, \quad (7.8)$$

fórmula esta que pode ser demonstrada facilmente por indução.

Demonstração. Para $m = 0$ temos

$$(p + q)^0 = \binom{0}{0} p^0 q^0 = 1 \quad (7.9)$$

e para $m = 1$

$$(p + q)^1 = \sum_{i=0}^1 \binom{1}{i} p^i q^{1-i} = p + q \quad (7.10)$$

e, uma vez que

$$\sum_{i=0}^m a_i = \sum_{i=1}^{m+1} a_{i-1} \quad (7.11)$$

e

$$\binom{m}{i-1} + \binom{m}{i} = \binom{m+1}{i} \quad (7.12)$$

que corresponde à propriedade fundamental no triângulo de Pascal, concluímos que

$$\begin{aligned} (p + q)^{m+1} &= (p + q) \sum_{i=0}^m \binom{m}{i} p^i q^{m-i} = \\ &= \sum_{i=0}^m \binom{m}{i} p^{i+1} q^{m-i} + \sum_{i=0}^m \binom{m}{i} p^i q^{m-i+1} = \\ &= p^{m+1} + \sum_{i=0}^{m-1} \binom{m}{i} p^{i+1} q^{m-i} + \sum_{i=1}^m \binom{m}{i} p^i q^{m-i+1} + q^{m+1} = \\ &= p^{m+1} + \sum_{i=1}^m \binom{m}{i-1} p^i q^{m-i+1} + \sum_{i=1}^m \binom{m}{i} p^i q^{m-i+1} + q^{m+1} = \\ &= p^{m+1} + \sum_{i=1}^m \left[\binom{m}{i-1} + \binom{m}{i} \right] p^i q^{m+1-i} + q^{m+1} = \\ &= p^{m+1} + \sum_{i=1}^m \binom{m+1}{i} p^i q^{m+1-i} + q^{m+1} = \\ &= \sum_{i=0}^{m+1} \binom{m+1}{i} p^i q^{m+1-i}. \end{aligned} \quad (7.13)$$

□

Corolário 7.2.1. *A probabilidade de tirar n bolas brancas e $m - n$ pretas é dada pelo termo correspondente do desenvolvimento de*

$$(p + q)^m. \quad (7.14)$$

Pacheco d'Amorim analisa depois a monotonia de $\mathbb{P}_{m,n}$ de forma a deduzir para que valor de n a probabilidade $\mathbb{P}_{m,n}$ atinge o seu máximo, isto é, o valor modal da distribuição binomial.

Teorema 7.3. *Supondo o número de provas de Bernoulli m fixo, podemos concluir:*

1. *A probabilidade $\mathbb{P}_{m,n}$ cresce com n desde zero até ao maior número inteiro inferior a $p(m+1)$;*
2. *A probabilidade $\mathbb{P}_{m,n}$ diminui desde o menor número inteiro n superior a $p(m+1) - 1$ até $n = m$;*
3. *Se $p(m+1) \in \mathbb{N}$ então $\mathbb{P}_{m,n}$ terá o seu valor máximo para n igual a $p(m+1) - 1$ e $p(m+1)$, caso contrário o valor máximo de $\mathbb{P}_{m,n}$ verificar-se-á para o único valor natural n compreendido entre $p(m+1) - 1$ e $p(m+1)$.*

Demonstração.

1. Se $n \leq p(m+1)$ temos

$$\frac{1}{p} \leq \frac{m+1}{n} \Leftrightarrow \frac{1}{p} - 1 \leq \frac{m+1}{n} - 1 \Leftrightarrow \frac{q}{p} \leq \frac{m-n+1}{n}, \quad (7.15)$$

logo,

$$1 \leq \frac{m-n+1}{n} \cdot \frac{p}{q} = \frac{\mathbb{P}_{m,n}}{\mathbb{P}_{m,n-1}} \Leftrightarrow \mathbb{P}_{m,n-1} \leq \mathbb{P}_{m,n}. \quad (7.16)$$

2. Se $n \geq p(m+1) - 1$ temos

$$\frac{1}{p} \geq \frac{m+1}{n+1} \Leftrightarrow \frac{p+q}{p} \geq \frac{m+1}{n+1} \Leftrightarrow 1 + \frac{q}{p} \geq \frac{m+1}{n+1} \Leftrightarrow \frac{q}{p} \geq \frac{m-n}{n+1}, \quad (7.17)$$

logo,

$$1 \geq \frac{m-n}{n+1} \cdot \frac{p}{q} = \frac{\mathbb{P}_{m,n+1}}{\mathbb{P}_{m,n}} \Leftrightarrow \mathbb{P}_{m,n} \geq \mathbb{P}_{m,n+1}. \quad (7.18)$$

3. Pelo que foi demonstrado nos pontos **1.** e **2.** deste teorema podemos concluir que a probabilidade $\mathbb{P}_{m,n}$ será máxima para o número inteiro n que satisfaz

$$p(m+1) - 1 \leq n \leq p(m+1), \quad (7.19)$$

que será representado por

$$n = p(m+1) - r \in \mathbb{N}, \quad \text{com } 0 \leq r \leq 1. \quad (7.20)$$

Nas situações em que $p(m+1)$ é um número inteiro existem dois valores que maximizam a probabilidade $\mathbb{P}_{m,n}$, correspondentes às situações onde $r = 0$ e $r = 1$, isto é, para $n = p(m+1) - 1$ e $n = p(m+1)$, pois, nestes casos, temos $\mathbb{P}_{m,p(m+1)-1} = \mathbb{P}_{m,p(m+1)}$ como facilmente se verifica

$$\frac{\mathbb{P}_{m,p(m+1)-1}}{\mathbb{P}_{m,p(m+1)}} = \frac{\frac{m!}{[p(m+1)-1]![q(m+1)]!} p^{p(m+1)-1} q^{q(m+1)}}{\frac{m!}{[p(m+1)]![q(m+1)-1]!} p^{p(m+1)} q^{q(m+1)-1}} = \frac{p(m+1)}{q(m+1)} \cdot \frac{q}{p} = 1,$$

e, como tal, qualquer uma destas probabilidades será máxima. $\square$

Após deduzir o valor modal da distribuição binomial, Pacheco d'Amorim analisa o comportamento da probabilidade desta combinação quando o número de provas de Bernoulli aumenta indefinidamente, recorrendo, para tal, à fórmula de Stirling

$$m! = m^m e^{-m} \sqrt{2\pi m} (1 + \varepsilon_m), \quad (7.21)$$

que fornece uma aproximação do factorial. Contudo, Pacheco d'Amorim não a demonstra nem efectua qualquer referência à sua origem. Esta fórmula de aproximação do factorial foi deduzida por Abraham de Moivre em 1830 no *Miscellanea Analytica*, todavia apenas deduz que $m! \approx km^m e^{-m} \sqrt{m}$, sendo depois o matemático escocês James Stirling (1692–1770) a concluir que $k = \sqrt{2\pi}$, publicando este resultado no mesmo ano em *Methodus Differentialis*. Esta fórmula é fundamental para posteriormente de Moivre deduzir

a aproximação da binomial pela gaussiana. De Moivre refere em 1738, na segunda edição do *The Doctrine of Chances*, que Stirling tinha melhorado a sua fórmula, aplicando-se uma vez mais a lei de eponímia de Stephen Stigler (1980) [*Stigler's Law of Eponymy*], segundo a qual nenhuma descoberta científica recebe o nome do seu inventor. É o que se verifica neste caso, pois, apesar de a fórmula ser original de de Moivre, é denominada fórmula de Stirling. Hald (2003) apresenta detalhadamente a evolução desta fórmula por de Moivre e por Stirling. Acrescentamos, de seguida, uma demonstração mais acessível deste resultado (actualmente existem diversas formas de o deduzir).

Demonstração. A fórmula de Stirling indica que

$$n! \sim \sqrt{2\pi n} n^n e^{-n} \quad (7.22)$$

no sentido que

$$\lim_{n \rightarrow \infty} \frac{n!}{\sqrt{2\pi n} n^n e^{-n}} = 1, \quad (4) \quad (7.23)$$

que é equivalente a ter

$$\begin{aligned} \lim_{n \rightarrow \infty} \log \left(\frac{n!}{\sqrt{2\pi n} n^n e^{-n}} \right) &= 0 \Leftrightarrow \\ \Leftrightarrow \lim_{n \rightarrow \infty} \log(n!) - \left(n + \frac{1}{2} \right) \log(n) + n - \frac{1}{2} \log 2\pi &= 0. \end{aligned} \quad (7.24)$$

Defina-se então

$$d_n = \log(n!) - \left(n + \frac{1}{2} \right) \log(n) + n \quad (7.25)$$

que verifica

$$d_n - d_{n+1} = \left(n + \frac{1}{2} \right) \log \left(\frac{n+1}{n} \right) - 1. \quad (7.26)$$

⁽⁴⁾ Notemos que $\lim_{n \rightarrow \infty} \frac{f(n)}{g(n)} = 1$ não implica que $\lim_{n \rightarrow \infty} [f(n) - g(n)] = 0$, bastando pensarmos nas funções $f(n) = n^2 + n$ e $g(n) = n^2$ que verificam a primeira condição, mas não a segunda.

Feller (1968) também utiliza a mesma expressão de d_n para demonstrar o resultado, justificando a sua utilização como a seguir se explana. Uma vez que a função logarítmica é crescente, teremos, para $k \geq 1$,

$$\int_{k-1}^k \log x \, dx \leq \log k \leq \int_k^{k+1} \log x \, dx, \quad (7.27)$$

que somando de $k = 1$ até n obtemos

$$\int_0^n \log x \, dx \leq \sum_{k=1}^n \log k \leq \int_1^{n+1} \log x \, dx \quad (7.28)$$

que pode ser simplificado para

$$n \log n - n \leq \log(n!) \leq (n+1) \log(n+1) - n. \quad (7.29)$$

Feller considera que uma boa aproximação para $\log(n!)$ é um valor aproximado da média aritmética dos limites de variação da expressão anterior, tendo proposto $(n + \frac{1}{2}) \log(n) - n$ (onde considerou que para valores elevados de n teremos $\log n \approx \log(n+1)$), sendo então d_n o erro cometido.

Para deduzirmos o resultado pretendido, vamos demonstrar que d_n tende para uma constante, para depois verificarmos que essa constante é igual a $\frac{1}{2} \log 2\pi$. Começemos por recordar que para $|t| < 1$ temos

$$\log(1+t) = t - \frac{t^2}{2} + \frac{t^3}{3} - \dots = \sum_{i=1}^{\infty} (-1)^{i+1} \frac{t^i}{i}, \quad (7.30)$$

logo,

$$\begin{aligned} \log\left(\frac{1+t}{1-t}\right) &= \log(1+t) - \log(1-t) = \\ &= 2\left(t + \frac{t^3}{3} + \frac{t^5}{5} + \dots\right) = 2 \sum_{i=1}^{\infty} \frac{t^{2i-1}}{2i-1}. \end{aligned} \quad (7.31)$$

Assim, podemos utilizar este resultado na expressão de $d_n - d_{n+1}$ (fórmula (7.26)) e, uma vez que

$$\frac{n+1}{n} = \frac{1 + \frac{1}{2n+1}}{1 - \frac{1}{2n+1}}, \quad (7.32)$$

utilizando $t = \frac{1}{2n+1}$ em (7.31) deduzimos

$$\begin{aligned} d_n - d_{n+1} &= (2n+1) \left(\frac{1}{2n+1} + \frac{1}{3(2n+1)^3} + \frac{1}{5(2n+1)^5} + \cdots \right) - 1 = \\ &= \frac{1}{3} \frac{1}{(2n+1)^2} + \frac{1}{5} \frac{1}{(2n+1)^4} + \frac{1}{7} \frac{1}{(2n+1)^6} + \cdots > 0. \end{aligned} \quad (7.33)$$

Assim, d_n é decrescente e, uma vez que

$$\begin{aligned} 0 < d_n - d_{n+1} &< \frac{1}{3} \left(\frac{1}{(2n+1)^2} + \frac{1}{(2n+1)^4} + \cdots \right) = \\ &= \frac{1}{3} \frac{1}{(2n+1)^2 - 1} = \frac{1}{12} \left(\frac{1}{n} - \frac{1}{n+1} \right), \end{aligned} \quad (7.34)$$

concluimos que $d_n - \frac{1}{12n}$ é crescente, pois

$$d_n - \frac{1}{12n} < d_{n+1} - \frac{1}{12(n+1)}. \quad (7.35)$$

Dado que d_n é decrescente e $d_n - \frac{1}{12n}$ é crescente, podemos concluir que d_n tende para uma constante K , logo,

$$\lim_{n \rightarrow \infty} d_n = K \quad (7.36)$$

e, consequentemente,

$$\lim_{n \rightarrow \infty} \frac{n!}{\sqrt{n} n^n e^{-n}} = e^K. \quad (7.37)$$

Falta agora apenas concluir que $e^K = \sqrt{2\pi}$ para deduzirmos o resultado pretendido. Recordando a fórmula de Wallis (1616–1703), publicada em 1665⁽⁵⁾,

⁽⁵⁾ Este resultado pode ser demonstrado primitivando por partes indefinidamente $v(n) = \int_0^{\frac{\pi}{2}} (\sin x)^n dx$, onde se conclui a fórmula recursiva $v(n) = \frac{n-1}{n} v(n-2)$ para $n \geq 2$. Uma vez que $v(0) = \frac{\pi}{2}$ e $v(1) = 1$ teremos

$$v(n) = \begin{cases} \frac{\pi}{2} \frac{1 \cdot 3 \cdots (n-3)(n-1)}{2 \cdot 4 \cdots (n-2)n} & n \text{ par} \\ \frac{2 \cdot 4 \cdots (n-3)(n-1)}{3 \cdot 5 \cdots (n-2)n} & n \text{ impar} \end{cases}.$$

Se simplificarmos $v(2n+2) \leq v(2n+1) \leq v(2n)$ (porque v é uma função decrescente,

que corresponde a

$$\begin{aligned}\frac{\pi}{2} &= \frac{2}{1} \cdot \frac{2}{3} \cdot \frac{4}{3} \cdot \frac{4}{5} \cdot \frac{6}{5} \cdot \frac{6}{7} \cdots \frac{2n}{2n-1} \cdot \frac{2n}{2n+1} \cdots = \\ &= \lim_{n \rightarrow \infty} \frac{2^{4n} (n!)^4}{(2n)! (2n+1)!} = \lim_{n \rightarrow \infty} \frac{2^{4n} (n!)^4}{[(2n)!]^2 (2n+1)},\end{aligned}\quad (7.38)$$

podemos utilizar a aproximação

$$n! \approx \sqrt{n} n^n e^{-n} e^K, \quad (7.39)$$

que é consequência de (7.37), e obtemos

$$\begin{aligned}&\lim_{n \rightarrow \infty} \frac{2^{4n} (n!)^4}{[(2n)!]^2 (2n+1)} = \\ &= \lim_{n \rightarrow \infty} \frac{2^{4n} (\sqrt{n} n^n e^{-n} e^K)^4}{(\sqrt{2n} (2n)^{2n} e^{-2n} e^K)^2 (2n+1)} = \\ &= \lim_{n \rightarrow \infty} \frac{n e^{2K}}{2(2n+1)} = \frac{e^{2K}}{4} = \frac{\pi}{2} \Rightarrow e^K = \sqrt{2\pi}.\end{aligned}\quad (7.40)$$

□

Recorrendo à mesma linguagem representativa de retirar m bolas de uma urna com bolas brancas e pretas, designaremos por combinação a totalidade de resultados desta experiência com o mesmo número de bolas brancas e pretas. Pacheco d'Amorim enuncia o teorema, que indica a convergência da probabilidade associada à combinação modal, da seguinte forma.

Teorema 7.4. *A probabilidade da combinação mais provável tende para zero quando o número de tiragens aumenta indefinidamente.*

pois $(\sin x)^{2n+2} \leq (\sin x)^{2n+1} \leq (\sin x)^{2n}$, obtemos

$$\frac{\pi}{2} \times \frac{2n+1}{2n+2} \leq \frac{2}{1} \cdot \frac{2}{3} \cdot \frac{4}{3} \cdot \frac{4}{5} \cdot \frac{6}{5} \cdot \frac{6}{7} \cdots \frac{2n}{2n-1} \cdot \frac{2n}{2n+1} \leq \frac{\pi}{2},$$

bastando tomar os limites quando n aumenta indefinidamente para obter o resultado pretendido.

Demonstração. Segundo o Teorema 7.3 (página 388) a probabilidade da combinação mais provável será dada por⁽⁶⁾

$$\mathbb{P} = \frac{m!}{[p(m+1)-r]![q(m+1)+r-1]!} p^{p(m+1)-r} q^{q(m+1)+r-1} \quad (7.41)$$

visto que, representando por $p(m+1)-r$ o número mais provável de bolas brancas, teremos que

$$m - [p(m+1)-r] = q(m+1)+r-1 \quad (7.42)$$

representará o número correspondente de bolas pretas. Desta forma, recorrendo à fórmula de Stirling,

$$m! = m^m e^{-m} \sqrt{2\pi m} (1 + \varepsilon_m), \quad (7.43)$$

onde ε_m é uma função de m que tende para zero quando m tende para infinito, para substituir os factoriais e desprezando $p-r$ em $p(m+1)-r$ e $q+r-1$ em $q(m+1)+r-1$, o que é legítimo quando analisamos m a crescer indefinidamente, obtemos

$$\mathbb{P} = \frac{m^m e^{-m} \sqrt{2\pi m} (1 + \varepsilon_m) p^{mp} q^{mq}}{(mp)^{mp} e^{-mp} \sqrt{2\pi mp} (1 + \varepsilon_{mp}) (mq)^{mq} e^{-mq} \sqrt{2\pi mq} (1 + \varepsilon_{mq})} \quad (7.44)$$

que, simplificando, resulta em

$$\mathbb{P} = \frac{1 + \xi_m}{\sqrt{2\pi mpq}}, \quad (7.45)$$

onde ξ_m é uma função que tende para zero quando m tende para infinito. Consequentemente, $\mathbb{P}$ tende também para zero quando m aumenta indefinidamente. $\square$

⁽⁶⁾ No caso de haverem duas combinações modais ($r=0$ e $r=1$) ambas fornecerão o mesmo valor para a probabilidade.

7.2 Teoremas de Jacob Bernoulli

Representemos por n_p o número de sucessos resultantes em m provas de Bernoulli e por n_q o respectivo números de insucessos ($n_q = m - n_p$). Pacheco d'Amorim, no Primeiro Teorema de Bernoulli, analisa a evolução das probabilidades para os possíveis valores que o rácio $\frac{n_p}{n_q}$ pode assumir para um determinado número m de provas de Bernoulli fixo.

Teorema 7.5 (Primeiro Teorema de Jacob Bernoulli). *A probabilidade varia inversamente com a distância entre $\frac{n_p}{n_q}$ e $\frac{p}{q}$.*⁽⁷⁾

Demonstração. No ponto número **3.** do Teorema 7.3 (página 388) o número de sucessos n_p mais provável em m provas de Bernoulli é o número inteiro da forma

$$n_p = p(m+1) - r, \quad (0 \leq r \leq 1), \quad (7.46)$$

e o número correspondente de acontecimentos contrários é

$$n_q = q(m+1) + r - 1, \quad (7.47)$$

sendo o quociente destes dois números dado por

$$\begin{aligned} \frac{n_p}{n_q} &= \frac{p(m+1) - r}{q(m+1) + r - 1} = \frac{q}{q} \cdot \frac{p(m+1) - r}{q(m+1) + r - 1} = \\ &= \frac{qp(m+1) - qr}{q(m+1) + r - 1} \cdot \frac{1}{q} = \frac{qp(m+1) + p(r-1) - p(r-1) - qr}{q(m+1) + r - 1} \cdot \frac{1}{q} = \\ &= \frac{p[q(m+1) + r - 1] - p(r-1) - qr}{q(m+1) + r - 1} \cdot \frac{1}{q} = \\ &= \frac{p}{q} + \frac{p(1-r) - qr}{q(m+1) + r - 1} \cdot \frac{1}{q}. \end{aligned} \quad (7.48)$$

⁽⁷⁾ Refira-se que este resultado não é totalmente verdadeiro, pois unicamente se verifica se considerarmos os casos em que há mais sucessos que na combinação modal separadamente das situações em que se verificam menos sucessos, isto é, se separarmos os casos em que a diferença $\frac{n_p}{n_q} - \frac{p}{q}$ é positiva dos que é negativa. Todavia, a apresentação dos três Teoremas de Jacob Bernoulli de Pacheco d'Amorim é análoga à efectuada por Montessus de Ballore (1908).

Supondo que o número de casos favoráveis cresce em uma unidade, a razão de probabilidade imediatamente inferior (pelo ponto número **2.** do Teorema 7.3) é⁽⁸⁾

$$\frac{n_p}{n_q} = \frac{p(m+1) - r + 1}{q(m+1) + r - 1 - 1} = \frac{p}{q} + \frac{p(1-r) - qr + 1}{q(m+1) + r - 1 - 1} \cdot \frac{1}{q}, \quad (7.49)$$

e a imediata a esta, que corresponde à situação em que o número de casos favoráveis cresce em duas unidades a combinação modal, será, recorrendo ao mesmo raciocínio, igual a

$$\frac{n_p}{n_q} = \frac{p(m+1) - r + 2}{q(m+1) + r - 1 - 2} = \frac{p}{q} + \frac{p(1-r) - qr + 2}{q(m+1) + r - 1 - 2} \cdot \frac{1}{q}, \quad (7.50)$$

sendo, quando o número mais provável de sucessos é excedido em α unidades, a razão em questão dada por

$$\begin{aligned} \frac{n_p}{n_q} &= \frac{p(m+1) - r + \alpha}{q(m+1) + r - 1 - \alpha} = \frac{p}{q} + \frac{p(1-r) - qr + \alpha}{q(m+1) + r - 1 - \alpha} \cdot \frac{1}{q} = \\ &= \frac{p}{q} + \frac{p - r + \alpha}{qm - p + r - \alpha} \cdot \frac{1}{q}. \end{aligned} \quad (7.51)$$

Deste resultado podemos concluir que

$$\frac{n_p}{n_q} - \frac{p}{q} = \frac{p - r + \alpha}{qm - p + r - \alpha} \cdot \frac{1}{q}, \quad (7.52)$$

expressão esta que, para $\alpha = 1, 2, \dots, q(m+1) + r - 1$, assume sempre valor positivo (pois no denominador temos uma probabilidade q a multiplicar pelo número de insucessos na situação em análise ($qm - p + r - \alpha$) e o numerador também é sempre positivo uma vez que $p - r > -1$) e, como facilmente se verifica, é crescente com α . Para $\alpha = -1, -2, \dots, -p(m+1) + r$ (situações em que se verifica menos sucessos que na combinação modal) a expressão, apesar de continuar a ser crescente com α , é negativa pelo facto de $p - r < 1$.

⁽⁸⁾ Apesar de Pacheco d'Amorim na página 105 da sua tese de doutoramento derivar correctamente a fórmula de $\frac{n_p}{n_q}$ que acabamos de deduzir, as fórmula que apresenta na página 106 não estão correctas, faltando multiplicar um dos termos por $\frac{1}{q}$.

Assim, uma vez que, pelos números **1.** e **2.** do Teorema 7.3, a probabilidade diminui à medida que $|\alpha|$ aumenta podemos garantir que, nas duas situações analisadas, a probabilidade diminui à medida que $\left| \frac{n_p}{n_q} - \frac{p}{q} \right|$ aumenta.⁽⁹⁾

α	$\left \frac{n_p}{n_q} - \frac{p}{q} \right $	α	$\left \frac{n_p}{n_q} - \frac{p}{q} \right $	α	$\left \frac{n_p}{n_q} - \frac{p}{q} \right $	α	$\left \frac{n_p}{n_q} - \frac{p}{q} \right $
1	0,083	-1	0,077	12	1,846	-12	0,648
2	0,174	-2	0,148	13	2,167	-13	0,684
3	0,273	-3	0,214	14	2,545	-14	0,718
4	0,381	-4	0,276	15	3,000	-15	0,750

Tabela 7.1: Valores de α e $\left| \frac{n_p}{n_q} - \frac{p}{q} \right|$ em 50 provas com $p = 0.5$

Notemos, contudo, que este resultado é verdadeiro se analisarmos os valores positivos e negativos de α separadamente, isto é, para $\alpha > 0$ o valor de $\left| \frac{n_p}{n_q} - \frac{p}{q} \right|$ aumenta com $|\alpha|$ e se $\alpha < 0$ o valor de $\left| \frac{n_p}{n_q} - \frac{p}{q} \right|$ cresce igualmente com o valor de $|\alpha|$. Porém, o crescimento de $\left| \frac{n_p}{n_q} - \frac{p}{q} \right|$ não acompanha o sentido de $|\alpha|$ no caso geral. Esta ideia é facilmente perceptível ao analisarmos os valores presentes na Tabela 7.1, obtidos utilizando uma distribuição binomial com $m = 50$ provas e probabilidade de sucesso $p = 0.5$. Nesta distribuição, apesar de a probabilidade ser igual para valores de α simétricos ($\mathbb{P}_\alpha = \mathbb{P}_{-\alpha}$), em relação às distâncias entre $\frac{n_p}{n_q}$ e $\frac{p}{q}$ não se observa tal simetria.

Desta forma, podemos concluir que a probabilidade só varia inversamente com a distância, conforme enunciado no teorema, se separarmos em duas situação distintas, quando $\alpha > 0$ e quando $\alpha < 0$. Pacheco d'Amorim parece considerar que a distância $\left| \frac{n_p}{n_q} - \frac{p}{q} \right|$ vai crescendo com o valor de $|\alpha|$ no caso geral, o que não se verifica. $\square$

⁽⁹⁾ Pacheco d'Amorim, para as situação em que há menos sucessos que na combinação modal ($\alpha < 0$), refere que se poderia analisar o quociente $\frac{n_q}{n_p}$ e obteríamos situação análoga à investigada para os casos em que α é positivo.

Após analisar a evolução da probabilidade para os possíveis valores da distância entre os rácios $\frac{n_p}{n_q}$ e $\frac{p}{q}$ supondo um determinado número de provas de Bernoulli fixo, Pacheco d'Amorim investiga o que acontece quando aumentamos o número de provas de Bernoulli indefinidamente, isto é, quando m tende para infinito, deduzindo desta forma o Segundo Teorema de Jacob Bernoulli.

Teorema 7.6 (Segundo Teorema de Jacob Bernoulli). *À medida que se multiplicam as experiências, a probabilidade de cada valor de $\frac{n_p}{n_q}$ vai diminuindo, mas tanto mais rapidamente quanto maior for a distância entre $\frac{p}{q}$ e $\frac{n_p}{n_q}$.*⁽¹⁰⁾

Demonstração. A probabilidade de $\frac{n_p}{n_q}$ será máxima quando o afastamento α for nulo (Teorema 7.5) e, como neste caso ela tende para zero quando m aumenta indefinidamente (Teorema 7.4), naturalmente, tenderão também para zero as probabilidades nas restantes situações. Além disso, temos⁽¹¹⁾

$$\begin{aligned} \frac{\mathbb{P}_{\alpha-1}}{\mathbb{P}_{\alpha}} &= \frac{\frac{m!}{[p(m+1)-r+\alpha-1]![q(m+1)+r-1-(\alpha-1)]!} p^{(m+1)p-r+\alpha-1} q^{(m+1)q+r-1-(\alpha-1)}}{\frac{m!}{[p(m+1)-r+\alpha]![q(m+1)+r-1-\alpha]!} p^{(m+1)p-r+\alpha} q^{(m+1)q+r-1-\alpha}} \\ &= \frac{p(m+1)-r+\alpha}{q(m+1)+r-1-\alpha} \cdot \frac{q}{p} = \frac{A+q\alpha}{B-p\alpha}, \end{aligned} \quad (7.53)$$

onde A e B não dependem de α . Esta relação mostra que, quanto maior for o valor de $|\alpha|$ maior é o valor do quociente $\frac{\mathbb{P}_{\alpha-1}}{\mathbb{P}_{\alpha}}$ e mais rapidamente diminui a probabilidade de cada valor de $\frac{n_p}{n_q}$.

Notemos que a relação expressa em (7.53) pode ser transcrita por

$$\mathbb{P}_{\alpha-1} = \frac{A+q\alpha}{B-p\alpha} \mathbb{P}_{\alpha} = \theta(\alpha) \mathbb{P}_{\alpha}, \quad (7.54)$$

⁽¹⁰⁾ Este resultado necessita da mesma correcção que foi efectuada no Primeiro Teorema de Bernoulli, pois Pacheco d'Amorim parte do princípio (incorrecto) de que $\left| \frac{n_p}{n_q} - \frac{p}{q} \right|$ cresce com o valor de $|\alpha|$.

⁽¹¹⁾ Notemos que Pacheco d'Amorim não efectua correctamente a dedução da expressão de $\frac{\mathbb{P}_{\alpha-1}}{\mathbb{P}_{\alpha}}$. Contudo, a ideia principal era concluir que este rácio é crescente e o autor, apesar de apresentar resultados estranhos, deriva conclusão análoga.

sendo $\theta(\alpha)$ uma função crescente. Para $\alpha \geq 1$ temos $\theta(\alpha) > 1$, uma vez que, nestas situações, se verifica $\mathbb{P}_{\alpha-1} > \mathbb{P}_\alpha$. Como, além disso, $\theta(\alpha)$ é crescente, significará que, quanto maior for o valor de α , maior será o desfasamento entre $\mathbb{P}_{\alpha-1}$ e $\mathbb{P}_\alpha$; logo, além de a probabilidade ir diminuindo com o aumento de α , essa diminuição será cada vez mais acentuada. Se $\alpha \leq -1$ teremos $\theta(\alpha) < 1$, consequência de nestes casos termos $\mathbb{P}_{\alpha-1} < \mathbb{P}_\alpha$. Uma vez que $\theta(\alpha)$ é crescente, conclui-se que, quanto menor for α (que neste caso corresponde a quanto maior for $|\alpha|$), maior será o desfasamento entre $\mathbb{P}_{\alpha-1}$ e $\mathbb{P}_\alpha$. Assim, podemos concluir que, quanto maior for o valor absoluto de α mais rapidamente a probabilidade converge para zero. $\square$

Representemos por $\mathbb{T}_i$ a probabilidade duma combinação que tem i casos favoráveis e por $\mathbb{T}_n$ a combinação mais provável, sendo a probabilidade $\mathbb{P}(\alpha)$, com $\alpha > 0$, definida por

$$\mathbb{P}(\alpha) = \mathbb{T}_{n-\alpha} + \cdots + \mathbb{T}_n + \cdots + \mathbb{T}_{n+\alpha} = \sum_{i=-\alpha}^{\alpha} \mathbb{T}_{n+i}, \quad (7.55)$$

que representa a probabilidade de haver um afastamento absoluto, em relação à combinação mais provável, no máximo igual a α (onde naturalmente α só assume valores positivos). Desta forma, podemos enunciar o Lema de Vallée-Poussin como se segue⁽¹²⁾.

⁽¹²⁾ Pacheco d'Amorim, como usual na sua tese, não efectua qualquer referência (além do seu nome) à origem deste Lema. Montessus de Ballore (1908) apresenta uma construção análoga à efectuada por Pacheco d'Amorim identificando a origem deste resultado num artigo de Charles-Jean de la Vallée-Poussin (1866–1962) intitulado “*Démonstration nouvelle du théorème de Bernoulli*” que foi publicado em 1907 nos *Annales de la Société Scientifique de Bruxelles* **31**, p. 219–236.

Lema 7.1 (Lema de Vallée-Poussin). *Representando por $\mathbb{P}(\alpha)$, com $\alpha > 0$, a probabilidade de haver um afastamento, em valor absoluto, menor ou igual a α em relação à situação modal, teremos então*

$$1 - \mathbb{P}(\alpha) < \frac{m}{\left(1 + \frac{\alpha}{pq(m+1)}\right)^{\frac{\alpha}{2}}}. \quad (7.56)$$

Demonstração. Utilizando um raciocínio análogo ao empregue na demonstração do Segundo Teorema de Bernoulli, podemos facilmente deduzir que

$$\frac{\mathbb{T}_{n+1}}{\mathbb{T}_n} = \frac{\frac{m!}{(n+1)!(m-n-1)!} p^{n+1} q^{m-n-1}}{\frac{m!}{n!(m-n)!} p^n q^{m-n}} = \frac{m-n}{n+1} \cdot \frac{p}{q}, \quad (7.57)$$

onde o número de sucessos que torna a probabilidade máxima, em m provas de Bernoulli, é $n = p(m+1) - r$ (Teorema 7.3) e, conseqüentemente, teremos

$$\begin{cases} n+1 &= p(m+1) - r + 1 \\ m-n &= q(m+1) + r - 1 \end{cases} \quad (7.58)$$

o que, utilizando ainda a substituição

$$1 - r = \varepsilon, (0 \leq \varepsilon \leq 1), \quad (7.59)$$

nos leva a obter

$$\frac{\mathbb{T}_{n+1}}{\mathbb{T}_n} = \frac{q(m+1) - \varepsilon}{p(m+1) + \varepsilon} \cdot \frac{p}{q} = \frac{1 - \frac{\varepsilon}{q(m+1)}}{1 + \frac{\varepsilon}{p(m+1)}}. \quad (7.60)$$

Uma vez que

$$\frac{1 - \frac{\varepsilon}{q(m+1)}}{1 + \frac{\varepsilon}{p(m+1)}} \leq \frac{1}{1 + \frac{\varepsilon}{pq(m+1)}}, \quad (7.61)$$

pois

$$\begin{aligned} & \left(1 - \frac{\varepsilon}{q(m+1)}\right) \left(1 + \frac{\varepsilon}{pq(m+1)}\right) \leq 1 + \frac{\varepsilon}{p(m+1)} \Leftrightarrow \\ \Leftrightarrow & 1 - \frac{\varepsilon}{q(m+1)} + \frac{\varepsilon}{pq(m+1)} - \frac{\varepsilon^2}{pq^2(m+1)} \leq 1 + \frac{\varepsilon}{p(m+1)} \Leftrightarrow \\ \Leftrightarrow & \frac{-pq\varepsilon + q\varepsilon - \varepsilon^2 - \varepsilon q^2}{pq^2(m+1)} \leq 0 \Leftrightarrow -pq\varepsilon + q\varepsilon - \varepsilon^2 - \varepsilon q^2 \leq 0 \Leftrightarrow \\ \Leftrightarrow & -q\varepsilon(p+q) + q\varepsilon - \varepsilon^2 \leq 0 \Leftrightarrow -\varepsilon^2 \leq 0, \end{aligned} \quad (7.62)$$

podemos, portanto, concluir que⁽¹³⁾

$$\frac{\mathbb{T}_{n+1}}{\mathbb{T}_n} \leq \frac{1}{1 + \frac{\varepsilon}{pq(m+1)}}. \quad (7.63)$$

De modo análogo podemos obter a desigualdade seguinte:

$$\begin{aligned} \frac{\mathbb{T}_{n+2}}{\mathbb{T}_{n+1}} &= \frac{(m-n-1)}{n+1+1} \cdot \frac{p}{q} = \frac{q(m+1) - \varepsilon - 1}{p(m+1) + \varepsilon + 1} \cdot \frac{p}{q} = \\ &= \frac{q(m+1) - (1+\varepsilon)}{q(m+1) + (1+\varepsilon)} \cdot \frac{p}{q} < \frac{1}{1 + \frac{\varepsilon+1}{pq(m+1)}}, \end{aligned} \quad (7.64)$$

e, de um modo geral, concluímos que⁽¹⁴⁾

$$\frac{\mathbb{T}_{n+\alpha+1}}{\mathbb{T}_{n+\alpha}} < \frac{1}{1 + \frac{\varepsilon+\alpha}{pq(m+1)}}. \quad (7.65)$$

Multiplicando membro a membro estas desigualdades e considerando posteriormente que $\varepsilon = 0$, obtemos

$$\begin{aligned} \frac{\mathbb{T}_{n+\alpha+1}}{\mathbb{T}_n} &= \frac{\mathbb{T}_{n+1}}{\mathbb{T}_n} \cdot \frac{\mathbb{T}_{n+2}}{\mathbb{T}_{n+1}} \cdots \frac{\mathbb{T}_{n+\alpha+1}}{\mathbb{T}_{n+\alpha}} < \\ &< \frac{1}{1 + \frac{\varepsilon}{pq(m+1)}} \cdot \frac{1}{1 + \frac{\varepsilon+1}{pq(m+1)}} \cdots \frac{1}{1 + \frac{\varepsilon+\alpha}{pq(m+1)}} < \\ &< \frac{1}{1 + \frac{1}{pq(m+1)}} \cdots \frac{1}{1 + \frac{\alpha}{pq(m+1)}} = \\ &= \prod_{x=1}^{\alpha} \frac{1}{1 + \frac{x}{pq(m+1)}}, \end{aligned} \quad (7.66)$$

que, invertendo a ordem a todos os factores deste produto, alcançamos

$$\frac{\mathbb{T}_{n+\alpha+1}}{\mathbb{T}_n} < \prod_{x=1}^{\alpha} \frac{1}{1 + \frac{\alpha+1-x}{pq(m+1)}}. \quad (7.67)$$

(13) A igualdade verifica-se unicamente na situação $\varepsilon = 0$.

(14) As desigualdades seguintes são estritas, uma vez que a igualdade unicamente se verifica quando $\varepsilon = 0$ e simultaneamente $\alpha = 0$.

Multiplicando membro a membro estas duas desigualdades obtemos

$$\left(\frac{\mathbb{T}_{n+\alpha+1}}{\mathbb{T}_n}\right)^2 < \prod_{x=1}^{\alpha} \left[\frac{1}{1 + \frac{x}{pq(m+1)}} \cdot \frac{1}{1 + \frac{\alpha+1-x}{pq(m+1)}} \right], \quad (7.68)$$

que, reparando que os termos presentes no denominador verificam

$$\begin{aligned} & \left[1 + \frac{x}{pq(m+1)}\right] \cdot \left[1 + \frac{\alpha+1-x}{pq(m+1)}\right] = \\ &= 1 + \frac{\alpha+1}{pq(m+1)} + \frac{x(\alpha+1-x)}{[pq(m+1)]^2} > \\ &> 1 + \frac{\alpha+1}{pq(m+1)}, \end{aligned} \quad (7.69)$$

nos leva a deduzir que

$$\begin{aligned} \left(\frac{\mathbb{T}_{n+\alpha+1}}{\mathbb{T}_n}\right)^2 &< \prod_{x=1}^{\alpha} \left[\frac{1}{1 + \frac{x}{pq(m+1)}} \cdot \frac{1}{1 + \frac{\alpha+1-x}{pq(m+1)}} \right] < \\ &< \prod_{x=1}^{\alpha} \left[1 + \frac{\alpha+1}{pq(m+1)}\right]^{-1} = \\ &= \left[1 + \frac{\alpha+1}{pq(m+1)}\right]^{-\alpha} < \\ &< \left[1 + \frac{\alpha}{pq(m+1)}\right]^{-\alpha}, \end{aligned} \quad (7.70)$$

que é equivalente a

$$\mathbb{T}_{n+\alpha+1} < \mathbb{T}_n \left[1 + \frac{\alpha}{pq(m+1)}\right]^{-\frac{\alpha}{2}}, \quad (7.71)$$

e, visto que $\mathbb{T}_n < 1$ (pois é uma probabilidade), concluimos que

$$\mathbb{T}_{n+\alpha+1} < \left[1 + \frac{\alpha}{pq(m+1)}\right]^{-\frac{\alpha}{2}}. \quad (7.72)$$

Por outro lado, como as probabilidades das combinações que não entram em $\mathbb{P}(\alpha)$ assumem como valor máximo $\mathbb{T}_{n+\alpha+1}$ (consequência do Teorema 7.3),

teremos, uma vez que existem $m - (2\alpha + 1)$ possibilidades que não estão incluídas em $\mathbb{P}(\alpha)$, a seguinte desigualdade

$$1 - \mathbb{P}(\alpha) < [m - (2\alpha + 1)]\mathbb{T}_{n+\alpha+1} < m\mathbb{T}_{n+\alpha+1} < \frac{m}{\left[1 + \frac{\alpha}{pq(m+1)}\right]^{\frac{\alpha}{2}}}. \quad (7.73)$$

□

Com base no Lema de Vallée-Poussin, Pacheco d'Amorim deduz o Terceiro Teorema de Bernoulli que é equivalente ao usual Teorema de Bernoulli ou Lei (Frac) dos Grandes Números como Poisson (1837, p. 7) apelidou este teorema fundamental na Teoria da Probabilidade e nas suas aplicações, original de Jacob Bernoulli (1713).

Teorema 7.7 (Terceiro Teorema de Jacob Bernoulli). *Tem-se uma probabilidade sempre crescente de que a razão do número de acontecimentos favoráveis n_p para o número de acontecimentos contrários n_q se não afastará da razão das suas probabilidades respectivas além de certos limites, i.e., representando por $\mathbb{P}_m(\varepsilon)$, com $m \in \mathbb{N}$ e $\varepsilon \in \mathbb{R}^+$ a probabilidade assim definida*

$$\mathbb{P}_m(\varepsilon) = \mathbb{P} \left[\left| \frac{n_p}{n_q} - \frac{p}{q} \right| < \varepsilon \right] \quad (7.74)$$

então $\mathbb{P}_m(\varepsilon)$ é crescente com m e, por mais apertados que estes limites sejam (por mais baixo que seja o valor de ε), a probabilidade de que se trata, aproximar-se-á da unidade tanto quanto se queira, logo que o número de experiências aumente suficientemente, i.e.

$$\lim_{m \rightarrow \infty} \mathbb{P}_m(\varepsilon) = 1, \forall \varepsilon > 0. \quad (7.75)$$

Pacheco d'Amorim considera que este teorema também pode ser assim enunciado: *a probabilidade, de o afastamento ser da ordem do número de experiências, tende para zero quando o número de experiências tende para infinito.* O significado deste enunciado alternativo do teorema tornar-se-á

mais claro após a secção 7.3, onde é analisada a ordem dos afastamentos em relação ao número de provas.

Demonstração. Como vimos na demonstração do Teorema 7.5, Primeiro Teorema de Bernoulli presente na página 395, caso o afastamento α seja positivo, teremos

$$\frac{n_p}{n_q} - \frac{p}{q} = \frac{p - r + \alpha}{mq - p + r - \alpha} \cdot \frac{1}{q} > 0 \quad (7.76)$$

e, portanto, nestas situações verifica-se

$$\left| \frac{n_p}{n_q} - \frac{p}{q} \right| = \frac{p - r + \alpha}{mq - p + r - \alpha} \cdot \frac{1}{q}. \quad (7.77)$$

Assim, quando $\alpha > 0$, teremos para qualquer $\varepsilon > 0$

$$\begin{aligned} \left| \frac{n_p}{n_q} - \frac{p}{q} \right| > \varepsilon &\Leftrightarrow \frac{p - r + \alpha}{mq - p + r - \alpha} \cdot \frac{1}{q} > \varepsilon \Leftrightarrow \\ &\Leftrightarrow p - r + \alpha > mq^2\varepsilon - pq\varepsilon + rq\varepsilon - \alpha q\varepsilon \Leftrightarrow \\ &\Leftrightarrow \alpha(1 + q\varepsilon) > mq^2\varepsilon - pq\varepsilon + rq\varepsilon - p + r \Leftrightarrow \\ &\Leftrightarrow \alpha > \frac{q^2\varepsilon}{1 + q\varepsilon}m + \left(\frac{q^2\varepsilon}{1 + q\varepsilon} - \frac{q^2\varepsilon}{1 + q\varepsilon} \right) + r - p \Leftrightarrow \\ &\Leftrightarrow \alpha > \kappa_A(m + 1) + \kappa_B \end{aligned} \quad (7.78)$$

onde κ_A e κ_B são constantes, sendo que $\kappa_A > 0^{(15)}$. Recorrendo ao Lema de Vallée-Poussin (Lema 7.1, presente na página 400), onde $\mathbb{P}(\alpha)$, com $\alpha > 0$, representa a probabilidade de o afastamento, em valor absoluto, ser menor ou igual que α (e consequentemente $1 - \mathbb{P}(\alpha)$ representará a probabilidade de o afastamento ser superior a α), e utilizando o resultado da expressão (7.78) podemos deduzir

$$\mathbb{P} \left(\left| \frac{n_p}{n_q} - \frac{p}{q} \right| > \varepsilon \right) = 1 - \mathbb{P}(\alpha) < \frac{m}{\left[1 + \frac{\alpha}{pq(m + 1)} \right]^{\frac{\alpha}{2}}} <$$

⁽¹⁵⁾ Pacheco d'Amorim apresenta, como é frequente ao longo deste capítulo, algumas incorrecções no seu desenvolvimento, contudo conclui igualmente que $\alpha > \kappa_A(m + 1) + \kappa_B$ sendo, desta forma, a sua demonstração idêntica à que aqui apresentamos.

$$\begin{aligned}
&< \frac{m}{\left[1 + \frac{\kappa_A(m+1) + \kappa_B}{pq(m+1)}\right]^{\frac{\kappa_A(m+1) + \kappa_B}{2}}} = \\
&= \frac{m}{\left[1 + \frac{\kappa_A}{pq} + \frac{\kappa_B}{pq(m+1)}\right]^{\frac{\kappa_A(m+1) + \kappa_B}{2}}} \quad (7.79)
\end{aligned}$$

que, uma vez que $\kappa_A > 0$, tende para zero quando m tende para infinito.

No que se refere às situações em que $\alpha < 0$, podemos demonstrar este resultado utilizando uma linha de raciocínio paralela. $\square$

O resultado patente no Terceiro Teorema de Jacob Bernoulli é análogo à Lei Fraca dos Grandes Números que é habitualmente apresentada através de

$$\forall \varepsilon > 0 : \lim_{m \rightarrow \infty} \mathbb{P}[|\hat{p} - p| < \varepsilon] = 1, \quad (7.80)$$

onde $\hat{p}$ representa a proporção de sucessos observados em m provas de Bernoulli e p a probabilidade de sucesso em cada prova de Bernoulli. Contudo estas equações são equivalentes, pois, se $\left|\frac{n_p}{n_q} - \frac{p}{q}\right| < \varepsilon$, então existirá um ε' tal que $|\hat{p} - p| < \varepsilon'$ uma vez que

$$\begin{aligned}
&\left|\frac{n_p}{n_q} - \frac{p}{q}\right| = \left|\frac{\hat{p}}{\hat{q}} - \frac{p}{q}\right| = \left|\frac{\hat{p}}{1 - \hat{p}} - \frac{p}{1 - p}\right| = \\
&= \frac{1}{(1 - p)(1 - \hat{p})} |\hat{p}(1 - p) - p(1 - \hat{p})| = \\
&= \frac{1}{(1 - p)(1 - \hat{p})} |\hat{p} - p| \quad (7.81)
\end{aligned}$$

e, consequentemente,⁽¹⁶⁾

⁽¹⁶⁾ Notemos que esta expressão demonstra claramente a ideia que defendemos aquando da apresentação do Primeiro Teorema de Bernoulli, que, enquanto $|\hat{p} - p| = \frac{|\alpha|}{m}$ aumenta com $|\alpha|$, $\left|\frac{n_p}{n_q} - \frac{p}{q}\right|$ não. O termo $(1 - \hat{p})$ que aparece no denominador em (7.81) faz com que esta simetria não se verifique para a expressão $\left|\frac{n_p}{n_q} - \frac{p}{q}\right|$.

$$\mathbb{P}\left(\left|\frac{n_p}{n_q} - \frac{p}{q}\right| < \varepsilon\right) = \mathbb{P}(|\hat{p} - p| < \varepsilon'), \quad (7.82)$$

onde $\varepsilon' = \varepsilon(1-p)(1-\hat{p})$.

No entanto, apesar de estas expressões serem equivalentes, devemos salientar que a análise efectuada por Pacheco d'Amorim é baseada nas vantagens (*odds*), um conceito que, tal como o de probabilidade, permite medir a incerteza. As vantagens de um acontecimento $\mathbf{A}$, que representaremos por $r(\mathbf{A})$, podem ser definidas por

$$r(\mathbf{A}) = \frac{\mathbb{P}(\mathbf{A})}{\mathbb{P}(\overline{\mathbf{A}})} = \frac{\mathbb{P}(\mathbf{A})}{1 - \mathbb{P}(\mathbf{A})}. \quad (7.83)$$

Esta quantidade assume uma relação bijectiva com a probabilidade, uma vez que, conhecendo as vantagens de determinado acontecimento, também podemos determinar a probabilidade desse acontecimento através de

$$\mathbb{P}(\mathbf{A}) = \frac{r(\mathbf{A})}{1 + r(\mathbf{A})}. \quad (7.84)$$

Desta forma, o conhecimento do valor de uma destas quantidades permite-nos determinar o valor da outra. Por exemplo, se estivermos sob a condição de aplicabilidade da definição clássica da probabilidade, a probabilidade de um dado acontecimento $\mathbf{A}$ será determinada pelo rácio entre o número de casos favoráveis à realização do acontecimento $\mathbf{A}$ e o número total de casos possíveis e, conseqüentemente, teremos que as vantagens do acontecimento $\mathbf{A}$ corresponderão ao número de caso favoráveis à realização do acontecimento $\mathbf{A}$ a dividir pelo número de caso desfavoráveis ao acontecimento $\mathbf{A}$.

A vantagens são normalmente associadas a ambientes de apostas, pois, se um jogador apostar um euro no acontecimento $\mathbf{A}$, que tem vantagens $r(\mathbf{A})$, então deverá receber $\frac{1}{r(\mathbf{A})}$ euros se o acontecimento $\mathbf{A}$ ocorrer ou, de forma semelhante, $r(\mathbf{A})$ corresponde à quantidade de euros que o jogador deverá apostar de forma a ganhar um euro (se ganhar a aposta). Notemos que, se

apostarmos E euros num acontecimento $\mathbf{A}$ que ocorrerá com probabilidade p , onde ganhamos S euros se o acontecimento ocorrer e perdemos o dinheiro da aposta caso o acontecimento $\mathbf{A}$ não ocorra, então, para que seja uma aposta justa, o ganho esperado deverá ser nulo, de onde concluímos

$$\mathbb{E}(G) = Sp - E(1 - p) = 0 \Leftrightarrow E = \frac{p}{1 - p}S \Leftrightarrow E = r(\mathbf{A})S. \quad (7.85)$$

Este conceito assume uma grande importância em algumas interpretações de probabilidade, nomeadamente na visão bayesiana personalista desenvolvida por Bruno de Finetti (1937). Para este autor a probabilidade de uma certa hipótese ou proposição pode ser interpretada como o grau de credibilidade do indivíduo relativamente a essa hipótese, podendo estes graus de credibilidade ser determinados pelo comportamento do indivíduo em situações onde existe incerteza, como é o caso de apostas. Segundo Bruno de Finetti, $SP_i(\mathbf{A})$ é definido como sendo o preço máximo que o indivíduo i está disposto a apostar na ocorrência do acontecimento $\mathbf{A}$, onde pode ganhar um prémio de S euros se, e somente se, o acontecimento $\mathbf{A}$ ocorrer. Nas situações em que o ganho da aposta assume uma unidade ($S = 1$), então $\mathbb{P}_i(\mathbf{A})$ é o preço máximo que o indivíduo i está disposto a apostar de forma a obter um euro caso o acontecimento $\mathbf{A}$ ocorra. Notemos que nesta situação o indivíduo i perde $\mathbb{P}_i(\mathbf{A})$ caso o acontecimento $\mathbf{A}$ não ocorra e ganha $1 - \mathbb{P}_i(\mathbf{A})$ caso o acontecimento ocorra, razão pela qual o rácio $r(\mathbf{A})$ destas duas quantidades é denominado por vantagens (*odds*) do acontecimento $\mathbf{A}$. Assim, nesta visão, se as vantagens de um determinado acontecimento $\mathbf{A}$ do indivíduo i são iguais a $\frac{1}{4}$, isto significa que a probabilidade que o indivíduo i associa ao acontecimento $\mathbf{A}$ ocorrer é quatro vezes inferior à probabilidade que associa ao acontecimento $\mathbf{A}$ não ocorrer e, por esta razão, está disposto a apostar um euro no acontecimento $\mathbf{A}$ para ganhar 4 euros no caso de este acontecimento ocorrer.

Na situação analisada, referente a provas de Bernoulli com probabilidade

de sucesso igual a p , as vantagens de ocorrer sucesso terão a forma

$$r(\mathbf{A}) = \frac{p}{1-p} = \frac{p}{q}. \quad (7.86)$$

Se, com base em n provas de Bernoulli, estimarmos as vantagens através das vantagens observadas (ou empíricas) definidas através de

$$\widehat{r(\mathbf{A})} = \frac{\hat{p}}{1-\hat{p}} = \frac{\hat{p}}{\hat{q}} = \frac{\frac{n_p}{n}}{\frac{n_q}{n}} = \frac{n_p}{n_q}, \quad (7.87)$$

então a distância que Pacheco d'Amorim emprega em todo o seu estudo de convergência representa a distância entre as vantagens observadas em n provas de Bernoulli e a vantagens teóricas, pois

$$\left| \frac{n_p}{n_q} - \frac{p}{q} \right| = \left| \widehat{r(\mathbf{A})} - r(\mathbf{A}) \right|. \quad (7.88)$$

Desta forma, o Terceiro Teorema de Bernoulli, na forma apresentada por Pacheco d'Amorim, estabelece que, seja qual for o valor ε que fixemos, a distância entre as vantagens observadas e teóricas será inferior a esse valor, com probabilidade a convergir para a unidade logo que o número de provas tenda para infinito. Esta forma de apresentar a Lei Fraca dos Grandes Números não é usual nas obras contemporâneas da tese de doutoramento de Pacheco d'Amorim. Contudo, uma vez que não conhecemos o conteúdo do artigo de Vallée-Poussin, onde provavelmente Pacheco d'Amorim se baseou, não podemos efectuar uma análise mais pormenorizada sobre a originalidade da sua construção do Terceiro Teorema de Bernoulli. Todavia, antes de comentarmos a sua construção, recordemos, de uma forma simplificada, um pouco da história da Lei Fraca dos Grandes Números⁽¹⁷⁾.

⁽¹⁷⁾ Spanos (1999) faz uma boa apresentação da história dos principais resultados de convergência da Teoria da Probabilidade, sendo, por esta razão, a principal fonte bibliográfica para a resenha histórica que apresentamos. Gnedenko e Kolmogoroff (1954) expõem também, no prefácio do seu livro, um resumo da história dos resultados existentes neste campo.

A primeira Lei dos Grandes Números foi demonstrada por Jacob Bernoulli (1654–1705), tendo sido publicada, em 1713, na sua obra póstuma *Ars Conjectandi*⁽¹⁸⁾. Bernoulli, que considerava este resultado importante, chamou-lhe *Teorema de Ouro*, sendo depois designada por *Lei dos Grandes Números* por Siméon Denis Poisson (1781–1840)⁽¹⁹⁾, em *Recherches sur la Probabilité des Jugements*, obra publicada em 1837, em contrapartida à sua Lei dos Pequenos Números (aproximação da distribuição binomial à distribuição de Poisson para valores de p pequenos). Recorrendo a conceitos actuais, para tornar a exposição mais clara, Bernoulli considerava uma sucessão $\{X_m\}_{m \geq 1}$ de variáveis aleatórias independentes com distribuição de Bernoulli e probabilidade de sucesso p , tendo obtido

$$\lim_{m \rightarrow +\infty} \mathbb{P} \left(\left| \frac{1}{m} \sum_{i=1}^m X_i - p \right| < \varepsilon \right) = 1, \forall \varepsilon. \quad (7.89)$$

Poisson (1837) notou que as hipóteses da Lei dos Grandes Números podiam ser relaxadas, demonstrando que, se considerarmos uma sucessão de variáveis aleatórias independentes $X_1, X_1, \dots, X_m, \dots$ com $X_i \sim \text{Ber}(p_i)$ então

$$\lim_{m \rightarrow +\infty} \mathbb{P} \left(\left| \frac{1}{m} \sum_{i=1}^m X_i - \frac{1}{m} \sum_{i=1}^m p_i \right| < \varepsilon \right) = 1, \forall \varepsilon > 0, \quad (7.90)$$

⁽¹⁸⁾ Esta obra publicada em 1713, oito anos após a sua morte, pelo seu sobrinho Nicolaus Bernoulli é constituída por quatro partes: I – *Tractatum Hugeni De Ratiociniis in Ludo Aleae, Cum Annotationibus Jacobi Bernoulli*, que é uma versão anotada de *De Ratiociniis in Ludo Aleae* de Huygens; II – *Doctrinam de Permutationibus & Combinationibus* onde demonstra o binómio de Newton; III – *Usum Praecedentis Doctrinae in variis Sortitionibus & Ludis Aleae* onde aplica elementos da parte II ao cálculo de probabilidades; IV – *Usum & Applicationem Praecedentis Doctrinae in Civilibus, Moralibus & Oeconomicis* onde desenvolve a Lei Fraca dos Grandes Números, a primeira demonstração rigorosa de um teorema limite.

⁽¹⁹⁾ Bru (2005) faz uma análise das principais contribuições de Poisson para a Teoria da Probabilidade.

não sendo, neste caso, obrigatório que todas as provas de Bernoulli tenham a mesma probabilidade de sucesso.

Posteriormente, em 1867, Pafnuty Chebycheff (1821–1894) relaxou ainda mais as hipótese da Lei Fraca dos Grandes Números, não efectuando imposições sobre a distribuição (não se restringindo à distribuição de Bernoulli), mas sobre a existência do segundo momento, que é a única condição exigida para a aplicação da desigualdade de Bienaymé-Chebycheff. Deste modo, se considerarmos uma sucessão de variáveis aleatórias $X_1, X_1, \dots, X_m, \dots$ independentes duas a duas⁽²⁰⁾, com variância finita $\mathbb{V}ar(X_i) = \sigma_i^2 < c < \infty$ (esta condição garante a existência de valor esperado finito — $\mathbb{E}(X_i) = \mu_i < \infty$ — uma vez que corresponde a um momento de ordem inferior) de forma a garantir que

$$\mathbb{V}ar\left(\frac{1}{m} \sum_{i=1}^m X_i\right) < \frac{c}{m} \xrightarrow{m \rightarrow \infty} 0 \quad (7.91)$$

podemos concluir, nestas condições, que

$$\lim_{m \rightarrow +\infty} \mathbb{P}\left(\left|\frac{1}{m} \sum_{i=1}^m X_i - \frac{1}{m} \sum_{i=1}^m \mu_i\right| < \varepsilon\right) = 1, \forall \varepsilon > 0 \quad (7.92)$$

recorrendo unicamente à desigualdade de Bienaymé-Chebycheff. Salientamos que este resultado, ao contrário dos anteriores de Bernoulli e Poisson, já não se restringe à soma de provas de Bernoulli independentes, sendo válido para a soma de quaisquer variáveis, discretas ou contínuas, cujos momentos satisfaçam as condições apresentadas. Desta forma é dado um passo fundamental na evolução dos Teoremas limites na Teoria da Probabilidade.

Em 1900 Andrei Markov (1856–1922)⁽²¹⁾ relaxou as hipóteses da Lei Fraca dos Grandes Números do seu professor Chebycheff, assumindo apenas a exis-

⁽²⁰⁾ Notemos que esta condições é mais fraca do que impor que as variáveis aleatórias tenham que ser todas independentes entre si. Contudo seria suficiente supor correlação (ou covariância) nula entre as variáveis, isto é, $\mathbb{C}ov(X_i, X_j) = 0$ para $i \neq j$.

⁽²¹⁾ Basharin, Langville e Naumov (2004) apresentam um estudo bibliográfico da vida

tência do momento de ordem $1 + \delta$, isto é considerando $\mathbb{E}(|X_i - \mu_i|^{1+\delta}) < \infty$ para um valor qualquer positivo δ (a existência do momento $1 + \delta$ implica a existência de μ). Posteriormente Markov relaxou a hipótese de independência⁽²²⁾ assumida pelos seus antecessores, assumindo a inexistência de correlação assintótica, isto é, impondo

$$\mathbb{V}ar\left(\frac{1}{m}\sum_{i=1}^m X_i\right) \xrightarrow{m \rightarrow \infty} 0 \quad (7.93)$$

que, sob a hipótese de variância finita, só se verifica se

$$\frac{1}{m^2} \sum_{i=1}^m \sum_{\substack{j=1 \\ j \neq i}}^m \mathbb{C}ov(X_i, X_j) \xrightarrow{m \rightarrow \infty} 0, \quad (7.94)$$

sendo, por este motivo, denominado não correlação assintótica. Em 1918 Bernstein considerou uma sucessão de variáveis aleatórias $\{X_i\}, i = 1, 2, \dots$ com valor esperado e variância finita (seja $\mathbb{E}(X_i) = \mu_i < v < \infty$ e $\mathbb{V}ar(X_i) = \sigma_i^2 < c < \infty$) e cujas correlações satisfaçam

$$\mathbb{C}orr(X_i, X_j) = \rho_{i,j} \leq \rho(|i - j|) \leq 1 \text{ e } \lim_{k \rightarrow \infty} \rho(k) = 0 \quad (7.95)$$

onde $\rho(0) = 1$, de forma a garantir que

$$\mathbb{V}ar\left(\frac{1}{m}\sum_{i=1}^m X_i\right) = \frac{1}{m^2} \sum_{i=1}^m \sum_{j=1}^m \rho_{i,j} \sigma_i \sigma_j \leq \frac{c}{m^2} \sum_{i=1}^m \sum_{j=1}^m |\rho(|i - j|)| \quad (7.96)$$

convirja para zero e, consequentemente, se verifique a condição de não correlação assintótica de Markov.

e obra de Andrei Markov focando, entre outros assuntos, a sua contribuição para a Lei Fraca dos Grandes Números e para as actualmente denominadas Cadeias de Markov.

⁽²²⁾ Em 1902 Pavel Nekrasov (1853–1924) publicou um artigo onde afirmava, erradamente, que a independência das variáveis é uma condição necessária para garantir a Lei Fraca dos Grandes Números. Markov, discordando desta premissa, começou a analisar o comportamento de somas de variáveis onde a hipótese de independência não se verificava, quer para a Lei Fraca dos Grandes Números quer para o Teorema Limite Central.

No final na década de 1920, Kolmogoroff apresenta condições necessárias e suficientes para a Lei Fraca dos Grandes Números, impondo condições ao comportamento da soma e não ao comportamento individual de cada variável (o que torna mais complicado a sua aplicabilidade). Assim, definindo $S_m = \sum_{i=1}^m (X_i - \mathbb{E}(X_i))$, temos

$$\begin{aligned} \lim_{m \rightarrow +\infty} \mathbb{E} \left(\frac{S_m^2}{m^2 + S_m^2} \right) = 0 &\Leftrightarrow \\ \Leftrightarrow \lim_{m \rightarrow +\infty} \mathbb{P} \left(\left| \frac{1}{m} \sum_{i=1}^m X_i - \frac{1}{m} \sum_{i=1}^m \mu_i \right| < \varepsilon \right) = 1, \forall \varepsilon > 0. \end{aligned} \quad (7.97)$$

Finalmente, em 1928, Khintchine demonstra que, se considerarmos uma sucessão de variáveis aleatórias $\{X_i\}, i = 1, 2, \dots$ independentes e identicamente distribuídas com valor esperado finito ($\mathbb{E}(X_i) = \mu < \infty$), então

$$\lim_{m \rightarrow +\infty} \mathbb{P} \left(\left| \frac{1}{m} \sum_{i=1}^m X_i - \mu \right| < \varepsilon \right) = 1, \forall \varepsilon > 0, \quad (7.98)$$

que corresponde à forma de apresentação mais usualmente utilizada.

Apesar de Pacheco d'Amorim apresentar de uma forma bastante detalhada a demonstração dos Teoremas de Bernoulli, parece-nos que deveria ter tentado ir mais longe neste capítulo, uma vez que a demonstração da Lei Fraca dos Grandes Números restrita à situação de provas de Bernoulli com probabilidade de sucesso constante é desenvolvida, com mais ou menos detalhe, em praticamente todas as obras de Probabilidade do século XIX. Mesmo a demonstração desta lei para provas de Bernoulli com probabilidade de sucesso variável já tinha sido demonstrada por Poisson em 1837. Estranha-se, por conseguinte, o facto de Pacheco d'Amorim não analisar esta situação. Os restantes desenvolvimentos da Lei Fraca dos Grandes Números previamente referidos, anteriores à tese de doutoramento de Pacheco d'Amorim, devem-se à escola russa de Probabilidade e, muito provavelmente, eram totalmente desconhecidos por Pacheco d'Amorim. Os resultados da escola russa libertaram a Lei dos Grandes Números da sua restrição à soma de provas de Bernoulli,

generalizando o resultado para somas de quaisquer variáveis. Esta nova abordagem, bastante mais geral que a anterior, levou à criação e desenvolvimento de técnicas bem mais sofisticadas que as necessárias para a análise do caso particular de somas de provas de Bernoulli, baseadas na fórmula de Stirling, única situação que Pacheco d'Amorim analisa nesta secção da sua tese de doutoramento. Contudo, no capítulo ESPERANÇA MATEMÁTICA E VALOR MÉDIO, após definir e apresentar algumas propriedades da esperança matemática e do valor médio, que define de modo distinto do habitual, o autor generaliza o resultado garantindo a convergência da média empírica de uma função f das observações, obtidas numa sequência de experiências com um número finito de resultados ou numa sequência de lançamentos de um ponto em regiões, para o valor médio (valor esperado) dessa função. A sua análise é distinta da usada pela escola russa de probabilidade, pois Pacheco d'Amorim não tenta deduzir a convergência através da imposição de condições aos momentos da distribuição, mas através de uma construção muito semelhante à do integral de Riemann, construção esta que comentaremos na secção 8.6. Porém, mesmo incluindo esta generalização, a análise da Lei Fraca dos Grandes Números por Pacheco d'Amorim é restrita à repetição de provas, quer sejam tiragens em classes finitas quer sejam lançamentos em regiões, idênticas e independentes. Refira-se ainda que estas condições, que actualmente designamos por variáveis i. i. d. (independentes e identicamente distribuídas), são fulcrais nas aplicações do Cálculo das Probabilidades que o autor desenvolve na conclusão da sua tese de doutoramento e que comentaremos na secção 9.

7.3 Ordem de convergência do número de experiências em relação aos afastamentos

Nesta secção Pacheco d'Amorim justifica, de forma original, a ordem de convergência do número de provas m em relação aos afastamentos absolutos α (distância à combinação modal). Deste modo o autor pretende determinar que valor de β faz com que a variável aleatória $\frac{\alpha^\beta}{m}$ convirja em distribuição para uma variável aleatória não degenerada, isto é, determinar o valor de β tal que

$$\lim_{m \rightarrow +\infty} \mathbb{P} \left(\frac{\alpha^\beta}{m} \leq \varepsilon \right) = F_X(\varepsilon) \quad (7.99)$$

onde F_X é uma função de distribuição de uma variável aleatória X não degenerada. No teorema que a seguir é apresentado Pacheco d'Amorim deduz que β não pode ser menor que 2.

Teorema 7.8. *Há uma probabilidade nula de que a ordem do número de experiências em relação ao afastamento absoluto α seja inferior à segunda, isto é, a probabilidade de que o afastamento absoluto α seja tal que*

$$\frac{\alpha^{n+1}}{m^n} > \varepsilon, \quad \forall \varepsilon > 0 \quad (7.100)$$

tende para zero quando m cresce, logo que $n > 1$.

Demonstração. De (7.100) conclui-se

$$\frac{\alpha^{n+1}}{m^n} > \varepsilon \Leftrightarrow \alpha > \varepsilon^{\frac{1}{n+1}} m^{\frac{n}{n+1}} = \varepsilon' m^{\frac{n}{n+1}}, \quad \text{com } n > 1 \quad (7.101)$$

onde $\varepsilon' = \varepsilon^{\frac{1}{n+1}}$. Aplicando o Lema de Vallée-Poussin a este resultado obtemos

$$1 - \mathbb{P}(\alpha) < \frac{m}{\left[1 + \frac{\alpha}{pq(m+1)}\right]^{\frac{\alpha}{2}}} < \frac{m}{\left[1 + \frac{\varepsilon' m^{\frac{n}{n+1}}}{pq(m+1)}\right]^{\frac{\varepsilon' m^{\frac{n}{n+1}}}{2}}}, \quad (7.102)$$

que, desprezando as parcelas finitas em presença das infinitas, pode ser aproximado por

$$\frac{m}{\left[1 + \frac{\varepsilon' m^{\frac{n}{n+1}}}{pq(m+1)}\right]^{\frac{\varepsilon' m^{\frac{n}{n+1}}}{2}}} \approx \frac{m}{\left[1 + \kappa_A m^{-\frac{1}{n+1}}\right]^{\kappa_B m^{\frac{n}{n+1}}}}. \quad (7.103)$$

O denominador presente na fórmula anterior pode ser desenvolvido pelo binómio de Newton, quando consideramos que m tende para infinito, através de

$$\begin{aligned} \left[1 + \kappa_A m^{-\frac{1}{n+1}}\right]^{\kappa_B m^{\frac{n}{n+1}}} &= \sum_{k=0}^{\kappa_B m^{\frac{n}{n+1}}} \binom{\kappa_B m^{\frac{n}{n+1}}}{k} \left(\kappa_A m^{-\frac{1}{n+1}}\right)^k = \\ &= 1 + \kappa_B m^{\frac{n}{n+1}} \left(\kappa_A m^{-\frac{1}{n+1}}\right) + \frac{\kappa_B m^{\frac{n}{n+1}} \left(\kappa_B m^{\frac{n}{n+1}} - 1\right)}{2} \left(\kappa_A m^{-\frac{1}{n+1}}\right)^2 + \\ &+ \dots + \binom{\kappa_B m^{\frac{n}{n+1}}}{j} \left(\kappa_A m^{-\frac{1}{n+1}}\right)^j + \dots \approx \\ &\approx 1 + \kappa_{A_1} m^{\frac{n}{n+1}} m^{-\frac{1}{n+1}} + \kappa_{A_2} m^{\frac{2n}{n+1}} m^{-\frac{2}{n+1}} + \dots + \kappa_{A_j} m^{\frac{jn}{n+1}} m^{-\frac{j}{n+1}} + \dots = \\ &= 1 + \kappa_{A_1} m^{\frac{n-1}{n+1}} + \kappa_{A_2} m^{\frac{2n-1}{n+1}} + \dots + \kappa_{A_j} m^{\frac{jn-1}{n+1}} + \dots \end{aligned} \quad (7.104)$$

onde, na penúltima passagem, foram desprezados termos finitos na presença de infinitos. Assim obtemos

$$\begin{aligned} \frac{m}{\left[1 + \kappa_A m^{-\frac{1}{n+1}}\right]^{\kappa_B m^{\frac{n}{n+1}}}} &\approx \frac{m}{1 + \kappa_{A_1} m^{\frac{n-1}{n+1}} + \kappa_{A_2} m^{\frac{2n-1}{n+1}} + \dots} = \\ &= \frac{1}{m^{-1} + \kappa_{A_1} m^{\frac{n-1}{n+1}-1} + \kappa_{A_2} m^{\frac{2n-1}{n+1}-1} + \dots + \kappa_{A_i} m^{i\frac{n-1}{n+1}-1} + \dots}. \end{aligned} \quad (7.105)$$

Uma vez que $n > 1$, haverá sempre um número inteiro i tal que

$$i\frac{n-1}{n+1} - 1 > 0, \quad (7.106)$$

consequentemente, a expressão (7.105) vai convergir para zero e, portanto, o mesmo se verificará com $1 - \mathbb{P}(\alpha)$. $\square$

Notemos que o Teorema 7.8, original de Pacheco d'Amorim segundo Manuel dos Reis (1929), é equivalente a termos

$$\mathbb{P}\left(\frac{\alpha^\beta}{m} > \varepsilon\right) \rightarrow 0, \quad \forall \varepsilon > 0, \quad (7.107)$$

quando m tende para infinito e $\beta < 2$, logo, nestas situações, $\frac{\alpha^\beta}{m}$ converge para uma distribuição degenerada no ponto zero. O seguinte teorema analisa os casos em que $\beta > 2$.

Teorema 7.9. *Há uma probabilidade nula de que a ordem do número de experiências em relação ao afastamento absoluto α , seja superior à segunda, isto é, a probabilidade de que*

$$\frac{\alpha}{\sqrt[n]{m}} < \varepsilon, \quad \forall \varepsilon > 0 \quad (7.108)$$

tende para zero quando m aumenta indefinidamente, se $n > 2$.

Demonstração. Ao demonstrarmos que a probabilidade da combinação mais provável tende para zero quando aumentamos indefinidamente o número de provas de Bernoulli (Teorema 7.4 na página 393) obtivemos

$$\mathbb{T}_{n-\alpha} < \mathbb{T}_n = \frac{1 + \xi_m}{\sqrt{2\pi pqm}}, \quad (7.109)$$

onde ξ_m é uma função que tende para zero quando m tende para infinito. Consequentemente, uma vez que $\mathbb{P}(\alpha)$ contém $2\alpha + 1$ combinações, teremos

$$\begin{aligned} \mathbb{P}(\alpha) &< (2\alpha + 1)\mathbb{T}_n = \frac{(2\alpha + 1)}{\sqrt{2\pi pqm}}(1 + \xi_m) = \\ &= \frac{\alpha}{\sqrt{m}} \cdot \frac{2(1 + \xi_m)}{\sqrt{2\pi pq}} + \frac{1 + \xi_m}{\sqrt{2\pi pqm}}, \end{aligned} \quad (7.110)$$

ou

$$\mathbb{P}(\alpha) < \frac{\alpha}{\sqrt{m}}\xi_1 + \xi_2, \quad \text{onde } \lim_{m \rightarrow \infty} \xi_1 = \kappa \text{ e } \lim_{m \rightarrow \infty} \xi_2 = 0, \quad (7.111)$$

sendo κ constante. Uma vez que pretendemos analisar as situações nas quais se verifica

$$\frac{\alpha}{\sqrt[n]{m}} < \varepsilon \Leftrightarrow \alpha < \varepsilon \sqrt[n]{m}, \quad \text{com } n > 2, \quad (7.112)$$

teremos então

$$\mathbb{P}(\alpha) < \frac{\alpha}{\sqrt{m}}\xi_1 + \xi_2 < \frac{\varepsilon \sqrt[n]{m}}{\sqrt{m}}\xi_1 + \xi_2 = m^{\frac{1}{n}-\frac{1}{2}} \varepsilon \xi_1 + \xi_2 \quad (7.113)$$

e, consequentemente,

$$\lim_{m \rightarrow +\infty} \mathbb{P}(\alpha) = 0 \quad \text{se} \quad \frac{1}{n} - \frac{1}{2} < 0, \quad (7.114)$$

isto é, $n > 2$. Assim concluimos que é nula a probabilidade de que o número de experiências seja de ordem superior à segunda em relação ao afastamento α . $\square$

O resultado patente no Teorema 7.9 é equivalente a definir uma variável aleatória $\frac{\alpha^\beta}{m}$, com $\beta > 2$, onde α é o afastamento absoluto em relação à combinação modal, e deduzir que para qualquer $\varepsilon > 0$ teremos

$$\mathbb{P}\left(\frac{\alpha^\beta}{m} < \varepsilon\right) \rightarrow 0 \quad (7.115)$$

quando m tende para infinito e $\beta > 2$.

Assim, nos dois teoremas anteriores (Teoremas 7.8 e 7.9) Pacheco d'Amorim deduz que é nula a probabilidade de que o número de experiências seja de ordem inferior ou superior à segunda, em relação ao afastamento absoluto α . Desta forma conclui que a ordem do afastamento terá de ser de segunda.

Teorema 7.10. *A ordem do número de experiências relativas ao afastamento absoluto α será a segunda, i.e.,*

$$\lim_{m \rightarrow +\infty} \mathbb{P}\left(\frac{\alpha^2}{m} \leq \varepsilon\right) = F_X(\varepsilon), \quad (7.116)$$

onde F_X é uma função de distribuição de uma variável aleatória X não degenerada.

Deste modo Pacheco d'Amorim demonstra que, para haver convergência em distribuição, que não seja para uma variável aleatória degenerada, temos que multiplicar os afastamentos absolutos α por $\frac{1}{\sqrt{m}}$, uma vez que

$$\frac{\alpha^2}{m} \leq \varepsilon \Leftrightarrow \frac{\alpha}{\sqrt{m}} \leq \sqrt{\varepsilon} = \varepsilon'. \quad (7.117)$$

Manuel dos Reis (1929), na sua tese de doutoramento, atribui a primazia destes resultados a Pacheco d'Amorim. Segundo este autor a fundamentação de que temos de dividir o afastamento por $\sqrt{m}$ para obtermos a convergência em distribuição é original de Pacheco d'Amorim. Notemos que, considerando X_i , com $i = 1, \dots, m$, variáveis aleatórias independentes com distribuição de Bernoulli, então

$$\alpha = \left| \sum_{i=1}^m X_i - mp \right| \quad (7.118)$$

e os Teoremas 7.8 e 7.9 são equivalentes a ter

$$\lim_{m \rightarrow +\infty} \mathbb{P} \left(\frac{\left| \sum_{i=1}^m X_i - mp \right|^\beta}{m} < \varepsilon \right) = \begin{cases} 1 & \text{se } \beta < 2 \\ 0 & \text{se } \beta > 2 \end{cases}, \forall \varepsilon > 0, \quad (7.119)$$

logo, a ordem do número de experiências é de segunda em relação aos afastamentos, isto é, temos que utilizar $\beta = 2$ para haver convergência em distribuição⁽²³⁾.

Contudo, Manuel dos Reis refere igualmente que estes resultados poderiam ser derivados, de uma forma bem mais acessível, através do Teorema

⁽²³⁾ Refira-se que, até agora, unicamente foi demonstrado que se houver algum valor real para β , tal que $\frac{\alpha^\beta}{m}$ convirja em distribuição para uma variável aleatória não degenerada, então esse valor unicamente poderá ser $\beta = 2$. Desta forma, o Teorema 7.10 só estará realmente demonstrado com os resultados presentes na secção 7.5 onde será deduzida a convergência para $\beta = 2$.

de Moivre-Laplace⁽²⁴⁾, pois, segundo este resultado anterior à tese de doutoramento de Pacheco d'Amorim, temos

$$\frac{\sum_{i=1}^m X_i - mp}{\sqrt{mpq}} \xrightarrow[n \rightarrow \infty]{d} Y \sim \text{gaussiana}, \quad (7.120)$$

onde $X \xrightarrow{d} Y$ significa que a variável aleatória X converge em distribuição para a variável aleatória Y . Consequentemente teremos

$$\lim_{m \rightarrow +\infty} \mathbb{P} \left(\frac{\left| \sum_{i=1}^m X_i - mp \right|}{\sqrt{mpq}} < \varepsilon \right) = 2\Phi(\varepsilon) - 1, \quad (7.121)$$

onde Φ representa a função de distribuição de uma variável aleatória gaussiana standard. Assim, pela simetria da distribuição gaussiana teremos $\Phi(0) = 0.5$ e, uma vez que Φ é uma função de distribuição, teremos ainda

$$\lim_{x \rightarrow +\infty} \Phi(x) = 1. \quad (7.122)$$

Por conseguinte, concluímos que

$$\begin{aligned} & \mathbb{P} \left(\frac{\left| \sum_{i=1}^m X_i - mp \right|^\beta}{m} < \varepsilon \right) = \mathbb{P} \left(\frac{\left| \sum_{i=1}^m X_i - mp \right|}{m^{\frac{1}{\beta}}} < \varepsilon^{\frac{1}{\beta}} \right) = \\ & = \mathbb{P} \left(\frac{\left| \sum_{i=1}^m X_i - mp \right|}{\sqrt{mpq}} < \frac{\varepsilon^{\frac{1}{\beta}}}{\sqrt{pq}} m^{\frac{1}{\beta} - \frac{1}{2}} \right) = \\ & = 2\Phi \left(\frac{\varepsilon^{\frac{1}{\beta}}}{\sqrt{pq}} m^{\frac{1}{\beta} - \frac{1}{2}} \right) - 1. \end{aligned} \quad (7.123)$$

⁽²⁴⁾ Pacheco d'Amorim demonstra este resultado posteriormente, sendo a sua análise comentada na secção 7.5.

Tomando os limites, na expressão (7.123), quando m aumenta indefinidamente obtemos

$$\begin{aligned} \lim_{m \rightarrow +\infty} \mathbb{P} \left(\left| \frac{\sum_{i=1}^m X_i - mp}{m} \right| < \varepsilon \right) &= \lim_{m \rightarrow +\infty} 2\Phi \left(\frac{\varepsilon^{\frac{1}{\beta}}}{\sqrt{pq}} m^{\frac{1}{\beta} - \frac{1}{2}} \right) - 1 = \\ &= \begin{cases} 1 & \text{se } \beta < 2 \\ 2\Phi \left(\sqrt{\frac{\varepsilon}{pq}} \right) - 1 & \text{se } \beta = 2 \\ 0 & \text{se } \beta > 2 \end{cases}, \forall \varepsilon > 0. \end{aligned} \quad (7.124)$$

Notemos que uma parte deste resultado, referente à situação $\beta < 2$ descrita no Teorema 7.8, também poderia ter sido deduzida através da desigualdade de Bienaymé-Chebycheff que, supondo a existência da variância σ^2 da variável aleatória X , corresponde a

$$\mathbb{P}(|X - \mathbb{E}(X)| \geq \varepsilon) \leq \frac{\sigma^2}{\varepsilon^2}, \quad \forall \varepsilon > 0. \quad (7.125)$$

Demonstração. Uma forma possível de demonstrar esta fórmula é considerar que esta desigualdade é um caso particular da desigualdade de Markov que pode ser expressa da forma a seguir apresentada. Seja X uma variável aleatória não negativa tal que $\mathbb{E}(X)$ exista, então

$$\mathbb{P}(X > t) \leq \frac{\mathbb{E}(X)}{t}, \quad \forall t > 0, \quad (7.126)$$

pois

$$\begin{aligned} \mathbb{E}(X) &= \int_0^{+\infty} x dF_X(x) = \int_0^t x dF_X(x) + \int_t^{+\infty} x dF_X(x) \geq \\ &\geq \int_t^{+\infty} x dF_X(x) \geq t \int_t^{+\infty} dF_X(x) = t \mathbb{P}(X > t). \end{aligned} \quad (7.127)$$

A desigualdade de Bienaymé-Chebycheff é um caso particular da de Markov, pois, considerando $Y = (X - \mu)^2$ (que é não negativa para qualquer variável

aleatória X), obtemos

$$\mathbb{P}(|X - \mu| \geq t) = \mathbb{P}((X - \mu)^2 \geq t^2) \leq \frac{\mathbb{E}[(X - \mu)^2]}{t^2} = \frac{\sigma^2}{t^2}. \quad (7.128)$$

□

Deste modo, considerando $X = \sum_{i=1}^m X_i$ na expressão (7.125) (desigualdade de Bienaymé-Chebycheff), teremos $\mathbb{E}(X) = mp$ e $\sigma^2 = mp(1-p)$. Desta forma obtemos

$$\begin{aligned} \mathbb{P}\left(|X - \mathbb{E}(X)|^\beta < \varepsilon m\right) &= \mathbb{P}\left(|X - \mathbb{E}(X)| < \varepsilon' m^{\frac{1}{\beta}}\right) \geq \\ &\geq 1 - \frac{mp(1-p)}{\left(m^{\frac{1}{\beta}} \varepsilon'\right)^2} = 1 - \frac{p(1-p)}{(\varepsilon')^2} m^{1-\frac{2}{\beta}}, \end{aligned} \quad (7.129)$$

que, para qualquer valor $\varepsilon > 0$ (ou $\varepsilon' > 0$ que é equivalente) e para $1 - \frac{2}{\beta} < 0$ (isto é $\beta < 2$), converge para a unidade quando o número de provas tende para infinito, isto é, para $\beta < 2$ verifica-se

$$\lim_{m \rightarrow +\infty} \mathbb{P}\left(\left|\frac{\sum_{i=1}^m X_i - mp}{m}\right|^\beta < \varepsilon\right) = 1, \quad \forall \varepsilon > 0. \quad (7.130)$$

Apesar de existirem outras formas, porventura mais elementares, de demonstrar estes resultados com os conhecimentos disponíveis na época, não deixa de ser inovador que Pacheco d'Amorim se tenha preocupado em investigar esta questão. Para facilitar o nosso comentário, apresentemos uma outra forma equivalente de representar as conclusões expressas nos Teoremas 7.8, 7.9 e 7.10,

$$\lim_{m \rightarrow +\infty} \mathbb{P}\left(\left|\frac{\sum_{i=1}^m X_i - mp}{m^{\frac{1}{\beta}}}\right| < \varepsilon\right) = \begin{cases} 1 & \text{se } \beta < 2 \\ 2\Phi\left(\frac{\varepsilon}{\sqrt{pq}}\right) - 1 & \text{se } \beta = 2 \\ 0 & \text{se } \beta > 2 \end{cases}, \quad (7.131)$$

para qualquer valor de ε positivo. Esta representação salienta algumas semelhanças entre a forma como Pacheco d'Amorim analisa a ordem de convergência e os conceitos concebidos por Paul Lévy alguns anos mais tarde, apesar de estes inserirem uma generalidade e sofisticação bastante superiores à análise de Pacheco d'Amorim, de domínio de atracção de somas de uma variável aleatória X para uma lei estável Y ⁽²⁵⁾.

Definição 7.1 (Lei estável para somas). Seja $X_1, \dots, X_m, \dots$ uma sucessão de variáveis aleatórias independentes e identicamente distribuídas a X e seja $S_m = \sum_{i=1}^m X_i$. Dizemos que X é uma **lei estável para somas** se existirem constantes normalizadoras $a_m \in \mathbb{R}$ e $b_m \in \mathbb{R}^+$ tais que

$$\frac{S_m - a_m}{b_m} \xrightarrow[m \rightarrow \infty]{d} X, \quad (7.132)$$

onde as constantes normalizadoras satisfazem $b_m = km^{\frac{1}{\alpha}}$, com $k > 0$ e $\alpha \in]0, 2]$. Denomina-se α por expoente característico da distribuição X .

Desta forma, as leis estáveis são aquelas que, se somarmos variáveis com essa distribuição, a soma será uma variável com o mesmo tipo de distribuição. Contudo, por vezes a convergência em distribuição da soma de um tipo de variáveis é para outro tipo de distribuição e, neste caso, estaremos a falar de domínios de atracção para somas.

Definição 7.2 (Domínio de atracção para somas). Seja $X_1, \dots, X_m, \dots$ uma sucessão de variáveis aleatórias independentes e identicamente distribuídas a X e seja $S_m = \sum_{i=1}^m X_i$. Dizemos que X está no **domínio de atracção** da variável aleatória (não degenerada) Y (que é necessariamente uma variável estável), que representamos por $X \in \mathcal{D}(Y)$, se existirem constantes de atracção ou normalizadoras $a_m \in \mathbb{R}$ e $b_m \in \mathbb{R}^+$ tais que

$$\frac{S_m - a_m}{b_m} \xrightarrow[m \rightarrow \infty]{d} Y. \quad (7.133)$$

⁽²⁵⁾ Consultar, por exemplo, Gnedenko e Kolmogoroff (1954) ou Lukacs (1970).

Se Y corresponder a uma variável aleatória com distribuição gaussiana, então dizemos que X pertence ao domínio de atracção da distribuição gaussiana e, neste caso, teremos $a_m = m\mathbb{E}(X)$ e $b_m = \sigma m^{\frac{1}{2}}$, onde σ representa o desvio padrão da variável aleatória X . Há resultados com condições necessárias e suficientes para que uma variável aleatória pertença ao domínio de atracção da distribuição gaussiana, nomeadamente

$$\frac{x^2 \int_{|X|>x} dF_X(t)}{\int_{|X|<x} t^2 dF_X(t)} = \frac{x^2 \mathbb{P}(|X| > x)}{\int_{|X|<x} t^2 dF_X(t)} \xrightarrow{x \rightarrow \infty} 0 \quad (7.134)$$

ou, recorrendo à teoria da variação regular de Karamata (1902–1967)⁽²⁶⁾, se o segundo momento truncado da variável aleatória X definido por

$$L(x) = \mathbb{E}(X^2, |X| \leq x) = \int_{-x}^x t^2 dF_X(t), \quad (7.135)$$

for uma função de variação lenta no sentido de Karamata, isto é,

$$\frac{L(tx)}{L(x)} \xrightarrow{x \rightarrow \infty} K \in [0, \infty[. \quad (7.136)$$

De salientar que estes resultados são posteriores à tese de doutoramento de Pacheco d'Amorim e se devem sobretudo aos trabalhos de Paul Lévy (1886–1971), Alexander Khinchin (1894–1959), William Feller (1906–1970), que, de forma independente, em 1935 estabelecem as condições necessárias e suficientes para que uma distribuição pertença ao domínio de atracção da distribuição gaussiana. Anos mais tarde, Boris Gnedenko (1912–1995) e Wolfgang Doeblin (1915–1940) estabelecem condições necessárias e suficientes para que uma distribuição pertença ao domínio de atracção de uma lei estável com expoente $\alpha \in]0, 2]$, onde $\alpha = 2$ corresponde à distribuição gaussiana e $\alpha = 1$ à distribuição de Cauchy. Estas condições são

$$\frac{\mathbb{P}(X \leq -x)}{\mathbb{P}(X > x)} = \frac{F_X(-x)}{1 - F_X(x)} \xrightarrow{x \rightarrow \infty} K \in [0, \infty[, \quad (7.137)$$

⁽²⁶⁾ Consultar, por exemplo, Bingham (2007), Seneta (2002) ou Feller (1971).

e para toda a constante $t > 0$

$$\frac{\mathbb{P}(|X| > tx)}{\mathbb{P}(|X| > x)} = \frac{[1 - F_X(tx) + F_X(-tx)]}{[1 - F_X(x) + F_X(-x)]} \xrightarrow{x \rightarrow \infty} t^{-\alpha} \quad (7.138)$$

isto é, a soma das caudas $\mathbb{P}(|X| > x)$ é uma função de variação regular com expoente $-\alpha$ onde, na expressão (7.133) teremos $b_m = \xi(m) m^{\frac{1}{\alpha}}$ sendo $\xi(m)$ uma função de variação lenta no sentido de Karamata e $\alpha \in (0, 2]$ o expoente característico da variável aleatória limite Y .

Parece-nos, perante isto, que Pacheco d'Amorim está a investigar se a distribuição de Bernoulli pertence ao domínio de atracção da distribuição gaussiana, procurando a forma da expressão de b_m que permita a convergência em distribuição de $\frac{S_m - a_m}{b_m}$ para uma variável aleatória gaussiana, concluindo que $b_m = \sqrt{m}$. A contraposição das ideias concebidas por Pacheco d'Amorim às de Paul Lévy não tem por objectivo equiparar o valor dos conceitos matemáticos concebidos por estes dois autores, nem atenuar as diferenças que os separam. Pretendemos unicamente salientar a ideia filosófica que está por detrás desta análise de Pacheco d'Amorim que, apesar de restrita à situação mais elementar de analisar a soma de provas de Bernoulli independentes com probabilidade constante, insere em si a ideia de identificar a expressão das constantes normalizadoras b_m de forma que se obtenha a convergência em distribuição para a lei de Gauss.

Entendemos referir que não encontrámos esta forma de analisar a ordem de convergência em qualquer outra obra anterior à tese de doutoramento de Pacheco d'Amorim por nós consultada e repetimos o testemunho de Manuel dos Reis (1929), que lhe atribui a primazia deste resultado. Assim, Pacheco d'Amorim, apesar de se restringir à análise da convergência de somas de Bernoulli com probabilidade de sucesso constante, demonstra uma vez mais a sua originalidade na forma como tenta resolver determinados problemas. Neste caso antecipa algumas ideias incutidas nos conceitos domínio de atrac-

ção e leis estáveis, que são posteriores à sua tese, sendo a origem destas ideias habitualmente atribuídas a Paul Lévy⁽²⁷⁾.

7.4 Probabilidade de um número racional

Pacheco d'Amorim dedica esta secção a deduzir que a probabilidade de um número, escolhido à sorte no intervalo $(0, 1)$, ser racional é igual a zero.

O autor começa por referir que qualquer número pode ser escrito, na sua forma decimal, através de um número infinito de casas, exemplificando com o número $\frac{1}{2}$ que corresponderá a $0.5000 \dots$. Pacheco d'Amorim, portanto, não se preocupou com os números que podem ser representados de duas formas infinitas distintas na base decimal, como é o caso de $\frac{1}{2}$ que pode ser escrito através de $0.5000 \dots$ (forma degenerada) e através de $0.49999 \dots$ (forma não degenerada). Assim, o autor não teve o cuidado de restringir as possíveis sequências $\mathcal{B}$, resultantes em infinitas provas de Bernoulli, de forma a obter uma bijecção com os pontos do intervalo $(0, 1)$, conforme efectuámos na secção 6.2.1, onde apresentamos o princípio de Borel e restringimos as representações às formas não degeneradas, pois, utilizando unicamente representações não degeneradas para cada número, garantimos uma relação biunívoca entre cada ponto do intervalo $(0, 1)$ e uma sucessão infinita de provas de Bernoulli com dez resultados possíveis.

Para deduzir a probabilidade de o número escolhido ser racional, Pacheco d'Amorim começa por propor o problema que passamos a expor.

Problema 7.1. Tira-se, à sorte, um número do intervalo $(0, 1)$. Qual a

⁽²⁷⁾ David e Edwards (2001) identificam a origem do conceito de lei estável e domínio de atracção com o trabalho desenvolvido por Paul Lévy na década de 1920. A livro *Calcul des Probabilités* de Paul Lévy, publicado em 1925, é uma obra notável, que introduz e desenvolve diversos conceitos fundamentais na Teoria da Probabilidade (cf. Loève (1973)).

probabilidade de que os seus algarismos se sucedam de modo a satisfazerem a Lei de Bernoulli⁽²⁸⁾?

Pacheco d'Amorim considera que, com as definições fornecidas na sua tese de doutoramento, nomeadamente as concebidas no capítulo PROBABILIDADE CONTÍNUA, não é possível resolver este problema. Desta forma, o autor entende que este problema, que parece pertencente à classe de problemas da probabilidade contínua, uma vez que estamos a escolher um ponto numa região, não pode ser resolvido com as definições apresentadas no capítulo em que estas probabilidades foram analisadas (capítulo 4 da PROBABILIDADE CONTÍNUA), pois só foram definidas probabilidades de regiões em relação a outras regiões e não podemos afirmar *a priori* que os números que satisfazem a lei de Bernoulli formem uma região. Borel (1909a) refere que a probabilidade geométrica, que corresponde à probabilidade contínua de Pacheco d'Amorim, se aplica a conjuntos que, na terminologia de Georg Cantor (1845–1918), têm a *puissance du continu*. De facto, não é imediato identificar se um conjunto, constituído pelos números que na sua representação na base binária apresentam determinada característica, tem ou não tem a dimensão do contínuo. Por exemplo, os números que podem ser representados de forma degenerada formam um conjunto infinito numerável (Adams e Guillemin (1996) demonstram este facto para a representação binária), enquanto o conjunto formado pelos números cuja representação na base ternária não possuem o número 1, denominado conjunto de Cantor, é fechado não numerável e, apesar de ter medida de Lebesgue nula, tem a mesma potência do contínuo, ainda que não contenha nenhum aberto nem seja um boreliano. Como estes exemplos ilustram, Pacheco d'Amorim tem razão nesta sua

(28) Pacheco d'Amorim pretende analisar a probabilidade de que a distância entre as frequências relativas de cada dígito d (com $d = 0, \dots, 9$) e 0.1 se mantenha inferior a ε para qualquer ε positivo.

preocupação de que não podemos afirmar *a priori* que o conjunto favorável forme uma região, para que possamos recorrer às definições de probabilidade contínua.

A forma encontrada pelo autor para resolver este problema foi propor um segundo problema, enunciado de seguida, que considera equivalente ao primeiro se tomarmos o limite de m a tender para infinito.

Problema 7.2. Qual a probabilidade de que os primeiros m algarismos do número, tirado à sorte do intervalo $(0, 1)$, satisfaçam a lei de Bernoulli?

SOLUÇÃO: Este problema já tem solução com as definições de probabilidade contínua, pois os números cujos primeiros m algarismos são idênticos formam uma região. Assim, temos de identificar as regiões favorável e possível e, uma vez que estamos a escolher um número à sorte numa região, intervalo $(0, 1)$, teremos equipossibilidade e, por conseguinte, poderemos recorrer à definição geométrica de probabilidade e esta será igual ao rácio entre a medida (comprimento) da região favorável e da região possível (que é igual à unidade).

Em primeiro lugar notemos que existem 10^m números distintos quando utilizamos m dígitos na base decimal e que a cada uma destas representações está associado um intervalo, cada um deles de amplitude igual a 0.1^m , logo, a probabilidade de ser escolhido cada um destes intervalos é 0.1^m .

Também podemos deduzir esta probabilidade recorrendo às definições de probabilidades em classes finitas. Assim, consideremos que estamos a escolher à sorte cada um dos dígitos do número, isto é, que temos uma urna com dez bolas, numeradas de 0 a 9, e que vamos escolher à sorte e com reposição m bolas. Qualquer número com m algarismos pode ser obtido por meio de m tiragens nessa urna, sendo a possibilidade de cada um desses números igual a 0.1^m , uma vez que em cada tiragem qualquer uma das 10 bolas terá uma

probabilidade de ser a escolhida igual a 0.1. Esta probabilidade pode ser justificada pela probabilidade em classes finitas, uma vez que temos dez elementos equipossíveis, mas também pode ser obtida com recurso à aplicação apresentada na secção 6.2 do capítulo do PONTO IMAGEM, utilizando a fórmula (6.19) presente na página 332, onde consideraríamos o ponto imagem $y = f(x) = x$ e imediatamente obteríamos $\mathbb{P} = 0.1$ para qualquer valor de a e de d .

Uma vez que Pacheco d'Amorim obteve a mesma probabilidade 0.1^m pelos dois métodos referidos, um pela probabilidade contínua utilizando a medida da região formada por cada sequência de m dígitos e o outro pela probabilidade discreta considerando m provas de Bernoulli independentes com probabilidade de sucesso 0.1, o autor conclui que, em relação ao problema em análise, é indiferente lançar à sorte um ponto no intervalo $(0, 1)$, como referido no enunciado, ou fazer m tiragens consecutivas, à sorte e com reposição, numa urna que contenha 10 bolas numeradas de 0 a 9. Contudo, nesta última formalização do problema, sabemos a resposta através dos Teoremas de Bernoulli — a probabilidade vai convergir para a unidade.

Pacheco d'Amorim afirma, assim, que a probabilidade pedida no segundo enunciado (Problema 7.2) tende para 1 à medida que m aumenta. Consequentemente, a probabilidade pretendida no primeiro enunciado (Problema 7.1) que, como previamente referido, corresponde ao limite quando m tende para infinito do segundo problema, é igual a 1.

O autor afirma ainda que podemos demonstrar de forma análoga que também é igual a 1 a probabilidade de que sejam satisfeitas todas as outras leis análogas às de Bernoulli. $\diamond$

Pacheco d'Amorim conclui, da resolução destes problemas, que a probabilidade de que o número, tirado à sorte no intervalo $(0, 1)$, seja racional é igual a zero, pois um número racional dá sempre lugar a uma dízima periódica.

Como tal, nestas situações, poderão ocorrer duas possibilidades distintas. Na primeira situação encontram-se os números que não possuem todos os dígitos em proporções iguais na sua dízima periódica e, conseqüentemente, não verificam a Lei de Bernoulli. Na segunda possibilidade teremos os números racionais que contenham todos os algarismos de 0 a 9 em proporções iguais e, deste modo, a distribuição dos algarismos satisfaz a Lei de Bernoulli. Contudo, neste último caso, a sequência de dígitos não satisfaz as leis análogas às de Bernoulli, dado que o afastamento absoluto α passará periodicamente pelos mesmos valores e, portanto, ficará inferior a certo máximo. Por esta razão contraria o Teorema 7.10, de que a ordem do número de experiências relativas ao afastamento absoluto α será de segunda, uma vez que, nesta situações, teremos obrigatoriamente

$$\frac{\alpha^2}{m} \xrightarrow{m \rightarrow \infty} 0 \quad (7.139)$$

como consequência de α , nestas condições, ser limitado e, por conseguinte, podemos concluir

$$\mathbb{P} \left(\frac{\alpha^2}{m} < \varepsilon \right) = 1, \forall \varepsilon > 0. \quad (7.140)$$

Desta forma os números racionais ou não satisfazem a Lei de Bernoulli ou não satisfazem as leis análogas, razão pela qual Pacheco d'Amorim considera que é nula a probabilidade de, ao ser escolhido à sorte um número no intervalo $(0, 1)$, ele ser racional.

7.4.1 Os números Normais de Borel

Pacheco d'Amorim, ao longo desta pesquisa, onde conclui que a probabilidade de um número escolhido aleatoriamente ser racional é igual a zero, não faz qualquer referência a outros trabalhos. Recordemos que esta análise foi efectuada num artigo de Borel (1909a), que se tornou célebre na História da

Teoria da Probabilidade por inserir a σ -aditividade na Teoria da Probabilidade, por apresentar a primeira versão do Lema de Borel-Cantelli, com o qual demonstra a Lei Forte dos Grandes Números para provas de Bernoulli independentes com probabilidade de sucesso igual a $p = \frac{1}{2}$ e por, desta lei, concluir que a probabilidade de, ao ser escolhido um número à sorte no intervalo $(0, 1)$, sair um número normal na base decimal é igual à unidade e, conseqüentemente, a probabilidade de sair um número racional é nula.

Na primeira parte do seu artigo Borel demonstra a primeira versão do Lema de Borel-Cantelli de que seguidamente apresentamos uma versão mais geral, demonstrada por Francesco Cantelli (1875–1966) em 1917⁽²⁹⁾.

Lema 7.2 (Lema de Borel-Cantelli). *Seja $\{\mathbf{A}_i\}$, com $i = 1, \dots, m, \dots$, uma sequência de acontecimentos associados a um espaço de probabilidade $(\Omega, \mathcal{A}, \mathbb{P})$ então*

$$\sum_{i=1}^{+\infty} \mathbb{P}(\mathbf{A}_i) < +\infty \Rightarrow \mathbb{P}\left(\limsup_{m \rightarrow +\infty} \mathbf{A}_m\right) = 0. \quad (7.141)$$

Se os acontecimentos $\{\mathbf{A}_i\}$, com $i = 1, \dots, m, \dots$, forem independentes dois a dois e se

$$\sum_{i=1}^{+\infty} \mathbb{P}(\mathbf{A}_i) \rightarrow +\infty \quad (7.142)$$

então

$$\mathbb{P}\left(\limsup_{m \rightarrow +\infty} \mathbf{A}_m\right) = 1. \quad (7.143)$$

Nas fórmulas anteriores temos

$$\limsup_{m \rightarrow +\infty} \mathbf{A}_m = \bigcap_{i=1}^{\infty} \bigcup_{k=m}^{\infty} \mathbf{A}_k = \{\mathbf{A}_m \text{ i.o.}\} \quad (7.144)$$

⁽²⁹⁾ Regazzini (2005) apresenta um estudo bibliográfico sobre a vida e a obra de Cantelli e a sua importância no aperfeiçoamento da Lei dos Grandes Números. Margherita Benzi, Michele Benzi e Eugene Seneta (2007) apresentam também uma síntese da obra deste autor.

onde $\mathbf{A}_m$ i.o. significa que os acontecimentos $\mathbf{A}_i$ ocorrem infinitas vezes (*infinitely often*).

A demonstração deste resultado pode ser consultada, por exemplo, em Feller (1967). Depois de deduzir o Lema 7.2, Borel apresenta o princípio de Borel que introduzimos na secção 6.2.1, onde ilustramos a sua aplicabilidade com alguns exemplos simples com o objectivo de comparar este princípio com a aplicação do conceito ponto imagem criado por Pacheco d'Amorim. Recordemos então que Borel cria uma bijecção entre o intervalo $(0, 1]$ e a colecção $\mathcal{B}'$ de sequências infinitas (não degeneradas) de provas de Bernoulli independentes com probabilidade $\frac{1}{2}$ no caso de utilizarmos a representação binária (cf. definição 6.2 presente na página 336). Sejam X_i , com $i = 1, \dots, m, \dots$, as provas de Bernoulli com probabilidade de sucesso igual a $\frac{1}{2}$; então, pelo Teorema Limite Central de Moivre-Laplace, concluímos que em $2m$ provas teremos

$$\mathbb{P} \left(m - \lambda_m \sqrt{m} < \sum_{i=1}^{2m} X_i < m + \lambda_m \sqrt{m} \right) = 2\Phi \left(\sqrt{2}\lambda_m \right) - 1, \quad (7.145)$$

onde Φ representa a função de distribuição de uma distribuição gaussiana standard. Borel considera que λ_m é uma função crescente com m , de tal forma que

$$\lim_{m \rightarrow +\infty} \frac{\lambda_m}{\sqrt{m}} = 0 \quad (7.146)$$

e propõe a utilização de

$$\lambda_m = \log(m). \quad (7.147)$$

Substituindo λ_m na expressão (7.145) obtemos

$$\begin{aligned} p_m &= \mathbb{P} \left(m - \log(m) \sqrt{m} < \sum_{i=1}^{2m} X_i < m + \log(m) \sqrt{m} \right) = \\ &= 2\Phi \left(\sqrt{2} \log(m) \right) - 1. \end{aligned} \quad (7.148)$$

Consideremos então que estamos perante uma sequência de provas de Bernoulli com probabilidade de sucesso p_m e, conseqüentemente, com probabilidade de insucesso $q_m = 1 - p_m$. Uma vez que

$$\sum_{i=1}^{+\infty} q_i < +\infty, \quad (7.149)$$

pelo Lema de Borel-Cantelli podemos concluir que existe uma probabilidade nula de que se verifiquem infinitos insucessos ou, dito de outra forma, existe uma probabilidade igual a um de que, a partir de determinada ordem, todas as provas sejam sucesso, isto é, que ocorra

$$-\log(m) \sqrt{m} < \sum_{i=1}^{2m} X_i - m < \log(m) \sqrt{m}. \quad (7.150)$$

Assim, existe uma probabilidade igual a um de que o rácio entre o número de zeros e de uns na representação diádica se situe no intervalo

$$\left(\frac{m - \log(m) \sqrt{m}}{m + \log(m) \sqrt{m}}, \frac{m + \log(m) \sqrt{m}}{m - \log(m) \sqrt{m}} \right) \quad (7.151)$$

que converge para o ponto 1. Deste resultado Borel conclui que a Probabilidade de que o quociente entre o número de zeros e o número de uns tenda para um, quando aumentamos o número de provas de Bernoulli, é igual à unidade.

Borel refere que, de igual forma, se poderia ter demonstrado este resultado para outra base, por exemplo a decimal. Desta forma conclui-se que as frequências relativas do dígito d (com $d = 1, \dots, 9$), que representaremos por $f_i(d)$, convergem para 0.1 com probabilidade igual a um. Para Borel um número é simplesmente normal (*simplement normal*) na base $q \in \mathbb{N}$, se na sua representação na base q cada dígito d , com $d = 0, \dots, q-1$, ocorrer com frequência relativa $f_i(d) = \frac{1}{q}$. Então, podemos concluir que, ao escolhermos um número à sorte no intervalo $(0, 1]$, ele será simplesmente normal na base 10 com probabilidade igual a um. Um número será inteiramente normal

(*intièrement normal*), ou apenas normal, na base $q \in \mathbb{N}$, se na sua representação na base q^p , para qualquer $p \in \mathbb{N}$, cada dígito d , com $d = 0, \dots, q^p - 1$, ocorrer com frequência relativa $f_i(d) = \frac{1}{q^p}$. Notemos que isto corresponde a analisar as frequências dos dígitos na base q por blocos, isto é, um número é inteiramente normal na base q se, na sua representação na base q , cada bloco de dígitos de qualquer dimensão ocorrer com igual frequência. Borel conclui que a probabilidade de um número escolhido aleatoriamente no intervalo $(0, 1]$ ser inteiramente normal é igual à unidade. Finalmente Borel define um número absolutamente normal se for normal em qualquer base $q \in \mathbb{N}$, concluindo que a probabilidade de um número, escolhido à sorte no intervalo $(0, 1]$, ser absolutamente normal é igual à unidade.

Existem diversas semelhanças entre a forma como Borel demonstra que é nula a probabilidade de um número escolhido aleatoriamente no intervalo $(0, 1)$ ser racional e a forma como Pacheco d'Amorim o faz. De entre elas salienta-se a ideia genial de construir uma correspondência entre o número escolhido à sorte no intervalo $(0, 1)$ e uma sequência infinita de provas de Bernoulli independentes com probabilidade $\frac{1}{q}$, onde q é a base em que representamos o número, tendo ambos os autores escolhido a decimal. É esta associação que lhes permite resolver o problema: a Borel permitiu-lhe recorrer ao Teorema Limite Central de Moivre-Laplace e concluir o resultado expresso em (7.145); a Pacheco d'Amorim permite-lhe recorrer aos teoremas de Bernoulli e análogos, que demonstrou no contexto de provas de Bernoulli independentes com probabilidade constante.

Contudo, após este paralelismo, os dois autores seguem por caminhos distintos. Borel recorre ao Lema de Borel-Cantelli, que apresenta a primeira versão no primeiro capítulo do mesmo artigo, e deduz que as frequências relativas $f_m(d)$ de qualquer dígito d , com $d = 0, \dots, 9$, convergem para 0.1

com probabilidade igual a um, i.e.

$$\mathbb{P} \left(\lim_{m \rightarrow +\infty} f_m(d) = \frac{1}{10} \right) = 1, \quad (7.152)$$

considerando que, de igual forma, se demonstra o mesmo resultado para as bases 10^p , para $p \in \mathbb{N}$, justificando desta forma que as dízimas periódicas de qualquer dimensão, que caracterizam os números racionais, têm probabilidade nula. Por exemplo, o número com dízima periódica 0123456789, apesar de cada dígito ter frequências relativas iguais a 0.1 na base 10, na base $10^2 = 100$ o dígito 00 terá frequências relativas iguais a zero, pois nunca ocorre. Assim, será um número simplesmente normal, mas não (inteiramente) normal no sentido de Borel e, por esta razão, terá probabilidade nula de ocorrer.

Pacheco d'Amorim, sem o recurso ao Lema de Borel-Cantelli, utilizando unicamente a Lei Fraca dos Grandes Números, conclui que a probabilidade de as frequências relativas $f_m(d)$ de qualquer dígito d , com $d = 0, \dots, 9$, se afastarem de 0.1 mais do que um valor ε , por mais pequeno que esse valor seja, converge para um, i.e.

$$\lim_{m \rightarrow +\infty} \mathbb{P} \left(\left| f_m(d) - \frac{1}{10} \right| < \varepsilon \right) = 1, \quad \forall \varepsilon > 0. \quad (7.153)$$

Desta forma, a conclusão de Pacheco d'Amorim, apesar de posterior, é mais fraca que a de Borel, pois o tipo de convergência deduzido por Borel implica a convergência demonstrada por Pacheco d'Amorim, não se verificando o contrário. Por esta razão Khintchine designou, em 1928, este resultado de Borel, tal como as suas generalizações posteriores, por Lei Forte dos Grandes Números. Outra diferença na análise destes dois autores é a justificação de que os números racionais, que possuem dízima periódica onde os dígitos d tenham todos igual proporção, convergindo então as suas frequências relativas para 0.1, têm também probabilidade nula. Borel, como referimos, justifica recorrendo à representação nas bases 10^p , Pacheco d'Amorim baseia-se nos

resultados que demonstrou relativos à ordem do número de experiências relativas ao afastamento absoluto α .

Borel, no seu artigo, afirma que os seus resultados poderiam ser demonstrados através da Teoria da Medida dos conjuntos (como Faber, em 1910, e posteriormente Hausdorff, em 1914, o fizeram). Deste modo, para salientar e ilustrar as potencialidades do princípio de Borel, que introduzimos na secção 6.2.1, vamos resolver alguns problemas um pouco mais delicados que os que nessa secção apresentámos.

Seja X_i o i -ésimo dígito do número ω escolhido à sorte em $(0, 1]$, que corresponde a uma prova de Bernoulli com probabilidade $\frac{1}{2}$. Para a resolução de alguns problemas é usual recorrer-se à função de Rademacher⁽³⁰⁾, que é definida por

$$R_n = 2X_i(\omega) - 1 = \begin{cases} +1 & \text{se } X_i = 1 \\ -1 & \text{se } X_i = 0 \end{cases}. \quad (7.154)$$

Se as variáveis X_i correspondem ao resultado do i -ésimo lançamento de uma moeda, podemos interpretar a função de Rademacher R_i como o resultado de uma aposta, onde o apostador ganha um euro ($R_i = +1$) se sair cara ($X_i = 1$) e perde um euro ($R_i = -1$) se sair coroa ($X_i = 0$) no i -ésimo lançamento da moeda. A soma $G_m = \sum_{i=1}^m R_i$ dos primeiros m termos desta sucessão pode ser interpretada como o ganho (ou perda no caso de assumir um valor negativo) de um jogador após m lançamentos da moeda. Analisemos algumas características desta função. Notemos que, para qualquer $m \in \mathbb{N}$, esta função verifica

$$\mathbb{P}(\omega : R_m = +1) = \mathbb{P}(\omega : X_m = 1) = \frac{1}{2} \quad (7.155)$$

e

$$\mathbb{P}(\omega : R_m = -1) = \mathbb{P}(\omega : X_m = 0) = \frac{1}{2}, \quad (7.156)$$

⁽³⁰⁾ Criadas por Hans Rademacher (1892–1969) na década de 1920.

logo, teremos ainda

$$\int_0^1 R_m \, d\omega = 0, \quad \forall m \in \mathbb{N}, \quad (7.157)$$

de onde se conclui que o ganho de cada jogador num jogo tem valor esperado nulo e, conseqüentemente,

$$\int_0^1 \sum_{i=1}^m R_i \, d\omega = \int_0^1 G_m \, d\omega = 0, \quad (7.158)$$

isto é, o jogador tem um ganho esperado nulo em m jogos (estamos perante um jogo justo). Podemos ainda concluir que, para quaisquer $i \neq j$ tais que $i, j \leq m$, teremos

$$\int_0^1 R_i R_j \, d\omega = 0 \Rightarrow \int_0^1 \sum_{i=1}^m \sum_{\substack{j=1 \\ i \neq j}}^m R_i R_j \, d\omega = 0, \quad (7.159)$$

pois teremos intervalos com igual amplitude onde se verifica $R_i R_j = +1$ e $R_i R_j = -1$. Uma vez que, para qualquer m , teremos $R_m^2 = 1$, então

$$\int_0^1 R_m^2 \, d\omega = 1, \quad \forall m \in \mathbb{N}, \quad (7.160)$$

e, conseqüentemente, verifica-se

$$\int_0^1 \sum_{i=1}^m R_i^2 \, d\omega = m. \quad (7.161)$$

Representando a soma de m variáveis aleatórias X_i através de S_m , i.e. $S_m = \sum_{i=1}^m X_i$, a Lei Fraca dos Grandes Números estabelece

$$\lim_{m \rightarrow +\infty} \mathbb{P} \left(\omega : \left| \frac{S_m}{m} - \frac{1}{2} \right| > \varepsilon \right) = 0, \quad (7.162)$$

pelo que, em termos da função de Rademacher, uma vez que $G_m = 2S_m - m$, teremos

$$\begin{aligned} \mathbb{P} \left(\omega : \left| \frac{S_m}{m} - \frac{1}{2} \right| > \varepsilon \right) &= \mathbb{P} \left(\omega : \left| \frac{G_m}{2m} \right| > \varepsilon \right) = \\ &= \mathbb{P} (\omega : |G_m| > 2m\varepsilon). \end{aligned} \quad (7.163)$$

Notemos que

$$\begin{aligned}
 \int_0^1 G_m^2 \, d\omega &= \int_{G_m^2 > (2m\varepsilon)^2} G_m^2 \, d\omega + \int_{G_m^2 \leq (2m\varepsilon)^2} G_m^2 \, d\omega \geq \\
 &\geq \int_{G_m^2 > (2m\varepsilon)^2} G_m^2 \, d\omega \geq \int_{G_m^2 > (2m\varepsilon)^2} (2m\varepsilon)^2 \, d\omega = \\
 &= (2m\varepsilon)^2 \mathbb{P} \left[\omega : G_m^2 > (2m\varepsilon)^2 \right] = \\
 &= (2m\varepsilon)^2 \mathbb{P} [\omega : |G_m| > 2m\varepsilon], \tag{7.164}
 \end{aligned}$$

que nos leva a concluir que

$$\mathbb{P} [\omega : |G_m| > 2m\varepsilon] \leq \frac{1}{4m^2\varepsilon^2} \int_0^1 G_m^2 \, d\omega. \tag{7.165}$$

Podemos ainda desenvolver o integral

$$\begin{aligned}
 \int_0^1 G_m^2 \, d\omega &= \int_0^1 \left(\sum_{i=1}^m R_i \right)^2 \, d\omega = \\
 &= \int_0^1 \sum_{i=1}^m R_i^2 \, d\omega + \int_0^1 \sum_{i=1}^m \sum_{\substack{j=1 \\ i \neq j}}^m R_i R_j \, d\omega = m \tag{7.166}
 \end{aligned}$$

pelos resultados expressos em (7.159) e (7.161). Desta forma a expressão (7.165) pode ser simplificada para

$$\mathbb{P} [\omega : |G_m| > 2m\varepsilon] \leq \frac{1}{4m\varepsilon^2} \tag{7.167}$$

que, em termos de probabilidade, corresponde à aplicação da desigualdade de Bienaymé-Chebycheff. Aplicando esta desigualdade ao resultado expresso em (7.162) com a função de Rademacher, conforme expressão (7.163), obtemos

$$\lim_{m \rightarrow +\infty} \mathbb{P} (\omega : |G_m| > 2m\varepsilon) \leq \lim_{m \rightarrow +\infty} \frac{1}{4m\varepsilon^2} = 0, \tag{7.168}$$

ficando, desta forma, demonstrada a Lei Fraca dos Grandes Números, para provas de Bernoulli independentes com probabilidade de sucesso igual a $\frac{1}{2}$, através da aplicação do princípio de Borel.

Para deduzirmos a Lei forte dos Grandes Número consideremos o conjunto $\mathbf{N}$ definido por

$$\mathbf{N} = \left\{ \omega : \lim_{m \rightarrow \infty} \frac{1}{m} \sum_{i=1}^m X_i = \frac{1}{2} \right\} \quad (7.169)$$

ou, de forma equivalente,

$$\mathbf{N} = \left\{ \omega : \lim_{m \rightarrow \infty} \frac{G_m}{m} = 0 \right\} \quad (7.170)$$

consistindo a Lei Forte dos Grandes Número em $\mathbb{P}(\mathbf{N}) = 1$. Os pontos ω que pertencem ao conjunto $\mathbf{N}$ são denominados números simplesmente normais na base binária. Desta forma, a Lei Forte dos Grandes Número estabelece que, quando escolhemos um número à sorte no intervalo $(0, 1]$, existe uma probabilidade nula de o número escolhido não ser simplesmente normal, isto é, de a frequência relativa de cada um dos seus dígitos ser diferente de $\frac{1}{2}$. A base da demonstração da Lei Forte dos Grandes Número é considerar a desigualdade

$$\mathbb{P} \left(\omega : \left| \frac{G_m}{m} \right| \geq 2\varepsilon \right) = \mathbb{P}(\omega : |G_m| \geq 2m\varepsilon) \leq \frac{1}{16m^4 \varepsilon^4} \int_0^1 G_m^4 \, d\omega, \quad (7.171)$$

pois

$$\begin{aligned} \int_0^1 G_m^4 \, d\omega &= \int_{G_m^4 > (2m\varepsilon)^4} G_m^4 \, d\omega + \int_{G_m^4 \leq (2m\varepsilon)^4} G_m^4 \, d\omega \geq \\ &\geq \int_{G_m^4 > (2m\varepsilon)^4} G_m^4 \, d\omega \geq \int_{G_m^4 > (2m\varepsilon)^4} (2m\varepsilon)^4 \, d\omega = \\ &= (2m\varepsilon)^4 \mathbb{P} \left[\omega : G_m^4 > (2m\varepsilon)^4 \right] = \\ &= (2m\varepsilon)^4 \mathbb{P}[\omega : |G_m| > 2m\varepsilon], \end{aligned}$$

onde

$$\int_0^1 G_m^4 \, d\omega = 3m^2 - 2m \leq 3m^2, \quad (7.172)$$

uma vez que $R_i^4 = 1$ e $R_i^2 R_j^2 = 1$ e, consequentemente, teremos

$$\int_0^1 \sum_{i=1}^m R_i^4 d\omega = m \quad (7.173)$$

e

$$\int_0^1 \sum_{i=1}^m \sum_{\substack{j=1 \\ j \neq i}}^m 3R_i^2 R_j^2 d\omega = 3m(m-1) = 3m^2 - 3m, \quad (7.174)$$

sendo os restantes termos de G_m^4 (que são da forma $R_i^3 R_j$, $R_i^2 R_j R_k$ e $R_i R_j R_k R_l$ para valores de i, j, k e l todos distintos e inferiores a m) nulos.

Associando os resultado (7.171) e (7.172) concluimos que

$$\mathbb{P}(\omega : |G_m| \geq 2m\varepsilon) \leq \frac{3m^2}{16m^4 \varepsilon^4} \leq \frac{1}{m^2 \varepsilon^4}. \quad (7.175)$$

Definição 7.3 (Conjunto de medida de Lebesgue nula). Um conjunto $\mathbf{A}$ tem medida de Lebesgue nula se, para cada valor $\delta > 0$, existe uma colecção finita ou infinita numerável de intervalos $\mathbf{I}_i$ tal que

$$\mathbf{A} \subset \bigcup_k \mathbf{I}_k \quad (7.176)$$

e

$$\sum_k \lambda(\mathbf{I}_k) < \delta. \quad (7.177)$$

Assim sendo, um conjunto de medida nula de Lebesgue corresponde a um intervalo que possui uma cobertura cuja soma das medidas dos intervalos que compõem essa cobertura é menor que qualquer valor arbitrariamente pequeno δ . Se definirmos

$$\mathbf{A}_m = \left\{ \omega : \left| \frac{1}{m} \sum_{i=1}^m X_i - \frac{1}{2} \right| \geq \varepsilon_m \right\} \quad (7.178)$$

ou, de forma equivalente,

$$\mathbf{A}_m = \{ \omega : |G_m| \geq 2m\varepsilon_m \}, \quad (7.179)$$

então por (7.175) concluímos que

$$\mathbb{P}(\mathbf{A}_m) \leq \frac{1}{m^2 \varepsilon_m^4}. \quad (7.180)$$

Notemos que os conjuntos $\mathbf{A}_m$ formam uma cobertura para o complementar do conjunto $\mathbf{N}$, que representaremos por $\overline{\mathbf{N}}$, isto é,

$$\overline{\mathbf{N}} \subset \bigcup_{k=1}^{\infty} \mathbf{A}_k, \quad (7.181)$$

que é equivalente a $\mathbf{N} \supset \bigcap_{k=1}^{\infty} \overline{\mathbf{A}_k}$. Notemos que se $\omega \in \bigcap_{k=1}^{\infty} \overline{\mathbf{A}_k}$, onde $\overline{\mathbf{A}_k} = \{\omega : |G_k| < 2k\varepsilon_k\}$, então para qualquer valor de m verifica-se $|G_m| < 2m\varepsilon_m$ e, uma vez que ε_m converge para zero, teremos $\frac{1}{m}|G_m|$ também a convergir para zero, que corresponde à definição do conjunto $\mathbf{N}$.

Procuremos então uma função ε_m que convirja para zero e garanta que a soma das probabilidades de $\mathbf{A}_m$ seja inferior a qualquer δ . Consideremos, por exemplo, $\varepsilon_m = c^{\frac{1}{4}} m^{-\frac{1}{8}}$ para uma constante qualquer $c \in \mathbb{R}^+$, pois

$$\begin{aligned} \sum_{k=1}^{\infty} \lambda(\mathbf{A}_k) &= \sum_{k=1}^{\infty} \mathbb{P}(\mathbf{A}_k) \leq \sum_{k=1}^{\infty} \frac{1}{m^2 \varepsilon_m^4} = \sum_{k=1}^{\infty} \frac{1}{m^2 \left(c^{\frac{1}{4}} m^{-\frac{1}{8}}\right)^4} = \\ &= \sum_{k=1}^{\infty} \frac{1}{c m^{\frac{3}{2}}}. \end{aligned} \quad (7.182)$$

Deste modo é possível escolher um valor para c de forma que a série anterior seja inferior a δ (um c tal que $c > \frac{1}{\delta} \sum_{k=1}^{\infty} m^{-\frac{3}{2}}$ uma vez que $\sum_{k=1}^{\infty} m^{-\frac{3}{2}} < \infty$). Desta forma demonstramos, segundo o princípio de atribuição de probabilidade de Borel e recorrendo à Teoria da Medida, que, quando escolhermos à sorte um número do intervalo $(0, 1]$, a probabilidade de o número escolhido ser simplesmente normal na base binária é igual à unidade, o que corresponde à Lei Forte dos Grandes Números para provas de Bernoulli independentes com probabilidade $p = \frac{1}{2}$.

Refira-se, para ilustrar a importância que os números normais assumiram no desenvolvimento da Teoria da Probabilidade, que na altura se considerava

que o facto de um número ser normal (no sentido de Borel) seria condição suficiente para garantir que a sequência de dígitos que representam esse número é aleatória. Esta ideia desmoronou-se em 1933, quando Chapernowne demonstrou que o número

$$0.1234567891011121314\cdots,$$

que resulta na concatenação dos números inteiros, é normal na base decimal. Copeland e Erdős, em 1946, demonstram que a concatenação dos números primos, i.e.,

$$0.23571113171923\cdots$$

também é um número normal na base decimal. Destes exemplos podemos facilmente concluir que o facto de um número ser normal na base q não significa que a sequência de dígitos que o representam nessa base constituam um sequência aleatória, isto é, o facto de as frequências relativas de cada resultado convergirem para a sua probabilidade não é condição suficiente para garantir a aleatoriedade das provas.

Para terminar esta secção vamos apresentar uma síntese de alguns desenvolvimentos nos teoremas limites que decorreram deste artigo de Borel. Como analisámos previamente, Borel (1909a) desenvolveu a Lei Forte dos Grandes Números considerando uma sucessão de variáveis aleatórias independentes $\{X_i\}, i = 1, 2, \cdots$ com distribuição de Bernoulli com parâmetro $\frac{1}{2}$, concluindo que

$$\mathbb{P}\left(\lim_{m \rightarrow +\infty} \frac{1}{m} \sum_{i=1}^m X_i = \frac{1}{2}\right) = 1, \quad (7.183)$$

que podemos representar por

$$\frac{1}{m} \sum_{i=1}^m X_i \xrightarrow{q.c.} \frac{1}{2}, \quad (7.184)$$

onde $q.c.$ significa quase certamente, isto é, excepto num conjunto de medida nula.

Cantelli em 1917 demonstra a primeira versão da Lei dos Grandes Números de âmbito geral. Em 1930 surge a primeira Lei Forte dos Grandes Números de Kolmogoroff, onde o resultado de Borel é generalizado para uma qualquer sucessão de variáveis aleatórias independentes $\{X_i\}, i = 1, 2, \dots$ com valor esperado e variância finita, $\mathbb{E}(X_i) = \mu_i < \infty$ e $\text{Var}(X_i) = \sigma_i^2 < \infty$, de forma que se verifique

$$\sum_{k=1}^{+\infty} \frac{\text{Var}(X_k)}{k^2} < \infty, \quad (7.185)$$

concluindo-se, nestas condições, que

$$\mathbb{P} \left(\lim_{m \rightarrow +\infty} \left| \frac{1}{m} \sum_{i=1}^m X_i - \frac{1}{m} \sum_{i=1}^m \mu_i \right| = 0 \right) = 1. \quad (7.186)$$

De acordo com a segunda Lei Forte dos Grandes Números de Kolmogoroff, se considerarmos uma sucessão de variáveis aleatórias independentes e identicamente distribuídas $\{X_i\}, i = 1, 2, \dots$ com valor esperado μ finito, então (são condições necessárias e suficientes)

$$\mathbb{P} \left(\lim_{m \rightarrow +\infty} \left| \frac{1}{m} \sum_{i=1}^m X_i - \mu \right| = 0 \right) = 1. \quad (7.187)$$

A Lei Forte dos Grandes Números está na origem de um outro resultado importante na Teoria da Probabilidade, a Lei do Logaritmo Iterado, que fornece o intervalo de flutuação da sequência estocástica $S_m = \frac{1}{m} \sum_{i=1}^m X_i$. Assim, considerando a existência de todos os momentos das variáveis aleatórias $\{X_i\}, i = 1, 2, \dots$, Felix Hausdorff (1868–1942) em 1914 demonstrou que

$$\mathbb{P} \left(\lim_{m \rightarrow +\infty} \frac{\left| \frac{1}{m} \sum_{i=1}^m X_i - \mu \right|}{m^{\frac{1}{2} + \epsilon}} \leq M \right) = 1, \quad (7.188)$$

para qualquer $\epsilon > 0$ e $M \in \mathbb{R}^+$. No mesmo ano, que também corresponde ao ano em que Pacheco d'Amorim defende a sua tese de doutoramento, Godfrey

Hardy (1877–1947) e John Littlewood (1885–1977) deduzem

$$\mathbb{P} \left(\lim_{m \rightarrow +\infty} \frac{\left| \frac{1}{m} \sum_{i=1}^m X_i - \mu \right|}{\sqrt{m \ln(m)}} \leq M \right) = 1. \quad (7.189)$$

Em 1924 Aleksandre Khintchine (1894–1959) deduz a Lei do Logaritmo Iterado considerando uma sequência de variáveis aleatórias independentes e identicamente distribuídas $\{X_i\}, i = 1, 2, \dots$, com valor esperado e variância finita ($\mathbb{E}(X_i) = \mu < \infty$ e $\mathbb{V}ar(X_i) = \sigma^2 < \infty, \forall i$), concluindo que

$$\mathbb{P} \left(\limsup_{m \rightarrow +\infty} \frac{\left| \frac{1}{m} \sum_{i=1}^m X_i - \mu \right|}{\sqrt{m \ln(\ln(m))}} = \sqrt{2\sigma^2} \right) = 1. \quad (7.190)$$

Intuitivamente a Lei do Logaritmo Iterado estabelece que a sequência estocástica $S_m = \sum_{i=1}^m X_i$ flutua dentro dos limites $\mu \pm \sqrt{2\sigma^2 m \ln(\ln(m))}$.

Não é certo que Pacheco d'Amorim tenha lido o artigo de Borel nesta secção comentado. Contudo, em relação à Lei Forte dos Grandes Números, Pacheco d'Amorim não apresenta qualquer análise ou referência. Uma vez mais, no que respeita a teoremas limites, Pacheco d'Amorim não analisa os resultados mais actuais nem apresenta qualquer tentativa de os melhorar. No que respeita à Lei dos Grandes Números, enquanto a análise de Pacheco d'Amorim se resume à convergência fraca restrita às provas de Bernoulli independentes com probabilidade constante, outros autores procuravam condições para garantir um tipo de convergência mais forte num âmbito bastante mais geral. Refira-se, contudo, que Borel no seu livro *Éléments de la théorie des probabilités* (mesmo na segunda edição, de 1910, aquela que consultámos e que também é a referida por Pacheco d'Amorim na sua tese de doutoramento), apesar de ser posterior ao seu artigo onde demonstra um caso particular da Lei Forte dos Grandes Números, não inclui nenhuma análise deste tipo de convergência.

7.5 Lei dos desvios

Pacheco d'Amorim, após ter demonstrado, dum modo rigoroso, o Terceiro Teorema de Jacob Bernoulli e outros análogos, relativos à ordem de grandeza dos afastamentos ou desvios, deduz uma relação aproximada entre os desvios e as suas probabilidades, que é apresentada no teorema a seguir apresentado, que corresponde ao Teorema Limite Central restrito à situação binomial, habitualmente apelidado por Teorema de Moivre-Laplace.

Teorema 7.11. *Seja $\mathbb{T}_n$ a probabilidade da combinação de probabilidade máxima (combinação modal ou normal) e $\mathbb{P}(k)$ a probabilidade de que o afastamento, em valor absoluto, seja inferior a k . Então haverá uma probabilidade sempre crescente com m de que seja*

$$\mathbb{P}(k) = \frac{2}{\sqrt{\pi}} \int_0^{\lambda_1} e^{-\lambda^2} d\lambda = \theta(\lambda_1) \quad (7.191)$$

onde

$$\lambda_1 = \frac{k}{\sqrt{2mpq}}, \quad (7.192)$$

sendo que $\theta(\cdot)$ denomina-se **lei dos desvios**, **lei dos afastamentos** ou **lei de Gauss** e λ_1 é o **afastamento relativo** (para se distinguir do **afastamento absoluto** k). Ao número $\sqrt{2mpq}$ denomina-se por **unidade de afastamento**⁽³¹⁾.

Demonstração. A probabilidade $\mathbb{P}(k)$ de que o afastamento em relação à combinação modal, em valor absoluto, seja inferior a k pode ser decomposto

⁽³¹⁾ Notemos que $\sqrt{2mpq} = \sqrt{2} \sigma_s$, onde σ_s^2 corresponde à variância da soma das m provas de Bernoulli independentes com probabilidade p , i.e. $S = \sum_{i=1}^m X_i$. Segundo David e Edwards (2001), o termo variância só surge em 1918 no trabalho de Ronald Fisher (1890–1962), apesar de identificarem o aparecimento do conceito desvio padrão com Karl Pearson (1857–1936) em 1894. Bachelier (1912), por exemplo, denomina $2mpq = 2 \sigma_s^2$ por *fonction d'instabilité*.

(tal como foi efectuado no Lema de Vallée-Poussin) em

$$\mathbb{P}(k) = \mathbb{T}_{n-k} + \cdots + \mathbb{T}_n + \cdots + \mathbb{T}_{n+k} = \sum_{i=-k}^k \mathbb{T}_{n+i}, \quad (7.193)$$

onde cada probabilidade $\mathbb{T}_{n+i}$ é determinada, pela função de probabilidade da distribuição binomial, através de

$$\mathbb{T}_{n+i} = \frac{m!}{(n+i)! (m-n-i)!} p^{n+i} q^{m-n-i}, \quad (7.194)$$

onde n representa o número de bolas brancas da combinação normal (ou modal) e, por isso, como deduzimos previamente, n terá a forma

$$n = p(m+1) - r, \quad \text{com } 0 \leq r \leq 1. \quad (7.195)$$

Supondo o número de provas de Bernoulli elevado (m a tender para infinito) podemos substituir $p(m+1) - r$ por mp (desprezando os termos finitos em conjunção com termos infinitos) e utilizar a fórmula de Stirling ($m! \simeq m^m e^{-m} \sqrt{2\pi m}$) para obter valores aproximados para os factoriais. Desta forma $\mathbb{T}_{n+i}$ será aproximadamente igual a

$$\begin{aligned} \mathbb{T}_{n+i} &= \frac{m!}{(n+i)! (m-n-i)!} p^{n+i} q^{m-n-i} \approx \\ &\approx \frac{m^m e^{-m} \sqrt{2\pi m} p^{mp+i} q^{mq-i}}{(mp+i)^{mp+i} e^{-mp-i} \sqrt{2\pi (mp+i)} (mq-i)^{mq-i} e^{-mq+i} \sqrt{2\pi (mq-i)}} = \\ &= \frac{\sqrt{m} p^{mp+i} q^{mq-i}}{\left(p + \frac{i}{m}\right)^{mp+i} \left(q - \frac{i}{m}\right)^{mq-i} \sqrt{2\pi (mp+i)(mq-i)}} = \\ &= \frac{p^{mp+i} q^{mq-i}}{\left(p + \frac{i}{m}\right)^{mp+i+\frac{1}{2}} \left(q - \frac{i}{m}\right)^{mq-i+\frac{1}{2}} \sqrt{2\pi m}} = \\ &= \frac{1}{\sqrt{2\pi mpq}} \cdot \frac{1}{\left(1 + \frac{i}{mp}\right)^{mp+i+\frac{1}{2}} \left(1 - \frac{i}{mq}\right)^{mq-i+\frac{1}{2}}} = \\ &= \frac{1}{\sqrt{2\pi mpq}} \cdot \mathbf{H} \end{aligned} \quad (7.196)$$

onde $\mathbf{H}$ assume a forma

$$\mathbf{H} = \left(1 + \frac{i}{mp}\right)^{-mp-i-\frac{1}{2}} \left(1 - \frac{i}{mq}\right)^{-mq+i-\frac{1}{2}} \quad (7.197)$$

e, conseqüentemente,

$$\begin{aligned} \log \mathbf{H} = & - \left(mp + i + \frac{1}{2}\right) \log \left(1 + \frac{i}{mp}\right) - \\ & - \left(mq - i + \frac{1}{2}\right) \log \left(1 - \frac{i}{mq}\right). \end{aligned} \quad (7.198)$$

Contudo, do Terceiro Teorema de Jacob Bernoulli, podemos concluir que a probabilidade de que $\left|\frac{i}{m}\right| > \varepsilon$ tende para zero quando m aumenta. Haverá, por conseguinte, uma probabilidade sempre crescente de que

$$\begin{aligned} \log \mathbf{H} &= - \left(mp + i + \frac{1}{2}\right) \left[\frac{i}{mp} - \frac{i^2}{2m^2p^2} + \frac{i^3}{3m^3p^3} - \dots \right] - \\ & - \left(mq - i + \frac{1}{2}\right) \left[-\frac{i}{mq} - \frac{i^2}{2m^2q^2} - \frac{i^3}{3m^3q^3} - \dots \right] = \\ &= -\frac{i^2}{m} \left[\frac{1}{p} + \frac{1}{q} \right] - \frac{1}{2} \cdot \frac{i}{m} \left[\frac{1}{p} - \frac{1}{q} \right] + \\ & + \frac{i^2}{2m} \left[\frac{1}{p} + \frac{1}{q} \right] + \frac{i^3}{2m^2} \left[\frac{1}{p^2} - \frac{1}{q^2} \right] + \\ & + \frac{1}{2} \frac{i^2}{2m^2} \left[\frac{1}{p^2} + \frac{1}{q^2} \right] + \dots \end{aligned} \quad (7.199)$$

que pode ser simplificado para

$$\log \mathbf{H} = -\frac{i^2}{2mpq} - \frac{i}{2mpq} + \frac{i^3}{2m^2p^2q^2} + \dots \quad (7.200)$$

Mas, pelo Teorema 7.8 (página 414), que garante uma probabilidade nula para que a ordem do número de experiências em relação ao afastamento absoluto seja inferior à segunda, podemos concluir que os termos

$$\frac{i}{m}, \frac{i^3}{m^2}, \dots, \frac{i^n}{m^n}, \dots, \frac{i^{n+1}}{m^n}, \quad (7.201)$$

têm uma probabilidade de se manterem superiores a uma quantidade qualquer $\varepsilon > 0$, por menor que seja o valor de ε , que tende para zero quando o

número de provas m aumenta indefinidamente⁽³²⁾. Assim haverá uma probabilidade sempre crescente com m de que $\mathbf{H}$ seja tal que

$$\log \mathbf{H} = -\frac{i^2}{2mpq} \quad (7.202)$$

e, portanto,

$$\mathbf{H} = e^{-\frac{i^2}{2mpq}}, \quad (7.203)$$

que, substituindo na expressão aproximada de $\mathbb{T}_{n+i}$ (fórmula (7.196) presente na página 445), obtemos

$$\mathbb{T}_{n+i} = \frac{1}{\sqrt{2\pi mpq}} e^{-\frac{i^2}{2mpq}}. \quad (7.204)$$

Utilizando este resultado na expressão de $\mathbb{P}(k)$ (fórmula (7.193) apresentada na página 445) obtemos

$$\begin{aligned} \mathbb{P}(k) &= \sum_{i=-k}^k \mathbb{T}_{n+i} = \frac{1}{\sqrt{2\pi mpq}} \sum_{i=-k}^k e^{-\frac{i^2}{2mpq}} = \\ &= \frac{2}{\sqrt{2\pi mpq}} \sum_{i=0}^k e^{-\frac{i^2}{2mpq}}; \end{aligned} \quad (7.205)$$

que, substituindo o somatório por um integral tomado entre os mesmos limites, obtemos

$$\mathbb{P}(k) = \frac{2}{\sqrt{2\pi mpq}} \int_0^k e^{-\frac{x^2}{2mpq}} dx. \quad (7.206)$$

⁽³²⁾ As demonstrações patentes nas obras contemporâneas a Pacheco d'Amorim utilizam um raciocínio análogo ao apresentado por este autor. Contudo, uma vez que os restantes autores não deduziram resultados semelhantes aos de Pacheco d'Amorim referentes à ordem dos números de experiências em relação aos afastamentos (comentados na secção 7.3), a justificação de que estes termos tendem para zero por estes apresentada é menos elaborada (consultar, por exemplo, Poincaré (1896, p. 91), Bertrand (1888, p. 74) ou Borel (1909, p. 62)).

Finalmente, aplicando a substituição de variável $x = \lambda\sqrt{2mpq}$ concluimos que

$$\mathbb{P}(k) = \frac{2}{\sqrt{2\pi mpq}} \int_0^{\lambda_1} e^{-\lambda^2} d\lambda \cdot \sqrt{2mpq} = \frac{2}{\sqrt{\pi}} \int_0^{\lambda_1} e^{-\lambda^2} d\lambda, \quad (7.207)$$

onde

$$\lambda_1 = \frac{k}{\sqrt{2mpq}}. \quad (7.208)$$

Deduzimos, então, a Lei dos Desvios ou Lei de Gauss

$$\theta(\lambda_1) = \frac{2}{\sqrt{\pi}} \int_0^{\lambda_1} e^{-\lambda^2} d\lambda. \quad (7.209)$$

□

Pacheco d'Amorim refere que a Lei dos Desvios é apenas uma aproximação e uma lei provável. Salienta, no entanto, que a sua probabilidade tende muito rapidamente para a unidade quando m aumenta, e os erros cometidos na sua dedução tendem rapidamente para zero. O autor destaca ainda a sua importância nas aplicações e os bons resultados que dela se usufruem quando aplicada, referindo que em muitas situações “*o resultado obtido é igual ao resultado verdadeiro*”. Refere ainda que a probabilidade de que a variável λ esteja compreendida entre 0 e ∞ será igual à unidade, i.e.

$$\mathbb{P} = \frac{2}{\sqrt{\pi}} \int_0^{\infty} e^{-\lambda^2} d\lambda = \frac{2}{\sqrt{\pi}} \frac{\sqrt{\pi}}{2} = 1. \quad (7.210)$$

Além das gralhas presentes nesta parte da obra de Pacheco d'Amorim, escrevendo por vezes $\frac{1}{\sqrt{\pi}}$ onde deveria escrever $\frac{2}{\sqrt{\pi}}$, o autor não justifica de onde vem o valor deste integral, isto é, por que razão $\int_0^{\infty} e^{-\lambda^2} d\lambda = \frac{\sqrt{\pi}}{2}$. Por exemplo Borel (1909, p. 52–53), em nota de rodapé, apresenta a demonstração *la plus simple* deste resultado, considerando a função J definida por

$$J = \frac{1}{\sqrt{\pi}} \int_{-\infty}^{+\infty} e^{-x^2} dx, \quad (7.211)$$

logo, o seu quadrado será

$$J^2 = \frac{1}{\pi} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} e^{-x^2-y^2} dx dy, \quad (7.212)$$

que, mudando para coordenadas polares, se obtém

$$J^2 = \frac{1}{\pi} \int_0^{2\pi} \int_0^{+\infty} e^{-\rho^2} \rho d\rho d\theta = \frac{1}{\pi} \int_0^{2\pi} \frac{1}{2} d\theta = 1 \quad (7.213)$$

e, conseqüentemente, uma vez que é positivo, teremos $J = 1$.

Pacheco d'Amorim apresenta depois tabelas dos valores da função $\theta(\lambda)$ para valores de λ a variarem em progressão aritmética de razão 0.01 de 0 até 4.00 e, depois, com razão 0.1 até ao valor 4.80. Até ao valor 3.45 os valores apresentados têm sete casas decimais e a partir deste ponto têm onze casas decimais. Na tradução exposta na PARTE II deste trabalho não apresentamos os valores patentes na tese de doutoramento de Pacheco d'Amorim, mas antes valores calculados com recurso a um computador utilizando o software *Mathematica 5.1*. Ao compararmos os valores determinados pelo computador e os apresentados na tese de doutoramento de Pacheco d'Amorim notamos que a diferença é quase nula. Na verdade, raramente ultrapassa o valor 2×10^{-7} e, a partir de 3.45, onde são utilizadas onze casas decimais, a diferença excepcionalmente passa os 0.1×10^{-11} . Além disso, não é frequente na época em que Pacheco d'Amorim apresentou a sua tese de doutoramento existirem tabelas tão pormenorizadas da função de Gauss, ainda que estes valores não sejam originais do autor, pois, por exemplo, Montessus de Ballore (1908) e Bachelier (1912) apresentam exactamente a mesma tabela sem referir (também) qual é a sua origem. Apesar de os valores apresentados na tabela não serem originais de Pacheco d'Amorim é, de facto, de admirar a minúcia, para a época, dos valores patentes na referida tabela.

Teorema 7.12. *A probabilidade de a distância entre $\frac{n_p}{n_q}$ e $\frac{p}{q}$ ser superior a um determinado valor positivo ε é dado por⁽³³⁾*

$$\mathbb{P} \left(\left| \frac{n_p}{n_q} - \frac{p}{q} \right| > \varepsilon \right) = 1 - \theta \left[\frac{\varepsilon q^2 \sqrt{m}}{\sqrt{2pq}(1 \pm \varepsilon q)} \right], \quad (7.214)$$

que tende para zero quando o número de provas de Bernoulli m tende para infinito.

Demonstração. Para que se verifique $\left| \frac{n_p}{n_q} - \frac{p}{q} \right| > \varepsilon$ teremos duas possibilidades: ou ocorrerá $\frac{n_p}{n_q} - \frac{p}{q} > \varepsilon$, nos casos em que o número de sucesso é superior à combinação modal; ou verificar-se-á $\frac{n_p}{n_q} - \frac{p}{q} < -\varepsilon$, se o número de sucessos for inferior ao da combinação modal. No primeiro caso, considerando um afastamento $k = \lambda \sqrt{2mpq} > 0$, teremos

$$\begin{aligned} \frac{n_p}{n_q} - \frac{p}{q} &= \frac{mp + \lambda \sqrt{2mpq}}{mq - \lambda \sqrt{2mpq}} - \frac{p}{q} = \frac{\lambda \sqrt{2mpq}}{mq^2 - \lambda q \sqrt{2mpq}} = \\ &= \frac{\lambda \sqrt{2pq}}{q^2 \sqrt{m} - \lambda q \sqrt{2pq}} > 0. \end{aligned} \quad (7.215)$$

Na segunda situação, supondo um afastamento $k = \lambda \sqrt{2mpq} < 0$ (logo $\lambda < 0$), teremos igualmente

$$\frac{n_p}{n_q} - \frac{p}{q} = \frac{\lambda \sqrt{2pq}}{q^2 \sqrt{m} - \lambda q \sqrt{2pq}} < 0. \quad (7.216)$$

Deste modo vamos obter

$$\begin{aligned} \left| \frac{n_p}{n_q} - \frac{p}{q} \right| > \varepsilon &\Leftrightarrow \frac{n_p}{n_q} - \frac{p}{q} > \varepsilon \vee \frac{n_p}{n_q} - \frac{p}{q} < -\varepsilon \Leftrightarrow \\ &\Leftrightarrow \frac{\lambda \sqrt{2pq}}{q^2 \sqrt{m} - \lambda q \sqrt{2pq}} > \varepsilon \vee \frac{\lambda \sqrt{2pq}}{q^2 \sqrt{m} - \lambda q \sqrt{2pq}} < -\varepsilon \Leftrightarrow \\ &\Leftrightarrow \lambda > \frac{\varepsilon q^2 \sqrt{m}}{\sqrt{2pq}(1 + \varepsilon q)} \vee \lambda < \frac{-\varepsilon q^2 \sqrt{m}}{\sqrt{2pq}(1 - \varepsilon q)}. \end{aligned} \quad (7.217)$$

⁽³³⁾ Este resultado apresentado por Pacheco d'Amorim está incorrecto, conforme destacaremos na demonstração deste teorema. Todavia, poderíamos enunciar

$$\mathbb{P} \left(\left| \frac{n_p}{n_q} - \frac{p}{q} \right| > \varepsilon \right) \leq 1 - \theta(\lambda_m)$$

onde $\lambda_m = \min \left\{ \frac{\varepsilon q^2 \sqrt{m}}{\sqrt{2pq}(1 + \varepsilon q)}, \frac{\varepsilon q^2 \sqrt{m}}{\sqrt{2pq}(1 - \varepsilon q)} \right\}$ e deduziríamos conclusão final, referente à convergência da probabilidade, igual.

Por conseguinte, recorrendo ao Teorema 7.11 (página 444) e utilizando a simetria da distribuição gaussiana, podemos concluir que

$$\begin{aligned}
 & \mathbb{P} \left(\left| \frac{n_p}{n_q} - \frac{p}{q} \right| > \varepsilon \right) = \\
 &= \mathbb{P} \left(\lambda > \frac{\varepsilon q^2 \sqrt{m}}{\sqrt{2pq}(1 + \varepsilon q)} \right) + \mathbb{P} \left(\lambda < \frac{-\varepsilon q^2 \sqrt{m}}{\sqrt{2pq}(1 - \varepsilon q)} \right) = \\
 &= \frac{1}{2} \left[1 - \theta \left(\frac{\varepsilon q^2 \sqrt{m}}{\sqrt{2pq}(1 + \varepsilon q)} \right) \right] + \frac{1}{2} \left[1 - \theta \left(\frac{\varepsilon q^2 \sqrt{m}}{\sqrt{2pq}(1 - \varepsilon q)} \right) \right] = \\
 &= 1 - \frac{1}{2} \left[\theta \left(\frac{\varepsilon q^2 \sqrt{m}}{\sqrt{2pq}(1 + \varepsilon q)} \right) + \theta \left(\frac{\varepsilon q^2 \sqrt{m}}{\sqrt{2pq}(1 - \varepsilon q)} \right) \right]. \quad (7.218)
 \end{aligned}$$

Esta expressão, conforme referimos, não corresponde à probabilidade (incorrecta) enunciada no teorema por Pacheco d'Amorim. Porém, podemos definir a quantidade λ_m através de

$$\lambda_m = \min \left\{ \frac{\varepsilon q^2 \sqrt{m}}{\sqrt{2pq}(1 + \varepsilon q)}, \frac{\varepsilon q^2 \sqrt{m}}{\sqrt{2pq}(1 - \varepsilon q)} \right\} \quad (7.219)$$

e, através desta quantidade, concluir que

$$\mathbb{P} \left(\left| \frac{n_p}{n_q} - \frac{p}{q} \right| > \varepsilon \right) \leq 1 - \theta(\lambda_m). \quad (7.220)$$

Desta forma, por mais pequeno que ε seja, esta probabilidade tenderá rapidamente para zero, por causa do factor $\sqrt{m}$. $\square$

Esta dedução efectuada por Pacheco d'Amorim é idêntica à apresentada por Borel (1909, p. 63–65), inclusive nos erros comentados. Borel demonstra o Teorema de Bernoulli ou Lei Fraca dos Grandes Números através deste teorema.

Pacheco d'Amorim ilustra a utilidade deste resultado aplicando-o na resolução do problema a seguir exposto.

Problema 7.3. Jogam-se 200 jogos de cara ou coroa, a tostão cada jogo. Qual a probabilidade de ganhar ou perder uma quantia superior a 10 tostões?

SOLUÇÃO: Uma vez que neste caso temos $p = q = \frac{1}{2}$, $m = 200$ e $k > 10$, logo, se

$$k = \lambda \sqrt{2 \cdot 200 \cdot \frac{1}{2} \cdot \frac{1}{2}} = 10\lambda > 10,$$

será $\lambda > 1$ e

$$1 - \theta(1) = \frac{16}{100} = \frac{4}{25},$$

que será o resultado aproximado do problema. $\diamond$

Por fim Pacheco d'Amorim refere que Borel (1909) generalizou a Lei dos desvios para provas de Bernoulli independentes com probabilidades distintas, isto é, para o caso de as tiragens serem feitas em urnas de composições diferentes. Desta forma, se representarmos por p_i e q_i respectivamente, as probabilidades de sucesso e de insucesso da urna i e efectuarmos m_i tiragens na urna i , com $m = m_1 + \dots + m_n$, o número mais provável de bolas brancas será

$$m_1 p_1 + m_2 p_2 + \dots + m_n p_n. \quad (7.221)$$

O autor refere, assim, que Borel demonstrou que, nesta situação, a Lei dos Desvios ainda terá a mesma forma; apesar de se tomar para unidade do afastamento h a raiz quadrada do dobro da soma dos quadrados dos afastamentos correspondentes às diversas urnas, isto é,

$$h = \sqrt{2 \sum_{i=1}^n m_i p_i (1 - p_i)}, \quad (7.222)$$

concluindo-se que

$$\mathbb{P}(k) = \mathbb{P}\left(\left|\sum_{i=1}^n n_i - \sum_{i=1}^n m_i p_i\right| \leq k\right) = \frac{2}{\sqrt{\pi}} \int_0^{\lambda_1} e^{-\lambda} d\lambda = \theta(\lambda), \quad (7.223)$$

onde n_i corresponde ao número de sucessos nas m_i provas de Bernoulli com probabilidade de sucesso p_i e

$$\lambda_1 = \frac{k}{\sqrt{2 \sum_{i=1}^n m_i p_i (1 - p_i)}}. \quad (7.224)$$

Para comentarmos a análise de Pacheco d'Amorim do Teorema Limite Central recordemos um pouco da sua história. O primeiro resultado referente ao Teorema Limite Central foi obtido por Abraham de Moivre em 1730 e refere-se a provas de Bernoulli independentes com probabilidade de sucesso constante. Assim, considerando que X_i , com $i = 1, \dots$, são provas de Bernoulli independentes com probabilidade de sucesso igual a p , de Moivre conclui que

$$\lim_{m \rightarrow +\infty} \mathbb{P} \left(\frac{\sum_{i=1}^m X_i - mp}{\sqrt{mp(1-p)}} \leq \varepsilon \right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\varepsilon} e^{-\frac{x^2}{2}} dx = \Phi(\varepsilon) \quad (7.225)$$

onde Φ representa a função de distribuição de uma variável gaussiana standard. Este resultado tornou-se mais conhecido depois dos trabalhos de Laplace (1749–1827) e de Gauss (1777–1855), que são posteriores ao de Moivre, sendo o resultado acima indicado usualmente denominado Teorema Limite Central de Moivre-Laplace e a distribuição limite denominada (como temos feito) distribuição de Gauss ou gaussiana. Desta forma, a estes resultados, aplica-se uma vez mais a lei de eponímia de Stephen Stigler (1980).

Este Teorema Limite Central supõe o valor de p fixo (provas de Bernoulli com probabilidade de sucesso igual) e fornece boas aproximações da distribuição binomial, desde que p não esteja próximo de zero nem da unidade. Por este motivo, para tais situações Poisson criou a lei que denominou Lei dos Pequenos Números (em comparação com a Lei dos Grandes Números que é um nome de que Poisson também é autor). Esta lei estabelece que, considerando a soma de variáveis aleatórias independentes $\{X_i\}_{i \geq 1}$ com distribuição de Bernoulli com parâmetro p_m e tomando os limites de $m \rightarrow \infty$ e $p_m \rightarrow 0$ simultaneamente, de forma a que $m p_m \rightarrow \lambda$, vamos obter

$$\mathbb{P} \left(\sum_{i=1}^m X_i = k \right) \longrightarrow e^{-\lambda} \frac{\lambda^k}{k!}, \quad (7.226)$$

que corresponde à função de probabilidade de uma variável aleatória com distribuição de Poisson. Este mesmo autor generaliza o Teorema Limite Central para provas de Bernoulli com probabilidade de sucesso variável, isto é, considerando uma sequência de provas de Bernoulli X_i independentes com probabilidade de sucesso p_i , então

$$\lim_{m \rightarrow +\infty} \mathbb{P} \left(\frac{\sum_{i=1}^m X_i - \sum_{i=1}^m p_i}{\sqrt{\sum_{i=1}^m p_i (1 - p_i)}} \leq \varepsilon \right) = \Phi(\varepsilon). \quad (7.227)$$

Na década de 1870 Chebycheff analisa o caso geral, abandonando a análise restrita às provas de Bernoulli, considerando uma sequência de variáveis aleatórias $\{X_i\}_{i=1}^{\infty}$ independentes com valor esperado nulo ($\mathbb{E}(X_i) = 0$) e variância $\text{Var}(X_i) = \sigma_i^2$ que fossem limitadas ($|X_k| \leq b$), então, uma vez que $\text{Var}(\sum_{i=1}^m X_i) = \sum_{i=1}^m \sigma_i^2$, obtemos

$$\lim_{m \rightarrow +\infty} \mathbb{P} \left(\frac{\sum_{i=1}^m X_i}{\sqrt{\sum_{i=1}^m \sigma_i^2}} \leq z \right) = \Phi(z). \quad (7.228)$$

Este resultado não é correcto. Posteriormente o seu aluno Markov demonstrou que era necessário impor que

$$\text{Var} \left(\sum_{i=1}^m X_i \right) \xrightarrow{m \rightarrow +\infty} +\infty. \quad (7.229)$$

Contudo, a imposição de que a variável tinha que ser limitada ($|X_k| \leq b$) era considerada demasiado restritiva, uma vez que muitas distribuições não são limitadas (como se sucede, por exemplo, com a distribuição gaussiana). Markov resolveu este problema introduzindo um método denominado método de truncagem. Este método consiste em considerar uma variável aleatória não limitada X e definir uma variável aleatória artificial $\tilde{X}$ através de

$$\tilde{X} = X \cdot \mathbb{I}_{\{|x| \leq b\}} = \begin{cases} X & \text{se } |X| \leq b \\ 0 & \text{se } |X| > b \end{cases}, \quad (7.230)$$

onde $\mathbb{I}$ representa a função indicatriz. Desta forma a variável aleatória $\tilde{X}$, por ser limitada, possui todos os momentos e, para concluirmos que X também tenha todos os momentos, temos que garantir que a diferença entre estas duas variáveis aleatórias é assintoticamente negligenciável, isto é,

$$\mathbb{P}(|X| \leq b) = 1 \quad \text{ou} \quad \mathbb{P}(|X| > b) = 0, \quad (7.231)$$

para um qualquer valor b positivo. Com este método podemos definir outra condição conhecida por condição de integrabilidade uniforme, útil no contexto da convergência de soma de variáveis aleatórias, através de

$$\lim_{b \rightarrow \infty} \left(\sup_{1 \leq k \leq m} \mathbb{E} \{ X_k \cdot \mathbb{I}_{\{|x| > b\}} \} \right) = 0. \quad (7.232)$$

Outra condição importante é combinar $\mathbb{P}(|X_k| > \varepsilon) = 0$, para qualquer $\varepsilon > 0$ e $k = 1, \dots$, com a condição de Markov expressa em (7.229), sendo conhecida como condição de integrabilidade uniforme assintótica

$$\lim_{m \rightarrow \infty} \left(\max_{1 \leq k \leq m} \mathbb{P} \left(\left| \frac{X_k - \mu_k}{\sqrt{\sum_{i=1}^m \sigma_i^2}} \right| > \varepsilon \right) \right) = 0, \quad (7.233)$$

onde cada variável aleatória X_k da sucessão $\{X_i\}_{i=1}^{\infty}$ é negligenciável em relação à soma de todas as variáveis aleatórias. Chegamos, deste modo, ao primeiro Teorema Limite Central de âmbito geral, que foi obtido por Lyapounov em 1901 e que pode ser enunciado da seguinte forma: consideremos uma sucessão $\{X_i\}_{i=1}^{\infty}$ de variáveis aleatórias independentes com $\mathbb{E}(X_k) = \mu_k$, $\text{Var}(X_k) = \sigma_k^2$ com momentos de ordem $2 + \delta$ limitados $\left(\mathbb{E}(|X_k|^{2+\delta}) < \infty \right)$, para um qualquer $\delta > 0$ e $k = 1, 2, \dots$, tal que

$$\frac{\sum_{k=1}^m \mathbb{E}(|X_k - \mu_k|^{2+\delta})}{\left(\sqrt{\sum_{i=1}^m \sigma_i^2} \right)^{2+\delta}} \xrightarrow{m \rightarrow \infty} 0, \quad (7.234)$$

então

$$\lim_{m \rightarrow \infty} \mathbb{P} \left(\frac{\sum_{k=1}^m (X_k - \mu_k)}{\sqrt{\sum_{i=1}^m \sigma_i^2}} \leq z \right) = \Phi(z), \forall z \in \mathbb{R}. \quad (7.235)$$

Posteriormente, em 1922, Lindeberg obtém o seguinte Teorema Limite Central: consideremos uma sucessão $\{X_i\}_{i=1}^{\infty}$ de variáveis aleatórias independentes com $\mathbb{E}(X_k) = \mu_k$, $\text{Var}(X_k) = \sigma_k^2 < \infty$ tal que

$$\forall \varepsilon > 0 : \lim_{m \rightarrow \infty} \left(\frac{\sum_{k=1}^m \mathbb{E} \left[(X_k - \mu_k)^2 \cdot \mathbb{I}_{\{|X_k - \mu_k| > \varepsilon s_n\}} \right]}{\sum_{i=1}^m \sigma_i^2} \right) = 0, \quad (7.236)$$

então teremos

$$\lim_{m \rightarrow \infty} \left(\max_{1 \leq k \leq m} \left(\frac{\sigma_k^2}{\sum_{i=1}^m \sigma_i^2} \right) \right) = 0 \quad (7.237)$$

e

$$\lim_{m \rightarrow \infty} \mathbb{P} \left(\frac{\sum_{k=1}^m (X_k - \mu_k)}{\sqrt{\sum_{i=1}^m \sigma_i^2}} \leq z \right) = \Phi(z), \forall z \in \mathbb{R}. \quad (7.238)$$

Finalmente, em 1935, Feller demonstra que as condições apresentadas no Teorema Limite Central de Lindeberg são suficientes e necessárias, isto é, se tivermos uma sucessão $\{X_i\}_{i=1}^{\infty}$ de variáveis aleatórias independentes com variância finita, a condição (7.236) é equivalente a (7.237) e (7.238), sendo por esta razão por vezes denominado por Teorema limite central de Lindeberg-Feller.

Em relação à análise de Pacheco d'Amorim, no que se refere ao Teorema Limite Central, a sua exposição é decepcionante e vai menos longe que, por exemplo, Bachelier e Poincaré. De facto, a exposição de Pacheco d'Amorim segue uma linha clássica de aproximar probabilidades (de um conjunto de pontos rodeando a moda de uma distribuição discreta, única situação que

aborda), usando a regra de Stirling. Não se ocupa do contexto mais geral de considerar uma sucessão de variáveis aleatórias convergente para uma variável aleatória gaussiana, pelo que não tem necessidade de usar ferramentas mais sofisticadas para justificar as aproximações. Bachelier e Poincaré, confrontando-se com o problema, árduo, de tratar somas de variáveis aleatórias com um número crescente de parcelas, desenvolvem instrumentos mais delicados para a sua abordagem. Poincaré, em particular, define a função característica (no seu caso, uma transformada de Laplace que não restringe ao eixo imaginário, como posteriormente viria a ser norma). Já Laplace, Poisson, Chebycheff, Lyapounov, antes se tinham apercebido das vantagens de trabalhar num espaço transformado, mas parece ser Poincaré⁽³⁴⁾ a merecer a prioridade na preferência pela transformada de Fourier. Não tem ainda, no entanto, os resultados fundamentais neste campo — os teoremas da convergência de Lévy e Cramér. Correndo assim o perigo de alguma injustiça, entendemos que Pacheco d’Amorim não terá tido os cuidados que modernamente seriam exigidos na procura de documentação actualizada. Refira-se, por exemplo, que um ano antes da apresentação da tese de doutoramento de Pacheco d’Amorim, para festejar o segundo centenário da Lei dos Grandes Números [Bernoulli (1713)], Markov (1913) apresenta um suplemento, escrito em francês, da terceira edição russa da sua obra de Cálculo das Probabilidades intitulado “*Démonstration du second théorème-limite du calcul des probabilités par la méthode des moments*”. Neste suplemento de 66 páginas Markov inclui a desigualdade de Bienaymé-Chebycheff, a sua generalização desta desigualdade (desigualdade de Markov), a Lei Fraca dos Grandes Números (não restrito à soma de provas de Bernoulli), o Teorema Limite

⁽³⁴⁾ Lévy (1925, p. 161) refere que a noção de função característica foi apresentada por Cauchy em 1853, mas o nome função característica só surgiu na obra de Poincaré para o valor esperado $\mathbb{E}\left(e^{tx}\right)$ e não de $\mathbb{E}\left(e^{itx}\right)$ como actualmente denominamos.

Central de Lyapounov, e apresenta ainda um estudo onde estende as leis do Cálculo das Probabilidades a uma sequência de provas onde existe dependência entre elas, investigando, num texto com 20000 letras, a probabilidade de uma prova (letra) ser vogal, de ser vogal depois de uma consoante e de ser uma vogal após outra vogal, num dos estudos pioneiros das actualmente apelidadas cadeias de Markov.

7.6 Comentário geral ao capítulo

Pacheco d'Amorim complementa o anterior desenvolvimento de alguns resultados exactos, correntes na literatura deste campo, em que a exposição da Lei dos Grandes Números é bastante mais completa que a do Teorema Limite Central.

O autor faz uma dedução rigorosa da Lei Fraca dos Grandes Números para a situação de somas de provas de Bernoulli com probabilidade constante, mas não analisa a situação de provas com probabilidade variável, que Poisson (1837) já tinha obtido, ou a situação mais geral de somas de qualquer variável sob determinadas condições sobre os momentos, tal como Chebycheff (1867) anteriormente efectuara, nem há uma tentativa de obter a Lei Forte dos Grandes Números, tal como Borel fizera em 1909 num artigo que provavelmente era do conhecimento de Pacheco d'Amorim. No que respeita ao Teorema Limite Central a análise efectuada é igualmente restrita à situação mais elementar de provas de Bernoulli com probabilidade constante, havendo apenas uma referência a Borel para a generalização para o caso de probabilidades variáveis, mas nenhuma tentativa de analisar o caso mais geral, onde a escola russa, nomeadamente Chebycheff e Lyapounov, já tinha obtido alguns resultados. De facto, Pacheco d'Amorim neste capítulo dedicado às convergências restringe-se à análise de somas de provas de Bernoulli com probabili-

dade de sucesso constante, não se ocupando do caso mais geral de somas de quaisquer variáveis aleatórias. Todavia, estas críticas serão atenuadas no que concerne à Lei Fraca dos Grandes Números, uma vez que o autor, no capítulo ESPERANÇA MATEMÁTICA E VALOR MÉDIO, que de seguida analisaremos, generaliza o resultado agora obtido para provas de Bernoulli independentes com probabilidade constante para provas idênticas e independentes com um número finito de resultados e, posteriormente, para lançamentos idênticos e independentes em regiões, numa curiosa dedução comparável à construção do integral de Riemann.

Refira-se, no entanto, que a tese de doutoramento de Pacheco d'Amorim foi apresentada numa época (anterior a 1920), em relação à qual Cramér (1976) tem uma postura muito crítica no que se refere ao desenvolvimento da Teoria da Probabilidade, comentando que os “*livros e artigos em problemas de probabilidade eram demasiado óbvios faltando-lhe rigor matemático*” (exceptuando os desenvolvimentos russos, nomeadamente os trabalhos realizados por Chebycheff e Lyapounov, que parecem ser totalmente desconhecidos quer para Pacheco d'Amorim quer para os autores franceses seus contemporâneos). Le Cam (1986), partilhando desta opinião, é também muito severo com os autores franceses desta época, escrevendo que “*Bertrand e Poincaré escreveram tratados sobre o cálculo de probabilidades, um assunto que nenhum dos dois parecia conhecer*”. Considerando que Pacheco d'Amorim se baseia principalmente (ou mesmo exclusivamente) nesta escola, a avaliar pelas poucas referências que faz ao longo da sua obra, poder-se-ia afirmar que, nos aspectos aqui comentados, Pacheco d'Amorim não consegue ir mais longe que outros autores seus precedentes, devendo incluir-se os *Elementos de Cálculo das Probabilidades* de Pacheco d'Amorim nos juízos críticos de Cramér e de Le Cam.

Contudo, apesar de Pacheco d'Amorim não desenvolver os resultados as-

simptóticos existentes na época, há a realçar neste capítulo a forma original como justifica a necessidade de dividir os afastamentos em relação à combinação modal por $\sqrt{m}$ para haver uma convergência em distribuição.

Desta forma, parece-nos que Pacheco d'Amorim poderia ter tentado ir mais longe, pois, a avaliar pela CONCLUSÃO da sua tese, onde arquitecta as aplicações da Teoria da Probabilidade, o autor considera a Lei dos Grandes Números (Teoremas de Bernoulli) e o Teorema Limite Central (Lei dos Desvios) dois resultados fundamentais na Teoria da Probabilidade. De facto, os resultados referentes a convergências assumem um papel preponderante na Teoria da Probabilidade, não só nas aplicações como na própria concepção de probabilidade, como demonstra a axiomática de Richard von Mises baseada em colectivos, de tal forma que Gnedenko e Kolmogoroff, no prefácio da sua obra sobre limites de somas de variáveis independentes, escreveram:

“In reality, however, the epistemological value of the theory of probability is revealed only by limit theorems. Moreover, without limit theorems it is impossible to understand the real content of the primary concept of all our sciences — the concept of probability. In fact, all epistemologic value of the theory of probability is based on this: that large-scale random phenomena in their collective action create strict, nonrandom regularity.”

[Gnedenko e Kolmogoroff, 1954, Preface, p. 1]

Capítulo 8

Esperança Matemática e Valor Médio

Pacheco d'Amorim começa este capítulo por definir esperança matemática e valor médio de uma função f em relação a uma tiragem à sorte numa classe $\mathbf{A}$ ou a um lançamento à sorte numa região $\mathbf{X}$. Para o autor o conceito esperança matemática refere-se unicamente a um subconjunto da classe (região) possível da escolha aleatória, sendo o valor médio equivalente à esperança matemática determinada em toda a classe (região) possível. De seguida o autor deduz algumas propriedades destas medidas, tais como a aditividade em relação a classes (regiões) disjuntas, aditividade em relação à função f em estudo, a multiplicação em selecções efectuadas em classes e regiões compostas $\mathbf{X} \times \mathbf{Y}$ quando é possível decompor a função f em duas funções tais que cada uma só dependa do resultado de uma das escolhas, i.e. $f = f_1(x) f_2(y)$. Demonstra, ainda, para a situação de um lançamento de um ponto $(x_1, \dots, x_n)$ na região $\mathbf{X} \subset \mathbb{R}^n$, a existência do valor esperado de uma transformação $f : \mathbf{X} \rightarrow \mathbf{Z} \subset \mathbb{R}$, e, consequentemente, de um ponto imagem $z = f(x_1, \dots, x_n)$, bem como de uma função qualquer desse ponto imagem $\varphi(z)$.

Pacheco d'Amorim deduz uma desigualdade semelhante à de Markov, onde conclui que, se pudermos garantir que o valor esperado de uma função não negativa seja inferior a uma quantidade, que pode assumir um valor tão baixo como pretendamos, i.e. $\mathbb{E}_{\mathbf{x}} [f(x)] < \delta$ para todo $\delta > 0$, então a probabilidade de a função ser positiva converge para zero. Seguidamente Pacheco d'Amorim resolve o problema que denominamos por problema da curva flexível e inextensível, concluindo que o valor esperado do quadrado da distância entre os dois pontos extremos da curva lançada à sorte é nulo. Consequentemente, com recurso à desigualdade deduzida pelo autor, podemos ainda depreender que é igual a um a probabilidade de o quadrado da distância entre os extremos da curva lançada à sorte ser igual a zero. Uma conclusão, à primeira vista, contra intuitiva mas que, no entanto, é verdadeira com as definições de lançamento de curva flexível e inextensível que o autor concebe.

O autor resolve ainda o problema da agulha de Buffon recorrendo ao conceito de esperança matemática, após deduzir que a probabilidade de a agulha intersectar um recta é independente da forma da agulha e é proporcional ao comprimento desta. Assim, recorrendo a um comprimento e formato adequado, proposto por Barbier em 1860, deduz facilmente a probabilidade pretendida para o caso em que a agulha possui um comprimento inferior à distância entre duas rectas paralelas.

Pacheco d'Amorim utiliza o conceito de valor esperado para demonstrar a Lei Fraca dos Grandes Números quando temos experiências aleatórias independentes com dois resultados possíveis e associamos a cada resultado um número. Desta forma o autor conclui, em relação a esta experiência, que

$$\lim_{m \rightarrow +\infty} \mathbb{P}(|\overline{X}_m - \mu| < \varepsilon) = 1, \quad \forall \varepsilon > 0. \quad (8.1)$$

Com este resultado Pacheco d'Amorim salienta a importância da média empírica na análise estatística. De seguida o autor generaliza este resultado

para a situação em que temos experiências com n modalidades, a que associamos n valores numéricos distintos e, finalmente, demonstra que o resultado também é válido para os lançamentos, isto é, para a probabilidade contínua. Esta última demonstração é baseada no caso anterior, onde temos n modalidades distintas, considerando que a região onde é efectuado o lançamento está dividido em n classes e comprovando a existência do limite quando o número de classes n tende para infinito e a amplitude de cada classe tende para zero, numa dedução bastante parecida à construção do integral de Riemann.

Por fim o autor analisa algumas propriedades das quantidades aleatórias λ , caracterizadas pela lei de Gauss, determinando o valor médio de algumas características destas quantidades, tais como o valor médio de λ , do valor absoluto de λ e de λ^2 . Após concluir que

$$\frac{\mathbb{M}(\lambda^2)}{[\mathbb{M}(|\lambda|)]^2} = \frac{\pi}{2} \quad (8.2)$$

refere, sem apresentar qualquer estudo, que este resultado pode ser utilizado para simular o valor de π .

8.1 Esperança Matemática e valor médio em Classes

Pacheco d'Amorim começa por definir a esperança matemática e o valor esperado de uma tiragem, à sorte, de um elemento de uma classe finita para depois deduzir algumas propriedades destes conceitos.

8.1.1 Definição

Consideremos, em relação a uma determinada tiragem à sorte, uma classe possível⁽¹⁾ $\mathbf{A} = \{a_1, \dots, a_{n_{\mathbf{A}}}\}$ com probabilidade de cada elemento a_i , em relação à classe $\mathbf{A}$, representada por $\mathbb{P}_{\mathbf{A}}(a_i)$. Seja $\mathbf{A}' = \{a'_1, \dots, a'_{n_{\mathbf{A}'}}\}$ uma classe contida em $\mathbf{A}$ e f uma função definida na classe $\mathbf{A}'$, que a cada elemento da classe $\mathbf{A}'$ faz corresponder um número. Se somarmos, para todos os elementos a'_i da classe $\mathbf{A}'$, os termos obtidos pela multiplicação da probabilidade $\mathbb{P}_{\mathbf{A}}(a'_i)$ pela sua imagem da função f , i.e. $f(a'_i)$, obtemos um número que é denominado por esperança matemática da função f em relação à classe $\mathbf{A}'$.

Definição 8.1 (Esperança matemática em classes). Representando a **esperança matemática da função f em relação à classe $\mathbf{A}'$** com o símbolo $\mathbb{E}_{\mathbf{A}'}[f(a)]$, esta quantidade é definida por

$$\mathbb{E}_{\mathbf{A}'}[f(a)] = \sum_{a_i \in \mathbf{A}'} a_i \mathbb{P}_{\mathbf{A}}(a_i). \quad (8.3)$$

Se, na definição de esperança matemática em classes, a classe $\mathbf{A}'$ corresponder com a classe $\mathbf{A}$, então obtemos o valor médio ou valor provável da função f na classe $\mathbf{A}$.

Definição 8.2 (Valor médio em classes). Representando com o símbolo $\mathbb{M}[f(a)]$ o **valor médio** ou **valor provável da função f** , este será definido por

$$\mathbb{M}[f(a)] = \sum_{a_i \in \mathbf{A}} a_i \mathbb{P}_{\mathbf{A}}(a_i). \quad (8.4)$$

⁽¹⁾ Notemos que Pacheco d'Amorim, neste capítulo, não se refere à classe total possível, mas unicamente a uma classe possível. Como tal, a classe $\mathbf{A}$ a que nos referimos poderá não corresponder à classe na qual foi efectuada a tiragem, podendo ser apenas um subconjunto desta. Contudo, a probabilidade que iremos utilizar será sempre relativa a esta classe $\mathbf{A}$. Desta forma, os conceitos que o autor apresenta são de esperança matemática e valor esperado condicionados a $\mathbf{A}$, isto é, $\mathbb{E}_{\mathbf{A}'}[\varphi(x)]$ corresponderá a $\mathbb{E}_{\mathbf{A}'}[\varphi(x) | x \in \mathbf{A}]$.

Pacheco d'Amorim ilustra a utilização destes dois novos conceitos com o exemplo a seguir apresentado.

Exemplo 8.1. Consideremos um lançamento de um dado. Seja f a função que a cada face dum dado faz corresponder o respectivo número de pontos. Desta forma, sendo $\mathbf{A} = \{1, 2, 3, 4, 5, 6\}$, o valor médio será

$$\mathbb{M}[f(a)] = \frac{1}{6} \cdot 1 + \frac{1}{6} \cdot 2 + \cdots + \frac{1}{6} \cdot 6 = 3.5. \quad (8.5)$$

A esperança matemática relativa à classe $\mathbf{A}' = \{1, 2\}$ (faces 1 e 2) é dada por

$$\mathbb{E}_{\mathbf{A}'}[f(a)] = \frac{1}{6} \cdot 1 + \frac{1}{6} \cdot 2 = 0.5. \quad (8.6)$$



Qual seria o sentido que Pacheco d'Amorim atribuía ao resultado obtido no exemplo anterior? Como justificar que a esperança matemática das faces $\{1, 2\}$ assumia o valor 0.5, que é inferior a qualquer dos possíveis resultados desta experiência? E qual o objectivo? Parece-nos que Pacheco d'Amorim procura criar uma medida que permita-lhe concluir a aditividade, podendo desta forma dividir o cálculo do valor médio em parcelas. Notemos que, utilizando a definição de esperança matemática, expressão (8.3), e a definição de probabilidade condicionada, teremos

$$\begin{aligned} \mathbb{E}_{\mathbf{A}'}[f(a)] &= \sum_{a_i \in \mathbf{A}'} a_i \mathbb{P}_{\mathbf{A}}(a_i) = \sum_{a_i \in \mathbf{A}'} a_i \mathbb{P}_{\mathbf{A}'}(a_i) \mathbb{P}_{\mathbf{A}}(\mathbf{A}') = \\ &= \mathbb{P}_{\mathbf{A}}(\mathbf{A}') \sum_{a_i \in \mathbf{A}'} a_i \mathbb{P}_{\mathbf{A}'}(a_i) = \mathbb{E}_{\mathbf{A}'}[f(a) \mid a \in \mathbf{A}'] \mathbb{P}_{\mathbf{A}}(\mathbf{A}'), \end{aligned} \quad (8.7)$$

onde recorreremos ao Teorema da Probabilidade Composta (cf. Proposição VII presente na página 195). Acrescente-se que, considerando que as classes $\mathbf{A}_1, \dots, \mathbf{A}_n$ definem uma partição de $\mathbf{A}'$ e recorrendo ao Teorema das Probabilidades Totais, obtemos

$$\mathbb{E}_{\mathbf{A}'}[f(a) \mid a \in \mathbf{A}'] = \sum_{i=1}^n \mathbb{E}_{\mathbf{A}_i}[f(a) \mid a \in \mathbf{A}_i] \mathbb{P}_{\mathbf{A}'}(\mathbf{A}_i). \quad (8.8)$$

Deste modo, parece que Pacheco d'Amorim procura atingir este resultado mas de uma forma mais acessível, isto é, onde possa decompor a esperança matemática de uma classe $\mathbf{A}'$ unicamente na soma das esperanças matemáticas das classes $\mathbf{A}_i$ que compõem a partição, obtendo a aditividade da esperança matemática

$$\mathbb{E}_{\mathbf{A}'} [f(a)] = \sum_{i=1}^n \mathbb{E}_{\mathbf{A}_i} [f(a)], \quad (8.9)$$

conforme analisaremos, com mais detalhe, no Teorema 8.1, presente na página 470. Assim, $\mathbb{E}_{\mathbf{A}'} [f(a)]$ não corresponde a um valor esperado condicionado à região $\mathbf{A}'$, pois não utilizamos no seu cálculo as probabilidades condicionadas à região $\mathbf{A}'$, mas, antes, uma espécie de valor esperado truncado. Refira-se que o que Pacheco d'Amorim ganha, em relação à fórmula usualmente utilizada na actualidade (cf. expressão (8.8)), na simplificação da apresentação do resultado, perde na interpretação, pois, de facto, afirmar $\mathbb{E}_{\{1,2\}} [f(a)] = 0.5$ não nos parece fazer muito sentido, enquanto afirmar $\mathbb{E}_{\{1,2\}} [f(a) | a \in \{1, 2\}] = 1.5$ tem um significado bem claro. Saliente-se, contudo, que Castelnuovo (1919), numa obra fantástica onde concilia os avanços, em Teoria da Probabilidade, da escola francesa com os provenientes da escola russa, que, repetimos, Pacheco d'Amorim parece desconhecer, utiliza exactamente o mesmo método.

Após definir e exemplificar estes dois conceitos, esperança matemática e valor médio, Pacheco d'Amorim refere, em nota de rodapé:

“Embora esta distinção entre esperança matemática e valor médio não costume vir explicitamente feita nos livro de Probabilidades, todos os autores dão a estes termos a significação que acabamos de atribuir-lhes.”

[Pacheco d'Amorim, 1914, p. 138]

Esta afirmação de Pacheco d'Amorim é surpreendente, pois, em relação às obras por nós consultadas, onde se incluem as poucas obras referidas pelo

autor na sua tese de doutoramento, não encontramos esta forma de definir esperança matemática e valor médio. No entanto, Guido Castelnuovo (1865–1952) na sua obra notável *Calcolo delle Probabilità*, de 1919⁽²⁾, afirma:

“Il valor medio da alcuni autori è detto valore probabile (con una locuzione che si presta ad equivoci, perchè talvolta usata in senso diverso); da altri speranza matematica. Effettivamente la speranza matematica inerente a guadagni aleatori (n.19) è il valor medio del detti guadagni, quando si tenga conto di tutte le alee che può correre il giuocatore (vincite e perdite)”

[Guido Castelnuovo, 1919, p. 32]

Desta forma Castelnuovo também considera que a esperança matemática é determinada utilizando unicamente uma parte dos valores que a variável assume, enquanto o valor médio engloba obrigatoriamente todos os valores. Todavia Castelnuovo, associando a esperança matemática com as origens da Teoria da Probabilidade nos jogos de azar, entende que a esperança matemática é determinada utilizando unicamente os ganhos de um jogador, enquanto

⁽²⁾ Esta obra de Castelnuovo, apesar de ser apenas cinco anos posterior à tese de doutoramento de Pacheco d’Amorim, num contexto em que a Europa está mergulhada numa guerra imensa, quando comparada com as obras desta época da escola francesa de probabilidade, tais como as de Bertrand, Poincaré, Bachelier ou Borel (que também refere), insere uma visão da Teoria da Probabilidade muito mais moderna. Castelnuovo está ciente das inovações provenientes da escola russa de Petersburgo, dando grande importância aos trabalhos de Chebycheff, Lyapounov e Markov. Com efeito, o autor declara: *“La grande opera di TCHEBYCHEF e della sua scola . . . (MARKOFF, LIAPOUNOFF,...) si accorgerà che essa costituisce il maggior contributo portato al calcolo delle probabilità dopo LAPLACE.”*. Este conhecimento dos avanços da escola russa devem-se, como o autor refere, à tradução para alemão *Wahrscheinlichkeits-Rechnung* em 1912 da obra de Markov sobre cálculo da probabilidade. Desta forma Castelnuovo insere, nesta obra, os resultados provenientes da escola francesa e da escola russa, bem como os desenvolvimentos efectuados pelo seu compatriota Francesco Paolo Cantelli.

o valor médio utiliza todos os valores que a variável pode assumir, os ganhos e as perdas (ou ganhos negativos). A definição de Pacheco d'Amorim insere a mesma ideia, que a esperança matemática é determinada utilizando unicamente uma parte, mas é mais geral, uma vez que pode ser aplicada a um subconjunto qualquer do suporte da variável aleatória e não restrita a ambientes de jogos de azar. Bertrand (1888, p. 47 e seguintes) define a esperança matemática de forma semelhante à de Castelnuovo:

“Cette égalité traduit la définition: le joueur échange sa mise contre une espérance mathématique. Si l'équivalence n'existe pas, le jeu n'est pas équitable.

L'espérance mathématique de celui qui a la probabilité p de recevoir la somme S est mesurée par le produit pS .”

[Bertrand, 1888, p. 48]

Assim, a esperança matemática na obra de Bertrand era restrita aos ganhos, de forma a obter o valor que o jogador está disponível a pagar para entrar no jogo. Borel (1909, p. 12) apresenta ideia semelhante e Poincaré (1896, p. 58) define a esperança matemática incluindo todos os valores positivos e negativos que um jogador pode ganhar, sendo o jogo equitativo (*équitable*) se todos os jogadores possuírem igual valor esperado. A mesma ideia é apresentada por Guimarães (1904, p. 22–24). Bachelier (1900, p. 32), Montessus de Ballore (1908, p. 75) e Bachelier (1912, p. 6), seguindo de igual forma o hábito de recurso a jogos de azar, definem esperança matemática total, como a medida que inclui todos os valores que a variável pode assumir, as perdas e os ganhos, e a esperança matemática benéfica e negativa, onde a diferença entre estas últimas quantidades corresponde à vantagem de um jogador ou esperança matemática total, que é nula num jogo equitativo.

Deste modo, a definição de esperança matemática, presente nestas obras, de facto não inclui todos os valores que a variável pode assumir. Contudo,

as definições por estes autores apresentadas não é idêntica à de Pacheco d'Amorim. Estes autores restringem-se a ambientes de jogos de apostas e pretendem, com esta definição, chegar ao conceito de jogo equitativo. Por esta razão analisam o ganho de cada jogador, que deverá ser igual para todos os jogadores, ou definem o ganho esperado para definirem a quantidade monetária que deve ser apostada. Na definição fornecida por Pacheco d'Amorim, sem recurso a jogos de apostas, o autor concebe este conceito para qualquer subconjunto de valores que a variável possa assumir, sendo, por esta razão, distinta das anteriores. Na nossa opinião, uma vez que o autor está a libertar-se da dependência dos jogos de apostas, presente, ainda no início do século XX, em muitas obras da Teoria da Probabilidade, não seria necessário efectuar uma distinção entre valor médio e esperança matemática, uma vez que o valor médio corresponde a um caso particular de esperança matemática (que poderia, por exemplo, apelidar de esperança matemática total tal como Bachelier). Apesar destas considerações, há a exaltar, principalmente em comparação com as obras da época da escola francesa, a libertação destes conceitos dos jogos de azar, pois Pacheco d'Amorim não só generaliza o valor esperado para qualquer subconjunto do suporte da variável aleatória, como a sua definição não se refere a qualquer tipo de jogo, podendo ser utilizada em qualquer contexto de experiências aleatórias.

8.1.2 Propriedades

Nesta secção vamos apresentar as propriedades deduzidas por Pacheco d'Amorim para a esperança matemática. Uma vez que o valor médio é, segundo as definições de Pacheco d'Amorim, um caso particular da esperança matemática, correspondente às situações onde se verifica $\mathbf{A}' = \mathbf{A}$, as propriedades que iremos deduzir como válidas para a esperança matemática serão igualmente válidas para o valor médio quando tal fizer sentido.

Na primeira propriedade o autor divide a classe $\mathbf{A}'$, em relação à qual é determinada a esperança matemática, nas sub-classes $\mathbf{A}_1, \dots, \mathbf{A}_n$, que formam uma partição de $\mathbf{A}'$, concluindo que a Esperança Matemática da classe $\mathbf{A}'$ é igual à soma das esperanças matemáticas de cada uma das sub-classes $\mathbf{A}_i$ que formam a partição.

Teorema 8.1 (Aditividade da esperança matemática em relação a classes disjuntas). *A esperança matemática duma classe é igual à soma das esperanças matemáticas das suas partes, isto é, sendo $\mathbf{A}' = \cup_{i=1}^n \mathbf{A}_i$ em que $\mathbf{A}_i \cap \mathbf{A}_j = \emptyset$ para $i \neq j$ (i.e. as classes $\mathbf{A}_1, \mathbf{A}_2, \dots, \mathbf{A}_n$ definem uma partição da classe $\mathbf{A}'$), então*

$$\mathbb{E}_{\mathbf{A}'} [f(a)] = \sum_{i=1}^n \mathbb{E}_{\mathbf{A}_i} [f(a)]. \quad (8.10)$$

Demonstração. Uma vez que $\mathbf{A}' = \cup_{i=1}^n \mathbf{A}_i$ e $\mathbf{A}_i \cap \mathbf{A}_j = \emptyset$ para $i \neq j$, pela definição de esperança matemática (Definição 8.1) teremos

$$\sum_{i=1}^n \mathbb{E}_{\mathbf{A}_i} [f(a)] = \sum_{i=1}^n \sum_{a_i \in \mathbf{A}_i} a_i \mathbb{P}_{\mathbf{A}}(a_i) = \sum_{a_i \in \mathbf{A}'} a_i \mathbb{P}_{\mathbf{A}}(a_i) = \mathbb{E}_{\mathbf{A}'} [f(a)].$$

□

A segunda propriedade deduzida por Pacheco d'Amorim decompõe a função f na soma de n funções f_i , sendo o valor esperado de f igual à soma dos valores esperados das funções f_i .

Teorema 8.2 (Aditividade da esperança matemática em relação à função). *A esperança matemática da soma é igual à soma das esperanças matemáticas, isto é, sendo*

$$f(a_j) = \sum_{i=1}^n f_i(a_j), \quad \forall a_j \in \mathbf{A}' \quad (8.11)$$

então

$$\mathbb{E}_{\mathbf{A}'} [f(a)] = \sum_{i=1}^n \mathbb{E}_{\mathbf{A}'} [f_i(a)]. \quad (8.12)$$

Demonstração. Esta propriedade é consequência imediata da definição de esperança matemática (Definição 8.1), pois

$$\begin{aligned} \sum_{i=1}^n \mathbb{E}_{\mathbf{A}'} [f_i(a)] &= \sum_{i=1}^n \sum_{a_j \in \mathbf{A}'} f_i(a_j) \mathbb{P}_{\mathbf{A}}(a_j) = \\ &= \sum_{a_j \in \mathbf{A}'} \mathbb{P}_{\mathbf{A}}(a_j) \sum_{i=1}^n f_i(a_j) = \mathbb{E}_{\mathbf{A}'} \left[\sum_{i=1}^n f_i(a) \right] = \mathbb{E}_{\mathbf{A}'} [f(a)]. \end{aligned}$$

□

De seguida o autor apresenta a forma de determinar a esperança matemática de uma classe composta $\mathbf{A} \times \mathbf{B}$, isto é, de uma classe obtida pelo produto cartesiano de duas classes $\mathbf{A}$ e $\mathbf{B}$. Recordemos que, desta forma, tal como referimos no capítulo das CLASSES FINITAS (capítulo 3), está subjacente ao resultado seguinte a hipótese de independência das escolhas dos dois elementos, isto é, a selecção à sorte de um elemento composto em $\mathbf{A} \times \mathbf{B}$ pode ser decomposto em duas tiragens à sorte independentes, uma na classe $\mathbf{A}$ e outra na classe $\mathbf{B}$ que não depende do elemento previamente escolhido na classe $\mathbf{A}$. Para que seja possível decompor o valor esperado de uma função f em relação a um complexo de classe $\mathbf{A} \times \mathbf{B}$ em dois valores esperados, cada um dependente apenas de uma das duas escolhas à sorte, a função f tem de ser decomponível na multiplicação de duas funções, f_1 e f_2 , cada uma delas unicamente dependente de uma das tiragens à sorte.

Teorema 8.3 (Multiplicação da esperança matemática em relação a tiragens e funções independentes). *Sejam $\mathbf{A} = \{a_1, \dots, a_{n_A}\}$ e $\mathbf{B} = \{b_1, \dots, b_{n_B}\}$ duas classes possíveis e $\mathbf{A} \times \mathbf{B}$ a classe composta das duas. Consideremos uma subclasse $\mathbf{A}' \times \mathbf{B}' \subset \mathbf{A} \times \mathbf{B}$, tal que $\mathbf{A}' \subset \mathbf{A}$ e $\mathbf{B}' \subset \mathbf{B}$ e seja f uma função tal que $f(a_i, b_j) = f_1(a_i) f_2(b_j)$ para $i = 1, \dots, n_{\mathbf{A}'}$ e $j = 1, \dots, n_{\mathbf{B}'}$. Nestas condições teremos*

$$\mathbb{E}_{\mathbf{A}' \times \mathbf{B}'} [f(a, b)] = \mathbb{E}_{\mathbf{A}'} [f_1(a)] \cdot \mathbb{E}_{\mathbf{B}'} [f_2(b)], \quad (8.13)$$

isto é, a esperança matemática da classe composta é igual ao produto das esperanças matemáticas das classes componentes se a função f for decomponível.

Demonstração. Pela definição de esperança matemática (Definição 8.1) temos

$$\mathbb{E}_{\mathbf{A}'}[f_1(a)] = \sum_{a_i \in \mathbf{A}'} f_1(a_i) \mathbb{P}_{\mathbf{A}}(a_i), \quad \mathbb{E}_{\mathbf{B}'}[f_2(b)] = \sum_{b_j \in \mathbf{B}'} f_2(b_j) \mathbb{P}_{\mathbf{B}}(b_j)$$

e

$$\mathbb{E}_{\mathbf{A}' \times \mathbf{B}'}[f(a, b)] = \sum_{(a_i, b_j) \in \mathbf{A}' \times \mathbf{B}'} f(a_i, b_j) \mathbb{P}_{\mathbf{A} \times \mathbf{B}}(a_i, b_j).$$

Através das hipóteses inerentes ao teorema, em relação à decomposição da função f e à independência dos lançamentos, obtemos

$$\begin{aligned} & \sum_{(a_i, b_j) \in \mathbf{A}' \times \mathbf{B}'} f(a_i, b_j) \mathbb{P}_{\mathbf{A} \times \mathbf{B}}(a_i, b_j) = \\ &= \sum_{a_i \in \mathbf{A}'} \sum_{b_j \in \mathbf{B}'} f_1(a_i) \cdot f_2(b_j) \mathbb{P}_{\mathbf{A}}(a_i) \mathbb{P}_{\mathbf{B}}(b_j) = \\ &= \sum_{a_i \in \mathbf{A}'} f_1(a_i) \mathbb{P}_{\mathbf{A}}(a_i) \sum_{b_j \in \mathbf{B}'} f_2(b_j) \mathbb{P}_{\mathbf{B}}(b_j) \end{aligned}$$

e, consequentemente, teremos

$$\mathbb{E}_{\mathbf{A}' \times \mathbf{B}'}[f(a, b)] = \mathbb{E}_{\mathbf{A}'}[f(a)] \cdot \mathbb{E}_{\mathbf{B}'}[f(b)].$$

□

O próximo resultado refere-se à esperança matemática e valor esperado de uma constante, isto é, à situação onde $f(a_i) = k$ para qualquer elemento a_i da classe $\mathbf{A}'$.

Teorema 8.4 (Esperança matemática e valor médio de uma constante).

A esperança matemática de uma constante, em relação a uma determinada

classe $\mathbf{A}'$, é igual à própria constante multiplicada pela probabilidade da classe $\mathbf{A}'$, isto é,

$$\mathbb{E}_{\mathbf{A}'}(k) = k \cdot \mathbb{P}_{\mathbf{A}}(\mathbf{A}'), \quad (8.14)$$

consequentemente, o valor médio da constante é a própria constante,

$$\mathbb{M}(k) = k. \quad (8.15)$$

Demonstração.

$$\mathbb{E}_{\mathbf{A}'}(k) = \sum_{a_i \in \mathbf{A}'} k \mathbb{P}_{\mathbf{A}}(a_i) = k \sum_{a_i \in \mathbf{A}'} \mathbb{P}_{\mathbf{A}}(a_i) = k \mathbb{P}_{\mathbf{A}}(\mathbf{A}')$$

e

$$\mathbb{M}(k) = \mathbb{E}_{\mathbf{A}}(k) = k \mathbb{P}_{\mathbf{A}}(\mathbf{A}) = k.$$

□

8.2 Esperança matemática e valor médio em Regiões

Pacheco d'Amorim nesta secção estende os conceitos esperança matemática e valor médio para a situação de lançamentos à sorte efectuados em regiões, isto é, para o contexto da probabilidade contínua.

8.2.1 Definição

Consideremos um determinado lançamento de um ponto à sorte, seja o ponto x que pode ser um ponto livre ou um ponto imagem (conforme definição presente no capítulo 6), numa região que contém a região $\mathbf{X}^{(3)}$, sendo $\mathbb{P}_{\mathbf{X}}(x)$

⁽³⁾ Tal como na análise à esperança matemática em classes finitas, na probabilidade contínua a região $\mathbf{X}$ não significará obrigatoriamente a região total possível em relação ao

a sua lei de probabilidade em relação à região $\mathbf{X}$ em cada ponto $x \in \mathbf{X}$. Consideremos ainda uma função φ , definida em $\mathbf{X}$, das coordenadas de x e seja $\mathbf{X}'$ uma região qualquer contida em $\mathbf{X}$. Denomina-se por valor esperado da função φ em relação à região $\mathbf{X}'$ o integral, avaliado na região $\mathbf{X}'$, da imagem da função φ a multiplicar pela respectiva lei de probabilidade.

Definição 8.3 (Esperança matemática em regiões). Representando a **esperança matemática da função φ em relação à região $\mathbf{X}'$** com o símbolo $\mathbb{E}_{\mathbf{X}'}[\varphi(x)]$, esta quantidade é definida por

$$\mathbb{E}_{\mathbf{X}'}[\varphi(x)] = \int_{\mathbf{X}'} \varphi(x) \mathbb{P}_{\mathbf{X}}(x) \, dx. \quad (8.16)$$

De forma análoga à efectuada em classes finitas, Pacheco d'Amorim define o valor médio ou valor provável da função φ na região $\mathbf{X}$ como sendo a esperança matemática da função φ na região $\mathbf{X}$, i.e., corresponde à situação $\mathbf{X}' = \mathbf{X}$ na definição de esperança matemática.

Definição 8.4 (Valor médio em regiões). Representando o **valor médio ou valor provável da função φ na região $\mathbf{X}$** com o símbolo $\mathbb{M}[\varphi(x)]$, esta quantidade é definida por

$$\mathbb{M}[\varphi(x)] = \int_{\mathbf{X}} \varphi(x) \mathbb{P}_{\mathbf{X}}(x) \, dx. \quad (8.17)$$

8.2.2 Propriedades

Tal como nas propriedades da esperança matemática e valor médio em classes, vamos restringir a exposição à esperança matemática. Contudo, uma vez que o valor médio é um caso particular de esperança matemática, estes resultados também poderão ser aplicados para o valor médio quando tal faça sentido.

lançamento à sorte do ponto x . Desta forma, as definições apresentadas pelo autor correspondem à situação de condicionamento, isto é, $\mathbb{E}_{\mathbf{X}'}[\varphi(x)]$ corresponde a $\mathbb{E}_{\mathbf{X}'}[\varphi(x) | x \in \mathbf{X}]$.

Teorema 8.5 (Aditividade da esperança matemática em relação a regiões disjuntas). *Sejam $\mathbf{X}_1, \dots, \mathbf{X}_n$ n regiões que definem uma partição de $\mathbf{X}'$ (isto é $\mathbf{X}' = \mathbf{X}_1 \cup \mathbf{X}_2 \cup \dots \cup \mathbf{X}_n$ e $\mathbf{X}_i \cap \mathbf{X}_j = \emptyset$ para $i \neq j$), então*

$$\mathbb{E}_{\mathbf{X}'} [\varphi(x)] = \mathbb{E}_{\mathbf{X}_1} [\varphi(x)] + \mathbb{E}_{\mathbf{X}_2} [\varphi(x)] + \dots + \mathbb{E}_{\mathbf{X}_n} [\varphi(x)]. \quad (8.18)$$

Demonstração. Uma vez que as regiões $\mathbf{X}_1, \dots, \mathbf{X}_n$ definem uma partição de $\mathbf{X}'$ teremos

$$\begin{aligned} \mathbb{E}_{\mathbf{X}'} [\varphi(x)] &= \int_{\mathbf{X}'} \varphi(x) \mathbb{P}_{\mathbf{x}}(x) \, dx = \\ &= \int_{\mathbf{X}_1} \varphi(x) \mathbb{P}_{\mathbf{x}}(x) \, dx + \int_{\mathbf{X}_2} \varphi(x) \mathbb{P}_{\mathbf{x}}(x) \, dx + \dots + \int_{\mathbf{X}_n} \varphi(x) \mathbb{P}_{\mathbf{x}}(x) \, dx = \\ &= \sum_{i=1}^n \mathbb{E}_{\mathbf{X}_i} [\varphi(x)]. \end{aligned}$$

□

Teorema 8.6 (Aditividade da esperança matemática em relação à função). *A esperança matemática da soma é igual à soma das esperanças matemáticas, isto é, se*

$$\varphi(x) = \sum_{i=1}^n \varphi_i(x), \quad \forall x \in \mathbf{X}, \quad (8.19)$$

então

$$\mathbb{E}_{\mathbf{X}'} [\varphi(x)] = \sum_{i=1}^n \mathbb{E}_{\mathbf{X}'} [\varphi_i(x)]. \quad (8.20)$$

Demonstração.

$$\begin{aligned} \mathbb{E}_{\mathbf{X}'} [\varphi(x)] &= \mathbb{E}_{\mathbf{X}'} \left[\sum_{i=1}^n \varphi_i(x) \right] = \int_{\mathbf{X}'} \left(\sum_{i=1}^n \varphi_i(x) \right) \mathbb{P}_{\mathbf{x}}(x) \, dx = \\ &= \sum_{i=1}^n \int_{\mathbf{X}'} \varphi_i(x) \mathbb{P}_{\mathbf{x}}(x) \, dx = \sum_{i=1}^n \mathbb{E}_{\mathbf{X}'} [\varphi_i(x)]. \end{aligned}$$

□

Tal como nas tiragens em classes compostas, nos lançamentos em regiões compostas $\mathbf{X} \times \mathbf{Y}$ (o que corresponde a ter dois lançamentos independentes, um na região $\mathbf{X}$ e outro na região $\mathbf{Y}$) podemos, nas situações em que a função φ verifique $\varphi(x, y) = \varphi_1(x) \cdot \varphi_2(y)$, decompor a esperança matemática da função φ em relação à região composta $\mathbf{X} \times \mathbf{Y}$ no produto de duas esperanças matemáticas, cada uma dependente de apenas um dos lançamentos.

Teorema 8.7 (Multiplicação da esperança matemática em relação a lançamentos e funções independentes). *Seja $\mathbf{X}$ uma região possível em relação ao lançamento do ponto x e $\mathbf{Y}$ uma região possível em relação ao lançamento do ponto y . Consideremos uma região composta $\mathbf{X}' \times \mathbf{Y}' \subset \mathbf{X} \times \mathbf{Y}$, com $\mathbf{X}' \subset \mathbf{X}$ e $\mathbf{Y}' \subset \mathbf{Y}$, e uma função φ tal que $\varphi(x, y) = \varphi_1(x) \cdot \varphi_2(y)$, então*

$$\mathbb{E}_{\mathbf{X}' \times \mathbf{Y}'} [\varphi(x, y)] = \mathbb{E}_{\mathbf{X}'} [\varphi_1(x)] \cdot \mathbb{E}_{\mathbf{Y}'} [\varphi_2(y)]. \quad (8.21)$$

Demonstração.

$$\begin{aligned} \mathbb{E}_{\mathbf{X}' \times \mathbf{Y}'} [\varphi(x, y)] &= \int_{\mathbf{X}' \times \mathbf{Y}'} \varphi(x, y) \mathbb{P}_{\mathbf{X} \times \mathbf{Y}}(x, y) \, d(x, y) = \\ &= \int_{\mathbf{Y}'} \int_{\mathbf{X}'} \varphi_1(x) \varphi_2(y) \mathbb{P}_{\mathbf{X}}(x) \mathbb{P}_{\mathbf{Y}}(y) \, dx \, dy \\ &= \int_{\mathbf{X}'} \varphi_1(x) \mathbb{P}_{\mathbf{X}}(x) \, dx \int_{\mathbf{Y}'} \varphi_2(y) \mathbb{P}_{\mathbf{Y}}(y) \, dy \\ &= \mathbb{E}_{\mathbf{X}'} [\varphi_1(x)] \cdot \mathbb{E}_{\mathbf{Y}'} [\varphi_2(y)]. \end{aligned}$$

□

Teorema 8.8 (Existência de valor esperado de uma transformação). *Seja x um ponto variando numa certa região que contém $\mathbf{X}$, f uma qualquer função⁽⁴⁾ das suas coordenadas, e $\mathbb{P}_{\mathbf{X}}(x)$ a sua lei de probabilidade relativamente*

⁽⁴⁾ Pacheco d'Amorim define desta forma a função f , não lhe impondo qualquer restrição. Contudo, para os seus resultados serem válidos, esta função deverá obedecer a

a $\mathbf{X}$. Aplicando a transformação $z = f(x)$ obtemos

$$\mathbb{E}_{\mathbf{x}} [f(x)] = \mathbb{E}_{\mathbf{z}}(z), \quad (8.22)$$

onde $\mathbf{Z} = f(\mathbf{X})$.

Demonstração. Consideremos um ponto $\mathbf{x}$ definido numa região $\mathbf{X} \subset \mathbb{R}^n$ com lei de probabilidade $\mathbb{P}_{\mathbf{x}}(x)$. Pela definição de esperança matemática (Definição 8.3) teremos

$$\mathbb{E}_{\mathbf{x}'} [f(x)] = \int_{\mathbf{X}'} f(x) \mathbb{P}_{\mathbf{x}}(x) \, dx,$$

para $\mathbf{X}' \subset \mathbf{X}$. Representando os valores da função $f : \mathbb{R}^n \rightarrow \mathbb{R}$ pela variável z (por hipótese do teorema teremos $z = f(x)$) e por $\mathbf{Z}' \subset \mathbb{R}$ a região definida pelos valores que z assume quando o ponto x varia no intervalo $\mathbf{X}'$ (projecção em $\mathbb{R}$ de $f(\mathbf{X}')$), obtemos então

$$\mathbb{E}_{\mathbf{x}'} [f(x)] = \int_{\mathbf{X}'} z \mathbb{P}_{\mathbf{x}}(x) \, dx = \int_{\mathbf{Z}'} z \int_{\mathbf{X}''} \mathbb{P}_{\mathbf{x}}(x) \, dx$$

sendo o segundo integral, em $\mathbf{X}'' \subset \mathbb{R}^{n-1}$, estendido à região de $\mathbf{X}'$ em que z toma valores no intervalo $(z, z + dz)$. Pela definição da lei de probabilidade $\mathbb{P}_{\mathbf{x}}(x)$, este integral corresponde à probabilidade de que z pertença ao intervalo $(z, z + dz)$, podendo o integral ser representado por $\mathbb{P}_{\mathbf{z}}(z)dz$, onde $\mathbb{P}_{\mathbf{z}}(z)$ é a lei da probabilidade de z . Assim teremos

$$\mathbb{E}_{\mathbf{x}'} [f(\mathbf{x})] = \int_{\mathbf{Z}'} z \mathbb{P}_{\mathbf{z}}(z) \, dz = \mathbb{E}_{\mathbf{z}'}(z).$$

determinadas condições de regularidade. Por exemplo, se a função f for $f : \mathbb{R} \rightarrow \mathbb{R}$, deverá satisfazer as condições que estabelecemos no capítulo dedicado à análise do conceito ponto imagem, isto é, que $z = f(x)$ estabeleça uma correspondência biunívoca e seja diferenciável, de forma a garantir a existência de f^{-1} e da função derivada de f .

Considerando o caso particular $\mathbf{X}' = \mathbf{X}$, teremos identicamente para o valor médio

$$\mathbb{M}_{\mathbf{x}} [f(x)] = \mathbb{E}_{\mathbf{x}} [f(x)] = \int_{\mathbf{z}} z \mathbb{P}_{\mathbf{z}}(z) \, dz = \mathbb{E}_{\mathbf{z}}(z) = \mathbb{M}_{\mathbf{z}} [z].$$

□

Tentemos decifrar o que Pacheco d'Amorim faz neste teorema e sua demonstração. Começemos por analisar a situação mais simples de termos apenas uma função $f : \mathbb{R} \rightarrow \mathbb{R}$. Notemos que tal resultado, neste caso particular, corresponde a aplicar a transformação $z = f(x)$ na expressão do valor esperado de $f(x)$ em relação a $\mathbf{X}'$, pois

$$\begin{aligned} \mathbb{E}_{\mathbf{x}'} [f(x)] &= \int_{\mathbf{x}'} f(x) \mathbb{P}_{\mathbf{x}}(x) \, dx = \\ &= \int_{f(\mathbf{x}')} z \mathbb{P}_{\mathbf{x}}(f^{-1}(z)) \left| \frac{df^{-1}(z)}{dz} \right| \, dz = \\ &= \int_{\mathbf{z}'} z \mathbb{P}_{\mathbf{z}}(z) \, dz = \mathbb{E}_{\mathbf{z}'} [z] \end{aligned} \quad (8.23)$$

onde utilizamos

$$\mathbb{P}_{\mathbf{z}}(z) = \mathbb{P}_{\mathbf{x}}(f^{-1}(z)) \left| \frac{d}{dz} f^{-1}(z) \right| = \frac{\mathbb{P}_{\mathbf{x}}(f^{-1}(z))}{|f'(f^{-1}(z))|}. \quad (8.24)$$

Desta forma Pacheco d'Amorim demonstra que podemos definir uma nova variável $z = f(x)$, como uma transformação de um lançamento x em $\mathbf{X}$, e de igual forma determinar o valor esperado desta nova variável através da definição. Notemos que, neste caso, não é obrigatório que o ponto x seja um ponto directamente lançado na região $\mathbf{X}$, caracterizado pela equipossibilidade, tal como acontecia no capítulo PONTO IMAGEM, pois o ponto x pode ele próprio ser já um ponto imagem caracterizado pela lei de probabilidade $\mathbb{P}_{\mathbf{x}}(x)$. Assim, podemos transformar o ponto x noutro ponto z através da

aplicação de uma função f e determinar o valor médio da função f através da lei de probabilidade da nova variável z . Parece-nos, todavia, que a ideia central, considerada por Pacheco d'Amorim neste teorema, é que a probabilidade é invariante para transformações, neste caso, de $\mathbb{R}^n$ em $\mathbb{R}$, isto é, se estivermos a analisar um ponto $\mathbf{x}$ variando na região $\mathbf{X} \subset \mathbb{R}^n$, caracterizado por uma determinada lei de probabilidade $\mathbb{P}_{\mathbf{x}}(x_1, \dots, x_n)$, e definirmos uma transformação $z = f(\mathbf{x})$, onde $\mathbf{x} = (x_1, \dots, x_n)$, obtemos um ponto z variando na região $\mathbf{Z} = f(\mathbf{X}) \subset \mathbb{R}$ e consequentemente teremos, para qualquer região $\mathbf{Z}' \subset \mathbf{Z}$, a igualdade

$$\mathbb{P}_{\mathbf{z}}(\mathbf{Z}') = \mathbb{P}_{\mathbf{x}}(\mathbf{X}'), \quad (8.25)$$

onde a região $\mathbf{X}' \subset \mathbb{R}^n$ é definida por

$$\mathbf{X}' = \{(x_1, \dots, x_n) \in \mathbf{X} : z = f(x_1, \dots, x_n) \in \mathbf{Z}'\}. \quad (8.26)$$

Por este motivo a lei de probabilidade de z será uma função $\mathbb{P}_{\mathbf{z}}(z)$, onde $\mathbf{Z}$ é o intervalo de variação do ponto z , tal que se verifica

$$\int_{\mathbf{X}'} \mathbb{P}_{\mathbf{x}}(x_1, \dots, x_n) \, d(x_1, \dots, x_n) = \int_{\mathbf{Z}'} \mathbb{P}_{\mathbf{z}}(z) \, dz \quad (8.27)$$

para qualquer região $\mathbf{Z}' \subset \mathbf{Z}$. Deste modo, uma vez que $f(\mathbf{x}) = z$, obtemos

$$\int_{\mathbf{X}'} f(\mathbf{x}) \mathbb{P}_{\mathbf{x}}(\mathbf{x}) \, d\mathbf{x} = \int_{\mathbf{Z}'} z \mathbb{P}_{\mathbf{z}}(z) \, dz, \quad (8.28)$$

isto é,

$$\mathbb{E}_{\mathbf{x}'}[f(\mathbf{x})] = \mathbb{E}_{\mathbf{z}'}[z]. \quad (8.29)$$

Um segunda forma de visualizarmos este resultado é aplicarmos uma substituição $g : \mathbf{X} \subset \mathbb{R}^n \rightarrow \mathbf{Y} \subset \mathbb{R}^n$, onde consideramos $y_i = g_i(\mathbf{x})$ verificando $y_1 = g_1(\mathbf{x}) = f(\mathbf{x}) = z$ e $g_i(\mathbf{x}) = x_i$ para $i = 2, \dots, n$, isto é, a variável y_1 corresponde à variável $z = f(\mathbf{x})$ que pretendemos analisar e as restantes

variáveis y_i , que correspondem a x_i , servem unicamente para obtermos a matriz jacobiana da transformação. Assim, representando por $\mathcal{J}(\mathbf{y})$ (conforme definição apresentada na página 363) a matriz jacobiana da transformação, teremos

$$\begin{aligned}
 \mathbb{E}_{\mathbf{x}'} [f(\mathbf{x})] &= \int_{\mathbf{x}'} f(\mathbf{x}) \mathbb{P}_{\mathbf{x}}(\mathbf{x}) \, d\mathbf{x} = \\
 &= \int_{\mathbf{x}'} f(x_1, \dots, x_n) \mathbb{P}_{\mathbf{x}}(x_1, x_2, \dots, x_n) \, d(x_1, x_2, \dots, x_n) = \\
 &= \int_{\mathbf{y}'} y_1 \mathbb{P}_{\mathbf{x}}(g^{-1}(x_1), \dots, g^{-1}(x_n)) |\mathcal{J}(y_1, \dots, y_n)| \, d(y_1, \dots, y_n) = \\
 &= \int_{\mathbf{y}'} y_1 \mathbb{P}_{\mathbf{y}}(y) \, d(y_1, y_2, \dots, y_n) \tag{8.30}
 \end{aligned}$$

onde $\mathbf{Y}' = f(\mathbf{X}')$. Pelo Teorema de Fubinni podemos escolher a ordem de integração, logo, a expressão anterior, representando agora a variável y_1 por z , como previamente referimos, é igual a

$$\int_{\mathbf{z}'} z \left(\int_{\mathbf{Y}'_z} \mathbb{P}_{\mathbf{y}}(\mathbf{y}) \, d(y_2, y_3, \dots, y_n) \right) dz = \int_{\mathbf{z}'} z \mathbb{P}_{\mathbf{z}}(z) \, dz = \mathbb{E}_{\mathbf{z}'} [z], \tag{8.31}$$

onde $\mathbf{Y}'_z$ corresponderá à região de variação das variáveis $(y_2, y_3, \dots, y_n)$ para cada valor da variável z .

Notemos, no entanto, que, em qualquer das abordagens apresentadas, estamos a obter a lei de probabilidade do ponto z através do Teorema da Probabilidade Total. Considerando, por exemplo, a expressão (8.31), temos que a probabilidade de z pertencer a um dado intervalo $[z_0, z_0 + dz)$ é dado por

$$\mathbb{P}_{\mathbf{z}}(z \in [z_0, z_0 + dz)) \approx \mathbb{P}_{\mathbf{z}}(z_0) dz = \int_{\mathbf{Y}'_{z_0}} \mathbb{P}_{\mathbf{y}}(\mathbf{y}) \, d\mathbf{y} \tag{8.32}$$

onde a região $\mathbf{Y}'_{z_0}$ é definida pelos pontos $\mathbf{x} \in \mathbf{X}$ tais que $z \in [z_0, z_0 + dz)$. Desta forma, utilizando a notação de lei *a priori* e lei *a posteriori* introduzida

no capítulo PONTO IMAGEM, estamos a aplicar a ideia

$$\mathbb{P}_{\mathbf{Z}}(z_0) = \int_{\mathbf{Y}'_{z_0}} p(z_0|\mathbf{y})a_{\mathbf{y}}(\mathbf{y}) \, d\mathbf{y}, \quad (8.33)$$

onde $\mathbf{Y}'_{z_0} = \{\mathbf{y} : z = z_0\}$, isto é, estamos a projectar a probabilidade, definida pela lei de probabilidade conjunta, no eixo Z . Para tal recorremos, mesmo que possa passar despercebido, a leis de probabilidade condicionadas a acontecimentos de medida nula. Mas, para que possamos recorrer a estas propriedades sem correr o risco de obter resultados divergentes, como diversos paradoxos ilustram, tal como o de Borel analisado na secção 4.6, determinadas condições de regularidade têm que ser verificadas⁽⁵⁾.

Pacheco d'Amorim refere ainda que este teorema é importante e que, na resolução do Problema 8.1, problema do lançamento da curva flexível apresentado na página 486, se tornar evidente esta utilidade, pois, para determinar uma probabilidade referente à distância entre os extremos de dois segmentos adjacentes lançados à sorte, onde necessita da lei de probabilidade da distância que é desconhecida, recorre à lei de probabilidade do ângulo formado pelos dois segmentos, que é fácil de determinar, uma vez que todos os valores têm igual possibilidade de ocorrer (distribuição uniforme na região $(0, 2\pi]$).

Teorema 8.9. *Seja x um ponto variando numa certa região que contém $\mathbf{X}$, f uma qualquer função das suas coordenadas, e $\mathbb{P}_{\mathbf{X}}(x)$ a sua lei de probabilidade relativamente a $\mathbf{X}$. Consideremos a variável z com lei de probabilidade $\mathbb{P}_{\mathbf{Z}}(z)$ e seja φ uma função qualquer. Nestas condições, a esperança matemática da*

⁽⁵⁾ Consultar, por exemplo, Rao (1988), que apresenta de forma rigorosa o problema, referindo que nem as axiomáticas de Kolmogoroff (1933) e de Rényi (1955) resolvem este problema de recorrermos a probabilidades condicionadas a conjuntos de medida nula no contexto da probabilidade contínua, quando consideramos uma variável a assumir um valor particular.

função φ , em relação à região $\mathbf{X}'$, é dada por

$$\mathbb{E}_{\mathbf{X}'}[f(x)] = \mathbb{E}_{\mathbf{Z}'}[\varphi(z)] = \int_{\mathbf{Z}'} \varphi(z) \mathbb{P}_{\mathbf{Z}}(z) \, dz, \quad (8.34)$$

onde $\mathbf{Z}' = \varphi^{-1}[f(\mathbf{X}')]$.

Demonstração. Pacheco d'Amorim refere que basta fazer $f(x) = \varphi(z)$ no teorema antecedente, e tomar ainda o segundo integral estendido à região em que z toma valores no intervalo $(z, z + dz)$, de forma a obtermos $\mathbb{P}_{\mathbf{Z}}(z)dz$. De facto, este resultado é semelhante ao anterior e pode ser obtido de qualquer das formas que deduzimos o teorema precedente, sendo suficiente utilizar $f(x) = \varphi(z)$, isto é $z = \varphi^{-1}[f(x)]$, nas substituições usadas. $\square$

O próximo resultado refere-se à esperança matemática e valor esperado de uma função constante.

Teorema 8.10 (Esperança matemática e valor médio de uma constante). *A esperança matemática de uma constante é igual à própria constante multiplicada pela probabilidade da região, isto é*

$$\mathbb{E}_{\mathbf{X}'}(k) = k \mathbb{P}_{\mathbf{X}}(\mathbf{X}'). \quad (8.35)$$

Consequentemente, o valor médio (caso particular da esperança em que se verifica $\mathbf{X} = \mathbf{X}'$) da constante é a própria constante,

$$\mathbb{M}(k) = k. \quad (8.36)$$

Demonstração.

$$\mathbb{E}_{\mathbf{X}'}(k) = \int_{\mathbf{X}'} k \mathbb{P}_{\mathbf{X}}(x) \, dx = k \int_{\mathbf{X}'} \mathbb{P}_{\mathbf{X}}(x) \, dx = k \mathbb{P}_{\mathbf{X}}(\mathbf{X}')$$

e

$$\mathbb{M}(k) = \mathbb{E}_{\mathbf{X}}(k) = k \mathbb{P}_{\mathbf{X}}(\mathbf{X}) = k.$$

$\square$

8.3 Desigualdade

Depois de apresentar as principais propriedades da esperança matemática e do valor médio, Pacheco d'Amorim deduz uma desigualdade muito semelhante à actualmente apelidada desigualdade de Markov, que de seguida apresentamos.

Teorema 8.11. *Seja f uma função não negativa cujo valor médio se pode tornar inferior a qualquer número positivo δ , por mais pequeno que δ seja. Nestas condições, a probabilidade de que f se mantenha superior a certo número ε é menor do que $\frac{\delta}{\varepsilon}$, isto é, definindo a região $\mathbf{X}' \subset \mathbf{X}$ através de*

$$\mathbf{X}' = \{x \in \mathbf{X} : f(x) \geq \varepsilon\} \quad (8.37)$$

então, se $\mathbb{E}_{\mathbf{x}} [f(x)] \leq \delta$, teremos

$$\mathbb{P}_{\mathbf{x}} (\mathbf{X}') \leq \frac{\delta}{\varepsilon}. \quad (8.38)$$

Como tal, uma vez que esta desigualdade é válida para qualquer δ positivo, a probabilidade $\mathbb{P}_{\mathbf{x}} (\mathbf{X}')$ será nula.

Demonstração. Consideremos uma função f tal que $f(x) \geq 0, \forall x \in \mathbf{X}$ e $\mathbb{E}_{\mathbf{x}} [f(x)] \leq \delta$, para qualquer $\delta > 0$. Seja $\mathbf{X}' \subset \mathbf{X}$ a região definida por $\mathbf{X}' = \{x \in \mathbf{X} : f(x) \geq \varepsilon\}$, então

$$\begin{aligned} \mathbb{E}_{\mathbf{x}} [f(x)] &= \int_{\mathbf{X}} f(x) \mathbb{P}_{\mathbf{x}}(x) \, dx = \\ &= \int_{\mathbf{X}-\mathbf{X}'} f(x) \mathbb{P}_{\mathbf{x}}(x) \, dx + \int_{\mathbf{X}'} f(x) \mathbb{P}_{\mathbf{x}}(x) \, dx \geq \\ &\geq \int_{\mathbf{X}'} f(x) \mathbb{P}_{\mathbf{x}}(x) \, dx \geq \varepsilon \mathbb{P}_{\mathbf{x}} (\mathbf{X}'), \end{aligned}$$

logo

$$\varepsilon \mathbb{P}_{\mathbf{x}} (\mathbf{X}') \leq \mathbb{E}_{\mathbf{x}} [f(x)] \leq \delta \Leftrightarrow \mathbb{P}_{\mathbf{x}} (\mathbf{X}') \leq \frac{\delta}{\varepsilon}.$$

Uma vez que a desigualdade é válida para qualquer valor de δ , por mais pequeno que seja, então podemos tornar a probabilidade $\mathbb{P}_{\mathbf{x}}(\mathbf{X}')$ tão próxima de zero quanto pretendamos, isto é, a probabilidade $\mathbb{P}_{\mathbf{x}}(\mathbf{X}')$ será nula. $\square$

Notemos que a desigualdade deduzida por Pacheco d'Amorim parece um corolário da desigualdade de Markov. Uma das formas habituais de apresentar a desigualdade de Markov é considerar uma função ϕ , tal que $\phi(x) \geq 0$ para qualquer $x \in \mathbf{X}$ e $\mathbb{E}_{\mathbf{x}}[\phi(x)] < \infty$, concluindo-se que sob estas hipóteses se verifica

$$\mathbb{P}_{\mathbf{x}}(\phi(x) \geq \varepsilon) \leq \frac{\mathbb{E}_{\mathbf{x}}[\phi(x)]}{\varepsilon}, \quad \forall \varepsilon > 0. \quad (8.39)$$

No capítulo anterior deduzimos que a famosa desigualdade de Bienaymé-Chebycheff é um caso particular desta, correspondente a considerar $\phi(X) = (X - \mathbb{E}(X))^2$ (consultar página 420). Para obter a desigualdade obtida por Pacheco d'Amorim da desigualdade de Markov basta considerar que $\mathbb{E}_{\mathbf{x}}[\phi(x)]$ assume um valor menor que δ , para qualquer δ positivo, e, consequentemente, a probabilidade $\mathbb{P}_{\mathbf{x}}(\phi(x) \geq \varepsilon)$ terá que ser nula, pois

$$\mathbb{P}_{\mathbf{x}}(\phi(x) \geq \varepsilon) \leq \frac{\mathbb{E}_{\mathbf{x}}[\phi(x)]}{\varepsilon} \leq \frac{\delta}{\varepsilon} \rightarrow 0. \quad (8.40)$$

Não se entende a razão pela qual Pacheco d'Amorim não utilizou o resultado na forma (8.39), cuja demonstração está inserida na dedução da sua desigualdade, ou o seu caso particular correspondente à desigualdade de Bienaymé-Chebycheff, pois esta desigualdade é suficiente para demonstrar, de forma bastante simples, a Lei Fraca dos Grandes Números para o caso geral, supondo a existência do segundo momento. Seja $X_1, X_2, \dots, X_n, \dots$ uma sucessão de variáveis aleatórias independentes e identicamente distribuídas com valor esperado $\mathbb{E}(X_i) = \mu < \infty$ e variância $\mathbb{V}ar(X_i) = \sigma^2$, então, se representarmos a média destas variáveis através de $\overline{X}_n = \frac{1}{n}(X_1 + \dots + X_n)$, obtemos

$$\lim_{n \rightarrow +\infty} \mathbb{P}(|\overline{X}_n - \mathbb{E}(X)| \geq k) \leq \lim_{n \rightarrow +\infty} \frac{\sigma^2}{nk^2} = 0, \quad (8.41)$$

uma vez que $\mathbb{E}(\overline{X}_n) = \mu$ e $\mathbb{V}ar(\overline{X}_n) = \frac{\sigma^2}{n}$. Assim se verifica que, de facto, Pacheco d'Amorim poderia ter demonstrado a Lei Fraca dos Grandes Números de uma forma bastante elementar. Refira-se, todavia, que, apesar de $\mathbb{E}(\overline{X}_n) = \mu$ ser imediato pela aditividade do valor esperado, propriedade que Pacheco d'Amorim também deduziu, as propriedades da variância através das quais se deduz $\mathbb{V}ar(\overline{X}_n) = \frac{\sigma^2}{n}$ não estariam disponíveis, uma vez que nem o conceito de variância era ainda utilizado⁽⁶⁾. Desta forma, a dedução de $\mathbb{V}ar(\overline{X}_n) = \frac{\sigma^2}{n}$, um exercício primário actualmente, não seria na época tão trivial, como somos levados a crer pela facilidade da sua dedução.

Pacheco d'Amorim, apesar de não recorrer a esta desigualdade, analisa posteriormente o caso geral da Lei dos Grandes Números, não restrito às provas de Bernoulli, tal como no quinto capítulo da sua tese, situação que comentaremos na secção 8.6. Em relação ao objectivo do autor, quando deduziu esta desigualdade, este não é claro uma vez que esta desigualdade não é utilizada em mais nenhum momento da sua tese de doutoramento.

8.4 Problema da curva flexível e inextensível

Pacheco d'Amorim, com a resolução deste problema, afirma que o valor provável do quadrado da distância que separa os pontos extremos de uma curva flexível e inextensível, lançada à sorte num plano, é nulo. Ora, uma vez que o quadrado da distância que separa os pontos extremos (que representaremos por d^2) só assume valores não negativos, aplicando a desigualdade deduzida por Pacheco d'Amorim (Teorema 8.11 da página 483), podemos concluir que,

⁽⁶⁾ Segundo David e Edwards (2001) o termo variância só surge em 1918 nos trabalhos de Ronald Fisher.

sendo ε um valor qualquer positivo, então teremos

$$\mathbb{P}(d^2 \geq \varepsilon) \leq \frac{\delta}{\varepsilon} \quad (8.42)$$

onde δ pode assumir qualquer valor positivo, visto que o valor esperado de d^2 é nulo. Desta forma concluímos que podemos obter um valor tão baixo quanto queiramos e, conseqüentemente, esta probabilidade é nula.

Começemos, então, por analisar o problema da curva flexível e inextensível de onde o autor deduziu esta surpreendente conclusão.

Problema 8.1. Consideremos um polígono articulado aberto, de lados $l_1, l_2, \dots, l_n$, com comprimentos respectivamente $\overline{l}_1, \overline{l}_2, \dots, \overline{l}_n$ e sejam a e b os seus pontos extremos. Lança-se, à sorte, esse polígono sobre um plano. Pretendemos determinar o valor médio $\mathbb{M}_n(d^2)$ onde d representa a distância entre a e b .

SOLUÇÃO: Pacheco d'Amorim começa por referir que, para calcular o valor médio do quadrado da distância entre a e b , i.e. $\mathbb{M}_n(d^2)$, podemos (Teorema 8.9, página 481) utilizar a lei da probabilidade de d . Além disso, a lei da probabilidade de d pode ser substituída por uma lei de probabilidade referente a qualquer ponto x que esteja convenientemente ligado a d (Teorema 8.8, página 476).

Após estas referências o autor demonstra o resultado por indução. Começemos por considerar o caso de o polígono ter um único lado l_1 . Nesta situação teremos o valor médio de uma constante e, conseqüentemente, pelo Teorema 8.10 presente na página 482, obtemos

$$\mathbb{M}_1(d^2) = \mathbb{M}_1(\overline{l}_1^2) = \overline{l}_1^2. \quad (8.43)$$

No caso em que o polígono possui dois lados teremos, pelo Teorema 8.9,

$$\begin{aligned}\mathbb{M}_2(d^2) &= \int_0^{\overline{l}_1 + \overline{l}_2} \mathbb{P}(d) \cdot d^2 \, dd = \\ &= \int_0^{2\pi} \frac{(\overline{l}_1^2 + \overline{l}_2^2 - 2\overline{l}_1 \overline{l}_2 \cos \alpha)}{2\pi} \, d\alpha = \overline{l}_1^2 + \overline{l}_2^2,\end{aligned}\quad (8.44)$$

onde α é o ângulo formado pelos lados l_1 e l_2 que tem como lei de probabilidade $\frac{1}{2\pi}$.

Consideremos que no caso de o polígono ser constituído por i lados se verifica

$$\mathbb{M}_i(d^2) = \overline{l}_1^2 + \overline{l}_2^2 + \cdots + \overline{l}_i^2 \quad (8.45)$$

e provemos que nesse caso teremos ainda

$$\mathbb{M}_{i+1}(d^2) = \overline{l}_1^2 + \overline{l}_2^2 + \cdots + \overline{l}_i^2 + \overline{l}_{i+1}^2. \quad (8.46)$$

Seja δ o segmento que une a origem de l_1 com a extremidade de l_i e sejam $\alpha_1, \alpha_2, \dots, \alpha_i$ os ângulos das articulações dos $i+1$ lados (os lados l_j e l_{j+1} formam o ângulo α_j para $j = 1, \dots, i$); então d^2 pode ser descrito por uma função $f(\alpha_1, \alpha_2, \dots, \alpha_i)$ desses ângulos e o valor médio de d^2 terá a forma

$$\begin{aligned}\mathbb{M}_{i+1} &= \int_{[0, 2\pi]^i} f(\alpha_1, \alpha_2, \dots, \alpha_i) \frac{d\alpha_1}{2\pi} \cdot \frac{d\alpha_2}{2\pi} \cdots \frac{d\alpha_i}{2\pi} = \\ &= \int_{[0, 2\pi]^{i-1}} \frac{d\alpha_1}{2\pi} \cdot \frac{d\alpha_2}{2\pi} \cdots \frac{d\alpha_{i-1}}{2\pi} \int_0^{2\pi} f(\alpha_1, \alpha_2, \dots, \alpha_i) \frac{d\alpha_i}{2\pi} = \\ &= \int_{[0, 2\pi]^{i-1}} \frac{d\alpha_1}{2\pi} \cdot \frac{d\alpha_2}{2\pi} \cdots \frac{d\alpha_{i-1}}{2\pi} \int_0^{2\pi} \frac{1}{2\pi} \left(\overline{\delta}^2 + \overline{l}_{i-1}^2 + 2\overline{\delta} \overline{l}_{i+1} \cos \alpha_i \right) d\alpha_i = \\ &= \int_{[0, 2\pi]^{i-1}} \frac{d\alpha_1}{2\pi} \cdot \frac{d\alpha_2}{2\pi} \cdots \frac{d\alpha_{i-1}}{2\pi} \left(\overline{\delta}^2 + \overline{l}_{i+1} \right) = \\ &= \mathbb{M}_i \left(\overline{\delta}^2 + \overline{l}_{i+1} \right) = \mathbb{M}_i(\delta^2) + \mathbb{M}_i(\overline{l}_{i+1}) = \\ &= \overline{l}_1^2 + \overline{l}_2^2 + \cdots + \overline{l}_i^2 + \overline{l}_{i+1}^2.\end{aligned}\quad (8.47)$$

Deste modo Pacheco d'Amorim demonstrou que

$$\mathbb{M}_n(d^2) = \overline{l}_1^2 + \overline{l}_2^2 + \cdots + \overline{l}_n^2. \quad (8.48)$$

◇

Se os lados forem todos iguais e representarmos o perímetro por L , o resultado obtido pode ser representado através de

$$\mathbb{M}_n = n\overline{l}_1^2 = \frac{L^2}{n}. \quad (8.49)$$

Se, além disso, determinarmos o limite de $\mathbb{M}_n$ quando o número de lados aumenta indefinidamente teremos

$$\lim_{n \rightarrow +\infty} \mathbb{M}_n = \lim_{n \rightarrow +\infty} \frac{L^2}{n} = 0, \quad (8.50)$$

resultado este que significa que o valor médio do quadrado da distância que separa os pontos extremos duma curva flexível e inextensível lançada, á sorte, sobre um plano, é nulo, qualquer que seja o comprimento L da curva (desde que este seja finito).

Se recorrermos à desigualdade previamente apresentada por Pacheco d'Amorim (Teorema 8.11) podemos então concluir que

$$\lim_{n \rightarrow +\infty} \mathbb{P}(d^2 \neq 0) = 0, \quad (8.51)$$

isto é, ao lançarmos á sorte uma curva flexível e inextensível num plano, a probabilidade de a curva obtida ser fechada converge para a unidade. Note-mos que, se representarmos a lei de probabilidade da distância d_∞ da curva obtida no limite por $\mathbb{P}(d_\infty)$ temos

$$\mathbb{M}_\infty = \int_{\mathbf{D}} d_\infty^2 \mathbb{P}(d_\infty) \, dd_\infty = 0 \quad (8.52)$$

onde $\mathbf{D}$ representa o intervalo de variação de d_∞ . Tendo em conta que $d_\infty \geq 0$, se definirmos a região $\mathbf{D}_1 \subset \mathbf{D}$ como $\mathbf{D}_1 = \{d_\infty \in \mathbf{D} : d_\infty > 0\}$, teremos

obrigatoriamente

$$\int_{\mathbf{D}_1} \mathbb{P}(d_\infty) \, dd_\infty = \mathbb{P}_{\mathbf{D}}(\mathbf{D}_1) = 0, \quad (8.53)$$

logo, a curva desta forma lançada será fechada, excepto num conjunto de medida (de probabilidade) nula.

Antes de comentarmos o resultado agora apresentado, estranho à primeira vista, analisemos este problema recorrendo a conceitos mais modernos. Assim, consideremos o seguinte problema que pensamos ser idêntico ao resolvido por Pacheco d'Amorim.

Problema 8.2. Consideremos n segmentos $l_1, l_2, \dots, l_n$, com comprimentos respectivamente $\bar{l}_1, \bar{l}_2, \dots, \bar{l}_n$, formando uma linha poligonal com perímetro total $L = \sum_{i=1}^n \bar{l}_i$, onde os ângulos entre o lado l_i e o eixo das abcissas é uma variável aleatória θ_i , com $i = 1, \dots, n$. As variáveis aleatórias θ_i são independentes com distribuição uniforme em $[0, 2\pi]$. Qual o valor esperado (valor médio ou provável na nomenclatura utilizada por Pacheco d'Amorim) do quadrado da distância entre os dois extremos do polígono assim formado e qual o valor do limite quando o número de lados n tende para infinito?

Antes de deduzirmos formalmente a resposta a este problema, com o objectivo de ilustrar o comportamento destes polígonos assim formados, recorreremos à simulação considerando que os lados têm todos o mesmo comprimento ($\bar{l}_i = \frac{L}{n}$ para $i = 1, \dots, n$). Apresentamos na Figura 8.1 os resultados obtidos em quatro simulações utilizando o comprimento total fixo (igual a 1) e o número de lados n assumindo os valores 100, 1000, 10000 e 100000, onde obtivemos como quadrado da distância entre as duas extremidades respectivamente 0.002431, 0.001348, 0.000358 e 0.0000005. A representação gráfica parece mostrar que a distância, ou o quadrado da distância, diminui à medida que aumentamos o número de lados mantendo o comprimento total (perímetro) do polígono fixo. Contudo, uma observação de cada situação,

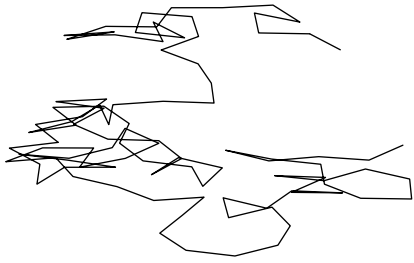
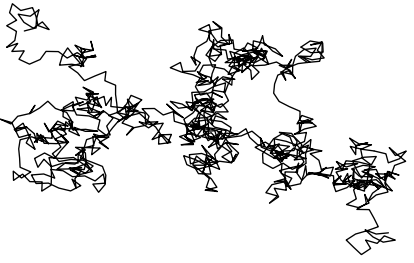
$n = 100$	$n = 1000$
 $d^2 = 0.002431$	 $d^2 = 0.001348$
$n = 10000$	$n = 100000$
 $d^2 = 0.000358$	 $d^2 = 0.0000005$

Figura 8.1: Simulação do Problema curva flexível

não é suficiente para tirarmos qualquer conclusão. Então, para acentuar esta tendência, de a distância entre as duas extremidades diminuir quando aumentamos o número de lados utilizados no polígono (considerando sempre fixo o perímetro do polígono resultante), efectuamos a simulação de 100000 réplicas com o número de lados n igual a 100, 1000, 10000 e 100000. Os resultados desta simulação encontram-se resumidos na Tabela 8.1, onde estão presentes algumas medidas para caracterizar os valores obtidos.

Dos valores presentes na tabela é clara a tendência de a distância diminuir à medida que aumentamos o número de lados do polígono, parecendo existir uma propensão para, cada vez que aumentamos (multiplicando por 10) o número de lados n utilizado, todas as medidas diminuírem, sendo os valores obtidos divididos por 10, nomeadamente a média, o desvio padrão, os quartis

Dimensão n	100	1000	10000	100000
Réplicas r	100000	100000	100000	100000
Média	0.009967	0.001002	0.000100	9.9×10^{-6}
Desvio Padrão	0.010159	0.001001	0.000100	9.9×10^{-6}
Mínimo	4.4×10^{-8}	3.9×10^{-8}	3.9×10^{-11}	9.4×10^{-12}
Quartil 1	0.002880	0.000290	0.000029	2.8×10^{-6}
Quartil 2	0.006927	0.000694	0.000070	6.9×10^{-6}
Quartil 3	0.013710	0.001387	0.000139	0.000014
Máximo	0.157082	0.015443	0.001420	0.000118

Tabela 8.1: Simulação do Problema curva flexível

e os extremos. Analisemos, no entanto, o problema formalmente.

SOLUÇÃO: Consideremos o ponto inicial, uma das extremidades do polígono definida quando lançamos o primeiro segmento no plano, o ponto $A_0 = (x_0, y_0)$ (poderia ser, como usual, a origem, pois o resultado final não depende do ponto considerado como ponto inicial). Estando fixa uma das extremidades do primeiro lado l_1 e formando este lado um ângulo θ_1 com o eixo das abcissas, a outra extremidade do lado l_1 corresponderá ao ponto

$$A_1 = (x_1, y_1) = (x_0 + \bar{l}_1 \cos(\theta_1), y_0 + \bar{l}_1 \sin(\theta_1)). \quad (8.54)$$

O segundo lado l_2 terá uma das extremidades ligadas ao lado l_1 no ponto A_1 e a outra extremidade dependerá do ângulo θ_2 que este lado forma com o eixo das abcissas, sendo a segunda extremidade do lado l_2 o ponto $A_2 = (x_2, y_2)$ dado por

$$A_2 = (x_0 + \bar{l}_1 \cos(\theta_1) + \bar{l}_2 \cos(\theta_2), y_0 + \bar{l}_1 \sin(\theta_1) + \bar{l}_2 \sin(\theta_2)), \quad (8.55)$$

e, de uma forma geral, vamos obter

$$A_k = (x_k, y_k) = \left(x_0 + \sum_{i=1}^k \bar{l}_i \cos(\theta_i), y_0 + \sum_{i=1}^k \bar{l}_i \sin(\theta_i) \right). \quad (8.56)$$

Deste modo a outra extremidade do polígono, obtida após o lançamento dos n lados, corresponderá ao ponto

$$A_n = (x_n, y_n) = \left(x_0 + \sum_{i=1}^n \bar{l}_i \cos(\theta_i), y_0 + \sum_{i=1}^n \bar{l}_i \sin(\theta_i) \right). \quad (8.57)$$

O quadrado da distância entre as duas extremidades é então dado por

$$d^2 = \left(\sum_{i=1}^n \bar{l}_i \cos(\theta_i) \right)^2 + \left(\sum_{i=1}^n \bar{l}_i \sin(\theta_i) \right)^2 \quad (8.58)$$

que, simplificando, obtemos

$$\begin{aligned} d^2 &= \left[\sum_{i=1}^n \bar{l}_i^2 \cos^2(\theta_i) + \sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n \bar{l}_i \bar{l}_j \cos(\theta_i) \cos(\theta_j) + \right. \\ &+ \left. \sum_{i=1}^n \bar{l}_i^2 \sin^2(\theta_i) + \sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n \bar{l}_i \bar{l}_j \sin(\theta_i) \sin(\theta_j) \right] = \\ &= \left[\sum_{i=1}^n \bar{l}_i^2 + \sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n \bar{l}_i \bar{l}_j (\cos(\theta_i) \cos(\theta_j) + \sin(\theta_i) \sin(\theta_j)) \right], \quad (8.59) \end{aligned}$$

uma vez que $\sin^2(\theta_i) + \cos^2(\theta_i) = 1$. Antes de determinarmos o valor esperado desta expressão, notemos que $\mathbb{E}[\cos(\theta_i) \cos(\theta_j)] = 0$ para $i \neq j$, pois

$$\mathbb{E}[\cos(\theta_i)] = \int_0^{2\pi} \frac{\cos(\theta_i)}{2\pi} d\theta_i = 0 \quad (8.60)$$

e, uma vez que as variáveis aleatórias θ_i e θ_j , com $i \neq j$, são independentes teremos ainda

$$\mathbb{E}[\cos(\theta_i) \cos(\theta_j)] = \mathbb{E}[\cos(\theta_i)] \mathbb{E}[\cos(\theta_j)] = 0. \quad (8.61)$$

De forma análoga podemos ainda concluir que

$$\mathbb{E}[\sin(\theta_i) \sin(\theta_j)] = 0. \quad (8.62)$$

Recorrendo aos resultados (8.59), (8.61) e (8.62), podemos deduzir que o valor esperado de d^2 é dado por

$$\begin{aligned}\mathbb{E}[d^2] &= \mathbb{E}\left[\sum_{i=1}^n \bar{l}_i^2 + \sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n \bar{l}_i \bar{l}_j (\cos(\theta_i) \cos(\theta_j) + \sin(\theta_i) \sin(\theta_j))\right] = \\ &= \sum_{i=1}^n \bar{l}_i^2 + \sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n \bar{l}_i \bar{l}_j (\mathbb{E}[\cos(\theta_i) \cos(\theta_j)] + \mathbb{E}[\sin(\theta_i) \sin(\theta_j)]) = \\ &= \sum_{i=1}^n \bar{l}_i^2.\end{aligned}\tag{8.63}$$

O limite, quando aumentamos indefinidamente o valor de n , será

$$\lim_{n \rightarrow +\infty} \mathbb{E}[d^2] = \lim_{n \rightarrow +\infty} \sum_{i=1}^n \bar{l}_i^2\tag{8.64}$$

que é sempre inferior a L^2 , pois o quadrado da soma de quantidades positivas é maior ou igual que a soma dos quadrados dessas quantidades, e assumirá o valor zero se, para qualquer ϵ positivo, tivermos

$$\max_{i=1, \dots, n} \bar{l}_i = o\left(n^{-(0.5+\epsilon)}\right), \quad \text{quando } n \rightarrow \infty,\tag{8.65}$$

pois nesta situação teremos

$$\begin{aligned}\lim_{n \rightarrow +\infty} \mathbb{E}[d^2] &= \lim_{n \rightarrow +\infty} \sum_{i=1}^n \bar{l}_i^2 \leq \lim_{n \rightarrow +\infty} \sum_{i=1}^n \left(n^{-(0.5+\epsilon)}\right)^2 = \\ &= \lim_{n \rightarrow +\infty} \sum_{i=1}^n \frac{1}{n^{1+2\epsilon}} = \lim_{n \rightarrow +\infty} \frac{1}{n^{2\epsilon}} = 0.\end{aligned}\tag{8.66}$$

Isto significa que, se, ao dividirmos os lados do polígono, o fizermos de forma que o comprimento máximo dos lados convirja para zero mais rapidamente que $n^{-(0.5+\epsilon)}$ para qualquer ϵ positivo, então garantimos que o valor esperado do quadrado da distância entre os dois pontos extremos, do polígono lançado à sorte no plano, também convirja para zero.

No caso de considerarmos, tal como Pacheco d'Amorim fez, os lados todos com igual comprimento, isto é $\bar{l}_i = \frac{L}{n}$ para $i = 1, \dots, n$, então facilmente

obtemos

$$\mathbb{E} \left[d^2 \right] = \sum_{i=1}^n \left(\frac{L}{n} \right)^2 = \frac{L^2}{n}, \quad (8.67)$$

e, conseqüentemente,

$$\lim_{n \rightarrow +\infty} \mathbb{E} \left[d^2 \right] = \lim_{n \rightarrow +\infty} \frac{L^2}{n} = 0. \quad (8.68)$$

Recorrendo à desigualdade patente no Teorema 8.11 podemos concluir ainda que

$$\lim_{n \rightarrow +\infty} \mathbb{P} (d^2 \neq 0) = 0, \quad (8.69)$$

e, portanto, temos uma probabilidade nula de obter um polígono aberto neste tipo de lançamento concebido por Pacheco d'Amorim. $\diamond$

Desta forma, podemos concluir que o resultado que Pacheco d'Amorim apresenta está correcto. Contudo, não é o resultado em si que nos surpreende, é a forma como o autor o anuncia, pois parece ter uma generalidade muito superior ao que realmente insere. Notemos que esta conclusão resulta do lançamento à sorte de uma curva flexível e inextensível que, segundo a Definição 5.14 (presente na página 313), se refere a uma curva que se vai *dobrando* sucessivamente, de forma que o seu número de lados aumente indefinidamente e o comprimento de cada lado diminua, mantendo-se constante o comprimento total (perímetro) do polígono. Em primeiro lugar, saliente-se, ao contrário do que Pacheco d'Amorim (1920, p. 4)⁽⁷⁾ afirma, as linhas curvas não podem todas ser obtidas por este processo. Pelo contrário, só um tipo muito específico de curvas que, por exemplo, não é diferenciável em qualquer ponto, é que pode ser obtido por este método. Além disso, na resolução deste problema, Pacheco d'Amorim considerou que os lados do polígono possuem todos o mesmo comprimento, o que não é obrigatório pela definição de lançamento de uma curva flexível e inextensível, logo, não estamos a incluir,

⁽⁷⁾ “A linha curva não passa dum linha quebrada com um número infinitamente grande de lados.”

na sua resolução do problema, todas as situações que verificam a definição de lançamentos de curvas flexíveis e inextensíveis. Há ainda a acrescentar a hipótese, subjacente na resolução, de independência entre os ângulos, que poderá ser factor essencial para o desfecho final e será investigada com mais detalhe na secção 8.4.1. Refira-se que esta hipótese de independência das variáveis θ_i também não nos parece compatível com a ideia de lançar um fio à sorte num plano, pois deve haver alguma dependência entre θ_i , ângulo formado entre o lado l_i e o eixo das abcissas, e o ângulo θ_{i-1} , de forma a suavizar a curva tornando impossível, por exemplo, “*voltar atrás*”.

Assim, consideramos alguma dependência entre estas variáveis θ_i , por exemplo, que a sua distribuição é uniforme no intervalo $[\theta_{i-1} - \frac{\pi}{k}, \theta_{i-1} + \frac{\pi}{k}]$ com $k > 1$. Para facilitar este estudo definimos as variáveis aleatórias θ'_i que correspondem ao ângulo entre o lado l_i e l_{i-1} , que terá distribuição uniforme no intervalo $[-\frac{\pi}{k}, \frac{\pi}{k}]$, onde $k = 1$ representa a independência entre os ângulos θ_i e $k = 4$ significará que, em relação ao lado anterior, cada novo lado lançado só pode fazer um ângulo de, no máximo, 45 graus (sendo todos os ângulos igualmente possíveis neste intervalo de variação). Efectuamos também uma simulação onde consideramos que o ângulo θ'_i , formado pelos lados l_i e l_{i-1} , é determinado pela expressão $(\theta''_i - 0.5) \frac{2\pi}{k}$, onde as variáveis aleatórias θ''_i são caracterizadas por uma distribuição Beta com parâmetros iguais a dois. O objectivo é, tal como na situação de utilização da distribuição uniforme, restringir a variação do ângulo que um lado forma com o lado precedente, pois um segmento em relação ao anterior só poderá formar um ângulo compreendido no intervalo $[-\frac{\pi}{k}, \frac{\pi}{k}]$. Contudo, em vez considerarmos equipossibilidade, ponderamos o centro deste intervalo com um peso maior. Os resultados desta análise serão apresentados na secção 8.4.1.

Outro factor importante nesta dedução é a forma como os comprimentos dos lados convergem para zero. Para concluirmos que o valor esperado do

quadrado da distância converge para zero tivemos de considerar que o comprimento máximo convergia para zero mais rapidamente que $n^{-(0.5+\epsilon)}$ para qualquer ϵ positivo. Na secção 8.4.2 analisaremos esta situação.

Deste modo, para as curvas construídas pelo método descrito no enunciado do Problema 8.2, de facto, o valor esperado do quadrado da distância, entre os dois pontos extremos da curva lançada à sorte, é nulo. Porém, não podemos generalizar este resultado para outras situações, como sugere a leitura do seguinte comentário de Pacheco d'Amorim:

“Os valor provável do quadrado da distância que separa os pontos extremos duma curva flexível, lançada à sorte sobre um plano, é nulo, qualquer que seja o comprimento da curva, logo que seja finito.”

[Pacheco d'Amorim, 1914, p. 147]

Desta forma parece que, se traçarmos uma curva num plano, com um determinado comprimento fixo, de forma aleatória, vamos certamente (excepto num conjunto com medida nula) obter uma curva fechada. Salientemos que, se criarmos outro método para lançarmos a curva de forma aleatória, distinto do apresentado por Pacheco d'Amorim, o resultado poderá ser distinto. Por exemplo, num passeio aleatório de uma partícula num reticulado de dimensão 2, que corresponde ao plano restrito apenas aos pontos definidos em $\mathbb{Z}^2$, onde habitualmente temos um ponto de partida (ponto A_0) e em cada iteração a partícula muda de localização para um dos quatro possíveis pontos vizinhos à sua localização, percorrendo uma determinada distância que é constante (igual a um) num sentido escolhido aleatoriamente em cada iteração, tendo os quatro possíveis sentidos igual probabilidade de ocorrerem. Notemos que este problema é semelhante ao de Pacheco d'Amorim, considerando que os únicos ângulos possíveis são 90, 180, 270 e 360 graus e os lados têm sempre um comprimento fixo. Porém, o perímetro percorrido não é fixo, como

Pacheco d'Amorim impôs no seu problema. Assim, como comparação ao estudo de Pacheco d'Amorim, apresentamos na secção 8.4.3 alguns resultados sobre este passeio aleatório, deduzidos por George Pólya (1887–1985) em 1921. Finalmente na secção 8.4.4 vamos deduzir que, através de um passeio aleatório num reticulado, podemos definir um processo estocástico contínuo, que corresponde ao processo de Wiener (ou movimento browniano), tal como Bachelier (1900) tinha apresentado na sua tese de doutoramento para uma dimensão. Para tal, é suficiente considerar que os passos dados pela partícula (comprimento dos lados) vão convergindo para zero a uma determinada ordem de convergência.

8.4.1 Abandonando a hipótese de independência

Nesta secção vamos considerar o enunciado do Problema 8.2, enunciado na página 489. No entanto, em vez de utilizarmos as variáveis θ_i , que representam os ângulos que os lados l_i formam com o eixo das abcissas, vamos recorrer às variáveis θ'_i , que caracterizam o ângulo formado pelo lado l_i e l_{i-1} . Esta substituição tem como objectivo utilizar variáveis independentes, bastante mais fáceis de manejar, em vez de um processo de Markov, onde a distribuição da variável aleatória θ_i depende do valor assumido pela variável anterior θ_{i-1} . Assim, se as variáveis θ'_i são caracterizadas por uma distribuição uniforme na região $[-\frac{\pi}{k}, \frac{\pi}{k}]$, as variáveis θ_i serão caracterizadas por uma distribuição uniforme na região $[\theta_{i-1} - \frac{\pi}{k}, \theta_{i-1} + \frac{\pi}{k}]$.

Problema 8.3. Consideremos n segmentos $l_1, l_2, \dots, l_n$, de comprimento respectivamente $\bar{l}_1, \bar{l}_2, \dots, \bar{l}_n$, formando uma linha poligonal com perímetro total $L = \sum_{i=1}^n \bar{l}_i$, onde o ângulo entre os lados l_i e l_{i-1} é caracterizado pela variável aleatória θ'_i , com $i = 1, \dots, n$. As variáveis aleatórias θ'_i , com $i = 1, \dots, n-1$, têm distribuição uniforme no intervalo $[-\frac{\pi}{k}, +\frac{\pi}{k}]$, onde consideramos, por conveniência, que o sentido do lado l_0 (lado fictício que ajudará a definir a

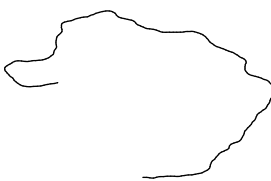
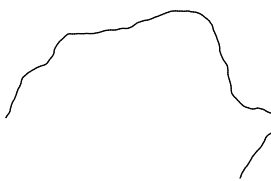
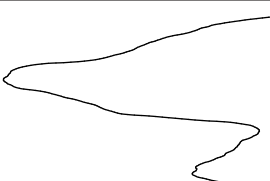
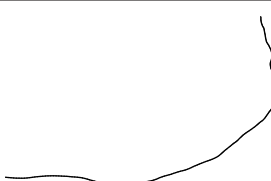
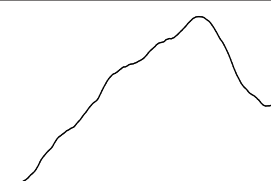
$k = 1$	$k = 2$	$k = 4$
		
$d^2 = 0.000062$	$d^2 = 0.000292$	$d^2 = 0.000794$
$k = 10$	$k = 50$	$k = 75$
		
$d^2 = 0.010393$	$d^2 = 0.069783$	$d^2 = 0.279937$
$k = 100$	$k = 180$	$k = 360$
		
$d^2 = 0.775115$	$d^2 = 0.601732$	$d^2 = 0.980094$

Figura 8.2: Simulação do Problema curva flexível com dependência

variável l_1) é o do eixo das abcissas. Qual será o valor esperado do quadrado da distância entre os dois extremos do polígono assim formado e qual o valor do limite quando o número de lados n tende para infinito? E, se os ângulos θ'_i forem determinados por $\theta'_i = (\theta''_i - 0.5) \frac{2\pi}{k}$ onde θ''_i são variáveis aleatórias independentes caracterizadas por uma distribuição Beta com parâmetros iguais a dois, a situação será diferente?

Antes de deduzirmos formalmente a resposta deste problema, vamos mostrar algumas simulações para diferentes valores de k , onde $k = 1$ significa independência entre as variáveis. Nas simulações representadas na Figura 8.2

foram utilizados $n = 10000$ lados e o sentido do primeiro lado foi escolhido à sorte (distribuição uniforme no intervalo $[0, 2\pi]$).

Valor de k	1	2	4	10
Média	0.000100	0.000448	0.001943	0.012132
Desvio Padrão	0.000100	0.000447	0.001976	0.011968
Mínimo	1.5×10^{-8}	5.5×10^{-8}	1.0×10^{-7}	8.1×10^{-8}
Quartil 1	0.000030	0.000128	0.000547	0.003579
Quartil 2	0.000070	0.000311	0.001358	0.008494
Quartil 3	0.000140	0.000629	0.002700	0.016697
Máximo	0.001016	0.004383	0.021147	0.102610
Valor de k	50	100	180	360
Média	0.254105	0.620028	0.849307	0.959300
Desvio Padrão	0.188072	0.219254	0.115433	0.034886
Mínimo	2.2×10^{-5}	2.7×10^{-8}	0.062850	0.655471
Quartil 1	0.094237	0.487715	0.804503	0.948835
Quartil 2	0.216789	0.673613	0.884487	0.970490
Quartil 3	0.387465	0.793463	0.930352	0.982139
Máximo	0.856584	0.955355	0.987720	0.996593

Tabela 8.2: Simulação do Problema curva flexível com dependência

Nas Tabelas 8.2 e 8.3 estão presentes os resultados obtidos por simulações com $r = 10000$ réplicas. Na Tabela 8.2 foi utilizado um número de lados igual a $n = 10000$ e na Tabela 8.3, com o objectivo de analisar a evolução quando aumentamos o número de lados do polígono, foi utilizado $n = 100000$ lados. Os resultados patentes na Tabela 8.2, uma vez que utilizam igual dimensão para as amostras e igual número de réplicas, demonstram claramente, como seria esperado, que o quadrado da distância varia no mesmo sentido que o

Valor de k	4	10	180	360
Média	0.000192	0.001213	0.317973	0.684803
Desvio Padrão	0.000190	0.001220	0.212791	0.197953
Mínimo	1.9×10^{-8}	4.5×10^{-8}	0.000215	0.000255
Quartil 1	0.000054	0.000350	0.131308	0.582783
Quartil 2	0.000134	0.000825	0.292425	0.739048
Quartil 3	0.000268	0.001668	0.488329	0.836122
Máximo	0.001596	0.010514	0.883033	0.964351

Tabela 8.3: Simulação do Problema curva flexível com dependência

valor de k , isto é, quanto maior (mais estreita) for a restrição da direcção do novo lado lançado maior é o valor esperado do quadrado da distância. Contudo, desta ilação nada podemos concluir sobre a convergência, pois podemos estar perante casos onde a convergência é apenas mais lenta, precisando de um maior número de lados. Assim, foram efectuadas simulações para algumas situações, correspondentes a $k = 4$, $k = 10$, $k = 180$ e $k = 360$, para averiguar se o aumento do número de lados, considerando sempre um perímetro total fixo (igual a uma unidade), provoca a diminuição do valor esperado do quadrado da distância. Os valores obtidos, apresentados na Tabela 8.3, indicam que esta afirmação parece ser verdadeira. Em relação aos valores $k = 4$ e $k = 10$, parece-nos uma vez mais haver uma tendência, em todas as medidas apresentadas na tabela, para os valores obtidos serem divididos por dez quando aumentamos 10 vezes o número de lados utilizados na formação do polígono. Para $k = 180$ e $k = 360$ não se verifica, pelo menos para os valores apresentados, esta tendência. Todavia, os valores obtidos claramente demonstram a propensão de o valor esperado do quadrado da distância entre os dois pontos extremos diminuir à medida que aumentamos o número n de lados.

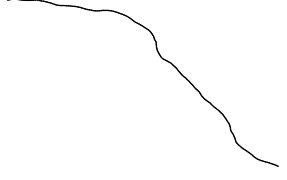
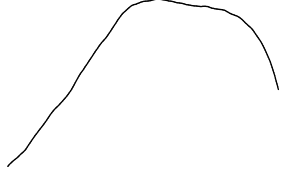
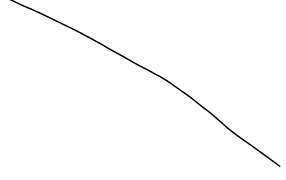
$k = 2$	$k = 4$	$k = 8$
		
$d^2 = 0.008495$	$d^2 = 0.001213$	$d^2 = 0.037510$
$k = 45$	$k = 90$	$k = 360$
		
$d^2 = 0.823240$	$d^2 = 0.783092$	$d^2 = 0.991751$

Figura 8.3: Passeios aleatórios utilizando distribuição Beta

Na Figura 8.3 apresentamos os resultados obtidos em simulações considerando polígonos com $n = 10000$ lados onde os ângulos dos lados foram determinados utilizando uma distribuição Beta com ambos os parâmetros assumindo o valor dois. Desta forma, o ângulo θ'_i entre os lados l_i e l_{i-1} é obtido pela função $(\theta''_i - 0.5) \frac{2\pi}{k}$ onde θ''_i segue uma distribuição Beta com ambos os parâmetros iguais a dois. Deste modo não só limitamos, consoante o valor de k , o campo de variação do ângulo, como tornamos mais prováveis na vizinhança de zero os valores que o ângulo θ'_i entre dois lados consecutivos forma. Uma vez mais os gráficos mostram, como era esperado, que a distância aumenta quando aumentamos o valor de k . Na Tabela 8.4 apresentamos a mesma situação para os ângulos θ'_i onde foram utilizados polígonos com n lados e r réplicas. Os resultados obtidos parecem indicar, de igual forma, que o valor esperado da distância diminui quando aumentamos o número de lados, mantendo o perímetro total fixo. Nem mesmo a utilização de $k = 360$ parece impedir a convergência do quadrado da distância para zero quando

k	2	4	360	360	360
n	10000	10000	10000	100000	1000000
r	10000	10000	10000	1000	100
Média	0.0032287	0.0130383	0.993670	0.939172	0.583819
Des. Padrão	0.0032413	0.0129391	0.005652	0.051486	0.229927
Mínimo	2.3×10^{-8}	2.2×10^{-6}	0.944348	0.625928	0.003998
Quartil 1	0.0009383	0.0038539	0.992055	0.922352	0.427542
Quartil 2	0.0022311	0.0094943	0.995470	0.955052	0.658625
Quartil 3	0.0044156	0.0178831	0.997336	0.975093	0.768295
Máximo	0.0276002	0.121705	0.999642	0.994849	0.918003

Tabela 8.4: Simulação do Problema curva flexível com dependência Beta

aumentamos indefinidamente o número de lados, pois, com o aumento do número de lados, mesmo nesta situação, o quadrado da distância tende a diminuir. Contudo, analisemos formalmente esta questão.

SOLUÇÃO: Vamos efectuar um desenvolvimento análogo ao utilizando na resolução do Problema 8.2, situação de independência das variáveis aleatórias θ_i , mas utilizando agora as variáveis θ'_i que caracterizam os ângulos entre dois lados consecutivos em vez do ângulo entre cada lado e o eixo das abcissas. Assim, representando o ponto inicial, uma das extremidades do polígono definida quando lançamos o primeiro segmento no plano, por $A_0 = (x_0, y_0)$, que corresponderá a uma das extremidades do primeiro lado l_1 , e formando este lado um ângulo θ'_1 com o eixo das abcissas (para a resolução deste problema é indiferente a direcção que consideramos para o primeiro lado), a outra extremidade do lado l_1 corresponderá ao ponto

$$A_1 = (x_1, y_1) = \left(x_0 + \bar{l}_1 \cos(\theta'_1), y_0 + \bar{l}_1 \sin(\theta'_1) \right). \quad (8.70)$$

O segundo lado l_2 terá uma das extremidades ligadas ao lado l_1 no ponto

A_1 e a outra extremidade dependerá do ângulo θ'_i que este lado forma com o lado l_1 , sendo a segunda extremidade do lado l_2 o ponto A_2 com coordenadas

$$\begin{aligned} A_2 &= (x_2, y_2) = (x_1, y_1) + \left(\bar{l}_2 \cos(\theta'_1 + \theta'_2), \bar{l}_2 \sin(\theta'_1 + \theta'_2) \right) = \\ &= \left(x_0 + \sum_{i=1}^2 \bar{l}_i \cos\left(\sum_{j=1}^i \theta'_j\right), y_0 + \sum_{i=1}^2 \bar{l}_i \sin\left(\sum_{j=1}^i \theta'_j\right) \right), \end{aligned} \quad (8.71)$$

e, de uma forma geral, vamos obter para as coordenadas do ponto A_k

$$(x_k, y_k) = \left(x_0 + \sum_{i=1}^k \bar{l}_i \cos\left(\sum_{j=1}^i \theta'_j\right), y_0 + \sum_{i=1}^k \bar{l}_i \sin\left(\sum_{j=1}^i \theta'_j\right) \right). \quad (8.72)$$

Deste modo a outra extremidade do polígono, obtida após o lançamento dos n lados, corresponderá ao ponto A_n com coordenadas

$$(x_n, y_n) = \left(x_0 + \sum_{i=1}^n \bar{l}_i \cos\left(\sum_{j=1}^i \theta'_j\right), y_0 + \sum_{i=1}^n \bar{l}_i \sin\left(\sum_{j=1}^i \theta'_j\right) \right). \quad (8.73)$$

O quadrado da distância entre as duas extremidades é então dado por

$$d^2 = \left(\sum_{i=1}^n \bar{l}_i \cos\left(\sum_{j=1}^i \theta'_j\right) \right)^2 + \left(\sum_{i=1}^n \bar{l}_i \sin\left(\sum_{j=1}^i \theta'_j\right) \right)^2 \quad (8.74)$$

que, simplificando, obtemos

$$\begin{aligned} d^2 &= \left[\sum_{i=1}^n \bar{l}_i^2 \cos^2\left(\sum_{k=1}^i \theta'_k\right) + \sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n \bar{l}_i \bar{l}_j \cos\left(\sum_{k=1}^i \theta'_k\right) \cos\left(\sum_{k=1}^j \theta'_k\right) + \right. \\ &\quad \left. + \sum_{i=1}^n \bar{l}_i^2 \sin^2\left(\sum_{k=1}^i \theta'_k\right) + \sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n \bar{l}_i \bar{l}_j \sin\left(\sum_{k=1}^i \theta'_k\right) \sin\left(\sum_{k=1}^j \theta'_k\right) \right] = \\ &= \left[\sum_{i=1}^n \bar{l}_i^2 + \sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n \bar{l}_i \bar{l}_j \left(\cos\left(\sum_{k=1}^i \theta'_k\right) \cos\left(\sum_{k=1}^j \theta'_k\right) + \right. \right. \\ &\quad \left. \left. + \sin\left(\sum_{k=1}^i \theta'_k\right) \sin\left(\sum_{k=1}^j \theta'_k\right) \right) \right], \end{aligned} \quad (8.75)$$

que, uma vez que

$$\cos(\alpha - \beta) = \cos(\alpha) \cos(\beta) + \sin(\alpha) \sin(\beta),$$

podemos simplificar um pouco mais para

$$\begin{aligned} d^2 &= \sum_{i=1}^n \bar{l}_i^2 + 2 \sum_{i=1}^{n-1} \sum_{j=i+1}^n \bar{l}_i \bar{l}_j \cos \left(\sum_{k=1}^j \theta'_k - \sum_{k=1}^i \theta'_k \right) = \\ &= \sum_{i=1}^n \bar{l}_i^2 + 2 \sum_{i=1}^{n-1} \sum_{j=i+1}^n \bar{l}_i \bar{l}_j \cos \left(\sum_{k=i+1}^j \theta'_k \right). \end{aligned} \quad (8.76)$$

Em termos de valor esperado teremos

$$\begin{aligned} \mathbb{E}(d^2) &= \sum_{i=1}^n \bar{l}_i^2 + 2 \sum_{i=1}^{n-1} \sum_{j=i+1}^n \bar{l}_i \bar{l}_j \mathbb{E} \left(\cos \left(\sum_{k=i+1}^j \theta'_k \right) \right) \\ &= \sum_{i=1}^n \bar{l}_i^2 + 2 \sum_{i=1}^{n-1} \sum_{j=i+1}^n \bar{l}_i \bar{l}_j \mathbb{E} (\cos (\mathbf{S}_{j-i})), \end{aligned} \quad (8.77)$$

onde $\mathbf{S}_{j-i} = \sum_{k=i+1}^j \theta'_k$ corresponde à soma de $j-i$ variáveis aleatórias independentes e identicamente distribuídas.

Recordemos que a função característica φ_X de uma variável aleatória X é definida por

$$\varphi_X(t) = \mathbb{E} \left[e^{itX} \right] = \mathbb{E} [\cos(tX)] + i \mathbb{E} [\sin(tX)], \quad (8.78)$$

que, no caso de as variáveis aleatórias serem simétricas, verifica

$$\varphi_X(t) = \mathbb{E} [\cos(tX)], \quad (8.79)$$

uma vez que nestes casos teremos $\mathbb{E} [\sin(tX)] = 0$ pelo facto de a função seno ser ímpar. Assim, considerando que as variáveis θ'_i são simétricas em torno de zero (o que faz todo o sentido no problema e as duas distribuições referidas no enunciado verificam esta condição), as variáveis $\mathbf{S}_n$ também serão simétricas em torno de zero (pois resultam da soma de variáveis aleatórias independentes simétricas em torno de zero) e teremos

$$\mathbb{E} (\cos (\mathbf{S}_n)) = \varphi_{\mathbf{S}_n} (1). \quad (8.80)$$

Uma vez que, se X e Y são variáveis independentes, a função característica da soma será dada por

$$\varphi_{X+Y}(t) = \varphi_X(t) \varphi_Y(t), \quad (8.81)$$

então teremos

$$\varphi_{\mathbf{S}_n}(t) = \left(\varphi_{\theta'}(t) \right)^n \quad (8.82)$$

e, consequentemente,

$$\mathbb{E}(\cos(\mathbf{S}_n)) = \varphi_{\mathbf{S}_n}(1) = \left(\varphi_{\theta'}(1) \right)^n = \left(\mathbb{E}[\cos(\theta')] \right)^n. \quad (8.83)$$

Este resultado também pode ser deduzido pelas propriedades do co-seno, visto que

$$\begin{aligned} \mathbb{E}[\cos(\alpha + \beta)] &= \mathbb{E}[\cos(\alpha) \cos(\beta)] - \mathbb{E}[\sin(\alpha) \sin(\beta)] = \\ &= \mathbb{E}[\cos(\alpha)] \mathbb{E}[\cos(\beta)], \end{aligned} \quad (8.84)$$

se as variáveis α e β forem independentes e tiverem distribuição simétrica em torno de zero pois, nestes casos, verifica-se $\mathbb{E}[\sin(\alpha)] = \mathbb{E}[\sin(\beta)] = 0$. Se representarmos por $\rho = \mathbb{E}(\cos(\theta'))$, então, pela expressão (8.77), teremos

$$\begin{aligned} \mathbb{E}(d^2) &= \sum_{i=1}^n \bar{l}_i^2 + 2 \sum_{i=1}^{n-1} \sum_{j=i+1}^n \bar{l}_i \bar{l}_j \mathbb{E}(\cos(\mathbf{S}_{j-i})) = \\ &= \sum_{i=1}^n \bar{l}_i^2 + 2 \sum_{i=1}^{n-1} \sum_{j=i+1}^n \bar{l}_i \bar{l}_j \left(\mathbb{E}[\cos(\theta')] \right)^{j-i} = \\ &= \sum_{i=1}^n \bar{l}_i^2 + 2 \sum_{i=1}^{n-1} \sum_{j=i+1}^n \bar{l}_i \bar{l}_j \rho^{j-i}, \end{aligned} \quad (8.85)$$

onde ρ assume valores no intervalo $[0, 1)$, uma vez que corresponde ao valor esperado de um co-seno de uma variável aleatória contínua simétrica em torno de zero⁽⁸⁾.

⁽⁸⁾ Retiramos desta análise as variáveis aleatórias degeneradas no ponto zero, que só assumem o valor zero (excepto num conjunto de medida nula), pois nesta situação teríamos $\rho = 1$.

Representando o comprimento máximo dos lados por $\overline{l}_M$, isto é,

$$\overline{l}_M = \max_{i=1, \dots, n} \overline{l}_i \quad (8.86)$$

então obtemos

$$\mathbb{E}(d^2) \leq \sum_{i=1}^n \overline{l}_M^{-2} + 2\overline{l}_M^{-2} \sum_{i=1}^{n-1} \sum_{j=i+1}^n \rho^{j-i}. \quad (8.87)$$

Notemos que nesta expressão temos uma soma de termos em progressão geométrica com razão igual a ρ , logo, podemos concluir que

$$\begin{aligned} \mathbb{E}(d^2) &\leq \overline{l}_M^{-2} \left[n + 2 \sum_{i=1}^{n-1} \rho \frac{1 - \rho^{n-i}}{1 - \rho} \right] = \\ &= \overline{l}_M^{-2} \left[n + \frac{2\rho}{(1 - \rho)} \left(n - 1 - \sum_{i=1}^{n-1} \rho^{n-i} \right) \right] = \\ &= \overline{l}_M^{-2} \left[n + \frac{2\rho(n-1)}{(1 - \rho)} - \frac{2\rho^2(1 - \rho^{n-1})}{(1 - \rho)^2} \right]. \end{aligned} \quad (8.88)$$

A convergência desta expressão depende da forma como os comprimentos dos lados convergem para zero quando aumentamos o número n de lados. Se, tal como utilizamos na análise com os ângulos independentes, impusermos

$$\overline{l}_M = \max_{i=1, \dots, n} \overline{l}_i = o\left(n^{-(0.5+\epsilon)}\right), \quad \text{quando } n \rightarrow \infty, \quad (8.89)$$

para um qualquer valor $\epsilon > 0$, então a expressão (8.88) converge para zero quando n aumenta indefinidamente, pois

$$\begin{aligned} &\lim_{n \rightarrow +\infty} \overline{l}_M^{-2} \left[n + \frac{2\rho(n-1)}{(1 - \rho)} - \frac{2\rho^2(1 - \rho^{n-1})}{(1 - \rho)^2} \right] \leq \\ &\leq \lim_{n \rightarrow +\infty} \left[\frac{n}{n^{(1+2\epsilon)}} + \frac{2\rho(n-1)}{n^{(1+2\epsilon)}(1 - \rho)} - \frac{2\rho^2(1 - \rho^{n-1})}{n^{(1+2\epsilon)}(1 - \rho)^2} \right] \leq \\ &\leq \lim_{n \rightarrow +\infty} \left[n^{-2\epsilon} + \frac{2n^{-2\epsilon}\rho}{(1 - \rho)} - \frac{2\rho^2(1 - \rho^{n-1})}{n^{(1+2\epsilon)}(1 - \rho)^2} \right] = 0, \end{aligned} \quad (8.90)$$

e, conseqüentemente, teremos

$$\lim_{n \rightarrow \infty} \mathbb{E} \left(d^2 \right) = 0. \quad (8.91)$$

Naturalmente o caso em que os lados têm todos igual comprimento, i.e. $\bar{l}_i = \frac{L}{n}$, está incluído nesta situação e teremos igualmente nestes casos $\lim_{n \rightarrow \infty} \mathbb{E} \left(d^2 \right) = 0$. Concluimos, desta forma, que, quando as variáveis aleatórias θ'_i , que descrevem os ângulos entre dois lados consecutivos, são independentes, identicamente distribuídas e simétricas em torno de zero, então, se o comprimento máximo dos lados convergir para zero mais rapidamente que $n^{-(0.5+\epsilon)}$, o valor esperado do quadrado da distância entre os dois pontos extremos também tenderá para zero quando aumentamos indefinidamente o número de lados utilizados no polígono.

Nesta situação enquadram-se os dois casos investigados na simulação, quer aquele em que utilizamos a distribuição uniforme quer aquele em que recorreremos à distribuição Beta. Para a situação em que os ângulos θ'_i são caracterizados por uma distribuição uniforme no intervalo $\left[-\frac{\pi}{k}, \frac{\pi}{k}\right]$, a função característica correspondente será dada por

$$\varphi_{\mathbf{S}_1}(t) = \frac{e^{\frac{it\pi}{k}} - e^{-\frac{it\pi}{k}}}{\frac{2it\pi}{k}} = \frac{\sin\left(\frac{t\pi}{k}\right)}{\frac{t\pi}{k}} \quad (8.92)$$

e, conseqüentemente, teremos

$$\varphi_{\mathbf{S}_n}(t) = \left(\frac{\sin\left(\frac{t\pi}{k}\right)}{\frac{t\pi}{k}} \right)^n, \quad (8.93)$$

que no ponto $t = 1$ verifica

$$\mathbb{E} [\cos(\mathbf{S}_n)] = \varphi_{\mathbf{S}_n}(1) = \left(\frac{\sin\left(\frac{\pi}{k}\right)}{\frac{\pi}{k}} \right)^n. \quad (8.94)$$

Na Figura 8.4 está representado o gráfico da função $f(x) = \frac{\sin(x)}{x}$ para valores do intervalo $(0, \pi]$, pois, uma vez que $k \geq 1$, teremos $\frac{\pi}{k} \in (0, \pi]$.

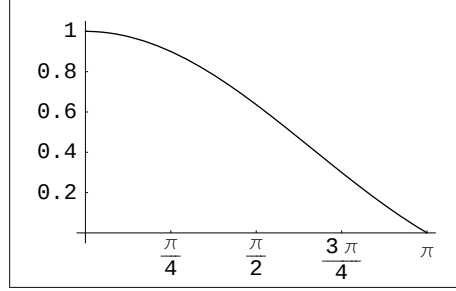


Figura 8.4: Gráfico de $f(x) = \frac{\sin(x)}{x}$

Desta representação podemos concluir que $0 \leq f(x) < 1$ ⁽⁹⁾ no intervalo referido e consequentemente $f^n(x) \rightarrow 0$ quando n aumenta. Deste modo podemos, partindo da expressão (8.85), determinar

$$\mathbb{E}(d^2) = \sum_{i=1}^n \bar{l}_i^2 + 2 \sum_{i=1}^{n-1} \sum_{j=i+1}^n \bar{l}_i \bar{l}_j \left(\frac{\sin\left(\frac{\pi}{k}\right)}{\frac{\pi}{k}} \right)^{j-i}. \quad (8.95)$$

Desta expressão podemos deduzir, tal como previamente efectuámos para o caso geral, que $\lim_{n \rightarrow \infty} \mathbb{E}(d^2) = 0$ se $\max_{i=1, \dots, n} \bar{l}_i = o\left(n^{-(0.5+\epsilon)}\right)$. Em relação à situação analisada por Pacheco d'Amorim, considerando lados com igual comprimento ($\bar{l}_i = \frac{L}{n}, \forall i$), podemos também concluir, nestas situações em que os ângulos entre lados são caracterizados pela distribuição uniforme no intervalo $[-\frac{\pi}{k}, \frac{\pi}{k}]$ com $k \in [1, +\infty)$, que o valor esperado da distância ao quadrado tende para zero⁽¹⁰⁾.

⁽⁹⁾ Igual conclusão poderíamos deduzir pelo estudo da derivada da função f . Como a derivada de f é negativa para $x \in (0, \pi]$, a função é sempre decrescente neste intervalo. Uma vez que a função é contínua e o limite da função, quando x tende para zero, é igual a um e a imagem da função no ponto π é zero, a função variará entre estes dois valores.

⁽¹⁰⁾ Retiramos da análise efectuada a situação $k \rightarrow \infty$, isto é, a situação em que a distribuição considerada para os ângulos entre dois lados consecutivos seria degenerada no ponto zero, pois nesta situação teríamos $\rho = 1$ e, consequentemente,

$$\mathbb{E}(d^2) = \sum_{i=1}^n \bar{l}_i^2 + \sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n \bar{l}_i \bar{l}_j = \sum_{i=1}^n \bar{l}_i^2 + \sum_{i=1}^n \bar{l}_i (L - \bar{l}_i) = \sum_{i=1}^n \bar{l}_i L = L^2,$$

A segunda situação investigada na simulação utiliza a variável aleatória θ''_i com distribuição Beta com ambos os parâmetros iguais a dois, cuja função densidade é dada por

$$f_{\theta''_i}(\theta'') = \frac{1}{\mathbf{B}(2, 2)} \theta''^{2-1} (1 - \theta'')^{2-1} \mathbb{I}_{[0,1]}(\theta'') = 6\theta'' (1 - \theta'') \mathbb{I}_{[0,1]}(\theta''), \quad (8.96)$$

logo, a função densidade de $\theta'_i = (\theta''_i - 0.5) \frac{2\pi}{k} = \theta''_i \frac{2\pi}{k} - \frac{\pi}{k}$ é dada por

$$\begin{aligned} f_{\theta'_i}(\theta') &= \frac{6}{\frac{2\pi}{k}} \left(\frac{\theta' + \frac{\pi}{k}}{\frac{2\pi}{k}} \right) \left(1 - \frac{\theta' + \frac{\pi}{k}}{\frac{2\pi}{k}} \right) \mathbb{I}_{[-\frac{\pi}{k}, \frac{\pi}{k}]}(\theta') = \\ &= \left(\frac{\pi}{k} \right)^{-3} \left(\theta' + \frac{\pi}{k} \right) \left(\frac{\pi}{k} - \theta' \right) \mathbb{I}_{[-\frac{\pi}{k}, \frac{\pi}{k}]}(\theta'). \end{aligned} \quad (8.97)$$

Uma vez que a variável θ'_i é simétrica em relação a zero, teremos

$$\begin{aligned} \mathbb{E} \left[e^{it\theta'_i} \right] &= \mathbb{E} \left[\cos(t\theta'_i) \right] = \\ &= \int_{-\frac{\pi}{k}}^{\frac{\pi}{k}} \cos(t\theta'_i) \left(\frac{\pi}{k} \right)^{-3} \left(\theta' + \frac{\pi}{k} \right) \left(\frac{\pi}{k} - \theta' \right) d\theta' = \\ &= 3 \left[\frac{\sin\left(\frac{t\pi}{k}\right)}{\left(\frac{t\pi}{k}\right)^3} - \frac{\cos\left(\frac{t\pi}{k}\right)}{\left(\frac{t\pi}{k}\right)^2} \right] \end{aligned} \quad (8.98)$$

e, para $t = 1$, temos

$$\mathbb{E} \left[\cos(\theta'_i) \right] = 3 \left[\frac{\sin\left(\frac{\pi}{k}\right)}{\left(\frac{\pi}{k}\right)^3} - \frac{\cos\left(\frac{\pi}{k}\right)}{\left(\frac{\pi}{k}\right)^2} \right]. \quad (8.99)$$

Notemos que, como ilustra a Figura 8.5, a função $g(x) = 3 \left(\frac{\sin(x)}{x^3} - \frac{\cos(x)}{x^2} \right)$ é positiva e inferior a um no intervalo $(0, \pi]$ (que corresponde a $k \geq 1$).

Outra forma de obter este resultado seria recorrer à função característica de uma variável aleatória θ''_i com distribuição Beta com ambos os parâmetros iguais a dois, pois esta é dada (consultar, por exemplo, Balakrishnan e Nevzorov (2003)) por

$$\varphi_{\theta''_i}(t) = 6 \left(\frac{2e^{it} - 2}{it^3} - \frac{e^{it} + 1}{t^2} \right) = \frac{6}{t^3} [-2i + t + e^{it}(2i + t)]. \quad (8.100)$$

como seria de esperar, uma vez que todos os lados teriam o mesmo sentido independentemente da forma como os comprimentos dos lados evoluam.

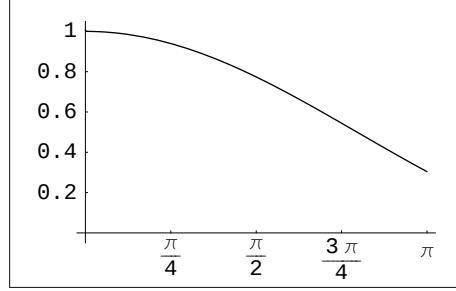


Figura 8.5: Gráfico de $g(x) = 3 \left(\frac{\sin(x)}{x^3} - \frac{\cos(x)}{x^2} \right)$

Uma vez que a função característica para transformações lineares $Y = aX + b$ é dada por $\varphi_Y(t) = e^{itb} \varphi_X(at)$ e como $\theta'_i = \theta''_i \frac{2\pi}{k} - \frac{\pi}{k}$, a sua função característica será

$$\begin{aligned} \varphi_{\theta'_i}(t) &= e^{-\frac{it\pi}{k}} \varphi_{\theta''_i} \left(\frac{2t\pi}{k} \right) = \\ &= e^{-\frac{it\pi}{k}} \frac{6}{\left(\frac{2t\pi}{k} \right)^3} \left[-2i + \frac{2t\pi}{k} + e^{\frac{2it\pi}{k}} \left(2i + \frac{2t\pi}{k} \right) \right] = \\ &= 3 \left[\frac{\sin \left(\frac{t\pi}{k} \right)}{\left(\frac{t\pi}{k} \right)^3} - \frac{\cos \left(\frac{t\pi}{k} \right)}{\left(\frac{t\pi}{k} \right)^2} \right] \end{aligned} \quad (8.101)$$

onde recorreremos à fórmula de Euler $e^{ix} = \cos x + i \sin x$.

Desta forma iríamos obter uma expressão para o valor esperado do quadrado da distância semelhante à obtida na situação em que a distribuição dos ângulos era uniforme (expressão (8.95)), que é

$$\mathbb{E}(d^2) = \sum_{i=1}^n \bar{l}_i^2 + 2 \sum_{i=1}^{n-1} \sum_{j=i+1}^n \bar{l}_i \bar{l}_j \left(3 \left[\frac{\sin \left(\frac{\pi}{k} \right)}{\left(\frac{\pi}{k} \right)^3} - \frac{\cos \left(\frac{\pi}{k} \right)}{\left(\frac{\pi}{k} \right)^2} \right] \right)^{j-i}, \quad (8.102)$$

sendo a única alteração a expressão de ρ . Todavia, esta alteração nada muda em relação às conclusões apresentadas, uma vez que o valor de ρ se situará ainda no intervalo $[0, 1)$.

Em geral os ângulos entre dois lados consecutivos são caracterizados por variáveis aleatórias independentes θ'_i com valor esperado nulo e limitadas,

pois só assumem valores num intervalo contido em $[-\pi, \pi]$ (logo, terão variância finita que representaremos por $\sigma_{\theta'}^2$). Como tal, mesmo que as variáveis não tenham distribuição simétrica, podemos aplicar o Teorema Limite Central e a variável $\mathbf{S}_n$, obtida como a soma de n destas variáveis, tenderá para a distribuição gaussiana de valor esperado nulo e variância $\sigma^2 = n\sigma_{\theta'}^2$. Consequentemente, a partir de certa altura, para m suficientemente grande, teremos

$$\mathbb{E}[\cos(\mathbf{S}_m)] \approx \int_{\mathbb{R}} \cos(x) \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{x^2}{2\sigma^2}} dx = e^{-\frac{\sigma^2}{2}} = e^{-\frac{m\sigma_{\theta'}^2}{2}}, \quad (8.103)$$

que facilmente obtemos pela função característica, pois, se a variável aleatória Y é caracterizada por uma distribuição gaussiana com média μ e variância σ^2 , então a sua função característica será dada por

$$\varphi_Y(t) = e^{it\mu - \frac{(t\sigma)^2}{2}}. \quad (8.104)$$

Representemos por $\rho(j-i) = \mathbb{E}(\cos(\mathbf{S}_{j-i}))$ que, para valores elevados de m , será aproximadamente $\rho(m) \approx e^{-\frac{m\sigma_{\theta'}^2}{2}}$ (que converge rapidamente para zero) e por $\bar{l}_M$ o comprimento máximo dos lados. Deste modo, considerando um valor de n tal que $n > m$, teremos para o valor esperado

$$\begin{aligned} \mathbb{E}(d^2) &= \sum_{i=1}^n \bar{l}_i^2 + 2 \sum_{i=1}^{n-1} \sum_{j=i+1}^n \bar{l}_i \bar{l}_j \mathbb{E}(\cos(\mathbf{S}_{j-i})) \leq \\ &\leq \sum_{i=1}^n \bar{l}_M^2 + 2\bar{l}_M^2 \sum_{i=1}^{n-1} \sum_{j=i+1}^n \rho(j-i) = \\ &= n\bar{l}_M^2 + 2\bar{l}_M^2 \sum_{i=1}^{n-1} (n-i)\rho(i) \approx \\ &\approx n\bar{l}_M^2 + 2\bar{l}_M^2 \left[\sum_{i=1}^m (n-i)\rho(i) + \sum_{i=m+1}^{n-1} (n-i)e^{-\frac{i\sigma_{\theta'}^2}{2}} \right] \leq \\ &\leq n\bar{l}_M^2 + 2\bar{l}_M^2 \left[\sum_{i=1}^m (n-i) + nK \right] = \\ &= \bar{l}_M^2 [n(1+2m+2K) - m^2 - m], \end{aligned} \quad (8.105)$$

onde

$$K = \lim_{n \rightarrow \infty} \sum_{i=m+1}^n e^{-\frac{i\sigma_{\theta'}^2}{2}}, \quad (8.106)$$

uma vez que a série é convergente. Se considerarmos o comprimento máximo a convergir para zero, de forma que $\overline{l}_M = o(n^{-(0.5+\epsilon)})$, com $\epsilon > 0$, teremos

$$\begin{aligned} & \lim_{n \rightarrow \infty} \overline{l}_M^{-2} [n(1 + 2m + 2K) - m^2 - m] \leq \\ & \leq \lim_{n \rightarrow \infty} \frac{n(1 + 2m + 2K) - m^2 - m}{n^{1+2\epsilon}} = 0 \end{aligned} \quad (8.107)$$

e, conseqüentemente,

$$\lim_{n \rightarrow \infty} \mathbb{E}(d^2) = 0. \quad (8.108)$$

Assim, se a convergência do comprimentos dos lados for tal que o comprimento máximo convirja para zero mais rapidamente que $n^{-(0.5+\epsilon)}$, tal como acontece na situação em que os lados têm todos o mesmo comprimento, o valor esperado do quadrado da distância converge para zero, desde que as variáveis aleatórias θ' , que caracterizam os ângulos entre os lados consecutivos do polígono, sejam independentes e identicamente distribuídas com valor esperado nulo e não sejam degeneradas. $\diamond$

8.4.2 Problema de escala

Concluimos, nas análises previamente apresentadas, que é nulo o valor esperado do quadrado da distância dos dois pontos extremos da curva flexível e inextensível lançada à sorte num plano. Para deduzirmos esta conclusão é essencial que o comprimento dos lados convirja para zero, pois, se os lados tiverem comprimento fixo não nulo, por mais pequeno que seja, o valor esperado do quadrado da distância dos dois pontos extremos já não será nulo. Desta forma, uma questão se coloca a este processo concebido por Pacheco d'Amorim. Será que nesta forma de construir o polígono, no limite, chegaremos a sair do mesmo ponto? Por outras palavras, no limite obteremos

um polígono, que certamente será fechado, ou simplesmente teremos sempre, como resultado desta construção, um único ponto? Notemos que a distância do ponto inicial a qualquer outro ponto do polígono é também nula, pois na construção da curva flexível e inextensível começamos com um segmento de comprimento L e vamos *dobrando* sucessivamente os lados do polígono de forma que o comprimento de cada lado convirja para zero. Ora, se fixarmos um ponto, seja p , do segmento inicial, que tenha uma distância $L_1 < L$ do ponto inicial, o quadrado da distância entre este ponto e o ponto inicial terá valor esperado nulo após as infinitas *dobragens* (pois será equivalente a considerar que o segmento inicial teria comprimento L_1). Uma vez que o valor esperado do quadrado da distância, entre o ponto inicial e qualquer outro ponto do polígono, é nulo, podemos concluir que a distância será igual a zero (excepto num conjunto de medida nula). Assim, não teremos um polígono fechado, mas apenas um ponto. Por esta razão não podemos considerar que o perímetro total do polígono se mantenha constante neste processo, pois um ponto tem perímetro nulo.

n	10	100	1000	10000
Uniforme, $k = 1$	0.068865	0.013417	0.000712	0.000046
Uniforme, $k = 9$	0.970272	0.175007	0.078634	0.003156
Uniforme, $k = 180$	0.999859	0.995492	0.995712	0.799578
Beta, $k = 1$	0.169517	0.040481	0.001222	0.000002
Beta, $k = 9$	0.968909	0.747073	0.006779	0.002663
Beta, $k = 180$	0.999801	0.998808	0.99233	0.90184

Tabela 8.5: Quadrado das distâncias entre os pontos extremos

Na Figura 8.6 estão representados quatro passeios aleatórios em cada gráfico, utilizando o número de lados n com os valores 10, 100, 1000 e 10000. Os ângulos entre lados consecutivos são caracterizados pela distribuição uni-

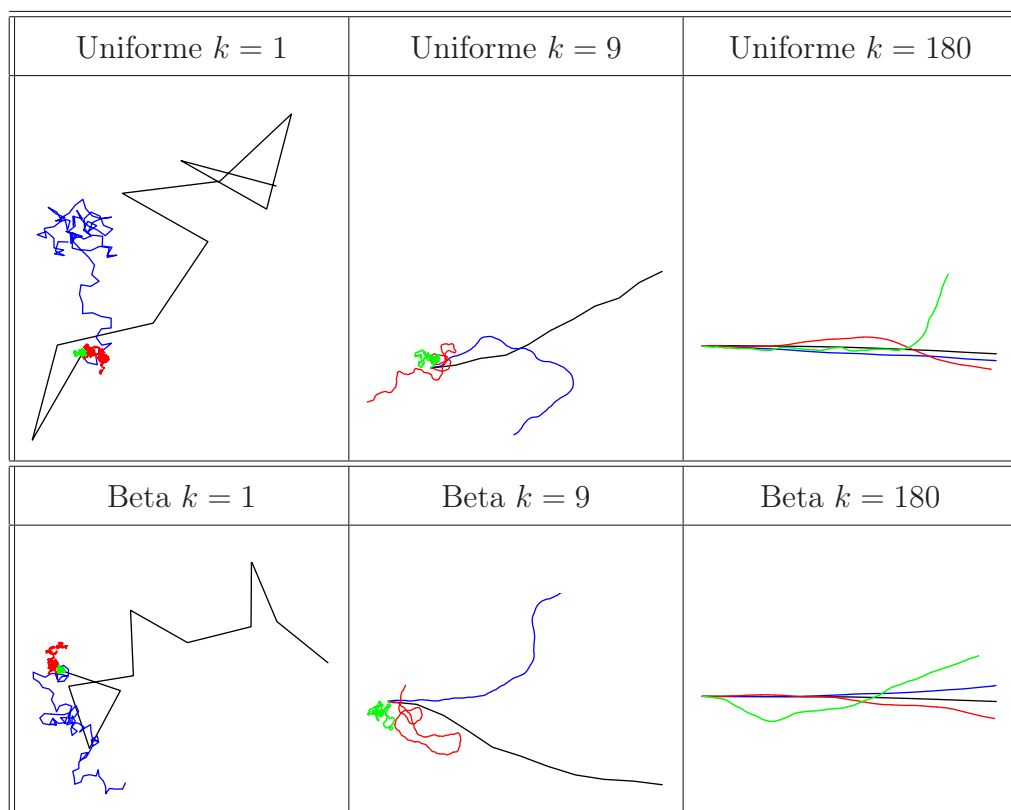


Figura 8.6: Passeios aleatórios com n distinto

forme e pela distribuição beta, conforme apresentado previamente, com k assumindo os valores 1, 9 e 180. Os gráficos mostram claramente, pelo menos para valores de k baixos, que a escala dos passeios aleatórios se vai modificando. Assim, à medida que o número de lados n vai aumentando o polígono parece ir diminuindo, sendo necessário alterar a escala para podermos visualizar os polígonos formados com um número de lados superior. Na Tabela 8.5 estão presentes os quadrados das distâncias entre os pontos extremos dos polígonos representados nos gráficos da Figura 8.6. Excepto nos casos onde $k = 180$, onde teríamos que utilizar valores mais elevados para o número de lados para visualizarmos esta convergência, os valores reafirmam a nossa ideia de que a escala está a diminuir com o aumento do número de lados.

Esta questão fez-nos pensar em, de certa forma, contrabalançar a dimi-

nuição do tamanho dos lados que definem o polígono, para que este não se desfigure obrigatoriamente num ponto, tal como foi utilizado no fluxo de Ricci normalizado. Em 2003 o matemático russo Gregori Parelman demonstrou a conjectura de Poincaré, enunciada em 1904, que, reconstituída de um modo muito simplificado, para evitar qualquer erro uma vez não ser a nossa área de especialização, afirma que todo o espaço tridimensional fechado e *sem buracos* pode ser deformado continuamente numa esfera, isto é, tem uma forma essencialmente esférica. Este problema tornou-se um dos maiores enigmas da matemática durante o século XX, sendo considerado um dos sete “*Problemas do Milénio*”, definidos pelo *Clay Mathematics Institute* que, cem anos depois, no mesmo local que Hilbert propôs os seus 23 problemas para o século XX, ofereceu um milhão de dólares a quem conseguisse resolver um dos problemas propostos (desde que publicado numa revista científica e submetido a dois anos para a detecção de qualquer erro).

Na demonstração da conjectura de Poincaré, Parelman recorre ao fluxo de Ricci normalizado. O fluxo de Ricci foi criado por Richard Hamilton na década de 1980 e tem o objectivo de transformar qualquer variedade numa geometria uniforme, isto é, com a curvatura constante. Hamilton denominou este processo por fluxo de Ricci em homenagem a Gregorio Ricci-Curbastro (1853–1925), um matemático italiano da área da geometria diferencial. A ideia da demonstração de Parelman, segundo Viana (2007), é, partindo de uma variedade qualquer M , de dimensão 3 e *sem buracos*, recorrer a uma métrica para deformar a variedade M , aumentando a curvatura onde esta é pequena e diminuindo a curvatura onde esta é grande, de forma que a curvatura convirja para uma geometria uniforme. Para que a conjectura de Poincaré seja verdadeira, este processo deveria convergir para uma esfera tridimensional, sendo neste processo utilizado o fluxo de Ricci que é definido por

$$\frac{d}{dt}g_{ij}(t) = -2R_{ij} \quad (8.109)$$

onde g_{ij} é a métrica utilizada, R_{ij} é o tensor da curvatura de Ricci e t é o tempo da deformação. Contudo, o fluxo de Ricci não preserva o volume da variedade, transformando a forma inicial sempre num ponto. Por esta razão Parelman definiu o fluxo de Ricci normalizado (*with surgery*) através de

$$\frac{d}{dt}g_{ij}(t) = -2R_{ij} + \lambda g_{ij}(t) \quad (8.110)$$

onde λ é a constante cosmológica que é definida de forma a preservar o volume da superfície.

Na transformação aplicada por Pacheco d'Amorim, partindo de um segmento de dimensão L , no limite teremos, sob determinadas condições previamente apresentadas, sempre um ponto. Por este motivo vamos efectuar uma normalização à transformação. Assim, quando aumentamos o número de lados a considerar, vamos efectuar um aumento no perímetro total da figura de forma a averiguar se existe algum equilíbrio.

Recordemos que, em todas as situações em que deduzimos a convergência do valor esperado do quadrado da distância entre os dois pontos extremos do polígono para zero, i.e. $\mathbb{E}(d^2) \rightarrow 0$, consideramos sempre que o comprimento máximos dos lados verificava $\overline{l}_M = o(n^{-(0.5+\epsilon)})$ com $\epsilon > 0$. Suponhamos agora que o perímetro total do polígono é dado por uma função $L(n) = L_0\sqrt{n}$, onde L_0 representa o comprimento inicial, e restrinjam a nossa análise ao caso em que os lados têm todos igual comprimento, tendo cada lado comprimento igual a $\frac{L_0}{\sqrt{n}}$. Da expressão (8.63) (página 493) podemos concluir que, quando os ângulos θ_i são variáveis independentes com distribuição uniforme no intervalo $[0, 2\pi]$, se verifica

$$\mathbb{E}[d^2] = \sum_{i=1}^n \overline{l}_i^2 = n \left(\frac{L_0}{\sqrt{n}} \right)^2 = L_0^2, \quad (8.111)$$

para qualquer valor de n . Desta forma, se considerarmos lados com comprimento igual a $\frac{L_0}{\sqrt{n}}$, teremos o valor esperado do quadrado da distância constante na situação de independência. Em relação à situação em que os ângulos θ'_i entre lados consecutivos são caracterizados por uma distribuição simétrica em torno de zero teremos (pela expressão (8.85), presente na página 505),

$$\begin{aligned}
 \mathbb{E}(d^2) &= \sum_{i=1}^n \left(\frac{L_0}{\sqrt{n}} \right)^2 + 2 \left(\frac{L_0}{\sqrt{n}} \right)^2 \sum_{i=1}^{n-1} \sum_{j=i+1}^n \rho^{j-i} = \\
 &= L_0^2 \left[1 + \frac{2}{n} \sum_{i=1}^{n-1} \rho \frac{1 - \rho^{n-i}}{1 - \rho} \right] = \\
 &= L_0^2 \left[1 + \frac{2\rho}{n(1-\rho)} \left(n - 1 - \sum_{i=1}^{n-1} \rho^{n-i} \right) \right] = \\
 &= L_0^2 \left[1 + \frac{2\rho(n-1)}{n(1-\rho)} - \frac{2\rho^2(1-\rho^{n-1})}{n(1-\rho)^2} \right]. \tag{8.112}
 \end{aligned}$$

Uma vez que $0 \leq \rho = \mathbb{E}[\cos(\theta')] < 1$, o limite, quando n aumenta indefinidamente, será dado por

$$\lim_{n \rightarrow +\infty} \mathbb{E}(d^2) = L_0^2 \left[1 + \frac{2\rho}{(1-\rho)} \right] = L_0^2 \left[\frac{1+\rho}{1-\rho} \right], \tag{8.113}$$

que assume valor mínimo L_0^2 , que se verifica nas situações em que $\rho = 0$, que corresponde, por exemplo, ao caso em que os ângulos θ'_i são caracterizados por uma distribuição uniforme no intervalo $[-\frac{\pi}{k}, \frac{\pi}{k}]$ com $k = 1$, ou seja, à situação em que os ângulos entre cada lado e o eixo das abcissas são independentes. O valor de $\lim_{n \rightarrow +\infty} \mathbb{E}(d^2)$ é, nesta situação, crescente com o valor de ρ e tenderá para infinito se ρ tender, por valores inferiores, para a unidade, caso que corresponde à situação de os ângulos terem distribuição degenerada no ponto zero, isto é, em que todos os lados têm o mesmo sentido. Como tal, nestas situações, obteremos um segmento cujo comprimento vai sempre aumentando. Nas situações em que $\rho \neq 1$, o limite será sempre o indicado na expressão (8.113). Assim, a equação (8.113) é válida desde que os ângulos

entre dois lados consecutivos sejam independentes e simétricos em torno de zero, pois nestas condições $\rho = \mathbb{E}[\cos(\theta')]$ corresponderá à imagem de uma função característica em $t = 1$. Nestes casos, se pretendermos fixar um limite, por exemplo L^2 , bastará escolher L_0 , comprimento do primeiro lado, igual a $L_0 = L\sqrt{\frac{1-\rho}{1+\rho}}$, onde $\rho = \mathbb{E}[\cos(\theta')]$ é determinado pela distribuição dos ângulos e obteremos $\lim_{n \rightarrow +\infty} \mathbb{E}(d^2) = L^2$. Analisemos o que acontece neste processo caso o nosso objectivo não se restrinja a determinar o limite.

Por exemplo, para uma distribuição uniforme no intervalo $[-\frac{\pi}{k}, \frac{\pi}{k}]$ com $k = 360$, que corresponde ao valor máximo considerado nas simulações (sem a normalização), onde um lado só pode formar um ângulo de, no máximo, meio grau com o lado anterior, teríamos um valor esperado para o quadrado da distância a tender para aproximadamente $157574L_0^2$.

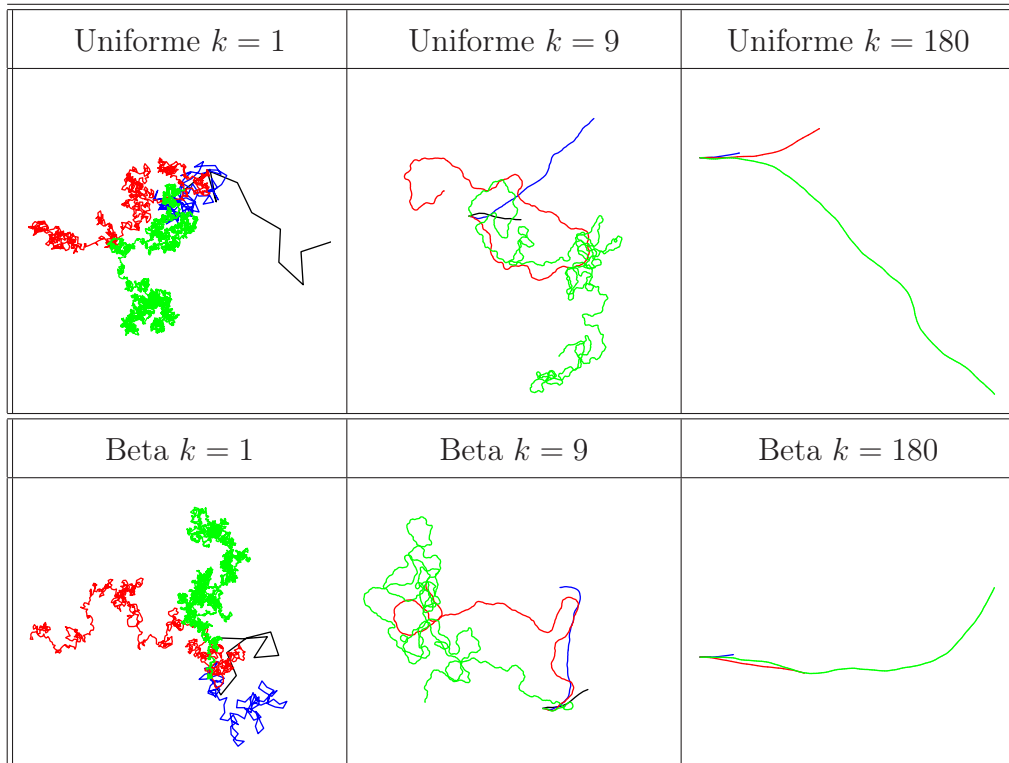


Figura 8.7: Passeios com normalização

Na Figura 8.7 mostramos 6 gráficos, cada um com quatro polígonos com

n	10	100	1000	10000
Uniforme, $k = 1$	2.24809	0.0862982	1.37117	1.28081
Uniforme, $k = 9$	9.59402	88.6543	4.39139	97.5701
Uniforme, $k = 180$	9.99856	99.5093	952.137	8985.41
Beta, $k = 1$	0.222216	0.396373	5.2863	2.25524
Beta, $k = 9$	9.42887	58.5562	115.842	54.3372
Beta, $k = 180$	9.99941	99.6059	996.49	8187.68

Tabela 8.6: Quadrado das distâncias entre os pontos extremos

a mesma caracterização no que se refere à distribuição utilizada, mas com o número de lados distinto. Os polígonos representados foram definidos utilizando lados com comprimento igual a $\frac{1}{\sqrt{n}}$, de forma a garantir que o limite do valor esperado do quadrado da distância seja um valor não nulo, conforme expressão (8.113). Os gráficos parecem evidenciar que, para valores baixos de k , os polígonos formados com distintos valores de n se situam à mesma escala; contudo, quando o valor de k é elevado ($k = 180$), notamos que existe já uma grande diferença entre a escala de cada polígono representado. Para analisarmos melhor estes exemplos, na Tabela 8.6 estão presentes os valores do quadrado da distância entre os dois pontos extremos de cada um dos polígonos representados nos gráficos da Figura 8.7. Notemos que, de facto, utilizando os lados com comprimento $\frac{1}{\sqrt{n}}$, conseguimos evitar o problema de escala, isto é, os polígonos tendem para um valor fixo, que podemos escolher *a priori*. Porém, os gráficos representados na Figura 8.7 e os valores patentes na Tabela 8.6 parecem evidenciar que, onde antes este problema era mais notório (valores de k baixos), conseguimos resolver o problema de escala, mas, nas situações onde anteriormente parecia não haver este problema (para valores de k elevados conforme gráficos na Figura 8.6), agora surgem. Para investigarmos esta situação recorreremos à simulação onde utili-

zamos $r = 10000$ réplicas e obtivemos os valores presentes na Tabela 8.7.

n	10	100	1000	10000
Unif., $k = 1$ Média	1.0212	1.01385	1.04724	1.00246
Unif., $k = 1$ Desv. Padrão	0.968133	1.02381	1.03111	0.999883
Unif., $k = 9$ Média	9.38237	57.0569	90.229	99.9522
Unif., $k = 9$ Desv. Padrão	0.512856	22.8362	82.4963	98.5065
Unif., $k = 360$ Média	9.99959	99.9576	995.848	9585.03
Unif., $k = 360$ Desv. Padrão	0.000329	0.035612	3.74747	341.246
Beta, $k = 1$ Média	1.71816	1.86948	1.84355	1.85484
Beta, $k = 1$ Desv. Padrão	1.4821	1.79769	1.88194	1.79754
Beta, $k = 9$ Média	9.60735	69.6632	154.21	159.169
Beta, $k = 9$ Desv. Padrão	0.318464	19.2391	130.116	156.077
Beta, $k = 360$ Média	9.99975	99.9747	997.619	9758.63
Beta, $k = 360$ Desv. Padrão	0.000219	0.021892	2.03097	211.823

Tabela 8.7: Simulação com processo com limite esperado fixo

Os valores presentes na Tabela 8.7, referentes ao quadrado da distância entre os pontos extremos, demonstram bem que, apesar de os valores convergirem para um limite, ainda variam bastante, pelo menos para valores de n baixos. Outra ideia presente nos resultados da simulação é que, quanto maior for o valor de k , a que corresponderá um valor maior de ρ , maior é a variação das médias obtidas quando aumentamos o número de lados utilizados na construção do polígono. Assim, se não pretendermos apenas analisar o limite e quisermos usar diversos valores de n , no modelo normalizado a distância esperada varia muito consoante o valor de n utilizado, como podemos constatar pela análise da expressão (8.112). Para ilustrar esta ideia, na Tabela 8.8 estão presentes alguns valores de $\mathbb{E}(d^2)$ em função de n e de ρ .

$\rho \downarrow \quad n \rightarrow$	10	100	1000	10000	100000	limite
0.1	1.1975	1.2197	1.2219	1.2222	1.2222	1.2222
0.25	1.5778	1.6578	1.6658	1.6666	1.6667	1.6667
0.5	2.3389	2.6066	2.6333	2.6361	2.6363	2.6364
0.75	4.7351	6.76	6.976	6.9976	6.9998	7
0.9	7.2762	17.2001	18.82	18.982	18.9982	19
0.95	8.5040	31.445	38.24	38.924	38.9924	39
0.99	9.6765	73.4744	179.201	197.0	198.802	199
0.999	9.9671	91.7471	735.655	1799.209	1979.02	1999
0.9999	9.9968	99.6675	967.482	7357.485	17999.29	19999

Tabela 8.8: Evolução de $\mathbb{E}(d^2)$ em função de n e ρ

Os valores apresentados na Tabela 8.8 demonstram a enorme variabilidade em função de n . Refira-se que para a distribuição uniforme no intervalo $[-\frac{\pi}{k}, \frac{\pi}{k}]$ temos $\rho = 0$ para $k = 1$, $\rho \approx 0,63662$ para $k = 2$, $\rho \approx 0,97982$ para $k = 9$ e $\rho \approx 0,999987$ para $k = 360$. Para a situação em que utilizamos a distribuição Beta teremos $\rho \approx 0,30396$ para $k = 1$, $\rho \approx 0,77404$ para $k = 2$, $\rho \approx 0,98787$ para $k = 9$ e $\rho \approx 0,999992$ para $k = 360$. Assim, se pretendermos utilizar distribuições onde o valor de ρ seja próximo da unidade e o número de lados do polígono não seja muito elevado, continuaremos a ter um problema de escala, pois o valor esperado do quadrado da distância vai aumentando bastante até atingir o seu limite. Para evitar este problema poderíamos utilizar um comprimento $L(n)$ para os lados que mantenha o valor esperado do quadrado da distância constante, que pela expressão (8.112) será dado por

$$\mathbb{E}(d^2) = \sum_{i=1}^n [L(n)]^2 + 2[L(n)]^2 \sum_{i=1}^{n-1} \sum_{j=i+1}^n \rho^{j-i} = L^2 \Leftrightarrow$$

$$\begin{aligned}
&\Leftrightarrow [L(n)]^2 \left[n + \frac{2\rho(n-1)}{(1-\rho)} - \frac{2\rho^2(1-\rho^{n-1})}{(1-\rho)^2} \right] = L^2 \Leftrightarrow \\
&\Leftrightarrow L(n) = L \left[n + \frac{2\rho(n-1)}{(1-\rho)} - \frac{2\rho^2(1-\rho^{n-1})}{(1-\rho)^2} \right]^{-\frac{1}{2}}. \quad (8.114)
\end{aligned}$$

Assim, se utilizarmos lados com comprimento dado pela expressão (8.114), teremos, independentemente do valor de n , o valor esperado do quadrado da distância, entre os dois pontos extremos do polígono, igual a L^2 , isto é, uma constante que escolhemos *a priori*.

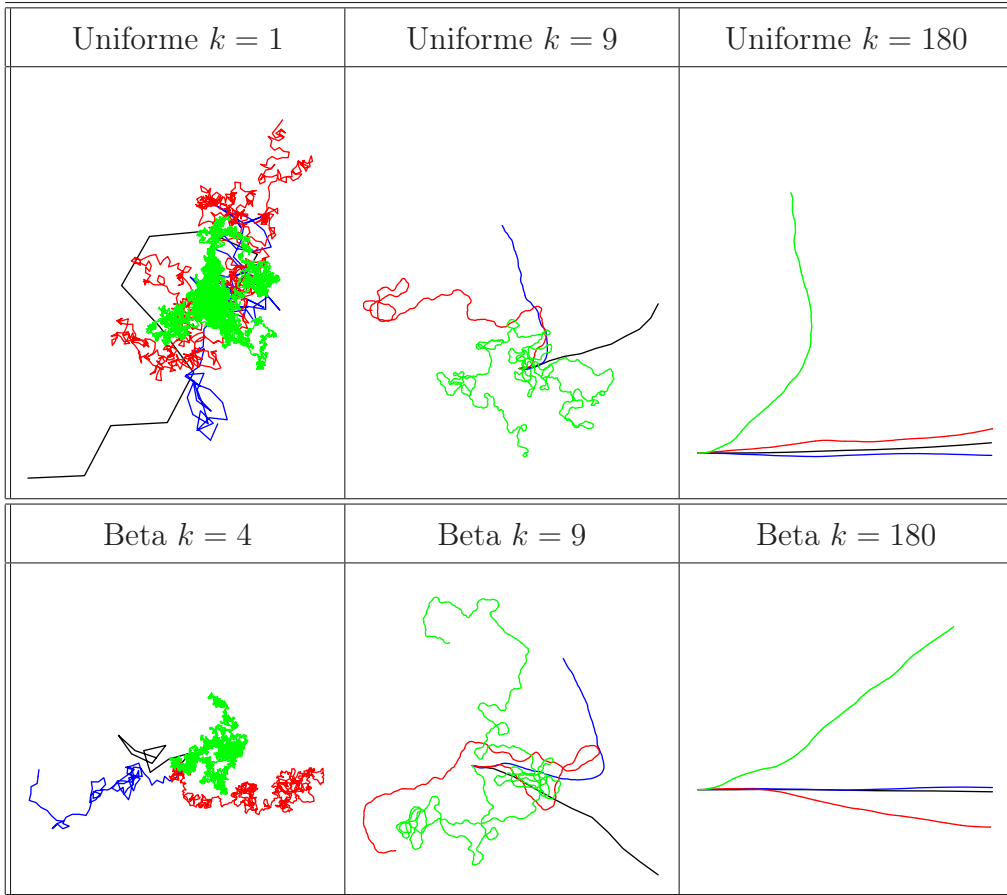


Figura 8.8: Passeios com distância esperada fixa

Na Figura 8.8, semelhante às Figuras 8.6 e 8.7, estão representados seis gráficos, cada um com quatro polígonos com o número de lados distinto, mas

com a mesma distribuição e utilizando lados com comprimento definido pela fórmula (8.114).

n	10	100	1000	10000
Uniforme, $k = 1$	1.48877	0.679048	1.06736	0.225169
Uniforme, $k = 9$	0.992787	1.06951	1.02311	0.545437
Uniforme, $k = 180$	0.999807	1.00038	1.00627	0.940506
Beta, $k = 1$	0.236091	1.44089	1.43936	0.902416
Beta, $k = 9$	1.00234	0.653142	0.531115	0.58078
Beta, $k = 180$	1.00002	1.00026	1.00127	1.02892

Tabela 8.9: Distâncias entre os pontos extremos

Notemos que os polígonos presentes em cada gráfico parecem ter uma distância entre os seus pontos extremos muito próxima. Para confirmar esta ideia, na Tabela 8.9 estão presentes os quadrados das distâncias entre os pontos extremos dos gráficos representados na Figura 8.8. Os valores, apesar de apresentarem algumas variações, estão todos situados relativamente perto da unidade, que corresponde ao valor esperado que fixamos ($L = 1$), e a variação observada não parece estar relacionada com o número de lados utilizado. Contudo, para confirmarmos esta ideia, simulamos $r = 10000$ réplicas de cada uma destas situações. Na Tabela 8.10 apresentamos um resumo dos resultados obtidos.

Os valores obtidos na simulação são compatíveis com os que esperávamos. Com esta correcção o valor esperado do quadrado da distância, independentemente do valor de n , será constante. Na simulação fixamos o valor pretendido na unidade (i.e. $L = 1$) e as médias obtidas estão todas próximas do valor teórico.

Salientemos que, com este modo de convergência, não estamos nas condi-

n	10	100	1000	10000
Unif., $k = 1$ Média	1.02377	1.02952	0.97736	1.04574
Unif., $k = 1$ Desv. Padrão	0.928878	1.08778	0.996464	1.029
Unif., $k = 9$ Média	0.998103	1.00133	1.01895	1.0098
Unif., $k = 9$ Desv. Padrão	0.054204	0.413321	0.907097	0.998602
Unif., $k = 360$ Média	1.00000	1.00000	1.00003	1.00127
Unif., $k = 360$ Desv. Padrão	0.000025	0.000359	0.003915	0.036109
Beta, $k = 1$ Média	0.990779	0.936404	0.953387	1.02671
Beta, $k = 1$ Desv. Padrão	0.84734	0.939707	0.997909	1.04152
Beta, $k = 9$ Média	0.999166	0.988368	1.00386	1.02963
Beta, $k = 9$ Desv. Padrão	0.035060	0.291304	0.859066	1.05594
Beta, $k = 360$ Média	1.00000	1.00001	0.999953	0.999144
Beta, $k = 360$ Desv. Padrão	0.000021	0.000224	0.000232	0.024278

Tabela 8.10: Simulação com distância esperada fixa

ções de lançamento de uma curva flexível e inextensível definidas por Pacheco d'Amorim, pois, apesar de o comprimento dos lados dos polígonos tender para zero, o perímetro total vai sempre aumentando, convergindo para infinito. Contudo, usando o método de construção de Pacheco d'Amorim, no limite obtemos um ponto que tem perímetro nulo e, por tal razão, não concordamos que o resultado no limite tenha o mesmo perímetro inicial. Além disso, o autor não mostra qual é o seu objectivo quando resolve este problema. Assim, caso o objectivo fosse obter um processo estocástico, o autor deveria ter utilizado outro enunciado, pois, nas condições do problema, vamos obter unicamente um ponto. Na secção 8.4.4 vamos verificar que o movimento browniano, ou processo de Wiener, que corresponde a um dos processos estocásticos mais utilizados na Física, pode ser obtido como um passeio aleatório

onde vamos diminuindo o comprimento de cada passo. A formalização matemática deste processo, recorrendo a conceitos da Teoria da Medida, surge em 1923 nos trabalhos de Wiener, sendo posterior à tese de doutoramento de Pacheco d'Amorim. Todavia, Bachelier (1900), numa obra notável mas ignorada durante muito tempo, já tinha deduzido diversas propriedades que caracterizam este processo e demonstra, inclusive, na sua tese de doutoramento de 1900, que podemos deduzir o movimento browniano através de um passeio aleatório⁽¹¹⁾.

8.4.3 Passeios aleatórios em reticulados

Analiseemos um problema semelhante ao investigado por Pacheco d'Amorim, mas onde o passeio é efectuado num reticulado. De uma forma geral este problema corresponde a considerar um espaço $\mathbb{Z}^p$, com $p \in \mathbb{N}$, onde uma partícula se encontra na origem no início (por simplificação, pois a localização inicial será indiferente para as conclusões que iremos apresentar) e se movimenta, em cada iteração, para umas das $2p$ localizações que lhe são vizinhas, tendo probabilidade igual a $\frac{1}{2p}$ de se deslocar para cada uma dessas possibilidades. Na Figura 8.9 apresentamos algumas simulações para o caso de $p = 2$, onde temos um passeio aleatório num plano e, em cada iteração, a partícula tem quatro possibilidades de destino, cada uma com probabilidade $\frac{1}{4}$. Chung

⁽¹¹⁾ Courtault *et al.* (2000) apresentam um resumo das principais ideias defendidas na tese de doutoramento de Bachelier, referindo: “*The thesis can be viewed as the origin of mathematical finance and of several important branches of stochastic calculus such as the theory of Brownian motion, Markov processes, diffusion processes, and even weak convergence in functional spaces. Of course, the reasoning was not rigorous but it was, on the intuitive level, basically correct.*”. Este artigo tem, como anexo, o relatório de avaliação da tese de doutoramento de Bachelier, escrita pelo seu orientador Henri Poincaré, o qual demonstra o quanto Poincaré apreciou este trabalho.

(2000, p. 195) refere que os passeios aleatórios num reticulados de dimensão unitária são semelhantes ao problema da ruína do jogador analisado por Laplace. Uma vez mais se verifica que os passeios aleatórios, tal como muitos conceitos na Teoria da Probabilidade, têm origem nos jogos de azar. Huygens (1657), com base nos problemas presentes na famosa correspondência entre Fermat e Pascal em 1654, introduz e explora o conceito de esperança matemática na resolução de uma colectânea de problemas. No final deste opúsculo o autor propõe 5 problemas para o leitor resolver, sendo o último o famoso problema da ruína do jogador. Os jogos propostos por Huygens eram considerados, nos séculos XVII e XVIII, um desafio para muitos matemáticos, tendo surgido diversas generalizações de alguns problemas. Edwards (1983) e Hald (2003) expõem as primeiras versões e resoluções deste problema, nomeadamente os estudos de Pascal, Fermat, Huygens, Montmort, Bernoulli, de Moivre, entre outros. Bachelier (1912) efectua uma análise exhaustiva de várias probabilidades e valores esperados associados a este problema nas suas diversas facetas. Uspensky (1937) e Feller (1968) realizam também um estudo pormenorizado do problema da ruína do jogador.

Pólya em 1921⁽¹²⁾ demonstrou que a probabilidade de um passeio aleatório num reticulado de dimensão dois (ou um) regressar ao ponto inicial (ser recursivo) é igual à unidade e, se o reticulado tiver dimensão superior à segunda, então esta probabilidade será menor que a unidade. Contudo, a probabilidade de, após n iterações (com n fixo *a priori*), a partícula se situar no ponto de partida vai diminuindo, tendendo para zero, à medida que o número de iterações n aumenta. Analisemos melhor esta questão para um passeio aleatório num reticulado de dimensão p .

⁽¹²⁾ Consultar, por exemplo, Chung (2000). Durrett (1986) refere que Pólya descobriu a resposta a este problema enquanto passeava num parque perto de Zurique e estranhou estar sempre a encontrar o mesmo casal. Feller (1968) também analisa este problema.

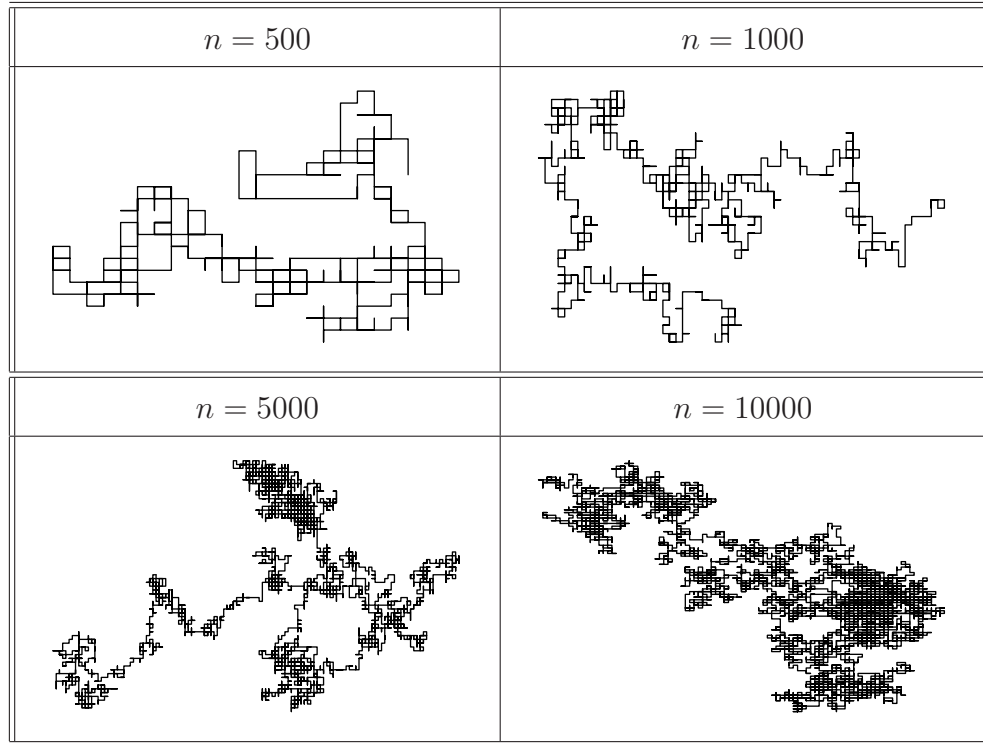


Figura 8.9: Passeios aleatórios em reticulados

Consideremos o acontecimento $\mathbf{A}$ que representa a partícula voltar alguma vez, durante o passeio aleatório, ao ponto de partida (ser recorrente); os acontecimentos $\mathbf{A}_k$ (com $k = 0, 1, \dots$) que correspondem a que a partícula se situe na casa de partida (origem, por hipótese) na iteração k , cuja probabilidade será representada por p_k (naturalmente $p_0 = 1$); e os acontecimentos $\mathbf{A}_k^{(1)}$, com probabilidade representada por $p_k^{(1)}$, que representam o passeio aleatório voltar pela primeira vez ao ponto de partida na iteração k (por conveniência considera-se $p_0^{(1)} = 0$). Com estas definições temos

$$\mathbf{A}_k^{(1)} = \mathbf{A}_k \cap \left(\bigcap_{i=1}^{k-1} \overline{\mathbf{A}_i} \right) \quad (8.115)$$

e

$$\mathbf{A} = \bigcup_{k=1}^{\infty} \mathbf{A}_k^{(1)} \quad (8.116)$$

e, conseqüentemente, uma vez que os acontecimentos $\mathbf{A}_k^{(1)}$ são disjuntos, te-

remos

$$\mathbb{P}(\mathbf{A}) = \sum_{k=0}^{+\infty} \mathbb{P}(\mathbf{A}_k^{(1)}) = \sum_{k=0}^{+\infty} p_k^{(1)}. \quad (8.117)$$

Seja $\mathbf{P}_{\mathbf{A}_k}(s)$ a função geradora de probabilidades dos acontecimentos $\mathbf{A}_k$ definida por

$$\mathbf{P}_{\mathbf{A}_k}(t) = \sum_{k=0}^{+\infty} t^k \mathbb{P}(\mathbf{A}_k) = \sum_{k=0}^{+\infty} t^k p_k, \quad |t| < 1 \quad (8.118)$$

onde $\sum_{k=0}^{+\infty} p_k$, uma vez que as probabilidades p_k não formam uma distribuição de probabilidade, pode assumir qualquer valor positivo e até convergir para infinito. Podemos interpretar p_k como o valor esperado de uma variável aleatória que assume o valor um, se após k iterações a partícula se situar na posição inicial, e zero, caso contrário, pelo que $\sum_{k=1}^{+\infty} p_k$ corresponderá ao número esperado de vezes que a partícula passará na posição inicial. Consideremos também a função geradora de probabilidades $\mathbf{P}_{\mathbf{A}_k^{(1)}}(t)$ dos acontecimentos $\mathbf{A}_k^{(1)}$, isto é,

$$\mathbf{P}_{\mathbf{A}_k^{(1)}}(t) = \sum_{k=0}^{+\infty} t^k \mathbb{P}(\mathbf{A}_k^{(1)}) = \sum_{k=0}^{+\infty} t^k p_k^{(1)}, \quad |t| \leq 1, \quad (8.119)$$

onde $\sum_{k=0}^{+\infty} p_k^{(1)} \leq 1$, uma vez que os acontecimentos $\mathbf{A}_k^{(1)}$ são disjuntos e, nas situações em que haja uma probabilidade não nula de a partícula não regressar à origem (de ser um passeio não recorrente), assume um valor inferior à unidade. Por este motivo a função $\mathbf{P}_{\mathbf{A}_k^{(1)}}(t)$ está definida para $t = 1$, pois neste caso teremos

$$\mathbf{P}_{\mathbf{A}_k^{(1)}}(1) = \sum_{k=0}^{+\infty} p_k^{(1)} = \mathbb{P}(\mathbf{A}). \quad (8.120)$$

Notemos que, das definições de $\mathbf{P}_{\mathbf{A}_k^{(1)}}(t)$ e de $\mathbf{P}_{\mathbf{A}_k}(t)$, podemos concluir que

$$\begin{aligned} & \mathbf{P}_{\mathbf{A}_k}(t) \mathbf{P}_{\mathbf{A}_k^{(1)}}(t) = \\ &= p_0^{(1)} p_0 + \left(p_0^{(1)} p_1 + p_1^{(1)} p_0 \right) t + \left(p_2^{(1)} p_0 + p_1^{(1)} p_1 + p_0^{(1)} p_2 \right) t^2 + \cdots + \\ &+ t^n \sum_{i=0}^n p_i^{(1)} p_{n-i} + \cdots, \end{aligned} \quad (8.121)$$

que, uma vez que $p_0^{(1)} = 0$, pode ser simplificado para

$$\begin{aligned} & \mathbf{P}_{\mathbf{A}_k}(t) \mathbf{P}_{\mathbf{A}_k^{(1)}}(t) = \\ &= p_1^{(1)} p_0 t + \left(p_1^{(1)} p_1 + p_2^{(1)} p_0 \right) t^2 + \cdots + t^n \sum_{i=1}^n p_i^{(1)} p_{n-i} + \cdots \quad (8.122) \end{aligned}$$

Assinalemos agora que as probabilidades p_k , para $k \geq 1$, podem ser expressas através de

$$p_k = p_1^{(1)} p_{k-1} + p_2^{(1)} p_{k-2} + \cdots + p_k^{(1)} p_0 = \sum_{i=1}^k p_i^{(1)} p_{k-i}, \quad (8.123)$$

então, da comparação destas últimas duas expressões, podemos concluir que

$$\mathbf{P}_{\mathbf{A}_k}(t) \mathbf{P}_{\mathbf{A}_k^{(1)}}(t) = \mathbf{P}_{\mathbf{A}_k}(t) - 1, \quad |t| < 1. \quad (8.124)$$

Nas situações em que se verifica $\sum_{i=0}^{+\infty} p_i < \infty$, a função geradora de probabilidades $\mathbf{P}_{\mathbf{A}_k}(t)$ também está definida para o ponto $t = 1$ e, por consequência, a igualdade (8.124) também é válida para $t = 1$, concluindo-se

$$\mathbf{P}_{\mathbf{A}_k}(1) \mathbf{P}_{\mathbf{A}_k^{(1)}}(1) = \mathbf{P}_{\mathbf{A}_k}(1) - 1. \quad (8.125)$$

Dado que pretendemos saber o valor de $\mathbb{P}(\mathbf{A}) = \mathbf{P}_{\mathbf{A}_k^{(1)}}(1)$, da expressão anterior obtemos

$$\mathbf{P}_{\mathbf{A}_k^{(1)}}(1) = \frac{\mathbf{P}_{\mathbf{A}_k}(1) - 1}{\mathbf{P}_{\mathbf{A}_k}(1)} < 1, \quad (8.126)$$

concluindo-se que $\mathbb{P}(\mathbf{A}) < 1$ se $\sum_{i=0}^{\infty} p_i < \infty$. Nos casos em que se verifica $\sum_{i=0}^{\infty} p_i = \infty$, uma vez que $\mathbf{P}_{\mathbf{A}_k}(t)$ não está definida para $t = 1$, podemos determinar $\mathbb{P}(\mathbf{A})$ através do limite quando t se aproxima da unidade (por valores inferiores), pois, nesta situação, $\mathbf{P}_{\mathbf{A}_k}(1)$ será crescente e, considerando um valor de n elevado, teremos

$$\sum_{i=0}^n p_i = \lim_{t \uparrow 1} \sum_{i=0}^n t^i p_i \leq \lim_{t \uparrow 1} \mathbf{P}_{\mathbf{A}_k}(1) = \sum_{i=0}^{+\infty} p_i = +\infty. \quad (8.127)$$

Por outro lado, pela expressão (8.124), podemos deduzir que

$$\lim_{t \uparrow 1} \mathbf{P}_{\mathbf{A}_k}(t) = \lim_{t \uparrow 1} \frac{1}{1 - \mathbf{P}_{\mathbf{A}_k^{(1)}}(t)} \quad (8.128)$$

que, para ser infinito, terá que acontecer $\lim_{t \uparrow 1} \mathbf{P}_{\mathbf{A}_k^{(t)}}(1) = 1$. Desta forma concluimos que

$$\mathbb{P}(\mathbf{A}) = \mathbf{P}_{\mathbf{A}_k^{(1)}}(1) = \sum_{i=1}^{+\infty} p_i^{(1)} \begin{cases} = 1 & \text{se } \sum_{i=0}^{+\infty} p_i = \infty \\ < 1 & \text{se } \sum_{i=0}^{+\infty} p_i < \infty \end{cases}. \quad (8.129)$$

Ainda que esta conclusão seja válida para um passeio aleatório num reticulado de qualquer dimensão, vamos restringir a análise à dimensão dois. Consideremos então um passeio aleatório num reticulado plano e analisemos as probabilidades p_i nesta situação. Suponhamos que foram dados $2n$ passos⁽¹³⁾; então, para que a partícula esteja na casa inicial, deve ter dado tantos passos para a direita como para a esquerda (sejam m) e tantos passos para cima como para baixo (sejam $n - m$), sendo a probabilidade pretendida dada por

$$\begin{aligned} p_{2n} &= \mathbb{P}(\mathbf{A}_{2n}) = \sum_{m=0}^n \frac{2n!}{m! m! (n-m)! (n-m)!} \left(\frac{1}{4}\right)^{2n} = \\ &= 4^{-2n} \binom{2n}{n} \sum_{m=0}^n \binom{n}{m} \binom{n}{n-m} = 4^{-2n} \binom{2n}{n}^2 \end{aligned} \quad (8.130)$$

obtendo-se, pela fórmula de Stirling, como valor aproximado

$$2^{-4n} \binom{2n}{n}^2 \approx 2^{-4n} \left(\frac{\sqrt{4\pi n} (2n)^{2n} e^{-2n}}{\sqrt{2\pi n} n^n e^{-n} \sqrt{2\pi n} n^n e^{-n}} \right)^2 = \frac{1}{\pi n}. \quad (8.131)$$

Deste modo, uma vez que

$$\sum_{i=0}^{+\infty} p_i = \sum_{k=0}^{+\infty} \mathbb{P}(\mathbf{A}_k) = \sum_{k=0}^{+\infty} \mathbb{P}(\mathbf{A}_{2k}) = +\infty, \quad (8.132)$$

pelos resultados apresentados na expressão (8.129), podemos concluir que a partícula vai voltar ao ponto de origem com probabilidade igual a um.

⁽¹³⁾ Se o número de iterações for ímpar, então a probabilidade de a partícula se situar na posição inicial é nula.

Refira-se que, de forma análoga, podemos demonstrar que, para um passeio num reticulado de dimensão um, em $\mathbb{Z}$, a probabilidade de a partícula voltar a passar pelo ponto de partida também será igual a um, enquanto para um passeio aleatório num reticulado de dimensão 3 ou superior, em $\mathbb{Z}^p$ com $p \geq 3$, a probabilidade será inferior à unidade, resultado que levou, segundo Durrett (1996, p. 186), o matemático japonês Shizuo Kakutani (1911–2004) a efectuar o seguinte comentário durante um colóquio.

“A drunk man will find his way home but a drunk bird may get lost forever”

[Shizuo Kakutani]

Este resultado, de certa forma, parece ir ao encontro das conclusões de Pacheco d’Amorim, pois garante que, para os passeios aleatórios efectuados num reticulado de dimensão dois (incluído num plano), a partícula regressa ao ponto de partida com probabilidade um. Podemos ainda demonstrar que a partícula regressa ao ponto de partida infinitas vezes com probabilidade um.

Consideremos os acontecimentos $\mathbf{A}_k^m$ que representam a partícula regressar à posição inicial pelo menos k vezes em m iterações, e o acontecimento $\mathbf{A}^\infty$ que representa a partícula regressar à posição de partida infinitas vezes, isto é

$$\mathbf{A}^\infty = \{\mathbf{A}_n \text{ i.o.}\} = \bigcap_{i=1}^{\infty} \bigcup_{k=i}^{\infty} \mathbf{A}_k, \quad (8.133)$$

então teremos

$$\mathbb{P}(\mathbf{A}^\infty) = \begin{cases} 1 & \text{se } \sum_{i=0}^{+\infty} p_i = +\infty \\ 0 & \text{se } \sum_{i=0}^{+\infty} p_i < +\infty \end{cases}. \quad (8.134)$$

Na situações em que $\sum_{i=0}^{\infty} p_i < \infty$ teremos

$$\begin{aligned} \mathbb{P}(\mathbf{A}^{\infty}) &= \mathbb{P}\left(\bigcap_{i=1}^{\infty} \bigcup_{k=i}^{\infty} \mathbf{A}_k\right) = \lim_{i \rightarrow \infty} \mathbb{P}\left(\bigcup_{k=i}^{\infty} \mathbf{A}_k\right) \leq \\ &\leq \lim_{i \rightarrow \infty} \sum_{k=i}^{\infty} \mathbb{P}(\mathbf{A}_k) = 0, \end{aligned} \quad (8.135)$$

situação que também poderíamos ter concluído com recurso ao Lema de Borel-Cantelli (consultar Lema 7.2 presente na página 430). No caso em que temos $\sum_{i=0}^{\infty} p_i = \infty$ não podemos recorrer ao Lema de Borel-Cantelli, pois, para tal, seria necessária a independência dos acontecimentos $\mathbf{A}_k$, que não se verifica. Nestas situações vamos recorrer aos acontecimentos $\mathbf{A}_k^m$ previamente definidos, que simbolizam a partícula regressar pelo menos k vezes à localização inicial durante m iterações. Notemos que

$$\mathbf{A}_k^{mk} \subset \mathbf{A}_k^{\infty}, \quad \forall m \in \mathbb{N} \quad (8.136)$$

logo, podemos concluir que⁽¹⁴⁾

$$[\mathbb{P}(\mathbf{A}_1^m)]^k \leq \mathbb{P}(\mathbf{A}_k^{mk}) \leq \mathbb{P}(\mathbf{A}_k^{\infty}). \quad (8.137)$$

Por outro lado, recorrendo a limites, obtemos

$$\mathbf{A}_1^m = \bigcup_{i=1}^m \mathbf{A}_i \xrightarrow{m \rightarrow \infty} \bigcup_{i=1}^{\infty} \mathbf{A}_i = \mathbf{A} \quad (8.138)$$

e, consequentemente, teremos

$$\mathbb{P}(\mathbf{A}_k^{\infty}) \geq [\mathbb{P}(\mathbf{A})]^k = 1, \quad (8.139)$$

(14) O acontecimento $\mathbf{A}_k^{mk}$ significa que a partícula regressou à posição inicial pelo menos k vezes em mk iterações. Notemos que uma das formas de ocorrer $\mathbf{A}_k^{mk}$ é dividir as mk iterações em k grupos de m iterações e considerar que em cada um destes grupos a partícula regressou à posição inicial pelo menos uma vez (isto é ocorrer $\mathbf{A}_1^m$ em cada grupo), sendo a probabilidade de isto ocorrer, uma vez que estamos a analisar iterações disjuntas, igual a $[\mathbb{P}(\mathbf{A}_1^m)]^k$. Há outras formas de obter o acontecimento $\mathbf{A}_k^{mk}$, logo, podemos concluir que $[\mathbb{P}(\mathbf{A}_1^m)]^k \leq \mathbb{P}(\mathbf{A}_k^{mk})$.

pois, como deduzimos previamente, $\mathbb{P}(\mathbf{A}) = 1$ quando $\sum_{i=0}^{\infty} p_i = \infty$ (cf. expressão (8.129)). Desta forma concluímos que $\mathbb{P}(\mathbf{A}_k^{\infty}) = 1$ para qualquer valor $k \in \mathbb{N}$ se $\sum_{i=0}^{\infty} p_i = \infty$ e, como

$$\mathbf{A}^{\infty} = \bigcap_{k=1}^{\infty} \mathbf{A}_k^{\infty}, \quad (8.140)$$

podemos concluir que, para a situação $\sum_{i=0}^{\infty} p_i = \infty$, temos $\mathbb{P}(\mathbf{A}^{\infty}) = 1$.

Com este resultado podemos deduzir que $\mathbb{P}(\mathbf{A}^{\infty})$ é igual a um para passeios aleatórios em reticulados de dimensão um⁽¹⁵⁾ ou dois e será nula para passeios em reticulados de dimensão superior. Deste resultado podemos concluir que, em relação à chalaça de Kakutani, um homem embriagado não só encontra o caminho de casa como, caso passe por ela sem se aperceber ou lhe apeteça passear um pouco mais para apanhar ar, ainda voltará certamente a encontrar a casa.

Apesar de estes resultados, em certa medida, parecerem ir ao encontro das conclusões de Pacheco d'Amorim, podemos também deduzir, pela probabilidade obtida na expressão (8.131), que a probabilidade aproximada de a partícula, após um determinado número n de iterações, se situar no ponto inicial converge para zero quando aumentamos o número de iterações. Como tal, se considerarmos um polígono formado pelos lados que caracterizam o passeio da partícula e tivermos um número n elevado de iterações, a probabilidade de a partícula se situar exactamente no ponto de partida, isto é, de o polígono assim formado ser fechado, será um infinitésimo, o que, de certa forma, contraria a ideia de que, se lançar à sorte uma curva flexível ela terá probabilidade igual a um de ser uma curva fechada. Refira-se que, nesta análise num reticulado, assumimos as direcções de cada movimento da

⁽¹⁵⁾ Um passeio aleatório num reticulado de dimensão um é equivalente a considerar a soma S_n das variáveis aleatórias independentes X_i , com $i \in \mathbb{N}$, que assumem o valor -1 com probabilidade $\frac{1}{2}$ e o valor 1 com probabilidade $\frac{1}{2}$.

partícula como independentes, tal como no problema enunciado por Pacheco d'Amorim; todavia, em vez de o número de sentidos possíveis ser em número infinito não numerável, que constitui a região $[0, 2\pi)$, os sentidos são apenas quatro, o que torna a probabilidade de voltar ao ponto de origem superior, pois esta probabilidade é decrescente com o número de direcções distintas utilizadas. Outra diferença entre estes dois problemas é que, enquanto no problema de Pacheco d'Amorim o comprimento dos lados vai diminuindo de forma que o perímetro total se mantenha imutável, no passeio aleatório da partícula no reticulado de dimensão dois os movimentos têm sempre a mesma amplitude e, quantas mais iterações houver, maior é o perímetro total percorrido pela partícula. Apesar de ser impossível comparar os resultados obtidos por estes dois problemas, a resolução deste problema de Pólya não deixa de justificar a nossa estranheza quando lemos pela primeira vez o resultado apresentado por Pacheco d'Amorim.

8.4.4 O movimento browniano

O movimento browniano foi descoberto pelo biólogo Robert Brown em 1827 enquanto investigava o movimento das partículas de pólen a flutuar na água através de um microscópio, verificando que estas estavam constantemente num movimento caótico. Inicialmente Brown considerava que os movimentos teriam vontade própria, isto é, que as partículas de pólen teriam vida e se movimentavam por si, sem que nenhuma força as empurrasse. Contudo, ao efectuar a mesma experiência com pó, obteve os mesmos resultados, tendo colocado de lado a hipótese de as partículas de pólen serem uma forma de vida. A justificação destes movimentos caóticos só foi explicada quase 80 anos depois. A primeira formalização matemática deste movimento, apesar de inserida num contexto distinto, devemos-la a Louis Bachelier (1870–1946), que na sua tese de doutoramento, publicada em 1900 e intitulada “*Théorie*

de la spéculation”, realizada sob a orientação de Henri Poincaré, deduz um processo estocástico contínuo (que corresponde ao movimento browniano de uma dimensão) para descrever as flutuações da bolsa de Paris. Finalmente em 1905, Albert Einstein (1879–1955) explicou o movimento da partícula de pólen fundamentando-se na teoria molecular. Segundo esta teoria, as pequenas partículas movimentam-se devido às colisões com as moléculas de água que estão em constante movimento, isto é, o movimento browniano é causado pelos choques das moléculas do fluido, pois estes choques, por serem em grande número e independentes, provocam o movimento aleatório das partículas. Finalmente, na década de 1920, o matemático Norbert Wiener (1894–1964) formaliza, recorrendo já a conceitos da Teoria da Medida, a actualmente denominada teoria do movimento browniano (ou processo de Wiener), que descreve matematicamente o movimento aleatório de uma partícula.

Um processo estocástico $\{\mathbf{W}(t), t \geq 0\}$, onde t habitualmente representa o tempo, é um movimento browniano (ou processo de Wiener) se verificar as seguintes condições:

1. a partícula situa-se na origem no início do processo, i.e. $\mathbf{W}(0) = 0$;
2. o processo $\mathbf{W}(t)$ possui incrementos independentes, isto é, para quaisquer tempos $t_0, t_1, \dots, t_n$, tais que $t_0 < t_1 < \dots < t_n$, as variáveis aleatórias $\mathbf{W}(t_i) - \mathbf{W}(t_{i-1})$, para $i = 1, \dots, n$, são independentes;
3. as variáveis aleatórias $\mathbf{W}(t_i) - \mathbf{W}(t_{i-1})$ são caracterizadas por uma distribuição gaussiana com média nula e variância $\sigma^2 = k(t_i - t_{i-1})$, onde k é uma constante qualquer positiva.⁽¹⁶⁾

⁽¹⁶⁾ Não é obrigatório o valor esperado ser nulo, pois o processo poderá ter uma tendência. Todavia, vamos restringir a nossa análise a esta situação.

O incremento de $\mathbf{W}(t_i) - \mathbf{W}(t_{i-1})$ pode ser interpretado como a soma de um elevado número de pequenos deslocamentos aleatórios e, por esta razão, podemos aplicar o Teorema Limite Central, obtendo-se a distribuição gaussiana com valor esperado nulo e variância proporcional ao tempo t em que analisamos o processo. Por este motivo um processo de Wiener pode ser obtido através de um passeio aleatório de dimensão 1, onde o número de iterações aumenta indefinidamente e o comprimento de cada passo tende para zero.

Para entendermos melhor esta ideia, comecemos por analisar um passeio aleatório de uma dimensão. Seja X_i uma sucessão de variáveis aleatórias independentes tais que

$$\mathbb{P}[X_i = -1] = \mathbb{P}[X_i = 1] = \frac{1}{2}, \quad i = 1, \dots, n, \quad (8.141)$$

sendo

$$\mathbf{S}_n = \sum_{i=1}^n \Delta_x X_i \quad (8.142)$$

um passeio aleatório de dimensão um onde Δ_x corresponde ao comprimento de cada passo dado pela partícula em cada iteração. Consideremos agora o processo estocástico contínuo $\{\mathbf{W}(t), t \geq 0\}$ que, sempre que o tempo t aumenta uma quantidade Δ_t , o processo $\mathbf{W}(t)$ aumenta ou diminui, com probabilidade $\frac{1}{2}$, uma determinada quantidade Δ_x (o processo está constante nos restantes valores de t que não são múltiplos de Δ_t). No período $t = n\Delta_t$ o processo é dado por

$$\mathbf{W}(t) = \sum_{i=1}^n X_i \Delta_x = \mathbf{S}_n, \quad (8.143)$$

que corresponde ao passeio aleatório previamente apresentado. Notemos que

$$\sigma_{\mathbf{W}(t)}^2 = \sigma_{\mathbf{S}_n}^2 = \Delta_x^2 \sum_{i=1}^n \sigma_{X_i}^2 = n\Delta_x^2 = \frac{t}{\Delta_t} \Delta_x^2. \quad (8.144)$$

Analisemos então o que acontece se fizermos Δ_x e Δ_t tenderem para zero, considerando um valor de t fixo, onde Δ_t convergir para zero implicará que n

aumente indefinidamente. Notemos que convém, para um valor qualquer de t fixo, que a variância resultante não seja nula, pois deixaríamos de ter um processo estocástico. Além disso, a variância deverá ser finita e proporcional ao tempo t , isto é, $\sigma_{\mathbf{W}(t)}^2 = kt$. Assim, da relação

$$\lim_{\substack{\Delta_x \rightarrow 0 \\ \Delta_t \rightarrow 0}} \frac{t}{\Delta_t} \Delta_x^2 = kt, \quad (8.145)$$

concluimos que devemos escolher, para o comprimento dos lados, $\Delta_x = \sqrt{k\Delta_t}$. Uma vez que, quando Δ_t tende para zero n tenderá para infinito, pelo Teorema Limite Central concluimos que $\mathbf{W}(t)$, para t fixo, é caracterizado aproximadamente por uma distribuição gaussiana com valor esperado nulo e variância kt com $k > 0$. Salientemos que este processo, resultante de tal limite, verifica as condições de um processo de Wiener previamente apresentadas, pois, se dividirmos o tempo em intervalos disjuntos através de $t_0 < t_1 < \dots < t_n$ facilmente concluimos que os incrementos $\mathbf{W}(t_i) - \mathbf{W}(t_{i-1})$ são independentes (resultam de somas de variáveis X_i independentes), a distribuição de $\mathbf{W}(t_i) - \mathbf{W}(t_{i-1})$ corresponde a uma distribuição gaussiana com média nula e variância $\sigma^2 = k(t_i - t_{i-1})$ com $k > 0$.

Deste modo, podemos visualizar o movimento browniano (processo de Wiener) como o limite de um passeio aleatório onde em qualquer espaço de tempo t , por mais pequeno que t seja, teremos sempre um grande número de passos independentes de pequeno comprimento Δ_x , que nos garante, pelo Teorema Limite Central, a convergência para a lei de gauss.

Bachelier (1900, 1912) já tinha considerado ideia semelhante. Representando por Δ_x a variação do valor dos prémios em cada período de tempo, Bachelier considerou que os prémios aumentavam Δ_x com probabilidade $p = \frac{1}{2}$ e diminuía Δ_x com probabilidade $q = \frac{1}{2}$ em cada período de tempo infinitesimal Δ_t , onde $t = n\Delta_t$ representa o período de tempo em que estamos a analisar o valor do prémio. Então, pelo Teorema Limite Central, uma vez que

para t fixo podemos considerar um Δ_t pequeno de forma que n assumira um valor elevado, teremos que a probabilidade de o prémio no período t assumir um valor compreendido entre x e $x + dx$ é dado por

$$\frac{1}{\sqrt{2\pi npq}} e^{-\frac{x^2}{2npq}} dx = \frac{2\sqrt{\Delta_t}}{\sqrt{2\pi}\sqrt{t}} e^{-\frac{2x^2\Delta_t}{t}} dx, \quad (8.146)$$

onde a função de instabilidade⁽¹⁷⁾ $\varphi(t)$ é dada por $\frac{t}{2\Delta_t}$. A esperança matemática de x , que para Bachelier corresponde ao valor esperado dos valores positivos, i.e. $\int_{\mathbb{R}^+} x \mathbb{P}_{\mathbb{R}}(x) dx$, é dada por

$$\int_0^{+\infty} \frac{2x\sqrt{\Delta_t}}{\sqrt{2\pi}\sqrt{t}} e^{-\frac{2x^2\Delta_t}{t}} dx = \frac{\sqrt{t}}{2\sqrt{2\pi}\sqrt{\Delta_t}}. \quad (8.147)$$

Bachelier considera que o mercado deve obedecer ao princípio de uniformidade (cf. Bachelier (1912, p. 286)), segundo o qual todos os elementos do tempo t devem contribuir de igual forma para a instabilidade, isto é, considerando

$$\varphi(t_1) = \int_0^{t_1} \varphi'(t) dt, \quad (8.148)$$

a função $\varphi'(t)$ deverá assumir o valor constante $4\pi k^2$ de forma que a sua esperança matemática⁽¹⁸⁾ seja

$$\frac{\sqrt{\varphi(t)}}{2\sqrt{\pi}} = k\sqrt{t}. \quad (8.149)$$

Este princípio surge na teoria do jogo, área a que Bachelier dedica uma boa parte da sua obra (leia-se, por exemplo, Bachelier(1901)), onde a função de

⁽¹⁷⁾ Bachelier define função de instabilidade por $2(\mathbb{E}^2(X) - \mathbb{E}(X^2))$, isto é, corresponde ao dobro da variância.

⁽¹⁸⁾ Recordemos que, se a variável aleatória X for caracterizada por uma distribuição gaussiana com valor médio nulo e variância σ_x^2 , então a sua esperança matemática (no sentido de Bachelier) será dada por $\frac{\sigma_x}{\sqrt{2\pi}}$.

instabilidade é proporcional ao número de jogos. Notemos que em n provas de Bernoulli teremos

$$\varphi = 2npq \Leftrightarrow \frac{\sqrt{\varphi}}{2\sqrt{\pi}} = \frac{\sqrt{pq}}{\sqrt{2\pi}}\sqrt{n} = k\sqrt{n}. \quad (8.150)$$

Desta forma Bachelier considera que a esperança matemática (expressão (8.147)) deverá ser igual a $k\sqrt{t}$ e, conseqüentemente, teremos

$$\frac{\sqrt{t}}{2\sqrt{2\pi}\sqrt{\Delta_t}} = k\sqrt{t} \Leftrightarrow \Delta_t = \frac{1}{2k\sqrt{2\pi}}, \quad (8.151)$$

que, substituindo na expressão (8.146), obtemos

$$\frac{1}{2\pi k\sqrt{t}} e^{-\frac{x^2}{4\pi tk^2}} dx, \quad (8.152)$$

que corresponde à lei de probabilidade da distribuição gaussiana com média nula e variância $\sigma^2 = 2\pi k^2 t = k't$. Bachelier obteve, deste modo, um processo de Wiener através de um passeio aleatório. Refira-se, como curiosidade, que Bachelier apresenta esta dedução após ter alcançado o processo de Wiener através das propriedades Markovianas. Se representarmos por $p_{x,t} dx$ a probabilidade de, no período t , o processo variar uma quantidade situada no intervalo $(x, x + dx)$, Bachelier (1900, p. 35) conclui que

$$p_{z,t_1+t_2} = \int_{-\infty}^{+\infty} p_{x,t_1} p_{z-x,t_2} dx, \quad (8.153)$$

que corresponde à propriedade fundamental dos processos de Markov, que segundo Kahane (1998), só foi introduzida por Markov em 1907 para os processos com tempo discreto⁽¹⁹⁾.

Feller (1968, p. 354) refere, em relação à dedução de um processo browniano através da passagem ao limite de um passeio aleatório, que esta estreita ligação entre o Processo de Wiener e o passeio aleatório fornece uma

⁽¹⁹⁾ Kahane (1998) apresenta uma análise à origem da fundamentação matemática do movimento browniano de 1900 até 1950, focando trabalhos de Bachelier, Borel, Einstein, Kolmogoroff, Lévy, Pólya, Steinhaus, Wiener, entre outros.

enorme contribuição para a compreensão de ambos os processos, acrescentando que esta ligação “*was fully exploited (though in a heuristic manner) by L. Bachelier, whose work has inspired A. Kolmogorov to develop the formal foundations of Markov processes*”.

Uma vez que Pacheco d’Amorim lança o polígono num plano, analisemos agora o caso bidimensional. O processo estocástico $\mathbf{W}(t) = (\mathbf{W}_1(t), \mathbf{W}_2(t))$, com $t \in [0, +\infty)$, é um processo bidimensional de Wiener se $\mathbf{W}_1(t)$ e $\mathbf{W}_2(t)$ são dois processos de Wiener independentes de uma dimensão. Assim, podemos concluir que um passeio aleatório em duas dimensões (cf. definido em (8.154)), com passos cada vez mais pequenos, tenderá para um processo de Wiener bidimensional. Esta conclusão tem como base, uma vez mais, o Teorema Limite Central, pois, em cada período de tempo t , teremos a soma de um grande número de pequenos passos da partícula, passos estes que são independentes e identicamente distribuídos. Deste modo, consideremos o passeio aleatório de dimensão dois descrito por

$$\mathbf{S}_n^{(2)} = (\mathbf{S}_n^{(x)}, \mathbf{S}_n^{(y)}) = \left(\sum_{i=1}^n \frac{\Delta_i}{\sqrt{2}} X_i, \sum_{i=1}^n \frac{\Delta_i}{\sqrt{2}} Y_i \right) \quad (8.154)$$

onde Δ_i corresponde ao comprimento de cada passo dado pela partícula⁽²⁰⁾ em cada iteração, X_i e Y_i são duas sucessões de variáveis aleatórias independentes que assumem o valor 1 e -1 com probabilidade $\frac{1}{2}$. Se efectuarmos raciocínio análogo ao que empregamos na construção do processo de uma

⁽²⁰⁾ Notemos que este passeio aleatório não corresponde ao usual passeio aleatório num reticulado de dimensão dois, em $\mathbb{Z}^2$, pois enquanto num passeio aleatório num reticulado a partícula dá passos de comprimento unitário na direcção dos eixos (em cada iteração uma e uma só das coordenadas aumenta ou diminui uma unidade), neste processo a partícula dá passos de comprimento Δ_i na direcção de uma das bissectrizes dos eixos (em cada iteração as duas coordenadas obrigatoriamente se alteram, de forma independente, em $\pm \frac{\Delta_i}{\sqrt{2}}$ percorrendo a partícula uma distância igual a Δ_i).

dimensão, concluiremos que, para um período de tempo t fixo, teremos

$$\mathbf{W}(t) = (\mathbf{W}_1(t), \mathbf{W}_2(t)) = \left(\sum_{i=1}^n \frac{\Delta_i}{\sqrt{2}} X_i, \sum_{i=1}^n \frac{\Delta_i}{\sqrt{2}} Y_i \right) = \mathbf{S}_n^{(2)}. \quad (8.155)$$

Desta expressão podemos deduzir que os processos $\mathbf{W}_i(t)$ ($i = 1, 2$), para cada espaço de tempo t , são obtidos pela soma de n variáveis independentes $\frac{\Delta_i}{\sqrt{2}} X_i$ (ou $\frac{\Delta_i}{\sqrt{2}} Y_i$), que assumem o valor $\frac{\Delta_i}{\sqrt{2}}$ e $-\frac{\Delta_i}{\sqrt{2}}$ com probabilidade $\frac{1}{2}$. Deste modo a variância de cada um dos processos de dimensão um será dado por

$$\sigma_{\mathbf{W}_1(t)}^2 = \left(\frac{\Delta_i}{\sqrt{2}} \right)^2 \sum_{i=1}^n \sigma_{X_i}^2 = \frac{n}{2} \Delta_i^2 = \frac{t}{2\Delta_t} \Delta_i^2, \quad (8.156)$$

obtendo-se igual valor para $\sigma_{\mathbf{W}_2(t)}^2$. Assim, tal como na situação de dimensão um, deveremos considerar o comprimento dos lados do tipo $\Delta_i = \sqrt{2k\Delta_t}$, com $k > 0$, para que as variâncias, para um valor de t fixo, verifiquem $0 < \sigma_{\mathbf{W}_i(t)}^2 = kt < \infty$. Notemos que os dois processos $\mathbf{W}_1(t)$ e $\mathbf{W}_2(t)$ são independentes, consequência de as variáveis aleatórias X_i e Y_i serem independentes. Deste modo, concluímos que o processo $\mathbf{W}(t) = (\mathbf{W}_1(t), \mathbf{W}_2(t))$ é um processo de Wiener bidimensional.

Podemos usar argumento idêntico para deduzir o Processo de Wiener através da dobragem de polígonos? Pacheco d'Amorim considera n segmentos $l_1, l_2, \dots, l_n$, com comprimentos respectivamente $\bar{l}_1, \bar{l}_2, \dots, \bar{l}_n$, que formam uma linha poligonal com perímetro total $L = \sum_{i=1}^n \bar{l}_i$, onde os ângulos entre o lado l_i e o eixo das abcissas é uma variável aleatória θ_i , com $i = 1, \dots, n$. As variáveis aleatórias θ_i são independentes e são caracterizadas por uma distribuição uniforme no intervalo $[0, 2\pi]$. Em relação a este problema (Problema 8.2 enunciado na página 489) demonstrámos que, sendo $A_0 = (x_0, y_0)$ o ponto inicial do polígono, a outra extremidade do polígono, após o lançamento de m lados, com $m \leq n$, será o ponto

$$A_m = (x_m, y_m) = \left(x_0 + \sum_{i=1}^m \bar{l}_i \cos(\theta_i), y_0 + \sum_{i=1}^m \bar{l}_i \sin(\theta_i) \right). \quad (8.157)$$

Restrinjamos a nossa análise à situação $A_0 = (0, 0)$ e $\bar{l}_i = \Delta_i$, isto é, quando o polígono começa a ser formado na origem e todos os lados têm igual comprimento. Notemos que podemos interpretar esta situação como se tratasse de um passeio aleatório de uma partícula. Deste modo, consideremos que uma partícula se situa no ponto A_0 no momento inicial e em cada iteração dá um passo de comprimento Δ_i num sentido escolhido aleatoriamente. Assim, considerando que a partícula se situa na origem no início do processo, a localização da partícula após m iterações (passos), com $m \leq n$, será dada por

$$A_m = (x_m, y_m) = \left(\sum_{i=1}^m \Delta_i \cos(\theta_i), \sum_{i=1}^m \Delta_i \sin(\theta_i) \right). \quad (8.158)$$

Deste modo, cada coordenada da localização da partícula é definida pela soma de m variáveis aleatórias independentes, pois teremos

$$A_m = \left(\mathbf{S}_m^{(x)}, \mathbf{S}_m^{(y)} \right). \quad (8.159)$$

Consequentemente, em relação à coordenada da abcissa, podemos definir um processo contínuo $\mathbf{W}_x(t)$ que varia uma quantidade $\Delta_i \cos(\theta_i)$ em cada intervalo de tempo Δ_i , mantendo-se constante nos restantes valores de t que não são múltiplos de Δ_i . Assim, o processo $\mathbf{W}_x(t)$, no período de tempo $t = m\Delta_i$, será descrito por

$$\mathbf{W}_x(t) = \sum_{i=1}^m \Delta_i \cos(\theta_i) = \mathbf{S}_m^{(x)}, \quad (8.160)$$

que tem variância igual a

$$\sigma_{\mathbf{W}_x(t)}^2 = \sigma_{\mathbf{S}_m^{(x)}}^2 = \Delta_i^2 \sum_{i=1}^m \sigma_{\cos(\theta_i)}^2 = \Delta_i^2 \frac{m}{2} = \Delta_i^2 \frac{t}{2\Delta_i}, \quad (8.161)$$

uma vez que $\mathbb{E}[\cos(\theta_i)] = 0$, $\mathbb{E}[\cos^2(\theta_i)] = \frac{1}{2}$ e consequentemente $\sigma_{\cos(\theta_i)}^2 = \frac{1}{2}$. Façamos agora Δ_i e Δ_t convergirem para zero, mantendo t fixo, de forma que a variância do processo convirja para um valor finito não nulo. Notemos que, se utilizarmos $\Delta_i = k\Delta_t$, iremos obter variância nula e, por conseguinte,

o processo será igual a zero com probabilidade um, que corresponde ao que Pacheco d'Amorim fez quando considerou perímetro fixo e lados com igual comprimento (pois $\Delta_l = k\Delta_t \Leftrightarrow \Delta_t = \frac{kt}{m}$). Desta forma, se considerarmos $\Delta_t = k'\sqrt{\Delta_t}$, sendo k' uma constante positiva, concluimos que a variância do processo é dada por

$$\sigma_{\mathbf{W}_x(t)}^2 = \frac{k'^2}{2} t = kt. \quad (8.162)$$

O Teorema Limite Central garante que o processo $\mathbf{W}_x(t)$ seja caracterizado por uma distribuição gaussiana, uma vez que o seu valor, mesmo para valores de t pequenos, resulta sempre de uma soma de muitos termos pequenos.

Notemos agora que $\mathbf{W}_x(t)$ é um processo de Wiener, pois o processo tem incrementos independentes, uma vez que a dois intervalos disjuntos corresponderão sempre duas somas de variáveis independentes e, consequentemente, para quaisquer valores $t_0, t_1, \dots, t_n$, tais que $t_0 < t_1 < \dots < t_n$, as variáveis aleatórias $\mathbf{W}_x(t_i) - \mathbf{W}_x(t_{i-1})$, para $i = 1, \dots, n$, são independentes. A variável $\mathbf{W}_x(t_i) - \mathbf{W}_x(t_{i-1})$ é caracterizada por uma distribuição gaussiana, dado ser obtida pela soma de muitas variáveis independentes e identicamente distribuídas com variância finita, tendo média nula e variância $\sigma^2 = k(t_i - t_{i-1})$ onde k é uma constante positiva (cf. expressão (8.162)).

Naturalmente, para a coordenada da ordenada podemos concluir, de forma análoga, que $\mathbf{W}_y(t)$ é também um processo de Wiener. Serão, no entanto, os processos $\mathbf{W}_x(t)$ e $\mathbf{W}_y(t)$ independentes? Começemos por destacar que os processos $\mathbf{W}_x(t)$ e $\mathbf{W}_y(t)$ têm correlação nula. Uma vez que $\mathbb{E}[\mathbf{W}_x(t)] = \mathbb{E}[\mathbf{W}_y(t)] = 0$, a covariância de $\mathbf{W}_x(t)$ e $\mathbf{W}_y(t)$ é dada por

$$\begin{aligned} \text{Cov}(\mathbf{W}_x(t), \mathbf{W}_y(t)) &= \mathbb{E} \left[\left(\sum_{i=1}^m \Delta_t \cos(\theta_i) \right) \left(\sum_{i=1}^m \Delta_t \sin(\theta_i) \right) \right] = \\ &= \mathbb{E} \left[\Delta_t^2 \sum_{i=1}^m \sum_{j=1}^m \cos(\theta_i) \sin(\theta_j) \right] = \Delta_t^2 \sum_{i=1}^m \mathbb{E}[\cos(\theta_i) \sin(\theta_i)], \end{aligned} \quad (8.163)$$

dado que $\mathbb{E}[\cos(\theta_i) \sin(\theta_j)] = 0$ pela independência das variáveis θ_i e θ_j

para $i \neq j$. Como

$$\mathbb{E}[\cos(\theta_i) \sin(\theta_i)] = \int_0^{2\pi} \frac{\cos(\theta_i) \sin(\theta_i)}{2\pi} d\theta_i = 0, \quad (8.164)$$

concluimos que

$$\mathbb{Cov}(\mathbf{W}_x(t), \mathbf{W}_y(t)) = 0 \quad (8.165)$$

e, por consequência, a correlação também será nula. O facto de a correlação entre duas variáveis ser nula não é suficiente para garantir a independência das variáveis, mas como, após aplicar o limite, $\mathbf{W}_x(t)$ e $\mathbf{W}_y(t)$ são caracterizados por uma distribuição gaussiana, então poderemos concluir que $\mathbf{W}_x(t)$ e $\mathbf{W}_y(t)$ são independentes. Desta forma, concluimos que o processo $\mathbf{W}(t) = (\mathbf{W}_x(t), \mathbf{W}_y(t))$ é um processo de Wiener bidimensional.

A característica que centrou a atenção de Pacheco d'Amorim na sua análise foi o valor esperado do quadrado da distância entre os dois pontos extremos do polígono. Feller (1971, p. 344) refere em relação a um Processo de Wiener que “*The most interesting variable in this processes is the distance $\mathbf{R}(t)$ from the origin ($\mathbf{R}^2 = \mathbf{X}^2 + \mathbf{Y}^2$)*”. Refira-se, como curiosidade, que, para analisar a distância entre os dois pontos extremos, poderíamos recorrer a um processo de Bessel, pois, se definirmos a norma $\|\mathbf{W}(t)\| = \sqrt{\mathbf{W}_1^2(t) + \mathbf{W}_2^2(t)}$, onde $\mathbf{W}_1^2(t)$ e $\mathbf{W}_2^2(t)$ são dois processos de Wiener independentes de dimensão um ($\mathbf{W}(t)$ é um processo de Wiener bidimensional), então o processo estocástico $\{\|\mathbf{W}(t)\|, t \geq 0\}$ será um processo de Bessel de duas dimensões.

Saliente-se, uma vez mais, que os processos aqui descritos não obedecem à definição de lançamento de uma curva flexível e inextensível concebida por Pacheco d'Amorim, pois não fixamos o perímetro total do polígono. Assim, em todas estas situações o perímetro total do polígono vai crescendo, ao contrário da análise de Pacheco d'Amorim, que mantém o perímetro total fixo, ou pelo menos tem esse objectivo. Se, este autor, tivesse utilizado outra forma de convergência, nomeadamente com os n lados a terem comprimento

$\bar{l}_i = \frac{L}{\sqrt{n}}$ ou $\bar{l}_i = L \left[n + \frac{2\rho(n-1)}{(1-\rho)} - \frac{2\rho^2(1-\rho^{n-1})}{(1-\rho)^2} \right]^{-\frac{1}{2}}$ (cf. expressão (8.114)), o perímetro total seria crescente com n , mas, nestes casos, Pacheco d'Amorim teria obtido um processo estocástico em vez de um ponto.

8.5 Problema da agulha de Buffon tratado como valores esperados

O problema da agulha de Buffon foi enunciado na secção 5.1.3, como Problema 5.4 (página 289), do capítulo referente ao lançamento, à sorte, de figuras. Passamos agora a mostrar a resolução apresentada por Pacheco d'Amorim deste mesmo problema, mas recorrendo ao conceito de esperança matemática.

O autor inicia esta resolução reflectindo sobre as características da esperança matemática, afirmando que a esperança matemática de uma região (ou classe) $\mathbf{X}$ pode ser obtida sem termos a informação do valor da esperança matemática de cada uma das partes desta região, isto é, podemos criar uma partição $\mathbf{X}_1, \dots, \mathbf{X}_n$ de $\mathbf{X}$ e, sem saber qual a esperança matemática em cada uma destas regiões (classes) $\mathbf{X}_i$, obter a esperança matemática em relação a $\mathbf{X}$. Podemos também obter a esperança matemática de uma região (ou classe) $\mathbf{X}$ sem saber a sua lei de probabilidade, situação que o autor vai ilustrar recorrendo ao problema da agulha de Buffon. Recordemos previamente que, como referimos na página 271, num lançamento de uma figura num campo de variação ilimitado, é indiferente efectuar o lançamento da figura de uma única vez ou executar o lançamento da figura por partes, efectuando uma partição e lançando uma parte de cada vez, como se de um lançamento de uma linha poligonal, fechada ou aberta, se tratasse. Notemos que, se dividirmos a agulha (segmento de comprimento γ) em segmentos γ_i

de igual comprimento, cada uma destas partes terá igual probabilidade de intersectar uma das rectas. Deste modo, se, em vez de utilizarmos uma recta dividida em partes de igual comprimento, utilizarmos um sistema articulado, a probabilidade de cada parte será ainda a mesma como consequência da definição do lançamento à sorte de uma figura variável. Consequentemente, a forma da figura não alterará essa mesma probabilidade e, como tal, podemos moldar o seu formato de forma a simplificar a obtenção da probabilidade pretendida. É este o panorama do problema da agulha de Buffon, onde a figura, agulha que corresponde a um segmento de comprimento γ , é lançada num plano que corresponde a uma folha de papel ilimitada (com rectas paralelas equidistantes, sendo δ a distância entre elas). Formalizemos então estas ideias.

Seja γ o comprimento da agulha que é dividida nos segmentos γ_i (que definem uma partição de γ) de comprimento $\frac{\gamma}{n}$ que se supõe inferior a δ . Deste modo, se aplicarmos uma função Ψ , que corresponde ao número de intersecções, esta associa, para cada parte da agulha com comprimento inferior à distância das rectas ($\gamma_i < \delta$), o número um se intersectar uma recta e o número zero nos casos contrários, pois, uma vez que o comprimento é inferior à distância das rectas, a função não pode assumir um valor superior a um. Deste modo, o valor esperado da função Ψ em relação a cada segmento γ_i será dado por

$$\mathbb{E}_{\gamma_i}(\Psi) = 0 \times \mathbb{P}_{\gamma_i}(\Psi = 0) + 1 \times \mathbb{P}_{\gamma_i}(\Psi = 1) = \mathbb{P}_{\gamma_i}(\Psi = 1), \quad (8.166)$$

que é igual para todos os segmentos γ_i visto que têm igual comprimento. Assim, pela aditividade do valor esperado, dado que $\gamma = \cup_{i=1}^n \gamma_i$ e os segmentos γ_i são disjuntos, teremos

$$\mathbb{E}_{\gamma}(\Psi) = \sum_{i=1}^n \mathbb{E}_{\gamma_i}(\Psi) = n\mathbb{E}_{\gamma_i}(\Psi) \quad (8.167)$$

que, caso a agulha tenha um comprimento inferior à distância entre as rectas, não permitindo que haja mais do que uma intersecção, verifica

$$\mathbb{M} = \mathbb{E}_\gamma (\Psi) = 0 \times \mathbb{P}_\gamma (\Psi = 0) + 1 \times \mathbb{P}_\gamma (\Psi = 1) = \mathbb{P}_\gamma (\Psi = 1) \quad (8.168)$$

pelo que, nesta situação, podemos concluir que

$$\mathbb{P}_\gamma (\Psi = 1) = \sum_{i=1}^n \mathbb{P}_{\gamma_i} (\Psi = 1) = n \mathbb{P}_{\gamma_i} (\Psi = 1). \quad (8.169)$$

Das expressões (8.167) e (8.169) podemos concluir, considerando que o número de segmentos que constituem a partição de γ aumenta indefinidamente, isto é, fazendo n tender para infinito, que a probabilidade (quando $\gamma < \delta$) e o valor médio (em qualquer situação) é proporcional ao comprimento da agulha,

$$\mathbb{E}_\gamma (\Psi) = k\gamma, \quad (8.170)$$

onde k , a constante de proporcionalidade, é independente da forma e do perímetro da figura considerada. Uma vez que é independente da forma, podemos lançar um segmento de comprimento γ ou, por exemplo, uma circunferência de perímetro γ , pois o valor esperado do número de intersecções será idêntico. Para determinarmos o valor de k , consideremos uma circunferência de diâmetro igual à distância de duas paralelas consecutivas, isto é, com diâmetro igual a δ (logo com perímetro $\gamma = \delta\pi$). Lançando, à sorte, esta circunferência sobre o plano das paralelas, ela encontrará sempre uma paralela e uma só em dois pontos⁽²¹⁾, tendo por esta razão o número de intersecções o valor médio igual a dois (pois trata-se do valor médio de uma constante). Assim podemos concluir que

$$\mathbb{E}_{\pi\delta} (\Psi) = k\pi\delta = 2 \quad (8.171)$$

⁽²¹⁾ Pacheco d'Amorim esqueceu-se da situação em que a circunferência é tangente a duas rectas. Contudo trata-se de uma situação com probabilidade nula, que a sua inclusão não altera as conclusões.

e, por conseguinte, o valor da constante de proporcionalidade será dado por

$$k = \frac{2}{\pi\delta}. \quad (8.172)$$

Substituindo, o valor de k , na expressão (8.170) obtemos

$$\mathbb{E}_\gamma(\Psi) = \frac{2\gamma}{\pi\delta}. \quad (8.173)$$

Assim, como concluímos previamente, este será o valor da probabilidade de haver um encontro caso a agulha tenha um comprimento inferior à distância entre as duas rectas da folha ($\gamma < \delta$), consequência de nesta situação ser impossível a agulha, em forma de segmento, ter duas ou mais intersecções com as rectas paralelas. Por esta razão o seu valor esperado será igual à sua probabilidade, tal como verificamos na expressão (8.168). Deste modo, para esta situação em que a distância das rectas é superior ao comprimento da agulha, poderíamos ter resolvido o problema recorrendo aos valores esperados e, naturalmente, teríamos obtido o mesmo valor que alcançámos quando resolvemos este mesmo problema no capítulo referente ao lançamento à sorte de figuras sem recurso ao conceito de valor esperado (consultar expressão (5.8) presente na página 291). No caso de o comprimento da circunferência ser superior à distância das rectas ($\gamma > \delta$), o número de intersecções poderá assumir valores superiores à unidade e, consequentemente, teremos

$$\begin{aligned} \mathbb{E}_\gamma(\Psi) &= 0 \times \mathbb{P}_\gamma(\Psi = 0) + 1 \times \mathbb{P}_\gamma(\Psi = 1) + 2 \times \mathbb{P}_\gamma(\Psi = 2) + \dots = \\ &= \sum_i i \mathbb{P}_\gamma(\Psi = i), \end{aligned} \quad (8.174)$$

razão pela qual, através da informação contida em (8.173), que continua válida nesta situação, nada podemos concluir sobre cada uma das probabilidades associadas a este problema.

Pacheco d'Amorim refere ainda que, se no problema da agulha de Buffon substituíssemos as rectas paralelas por círculos concêntricos e equidistantes,

obteríamos o mesmo valor para a esperança matemática, mas, nesta situação, nada poderíamos concluir acerca da probabilidade do encontro dum segmento rectilíneo com as circunferências. No problema de Buffon é possível deduzir a probabilidade através do valor esperado pelo facto de a função Ψ , que associa o número de intersecções, só assumir os valores zero e um, sob a hipótese de o comprimento da recta ser inferior à distância das rectas, enquanto neste problema, em que as rectas equidistantes são substituídas por círculos equidistantes, o número de intersecções poderá, mesmo que o comprimento da recta seja inferior à distância dos círculos, assumir pelo menos o valor dois. Como tal, de facto nesta situação a informação do valor esperado não é suficiente para deduzir as probabilidades associadas ao problema.

Como referimos na secção 5.1.3, esta forma de resolução do problema da agulha de Buffon é original de Barbier (1860), aplicando a esperança matemática e, recorrendo às palavras de Bertrand (1888, p. 53), utilizando uma “*génieuse substitution*” da agulha por um círculo de perímetro adequado. Esta resolução é utilizada por diversos autores contemporâneos de Pacheco d’Amorim, tais como Czuber (1884, p. 116–117), Bertrand (1888, p. 52–53), Poincaré (1896, p. 129–130), Borel (1909, p. 107) e (1914, p. 80–82), entre outros. Refira-se que o resultado deste problema, como consequência da Lei dos Grandes Números que garante a convergência da proporção amostral para a probabilidade, pode ser utilizado para simular o valor de π através de um grande número de lançamentos de uma agulha num soalho com rectas paralelas equidistantes. Na secção 8.7 apresentaremos alguns resultados obtidos em simulação.

8.6 Lei dos Grandes Números

Nesta secção vamos analisar a generalização efectuada por Pacheco d'Amorim para os resultados referentes à Lei Fraca dos Grandes Números, que deduziu, para a situação de provas de Bernoulli, no capítulo antecedente.

Com este objectivo o autor começa por considerar uma experiência aleatória com unicamente dois resultados possíveis, tendo probabilidade igual a p de ocorrer a modalidade a_1 e probabilidade igual a $q = 1 - p$ de ocorrer a modalidade a_2 . A esta experiência associamos uma função ξ que faz corresponder a cada modalidade a_i , com $i = 1, 2$, o número $\xi(a_i)$. O valor médio da função assim definida é

$$\mu = \mathbb{M}(\xi(a)) = p\xi(a_1) + q\xi(a_2). \quad (8.175)$$

Se repetirmos esta experiência aleatória m vezes, das quais se observam n_{a_i} vezes a modalidade a_i , verificando-se $m = n_{a_1} + n_{a_2}$, a **média aritmética das observações** ou **média empírica** é definida por

$$\overline{X}_m(\xi) = \overline{X}_m = \frac{n_{a_1}\xi(a_1) + n_{a_2}\xi(a_2)}{n_{a_1} + n_{a_2}} = \hat{p}\xi(a_1) + \hat{q}\xi(a_2), \quad (8.176)$$

onde $\hat{p} = \frac{n_{a_1}}{m}$ e $\hat{q} = \frac{n_{a_2}}{m}$ correspondem à proporção, respectivamente, da modalidade a_1 e a_2 nas m provas observadas.

Teorema 8.12. *A probabilidade de que a distância entre o valor médio μ e a média aritmética $\overline{X}_m$ seja superior a determinada quantidade positiva ε , por mais pequeno que seja o valor de ε , tende para zero quando o número de experiências aleatórias aumenta indefinidamente, isto é,*

$$\lim_{m \rightarrow +\infty} \mathbb{P}(|\overline{X}_m - \mu| > \varepsilon) = 0, \quad \forall \varepsilon > 0. \quad (8.177)$$

Demonstração.

$$\begin{aligned}
 |\overline{X}_m - \mu| &= |\hat{p}\xi(a_1) + \hat{q}\xi(a_2) - [\xi(a_1)p + \xi(a_2)q]| = \\
 &= |\xi(a_1)(\hat{p} - p) + \xi(a_2)(\hat{q} - q)| \leq |\xi(a_1)(\hat{p} - p)| + |\xi(a_2)(\hat{q} - q)| = \\
 &= |\xi(a_1)| |\hat{p} - p| + |\xi(a_2)| |\hat{q} - q| = (|\xi(a_1)| + |\xi(a_2)|) |\hat{p} - p| \quad (8.178)
 \end{aligned}$$

uma vez que $\hat{q} - q = p - \hat{p}$. Pelo Terceiro Teorema de Bernoulli (Teorema 7.7 presente na página 403) sabemos que

$$\lim_{m \rightarrow +\infty} \mathbb{P}(|\hat{p} - p| > \varepsilon) = 0, \quad \forall \varepsilon > 0 \quad (8.179)$$

logo

$$\begin{aligned}
 &\lim_{m \rightarrow +\infty} \mathbb{P}(|\overline{X}_m - \mu| > \varepsilon) \leq \\
 &\leq \lim_{m \rightarrow +\infty} \mathbb{P}((|\xi(a_1)| + |\xi(a_2)|) \cdot |\hat{p} - p| > \varepsilon) = \\
 &= \lim_{n \rightarrow +\infty} \mathbb{P}(|\hat{p} - p| > \varepsilon') = 0, \quad \forall \varepsilon' > 0, \quad (8.180)
 \end{aligned}$$

onde $\varepsilon' = \frac{\varepsilon}{|\xi(a_1)| + |\xi(a_2)|}$ e os valores $\xi(a_1)$ e $\xi(a_2)$ não são simultaneamente nulos (caso em que associaríamos o número zero a qualquer resultado da experiência aleatória e obrigatoriamente teríamos $\overline{X}_m = \mu$). Assim, por menor que seja o valor de ε , a probabilidade de que a distância entre a média empírica e o valor médio, i.e. $|\overline{X}_m - \mu|$, se mantenha superior a ε tenderá para zero, à medida que o número de experiências aleatórias m tenda para infinito. $\square$

Pacheco d'Amorim afirma que este resultado é válido para uma experiência aleatória com um qualquer número finito de modalidades e é igualmente válido para a esperança matemática de qualquer classe. Consideremos então uma experiência aleatória com n modalidades a_i que compõem a classe finita $\mathbf{A}$, com probabilidade de cada modalidade $p_i = \mathbb{P}_{\mathbf{A}}(a_i)$, verificando naturalmente $\sum_{i=1}^n p_i = 1$. A esta experiência é aplicada uma função ξ que a cada

modalidade a_i associa um valor numérico $\xi(a_i)^{(22)}$. Recordemos que, nesta experiência, a esperança matemática da função ξ em relação à classe $\mathbf{A}' \subset \mathbf{A}$ é dada por

$$\mathbb{E}_{\mathbf{A}'} [\xi(a)] = \sum_{a_i \in \mathbf{A}'} \xi(a_i) \mathbb{P}_{\mathbf{A}}(a_i) = \sum_{a_i \in \mathbf{A}'} \xi(a_i) p_i. \quad (8.181)$$

A média aritmética observada em m experiências em relação à classe $\mathbf{A}'$, representada por $\overline{X}_{m|\mathbf{A}'}$, é definida através de

$$\overline{X}_{m|\mathbf{A}'} = \sum_{a_i \in \mathbf{A}'} \xi(a_i) \frac{n_{a_i}}{m} = \sum_{a_i \in \mathbf{A}'} \xi(a_i) \hat{p}_i \quad (8.182)$$

onde n_{a_i} representa o número de observações da modalidade a_i nas m experiências efectuadas e $\hat{p}_i = \frac{n_{a_i}}{m}$ a proporção observada da modalidade a_i nessas m experiências. Nestas condições podemos deduzir os resultados que se seguem.

Teorema 8.13. *A probabilidade de a distância entre a esperança matemática de ξ em relação à classe $\mathbf{A}'$, representada por $\mathbb{E}_{\mathbf{A}'} [\xi(a)]$, e a média aritmética de ξ em relação à mesma classe, representada por $\overline{X}_{m|\mathbf{A}'}$, seja superior a determinada quantidade positiva ε , por mais pequeno que seja o valor de ε , tende para zero quando o número de experiências aleatórias aumenta indefinidamente, isto é,*

$$\lim_{m \rightarrow +\infty} \mathbb{P} \left(\left| \overline{X}_{m|\mathbf{A}'} - \mathbb{E}_{\mathbf{A}'} [\xi(a)] \right| > \varepsilon \right) = 0, \quad \forall \varepsilon > 0. \quad (8.183)$$

⁽²²⁾ Notemos que esta ideia está na origem do conceito de variável aleatória que foi pela primeira vez utilizado, segundo David e Edwards (2001), por Cantelli (1916), que o denominava *variabile casuale*, tendo-se tornado um conceito fundamental na Teoria da Probabilidade. A função ξ está a associar a cada resultado ω_i do espaço amostral Ω um valor numérico, pelo que a função ξ será uma aplicação $\xi : \Omega \rightarrow \mathbb{R}$, tal como a variável aleatória que corresponde a uma função mensurável. Contudo, Pacheco d'Amorim não se preocupa, como é seu estilo ao longo da sua tese de doutoramento, com as características a que esta função tem de obedecer para que não surjam problemas. Deste modo, nesta obra, não há qualquer imposição sobre a função ξ .

Demonstração. Utilizando raciocínio análogo ao empregue na situação com apenas duas modalidades, podemos concluir

$$\begin{aligned} \left| \bar{X}_{m|\mathbf{A}'} - \mathbb{E}_{\mathbf{A}'} [\xi(a)] \right| &= \left| \sum_{a_i \in \mathbf{A}'} \xi(a_i) \hat{p}_i - \sum_{a_i \in \mathbf{A}'} \xi(a_i) p_i \right| \leq \\ &\leq \sum_{a_i \in \mathbf{A}'} |\xi(a_i)| (\hat{p}_i - p_i) = \sum_{a_i \in \mathbf{A}'} |\xi(a_i)| \cdot |\hat{p}_i - p_i|. \end{aligned} \quad (8.184)$$

Em termos de probabilidade teremos então

$$\begin{aligned} \mathbb{P} \left(\left| \bar{X}_{m|\mathbf{A}'} - \mathbb{E}_{\mathbf{A}'} [\xi(a)] \right| > \varepsilon \right) &\leq \mathbb{P} \left(\sum_{a_i \in \mathbf{A}'} |\xi(a_i)| \cdot |\hat{p}_i - p_i| > \varepsilon \right) \\ &\leq \sum_{a_i \in \mathbf{A}'} \mathbb{P} \left(|\xi(a_i)| \cdot |\hat{p}_i - p_i| > \frac{\varepsilon}{\#\mathbf{A}'} \right) = \sum_{a_i \in \mathbf{A}'} \mathbb{P} (|\hat{p}_i - p_i| > \varepsilon'_i) \end{aligned} \quad (8.185)$$

onde $\varepsilon'_i = \frac{\varepsilon}{|\xi(a_i)| \cdot \#\mathbf{A}'}$ no caso de $\xi(a_i) \neq 0$ ⁽²³⁾. Uma vez que, pelo Terceiro Teorema de Bernoulli, cada uma das probabilidades $\mathbb{P} (|\hat{p}_i - p_i| > \varepsilon'_i)$ tende para zero quando m tende para infinito e na expressão (8.185) temos uma soma finita de quantidades a convergir para zero (se a classe $\mathbf{A}$ é finita a classe $\mathbf{A}'$ também o é), podemos concluir que

$$\lim_{m \rightarrow +\infty} \sum_{a_i \in \mathbf{A}'} \mathbb{P} (|\hat{p}_i - p_i| > \varepsilon'_i) = 0, \quad (8.186)$$

e, consequentemente,

$$\lim_{m \rightarrow +\infty} \mathbb{P} \left(\left| \bar{X}_{m|\mathbf{A}'} - \mathbb{E}_{\mathbf{A}'} [\xi(a)] \right| > \varepsilon \right) = 0. \quad (8.187)$$

Desta forma concluímos que o limite da soma das imagens da função ξ para os elementos observados na classe $\mathbf{A}'$ a dividir pelo número total de experiências converge, quando aumentamos indefinidamente o número de provas, para o valor esperado da função ξ em relação à classe $\mathbf{A}'$. $\square$

⁽²³⁾ No caso de $\xi(a_i) = 0$ a probabilidade correspondente será $\mathbb{P} \left(0 > \frac{\varepsilon}{\#\mathbf{A}'} \right) = 0$ e podemos tirar este termo da soma.

Nas situações em que a classe $\mathbf{A}'$ corresponde à classe $\mathbf{A}$, a esperança matemática corresponderá ao valor médio, podendo o resultado ser enunciado da forma que se passa a expor e ser considerado um corolário do teorema anterior, correspondente à situação $\mathbf{A}' = \mathbf{A}$.

Teorema 8.14 (Lei dos Grandes Números - Caso Discreto). *A probabilidade de a distância entre o valor médio e a média aritmética da função ξ ser superior a determinada quantidade positiva ε , por mais pequeno que seja o valor de ε , tende para zero quando o número de experiências aleatórias aumenta indefinidamente, isto é,*

$$\lim_{m \rightarrow +\infty} \mathbb{P}(|\bar{X}_m - \mu| > \varepsilon) = 0, \quad \forall \varepsilon > 0. \quad (8.188)$$

Isto significa que o limite da média aritmética dos valores encontrados para a função ξ , que assume um número finito de valores, numa série de experiências cujo número aumenta indefinidamente, corresponde ao valor médio dessa função.

Pacheco d'Amorim, após apresentar estes resultados referentes à Lei Fraca dos Grandes Números para classes finitas, destaca a vital importância destes, considerando que desempenham um papel fundamental na Estatística, pois é a estes resultados que a esperança matemática e o valor médio devem toda a sua importância, tal como a média aritmética nas aplicações do Cálculo das Probabilidades. Desta forma o autor destaca, desde logo, o papel essencial que estes resultados, em associação com os Teoremas de Bernoulli, que estudamos no capítulo precedente, desempenham nas aplicações do Cálculo das Probabilidades que Pacheco d'Amorim desenvolve na CONCLUSÃO da sua tese de doutoramento. Assim, para o autor, a importância capital assumida pela média aritmética deve-se então à sua convergência para o valor médio.

O autor considera ainda que estes teoremas podem ser generalizados para uma função, variando dum modo contínuo numa dada região, situação que

analisaremos doravante. Consideremos um ponto variando numa região $\mathbf{X}$, f uma função definida nessa região e $\mathbb{P}_{\mathbf{X}}(x)$ a lei de probabilidade do ponto x em relação à região $\mathbf{X}$. O valor médio da função f , conforme definição 8.4, presente na página 474, é dado por

$$\mu = \mathbb{M}(f(x)) = \int_{\mathbf{X}} f(x) \mathbb{P}_{\mathbf{X}}(x) \, dx. \quad (8.189)$$

Se efectuarmos m lançamentos à sorte do ponto x , a média aritmética observada da função f é dada por

$$\overline{X}_m(f) = \overline{X}_m = \frac{1}{m} \sum_{i=1}^m f(x_i), \quad (8.190)$$

onde x_i , com $i = 1, \dots, m$, representam as observações dos m lançamentos efectuados.

Teorema 8.15 (Lei dos Grandes Números - Caso contínuo). *Seja x um ponto, livre ou imagem, variando na região $\mathbf{X}$, f uma função definida em $\mathbf{X}$ e $\mathbb{P}_{\mathbf{X}}(x)$ a sua lei de probabilidade em relação à região $\mathbf{X}$, então*

$$\lim_{m \rightarrow \infty} \mathbb{P}(|\overline{X}_m - \mu| > \varepsilon) = 0, \quad \forall \varepsilon > 0. \quad (8.191)$$

Desta forma, o limite da média aritmética dos valores encontrados para a função f numa série de lançamentos, cujo número aumenta indefinidamente, corresponde ao valor médio dessa função.

Demonstração. Se dividirmos a região $\mathbf{X}$ em n partes, isto é, se as regiões $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n$ definirem uma partição da região $\mathbf{X}$, a esperança matemática da região parcial $\mathbf{X}_i$, será

$$\mathbb{E}_i = \mathbb{E}_{\mathbf{X}_i}(f) = \int_{\mathbf{X}_i} f(x) \mathbb{P}_{\mathbf{X}}(x) \, dx \quad (8.192)$$

onde, supondo continuidade da função f , o Teorema do Valor Médio garante a existência de um ponto x_i , da região $\mathbf{X}_i$, tal que se verifica

$$\mathbb{E}_i = f(x_i) \int_{\mathbf{X}_i} \mathbb{P}_{\mathbf{X}}(x) \, dx = f(x_i) \cdot \mathbb{P}_{\mathbf{X}}(\mathbf{X}_i). \quad (8.193)$$

Pela aditividade da esperança matemática podemos obter o valor médio μ através de

$$\mu = \mathbb{M}(f(x)) = \sum_{i=1}^n \mathbb{E}_i = \sum_{i=1}^n f(x_i) \cdot \mathbb{P}_{\mathbf{X}}(\mathbf{X}_i). \quad (8.194)$$

Consideremos que agrupamos os valores em n classes, correspondentes às regiões $\mathbf{X}_1, \dots, \mathbf{X}_n$ e que, a cada classe $\mathbf{X}_i$, associamos o valor numérico $f(x'_i) = f(x_i) + \varepsilon_i$ onde $x'_i \in \mathbf{X}_i$ e ε_i é infinitamente pequeno em relação a $f(x_i)$ pela continuidade da função f , i.e. $\varepsilon_i = o(f(x_i))$, e os valores de ε_i vão tender para zero se a medida das regiões $\mathbf{X}_i$ também tender para zero, isto é, $\mathbf{X}_i \rightarrow x_i$. A ideia central é utilizar $f(x'_i)$ como uma aproximação de $f(x_i)$, presente na expressão (8.194), cometendo-se um erro igual a ε_i que tende para zero à medida que a amplitude da classe diminui. Assim teremos

$$\begin{aligned} \bar{X}_m(f) &= \frac{1}{m} \sum_{i=1}^m f(x'_i) = \frac{1}{m} \sum_{i=1}^m (f(x_i) + \varepsilon_i) = \\ &= \frac{1}{m} \sum_{i=1}^m f(x_i) + \frac{1}{m} \sum_{i=1}^m \varepsilon_i. \end{aligned} \quad (8.195)$$

Desta forma, a média aritmética dos valores de $f(x)$ pode ser decomposta em duas parcelas: a primeira corresponde aos valores $f(x_i)$ e a segunda correspondente aos valores de ε_i . A primeira parcela, $\frac{1}{m} \sum_{i=1}^m f(x_i)$, tende, pela Lei dos Grandes Números do caso discreto (Teorema 8.14 presente na página 554), para $\sum_{i=1}^n f(x_i) \mathbb{P}_{\mathbf{X}}(\mathbf{X}_i)$ quando o número de lançamentos aumenta e, conseqüentemente, para μ , qualquer que seja o modo da partição de $\mathbf{X}$ (cf. expressão (8.194)). A segunda parcela, $\frac{1}{m} \sum_{i=1}^m \varepsilon_i$, que converge para $\sum_{i=1}^n \varepsilon_i \mathbb{P}_{\mathbf{X}}(\mathbf{X}_i)$, tenderá para zero, pois, se fixarmos um valor ε positivo, por mais pequeno que seja o valor de ε , é sempre possível criar uma partição de $\mathbf{X}$ em regiões com uma medida tão pequena quanto necessitemos para que o valor absoluto de $\sum_{i=1}^n \varepsilon_i \mathbb{P}_{\mathbf{X}}(\mathbf{X}_i)$ seja inferior a ε .

Deste modo, concluímos que o limite da média dos valores da função f existe e é igual ao valor médio de $f(x)$. $\square$

Pacheco d'Amorim demonstra desta forma a Lei Fraca dos Grandes Números para funções de lançamentos de pontos em regiões. Contudo, a sua análise, tal como se verifica por diversas vezes ao longo da sua tese de doutoramento, não é totalmente satisfatória, dado que não mostra, de forma clara, todas as hipóteses que considera nas suas demonstrações. Que características tem a função f ou a lei de probabilidade $\mathbb{P}_x(x)$ do ponto x que obedecer para que, de facto, possamos garantir a convergência da média aritmética para o valor médio? Supõe-se que o valor de $\mu = \mathbb{M}(f(x))$ tem que existir e que a função f tem que ser contínua, mas serão estas condições suficientes? Podemos também inferir, pelo facto de as probabilidades p_i serem constantes, que estamos a analisar a média de variáveis aleatórias independentes e identicamente distribuídas. Assim, a Lei Fraca dos Grandes Números deduzida por Pacheco d'Amorim é válida para as variáveis aleatórias discretas independentes que assumam um número finito de valores⁽²⁴⁾. No caso contínuo, é válida para qualquer função contínua f das observações das variáveis aleatórias independentes, desde que $\mathbb{E}(f)$ exista. Pacheco d'Amorim terá como base este resultado nas aplicações que apresenta na conclusão da sua tese de doutoramento. Para o autor é a esta convergência das características (funções) das observações de uma amostra aleatória para a mesma característica da população que se deve a importância do Cálculo das Probabilidades nas aplicações. O autor, com esta demonstração, segue um caminho distinto da escola russa de probabilidades. Enquanto Chebycheff, Lyapounov e Markov tentavam, impondo restrições aos momentos, deduzir a convergência da soma de variáveis, quer a Lei Fraca dos Grandes Números quer o Teorema Limite Central, Pacheco d'Amorim tenta obter a Lei Fraca dos Grandes Números

⁽²⁴⁾ Recordemos que Pacheco d'Amorim não faz qualquer análise da situação de classes infinitas, isto é, dos casos em que a variável aleatória assume um número infinito numerável de valores.

através de uma abordagem distinta. Em primeiro lugar, o autor não analisa apenas a média das observações, pois considera a média de uma função qualquer das observações (que é contínua na situação de lançamentos em regiões), apesar de este problema poder ser reduzido ao anterior, pois, se estamos a analisar a variável x , então $y = f(x)$ será outra variável (um ponto imagem na nomenclatura utilizada pelo autor). Por outro lado, Pacheco d'Amorim não faz qualquer imposição sobre os momentos, excluindo a existência do valor médio de $f(x)$ que temos sempre presente, apesar de nada ser referido (que é equivalente à existência de valor médio da variável y). Assim, o autor parte da existência da convergência da média empírica das imagens de uma função f (da variável y) em tiragens, idênticas e independentes, em classes finitas, para generalizar, através do limite, o resultado para uma qualquer função contínua f onde as observações são obtidas através de lançamentos idênticos e independentes numa qualquer região $\mathbf{X}$. A ideia é muito semelhante à construção do integral de Riemann, onde, para deduzirmos a existência do integral de uma qualquer função contínua numa região $[a, b]$, dividimos a região em n sub-intervalos $\mathbf{X}_i = [x_{i-1}, x_i)$, tais que

$$a = x_0 < x_1 < \cdots < x_n = b, \quad (8.196)$$

que definem uma partição do intervalo $[a, b]$ e, considerando a amplitude de cada um destes intervalos representada por

$$\Delta_i = x_i - x_{i-1}, \quad i = 1, \cdots, n, \quad (8.197)$$

então as somas de Riemann podem ser definidas por

$$\sum_{k=1}^n f(x'_i) \Delta_i, \quad (8.198)$$

onde $x'_i \in \mathbf{X}_i$. Representando a norma da partição através de

$$\Delta^{(n)} = \max_{1 \leq i \leq n} \Delta_i, \quad (8.199)$$

o integral de Riemann é definido por

$$\int_a^b f(x) \, dx = \lim_{\substack{n \rightarrow \infty \\ \Delta^{(n)} \rightarrow 0}} \sum_{k=1}^n f(x'_i) \Delta_i, \quad (8.200)$$

onde para funções contínuas (excepto num conjunto de medida nula) e limitadas o limite não depende da forma como a partição é considerada, isto é, o limite será igual desde que a norma da partição convirja para zero. Pacheco d'Amorim, no contexto da probabilidade contínua, para demonstrar a convergência da média aritmética de uma função f para o seu valor médio aplicou ideia análoga.

De salientar, uma vez mais, que as generalizações da Lei Fraca dos Grandes Números efectuadas por Pacheco d'Amorim neste capítulo assumem, na CONCLUSÃO da sua tese, um papel primordial na concepção efectuada pelo autor das aplicações do Cálculo das Probabilidades.

8.7 Propriedades da Lei de Gauss

Pacheco d'Amorim dedica o final deste capítulo para deduzir o valor médio de algumas características de uma variável que seja caracterizada pela lei de Gauss, começando por referir que no capítulo precedente tinha sido demonstrado que a probabilidade de que o afastamento relativo λ se mantenha, em valor absoluto, inferior a λ_0 é dada por

$$\theta(\lambda_0) = \frac{2}{\sqrt{\pi}} \int_0^{\lambda_0} e^{-\lambda^2} \, d\lambda, \quad (8.201)$$

sendo a lei de probabilidade da variável λ dada por

$$\mathbb{P}(\lambda) = \frac{1}{\sqrt{\pi}} e^{-\lambda^2}. \quad (8.202)$$

Pacheco d'Amorim apresenta a lei de probabilidade de λ unicamente neste capítulo, pois, até este momento, o autor apenas deduziu a lei de probabilidade do valor absoluto do afastamento. Pacheco d'Amorim não apresenta qualquer justificação para esta passagem da expressão (8.201), lei de probabilidade de $|\lambda|$, para a expressão (8.202), onde temos a lei de probabilidade de λ . Naturalmente, a simetria, em torno de zero, de λ é suficiente para justificar esta dedução, mas não será válida em geral esta passagem de obter a lei de probabilidade de uma variável aleatória x dividindo a lei de probabilidade do seu valor absoluto $|x|$ por dois. Recordemos que no único exemplo que o autor mostra na sua tese, de aplicação desta lei, Problema 7.3 referido na página 451, a probabilidade procurada refere-se a um valor absoluto, pois pretende-se determinar a probabilidade de um jogador ganhar ou perder uma quantia superior a 10 tostões. Desta forma, parece-nos que o autor devia ter deduzido esta lei de probabilidade para o afastamento, positivo ou negativo, no capítulo anterior, onde apresentou a Lei de Gauss.

O primeiro valor médio deduzido pelo autor foi o de λ , tendo obtido

$$\mathbb{M}(\lambda) = \frac{1}{\sqrt{\pi}} \int_{-\infty}^{+\infty} e^{-\lambda^2} \lambda \, d\lambda = \frac{1}{\sqrt{\pi}} \left[-\frac{e^{-\lambda^2}}{2} \right]_{-\infty}^{\infty} = 0, \quad (8.203)$$

sendo o valor médio dos afastamentos nulo, como seria esperado. Sem esta informação Pacheco d'Amorim nunca poderia ter deduzido (8.202) através de (8.201). Depois Pacheco d'Amorim determina o valor médio do afastamento absoluto $|\lambda|$, que é

$$\begin{aligned} \mathbb{M}(|\lambda|) &= \frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} e^{-\lambda^2} |\lambda| \, d\lambda = \frac{1}{\sqrt{\pi}} \int_0^{\infty} 2e^{-\lambda^2} \lambda \, d\lambda = \\ &= \frac{1}{\sqrt{\pi}} \left[-e^{-\lambda^2} \right]_0^{\infty} = \frac{1}{\sqrt{\pi}}. \end{aligned} \quad (8.204)$$

O valor médio de λ^2 é dado por

$$\begin{aligned}\mathbb{M}(\lambda^2) &= \frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} e^{-\lambda^2} \lambda^2 \, d\lambda = \frac{1}{\sqrt{\pi}} \int_0^{\infty} \lambda e^{-\lambda^2} 2\lambda \, d\lambda = \\ &= \frac{1}{\sqrt{\pi}} \left[-\lambda e^{-\lambda^2} + \int e^{-\lambda^2} \, d\lambda \right]_0^{\infty} = \frac{1}{\sqrt{\pi}} \cdot \frac{\sqrt{\pi}}{2} = \frac{1}{2},\end{aligned}\quad (8.205)$$

onde o autor utiliza

$$\int_0^{\infty} e^{-\lambda^2} \, d\lambda = \frac{\sqrt{\pi}}{2}.\quad (8.206)$$

Notemos que Pacheco d'Amorim na página 126 da sua tese de doutoramento, onde deduz que a probabilidade total na Lei de Gauss para o afastamento absoluto é igual à unidade⁽²⁵⁾, também recorre a este resultado, embora nunca o justifique.

Estas deduções das características da Lei de Gauss aparecem em quase todas as obras contemporâneas à tese de doutoramento de Pacheco d'Amorim. Anotemos que podemos facilmente deduzir algumas destas propriedades através da Função Gama, ou integral de Euler de segunda espécie, que é definido por

$$\begin{aligned}\Gamma(\alpha) &= \int_0^{+\infty} x^{\alpha-1} e^{-x} \, dx = \\ &= \left(\frac{x^\alpha}{\alpha} e^{-x} \right) \Big|_0^{+\infty} - \int_0^{+\infty} -\frac{x^\alpha}{\alpha} e^{-x} \, dx = \\ &= \frac{1}{\alpha} \int_0^{+\infty} x^\alpha e^{-x} \, dx = \frac{1}{\alpha} \Gamma(\alpha + 1),\end{aligned}\quad (8.207)$$

logo, teremos

$$\Gamma(\alpha + 1) = \alpha \Gamma(\alpha),\quad (8.208)$$

⁽²⁵⁾ Apesar de nessa dedução o autor apresentar uma gralha, que é frequente ao longo desse capítulo, pois ora apresenta $\theta(\lambda_0) = \frac{1}{\sqrt{\pi}} \int_0^{\lambda_0} e^{-\lambda^2} \, d\lambda$ ora apresenta (correctamente) $\theta(\lambda_0) = \frac{2}{\sqrt{\pi}} \int_0^{\lambda_0} e^{-\lambda^2} \, d\lambda$ para a função de distribuição do afastamento absoluto λ .

e dado que, como facilmente se verifica, $\Gamma(1) = 1$, concluímos que para $\alpha \in \mathbb{N}$ verifica-se

$$\Gamma(\alpha + 1) = \alpha(\alpha - 1) \cdots \Gamma(1) = \alpha!. \quad (8.209)$$

Assim, a função Gama é uma extensão da função factorial de $\mathbb{N}$ para $\mathbb{R}^+$.

Outro valor útil da função Gama é a imagem do ponto $\frac{1}{2}$ que é dada por

$$\Gamma(0.5) = \int_0^{+\infty} x^{-\frac{1}{2}} e^{-x} dx = \sqrt{\pi} \quad (8.210)$$

uma vez que, tal como concluímos no capítulo precedente ao demonstrarmos que o integral da lei de probabilidade de Gauss é igual à unidade, temos

$$\frac{1}{\sqrt{\pi}} \int_{-\infty}^{+\infty} e^{-x^2} dx = 1 \Leftrightarrow \int_0^{+\infty} e^{-x^2} dx = \frac{\sqrt{\pi}}{2}, \quad (8.211)$$

então, se aplicarmos a substituição $y = x^2$ na expressão da função Gamma, teremos

$$\frac{\sqrt{\pi}}{2} = \int_0^{+\infty} e^{-x^2} dx = \int_0^{+\infty} e^{-y} \frac{1}{2\sqrt{y}} dy = \frac{1}{2} \int_0^{+\infty} y^{-\frac{1}{2}} e^{-y} dy = \frac{1}{2} \Gamma\left(\frac{1}{2}\right) \quad (8.212)$$

de onde concluímos que $\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$. Notemos agora que

$$\begin{aligned} \mathbb{M}\left(|\lambda|^k\right) &= \frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} |\lambda|^k e^{-\lambda^2} d\lambda = \frac{2}{\sqrt{\pi}} \int_0^{\infty} \lambda^k e^{-\lambda^2} d\lambda = \\ &= \frac{1}{\sqrt{\pi}} \int_0^{\infty} y^{\frac{k}{2}-\frac{1}{2}} e^{-y} dy = \frac{1}{\sqrt{\pi}} \Gamma\left(\frac{k+1}{2}\right), \end{aligned} \quad (8.213)$$

onde para k par teremos os momentos $\mathbb{M}(\lambda^k)$ (pois os momentos $\mathbb{M}(\lambda^k)$ de ordem ímpar são nulos pela simetria da lei de Gauss), e para valores de k ímpares teremos os valores médios em relação ao valor absoluto. Assim, poderíamos facilmente deduzir

$$\mathbb{M}(1) = \frac{1}{\sqrt{\pi}} \Gamma\left(\frac{1}{2}\right) = \frac{1}{\sqrt{\pi}} \sqrt{\pi} = 1; \quad (8.214)$$

$$\mathbb{M}(|\lambda|) = \frac{1}{\sqrt{\pi}} \Gamma(1) = \frac{1}{\sqrt{\pi}}; \quad (8.215)$$

$$\mathbb{M}(\lambda^2) = \frac{1}{\sqrt{\pi}}\Gamma\left(\frac{3}{2}\right) = \frac{1}{\sqrt{\pi}}\frac{1}{2}\Gamma\left(\frac{1}{2}\right) = \frac{1}{2}; \quad (8.216)$$

$$\mathbb{M}(\lambda^k) = \frac{1}{\sqrt{\pi}}\Gamma\left(\frac{k+1}{2}\right) = \frac{k-1}{2} \frac{k-3}{2} \cdots \frac{1}{2}, \quad k \text{ par.} \quad (8.217)$$

Desta forma podemos concluir que a variância da lei de Gauss utilizada por Pacheco d’Amorim é $\frac{1}{2}$, o quarto momento $\mathbb{M}(\lambda^4)$ é igual a $\frac{3}{4}$, logo, o coeficiente de achatamento de Pearson será $\frac{\mathbb{M}(\lambda^4)}{\mathbb{M}^2(\lambda^2)} = 3$, como é bem conhecido.

Por fim Pacheco d’Amorim, de forma bastante semelhante à de Borel (1909, p. 58–59), refere que das expressões (8.204) e (8.205) podemos concluir que

$$\frac{\mathbb{M}(\lambda^2)}{[\mathbb{M}(|\lambda|)]^2} = \frac{\frac{1}{2}}{\frac{1}{\pi}} = \frac{\pi}{2}. \quad (8.218)$$

O autor afirma que a Lei Fraca dos Grandes Números fornece a este resultado um significado notável, de que “*podemos rectificar a circunferência, por meio de lançamentos à sorte*”. Pacheco d’Amorim, na primeira edição do seu Compêndio de Geometria, narra o seguinte.

“como se faz a um fio que se tem enrolado e que se rectifica, puxando-lhe pelas pontas. (...) A circunferência é uma das curvas que não podem ser rectificadas por meio da régua e do compasso, mas podemos medir o seu comprimento, experimentalmente, ajustando-lhe um fio, que a seguir se estende e mede.”

[Pacheco d’Amorim 1920, p. 47]

Desta forma o autor diz que este resultado pode ser utilizado para simular o valor de π . Fazemos algumas simulações para averiguar o que acontece.

Na Tabela 8.11 podemos consultar uma síntese dos resultados obtidos na simulação onde, para obter o valor de π , foram utilizadas n observações

n	5 000	10 000	100 000	1 000 000	10 000 000
r	10 000	10 000	10 000	1 000	1 000
Média	0.000165	0.000106	-0.000014	0.000029	-0.000022
Des. Padrão	0.023637	0.016709	0.005296	0.001654	0.000511
DMA	0.018908	0.013350	0.004224	0.001323	0.000401
Mínimo	-0.09655	-0.06167	-0.02088	-0.00583	-0.00192
Quartil 1	-0.01642	-0.01104	-0.00358	-0.00107	-0.00037
Quartil 2	-0.00026	-0.00008	-0.00010	-0.00004	-0.000017
Quartil 3	0.015655	0.011427	0.003543	0.001181	0.000296
Máximo	0.103579	0.064383	0.018331	0.004940	0.001829

Tabela 8.11: Simulação de π pela Lei de Gauss

independentes x_i de uma distribuição gaussiana com média nula e variância $\frac{1}{2}$, que corresponde à lei de Gauss que Pacheco d'Amorim utiliza, sendo os valores apresentados na tabela referentes aos erros de estimação definidos por

$$\hat{\pi}_1 - \pi = \frac{\overline{X}_n(x^2)}{[\overline{X}_n(|x|)]^2} - \pi, \quad (8.219)$$

obtidos com recurso ao software *Mathematica*, tendo sido utilizadas r réplicas. Dos valores verificados, por exemplo, pelos valores do desvio médio absoluto (DMA), este método não nos parece uma boa forma de obter uma aproximação para o valor de π , nem mesmo utilizando 10 000 000 de observações. Refira-se ainda que utilizámos *software* que obtém uma observação da lei de Gauss de forma imediata e que o autor só apresenta formas de obter a lei de Gauss como convergência de outras distribuições. Ora, Pacheco d'Amorim não explica como poderia efectuar esta rectificação da circunferência, uma vez que não explica como pode obter uma observação de uma lei de Gauss. Porventura, para a época, seria mais fácil recorrer a uma agulha e simular o valor de π através do resultado que obtivemos na resolução do problema

da agulha de Buffon. Para confirmar esta ideia simulamos o valor de π com recurso a n lançamentos de uma agulha.

n	5 000	10 000	100 000	1 000 000	10 000 000
r	10 000	10 000	10 000	1 000	1 000
Média	0.002017	0.000975	0.000091	-0.000099	0.000027
Des. Padrão	0.064559	0.046186	0.014433	0.004707	0.001460
DMA	0.051368	0.036658	0.011517	0.003791	0.001175
Mínimo	-0.21762	-0.16008	-0.05203	-0.01282	-0.00424
Quartil 1	-0.04178	-0.03021	-0.00974	-0.00330	-0.00095
Quartil 2	0.000108	-0.00089	-0.00020	-0.00012	0.000040
Quartil 3	0.045151	0.030996	0.009900	0.002967	0.001053
Máximo	0.269049	0.207369	0.054419	0.020663	0.004591

Tabela 8.12: Simulação de π pela agulha de Buffon com $\delta = 1$ e $\gamma = 0.5$

Na Tabela 8.12 estão presentes os resultados obtidos em r réplicas de n lançamentos de uma agulha de dimensão $\gamma = 0.5$ considerando rectas paralelas com uma distância igual a $\delta = 1$. Os valores apresentados referem-se aos erros de estimação definidos por

$$\hat{\pi}_2 - \pi = \frac{2\gamma}{\delta\hat{p}} - \pi, \quad (8.220)$$

onde $\hat{p}$ corresponde à proporção de vezes que a agulha intersectou uma recta nos n lançamentos. Os valores obtidos parecem ser ainda piores que os primeiros onde utilizamos a lei de Gauss. Na Tabela 8.13 apresentam-se os resultados análogos, para os erros de estimação obtidos pelo terceiro estimador $\hat{\pi}_3 - \pi$, obtidos em lançamentos de agulhas de dimensão $\gamma = 1$ num chão com rectas paralelas com distância igual a $\delta = 1$. Os resultados melhoraram, se bem que ainda teremos de lançar a agulha umas 10 milhões de vezes

para obter uma estimativa que tem um erro médio aproximadamente igual a 0.000574.

n	5 000	10 000	100 000	1 000 000	10 000 000
r	10 000	10 000	10 000	1 000	1 000
Média	0.000107	0.000062	0.000060	0.000017	0.000034
Des. Padrão	0.033786	0.023796	0.007449	0.002362	0.000721
DMA	0.027028	0.018955	0.005937	0.001877	0.000574
Mínimo	-0.10853	-0.08489	-0.02851	-0.00813	-0.00204
Quartil 1	-0.02341	-0.01610	-0.00503	-0.00162	-0.00046
Quartil 2	-0.00089	-0.00010	-10^{-6}	0.000068	0.000046
Quartil 3	0.022964	0.015970	0.005040	0.001544	0.000517
Máximo	0.133874	0.097274	0.031248	0.007017	0.002681

Tabela 8.13: Simulação de π pela agulha de Buffon com $\delta = \gamma = 1$

Podemos ainda recorrer a um terceiro método que corresponde à aplicação elementar da Probabilidade Geométrica, lançando um ponto à sorte no quadrado $[-1, 1] \times [-1, 1]$, que corresponde a lançar dois pontos à sorte em $[-1, 1]$, um para cada coordenada, a proporção de pontos que cai dentro do círculo de raio unitário e centro em $(0, 0)$ irá convergir, pela Lei dos Grandes Números, para a probabilidade desse acontecimento. Esta probabilidade é facilmente obtida pela definição geométrica de probabilidade e teremos o quociente entre a área do círculo (π) e a do quadrado, logo teremos $p = \frac{\pi}{4}$. Assim, um estimador para o valor de π será dado por $\hat{\pi}_4 = 4\hat{p}$. Na Tabela 8.14 estão presentes algumas medidas dos valores obtidos para os erros de estimação, $\hat{\pi}_4 - \pi$, na simulação efectuada.

Os valores obtidos para o estimador $\hat{\pi}_4$ também não parecem ser convincentes, pois continuamos a ter de lançar 10 milhões de pontos num quadrado

n	5 000	10 000	100 000	1 000 000	10 000 000
r	10 000	10 000	10 000	1 000	1 000
Média	0.000456	0.000134	0.000034	-0.000005	0.000010
Des. Padrão	0.023069	0.012679	0.005220	0.001709	0.000521
DMA	0.018429	0.013004	0.004158	0.001358	0.000418
Mínimo	-0.08799	-0.05719	-0.01995	-0.004788	-0.00158
Quartil 1	-0.01519	-0.01119	-0.00351	-0.001215	-0.00036
Quartil 2	0.000807	0.000007	-0.00007	0.000057	-0.000003
Quartil 3	0.016007	0.010807	0.003487	0.001115	0.000365
Máximo	0.084007	0.059207	0.020127	0.005255	0.001657

Tabela 8.14: Simulação de π pelo lançamento de um ponto

para obtermos um erro médio aproximadamente igual a 0.000418. Contudo, os resultados obtidos por este método de lançar pontos num quadrado apresentam uma melhor aproximação do valor de π que os verificados pelo lançamento de agulhas e são muito semelhantes aos alcançados pela lei de Gauss.

Notemos que, neste caso, tendo uma máquina de calcular disponível, provavelmente seria melhor apelar a métodos não aleatórios. Se recorrêssemos à fórmula de Wallis (consultar página 393), que é anterior à tese de doutoramento de Pacheco d'Amorim, e estimássemos o valor de π através de

$$\pi \approx 2 \cdot \frac{2}{1} \cdot \frac{2}{3} \cdot \frac{4}{3} \cdot \frac{4}{5} \cdots \frac{2n}{2n-1} \cdot \frac{2n}{2n+1} = \frac{2^{4n+1} (n!)^4}{[(2n)!]^2 (2n+1)} = \hat{\pi}_5(2n), \quad (8.221)$$

obteríamos para o erro de *estimação*, após alguma paciência a efectuar as contas (que, apesar de tudo, são bastante menores que as necessárias nas simulações anteriores), os valores apresentados na Tabela 8.15. Assim, para cometermos um erro inferior a 0.001 teríamos de utilizar um valor de n superior a 1000.

Contudo, uma vez que a fórmula de Wallis fornece, através do estimador

n	50	100	500	1 000	10 000
$\hat{\pi}_5 - \pi$	-0,030647	-0,015514	-0,003134	-0,001569	-0,000157
$\hat{\pi}_6 - \pi$	-0,000148	-0,000038	-0,000002	-4×10^{-7}	-4×10^{-9}

Tabela 8.15: Erros de estimação pela fórmula de Wallis

$\hat{\pi}_5(m)$ previamente definido, para um valor de m par (único caso definido em (8.221)) uma aproximação sempre inferior a π , podemos tentar corrigir este erro sistemático. Se definirmos para qualquer valor de m ímpar o *estimador*

$$\pi \approx 2 \cdot \frac{2}{1} \cdot \frac{2}{3} \cdot \frac{4}{3} \cdot \frac{4}{5} \cdots \frac{2n}{2n-1} \cdot \frac{2n}{2n+1} \cdot \frac{2n+2}{2n+1} = \hat{\pi}_5(2n+1), \quad (8.222)$$

que fornece sempre um valor superior a π , podemos utilizar o *estimador* $\hat{\pi}_6(n) = \frac{\hat{\pi}_5(n) + \hat{\pi}_5(n+1)}{2}$. Na Tabela 8.15 estão apresentados os valores dos erros cometidos, $\hat{\pi}_6(n) - \pi$, quando utilizamos esta aproximação. Assim, o método determinístico baseado na fórmula de Wallis parece fornecer melhor resultado. Contudo, para determinar estes valores, foi necessário efectuar numerosas multiplicações e divisões nas quais, sem recurso a um computador, seria impossível obter um resultado fidedigno. Deste modo, os estimadores associados ao lançamento da agulha de Buffon ou do lançamento à sorte de um ponto no quadrado, apesar de eventualmente ser difícil encontrar as condições ideais para efectuar as experiências, condições que garantam de facto a aleatoriedade dos resultados, parecem mais acessíveis. Com eles seria possível efectuar a simulação, lançar as agulhas ou os pontos à sorte, e, no fim, praticamente bastaria determinar uma proporção para obter uma estimativa de π . Assim sendo, para a época em que foi escrita a tese de doutoramento de Pacheco d'Amorim, apenas nos parecem válidos os estimadores $\hat{\pi}_2$, $\hat{\pi}_3$ e $\hat{\pi}_4$. Em 1901 o matemático italiano Lazzerini realizou 34080 lançamentos da agulha, dos quais 10848 intersectaram as rectas, *estimando* o valor de π em $\frac{34080}{10848} = \frac{355}{113} \approx 3.14159292$, cujo erro é aproximadamente igual a 0.00000026,

logo, obteve um valor correcto até à sexta casa decimal. Além de muita paciência poderemos afirmar, em comparação com os valores obtidos na simulação que efectuamos, que Lazzerini teve também muita sorte no valor que obteve. Curiosamente Castelnuovo (1919, p. 148) não refere este fantástico resultado e menciona que o astrónomo Wolf em 1850 obteve uma estimativa de 3.1596 utilizando 5000 lançamentos da agulha, M. A. Smith em 1855 efectuou 3204 provas e obteve 3.1553 e Lazzerini em 1902 com 2000 provas obteve 3.1446.

Em relação ao primeiro estimador, $\hat{\pi}_1$, que é o referido por Pacheco d'Amorim, consideramos praticamente impossível efectuar qualquer experiência na época, pois não existiria nenhum método para obter, pelo menos de forma fácil, observações (simulações) da lei de Gauss. A única situação possível seria gerar observações da lei de Gauss com recurso ao Teorema Limite Central, usando, por exemplo, um elevado número de lançamentos de uma moeda ao ar para obter cada observação gaussiana.

8.8 Comentário geral ao capítulo

Pacheco d'Amorim começa por definir esperança matemática e valor médio de forma distinta. Para o autor a esperança matemática de uma função em relação a uma região (classe) $\mathbf{X}'$ avalia apenas os valores que a variável assume nessa região (classe) $\mathbf{X}'$, mas que não correspondem a valores esperados condicionados, uma vez que a lei de probabilidade utilizada não é condicionada a $\mathbf{X}'$. Trata-se, assim, como se fosse uma truncagem do cálculo habitual de valor esperado. Em contrapartida, o conceito de valor médio, ou valor provável, inclui todos os valores que a variável pode assumir na região (classe) $\mathbf{X}$ em estudo e, uma vez que $\mathbf{X}$ poderá não corresponder à região (classe) total possível, então o valor médio corresponde ao valor esperado condicionado à

região (classe) $\mathbf{X}$. O autor deduz depois algumas propriedades usuais destas medidas, entre as quais destacamos a demonstração de que, para determinar o valor médio de uma função f que depende dos lançamentos dos pontos $x_1, \dots, x_n$, podemos utilizar a definição de valor médio ou definir um ponto $z = f(x_1, \dots, x_n)$ e utilizar este lançamento para determinar o valor médio, isto é, podemos recorrer a transformações que o valor esperado não será alterado, pois para o ponto $\mathbf{x}$ variando na região $\mathbf{X} \subset \mathbb{R}^n$ teremos uma função densidade conjunta a que corresponderá para $z = f(x_1, \dots, x_n)$, variando em $\mathbf{Z} \subset \mathbb{R}$ uma função densidade tal que $\mathbb{E}_{\mathbf{x}}(f) = \mathbb{E}_{\mathbf{z}}(z)$. Este resultado é mais um exemplo, tal como o seu conceito de ponto imagem, de que o autor considera que a medida de probabilidade é invariante para transformações biunívocas.

Com recurso ao conceito de valor médio Pacheco d'Amorim deduz uma desigualdade semelhante à actualmente denominada desigualdade de Markov, concluindo que, se uma função não negativa tem valor esperado que é inferior a qualquer δ positivo, então a probabilidade de a função assumir valores superiores a um determinado ε será, no máximo, $\frac{\delta}{\varepsilon}$. Com este resultado podemos deduzir que uma variável aleatória, não negativa, que tenha valor esperado nulo, terá probabilidade nula de assumir valores positivos, e, se o valor esperado de uma função f , não negativa, de um conjunto de variáveis aleatórias $X_1, X_2, \dots, X_n$ tender para zero, à medida que n aumenta indefinidamente, então a probabilidade de f assumir valores positivos convergirá igualmente para zero.

Pacheco d'Amorim resolve um problema de lançamento de uma curva flexível e inextensível onde conclui que o valor esperado do quadrado da distância converge para zero. Recorrendo à desigualdade que o autor deduziu previamente, podemos concluir que a probabilidade de no lançamento resultar uma curva fechada convergirá para a unidade. Apesar da desconfi-

ança com que lemos pela primeira vez este resultado, tendo como referência, por exemplo, os resultados obtidos em passeios aleatórios em reticulados, ele está correctamente deduzido e é válido para a forma específica como o autor define o lançamento de uma curva flexível e inextensível. Contudo, será necessário o cuidado extremo de não extrapolar este resultado para qualquer curva lançada à sorte no plano. Apesar disto, do que é exposto na tese de doutoramento de Pacheco d'Amorim não conseguimos determinar qual era o objectivo do autor ao apresentar este problema. Refira-se que, pelo método descrito pelo autor para efectuar o lançamento de uma curva flexível e inextensível, vamos sempre obter um único ponto e, como tal, obrigatoriamente o valor esperado do quadrado da distância entre os dois pontos extremos é nulo. Porém, a ideia patente nestes lançamentos de que o perímetro total é fixo também não corresponde ao resultado, pois o perímetro de um ponto é nulo. Deste modo, se Pacheco d'Amorim tivesse investigado outras situações, por exemplo, que os n lados que constituem o polígono tivessem comprimento igual a $\frac{1}{\sqrt{n}}$, teria obtido processos estocásticos que não convergiriam para um ponto, sendo inclusive possível definir, através deste limite, um processo estocástico contínuo, actualmente denominado por processo de Wiener.

O autor resolve o problema da agulha de Buffon recorrendo ao conceito de esperança matemática, notando que o valor esperado do número de intersecções é proporcional ao comprimento da agulha e que não depende da forma desta, razão pela qual podemos modelar o comprimento e o formato da agulha para facilmente obtermos o valor esperado e com ele determinar a constante de proporcionalidade. Com esta ideia genial de Barbier (1860) Pacheco d'Amorim deduziu a probabilidade para as situações em que o comprimento da agulha é inferior à distância das rectas, pois, nestes casos, o valor médio coincide com a probabilidade de haver intersecção.

Por fim Pacheco d'Amorim generaliza a Lei Fraca dos Grandes Números,

que é fundamental para toda a construção das aplicações que o autor faz na conclusão da sua tese de doutoramento. Deste modo Pacheco d'Amorim deduz que, se aplicarmos uma função, que associa um valor numérico a cada resultado de uma prova de Bernoulli, então a média empírica das imagens desta função, quando executamos provas de Bernoulli independentes, converge para o seu valor médio quando aumentamos indefinidamente o número de provas (consequência do Terceiro Teorema de Bernoulli analisado no capítulo antecedente). Este resultado é generalizado para provas com um número qualquer finito de modalidades e, posteriormente, para lançamentos em regiões (probabilidade contínua) numa construção muito semelhante à construção do integral de Riemann. Apesar de toda esta construção não ser, como é seu estilo, muito clara, pois o autor não apresenta de forma evidente todas as hipóteses que utiliza, a sua construção não deixa de ser singular, sendo as suas conclusões fundamentais para as aplicações que apresenta na conclusão da sua tese. As aplicações do Cálculo das Probabilidades propostas por Pacheco d'Amorim são essenciais para compreender toda a sua construção e serão analisadas e comentadas no próximo capítulo.

Capítulo 9

Conclusão

Pacheco d'Amorim, na CONCLUSÃO da sua tese de doutoramento, analisa as aplicações da Probabilidade. Toda a construção efectuada por Pacheco d'Amorim nos capítulos anteriores é concebida com base na selecção, à sorte, de um elemento do espaço amostra sob as hipóteses do fenómeno padrão, supondo que somos nós os agentes da selecção, de forma a garantir a aleatoriedade desta, e que possuímos total conhecimento do espaço amostra, através do qual podemos deduzir a possibilidade de cada elemento do espaço amostra. Deste modo, como epílogo, o autor analisa os casos onde estas condições não se verificam. O autor expõe, assim, a sua visão sobre as aplicações da Probabilidade, isto é, a sua concepção de Estatística.

Nesta exposição o autor começa por dividir o campo de análise de um fenómeno em três grupos que caracterizam o agente da selecção à sorte e, como tal, qualificam a primeira condição inerente ao fenómeno padrão. Cada grupo é depois dividido em dois sub-grupos que correspondem à tiragem à sorte de um elemento de uma classe (número finito de modalidades) e ao lançamento de um ponto numa região (caso contínuo). Finalmente, cada sub-grupo será dividido em três casos que descrevem o nosso grau de conhecimento sobre

o fenómeno em estudo, relativamente ao conhecimento qualitativo e quantitativo (análise da segunda hipótese do fenómeno padrão). O objectivo de Pacheco d'Amorim é analisar as condições que nos permitem reduzir todas as situações ao fenómeno padrão, de forma a podermos utilizar a teoria que desenvolveu nos capítulos precedentes para modelar os fenómenos em análise. Para justificar estas transformações, o autor fundamenta-se nas Leis de Bernoulli e análogas, considerando que estas nos permitem, sob determinadas condições, passar todas estas situações para o fenómeno padrão, uma vez que estas leis garantem que, se fixarmos um erro máximo ε para a distância entre o valor observado numa amostra, de um qualquer fenómeno, e o seu valor teórico (isto é, o valor desse fenómeno na população), então a probabilidade de cometermos um erro superior ao fixado vai convergir para zero à medida que aumentamos a dimensão da amostra. Deste modo, desde que possamos obter uma amostra de grande dimensão conseguimos sempre obter um resultado aproximado e provável. Assim, Pacheco d'Amorim salienta que este resultado, além de aproximado, é provável, pois, apesar de podermos garantir que o erro máximo cometido é ε com determinada probabilidade p , por mais que a probabilidade p se aproxime da unidade nunca podemos garantir *a priori*, com total certeza, esta majoração do erro. Apesar disto, o autor considera que esta aproximação provável é suficiente para modelar os fenómenos aleatórios, sendo a Teoria da Probabilidade a Ciência que nos deve guiar na tomada de decisão sob incerteza. Pacheco d'Amorim considera ainda que, caso não sejamos nós os agente da escolha à sorte, se o fenómeno se comportar em harmonia com as Leis de Bernoulli e análogas podemos considerá-lo como se fosse proveniente de uma escolha feita por nós próprios.

9.1 Redução ao fenómeno padrão

Pacheco d'Amorim, ao longo da sua tese de doutoramento, definiu e deduziu as propriedades da probabilidade, quer esta se refira a uma tiragem à sorte (fenómeno discreto) quer se compare com o lançamento, à sorte, de um ponto (fenómeno contínuo). Contudo, todas as propriedades derivadas são unicamente válidas sob as duas hipóteses inerentes ao fenómeno padrão, escolha à sorte de um elemento do espaço amostra, apresentado na INTRODUÇÃO da sua tese de doutoramento. Em primeiro lugar consideramos sempre que nós somos os agentes da escolha aleatória, pois, desta forma, temos garantido a aleatoriedade da escolha. A segunda condição refere-se ao total conhecimento do espaço amostra, isto é, a conhecermos a classe **A** de onde é tirado um elemento ou conhecermos a região **X** onde é efectuado o lançamento à sorte de um ponto. Estas duas condições permitem determinar a possibilidade de qualquer elemento do espaço amostra sem ambiguidade. Vamos começar por descrever quais são as outras situações, distintas do fenómeno padrão, que são equacionadas pelo autor.

Pacheco d'Amorim divide os fenómenos em análise em três grupos distintos, que caracterizam o agente da selecção aleatória. Deste modo, teremos as seguintes situações:

1. Nós somos os agentes da selecção;
2. A selecção é efectuada por um agente semelhante a nós;
3. A selecção é efectuada por um agente de outra natureza.

No primeiro grupo encontramos o fenómeno padrão, onde temos garantida a aleatoriedade da selecção; no segundo e terceiro grupos nada nos garante *a priori* que a escolha seja, de facto, efectuada à sorte.

Cada um destes três grupos é dividido em dois sub-grupos, de forma a distinguir as tiragens de elementos em classes (probabilidade discreta) dos lançamentos de pontos em regiões (probabilidade contínua). O primeiro sub-grupo corresponderá, assim, a uma tiragem à sorte de um elemento numa classe finita (número finito de modalidades) e o segundo sub-grupo ao lançamento à sorte de um ponto numa região (modalidade formando um contínuo).

Por fim cada sub-grupo será dividido em três casos, que caracterizam a quantidade de informação que possuímos em relação ao fenómeno em análise, conforme definido na INTRODUÇÃO da tese de doutoramento de Pacheco d'Amorim. Desta forma, cada sub-grupo é dividido nos três seguintes casos:

1. Conhecimento qualitativo e quantitativo do fenómeno — conhecemos a lei de probabilidade e o seu campo de existência;
2. Conhecimento qualitativo e desconhecimento quantitativo do fenómeno — conhecemos o seu campo de existência, mas desconhecemos a sua lei de probabilidade;
3. Desconhecimento qualitativo e quantitativo do fenómeno — desconhecemos o seu campo de existência e a sua lei de probabilidade.

Recorrendo a conceitos mais actuais, no primeiro caso conhecemos o suporte e a distribuição da variável que caracteriza o fenómeno em análise, no segundo caso conhecemos o suporte da variável mas desconhecemos a sua distribuição e, finalmente, no terceiro e último caso, desconhecemos quer a distribuição quer o suporte da variável.

O fenómeno padrão que Pacheco d'Amorim utilizou como base ao longo da sua construção corresponde ao primeiro caso do primeiro grupo (o sub-grupo só nos indica se estamos a trabalhar com probabilidade discreta ou contínua).

Neste capítulo o autor vai expor como podemos utilizar o Cálculo das Probabilidades nas outras situações, distintas do fenómeno padrão, e como podemos reduzir estes casos às condições do fenómenos padrão. Desta forma, o objectivo de Pacheco d'Amorim é expandir as conclusões que deduziu ao longo da sua tese para as situações em que não se verificam as hipóteses inerentes ao fenómeno padrão, referindo que “*todo o fenómeno, para que possa fazer parte do estudo desta ciência, deve poder reduzir-se a este*” (fenómeno padrão). É esta concepção que vamos comentar ao longo deste capítulo, destacando as principais ideias defendidas pelo autor. Assim, vamos começar por destacar o papel fundamental das Leis de Bernoulli e análogas (secção 9.2) na construção de Pacheco d'Amorim, pois será através destes resultados que o autor pretende obter valores aproximados e prováveis. Na secção 9.3 destacaremos a distinção entre ter probabilidade próxima da unidade e *certeza*, pois, para o autor, apesar desta distinção e de os resultados obtidos serem somente prováveis, estes não deixarão de ter um valor prático. Na secção 9.4 analisaremos as propostas de Pacheco d'Amorim para os fenómenos que não se localizam no primeiro caso, isto é, quando não temos informação completa acerca do espaço amostra. Na secção 9.5 examinaremos os fenómenos do segundo e terceiro grupo, quando não somos nós os agentes da selecção e, como tal, não podemos garantir a sua aleatoriedade. Finalmente, na secção 9.6 apresentaremos uma caracterização sucinta dos principais contributos para o desenvolvimento da Estatística *circa* 1914.

9.2 Leis de Bernoulli e análogas

Pacheco d'Amorim apresentou e deduziu a Lei Fraca dos Grandes Números (Leis de Bernoulli e análogas) e o Teorema Limite Central restrito às provas de Bernoulli no capítulo TEOREMA DE JACOB BERNOULLI E LEI DOS

DESVIOS e generalizou a Lei Fraca dos Grandes Números no capítulo ESPERANÇA MATEMÁTICA E VALOR MÉDIO, onde concluiu que, se estivermos a analisar a característica ξ de um fenómeno, então

$$\lim_{m \rightarrow \infty} \mathbb{P}(|\overline{X}_m(\xi) - \mu(\xi)| > \varepsilon) = 0, \quad \forall \varepsilon > 0, \quad (9.1)$$

isto é, a média amostral das imagens da função ξ convergem para o valor esperado dessa função ou, por outras palavras, a média aritmética dos valores observados da característica ξ de uma amostra converge para o valor esperado dessa característica. Desta forma, o limite da média aritmética dos valores encontrados para a função ξ numa série de tiragens ou lançamentos, cujo número aumenta indefinidamente, corresponde ao valor esperado dessa função.

Recordemos que as condições do teorema correspondem a selecções (tiragens ou lançamentos) efectuados à sorte nas mesmas circunstâncias, o que corresponde ao que usualmente referimos na actualidade como variáveis aleatórias i. i. d. — independentes e identicamente distribuídas. Desta forma Pacheco d'Amorim supõe, em toda a sua análise, que as classes (ou regiões) em que as tiragens (lançamentos) são efectuados se mantêm qualitativa e quantitativamente invariáveis, garantindo assim a independência e a semelhança entre as selecções. Pacheco d'Amorim refere que, caso não se verificassem estas hipóteses, então unicamente se a variação ocorrer de uma forma “*lenta e regular*”, de tal forma que possibilite a determinação da lei de variação, é que poderemos incluir no modelo as devidas correcções. Deste modo, o autor considera fundamental a hipótese de invariância do suporte e da lei de probabilidade durante o processo, o que equivale a afirmar que as observações têm que ser independentes e identicamente distribuídas.

Considerando que as condições de aplicabilidade estão satisfeitas e que podemos obter uma amostra de dimensão tão grande quanto necessitemos, as Leis de Bernoulli e análogas permitem obter uma aproximação tão boa quanto

seja o nosso objectivo. Assim sendo, se fixarmos um erro máximo ε para a aproximação, podemos obter um *estimador* que garanta que a *estimativa* obtida comete um erro inferior a ε com uma probabilidade tão próxima da unidade quanto seja pretendido. Pacheco d'Amorim salienta a distinção entre este tipo de aproximação e as usuais utilizadas em Matemática, pois esta, ao contrário das outras, é uma aproximação provável. Desta forma, na utilização deste método para estimação, por mais próxima que a probabilidade se situe da unidade, nunca teremos a certeza de que o erro cometido seja de facto inferior a ε . Comentaremos com mais detalhe este aspecto na secção 9.3, se bem que o autor considere que, mesmo não tendo a certeza do erro máximo cometido na estimação, esta não deixa de ter um valor prático relevante.

Desta forma, a base das aplicações apresentadas por Pacheco d'Amorim corresponde às Leis de Bernoulli e análogas, pois o seu objectivo é recorrer à *inversa* destas leis de forma a poder obter uma aproximação da realidade, valor da característica em estudo na população, com base no valor da média aritmética dessa mesma característica numa amostra. A dimensão da amostra utilizada definirá a qualidade da aproximação, pois, quanto maior for a amostra, mais provável se torna que o erro de aproximação cometido seja menor que o valor ε fixado. Deste modo, Pacheco d'Amorim salienta a importância da média aritmética dos valores obtidos numa amostra, pois é na convergência desta medida para o verdadeiro valor (valor esperado) que Pacheco d'Amorim vai fundamentar as aplicações que propõe.

Destaquemos que a utilização da *inversa* das Leis de Bernoulli, para fundamentar as aplicações, era bastante usual na época. Por exemplo, Alcayde y Carvajal (1908) denomina por Teorema Directo de Bernoulli o resultado que nos garante que, quando temos uma série de experiências independentes com probabilidade de sucesso p conhecida, a proporção de sucessos amostral $\hat{p}$ se situará no intervalo $(p - \varepsilon, p + \varepsilon)$ com uma probabilidade que, aumen-

tando o número de experiências, se aproxima indefinidamente da certeza. O Teorema Inverso de Bernoulli, para Alcayde y Carvajal, garante que, quando observamos sequências de experiências com a probabilidade de sucesso p desconhecida, a probabilidade da proporção de sucessos amostral $\hat{p}$ diferir de p numa quantidade ε , tão pequena quanto se queira, aumenta indefinidamente com o número de experiências, convergindo para a certeza. Guimarães (1904, p. 20) apresenta ideia semelhante, denominando por Teorema inverso do de Bernoulli as situações em que a probabilidade p é desconhecida mas se supõe que é constante. Borel (1914) utiliza implicitamente este resultado para definir a probabilidade estatística. Para Borel existe a probabilidade matemática, baseada na equiprobabilidade e correspondente à definição clássica de probabilidade, e a definição estatística que corresponde à proporção amostral e que será, segundo o autor, uma aproximação da anterior. Com efeito, Borel escreve:

“En définitive, si parmi un grand nombre d'événements analogues, nous en distinguons certains sous le nom de cas favorables, nous dirons que la probabilité est le rapport du nombre des cas favorables ou nombre total des événements. Si le cas favorable est la naissance d'un garçon et s'il est né 51.200 garçons sur 100.000 naissances, la probabilité de la naissance d'un garçon est 0,512.”

[Borel 1914, p. 12]

De facto, as Lei de Bernoulli e análogas são o principal elo de ligação entre a Teoria da Probabilidade e a realidade, sendo esta ligação explorada desde a obra de Bernoulli (1713). É nesta ordem assintótica dos fenómenos aleatórios, formalizada pela primeira vez na Lei Fraca dos Grandes Números de Bernoulli, que encontramos o principal fundamento da Estatística.

9.3 Probabilidade *versus* certeza

Pacheco d'Amorim destaca, ao longo da sua apresentação das aplicações, que as conclusões obtidas são *aproximadas* e *prováveis*, uma consequência imediata de recorrer às Leis de Bernoulli que consideram sempre um erro máximo ε com uma probabilidade que converge para a unidade. Porém, por mais próxima que essa probabilidade se situe da unidade, não poderemos garantir efectivamente que não vamos cometer um erro superior ao erro máximo fixado ε . O autor refere então que, ao recorrermos às Leis de Bernoulli e análogas para obter uma aproximação, o valor obtido será distinto do verdadeiro valor “*pelo hiato que separa a probabilidade da certeza*”, salientando, com referências a *Ars Conjectandi* de Jacob Bernoulli e ao *Essai Philosophique sur les probabilités* de Laplace, a diferença entre *certeza* e uma probabilidade igual a um (ou tão próxima da unidade quanto pretendamos). Pacheco d'Amorim refere que Laplace faz uma clara distinção entre certeza e probabilidade na vizinhança da unidade, pois certeza para Laplace verifica-se unicamente quanto a experiência só tem um único resultado possível e, consequentemente, o resultado está rigorosamente demonstrado, enquanto a probabilidade, por mais próxima da unidade que esteja, pode mesmo assim dar origem a um pequeno erro. O autor refere ainda a opinião de Bernoulli (1713), para quem esta diferença não existe, considerando que a probabilidade é um grau de certeza.

Pacheco d'Amorim, como previamente tinha referido na INTRODUÇÃO da sua tese, considera que são estas probabilidades de valor elevado, vizinhas da unidade, que nos podem servir de guia para a tomada de decisão. Na verdade, podemos não obter com toda a certeza a aproximação pretendida, mas o facto de conhecermos uma aproximação com uma probabilidade elevada servirá de informação para nos auxiliar na tomada de decisões sob meios de incerteza. Pacheco d'Amorim considera, por conseguinte, que é a estas probabilidades próximas da certeza ou, o que é o mesmo se pensarmos no

complementar, às probabilidades próximas da impossibilidade que devemos as aplicações da probabilidade. Recordemos que, na INTRODUÇÃO da sua tese de doutoramento, Pacheco d'Amorim nos dá claramente a sua opinião.

“É nessa previsão que a intuição nos sugere quando o número de bolas brancas é muito maior que o das pretas que reside o valor prático do Cálculo de Probabilidades.

Essa intuição ficará irremediavelmente separada da certeza, por mais que da unidade se aproxime a percentagem das bolas brancas. Mas nem por isso ela deixa de ter para nós um valor prático real.”

[Pacheco d'Amorim, 1914, Introdução, p. 8]

Deste modo, Pacheco d'Amorim considera nitidamente que não poderá, no estudo dos fenómenos aleatórios, atingir a certeza, mas este facto não invalida a utilidade prática do Cálculo das Probabilidades.

Esta ideia também já é partilhada em obras anteriores à tese de doutoramento de Pacheco d'Amorim. Por exemplo, Cournot, um dos primeiros autores a salientar a importância da Teoria da Probabilidade na análise Estatística, apresenta na sua obra *Exposition de la Théorie des Chances et des Probabilités*, publicada em 1843, um raciocínio semelhante com o seu princípio da impossibilidade, onde considera que acontecimentos com pequena probabilidade geralmente não acontecem e que este facto é a única forma de aplicar a Teoria da Probabilidade à realidade.

“L'événement physiquement impossible est donc celui dont la probabilité mathématique est infiniment petite; et cette seule remarque donne une consistance, une valeur objective et phénoménale à la théorie de la probabilité mathématique. [...] Sans doute la notion de l'impossibilité physique diffère essentiellement de celle de l'impossibilité mathématique ou métaphysique, et il n'y a au-

cun moyen d'établir la transition de l'une à l'autre."

[Cournot, 1843, p. 78–79]

Salientemos que, mesmo considerando obras posteriores à tese de doutoramento de Pacheco d'Amorim, é usual aparecer esta concepção. Por exemplo, Paul Lévy (1925) refere que a probabilidade é baseada em dois princípios.

“Tout ce qui précède repose sur deux notions fondamentales. La première est celle d'événements également probables; elle conduit à la définition de la probabilité et suffit pour la théorie. La seconde est celle d'événement très peu probable, grâce à laquelle les résultats de la théorie peuvent, dans certains cas, prendre un intérêt pratique.”

[Lévy, 1925, p. 34]

Assim, para estes autores, tal como para Pacheco d'Amorim, a aplicabilidade do Cálculo das Probabilidades deve-se à existência de acontecimentos com probabilidade pequena que podem ser ignorados na prática ou, o que é análogo, a aproximações com uma probabilidade próxima da unidade, tal como as que são obtidas com recurso à *inversa* das Leis de Bernoulli (em grandes amostras), que nos permitem ignorar o hiato existente entre probabilidade e certeza nessa aproximação. Salientemos apenas que esta ideia, de ignorar os acontecimentos quase impossíveis (de probabilidade quase nula), é geral em muitas concepções e aplicações em Estatística, pois não é idealizado efectuar uma previsão perfeita (caso fosse possível o nosso fenómeno não seria realmente aleatório) quando aplicamos resultados assintóticos tais como a Lei dos Grande Números ou o Teorema Limite Central. Contudo, em algumas áreas da Estatística, de que a Teoria dos Extremos é um exemplo, os acontecimentos com probabilidade quase nula não podem ser ignorados. Refira-se, no entanto, que esta teoria, apesar de a procura do máximo e/ou mínimo de um conjunto de variáveis aleatórias independentes e identicamente

distribuídas já aparecer em obras muito anteriores⁽¹⁾, começa por se salientar a partir dos trabalhos de Ladislaus von Bortkiewicz (1868–1931), Richard von Mises (1883–1953), Maurice Fréchet (1878–1973) e Emil Julius Gumbel (1891–1966), de quem é famosa a afirmação “*Il est impossible que l’improbable n’arrive jamais*”, trabalhos esses posteriores à tese de doutoramento de Pacheco d’Amorim.

Na axiomática de Kolmogoroff (1933), apesar de esta fornecer toda a estrutura matemática para lidarmos com probabilidades, a probabilidade é vista apenas como um conceito matemático, uma medida, pois qualquer função que satisfaça os axiomas é uma medida de probabilidade, não explicando como se determinam as probabilidades através de uma experiência real. Assim, numa experiência aleatória podemos determinar relações entre as probabilidades de diversos acontecimentos consoante as suas características, tais como independência, incompatibilidade, complementaridade, entre outras. Contudo, a axiomática de Kolmogoroff não explica como podemos determinar as probabilidades de experiências elementares⁽²⁾, tais como num lançamento de uma moeda a probabilidade de sair cada uma das faces, pois unicamente permite afirmar que a soma das duas probabilidades é igual à unidade se considerarmos que a moeda cai sempre com um lado virado para cima, e outras relações deste tipo (para deduzirmos as probabilidades teríamos que supor que os acontecimentos são equiprováveis). Desta forma, a axiomática de Kolmogoroff permite manipular as probabilidades de vários acontecimentos, mas não determiná-las e/ou interpretá-las. Assim, Kolmogo-

(1) David e Edwards (2001) referem alguns marcos importantes da história da Teoria dos Extremos onde mencionam que Nicolau Bernoulli em 1709 já tinha apresentado raciocínio semelhante.

(2) As únicas probabilidades que permite determinar são as do conjunto vazio e do universo.

roff não ambiciona apresentar uma axiomática que explique a aplicabilidade deste conceito (Estatística), referindo na sua obra que, com este objectivo, existem outras propostas de axiomatização da probabilidade.

“There are other postulational systems of the theory of probability, particularly those in which the concept of probability is not treated as one of the basic concepts, but is itself expressed by means of other concepts. (For example R. von Mises and S. Bernstein)”

[Kolmogoroff, 1933, p. 2]

Por esta razão, Kolmogoroff dedica apenas uma pequena secção de duas páginas no primeiro capítulo da sua obra à ligação entre a Teoria da Probabilidade que constrói e a realidade, isto é, à aplicabilidade do conceito de probabilidade. Nesta secção o autor considera uma experiência aleatória que pode ser repetida um qualquer número de vezes que se supõe sob o mesmo conjunto de condições. Representemos por $\Omega = \{\omega_1, \dots, \omega_n\}$ o conjunto de todas as possibilidades de resultado da experiência, onde alguns dos resultados podem, em geral, não ocorrer, mas serão incluídos no espaço-amostra Ω todos os resultados que pareçam *a priori* possíveis. Seja $\mathbf{A}$ um conjunto desses resultados, com $\mathbf{A} \subset \Omega$. Assim, sob determinadas condições (que Kolmogoroff não explicita), podemos assumir que é associado ao acontecimento $\mathbf{A}$ o número real $\mathbb{P}(\mathbf{A})$ com as seguintes características (cf. Kolmogoroff (1933, p. 3)):

1. É quase certo que, se as condições da experiência são repetidas um grande número de vezes, seja n , então se m é o número de ocorrências do acontecimento $\mathbf{A}$, o rácio $\frac{m}{n}$ vai diferir muito pouco de $\mathbb{P}(\mathbf{A})$, isto é, as frequências relativas do acontecimento $\mathbf{A}$ vão diferir pouco da sua probabilidade;
2. Se $\mathbb{P}(\mathbf{A})$ é muito pequeno, podemos estar praticamente certos de que,

se efectuarmos a experiência apenas uma vez, o acontecimento **A** não ocorrerá.

Deste modo, as ideias expostas por Kolmogoroff, no que se refere à aplicação ou à ligação do Cálculo das Probabilidades à realidade, não diferem muito das propostas por Pacheco d'Amorim duas décadas antes. As condições que refere, possibilidades de repetição do fenómeno sob as mesmas condições, são equivalentes às de Pacheco d'Amorim, correspondendo a primeira característica às Leis de Bernoulli e análogas e a segunda à ideia, também partilhada por Pacheco d'Amorim, de que para a tomada de decisão unicamente nos preocupamos com o que acontece *normalmente* pois podemos, para fins práticos, ignorar os acontecimentos com probabilidade ínfima.

9.4 Informação incompleta

Pacheco d'Amorim começa por analisar o primeiro grupo que é caracterizado, como referido previamente, pelas situações em que somos nós os agentes das tiragens ou lançamentos e, por este motivo, temos garantida a aleatoriedade da selecção. Assim, vamos analisar o que o autor propõe, nestas condições, caso não tenhamos informação completa do espaço amostra de forma a podermos determinar a possibilidade de cada resultado.

9.4.1 Primeiro sub-grupo — Caso discreto

Pacheco d'Amorim considera que qualquer fenómeno que assuma um número finito de modalidades pode ser associado a tiragens à sorte de elementos de uma classe finita **A** ou a lançamentos feitos numa região dividida num nú-

mero finito de partes⁽³⁾. Se, em relação ao fenómeno em análise, tivermos total conhecimento (primeiro caso), então estaremos nas condições do fenómeno padrão. Desta forma teremos o conhecimento qualitativo, quais as características dos elementos (modalidades) que a classe **A** contém e que podem ocorrer nesta escolha aleatória, e conhecimento quantitativo, isto é, quantos elementos de cada modalidade estão contidos na classe **A**. Assim, com esta informação, podemos determinar a possibilidade de cada modalidade e, como referido, estaremos nas condições do fenómeno padrão. Consequentemente, poderemos utilizar as conclusões expostas nos capítulos 3, 4 e 5. Assim, o objectivo do autor é reduzir os restantes dois casos a este que já foi devidamente investigado.

No segundo caso, onde possuímos conhecimento qualitativo mas não quantitativo do espaço amostra (conhecemos as possíveis modalidades desta escolha mas não conhecemos a probabilidade de cada modalidade), para reduzir ao primeiro caso teremos de determinar quantitativamente a classe em que as tiragens são feitas, isto é, achar o número de elementos de cada modalidade, ou melhor, a possibilidade ou probabilidade de cada modalidade. Recorrendo às Leis de Bernoulli e análogas, sabemos que a proporção amos-

⁽³⁾ Pacheco d'Amorim considera, desta forma, que a escolha à sorte de um elemento de entre um número finito de modalidades pode ser um ponto imagem de um lançamento à sorte numa região **X**, que o autor não formalizou no capítulo que dedicou a este conceito, mas esta aplicação do contínuo para o discreto por vezes estava implícita na sua exposição (conforme referimos aquando da análise do capítulo PONTO IMAGEM da tese de doutoramento de Pacheco d'Amorim). Notemos que, na construção de Pacheco d'Amorim, o lançamento de pontos em regiões ilimitadas (suporte do lançamento corresponder a uma região ilimitada **Y**) só está definido nos pontos imagem, pois os pontos livres são sempre lançados em regiões limitadas. Por conseguinte, o autor poderia ter desenvolvido esta ideia de aplicação do contínuo para o discreto para tentar fundamentar a probabilidade em conjuntos infinitos não numeráveis, que não analisa na sua tese, como em $\mathbb{N}$, dividindo a região onde é efectuado o lançamento num número infinito numerável de partes.

tral de cada modalidade $\hat{p}_i$ converge para a respectiva probabilidade p_i , isto é,

$$\mathbb{P}(|\hat{p}_i - p_i| < \varepsilon) \longrightarrow 1, \quad (9.2)$$

quando a dimensão da amostra aumenta indefinidamente. Desta forma, a proporção de cada modalidade na população p_i pode ser *estimada* através da proporção amostral $\hat{p}_i$ com uma “*probabilidade e aproximação tão grande quanto se queira*”, pois, desde que possamos efectuar tantas tiragens à sorte quantas as necessárias, obteremos uma aproximação com uma probabilidade tão próxima da unidade quanto seja pretendido. Assim, desde que seja possível recolher uma amostra de grande dimensão, poderemos sempre obter uma boa estimativa, uma vez que, mesmo que fixemos um erro de estimação ε pequeno, haverá sempre uma dimensão da amostra que nos garanta um erro menor que ε com uma probabilidade tão elevada quanto seja nosso objectivo. Como analisámos previamente, Pacheco d’Amorim refere que a determinação das probabilidades é *aproximada*, pois estamos a cometer um erro máximo ε , e *provável* pois a probabilidade de não excedermos o erro ε na estimação é tão próxima da unidade quanto pretendamos. Deste modo, como destacámos anteriormente, a base da aproximação são as Leis de Bernoulli e análogas que nos garantem que a estimação comete um erro, em valor absoluto, inferior a ε , com uma probabilidade que diferirá da unidade tão pouco quanto seja o nosso propósito. Desta forma, este segundo caso, onde desconhecemos as probabilidades de cada modalidade, é distinto do primeiro caso (fenómeno padrão) “*pelo hiato que separa a probabilidade da certeza*”, conforme comentamos na secção 9.3.

No terceiro caso desconhecemos qualitativamente e quantitativamente a classe **A**, não existindo qualquer informação disponível sobre o espaço amostra da experiência em investigação. Nestas situações devemos utilizar o mesmo método que empregamos no segundo caso, pois a estimação da pro-

babilidade implica a identificação das possíveis modalidades. Desta forma, a correspondência entre o terceiro caso e o primeiro é feita de forma análoga, ainda que, em vez de identificarmos unicamente a quantidade (probabilidade) das modalidades tenhamos também de identificar qualitativamente a classe **A**, isto é, identificar as modalidades. Assim, Pacheco d'Amorim considera que o processo da sua determinação é o mesmo do caso antecedente. Saliente-se, uma vez mais, que toda esta análise só é válida nos casos em que se verificam as condições de aplicabilidade das Leis de Bernoulli, isto é, quando as tiragens são qualitativa e quantitativamente invariáveis durante todo o processo em que o fenómeno é analisado, sendo as conclusões válidas unicamente enquanto estas condições se mantiverem.

Refira-se que este tipo de estimação de probabilidades, referentes a um fenómeno que assume um número finito de modalidades, com base numa amostra de dimensão elevada e no Teorema Inverso de Bernoulli, já era muito frequente no século XIX.

9.4.2 Segundo sub-grupo — Caso contínuo

Pacheco d'Amorim considera que no segundo sub-grupo estão incluídos quaisquer fenómenos que assumam um número infinito não numerável de modalidades, isto é, que formam um contínuo de segunda espécie na terminologia que Poincaré apresenta na sua obra *La Science et l'hypothèse*. Nestas situações podemos associar o fenómeno ao lançamento de um ponto numa região **X**, com o número de dimensões conveniente para modelar o fenómeno. Se, em relação ao fenómeno em análise, possuímos conhecimento total do espaço amostra (primeiro caso), estaremos nas condições do fenómeno padrão, pois teremos o conhecimento do suporte (região **X** que o autor denomina por campo de existência) e a lei de possibilidade π_x (ou lei de probabilidade). Verificando-se, neste primeiro caso, as condições do fenómeno padrão, pode-

mos recorrer aos capítulos 4 e 5 e 6 para estudar o fenómeno. Desta forma, o objectivo do autor é reduzir os restantes casos a este que já foi devidamente investigado.

O segundo caso supõe conhecido o campo de existência (região $\mathbf{X}$), mas desconhecida a sua lei de probabilidade. Desta forma, para reduzir este caso ao fenómeno padrão teremos de estimar a sua lei de probabilidade. Para analisar esta situação o autor considera que devemos definir uma partição da região $\mathbf{X}$ através das regiões $\mathbf{X}_1, \dots, \mathbf{X}_n$. Se os lançamentos forem efectuados directamente na região $\mathbf{X}$ (se x for um ponto livre), então, pelas Lei de Bernoulli e análogas, numa série de lançamentos obteremos aproximadamente o mesmo número de lançamentos em cada região $\mathbf{X}_i$ se estas tiverem igual dimensão (caso contrário o número será aproximadamente proporcional às suas dimensões). Isto é consequência de, quando temos lançamentos directos (pontos livres), termos obrigatoriamente a distribuição uniforme, uma vez que todos os pontos da região $\mathbf{X}$ (que neste caso tem que ser limitada) são equipossíveis. Se o ponto não for directamente lançado na região, mas caso se trate de um ponto imagem (conforme definições apresentadas no capítulo 6), então a proporção $\hat{p}_{\mathbf{X}_i}$, de pontos de uma sequência de lançamentos (amostra) que caem na região $\mathbf{X}_i$, será aproximadamente igual à probabilidade dessa região e, como tal, teremos

$$\hat{p}_{\mathbf{X}_i} \approx p_{\mathbf{X}_i} = \int_{\mathbf{X}_i} \pi_x \, dx, \quad (9.3)$$

pois, segundo as Leis de Bernoulli e análogas, verificar-se-á

$$\mathbb{P} \left(\left| \hat{p}_{\mathbf{X}_i} - p_{\mathbf{X}_i} \right| < \varepsilon \right) \longrightarrow 1, \quad (9.4)$$

quando a dimensão da amostra aumenta indefinidamente. Desta forma a proporção amostral $\hat{p}_{\mathbf{X}_i}$ convergirá para a probabilidade.

Com base nesta ideia Pacheco d'Amorim propõe duas hipóteses de aplicação em relação aos lançamentos (caso contínuo). Em primeiro lugar, con-

sidera que, se tivermos alguma informação sobre o fenómeno em análise, podemos adoptar uma lei *a priori* e testar se as observações estão de acordo com essa lei. Para efectuar este teste devemos realizar um grande número de sequências de lançamentos, onde cada sequência deverá ser constituída por um número elevado de lançamentos. O número de regiões que definem a partição, o número de sequências efectuado e o número de lançamentos que constitui cada sequência serão determinados pela qualidade da precisão pretendida para a aproximação. Através dos valores obtidos devemos determinar a proporção p de sequências que estão de acordo com a lei adoptada, pois, segundo as Leis de Bernoulli e análogas, esta proporção convergirá para a unidade se a lei adoptada *a priori* for verdadeira. Se a proporção observada de sequências que se harmonizam com a lei proposta p , que Pacheco d'Amorim denomina por probabilidade dessa lei, for suficientemente elevada, poderemos admitir a lei para modelar o fenómeno, caso contrário devemos rejeitá-la. Salientemos que, para o autor conseguir atingir esta meta, teria de ter disponíveis resultados sobre distribuições mais potentes que os que mostra, resultados que são indispensáveis para analisar se as distâncias entre as proporções observadas $\hat{p}_{\mathbf{x}_i}$ e as probabilidades $p_{\mathbf{x}_i}$ são devidas à sorte (aleatoriedade do fenómeno) ou simplesmente ao facto de a lei admitida *a priori* não se adequar aos dados observados. Por isso sublinhamos que Pacheco d'Amorim precisaria de mais resultados, que não apresenta nem estão patentes nas poucas obras que refere, para formalizar os testes não paramétricos que propõe. Tais resultados permitir-lhe-iam concluir se uma amostra está ou não em harmonia com determinada lei de probabilidade, tais como os testes do qui-quadrado formalizados por Karl Pearson no início do século XX, que ignoramos se eram do conhecimento de Pacheco d'Amorim. Apesar de não termos informação sobre o conhecimento dos trabalhos de Karl Pearson por parte de Pacheco d'Amorim, a ideia de testar se a distribuição das

observações do fenómeno se harmonizam com uma lei proposta *a priori* π_y pode ter sido baseada no artigo de Karl Pearson (1900). Neste artigo Pearson constrói o teste do qui-quadrado, que permite determinar se as observações de uma amostra se harmonizam com uma determinada lei de probabilidade.

“The question we wish to determine is whether the sample may be reasonably considered to represent a random system of deviations from the theoretical frequency distribution of the general population”

[Pearson, 1900, p. 164]

Contudo, a filosofia do teste apresentado por Pacheco d'Amorim é distinta da aplicada por Karl Pearson. Este utiliza uma amostra para determinar uma quantidade, baseada nas diferenças entre o número de observações na amostra e o número esperado de observações em cada modalidade (ou intervalo de valores), que determinará se esta distância pode ou não ser considerada proveniente unicamente da aleatoriedade do fenómeno. Caso esta distância não possa ser considerada unicamente devida à aleatoriedade, então é porque a lei considerada *a priori* não modela o fenómeno. Pacheco d'Amorim não utiliza uma única amostra, mas antes *um grande número de séries de lançamentos, sendo cada série formada por um número de lançamentos suficientemente grande*, isto é, um elevado número de amostras de grande dimensão, analisando a proporção de amostras que se harmonizam com a lei proposta. Nesta apresentação implicitamente estará um qualquer método de determinação se cada uma das amostras se harmoniza ou não com a lei proposta. Este método de classificação binária não é descrito pelo autor, mas provavelmente considera um erro máximo ε fixo, para a distância entre $\hat{p}_{\mathbf{x}_i}$ e $p_{\mathbf{x}_i}$, e, em cada amostra, analisa se as distâncias, para cada parte X_i , são efectivamente menores que ε . Caso o sejam, o autor considera que os dados desta amostra se harmonizam com a lei. Assim, podemos determinar

a proporção p de amostras que se harmonizam com a lei (probabilidade da lei) que, se assumir um valor elevado (perto da unidade), poderemos admitir a lei proposta como válida para a modelação do fenómeno, caso contrário devemos rejeitá-la. Este método parece-nos pouco praticável, uma vez que a obtenção de um grande número de amostras de grande dimensão raramente é possível.

Refira-se que a ideia de testar hipóteses através da observação é uma concepção que é usada desde os tempos remotos, tendo sido efectuada em todas as áreas da ciência e, por isso, não é novidade. No campo da Estatística, David e Edwards (2001) apresentam comentários ao artigo “*An argument for Divine Providence taken from the constant regularity observed in the births of both sexes*” de John Arbuthnott, publicado em 1710, num capítulo que denominam *The First Formal Test of Significance*. Desta forma, apesar da metodologia actualmente empregue nos testes de significância ter sido introduzida principalmente por Fisher a partir da década de 1920, este tipo de testes não é novidade no início do século XX.

A segunda aplicação proposta por Pacheco d’Amorim refere-se às situações onde não temos conhecimento suficiente sobre o fenómeno em estudo de forma a podermos propor uma distribuição *a priori*. Assim, quando investigarmos um fenómeno aleatório e não tivermos nenhuma razão para propor uma lei *a priori* que nos permita modelar o fenómeno, devemos estimar a lei de possibilidade criando uma partição $\mathbf{X}_1, \dots, \mathbf{X}_n$ do espaço amostra e lançando um número elevado de pontos. Deste modo, como previamente referimos, a proporção de pontos que cairá em cada região $\mathbf{X}_i$, representada por $\hat{p}_{\mathbf{X}_i}$, terá um valor aproximado da sua probabilidade $p_{\mathbf{X}_i}$, isto é,

$$\hat{p}_{\mathbf{X}_i} \approx p_{\mathbf{X}_i} = \int_{\mathbf{X}_i} \pi_x \, dx, \quad i = 1, \dots, n, \quad (9.5)$$

pelas Leis de Bernoulli. Uma vez que existe uma infinidade de funções π_x que

satisfazem as condições (9.5), Pacheco d'Amorim considera que deveremos escolher a função π_x que melhor satisfaça os nossos propósitos. Se não houver nenhuma razão para escolhermos uma função π_x em detrimento das outras funções candidatas, deveremos aplicar o princípio da parcimónia — quanto mais simples for a função que modela o fenómeno em estudo, tanto melhor.

Este tipo de estimação, da lei de probabilidade que caracteriza o fenómeno, também pode ter sido baseada nos trabalhos de Karl Pearson, apesar de estes se inserirem em uma filosofia distinta. Karl Pearson (1895) apresenta um sistema de distribuições onde dividiu as possíveis distribuições em vários tipos caracterizados pelos primeiros quatro momentos. Assim, recorrendo ao método dos momentos, podemos estimar os primeiros quatro momentos da distribuição utilizando os quatro momentos empíricos (amostrais) que, por sua vez, determinarão a distribuição que caracteriza os dados. A família de distribuições de Pearson⁽⁴⁾ é composta pelas distribuições cuja função densidade $p(x)$ é solução da equação diferencial:

$$\frac{1}{p} \frac{dp}{dx} = -\frac{a+x}{c_0 + c_1x + c_2x^2}, \quad (9.6)$$

ou, no que se refere às variáveis aleatórias discretas, da seguinte equação às diferenças:

$$p_k - p_{k-1} = -p_k \frac{a+x}{c_0 + c_1x + c_2x(1-x)}. \quad (9.7)$$

As soluções destas equações, juntamente com as restrições inerentes às funções de densidade, tais como $p(x) \geq 0$ e $\int_{-\infty}^{+\infty} p(x) dx = 1$ (ou $\sum p(k) = 1$ no caso discreto), determinarão os valores de a , c_0 , c_1 e c_2 , que estão relacionados com os quatro primeiros momentos (logo com a localização, dispersão, assimetria e achatamento) que, por sua vez, caracterizam o tipo de distribuição. Assim, a equação diferencial (9.6) gera diversas funções de densidade

⁽⁴⁾ Johnson, Kotz e Balakrishnan (1994) fazem uma boa descrição da família de distribuições de Pearson.

conhecidas tais como t-Student, Laplace, Pareto, gama, beta, Cauchy, gaussiana (que é considerada como um limite das outras distribuições), entre outras. No caso discreto, a equação às diferenças (9.7) gera a função de probabilidade das distribuições binomial, binomial negativa, hipergeométrica, Poisson, entre outras.

Notemos que, apesar de Pacheco d'Amorim usar um método para a estimação da distribuição que caracteriza o fenómeno, distinto do de Pearson, que era baseado nos primeiros quatro momentos da distribuição, o objectivo é idêntico.

Nesta aplicação de Pacheco d'Amorim, tal como no caso anterior, o número de elementos que constituem a partição e o número de lançamentos efectuados para obter a lei de probabilidade serão definidos segundo a precisão que pretendemos para os resultados. Apontemos que o autor, apesar de apresentar as ideias fundamentais da estimação e de salientar os factores que influenciam a sua qualidade, não constrói formalmente o método.

No terceiro caso, onde desconhecemos igualmente o suporte do lançamento, desconhecimento qualitativo, Pacheco d'Amorim considera que podemos considerar a região $\mathbf{X}$ ilimitada e com o número de dimensões adequado para a modelação do fenómeno, uma vez que a estimação da lei de possibilidade identificará as regiões onde a lei de probabilidade é nula. Desta forma, tal como no primeiro sub-grupo, o terceiro caso reduz-se ao segundo, pois a determinação da lei também identifica o suporte que, por sua vez, com a estimação da lei de probabilidade, se reduz ao primeiro, que corresponde ao fenómeno padrão.

Pacheco d'Amorim salienta, uma vez mais, que toda esta análise foi efectuada supondo que as regiões em que os lançamentos são efectuados se mantêm qualitativa e quantitativamente invariáveis. Deste modo, o autor sublinha a hipótese de que o suporte e a lei de probabilidade não podem ser

alterados durante o processo, isto é, as observações têm de ser independentes e identicamente distribuídas. Caso haja alterações, deveremos determinar a lei de variação, se tal for possível, de forma a corrigir as nossas conclusões.

9.5 Aleatoriedade

Na INTRODUÇÃO da sua tese de doutoramento Pacheco d'Amorim refere que o conceito primitivo *tirar, à sorte, um elemento numa classe* tem para nós um sentido inequívoco quando nós somos os agentes da tiragem. Se não formos nós os agentes da selecção à sorte, não podemos garantir a aleatoriedade dessa mesma escolha e, consequentemente, não poderemos aplicar as propriedades deduzidas que são unicamente válidas para fenómenos aleatórios. Assim, para o autor, a questão é qual o significado da frase *tirar, à sorte, um elemento numa classe* quando o agente da selecção for outro ser semelhante a nós (segundo grupo). Pacheco d'Amorim considera que esta frase corresponde unicamente a *tirar um elemento numa classe*, uma vez que não sabemos como foi efectuada a escolha. Contudo, Pacheco d'Amorim considera que é legítimo, em alguns casos, reduzir esta situação à anterior. Para tal, começa por analisar quais são as condições que garantem que, quando somos nós os agentes da escolha, ela seja efectivamente realizada à sorte, destacando duas condições. Em primeiro lugar considera que devemos desconhecer, por completo, a distribuição dos elementos na classe em que a tiragem é feita. Em segundo lugar, a escolha deverá ser efectuada de forma que não possamos prever o elemento que vai sair, nem tão pouco a sua qualidade, e que essa previsão não possa ser efectuada por nós nem por nenhum ser semelhante a nós. Assim, para Pacheco d'Amorim, sempre que qualquer escolha efectuada por um ser semelhante a nós verifique estas características, não existe qualquer razão para não a considerarmos *a priori* como se fossem

feitas por nós próprios e, desta forma, reduzir ao primeiro grupo, onde temos o fenómeno padrão.

O autor ilustra esta ideia com um exemplo de urnas. Se considerarmos uma urna contendo nove décimos de bolas brancas para um décimo de bolas pretas e dessa urna nós tirarmos à sorte uma bola, não hesitaremos em apostar que sairá uma bola branca. Caso não sejamos nós os agente da escolha mas um agente semelhante a nós, supondo satisfeitas as duas condições previamente referidas, então continuaremos indubitavelmente a apostar nessa mesma cor. Para esta aposta é fundamental que a escolha seja efectuada de modo que o agente da tiragem não consiga prever o elemento que vai tirar, nem ter informação sobre a distribuição das bolas dentro da urna para garantir que todas as bolas tenham igual possibilidade de ser escolhidas. Pacheco d'Amorim refere que estas hipóteses são a base de todos os jogos de azar e é para a justificação dela que, por exemplo, se baralham as cartas antes de serem dadas e se dá ao verso das mesmas uma aparência idêntica. Deste modo, Pacheco d'Amorim considera que, para que uma sequência de tiragens possa ser considerada aleatória, os resultados têm de decorrer de acordo com as Leis de Bernoulli e análogas, pois o que parece aleatório nem sempre o é. Desta forma, o autor aponta duas soluções para os fenómenos enquadrados neste segundo grupo: ou se identificam com o primeiro, se satisfazerem as Leis de Bernoulli e análogas, e neste caso aplicam-se os mesmos procedimentos previamente expostos; ou então não se harmonizam com as Leis de Bernoulli e, como tal, não poderão ser considerados aleatórios. Em tais casos estes fenómenos não podem ser reduzidos ao primeiro grupo e, consequentemente, não fazem parte do domínio do Cálculo das Probabilidades.

No terceiro grupo de fenómenos Pacheco d'Amorim considera os fenómenos análogos a escolhas feitas à sorte por agentes de outra natureza. O autor considera que, como referido, sempre que temos uma sequência de escolhas

aleatórias, os elementos resultantes apresentam uma grande desordem, de tal forma que não existe nenhuma lei que permita prever cada observação. O caso contrário seria antagónico à nossa ideia intuitiva de escolha à sorte, pois, numa escolha aleatória, cada observação é imprevisível. Assim, se presenciarmos um fenómeno cujas observações parecem não ter qualquer ordem, será legítimo considerar que é proveniente de uma escolha aleatória? E poderemos nós reduzir estas situações às anteriores, de forma a investigarmos estes fenómenos através do Cálculo das Probabilidades, determinando, tal como nos grupos anteriores, qualitativamente e quantitativamente, a classe (região) das modalidades? Pacheco d'Amorim admite que sim. O autor, de forma análoga à que efectuou no segundo grupo, considera que uma sequência de observações que se harmonize com as Leis de Bernoulli pode ser considerada proveniente de uma escolha aleatória. Assim, nestes casos, devemos testar se a escolha é, ou não, aleatória, isto é, se as observações se harmonizam com as Leis de Bernoulli. E, caso possamos admitir que seja, deveremos estimar qualitativamente e quantitativamente a escolha aleatória, pois, com o aumento do número de observações, sabemos que, pelas Leis de Bernoulli e análogas, a proporção de cada modalidade vai convergir para um valor que corresponde à sua probabilidade. Assim sendo, o autor considera que numa experiência aleatória, enquanto tivermos poucas observações, haverá uma grande irregularidade nas proporções observadas para cada modalidade, mas, à medida que a dimensão da amostra aumenta, essa irregularidade irá desaparecer e a distância entre as proporções observadas e o valor da probabilidade irá permanecer dentro de determinados limites com uma probabilidade cada vez mais elevada. Em suma, são estas as características dos fenómenos aleatórios, que são o objecto de estudo da Teoria da Probabilidade. Caso o fenómeno não apresente estas especificidades, então não faz parte dos objectos de estudo desta ciência.

A Teoria da Probabilidade investiga as características dos fenómenos aleatórios, sujeitos à influência do acaso, que são caracterizados pela sua imprevisibilidade individual, pois é impossível prever com exactidão (certeza) qual será o resultado de um fenómeno aleatório, e pela sua regularidade agregada, quando analisamos um grande número de resultados de um fenómeno aleatório, observamos que as frequências relativas tendem a estabilizar-se evidenciando uma ordem. Estas regularidades dos fenómenos aleatórios começaram por ser notadas nos jogos de azar, onde se verifica que as frequências relativas de cada um dos possíveis resultados tendem a estabilizar-se num determinado valor, sendo este valor utilizado, por exemplo, para fixar os rácios das apostas. Quer associado aos resultados dos jogos, quer à colecção de observações, tais como as tabelas de mortalidade, nestes agrupamentos de dados, apesar da imprevisibilidade existente em cada observação, é verificada uma notável estabilidade das frequências relativas quando observamos um grupo de observações. Este facto tornou evidente a existência de um padrão no comportamento agregado e criou, a partir do século XVIII, um novo alento na Estatística, que deixa de se restringir ao seu objectivo unicamente descritivo, ambicionando justificar estas regularidades dos fenómenos aleatórios. É esta a ideia das Leis de Bernoulli e análogas e o fundamento da Estatística.

Para Pacheco d'Amorim, se as observações do fenómeno não se harmonizarem com as Leis de Bernoulli e análogas, será porque as hipóteses implícitas não se verificarão, ou porque a aleatoriedade do fenómeno não é legítima ou porque a lei de probabilidade varia durante o processo em que analisamos o fenómeno. Caso as observações do fenómeno se harmonizem com as Leis de Bernoulli e análogas, então o fenómeno investigado pode ser identificado com uma tiragem ou lançamento à sorte e este grupo reduzir-se-á aos grupos precedentes. Esta conclusão será válida, segundo Pacheco d'Amorim, pelo menos durante o período de tempo que demoramos a estimar e a testar as

hipóteses e enquanto as condições dos fenómenos não se alterem. Caso haja alguma alteração, teremos de recomeçar o processo. Pacheco d'Amorim considera, assim, que um conjunto de observações, de um determinado fenómeno, que não verifique estas características, não pode ser considerado proveniente de uma escolha aleatória e, conseqüentemente, não faz parte do campo de análise da Teoria da Probabilidade. Se as condições em que se observa o fenómeno forem alteradas, a hipótese de aleatoriedade poderá manter-se válida. No entanto, deveremos determinar a classe (região) da escolha e determinar a sua lei de probabilidade tantas vezes quantas as necessárias para determinar a lei de variação (caso exista).

Desta forma, para Pacheco d'Amorim, desde que as observações obedeçam às Leis de Bernoulli e análogas, o fenómeno em análise poderá ser modelado como se tratasse de uma escolha à sorte efectuada por nós próprios. Esta ideia defendida por Pacheco d'Amorim é muito semelhante à que utilizamos quando recorremos a números pseudo-aleatórios, cada vez mais importantes na actualidade e que estão na base, por exemplo, das simulações que efectuamos ao longo desta tese. Os números pseudo-aleatórios são gerados por algoritmos e, como tal, não são aleatórios, pois, conhecendo o algoritmo e o número inicial, podemos determinar toda a sequência de números pseudo-aleatórios gerada. Contudo, o comportamento destes números é semelhante ao dos números aleatórios e, por esta razão, podem ser utilizados para todos os propósitos, como se efectivamente fossem números aleatórios. Para Pacheco d'Amorim também não é importante que os fenómenos que são escolhidos à sorte por outros agentes (segundo e terceiro grupo) sejam efectivamente aleatórios, pois, se se comportarem em harmonia com as leis de Bernoulli e análogas, terão um comportamento semelhante ao dos fenómenos aleatórios e, como tal, podem ser modelados como se fossem realmente fenómenos aleatórios. Refira-se, uma vez mais, que estas ideias estão na base da

utilização do números pseudo-aleatórios na simulação, tendo esta ideia sido unicamente aplicada mais tarde. Porém, algumas ideias sobre simulação já existiam, por exemplo, no século XIX, em relação a lançamentos de moedas, dados ou à simulação do valor de π recorrendo ao problema da agulha de Buffon, tal como comentamos na secção 8.7. Stigler (1991) refere que, apesar de muitos autores identificarem a primeira utilização da simulação com o artigo de Student (1908)⁽⁵⁾, existem diversas experiências efectuadas durante o século XIX onde está presente a ideia de simulação, tais como as efectuadas por Erastus de Forest, George Darwin e Francis Galton.

9.6 Estatística *circa* 1914

A construção proposta por Pacheco d'Amorim para a Teoria da Probabilidade e suas aplicações, que formalizaremos recorrendo a conceitos mais actuais na secção 10.1, principia com a fundamentação da probabilidade, onde temos a aleatoriedade garantida e distribuição conhecida, para depois, supondo que (pelo menos) uma destas condições não se verifica, apresentar os procedimentos que devemos implementar para inferir sobre a característica desconhecida. Desta forma, Pacheco d'Amorim propõe uma ligação entre a Probabilidade e a Estatística, numa construção pouco comum para a sua época, tal como posteriormente outros autores, de onde se destacam Richard von Mises, Bruno de Finetti e Maynard Keynes, viriam a efectuar. Deste modo, saliente-se a ideia de que a maioria das obras, em Probabilidades, uni-

⁽⁵⁾ Student (1908, secção VI) gerou 3000 observações através da escolha aleatória de cartões que tinham a altura e o comprimento do dedo médio da mão esquerda de prisioneiros, tendo depois dividido as observações em 750 amostras de dimensão 4 para analisar as suas conclusões. Refira-se que Gosset utilizou estes dados pelo facto de a distribuição destas características ser aproximadamente gaussiana, dado que os coeficientes de assimetria serem quase nulos e os de achatamento próximos de 3.

camente apresentam aplicações muito específicas. Ilustremos esta ideia com alguns exemplos com o objectivo de destacar o valor da visão da interligação entre as Probabilidades e a Estatística apresentada por Pacheco d'Amorim. Condorcet (1785) aplica as probabilidades às decisões em tribunal, tentando determinar a probabilidade de um tribunal chegar ao veredicto verdadeiro ao comparar a decisão de um julgamento, que só poderá ser a de inocente ou a de culpado, com tiragens independentes de bolas de uma urna com bolas brancas e pretas. Assim, supondo que existe uma probabilidade superior a cinquenta por cento de cada jurado acertar no veredicto correcto, concluiu que, se for tomada a decisão que a maioria dos jurados optar, muito provavelmente será tomada a decisão correcta, aumentando esta probabilidade com o adicionar de mais jurados. Com base nesta teoria podemos concluir que, em situações onde há muitas pessoas a dar a sua opinião, se torna mais provável que a resposta da maioria seja a correcta e, à medida que aumentamos o número de intervenientes, a resposta da maioria tenderá para a verdade com probabilidade um (Leis de Bernoulli). Poisson (1837) desenvolveu igualmente as aplicações da probabilidade aos julgamentos. Gouraud (1848), após 140 páginas onde expõe a história da Teoria da Probabilidade e suas aplicações, apresenta em meia dúzia de páginas a sua tese, referindo na quarta proposição de entre seis proposições:

“L’application du Calcul des probabilités aux sciences morales, et notamment à la critique historique, à la jurisprudence, à la législation, à l’économie sociale, à la métaphysique, est une des plus grandes erreurs où soit tombé l’esprit humain”

[Gouraud, 1848, p. 147]

Esta proposição mostra claramente que, durante os séculos XVII e XVIII, as aplicações das probabilidades eram bastante controversas. Foquemos, então, algumas obras mais recentes. Bertrand (1888), no Capítulo XII, *Les Lois de la statistique*, comentando as aplicações às tabelas de mortalidade, salienta

que a primeira condição é a invariância do número de observações favoráveis em relação ao número de provas, isto é, se vamos comparar um fenómeno qualquer com tiragens de bolas de urnas de forma a podermos aplicar os resultados deduzidos neste contexto, teremos que impor que a urna estará sempre nas mesmas condições. Esta ideia é semelhante à invariância imposta por Pacheco d'Amorim nas condições da selecção aleatória e corresponde a uma das condições para a dedução das Leis de Bernoulli que actualmente denominaríamos por variáveis aleatórias independentes e identicamente distribuídas. Bertrand dedica ainda o Capítulo XIII, *Probabilités des décisions*, para comentar as aplicações do cálculo das probabilidades aos julgamentos, sendo muito crítico em relação aos trabalhos de Laplace, Condorcet, Poisson, Cournot entre outros⁽⁶⁾.

Outras obras há que, apesar de serem dedicadas às aplicações do Cálculo das Probabilidades, também apresentam uma visão muito restrita, focando somente um ou outro tópico das aplicações do Cálculo das Probabilidades sem apresentarem uma visão geral da ligação entre as Probabilidades e a Estatística. Alcayde y Carvajal (1907), um engenheiro militar, apresenta aplicações à topografia, geodésicas, balística e fortificação. Broggi (1907), cujo objectivo é aplicar o Cálculo das Probabilidades aos seguros de vida, destaca o Teorema de Bernoulli, a sua generalização de Poisson e a desigualdade de Bienaymé-Chebycheff. Apresenta ainda a teoria estatística da mortalidade, focando a construção das tabelas de mortalidade e hipóteses

⁽⁶⁾ Bertrand refere, em relação à obra de Condorcet, que “*Aucun de ses principes n'est acceptable, aucune de ses conclusions n'approche de la vérité*” e confrontando os trabalhos de Laplace, Condorcet e Poisson, pois todos estes autores compararam a decisão de um tribunal com a tiragem de bolas de uma urna, refere que “*Laplace a rejeté les résultats de Condorcet, Poisson n'a pas accepté ceux de Laplace; ni l'un ni l'autre ont pu soumettre au calcul ce qui y échappe essentiellement: les chances d'erreur d'un esprit plus ou moins éclairé, devant des faits mal connus et des droits imparfaitement définis.*”

adjacentes. Com base nestas noções, desenvolve um capítulo de seguros de vida (cálculo actuarial) e teoria do risco. Laurent (1908), numa obra intitulada *Statistique Mathématique*, salienta a importância do Cálculo das Probabilidades na Estatística e apresenta aplicações a seguros com recurso a tabelas de mortalidade⁽⁷⁾ e à Demografia. Montessus de Ballore (1908) apresenta também, no final da sua obra, alguns capítulos dedicados à aplicação do Cálculo das Probabilidades, nomeadamente ao tiro do canhão, a tabelas de mortalidade, aos seguros e às ciências morais e económicas. Borel (1909, 1914), uma das referências de Pacheco d'Amorim, também analisa aplicações às ciências sociológicas, biológicas e físicas, mas não apresenta uma visão geral. Carvalho (1912) será mais um exemplo de uma obra que analisa o sexo nos nascimentos, as tabelas de mortalidade, os erros de medida, mas não propõe nenhum modelo de índole geral que caracterize a Estatística, como a aplicação das Probabilidades. Deste modo, consideramos que nenhuma destas obras apresenta a Estatística com uma visão geral como a proposta de Pacheco d'Amorim.

Há a salientar alguns progressos no século XIX, no campo da Estatística, alcançados por investigadores mais ligadas ao estudo de grandes quantidades de dados que ao Cálculo das Probabilidades⁽⁸⁾. Destaquemos, de uma forma sintetizada, alguns nomes importantes neste campo com o único objectivo de situar na época a visão de Pacheco d'Amorim. O estatístico belga Lambert Quételet⁽⁹⁾ (1796–1874) estudou probabilidades em Paris sob orientação de Joseph Fourier e Pierre Laplace e, provavelmente influenciado

⁽⁷⁾ Neste tema Laurent segue a obra de Broggi, referindo “*le traité de M. Broggi comme un excellent ouvrage pour ne pas dire le meilleur qui ait été publié sur la question.*”.

⁽⁸⁾ Heyde e Seneta (2001) apresentam um resumo biográfico das principais personalidades que contribuíram para a Teoria da Probabilidade e Estatística.

⁽⁹⁾ Driesbeke (1898) analisa a obra de Quételet.

pelos seus orientadores, generalizou o uso da distribuição gaussiana. Deste modo, Quételet não se restringiu à análise de erros e aplicou a distribuição gaussiana ao estudo de características humanas, tais como a altura e o peso, sendo famosa a sua concepção de *homem médio*⁽¹⁰⁾ definido através do valor médio das características humanas que são agrupadas de acordo com uma curva gaussiana, tendo publicado estes resultados em 1835 na sua obra mais famosa, *Sur l'homme et le développement de ses facultés, essai d'une physique sociale*. Salientemos que, provavelmente, foi a primeira vez que a distribuição gaussiana foi aplicada para modelar características de um qualquer fenómeno, neste caso o corpo humano, e não somente os erros de medida (notemos que o próprio Pacheco d'Amorim denomina esta distribuição por lei dos desvios). Quételet compilou e analisou grandes quantidades de dados referentes a diversas áreas, tais como o crime, a mortalidade, a Meteorologia e a Astronomia. Quételet deu, desta forma, uma contribuição vital para o desenvolvimento da *Estatística*⁽¹¹⁾. Outros estatísticos importantes no século XIX foram o alemão Wilhelm Lexis (1837–1914), pela sua análise da dispersão, que nos conduziu à análise da variância, o dinamarquês Thorvald Thiele (1838–1910), a quem devemos, por exemplo, a ideia de cumulantes que denominava por *half-invariants*⁽¹²⁾. Francis Galton (1822–1911) introduz a

⁽¹⁰⁾ Bertrand (1888) é muito severo com a construção de Quételet, questionando: “*l'homme dont la taille est égale à la taille moyenne, le poids au poids moyen, etc., peut-il exister? N'est-il pas un monstre?*”.

⁽¹¹⁾ Refira-se que o I. M. C. — Índice de Massa Corporal (ou B.M.I. — Body Mass Index) usado internacionalmente na actualidade como uma medida de obesidade foi criado por Quételet sendo definido por $IMC = \frac{P}{A^2}$ onde P representa o peso em quilogramas e A a altura em metros.

⁽¹²⁾ A função geradora de cumulantes da variável aleatória X é definida por $\mathbb{k}_X(t) = \ln(\psi_X(t))$ onde $\psi_X(t)$ representa a função geradora de momentos da variável aleatória X , isto é, $\psi_X(t) = \mathbb{E}(e^{tX})$. Deste modo teremos $\mathbb{k}_X(t) = \ln(\psi_X(t)) = \sum_{j=1}^{\infty} \kappa_j \frac{t^j}{j!}$,

regressão, sendo famoso o seu artigo de 1885, intitulado *Regression towards Mediocrity in Hereditary Stature*, onde conclui que os filhos de pais mais altos que a média tendem a ser mais altos que a média mas mais baixos que os pais e filhos de pais mais baixos que a média tendem a ser mais baixos que a média e mais altos que os pais. Francis Edgeworth (1845–1926) publica em 1885 *Methods of Statistics*, uma obra que já contém alguns resultados de inferência estatística, incluindo, por exemplo, um teste da qualidade do ajustamento semelhante ao que Karl Pearson viria a propor uns anos mais tarde⁽¹³⁾. Contudo as bases sólidas só começaram a ser definitivamente construídas por Karl Pearson (1857–1936), que introduziu ideias tais como a análise de regressão, o coeficiente de correlação, a família de distribuições de Pearson, o método dos momentos, o teste do qui-quadrado, entre outros. Recordamos que focamos a família de distribuições de Pearson e o teste do qui-quadrado na apresentação da concepção das aplicações do Cálculo das Probabilidades de Pacheco d'Amorim que, possivelmente, poderá ter sido influenciada por trabalhos de Pearson. Outra personalidade importante no nascimento da Estatística Matemática foi William Gosset (1876–1937), que, trabalhando com amostras pequenas no controlo de qualidade da cerveja Guinness, publicou em 1908, sob o pseudónimo de Student, um artigo onde apresenta a estatística actualmente denominada por *t* de Student, depois de investigar o comportamento da média de gaussianas em amostras pequenas,

onde κ_j , para $j = 1, 2, \dots$, são os cumulantes da distribuição X que podem ser obtidos pelas derivadas da função geradora no ponto zero. Os cumulantes gozam de diversas propriedades, de entre as quais se destacam que o cumulante de ordem r , para $r = 1, 2, \dots$, é uma função dos momentos de ordem menor ou igual a r , é homogéneo de ordem r , i.e. $\kappa_r(\alpha X) = \alpha^r \kappa_r(X)$, e gozam da propriedade aditiva para variáveis aleatórias X_i independentes, pois teremos $\kappa_r(\sum_{i=1}^n X_i) = \sum_{i=1}^n \kappa_r(X_i)$.

⁽¹³⁾ Stigler (1978) efectua uma análise detalhada da vida e obra de Edgeworth.

sendo o objectivo do artigo a dedução da distribuição de

$$T = \frac{\overline{X} - \mu}{S}, \quad (9.8)$$

onde $\overline{X}$ representa a média das variáveis aleatórias $X_1, \dots, X_n$ que são independentes e caracterizadas pela distribuição gaussiana com média μ e variância σ^2 e $S^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \overline{X})^2$. Lehmann (1999) salienta que, além da dedução da distribuição da expressão (9.8) (apesar de não ser rigorosa), o grande contributo para a Estatística deste artigo de Gosset é a sua ideia de obter distribuições exactas para pequenas amostras através da imposição de condições iniciais, tal como a distribuição ser conhecida (neste caso a gaussiana). Esta visão é bem distinta da usual na época, e seguida por Pacheco d'Amorim, para quem a única forma de efectuar uma análise de um fenómeno aleatório é através dos resultados assintóticos, nomeadamente as Leis dos Grandes Números e o Teorema Limite Central. Finalmente, com principal relevo a partir de 1920, Ronald Fisher (1890–1962) revoluciona a Estatística introduzindo diversos conceitos, tais como máxima verosimilhança, estatísticas suficientes, estatísticas ancilares, consistência, eficiência, quantidade de informação, testes de significância, distribuições amostrais, variância, entre muitos outros exemplos, que se tornaram fundamentais na Estatística. Fisher (1922, p. 313) considera que existem três tipos distintos de problemas em Estatística:

1. Problemas de especificação — a escolha do modelo estatístico apropriado;
2. Problemas de estimação — a escolha das estatísticas para analisar os parâmetros desconhecidos;
3. Problemas de distribuições — a dedução das distribuições amostrais das estatísticas adoptadas no ponto anterior.

O seu livro *Statistical Methods for Research Workers*, publicado em 1925, é um marco na história da Estatística. Notemos que enquanto Karl Pearson trabalha num modelo mais geral, tal como Pacheco d'Amorim, determinando a distribuição que caracteriza o fenómeno, Fisher emprega essencialmente um modelo paramétrico, supondo conhecido o tipo de distribuição e determinando o(s) parâmetro(s) desconhecido(s). De salientar também os trabalhos de Jerzy Neyman (1894–1981) e Egon S. Pearson (1885–1980), a partir de 1928, relativamente a testes de hipótese, que se destacam dos propostos por Fisher por incluírem uma hipótese alternativa, e, consequentemente, recorrem a um nível de significância α (probabilidade de erro do tipo I, isto é, rejeitar indevidamente a hipótese nula) e potência do teste $1 - \beta$ (onde β corresponde à probabilidade de um erro do tipo II, isto é, aceitar indevidamente a hipótese nula). Diferentemente, os testes de significância de Fisher, baseados no *p-value* que mede a probabilidade de encontrarmos uma amostra da mesma dimensão mais desfavorável para a hipótese testada que a observada considerando (condicionada a) que a hipótese nula é verdadeira (como tal, quanto menor for o seu valor maior é a evidência que a amostra apresenta contra a hipótese nula), não apresentam qualquer hipótese alternativa, pois pretendem unicamente validar um modelo. Lehmann (1993) analisa as diferenças entre estas duas filosofias de testes, que, aliás, geraram diversas discussões entre os seus autores.

Em relação às propostas de aplicação do Cálculo das Probabilidades de Pacheco d'Amorim salientemos, após esta sumária apresentação do contributo de algumas personalidades para a formação da Estatística, que não é, para nós, na apresentação de novos métodos de estimação ou de novos testes de significância que consideramos estar o seu valor, mas antes na ideia filosófica que lhe permite obter a ligação que entre o Cálculo das Probabilidades e a Estatística. De facto, na época, já existiam diversas aplicações de

estimação, paramétrica ou não paramétrica, e de testes de significância. Contudo, não encontramos em nenhuma obra uma construção que fundamente a ligação entre o Cálculo das Probabilidades e a Estatística abarcando uma visão geral, que não seja restrita a um tópico específico de aplicação. Não é demais salientar que, apesar de não conter conceitos actualmente capitais no elo entre estas duas ciências, apresenta grande semelhança com a concepção moderna de Estatística. Pinto (1913), na sua obra inacabada, também apresenta ideias bastante interessantes sobre a construção do conhecimento e a utilização do Cálculo das Probabilidade para modelar fenómenos casuais, numa construção muito rica mas bastante distinta da de Pacheco d'Amorim. Refira-se que, com este objectivo de obter uma ligação entre as Probabilidades e a Estatística, existiram diversas construções, tais como Bernstein (1917)⁽¹⁴⁾, Richard von Mises (1919), John Maynard Keynes (1921), Bruno de Finetti (1937), Jeffreys (1939), entre outros. Estas visões, todas distintas, são posteriores à tese de doutoramento de Pacheco d'Amorim e contribuíram para o desenvolvimento da Estatística, ocupando actualmente um lugar de destaque na História da Teoria da Probabilidade e Estatística. A Tese de Pacheco d'Amorim, escrita numa língua com pouco impacto internacional, ficou esquecida e, por este motivo, não terá tido a mesma influência que as axiomáticas referidas. Contudo, na nossa opinião, também deveria ser uma referência nos livros de História das Probabilidades e da Estatística.

A Estatística clássica ou frequentista, fundamentada na axiomática de Richard von Mises (que apresentaremos resumidamente na secção 10.3), é baseada em hipóteses semelhantes às inerentes à concepção de Pacheco d'Amorim,

⁽¹⁴⁾ Este artigo, em russo, de Bernstein corresponde a uma das primeiras tentativas de axiomatização da Probabilidade, sendo referido por Kolmogoroff (1933). Maistrov (1974) apresenta uma análise da proposta de Bernstein, enquanto Shafer e Vovk (2005) referem a axiomática de Bernstein na sua análise às origens da axiomática de Kolmogoroff.

nomeadamente na repetição dos fenómenos em estudo em condições idênticas.

9.7 Comentário geral ao capítulo

Pacheco d'Amorim, na CONCLUSÃO da sua tese de doutoramento, apresenta a sua concepção da ligação entre o Cálculo das Probabilidades e a Estatística. Enquanto no Cálculo das Probabilidades temos total conhecimento do fenómeno em estudo, porque estará nas condições inerentes ao fenómeno padrão e, como tal, conhecemos a sua distribuição, na Estatística temos que inferir, através da observação repetida do fenómeno, as sua lei de possibilidade (probabilidade). Assim, neste capítulo, Pacheco d'Amorim analisa a metodologia de investigação de um fenómeno que não se enquadre no fenómeno padrão, ou por não termos sido nós os agentes ou por não possuímos total conhecimentos da distribuição que caracteriza o fenómeno. Nesta análise o autor considera que, com base nas Leis de Bernoulli e análogas, podemos obter valores (estimativas) aproximados e prováveis, uma vez que fixamos o erro máximo para a aproximação (estimação) com uma probabilidade tão próxima da unidade quanto seja pretendido. Desta forma, como usual na época, o autor destaca o papel fundamental dos teoremas limites (Leis de Bernoulli e análogas) na ligação entre a Teoria da Probabilidade e a Estatística, isto é, entre o Cálculo das Probabilidades e a realidade. Nesta visão Pacheco d'Amorim considera que, apesar de podermos fixar o erro máximo cometido na aproximação com uma probabilidade que converge para a unidade, por mais próxima da unidade que esta probabilidade se situe nunca poderemos garantir com certeza que a aproximação obtida, através da observação repetida do fenómeno, não comete um erro maior que o estabelecido. Porém, Pacheco d'Amorim afirma que este facto não retira nenhum valor prático à

aproximação obtida, pois podemos ignorar esse *hiato* que separa a certeza da probabilidade e, por este motivo, devemos utilizar o valor aproximado na modelação dos fenómenos aleatórios de forma a ajudar-nos na tomada de decisão.

Desta forma Pacheco d'Amorim analisa as situações onde não se verificam as condições do fenómeno padrão. Em primeiro lugar estuda os casos em que não temos informação completa, isto é, quando não conhecemos a lei de possibilidade e/ou o suporte da escolha aleatória. Nesta metodologia, o autor concebe formas de estimação, paramétrica e não paramétrica, e testes de significância onde, apesar de não formalizar ou explicitar distintamente os processos, apresenta as ideias essenciais dessas aplicações. Neste aspecto as ideias patentes nas aplicações não serão completamente inovadoras, uma vez que estimação baseada nas Lei de Bernoulli e análogas e os testes de significância eram já utilizados por outros autores anteriores à tese de doutoramento de Pacheco d'Amorim. Depois Pacheco d'Amorim analisa os casos em que não somos nós os agentes da escolha à sorte, pois nestas situações não temos como certa a aleatoriedade da escolha. Nesta análise há a salientar a inclusão da concepção de testar se uma sequência de observações, independentemente da sua origem, pode ser considerada como aleatória, uma ideia presente na utilização de números pseudo-aleatórios na simulação e que não encontramos em mais obra nenhuma da época.

Na nossa opinião, é na concepção que Pacheco d'Amorim propõe como elo de ligação entre o Cálculo das Probabilidades e a Estatística que a sua construção se destaca. Consideramos, deste modo, que Pacheco d'Amorim ao fundamentar a ligação entre a Teoria da Probabilidade, quando estamos perante um fenómeno padrão, e a Estatística, nas restantes situações, apresenta uma visão muito moderna, não muito distinta da visão actual, quanto comparada com as obras da mesma época. Com este objectivo, de ligar es-

tes dois campos opostos de análise dos fenómenos casuais, existiram outras fundamentações, com visões distintas da de Pacheco d'Amorim, tais como Bernstein (1917), von Mises (1919), Keynes (1921), de Finetti (1937), Jeffreys (1939), entre outras, mas posteriores à sua tese. Refira-se que nas obras em Teoria da Probabilidade anteriores à tese de doutoramento de Pacheco d'Amorim por nós consultadas, que apresentam aplicações da Probabilidade, tais como Condorcet (1785), Cournot (1838, 1843), Bertrand (1888), Broggi (1907), Laurent (1908), Montessus de Ballore (1908), Borel (1909, 1914), Carvallo (1912), entre outras, não encontramos nenhuma análise que sugira uma ligação entre o Cálculo das Probabilidades e a Estatística de índole geral como a proposta por Pacheco d'Amorim. Neste aspecto é, de facto, notável a semelhança entre as aplicações propostas por Pacheco d'Amorim e a forma como pensamos em Estatística na actualidade, pois muitas ideias, contidas na sua concepção, só posteriormente emergiram.

Parte IV

CONCLUSÃO

Capítulo 10

Comentário Geral à Construção de Pacheco d'Amorim

Como epílogo deste trabalho, após termos comentado a tese de doutoramento de Pacheco d'Amorim capítulo a capítulo, vamos efectuar uma análise geral à sua proposta para a construção do Cálculo das Probabilidades e suas aplicações. Principiaremos esta parte formalizando, com recurso a conceitos mais modernos, a fundamentação proposta por Pacheco d'Amorim na sua tese. Esta formalização pretende evidenciar as semelhanças entre a estruturação da Probabilidade concebida pelo autor no início do século XX e as propostas de fundamentação que surgiram posteriormente e que continuam a ser referência na actualidade. Seguidamente, para justificar a nossa opinião, iremos expor as principais ideias inerentes às diversas propostas, posteriores a 1914, de definição de Probabilidade e sua ligação com as aplicações. Apresentaremos, por isso, os fundamentos de Kolmogoroff (1933), baseados na Teoria da Medida, que representam a primeira axiomática que resolve o sexto problema de Hilbert no que concerne à construção da Probabilidade. Esta axiomática contém as ferramentas matemáticas necessárias para o estudo das Probabilidades, mas não fundamenta uma ligação deste conceito com a realidade, isto

é, com as aplicações. Mostraremos depois as ideias centrais inerentes às duas principais teorias que permitem a ligação entre a Teoria da Probabilidade e a Estatística segundo Plato (1994) e Spanos (1999). Deste modo, analisaremos a visão frequencista de probabilidade de Richard von Mises, baseada em *colectivos*, que surgiu a partir de 1919 e representa a base da Estatística clássica. Examinaremos igualmente a interpretação personalista de probabilidade de Bruno de Finetti, que surge em 1931, mas só se torna mais conhecida a partir do seu famoso artigo *La Prévision: ses Lois Logiques, ses Sources Subjectives*, publicado em 1937, após apresentar um conjunto de conferências no Instituto Henri Poincaré a convite de Fréchet. As ideias de Bruno de Finetti, nomeadamente o seu conceito de permutabilidade, representam o emergir das bases da visão bayesiana subjectiva de Probabilidade e Estatística. Naturalmente outras tentativas de axiomatização, ou propostas de ligação entre a Probabilidade e a realidade existiram mas, neste trabalho, iremos centrar a nossa atenção sobre estas. Não deixaremos de efectuar algumas referências a outras concepções importantes na História da Probabilidade, tais como a axiomática de Bernstein (1917), a visão lógica de Probabilidade, apresentada por John Maynard Keynes e Harold Jeffreys a partir da década 1920, associada à interpretação bayesiana objectiva da Estatística, a interpretação fiducial de Fisher (1930, 1935), a versão propensista de Popper (1959) e a proposta de Rényi, que surge em 1955, baseada no conceito de Probabilidade Condicionada e numa visão frequencista de Estatística, tal como a de Pacheco d'Amorim. Alguns conceitos e/ou interpretações que iremos expor nesta conclusão foram utilizadas anteriormente neste trabalho. Porém, uma vez que pretendemos, neste capítulo, investigar as principais visões de Probabilidade, não nos preocupamos se as ideias que agora apresentamos já tinham sido mostradas previamente, só com a clareza da nossa actual exposição.

Finalizaremos este trabalho com um comentário geral à concepção da

fundamentação da Teoria das Probabilidade e suas aplicações proposta por Pacheco d'Amorim na sua tese de doutoramento.

10.1 Formalização da concepção geral de Pacheco d'Amorim

A formalização que propomos, para descrever a visão de Estatística e a sua ligação com a Teoria da Probabilidade proposta por Pacheco d'Amorim, é baseada em modelos apresentados por Spanos (1999). Consideremos um fenómeno aleatório que é caracterizado por um espaço de medida $(\Omega, \mathcal{A}, \mathbb{P})$ onde, como habitualmente, Ω representa o espaço-amostra ou espaço de resultados, $\mathcal{A}$ a σ -álgebra gerada por uma classe $\mathcal{C}$ de subconjuntos de Ω que corresponde ao espaço dos acontecimentos e $\mathbb{P}$ a medida de probabilidade associada ao par $(\Omega, \mathcal{A})$, conforme a axiomática de Kolmogoroff (1933), que será analisada com mais detalhe na secção 10.2. Deste fenómeno aleatório pretendemos investigar uma característica que é descrita por uma variável aleatória X . Assim, aplicamos a variável aleatória X , que é uma função, ao espaço de probabilidade inicial $(\Omega, \mathcal{A}, \mathbb{P})$, obtendo um novo espaço de probabilidade $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathbb{P}_X)$. Este espaço é distinto do primeiro, porquanto deixamos de trabalhar num espaço abstracto Ω , difícil de caracterizar, para passarmos a operar em $\mathbb{R}$, um espaço bem nosso conhecido, sendo $\mathcal{B}(\mathbb{R})$ a σ -álgebra de Borel sobre $\mathbb{R}$ e $\mathbb{P}_X$ a medida de probabilidade $\mathbb{P}$ transposta para $\mathbb{R}$ pela variável aleatória X . Uma vez que estamos a trabalhar em $\mathbb{R}$, sendo $\mathbb{P}_X$ uma função de conjuntos de $\mathcal{B}(\mathbb{R})$ (logo de $\mathbb{R}$), podemos caracterizar plenamente este espaço de probabilidade por um modelo de probabilidade $\Phi = \{F(x|\theta), \theta \in \Theta, x \in \mathbb{S}_X\}$ que é descrito por três características:

1. a forma particular da função de distribuição $F_X(x|\theta)$ associada a pa-

râmetros desconhecidos θ (onde $\theta = (\theta_1, \dots, \theta_m)$ é um vector de parâmetros);

2. o espaço dos parâmetros — Θ (o domínio do vector de parâmetros θ);
3. o suporte $\mathbb{S}_X$ da função densidade⁽¹⁾ que pode ser definido por

$$\mathbb{S}_X = \{x \in \mathbb{R} : \mathbb{P}_X [x - \epsilon < X(\omega) < x + \epsilon] > 0, \forall \epsilon > 0\}, \quad (10.1)$$

que, em relação à função densidade, corresponderá a

$$\mathbb{S}_X = \{x \in \mathbb{R} : f(x|\theta) > 0\}. \quad (10.2)$$

Assim, $\mathbb{S}_X$ representa os valores de $\mathbb{R}$ que a variável aleatória pode assumir, mesmo que seja com probabilidade nula, como acontece com qualquer ponto numa variável aleatória contínua.

Temos, desta forma, formalizada uma visão paramétrica da Estatística. De forma semelhante, mas mais geral, podemos pensar num modelo não paramétrico, onde temos que definir a função de distribuição $F(x) \in \mathcal{P}_F \subset \mathcal{P}$, onde $\mathcal{P}$ representa o conjunto de todas as distribuições e $\mathcal{P}_F$ representa um subconjunto de $\mathcal{P}$ apropriado para modelar o fenómeno em análise, sendo definido pelas características deste (o suporte $\mathbb{S}_X$ do fenómeno, se é discreto ou contínuo, o número de dimensões, simetria, achatamento, entre outras características). O objectivo de $\mathcal{P}_F$ é diminuir o conjunto de distribuições disponíveis para um conjunto mais adequado. Por exemplo, na situação anterior correspondente a um modelo paramétrico, teremos $\mathcal{P}_F = \Phi$, onde reduzimos as possíveis distribuições a uma única distribuição dependente do valor de um parâmetro (ou vector de parâmetros) θ . Salientemos que, normalmente, recorrendo a conceitos actuais, quanto mais restrita for a colecção

⁽¹⁾ $\mathbb{S}_\mu$ é o suporte de uma medida definida no espaço $(X, \mathcal{F}, \mu)$ se e só se $\mu(\mathbf{A}) = \mu(\mathbf{A} \cap \mathbb{S}_\mu)$ para qualquer $\mathbf{A}$ em $\mathcal{F}$, ou seja, se e só se $\mu(\overline{\mathbb{S}_\mu}) = 0$.

$\mathcal{P}_F$ mais precisas serão as nossas conclusões de inferência, pois obteremos estimadores com menor variância, testes mais potentes, intervalos de confiança e de previsão com menor amplitude. Por outro lado, quanto mais restritiva for a colecção $\mathcal{P}_F$ maior será a probabilidade de cometermos erros de especificação do modelo e, dependendo da robustez dos métodos utilizados na inferência em relação às hipóteses indevidamente incluídas no modelo, as conclusões poderão ser completamente erróneas.

Desta formalização destaque-se a ideia de que é bastante mais atractiva a análise efectuar-se no modelo probabilista, seja paramétrico ou não, como usualmente operamos em Estatística, que no espaço de probabilidade inicial $(\Omega, \mathcal{A}, \mathbb{P})$, tendo a variável aleatória e a função de distribuição um papel fundamental nesta transformação. Com esta ideia podemos, de forma simplificada⁽²⁾, visualizar o objectivo da Estatística na determinação (estimação), recorrendo a n concretizações independentes da variável aleatória X (que correspondem a uma amostra aleatória $X_1, X_2, \dots, X_n$ de dimensão n), da função de distribuição F_X , que servirá para modelar o fenómeno investigado, ou no testar se uma função de distribuição assumida *a priori* se harmoniza com os dados observados. Desta forma, a Estatística tem este objectivo de especificação da função de distribuição F_X , para depois testar a especificação efectuada e, se necessário, efectuar uma nova especificação. Esta ideia cíclica de especificação, teste e re-especificação, típica de qualquer ciência, é muito bem apresentada na obra póstuma e incompleta de José Pinto⁽³⁾, con-

⁽²⁾ Estamos a restringir, nesta apresentação, à situação de independência das variáveis aleatórias utilizadas para a inferência, que corresponde à situação abordada por Pacheco d'Amorim. Contudo o papel da Estatística é bastante mais vasto, não se restringindo a modelar uma variável aleatória mas, muitas vezes, tentando modelar sequências de variáveis interligadas como, por exemplo, quando modelamos um processo estocástico.

⁽³⁾ José Freire de Sousa Pinto (1855–1911) foi professor da Faculdade de Matemática da Universidade de Coimbra.

temporânea à tese de doutoramento de Pacheco d'Amorim e publicada nos *Annaes da Academia Polytechnica do Porto* em 1913, na qual é apresentada uma visão geral das aplicações do Cálculo das Probabilidades. Esta obra, apesar de ser direccionada para a Teoria dos Erros, que o autor considera uma parte integrante da Estatística, apresenta uma visão da fundamentação das aplicações da Teoria da Estatística. O autor começa por expor o processo de construção do conhecimento em qualquer ciência, concluindo que o avanço do conhecimento advém da combinação entre a observação do fenómeno e o raciocínio teórico. Assim, qualquer teoria começa pela observação do fenómeno, através da qual se forma uma teoria, que inicialmente será uma aproximação. Depois observamos uma vez mais o fenómeno. Esta nova observação servirá para testar a primeira teoria formada, que é confirmada ou aperfeiçoada com o novo conhecimento proveniente desta nova observação ou alterada, caso a teoria inicial venha a ser considerada errónea. É nesta cadeia sucessiva de duplo esforço, observação e raciocínio, que o conhecimento científico avança no sentido do conhecimento completo do fenómeno, onde a teoria formalizada acontecerá de certeza sempre que o fenómeno ocorra. Contudo, Pinto considera que existem alguns fenómenos que não podem ser totalmente teorizados, isto é, que não podemos “*passar à categoria de certeza*”, pois correspondem a fenómenos casuais. O autor destaca então dois tipos de probabilidades: as probabilidades aleatórias, referentes a jogos de azar, onde a equipossibilidade é clara, e a probabilidade dos fenómenos, onde é através da repetição dos fenómenos que devemos determinar o valor da probabilidade. Assim, o autor considera que a repetição continuada dos fenómenos, que tenha, por exemplo, dois resultados possíveis, **A** e **A'**, deverá mostrar a sua relação teleológica, isto é, como se ao acontecimento **A** estivessem associados α resultados favoráveis e ao acontecimento **A'** estivessem α' e, à medida que vamos observando o fenómeno, o rácio entre o número de vezes

que ocorrerá $\mathbf{A}$, seja n , e o número de vezes que ocorrerá $\mathbf{A}'$, seja n' , tenderá a manifestar a relação $\frac{\alpha}{\alpha'}$. Assim, o rácio $\frac{n}{n'}$ vai convergir, inicialmente de forma desordenada porém aproximando-se cada vez mais de um estado médio que representa a sua relação teleológica $\frac{\alpha}{\alpha'}$, supondo naturalmente invariabilidade do fenómeno. José Pinto fundamenta a sua posição:

“ao passo que as Probabilidades de LAPLACE, apesar do superior desenvolvimento da sua analyse, representam um meio restricto, artificial e como quasi separado do campo scientifico, a orientação iniciada por BERNOULLI na analyse dos seus theoremas e generalizada e accentuada mais tarde nas mathematicas philosophicas de WRONSKI, converteram, como dissémos, as Probabilidades num fecundo recurso scientifico, cujo emprego na actualidade se tem tornado geral, pela criação dos estudos estatísticos, que nellas tem o seu fundamento”

[Pinto, 1913, p. 16]

Deste modo, Pinto claramente considera que as Leis de Bernoulli são a ferramenta para a determinação das probabilidades quando não podemos aplicar a definição clássica de probabilidade, cuja utilidade restringe praticamente aos jogos de azar. Assim, a estas Leis devemos a importância da Estatística como ciência para conhecermos os fenómenos. A visão de Pinto é, em certa parte, semelhante à de Pacheco d'Amorim, pois ambos os autores apresentam uma visão geral baseada nas Leis de Bernoulli, considerando que a partir destas podemos determinar as probabilidades associadas ao fenómeno casual, através da sua realização repetida, supondo o fenómeno invariável. Contudo, apesar de Pinto salientar as Leis de Bernoulli para a determinação das probabilidades quando não é aplicável a definição clássica de probabilidade, não analisa a fundamentação do Cálculo das Probabilidades nem foca o caso contínuo, como Pacheco d'Amorim faz. Salientemos, todavia, a ri-

queza da construção filosófica da fundamentação para a determinação das probabilidades com recurso à inversa das Leis de Bernoulli, efectuada por Pinto (1913), numa das poucas obras Portuguesas dedicadas ao Cálculo das Probabilidades e à Estatística na época.

Após esta merecida referência histórica, retomemos à nossa formalização. Caso, a certa altura, no ciclo de especificação, testes de especificação e re-especificação, o teste de especificação sobre a validade da distribuição F_X para caracterizar um fenómeno não a rejeite, considera-se o modelo estatisticamente, ou empiricamente, adequado, uma vez que se harmoniza com os dados observados e, como tal, a função F_X pode ser utilizada para modelar o fenómeno. Por este motivo, nesta asserção, é usualmente evitado o termo *verdadeiro modelo* e usado o termo *modelo validado*, uma ideia que Pacheco d'Amorim deixa bem clara na CONCLUSÃO da sua tese.

Consideramos que, nesta perspectiva, a função F_X é suficiente para descrever o fenómeno e, como tal, a ponte de ligação entre as regularidades empíricas e a probabilidade é a relação entre a função de distribuição empírica (da amostra) $\hat{F}(x)$ e a função de distribuição teórica (da população) F_X , assumindo os Teoremas Limites que ligam $\hat{F}(x)$ e F_X uma importância capital (tal como as Leis de Bernoulli e análogas). Desta forma, entendemos que qualquer faceta da inferência estatística (estimação, testes de hipótese e/ou de significância e previsão) pode ser interpretada por esta relação entre a função de distribuição empírica $\hat{F}(x)$ e a função de distribuição teórica F_X . Esta especificação dependerá das características que as observações, amostra aleatória $(X_1, \dots, X_n)$, possuem, nomeadamente em relação à independência e homogeneidade distribucional. Assim, no caso típico de uma amostra aleatória, onde as variáveis aleatórias X_i são independentes e identicamente distribuídas, representando por $f(X_1, \dots, X_n|\theta)$ a função densidade conjunta

que depende de um vector de parâmetros θ , teremos

$$f(X_1, \dots, X_n | \theta) = \prod_{i=1}^n f_i(X_i | \theta) = f^n(X | \theta), \quad (10.3)$$

onde $f_i(X_i | \theta) = f(X | \theta)$ representa a função densidade de cada variável aleatória X_i supondo igualdade distribucional. No caso geral, não impondo independência nem homogeneidade, teríamos

$$f(X_1, \dots, X_n | \theta) = f_1(X_1 | \theta_1) \prod_{i=2}^n f_i(X_i | \theta_i, X_{i-1}, \dots). \quad (10.4)$$

Na Estatística clássica os estimadores e as estatísticas de testes correspondem a funções $h(X_1, \dots, X_n, \theta)$ da amostra aleatória $(X_1, \dots, X_n)$ e de parâmetros θ e são, por este motivo, variáveis aleatórias. Deste modo, as suas distribuições (distribuições amostrais) assumem um papel fundamental na inferência estatística.

Na nossa opinião, a ideia de Pacheco d'Amorim é muito semelhante à visão que acabamos de expor e que corresponde à que habitualmente usamos nas aplicações da Teoria da Probabilidade na Estatística clássica. Assim, em primeiro lugar, não efectuamos qualquer referência ao espaço inicial e unicamente nos preocupamos com a especificação da distribuição da variável aleatória X . Para Pacheco d'Amorim, apesar de não recorrer aos conceitos da Teoria da Medida, que era uma área ainda muito incipiente na época, qualquer fenómeno aleatório provém de uma tiragem de um elemento à sorte numa classe finita $\mathbf{A}$ ou de um lançamento de um ponto à sorte numa região limitada $\mathbf{X}$, que são caracterizados pela equipossibilidade, a que poderemos associar o espaço inicial $(\Omega, \mathcal{A}, \mathbb{P})$. Recordemos que, tal como defendemos na análise ao capítulo PONTO IMAGEM, o espaço de probabilidade abstracto $(\Omega, \mathcal{A}, \mathbb{P})$ pode ser caracterizado pela equipossibilidade dos elementos ω de Ω . No entanto, a característica que observamos do fenómeno aleatório poderá não corresponder directamente à tiragem ou lançamento inicial em $\mathbf{X}$,

mas a uma sua imagem $y = f(x)$, proveniente de uma escolha à sorte em $\mathbf{Y} = f(\mathbf{X})$, que corresponderá ao espaço de probabilidade $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathbb{P}_X)$. Assim, o ponto imagem, tal como a variável aleatória X , permite-nos passar do espaço $(\Omega, \mathcal{A}, \mathbb{P})$, caracterizado pela equipossibilidade, para o espaço de probabilidade $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathbb{P}_X)$, que poderá ser caracterizado por uma qualquer distribuição. Para Pacheco d'Amorim o ponto imagem, função do ponto lançado à sorte, desempenha o mesmo papel que as variáveis aleatórias, funções do espaço inicial, pois permite passar de uma tiragem ou lançamento directo, obrigatoriamente caracterizado pela equipossibilidade, para uma sua imagem que poderá ser caracterizada por qualquer lei de possibilidade. Refira-se ainda que, enquanto a classe ou região $\mathbf{X}$ tem de ser finita (restrição inerente à equipossibilidade), a classe ou região $\mathbf{Y}$ pode ser ilimitada. Finalmente, como expusemos previamente, na prática só nos preocupamos com a determinação da função de distribuição da variável aleatória X . Pacheco d'Amorim também se preocupa somente com a determinação da lei de possibilidade π_y (e com o seu suporte $\mathbf{Y}$ que é definido na determinação de π_y), não se interessando pelo região inicial $\mathbf{X}$. Desta forma, o modelo de Pacheco d'Amorim corresponde à visão mais geral (não paramétrica) que apresentamos. Note-mos que a única diferença em relação às três características essenciais num modelo de probabilidade paramétrico $\Phi = \{F(x|\theta), \theta \in \Theta, x \in \mathbb{S}_X\}$, provavelmente mais usual nas aplicações na actualidade, é que não há referência a parâmetros nem naturalmente a espaço de parâmetros. Esta distinção é justificada pelo facto de Pacheco d'Amorim não ter disponível as famílias de distribuições⁽⁴⁾, que unicamente se distinguem pelos valores assumidos pe-

⁽⁴⁾ Pacheco d'Amorim, como referimos diversas vezes ao longo deste trabalho, só efectua cinco referências em toda a sua tese de doutoramento, sendo, por este motivo, difícil avaliar as suas influências. Deste modo, é-nos impossível avaliar se os trabalhos de Karl Pearson, nomeadamente em relação à sua família de distribuições e teste do qui-quadrado, eram do seu conhecimento nesta época. Contudo, ao longo da sua tese, Pacheco d'Amorim

los parâmetros, nomeadamente de escala e localização. Deste modo, para Pacheco d'Amorim, não existe a ideia de estimação paramétrica, onde já conhecemos a forma da função $F(x|\theta)$ mas desconhecemos os valores dos parâmetros θ , existindo unicamente a estimação não paramétrica, isto é, a situação onde desconhecemos totalmente a forma da lei de probabilidade π_y . Assim, teremos que estimar π_y com base numa sequência de tiragens ou lançamentos idênticos, que correspondem às n concretizações independentes da variável aleatória X , isto é, à nossa amostra aleatória $(X_1, \dots, X_n)$.

O que foi exposto permite-nos concluir que a visão que Pacheco d'Amorim propõe na sua tese de doutoramento é, apesar de não conter a formalização matemática necessária para conseguirmos efectuar as estimações e testes que indica, em muitos aspectos semelhante à que actualmente utilizamos. Contudo, não deixamos de salientar esta sua visão moderna de Estatística, pelo facto de a maior parte das obras que focam as aplicações do Cálculo das Probabilidades, por nós consultadas, não inserirem uma generalidade de tratamento nem uma visão de Estatística e das suas ligações com a Probabilidade que seja comparável à proposta na tese de doutoramento de Pacheco d'Amorim (conforme comentamos na secção 9.6).

10.2 *Grundbegriffe* de Kolmogoroff (1933)

“The theory of probability, as a mathematical discipline, can and should be developed from axioms in exactly the same way as Geometry and Algebra. This means that after we have defined the elements to be studied and their basic relations, and have stated the axioms by which these relations are to be governed, all further exposition must be based exclusively on these axioms, independent

unicamente utiliza as distribuições binomial e gaussiana.

of the usual concrete meaning of these elements and their relations.”

[Kolmogoroff 1933, p. 1]

Kolmogoroff, em 1933, apresenta a primeira axiomática aceite pela generalidade de seus pares, resolvendo assim o problema proposto por Hilbert em 1900. Refira-se, contudo, que nem todos os axiomas apresentados por Kolmogoroff são de pacífica aceitação, nomeadamente o sexto, axioma da continuidade da medida, que nos garante a σ -aditividade, que não é aceite em diversas interpretações de probabilidade, como, por exemplo, a visão apresentada por Bruno de Finetti (1937), que analisaremos na secção 10.4. Com esta axiomática Kolmogoroff consegue criar a base matemática que serve de alicerce para o desenvolvimento, sem paradoxos, da Teoria da Probabilidade. Nesta secção vamos apresentar a axiomática proposta por Kolmogoroff mais detalhadamente, estabelecendo um compromisso entre a simbologia adoptada pelo autor e uma simbologia mais moderna e efectuando um paralelo entre a sua fundamentação e a Teoria da Medida. Na exposição tentaremos focar também a origem de algumas patologias existentes na Teoria da Probabilidade antes do tratamento da probabilidade como uma medida, como, por exemplo, a existência de conjuntos não mensuráveis.

10.2.1 Contextualização

Conforme referimos na INTRODUÇÃO deste trabalho, no Congresso Internacional de Matemática de Paris de 1900, David Hilbert, na sua famosa alocução, expõe um conjunto de 23 problemas por resolver que deveriam orientar a investigação em Matemática durante o século XX. Um desses problemas, o sexto, consistia na axiomatização das áreas da Física onde a Matemática desempenha um papel predominante, sendo o Cálculo das Probabilidades e a Mecânica as primeiras da lista.

“To treat in the same manner, by means of axioms, those physical sciences in which mathematics plays an important part; in the first rank are the theory of probabilities and mechanics.”

[Hilbert, 1902, pág. 418]

Nesta época, a Teoria da Probabilidade (ou Cálculo das Probabilidades, como nessa altura é apelidada, nomeadamente pela escola francesa) é assombrada por diversos paradoxos, tais como os famosos paradoxos de Bertrand (1888)⁽⁵⁾, que ilustram a ambiguidade existente em diversas noções básicas tais como a *escolha ao acaso* (*au hazard*), sendo necessária a construção de uma teoria para clarificar estes conceitos. Existiram diversas tentativas de axiomatização da Probabilidade anteriores ao tratado de Kolmogoroff, algumas delas são analisadas em Plato (1994), Bingham (2000), Shafer e Vovk (2006) e em Santos (2005). Nesta tese, além dos fundamentos de Kolmogoroff, desenvolveremos apenas duas destas primeiras tentativas de fundamentação da probabilidade, a visão de Richard von Mises, que surge por volta de 1919, e a de Bruno de Finetti, que aparece no início da década de 1930. Contudo, a primeira axiomatização de probabilidade que definitivamente resolveu os paradoxos patentes nesta teoria foi apresentada por Kolmogoroff em 1933, intitulada *Grundbegriffe der Wahrscheinlichkeitstheorie* e traduzida para inglês em 1950 por *Foundations of the Theory of Probability*.

A axiomática de Kolmogoroff baseia-se nas características das funções de conjuntos, nomeadamente na área actualmente denominada por Teoria da Medida, podendo definir-se probabilidade como uma medida em que a probabilidade do universo é unitária. Uma vez que a Probabilidade é uma medida e, como tal, uma função de conjuntos, é importante analisar todas as

⁽⁵⁾ Na secção 5.1.2 expusemos o paradoxo da escolha, à sorte, de uma corda num círculo, um dos mais notáveis paradoxos criados por Bertrand. Székely (1986) explora bem diversos paradoxos na evolução da Teoria da Probabilidade e da Estatística.

características desta função, nomeadamente o seu domínio, o seu contradomínio e as suas especificidades, que a tornam numa função tão fundamental na aplicabilidade da ciência à realidade. Por este motivo, vamos analisar e apresentar os resultados que consideramos essenciais para a compreensão da axiomática de Kolmogoroff.

“The fact that our additive set function $\mathbb{P}(A)$ is non-negative and satisfies the condition $\mathbb{P}(\Omega) = 1$, does not itself cause new difficulties. Random variables (see Chap. III) from a mathematical point of view represent merely functions measurable with respect to $\mathbb{P}(A)$, while their mathematical expectations are abstract Lebesgue integrals.”

[Kolmogoroff 1933, p. 8]

As bases para que a probabilidade fosse tratada como uma medida foram criadas, essencialmente, por Émile Borel (1871–1956), Henri Lebesgue (1875–1941), Maurice Fréchet (1878–1973), Constantin Carathéodory (1873–1950), Johann Radon (1887–1956), Otto Nikodym (1889 – 1974), entre outros. Nos finais dos século XIX eram sentidas diversas dificuldades teóricas associadas ao integral de Riemann, nomeadamente no que se refere à operação de passagem ao limite, havendo a necessidade de estender a teoria da integração a uma classe superior. Para que tal fosse possível, foram essenciais diversos desenvolvimentos teóricos. Por exemplo, Borel, na sua tese de doutoramento defendida em 1895, introduz o conceito de funções de conjuntos que gozam da aditividade numerável ou σ -aditividade, isto é, funções μ onde se verifica

$$\mu \left(\bigcup_{i=1}^{+\infty} \mathbf{A}_i \right) = \sum_{i=1}^{+\infty} \mu(\mathbf{A}_i), \quad (10.5)$$

para conjuntos $\mathbf{A}_i$ disjuntos dois a dois ($\mathbf{A}_i \cap \mathbf{A}_j = \emptyset, \forall i \neq j$). Borel, nessa obra, apresenta também o conceito de conjuntos de medida nula, que é uma noção essencial na atribuição de medida a conjuntos. Estes dois conceitos

são fundamentais para o desenvolvimento da Teoria da Medida. Assim, na sua tese de doutoramento, Borel tenta determinar a classe $\mathcal{C}_n$ de subconjuntos de $\mathbb{R}^n$, que seja uma álgebra fechada em relação a uniões numeráveis (ou seja, na terminologia actual, $\mathcal{C}_n$ é uma σ -álgebra sobre $\mathbb{R}^n$) e que contenha os conjuntos elementares (integráveis à Riemann), e uma função $m_n : \mathcal{C}_n \rightarrow [0, +\infty]$ que seja σ -aditiva e que, para qualquer conjunto elementar $E \subset \mathbb{R}^n$ verifique $m_n(E)$ igual ao valor do integral de Riemann do conjunto E . Em suma, Borel pretende, desta forma, encontrar uma extensão da integração à Riemann. Lebesgue (1902), também na sua tese de doutoramento “*Intégrale, longueur, aire*”, apresenta a sua definição de medida e de conjuntos mensuráveis, utilizando as ideias anteriores de Jordan e as novas ideias de Borel, nomeadamente na utilização de funções σ -aditivas definidas numa classe adequada, uma σ -álgebra. Deste modo, caso a função σ -aditiva seja não negativa teremos uma medida. Baseando-se na sua medida construiu o integral de Lebesgue, que simplifica bastante o processo de operar com limites no integral, demonstrando alguns teoremas de convergência, tais como o Teorema da Convergência Dominada e da Convergência Monótona. Estes resultados mostram bem as vantagens de recorrer ao integral de Lebesgue, em vez de utilizarmos o integral de Riemann, pois permite-nos ter, sob algumas condições, os conjuntos limites como mensuráveis. Contudo, estes trabalhos de Lebesgue ainda são restritos ao espaço Euclidiano. Fréchet, na sua tese de doutoramento de 1906, introduz a axiomática dos espaços métricos, começando a abstracção da medida de Lebesgue com a libertação da sua ligação aos espaços Euclidianos, sendo posteriormente a separação finalizada com os trabalhos de Carathéodory. Radon, em 1913, contribuiu para a generalização destas visões, considerando que a medida de Lebesgue não é mais que um caso particular de um objecto matemático, hoje denominado por medida, e considerando ainda que qualquer medida pode ser utilizada

para definir integrais de funções. Desta forma, conjuntamente com os avanços de Borel e Lebesgue, estavam lançadas as sementes para o nascimento da axiomática da Teoria da Medida.

Borel (1909a), num artigo que comentámos previamente, introduz a σ -aditividade na probabilidade, o princípio de Borel e a Lei Forte dos Grandes Números. Hugo Steinhaus (1887–1972) e Antonine Lomnicki (1881–1941) apresentam, em 1923, duas propostas de axiomatização da probabilidade, baseadas no princípio de Borel e na análise dos números normais. Steinhaus (1923) propõe uma axiomática baseada em cinco axiomas referentes ao conjunto Ω de todas as sequências infinitas de ρ 's (*rouge*) e η 's (*noir*) (que podem ser zeros e uns, bolas branca e pretas, etc.), a uma classe $\mathcal{A}$ constituída por todos os subconjuntos de Ω e uma função de conjuntos μ , definida para todos os conjuntos $\mathbf{E}$ pertencente a uma classe $\mathcal{B}$ contida em $\mathcal{A}$, tal que:

1. $\mathbb{P}(\mathbf{E}) \geq 0, \quad \forall \mathbf{E} \in \mathcal{B};$
2. (a) Seja e uma sequência finita de n elementos ρ 's e η 's — o conjunto $\mathbf{E}$ que consiste em todas as sequência infinitas que começam com e pertence a $\mathcal{B}$;
- (b) Se duas sequências e_1 e e_2 diferem unicamente numa casa, então os seus correspondentes conjuntos $\mathbf{E}_1$ e $\mathbf{E}_2$ terão a mesma probabilidade, isto é $\mu(\mathbf{E}_1) = \mu(\mathbf{E}_2)$;
- (c) $\mu(\Omega) = 1$;
3. $\mathcal{B}$ é fechado para uniões finitas e contáveis de elementos disjuntos, sendo μ aditiva e σ -aditiva;
4. Se $\mathbf{E}_1 \supset \mathbf{E}_2$ e $\mathbf{E}_1$ e $\mathbf{E}_2$ pertencem a $\mathcal{B}$, então $\mathbf{E}_1 - \mathbf{E}_2$ também pertence a $\mathcal{B}$;

5. Se $\mathbf{E} \in \mathcal{B}$ e $\mu(\mathbf{E}) = 0$, então qualquer subconjunto de $\mathbf{E}$ também pertence a $\mathcal{B}$.

Notemos que Steinhaus claramente separa a aditividade da σ -aditividade, isto é, não considera, tal como muitos autores anteriores, por exemplo Broggi, que a aditividade finita implica a aditividade numerável. Segundo Shafer e Vovk (2006), no mesmo ano Sierpinski apresenta uma axiomática para a medida de Lebesgue que consiste nos axiomas apresentados por Steinhaus, retirando o segundo axioma, que é substituído por outro que indica que a medida $\mu(\mathbf{I})$, do intervalo $\mathbf{I} \subset \mathbb{R}$, corresponde ao seu comprimento. Lomnicki (1923) apresenta uma construção também baseada no artigo de Borel (1909a) e na σ -aditividade, dividindo o Cálculo das Probabilidades em quatro fases. Deste modo, Lomnicki considera que, para determinar qualquer probabilidade, deveremos especificar:

1. O conjunto $\mathcal{M}$ onde a probabilidade é examinada (conjunto dos “*casos possíveis*”);
2. O conjunto $m \subset \mathcal{M}$ onde pretendemos determinar a probabilidade (conjunto dos “*casos favoráveis*”);
3. a distribuição dos *pesos*, isto é, a função que a cada elemento $e \in \mathcal{M}$ associa um número $\varphi(e)$ (grau de possibilidade de cada caso);
4. o modo de *mensuração* dos conjuntos de elementos de $\mathcal{M}$.

Lomnicki representa a probabilidade de m dentro de $\mathcal{M}$ por $p_{\mathcal{M}}(m, \varphi_{\mathcal{M}}(e))$ que, apesar de habitualmente ser representada através de $p_{\mathcal{M}}(m)$, considera ser uma função de três argumentos: o conjunto dos casos possíveis, o conjunto dos casos favoráveis e a função que atribui um peso a cada elemento dos casos possíveis. Nesta construção a probabilidade é sempre condicionada e

a função $\varphi(e)$ é uma função qualquer arbitrária (que será constante quando existe equipossibilidade), definida para qualquer elemento $e \in \mathcal{M}$, não sendo obrigatório que, por exemplo, no caso finito, a soma dos pesos dos elementos de $\mathcal{M}$ seja igual à unidade. Lomnicki divide a sua análise em três situações consoante as especificações do conjunto $\mathcal{M}$, correspondendo aos conjuntos finitos, infinitos numeráveis e não numeráveis.

Contudo, alguns resultados indispensáveis para a construção de Kolmogoroff só surgiram posteriormente. O Teorema da Extensão de Carathéodory, demonstrado em 1930 por Carathéodory, é essencial para garantir a existência e unicidade de uma medida no espaço de probabilidade quando o espaço de resultados é infinito não numerável. Outro resultado capital para a construção de Kolmogoroff é o Teorema de Radon-Nikodym, demonstrado por Radon, em 1913, para o espaço Euclidiano, sendo somente demonstrado em 1930 o caso geral por Nikodym. Este teorema é o fundamento da definição da probabilidade condicionada de Kolmogoroff, que, desta forma, garante a existência de um espaço de probabilidade bem definido para a probabilidade condicionada, evitando paradoxos existentes devido à ausência de uma definição rigorosa deste conceito essencial na construção da Teoria da Probabilidade e da Estatística (inclusive na construção de Pacheco d'Amorim).

Após este breve resumo histórico de alguns acontecimentos importantes para a obra de Kolmogoroff (1933), vamos expor as ideias principais presentes nos *fundamentos* deste autor⁽⁶⁾.

⁽⁶⁾ *Kolmogoropv's Heritage in Mathematics*, editado por Charpentier, Lesne e Nikolski (2007), apresenta uma resenha das principais contribuições de Kolmogoroff para o desenvolvimento da Matemática. O terceiro capítulo desta obra é dedicado ao seu trabalho na Teoria da Probabilidade.

10.2.2 Axiomática

Vamos apresentar a forma como Kolmogoroff construiu a sua medida de probabilidade, primeiro para as experiências aleatórias cujo espaço de resultados é finito, recorrendo a cinco axiomas, e depois para as que possuem um espaço de resultados infinito (numerável ou não), acrescentando o axioma da continuidade da medida aos axiomas anteriores e recorrendo ao Teorema de Extensão de Carathéodory. Apesar de esta axiomática, na actualidade, ser habitualmente apresentada recorrendo unicamente a três axiomas (ou quatro se considerarmos que o terceiro axioma é distinto consoante trabalhemos com um espaço de resultados finito ou infinito), que são equivalentes, vamos utilizar a sua forma original. Apontamos ainda algumas consequências da axiomática de Kolmogoroff, revelando as propriedades essenciais para o desenvolvimento da Teoria da Probabilidade.

Caso finito

Começemos por considerar uma experiência aleatória $\mathcal{E}$ com possíveis resultados $\Omega = \{\omega_1, \omega_2, \dots, \omega_n\}$, onde ω_i , $i = 1, \dots, n$ representam os possíveis resultados da experiência $\mathcal{E}$, que são denominados por acontecimentos elementares, sendo Ω denominado por universo, espaço de resultados, espaço amostra ou espaço dos acontecimentos elementares ω_i . Consideremos ainda uma classe $\mathcal{A}$, não vazia, de subconjuntos de Ω , sendo os elementos deste conjunto denominados por acontecimentos aleatórios. Neste contexto Kolmogoroff em 1933 apresenta os seguintes axiomas:

Ai) $\mathcal{A}$ é um corpo de conjuntos;⁽⁷⁾

⁽⁷⁾ Um sistema de conjuntos é denominado um corpo se for fechado para as operações habituais de conjuntos, isto é, se o conjunto obtido através da união, da intersecção ou da diferença entre dois conjuntos do sistema também pertencer ao sistema.

Aii) $\Omega \subset \mathcal{A}$;

Aiii) Para cada conjunto $\mathbf{A} \subset \mathcal{A}$ é associado um número real não negativo $\mathbb{P}(\mathbf{A})$ denominado probabilidade do acontecimento $\mathbf{A}$;

Aiv) $\mathbb{P}(\Omega) = 1$;

Av) Aditividade — Sejam $\mathbf{A}$ e $\mathbf{B}$ subconjuntos de $\mathcal{A}$, se $\mathbf{A}$ e $\mathbf{B}$ não têm elementos comuns ($\mathbf{A} \cap \mathbf{B} = \emptyset$), então

$$\mathbb{P}(\mathbf{A} \cup \mathbf{B}) = \mathbb{P}(\mathbf{A}) + \mathbb{P}(\mathbf{B}). \quad (10.6)$$

Desta forma, em primeiro lugar temos de construir o espaço amostra, onde estão definidos todos os possíveis resultados da experiência aleatória $\mathcal{E}$. Depois temos de criar o espaço de acontecimentos, que define quais são os subconjuntos de Ω que vão ser probabilizáveis, isto é, que terão atribuída uma probabilidade. Este espaço está definido nos dois primeiros axiomas (que, na apresentação usual da axiomática de Kolmogoroff não são utilizados, pois partimos logo de um espaço mensurável, não sendo necessário definir, através de axiomas, as características do domínio da função $\mathbb{P}$), correspondendo ao que actualmente designamos por uma álgebra sobre Ω ⁽⁸⁾. Notemos que considerar $\mathcal{A}$ um corpo e simultaneamente considerar $\Omega \subset \mathcal{A}$ corresponde a afirmar que $\mathcal{A}$ é uma álgebra sobre Ω . Uma forma possível de construir esta classe de conjuntos que definem o domínio da função $\mathbb{P}$, no caso finito, é utilizar o conjunto potência do conjunto Ω , ou conjunto de todas as partes de Ω , que corresponde ao conjunto de todos os possíveis subconjuntos de Ω , sendo representado por 2^Ω ou $\mathcal{P}(\Omega)$. Desta forma, $\mathcal{P}(\Omega)$ contém todos os possíveis subconjuntos de Ω . A notação 2^Ω advém da situação em que

⁽⁸⁾ Considerando um conjunto $\mathbf{X}$, não vazio, e $\mathcal{F}$ uma família de subconjuntos de $\mathbf{X}$, dizemos que $\mathcal{F}$ é uma **álgebra** sobre $\mathbf{X}$ se e só se $\mathbf{X} \in \mathcal{F}$, $\mathcal{F}$ for fechada em relação à união ($\mathbf{A}, \mathbf{B} \in \mathcal{F} \Rightarrow \mathbf{A} \cup \mathbf{B} \in \mathcal{F}$) e em relação à diferença ($\mathbf{A}, \mathbf{B} \in \mathcal{F} \Rightarrow \mathbf{A} - \mathbf{B} \in \mathcal{F}$).

Ω é finito, constituído por n elementos, pois nestes casos o conjunto potência é constituído por 2^n elementos. Notemos que, considerando o conjunto $\Omega = \{\omega_1, \omega_2, \dots, \omega_n\}$, teremos 2^Ω constituído por $\binom{n}{0} = 1$ conjunto com zero elementos (conjunto vazio — $\emptyset$); $\binom{n}{1} = n$ conjuntos com um elemento (que correspondem aos acontecimentos elementares $\{\omega_i\}, i = 1, \dots, n$); $\binom{n}{2}$ conjuntos possíveis de formar com dois acontecimentos elementares $\{\omega_i, \omega_j\}$, para $i, j = 1, \dots, n$ e $i \neq j$; $\dots$; $\binom{n}{k}$ conjuntos possíveis de formar com k acontecimentos elementares distintos $\{\omega_{i_1}, \omega_{i_2}, \dots, \omega_{i_k}\}$ onde $\{i_1, i_2, \dots, i_k\}$ é um sub-conjunto qualquer de $\{1, 2, \dots, n\}$; $\dots$ e $\binom{n}{n} = 1$ conjunto constituído por todos os acontecimentos elementares, isto é, o universo Ω . Assim, o conjunto potência é constituído por

$$\sum_{i=0}^n \binom{n}{i} = 2^n \quad (10.7)$$

conjuntos distintos.

Exemplo 10.1. Num lançamento de uma moeda ao ar, representando por $\mathbf{C}$ o acontecimento saiu cara virada para cima e $\overline{\mathbf{C}}$ o acontecimento saiu coroa virada para cima (supondo que sai sempre uma das faces), teremos

$$\Omega = \{\mathbf{C}, \overline{\mathbf{C}}\} \Rightarrow 2^\Omega = \{\Omega, \emptyset, \mathbf{C}, \overline{\mathbf{C}}\}. \quad (10.8)$$

Se considerarmos uma experiência aleatória com quatro resultados possíveis, isto é, com quatro acontecimentos elementares representados por $\mathbf{A}_1, \mathbf{A}_2, \mathbf{A}_3, \mathbf{A}_4$ teremos

$$\begin{aligned} \Omega &= \{\mathbf{A}_1, \mathbf{A}_2, \mathbf{A}_3, \mathbf{A}_4\} \Rightarrow \\ \Rightarrow 2^\Omega &= \{\emptyset, \mathbf{A}_1, \mathbf{A}_2, \mathbf{A}_3, \mathbf{A}_4, \{\mathbf{A}_1, \mathbf{A}_2\}, \{\mathbf{A}_1, \mathbf{A}_3\}, \\ &\quad \{\mathbf{A}_1, \mathbf{A}_4\}, \{\mathbf{A}_2, \mathbf{A}_3\}, \{\mathbf{A}_2, \mathbf{A}_4\}, \\ &\quad \{\mathbf{A}_3, \mathbf{A}_4\}, \{\mathbf{A}_1, \mathbf{A}_2, \mathbf{A}_3\}, \{\mathbf{A}_1, \mathbf{A}_2, \mathbf{A}_4\}, \\ &\quad \{\mathbf{A}_1, \mathbf{A}_3, \mathbf{A}_4\}, \{\mathbf{A}_2, \mathbf{A}_3, \mathbf{A}_4\}, \Omega\}. \end{aligned}$$



Na análise a uma experiência aleatória não é obrigatório utilizarmos o conjunto potência de Ω como espaço dos acontecimentos. Por exemplo, se estivermos interessados em analisar se num lançamento de um dado sai um número par ou ímpar, então, para esta experiência aleatória específica, será suficiente considerarmos como espaço de acontecimentos $\mathcal{A} = \{\emptyset, \{1, 3, 5\}, \{2, 4, 6\}, \Omega\}$, pois contém todos os acontecimentos a que pretendemos atribuir uma probabilidade.

Nas experiências aleatórias em que o espaço-amostra é finito, bastará conhecermos as probabilidades dos acontecimentos elementares ω_i para determinarmos a probabilidade de qualquer acontecimento do espaço de acontecimentos (mesmo considerando o conjunto potência de Ω). Esta circunstância resulta do facto de qualquer acontecimento $\mathbf{A}$, com $\mathbf{A} \subset \Omega$, poder ser decomposto (de uma única forma) na união finita de acontecimentos elementares. Consequentemente, pelo axioma da aditividade, a probabilidade do acontecimento $\mathbf{A}$ pode ser obtida pela soma das probabilidades dos acontecimentos elementares que o compõem (pois os acontecimentos elementares são necessariamente disjuntos e o axioma da aditividade é facilmente generalizado, por recorrência, para a união de qualquer número finito de conjuntos). Porém, não são as experiências com um número finito de resultados que originaram a necessidade de recorrer aos conceitos da Teoria da Medida, mas antes as que possuem um número infinito de resultados que estão na origem do diversos paradoxos que assombravam a Teoria da Probabilidade.

Refira-se, todavia, que a definição clássica de probabilidades, por diversas vezes comentada ao longo deste trabalho, pode ser obtida considerando um caso particular da definição axiomática de Kolmogoroff, bastando para tal considerarmos um universo finito (único caso analisado até ao momento) e supormos acontecimentos elementares equiprováveis, isto é, que a função $\mathbb{P}$ associa o mesmo número real a cada acontecimento elementar ω_i de Ω . As-

sim, consideremos um espaço de probabilidade $(\Omega, \mathcal{A}, \mathbb{P})$ e, dado que uma das hipóteses subjacentes à definição clássica é a do número de resultados ser finito, seja n o número de acontecimentos elementares distintos — $\Omega = \{\omega_1, \dots, \omega_n\}$. A segunda hipótese inerente à definição clássica corresponde à equiprobabilidade dos acontecimentos elementares, que terá como consequência $\mathbb{P}(\omega_i) = p, \forall i$. Pela probabilidade do universo ($\mathbb{P}(\Omega) = 1$) e pela aditividade da medida de probabilidade, obtemos

$$\mathbb{P}(\Omega) = 1 \Leftrightarrow \sum_{i=1}^n \mathbb{P}(\omega_i) = 1 \Leftrightarrow p = \frac{1}{n}. \quad (10.9)$$

Assim, a probabilidade de qualquer acontecimento elementar é igual a $\frac{1}{n}$. Como qualquer acontecimento $\mathbf{A}$, associado a esta experiência, pode ser obtido pela união de acontecimentos elementares de uma única forma, isto é, como existe um (e um só) subconjunto $\{k_1, \dots, k_{n_{\mathbf{A}}}\}$, com $n_{\mathbf{A}} = \#\mathbf{A} < n$, de $\{1, \dots, n\}$ tal que $\mathbf{A} = \cup_{j=1}^{n_{\mathbf{A}}} \omega_{k_j}$, temos

$$\mathbb{P}(\mathbf{A}) = \sum_{j=1}^{n_{\mathbf{A}}} \mathbb{P}(\omega_{k_j}) = \sum_{j=1}^{n_{\mathbf{A}}} \frac{1}{n} = \frac{n_{\mathbf{A}}}{n}, \quad (10.10)$$

que naturalmente corresponde ao quociente entre o número de casos favoráveis (de acontecimentos elementares que compõem o acontecimento $\mathbf{A}$) e o número de casos possíveis (de acontecimentos elementares que compõem o universo Ω).

Investiguemos agora as experiências que assumem um número infinito de resultados.

Caso infinito

Os cinco axiomas apresentados são suficientes para a análise de qualquer experiência aleatória cujo espaço de resultados seja finito. No entanto, nas experiências aleatórias cujos espaços de resultados Ω são constituídos por

um número infinito de resultados, devemos acrescentar, aos axiomas anteriormente apresentados, o axioma da continuidade da medida de probabilidade, que corresponde ao sexto axioma. Com o intuito de facilitar a apresentação, vamos representar por $\mathbf{A}_n \nearrow \mathbf{A}$ a sucessão de conjuntos $\{\mathbf{A}_n\}_{n \in \mathbb{N}}$ crescentes (isto é $\mathbf{A}_1 \subset \mathbf{A}_2 \subset \cdots \subset \mathbf{A}_n \subset \cdots$) onde $\lim_{n \rightarrow +\infty} \mathbf{A}_n = \mathbf{A}$, ou seja, $\mathbf{A} = \bigcup_{i=1}^{+\infty} \mathbf{A}_i$ e, de forma análoga, vamos representar por $\mathbf{A}_n \searrow \mathbf{A}$ a sucessão de conjuntos $\{\mathbf{A}_n\}_{n \in \mathbb{N}}$ decrescentes (isto é $\mathbf{A}_1 \supset \mathbf{A}_2 \supset \cdots \supset \mathbf{A}_n \supset \cdots$) onde $\lim_{n \rightarrow +\infty} \mathbf{A}_n = \mathbf{A}$, ou seja, $\mathbf{A} = \bigcap_{i=1}^{+\infty} \mathbf{A}_i$. Deste modo o axioma da continuidade da medida de probabilidade pode ser apresentado da seguinte forma:

Avi) Seja $\{\mathbf{A}_n\}_{n \geq 1}$ uma sucessão de acontecimentos de $\mathcal{A}$ tal que $\mathbf{A}_n \searrow \emptyset$ então

$$\lim_{n \rightarrow \infty} \mathbb{P}(\mathbf{A}_n) = 0. \quad (10.11)$$

A σ -aditividade, que actualmente corresponde ao terceiro axioma quando apresentamos a axiomática de Kolmogoroff para o caso infinito, é apresentada por Kolmogoroff como teorema — Teorema da Aditividade Generalizada, sendo a sua demonstração baseada no axioma da continuidade.

Teorema 10.1 (Teorema da Aditividade Generalizada ou σ -aditividade). *Seja $\{\mathbf{A}_n\}_{n \geq 1}$ uma sucessão de acontecimentos pertencentes a $\mathcal{A}$, disjuntos dois a dois ($\mathbf{A}_i \cap \mathbf{A}_j = \emptyset, \forall i \neq j$), então*

$$\mathbb{P}\left(\bigcup_{i=1}^{+\infty} \mathbf{A}_i\right) = \sum_{i=1}^{+\infty} \mathbb{P}(\mathbf{A}_i). \quad (10.12)$$

Demonstração. Considerando as hipótese do Teorema da Aditividade Generalizada e definindo $\mathbf{R}_n = \bigcup_{i > n} \mathbf{A}_i$ e $\mathbf{A} = \bigcup_{i=1}^{+\infty} \mathbf{A}_i$, obtém-se pela aditividade finita (axioma **Av**)

$$\mathbb{P}(\mathbf{A}) = \mathbb{P}\left(\bigcup_{i=1}^{\infty} \mathbf{A}_i\right) = \mathbb{P}\left(\bigcup_{i=1}^n \mathbf{A}_i\right) + \mathbb{P}(\mathbf{R}_n) = \sum_{i=1}^n \mathbb{P}(\mathbf{A}_i) + \mathbb{P}(\mathbf{R}_n). \quad (10.13)$$

Por outro lado, a sucessão $\mathbf{R}_n$ é uma sucessão de acontecimentos de $\mathcal{A}$ que são decrescentes e verificam $\cap_{n=1}^{+\infty} \mathbf{R}_n = \emptyset$, isto é, $\mathbf{R}_n \searrow \emptyset$. Aplicando o axioma da continuidade da medida de probabilidade vem

$$\lim_{n \rightarrow +\infty} \mathbb{P}(\mathbf{R}_n) = 0, \quad (10.14)$$

obtendo-se o resultado pretendido. $\square$

Uma diferença se destaca entre a apresentação original de Kolmogoroff e a que actualmente é usual. Enquanto Kolmogoroff apresenta a σ -aditividade como teorema, que é uma consequência do seu axioma da continuidade da medida de probabilidade (juntamente com os restantes axiomas), actualmente esta propriedade corresponde ao terceiro axioma (quando temos espaços infinitos). Contudo, a σ -aditividade (com os primeiros dois axiomas da apresentação actual que correspondem a $0 \leq \mathbb{P}(\mathbf{A}) \leq \mathbb{P}(\Omega) = 1, \forall \mathbf{A} \in \mathcal{A}$) também implica a continuidade da medida de probabilidade, sendo, por este motivo, indiferente, para a construção da teoria, qual destas propriedades é apresentada como axioma. Notemos que, se considerarmos que a medida de probabilidade é σ -aditiva e que $\{\mathbf{A}_n\}_{n \geq 1}$ é uma sucessão decrescente de conjuntos de $\mathcal{A}$ com $\cap_{i=1}^{\infty} \mathbf{A}_i = \emptyset$ ($\lim_{n \rightarrow \infty} \mathbf{A}_i = \emptyset$) então, definindo $\mathbf{B}_n = \mathbf{A}_n \setminus \mathbf{A}_{n+1}, \forall n \in \mathbb{N}$ temos $\mathbf{B}_n \in \mathcal{A}, \mathbf{B}_i \cap \mathbf{B}_j = \emptyset \forall i \neq j$ e

$$\mathbf{A}_m = \bigcup_{i=m}^{\infty} \mathbf{B}_i \Rightarrow \mathbb{P}(\mathbf{A}_m) = \mathbb{P}\left(\bigcup_{i=m}^{\infty} \mathbf{B}_i\right) = \sum_{i=m}^{\infty} \mathbb{P}(\mathbf{B}_i) \leq 1, \forall m \in \mathbb{N}. \quad (10.15)$$

Uma vez que a série é convergente, a partir de determinado valor de m a série tem que se aproximar de zero, logo $\mathbb{P}$ é contínua em $\emptyset$, isto é, $\lim_{n \rightarrow \infty} \mathbb{P}(\mathbf{A}_n) = 0$.

Assim, no caso de o espaço de resultados não ser finito, Kolmogoroff considera que só podemos definir a medida de probabilidade para as situações em que o espaço dos acontecimentos $\mathcal{A}$ for um corpo de Borel que contenha Ω , provando que podemos estender a medida para a σ -álgebra associada

pelo Teorema de Extensão de Carathéodory, que nos permite passar da medida definida numa álgebra para uma σ -álgebra. Um corpo de conjuntos $\mathcal{A}$ é um corpo de Borel se for fechado para a união numerável, isto é, se $\mathbf{A}_1, \dots, \mathbf{A}_n, \dots \in \mathcal{A} \Rightarrow \cup_{i=1}^{\infty} \mathbf{A}_i \in \mathcal{A}$. Se a um corpo de Borel $\mathcal{A}$, constituído por subconjuntos de Ω , acrescentarmos a condição $\Omega \in \mathcal{A}$, obtemos o que actualmente denominamos por σ -álgebra sobre Ω ⁽⁹⁾. Uma outra forma de definirmos σ -álgebra é considerarmos que $\mathcal{F}$ é uma σ -álgebra sobre $\mathbf{X}$ (com $\mathbf{X}$ não vazio) se se verificarem as seguintes três condições:

$$i) \mathbf{X} \in \mathcal{F};$$

ii) se $\mathbf{A}$ pertence a $\mathcal{F}$ então o seu complementar também pertence, isto é,

$$\mathbf{A} \in \mathcal{F} \Rightarrow \overline{\mathbf{A}} \in \mathcal{F};$$

iii) a união numerável de elementos de $\mathcal{F}$ também pertence a $\mathcal{F}$, isto é,

$$\mathbf{X}_1, \mathbf{X}_2, \dots \in \mathcal{F} \Rightarrow \cup_{i=1}^{\infty} \mathbf{X}_i \in \mathcal{F}.$$

Desta forma uma σ -álgebra sobre um conjunto $\mathbf{X}$ é uma colecção $\mathcal{F}$ de subconjuntos de $\mathbf{X}$ que contém $\mathbf{X}$ e é fechada para um número contável de operações de conjuntos, isto é, qualquer conjunto obtido através de um número contável de intersecções, uniões e passagens ao complementar de conjuntos de $\mathcal{F}$ também pertence a $\mathcal{F}$. Uma σ -álgebra que assume grande importância é a menor σ -álgebra que contenha uma classe $\mathcal{C}$ de conjuntos de Ω , denominada por σ -álgebra gerada por $\mathcal{C}$. Assim, considerando $\mathcal{C}$ uma classe de conjuntos de Ω , a **σ -álgebra gerada** por $\mathcal{C}$, representada por $\sigma(\mathcal{C})$, corresponde à intersecção de todas as σ -álgebras que contêm $\mathcal{C}$ (que também é uma σ -álgebra). Esta σ -álgebra é igualmente denominada por σ -álgebra

⁽⁹⁾ Se $\mathcal{F}$ é uma álgebra de conjuntos sobre $\mathbf{X}$ então $\mathcal{F}$ é uma **σ -álgebra** sobre $\mathbf{X}$ se e só se $\mathbf{A}_1, \mathbf{A}_2, \dots, \mathbf{A}_n, \dots \in \mathcal{F} \Rightarrow \cup_{i=1}^{\infty} \mathbf{A}_i \in \mathcal{F}$.

minimal pelo facto de corresponder à menor σ -álgebra que contém $\mathcal{C}$ ⁽¹⁰⁾. Assim, de entre todas as σ -álgebras de acontecimentos sobre $\mathcal{C}$, optamos por trabalhar com aquela que contém os acontecimentos de interesse para a construção do nosso modelo, evitando os subconjuntos de Ω desnecessários para o nosso fim (que podem originar paradoxos, conforme verificaremos mais adiante). Apresentemos agora o Teorema de Extensão de Carathéodory, um resultado fundamental na construção de Kolmogoroff, que nos permite estender a medida definida numa álgebra para uma σ -álgebra, tendo sido demonstrado (na sua versão mais geral utilizada por Kolmogoroff) em 1930 por Carathéodory.

Teorema 10.2 (Teorema de Extensão de Carathéodory). *Seja $\mathbf{X}$ um conjunto, $\mathcal{A}$ uma álgebra sobre $\mathbf{X}$ e $\mu : \mathbf{X} \rightarrow \mathbb{R}$ uma função não negativa, σ -aditiva, então μ pode ser extendida para uma medida definida na σ -álgebra gerada por $\mathcal{A}$ (seja $\sigma(\mathcal{A})$). Se μ é uma medida σ -finita⁽¹¹⁾ então esta extensão é única.*

Desta forma, se μ' é a extensão de μ obtida pelo Teorema 10.2 da álgebra $\mathcal{A}$ para a σ -álgebra gerada por $\mathcal{A}$, representada por $\sigma(\mathcal{A})$, teremos que μ' é uma medida em $\sigma(\mathcal{A})$ onde se verificará $\mu'(\mathbf{A}) = \mu(\mathbf{A}), \forall \mathbf{A} \in \mathcal{A}$. O Teorema de Extensão de Carathéodory permite-nos então probabilizar (atribuir um valor para a medida de probabilidade) alguns conjuntos que podem ser determinados pela medida dos que pertencem à álgebra $\mathcal{A}$. Por exemplo, se

⁽¹⁰⁾ Esta σ -álgebra é única, no entanto, não existe nenhum método construtivo para descrever a σ -álgebra gerada por $\mathcal{C}$.

⁽¹¹⁾ Uma medida é **σ -finita** se existirem conjuntos $\mathbf{A}_i$ tais que $\Omega = \cup_{i=1}^{\infty} \mathbf{A}_i$ que verifiquem $\mu(\mathbf{A}_i) < \infty$. Uma medida μ será **finita** se se verificar ainda $\mu(\Omega) < \infty$. Como $\mathbb{P}(\Omega) = 1$ (e consequentemente $\mathbb{P}(\mathbf{A}) \leq 1, \forall \mathbf{A} \in \Omega$), a medida de probabilidade é finita e consequentemente σ -finita (facilmente se comprova que uma medida de probabilidade é finita e que qualquer medida finita é σ -finita).

$\mathbf{A} \subset \mathbf{B} \subset \mathbf{C}$, com $\mathbf{A}$ e $\mathbf{C}$ pertencentes a $\mathcal{A}$ e $\mathbf{B}$ não, e $\mathbb{P}(\mathbf{A}) = \mathbb{P}(\mathbf{C})$, então $\mathbb{P}(\mathbf{B})$ também deveria ter probabilidade igual. Assim, este teorema permite passar de uma medida definida numa álgebra para uma σ -álgebra, o que, por exemplo, vai permitir probabilizar os resultados limites, isto é, atribuir a probabilidade aos acontecimentos $\mathbf{A}$ que podem ser obtidos através de um limite $\mathbf{A}_i \rightarrow \mathbf{A}$.

Contudo, que σ -álgebra devemos utilizar? Poderemos utilizar o conjunto potência (que é uma σ -álgebra) para modelar? No caso de o espaço dos resultados ser infinito numerável também se pode recorrer ao espaço potência. Contudo, quando Ω não é contável⁽¹²⁾, pode não ser possível atribuir probabilidade a todos os conjuntos de $\mathcal{P}(\Omega)$ de forma compatível com os axiomas, pois, neste caso, não existe uma função $\mathbb{P}$ (distinta da função nula que não verifica $\mathbb{P}(\Omega) = 1$), satisfazendo a σ -aditividade. Desta forma podemos concluir que nem todos os conjuntos são mensuráveis, pois, se considerarmos como domínio da nossa medida o conjunto potência, então irá surgir uma série de patologias nessa medida. Por exemplo, se considerarmos a classe $\mathcal{P}(\mathbb{R}^n)$, não existe nenhuma função que verifique a σ -aditividade para $n = 1$ e $n = 2$, e, no caso de $n = 3$, nem sequer existe uma função que verifique a aditividade finita, como ilustra o paradoxo de Tarski-Banach⁽¹³⁾. No paradoxo (teorema) de Tarski-Banach podemos dividir uma esfera em um número finito de peças e depois reagrupar de forma a formarem duas esferas cada uma do mesmo tamanho da original. Não há uma prova construtivista deste teorema, isto é, que descreva a maneira como a esfera deve ser repartida. A

⁽¹²⁾ Um conjunto é contável se for um conjunto finito ou um conjunto infinito numerável.

⁽¹³⁾ Alfred Tarski (1902–1983) e Stefan Banach (1892–1945) são dois matemáticos polacos. Banach é famoso pelos seus trabalhos que estão na origem da Análise Funcional.

demonstração faz uso do axioma da escolha⁽¹⁴⁾, um axioma controverso da Teoria dos Conjuntos, formulado em 1904 por Ernst Zermelo (1871–1953). Este teorema evidencia que nem todos os conjuntos são mensuráveis, isto é, se considerarmos uma medida com as características anteriormente referidas, nomeadamente a aditividade (ou σ -aditividade), é impossível atribuir uma medida de forma coerente (definir uma função que obedeça às características de uma medida) se considerarmos que todos os conjuntos têm atribuída uma medida. Por esta razão o espaço potência $\mathcal{P}(\Omega)$ não pode ser usado quando Ω é infinito não numerável.

Em suma, Kolmogoroff define Probabilidade como uma medida $\mathbb{P}$ tal que $\mathbb{P}(\Omega) = 1$. Uma medida, para estar bem definida, evitando qualquer patologia, é definida num determinado domínio que é denominado por espaço mensurável. Assim, um **espaço mensurável** é um par $(\mathbf{X}, \mathcal{F})$, onde $\mathbf{X}$ é um conjunto não vazio e $\mathcal{F}$ é uma σ -álgebra sobre $\mathbf{X}$. Dizemos que $\mathbf{A}$, com $\mathbf{A} \subset \mathbf{X}$, é um conjunto **$\mathcal{F}$ -mensurável** se e só se $\mathbf{A} \in \mathcal{F}$. Deste modo, teremos $\mathbf{X} = \Omega$ o espaço de resultados (ou espaço-amostra constituído pelos possíveis resultados — acontecimentos elementares) associado à experiência aleatória $\mathcal{E}$, e $\mathcal{F}$ será uma σ -álgebra sobre Ω , designada σ -álgebra dos acontecimentos (cada elemento deste conjunto é denominado por acontecimento), sendo o nosso espaço mensurável $(\Omega, \mathcal{F})$. Após definirmos o domínio das funções de conjuntos em estudo, analisamos as características destas funções onde se destaca a σ -aditividade. Recordemos que uma função $\mu : \mathcal{A} \rightarrow [0, +\infty]$ é uma medida real, definida no espaço mensurável $(\mathbf{X}, \mathcal{F})$ se e só se μ for uma função σ -aditiva, isto é,

$$\forall \mathbf{A}_n \in \mathcal{F} : \mathbf{A}_i \cap \mathbf{A}_j = \emptyset, \forall i \neq j \Rightarrow \mu \left(\bigcup_{i=1}^{\infty} \mathbf{A}_i \right) = \sum_{i=1}^{\infty} \mu(\mathbf{A}_i). \quad (10.16)$$

⁽¹⁴⁾ Axioma da escolha: Seja $\mathcal{C}$ uma classe de conjuntos não vazios e $T = \cup_{\mathbf{C} \in \mathcal{C}} \mathbf{C}$. Então existe uma função $f : \mathcal{C} \rightarrow T$ tal que $f(\mathbf{C}) \in \mathbf{C}$ para qualquer $\mathbf{C} \in \mathcal{C}$.

Uma medida real μ que verifique $\mu(\emptyset) = 0$ ⁽¹⁵⁾ é denominada **medida de Radon**. Assim, de forma sucinta, a Teoria da Medida pretende analisar as características das funções σ -aditivas definidas em σ -álgebras, que denominamos por medidas. A medida de probabilidade $\mathbb{P}$ definida no espaço mensurável $(\Omega, \mathcal{F})$ é uma medida de Radon tal que $\mathbb{P}(\Omega) = 1$. Agrupando o espaço mensurável e a medida obtemos o espaço de medida, que é um terno $(\mathbf{X}, \mathcal{F}, \mu)$, onde $(\mathbf{X}, \mathcal{F})$ é um espaço mensurável e μ é uma medida definida em $\mathcal{F}$. O espaço de medida de probabilidade corresponde ao terno $(\Omega, \mathcal{A}, \mathbb{P})$ onde $\mathbb{P}(\Omega) = 1$. Assim, para trabalhar em espaços com a medida bem definida, Kolmogoroff restringiu-se aos espaços de probabilidade cujo espaço dos acontecimentos é um espaço de Borel (σ -álgebra gerada por Ω), recorrendo, para tal, ao Teorema da Extensão de Carathéodory, que lhe permitiu estender a função de medida definida numa álgebra para a σ -álgebra que lhe está associada.

A axiomática de Kolmogoroff e o consequente tratamento da probabilidade como uma medida não trouxeram nada de revolucionário ao Cálculo das Probabilidades no caso finito, uma vez que não era nestas condições que surgiam os paradoxos. Por outro lado, quando Ω é infinito (numerável ou não), o facto de a axiomática de Kolmogoroff claramente definir os conjuntos mensuráveis, aos quais é possível atribuir uma medida de probabilidade, e o recurso à propriedade σ -aditiva da função $\mathbb{P}$, em muito contribuíram para resolver os problemas que assombravam a Teoria da Probabilidade no início do século XX. Refira-se que esta construção de Kolmogoroff, com a definição do espaço amostra e da função $\mathbb{P}$, permite clarificar alguns paradoxos, tais como o paradoxo de Bertrand (analisado na secção 5.1.2), cujas três soluções propostas estão associadas a três espaços de probabilidade $(\Omega, \mathcal{A}, \mathbb{P})$ distin-

⁽¹⁵⁾ Por vezes esta condição é substituída por μ não ser identicamente ∞ , que é condição equivalente.

tos. Desta forma a contradição, neste problema, surge pela inexactidão do enunciado do problema, que é suficientemente lato para permitir pelo menos três medidas de probabilidade distintas. Este problema, como salientamos na sua análise, não foi entendido por Pacheco d'Amorim.

Consequências da axiomática

Nesta secção vamos apresentar as principais consequências desta definição, isto é, algumas propriedades da medida de probabilidade que são consequência dos axiomas apresentados e que são fundamentais para o Cálculo das Probabilidades. Começemos por algumas propriedades simples que servem para manipular probabilidades em relação a dois acontecimentos de um espaço de probabilidade. Seja $(\Omega, \mathcal{F}, \mathbb{P})$ um espaço de probabilidade, então para quaisquer acontecimentos $\mathbf{A}, \mathbf{B} \in \mathcal{F}$ teremos

$$\mathbb{P}(\overline{\mathbf{A}}) = 1 - \mathbb{P}(\mathbf{A}), \quad (10.17)$$

onde $\overline{\mathbf{A}}$ representa o acontecimento complementar a $\mathbf{A}$ ($\overline{\mathbf{A}} = \Omega - \mathbf{A}$), pois pela aditividade verifica-se $1 = \mathbb{P}(\Omega) = \mathbb{P}(\mathbf{A} \cup \overline{\mathbf{A}}) = \mathbb{P}(\mathbf{A}) + \mathbb{P}(\overline{\mathbf{A}})$. Consequentemente verifica-se $\mathbb{P}(\emptyset) = 0$, uma vez que $\Omega = \overline{\emptyset}$. De igual forma podemos concluir que

$$\mathbb{P}(\mathbf{A}) = \mathbb{P}(\mathbf{A} - \mathbf{B}) + \mathbb{P}(\mathbf{A} \cap \mathbf{B}), \quad (10.18)$$

que tem como consequência a monotonia, isto é, se $\mathbf{A} \subset \mathbf{B}$, teremos

$$\mathbb{P}(\mathbf{A}) \leq \mathbb{P}(\mathbf{B}), \quad (10.19)$$

pois $\mathbb{P}(\mathbf{B}) = \mathbb{P}(\mathbf{A}) + \mathbb{P}(\mathbf{B} - \mathbf{A}) \geq \mathbb{P}(\mathbf{A})$. Da monotonia conclui-se que, para qualquer acontecimento $\mathbf{A} \in \mathcal{F}$, teremos

$$0 = \mathbb{P}(\emptyset) \leq \mathbb{P}(\mathbf{A}) \leq \mathbb{P}(\Omega) = 1. \quad (10.20)$$

Podemos ainda deduzir que

$$\mathbb{P}(\mathbf{A} \cup \mathbf{B}) = \mathbb{P}(\mathbf{A}) + \mathbb{P}(\mathbf{B}) - \mathbb{P}(\mathbf{A} \cap \mathbf{B}), \quad (10.21)$$

pois $\mathbb{P}(\mathbf{A} \cup \mathbf{B}) = \mathbb{P}(\mathbf{A}) + \mathbb{P}(\mathbf{B} - \mathbf{A}) = \mathbb{P}(\mathbf{A}) + \mathbb{P}(\mathbf{B}) - \mathbb{P}(\mathbf{A} \cap \mathbf{B})$. Este resultado pode ser generalizado para mais acontecimentos. Considerando que $\mathbf{A}_1, \dots, \mathbf{A}_n$ pertencem a $\mathcal{F}$ teremos

$$\begin{aligned} \mathbb{P}\left(\bigcup_{k=1}^n \mathbf{A}_k\right) &= \sum_{k=1}^n \mathbb{P}(\mathbf{A}_k) - \sum_{i_1 < i_2} \mathbb{P}(\mathbf{A}_{i_1} \cap \mathbf{A}_{i_2}) + \\ &+ \sum_{i_1 < i_2 < i_3} \mathbb{P}(\mathbf{A}_{i_1} \cap \mathbf{A}_{i_2}) - \dots + (-1)^{n+1} \mathbb{P}\left(\bigcap_{k=1}^n \mathbf{A}_k\right), \end{aligned}$$

habitualmente denominado por Regra da inclusão-exclusão e que pode ser demonstrada por indução⁽¹⁶⁾. Todos estes resultados foram, ou podem ser, igualmente demonstrados pela construção de Pacheco d'Amorim. Outros resultados fundamentais que utilizam um número maior de acontecimentos, para os quais Pacheco d'Amorim não apresentou qualquer demonstração e, em muitos casos, a sua construção não seria suficiente, serão os seguintes teoremas referentes a um número infinito de acontecimentos.

Teorema 10.3 (Teorema de Cobertura). *Seja $(\Omega, \mathcal{F}, \mathbb{P})$ um espaço de probabilidade e sejam $\mathbf{A}_1, \dots, \mathbf{A}_n, \dots$ quaisquer acontecimentos pertencentes a $\mathcal{F}$. Se definirmos um acontecimento $\mathbf{A} \subset \bigcup_{k=1}^{+\infty} \mathbf{A}_k$ então teremos*

$$\mathbb{P}(\mathbf{A}) \leq \sum_{k=1}^{\infty} \mathbb{P}(\mathbf{A}_k) \quad (10.22)$$

Demonstração. Podemos decompor o acontecimento $\mathbf{A}$ numa união numerá-

⁽¹⁶⁾ Outra forma de demonstrar esta propriedade é recorrer ao valor esperado das funções indicatrizes dos acontecimentos $\mathbf{A}_i$, pela identidade que existe entre o valor esperado da função indicatriz do acontecimento $\mathbf{A}$ e a sua probabilidade (consultar, por exemplo, Pestana e Velosa (2002)).

vel de conjuntos disjuntos através de

$$\begin{aligned} \mathbf{A} = & (\mathbf{A} \cap \mathbf{A}_1) \cup (\mathbf{A} \cap (\mathbf{A}_2 - \mathbf{A}_1)) \cup (\mathbf{A} \cap (\mathbf{A}_3 - (\mathbf{A}_2 \cup \mathbf{A}_1))) \cup \cdots \cup \\ & \cup (\mathbf{A} \cap (\mathbf{A}_k - (\cup_{i=1}^{k-1} \mathbf{A}_i))) \cup \cdots, \end{aligned} \quad (10.23)$$

que, pela σ -aditividade, teremos

$$\begin{aligned} \mathbb{P}(\mathbf{A}) &= \mathbb{P}(\mathbf{A} \cap \mathbf{A}_1) + \mathbb{P}(\mathbf{A} \cap (\mathbf{A}_2 - \mathbf{A}_1)) + \cdots \leq \\ &\leq \mathbb{P}(\mathbf{A}_1) + \mathbb{P}(\mathbf{A}_2) + \cdots = \sum_{k=1}^{\infty} \mathbb{P}(\mathbf{A}_k). \end{aligned} \quad (10.24)$$

□

Notemos que, se definíssemos o acontecimento $\mathbf{A}$ através de $\mathbf{A} = \cup_{k=1}^{+\infty} \mathbf{A}_k$, teríamos um caso particular do teorema anterior e obteríamos a subaditividade

$$\mathbb{P}\left(\bigcup_{k=1}^{\infty} \mathbf{A}_k\right) \leq \sum_{k=1}^{\infty} \mathbb{P}(\mathbf{A}_k). \quad (10.25)$$

Este resultado também é válido para um número finito de acontecimentos, obtendo-se

$$\mathbb{P}\left(\bigcup_{k=1}^n \mathbf{A}_k\right) \leq \sum_{k=1}^n \mathbb{P}(\mathbf{A}_k), \quad (10.26)$$

bastando, para tal, considerar $\mathbf{A}_i = \emptyset$, para $i > n$ na expressão (10.25).

Teorema 10.4 (Teorema da convergência monótona de Lebesgue). *Seja $(\mathbf{X}, \mathcal{F}, \mu)$ um qualquer espaço de medida. Seja $\{\mathbf{A}_n\}_{n \geq 1}$ uma sucessão de conjuntos de $\mathcal{F}$ tal que $\mathbf{A}_n \nearrow \mathbf{A}$, então $\mathbf{A} \in \mathcal{F}$ e $\mu(\mathbf{A}_n) \rightarrow \mu(\mathbf{A})$.*

Demonstração. Se definirmos a sucessão $\mathbf{B}_n = \mathbf{A}_n - \mathbf{A}_{n-1}$ para $n \geq 2$ e $\mathbf{B}_1 = \mathbf{A}_1$ (logo os conjuntos $\mathbf{B}_n$ são μ -mensuráveis, isto é, $\mathbf{B}_i \in \mathcal{F}$), verificamos que

$$\mathbf{B}_i \cap \mathbf{B}_j = \emptyset, \quad \forall i \neq j, \quad \mathbf{A}_n = \bigcup_{i=1}^n \mathbf{B}_i, \quad \mathbf{A} = \bigcup_{i=1}^{\infty} \mathbf{A}_i = \bigcup_{i=1}^{\infty} \mathbf{B}_i. \quad (10.27)$$

Como μ é uma medida e os conjuntos $\mathbf{B}_i$ são disjuntos dois a dois, temos

$$\mu(\mathbf{A}_n) = \mu\left(\bigcup_{i=1}^n \mathbf{B}_i\right) = \sum_{i=1}^n \mu(\mathbf{B}_i), \quad (10.28)$$

e

$$\mu(\mathbf{A}) = \mu\left(\bigcup_{i=1}^{\infty} \mathbf{B}_i\right) = \sum_{i=1}^{\infty} \mu(\mathbf{B}_i), \quad (10.29)$$

logo $\mu(\mathbf{A}_n) \rightarrow \mu(\mathbf{A})$. Uma vez que $\mathbf{A} = \bigcup_{i=1}^{\infty} \mathbf{B}_i$ e $\mathbf{B}_i \in \mathcal{F}$, então $\mathbf{A} \in \mathcal{F}$. $\square$

Teorema 10.5. *Seja $(\mathbf{X}, \mathcal{F}, \mu)$ um qualquer espaço de medida. Se os conjuntos $\mathbf{A}_n \in \mathcal{F}$ e $\mathbf{A}_n \searrow \mathbf{A}$ então $\mathbf{A} \in \mathcal{F}$. Caso se verifique ainda que $\mu(\mathbf{A}_1) \neq +\infty$ (que qualquer medida de probabilidade verifica) então $\mu(\mathbf{A}_n) \rightarrow \mu(\mathbf{A})$.*

Demonstração. Se definirmos a sucessão $\mathbf{B}_n = \mathbf{A}_1 - \mathbf{A}_n$ verificamos que os conjuntos $\mathbf{B}_n$ são μ -mensuráveis e $\mathbf{B}_1 \subset \mathbf{B}_2 \subset \cdots \subset \mathbf{B}_n \subset \cdots$ (formam uma sucessão crescente), logo $\mathbf{B}_n \rightarrow \bigcup_{i=1}^{+\infty} \mathbf{B}_i$, ou seja, teremos (pelo Teorema 10.4)

$$\mu(\mathbf{A}_1 - \mathbf{A}_n) = \mu(\mathbf{B}_n) \rightarrow \mu\left(\bigcup_{i=1}^{+\infty} \mathbf{B}_i\right) \quad (10.30)$$

que por sua vez será

$$\mu\left(\bigcup_{i=1}^{+\infty} \mathbf{B}_i\right) = \mu\left(\bigcup_{i=1}^{+\infty} (\mathbf{A}_1 - \mathbf{A}_i)\right) = \mu\left(\mathbf{A}_1 - \bigcap_{i=1}^{+\infty} \mathbf{A}_i\right). \quad (10.31)$$

Como $\mu(\mathbf{A}_1) \neq +\infty$ (e consequentemente $\mu(\mathbf{A}_i) \neq +\infty, \forall i$) temos, uma vez que $\mathbf{A}_n \subset \mathbf{A}_1$,

$$\mu(\mathbf{A}_n) = \mu(\mathbf{A}_1) - \mu(\mathbf{A}_1 - \mathbf{A}_n), \quad (10.32)$$

que tomando os limites teremos

$$\mu\left(\bigcap_{i=1}^{+\infty} \mathbf{A}_i\right) = \mu(\mathbf{A}_1) - \mu\left(\mathbf{A}_1 - \bigcap_{i=1}^{+\infty} \mathbf{A}_i\right), \quad (10.33)$$

e, consequentemente, através de (10.31), concluimos

$$\mu(\mathbf{A}_n) \rightarrow \mu\left(\bigcap_{i=1}^{+\infty} \mathbf{A}_i\right) = \mu(\mathbf{A}). \quad (10.34)$$

$\square$

Estes dois últimos resultados garantem que, considerando um espaço de probabilidade $(\Omega, \mathcal{F}, \mathbb{P})$ (que é um espaço de medida), se uma sequência de acontecimentos $\mathbf{A}_n \in \mathcal{F}$ verificarem $\mathbf{A}_n \rightarrow \mathbf{A}$ de forma monótona (os acontecimentos $\mathbf{A}_n$ são crescentes ou decrescentes), então temos garantido que $\mathbf{A} \in \mathcal{F}$ (logo pertence ao domínio da função $\mathbb{P}$, isto é, o acontecimento $\mathbf{A}$ é probabilizável) e $\mathbb{P}(\mathbf{A}_n) \rightarrow \mathbb{P}(\mathbf{A})$ (a probabilidade do acontecimento $\mathbf{A}$ pode ser definida pelo limite das probabilidades dos acontecimentos $\mathbf{A}_n$). Estes resultados são fundamentais na determinação de resultados de convergência.

Destes resultados podemos concluir que a axiomática de Kolmogoroff nos permite deduzir, de forma bastante mais acessível, as propriedades da probabilidade que Pacheco d'Amorim utilizou na sua obra (apesar de nesta apresentação apenas nos restringirmos a algumas propriedades da probabilidade absoluta) e permite-nos ir mais longe no caso infinito, situação que Pacheco d'Amorim não analisa na sua tese.

10.2.3 Variáveis Aleatórias e Função de Distribuição

Na secção precedente deduzimos que o espaço de probabilidade é definido pelo terno $(\Omega, \mathcal{A}, \mathbb{P})$, onde Ω representa o espaço-amostra ou espaço de resultados (conjunto constituído pelos possíveis resultados da experiência aleatória), $\mathcal{A}$ o espaço dos acontecimentos (σ -álgebra gerada por Ω) e $\mathbb{P}$ a função de probabilidade associada ao par $(\Omega, \mathcal{A})$, satisfazendo os axiomas apresentados. O espaço de probabilidade tem, desta forma, toda a estrutura matemática necessária para ser o alicerce da Teoria da Probabilidade. Contudo, para atingirmos o nosso objectivo de modelar experiências aleatórias, é conveniente transpor a medida de probabilidade para outro espaço mais apelativo de operar (esta ideia foi explorada quando comentamos o capítulo PONTO IMAGEM). Com este propósito existem as variáveis aleatórias que correspondem a funções mensuráveis em Teoria de Medida.

Definição 10.1 (Função mensurável). Sejam $(\mathbf{A}, \sigma(\mathbf{A}), \mathbb{P}_{\mathbf{A}})$ e $(\mathbf{B}, \sigma(\mathbf{B}), \mathbb{P}_{\mathbf{B}})$ dois espaços de medida. Dizemos que a função f , definida em $\mathbf{A}$, é $(\sigma(\mathbf{A}), \sigma(\mathbf{B}))$ -mensurável (ou apenas mensurável caso não haja possibilidade de confusão sobre as σ -álgebras com que estamos a trabalhar) se e só se

$$\forall \mathbf{C} \in \sigma(\mathbf{B}) \Rightarrow f^{-1}(\mathbf{C}) \in \sigma(\mathbf{A}). \quad (10.35)$$

Desta forma, a medida $\mathbb{P}_{\mathbf{B}}$ corresponde à medida $\mathbb{P}_{\mathbf{A}}$, transposta dos conjuntos de $\sigma(\mathbf{A})$ para os conjuntos em $\sigma(\mathbf{B})$, pois a cada conjunto $\mathbf{C} \in \sigma(\mathbf{B})$ está associado um conjunto $f^{-1}(\mathbf{C}) \in \sigma(\mathbf{A})$ que terá a mesma medida, isto é, $\mathbb{P}_{\mathbf{B}}(\mathbf{C}) = \mathbb{P}_{\mathbf{A}}(f^{-1}(\mathbf{C}))$. Isto quer dizer que podemos transpor a medida de probabilidade $\mathbb{P}$ definida para subconjuntos de Ω , que é um qualquer conjunto abstracto, para um conjunto que nos seja mais familiar. Com este objectivo surgem as variáveis aleatórias na Teoria da Probabilidade, que são uma aplicação $X(\cdot)$ do espaço de probabilidade $(\Omega, \mathcal{A}, \mathbb{P})$ no espaço $(\mathbb{R}, \mathcal{B}, \mathbb{P}_X)$, isto é,

$$(\Omega, \mathcal{A}, \mathbb{P}) \xrightarrow{X(\cdot)} (\mathbb{R}, \mathcal{B}, \mathbb{P}_X). \quad (10.36)$$

Nesta aplicação deixamos de trabalhar num espaço de resultados abstracto Ω para transpormos a medida de probabilidade para conjuntos de números reais $\mathbb{R}$. O conjunto $\mathcal{B}(\mathbb{R})$ representa a σ -álgebra de Borel sobre $\mathbb{R}$. Se representarmos por $\mathcal{B}(\mathbb{R}^n)$ a σ -álgebra gerada pelos abertos de $\mathbb{R}^n$, denominamos por conjuntos **Borel-mensuráveis** ou **borelianos** os conjuntos $\mathbf{A} : \mathbf{A} \in \mathcal{B}(\mathbb{R}^n)$. Nas situações $n = 1$ representamos as classes dos conjuntos borelianos apenas por $\mathcal{B}$. Apesar de a σ -álgebra de Borel sobre $\mathbb{R}$ (borelianos) ser, muitas vezes, definida como sendo a menor σ -álgebra que contém todos os conjuntos abertos de $\mathbb{R}$, também pode ser definida como sendo a menor σ -álgebra gerada pelas semi-rectas $] - \infty, x]$ com $x \in \mathbb{R}$. Pode-se demonstrar que esta também pode ser obtida por todos os conjuntos fechados ou por todos os conjuntos da forma $]a, b]$ ou da forma $] - \infty, a]$, entre outros. Em probabilidade, optamos

usualmente pelos conjuntos $] - \infty, x]$ por corresponderem ao caso de maior importância na Teoria da Probabilidade, pois estes conjuntos estão na base da definição da função de distribuição, definição fundamental na Teoria da Probabilidade e Estatística.

Como curiosidade, e com o objectivo de ilustrar a facilidade de conceber borelianos, refira-se que todos os intervalos obtidos pela utilização de complementares, uniões e/ou intersecções finitas ou infinitas numeráveis (as operações permitidas nas sigma-álgebras) de conjuntos da forma $\{a\}$, $[a, b]$, $]a, b[$, $] - \infty, a]$ ou $[a, +\infty[$ são borelianos. Com estas operações podemos obter quase todos os intervalos, sendo este quase que nos complica a Teoria da Medida, pois existem subconjuntos de $\mathbb{R}$ que não são borelianos, como o conjunto ternário de Cantor. Recordemos, conforme referimos na secção anterior, que é impossível definir uma medida para todos os conjuntos de $\mathbb{R}$, isto é, para $2^{\mathbb{R}}$.

Exemplo 10.2 (Conjunto ternário de Cantor). Para construir o conjunto ternário de Cantor começamos com o intervalo fechado $\mathbf{C}_0 = [0, 1]$, ao qual retiramos o seu intervalo central (aberto) com dimensão $\frac{1}{3}$, ou seja, extraímos o intervalo $] \frac{1}{3}, \frac{2}{3} [$ e ficamos com o intervalo $\mathbf{C}_1 = [0, \frac{1}{3}] \cup [\frac{2}{3}, 1]$ que é obtido pela união de dois intervalos fechados de dimensão $\frac{1}{3}$. Numa segunda iteração, a cada um destes dois intervalos vamos retirar o intervalo central com dimensão $(\frac{1}{3})^2$, obtendo-se $\mathbf{C}_2 = [0, \frac{1}{9}] \cup [\frac{2}{9}, \frac{1}{3}] \cup [\frac{2}{3}, \frac{7}{9}] \cup [\frac{8}{9}, 1]$. Na iteração n , o intervalo $\mathbf{C}_n$ é obtido retirando a cada um dos 2^{n-1} intervalos que constituem $\mathbf{C}_{n-1}$ o intervalo central com dimensão $(\frac{1}{3})^n$. Procedendo assim, sucessivamente, constrói-se uma família decrescente $\mathbf{C}_n$ de subconjunto de $[0, 1]$, onde cada $\mathbf{C}_n$ é uma união de 2^n intervalos fechados de dimensão $(\frac{1}{3})^n$, sendo o seu comprimento $(\frac{2}{3})^n$. O conjunto ternário de Cantor, que aqui representamos por $\mathbf{C}$, é o conjunto que resulta deste processo após infinitas iterações, podendo-se definir por $\mathbf{C} = \lim_{n \rightarrow \infty} \mathbf{C}_n = \bigcap_{i=0}^{\infty} \mathbf{C}_i$. Note-se que este conjunto

não é vazio (os pontos extremos dos intervalos que vão sendo produzidos sobrevivem a este processo), é um conjunto fechado cujo interior é vazio (não contém nenhum aberto), tendo medida de Lebesgue nula ($\lim_{n \rightarrow \infty} \left(\frac{2}{3}\right)^n = 0$ que também pode ser obtida analisando a medida dos conjuntos retirados que verifica $\sum_{n=1}^{\infty} 2^{n-1} \left(\frac{1}{3}\right)^n = 1$). Os elementos deste conjunto podem ser determinados por $\sum_{n=1}^{\infty} \frac{t_n}{3^n}$ onde $t_n \in \{0, 2\}$, ou seja, o conjunto é constituído pelos números que podem ser escritos na base ternária utilizando apenas os algarismos 0 e 2. Este conjunto não é um boreliano e, apesar de não ser numerável, pois tem a mesma potência do contínuo, tem medida nula.

Seja f_n a função indicatriz⁽¹⁷⁾ do conjunto $\mathbf{C}_n$ e defina-se a função $\zeta_n(x) = \int_0^x f_n(t) dt$ que é crescente e verifica $\zeta_n(0) = 0$ e $\zeta_n(1) = \left(\frac{2}{3}\right)^n$. Consideremos ainda as funções $g_n(x)$ definidas por

$$g_n(x) = \left(\frac{3}{2}\right)^n \zeta_n(x) = \left(\frac{3}{2}\right)^n \int_0^x f_n(t) dt. \quad (10.38)$$

A função g_n converge uniformemente para uma função G , que é contínua, crescente e verifica $G(0) = 0$ e $G(1) = 1$. Esta função G é denominada Escada do Diabo ou função de Cantor. Notemos que

$$1 = G(1) - G(0) \neq \int_0^1 G'(t) dt = 0, \quad (10.39)$$

o que contraria o Teorema Fundamental de Integração. ◆

⁽¹⁷⁾ Chamamos função indicatriz do conjunto $\mathbf{A}$ à função

$$\mathbb{I}_{\mathbf{A}}(x) = \begin{cases} 1 & \text{se } x \in \mathbf{A} \\ 0 & \text{se } x \notin \mathbf{A} \end{cases}. \quad (10.37)$$

Esta função também é usualmente apelidada por função característica do conjunto $\mathbf{A}$; contudo, como a função característica, em Teoria da Probabilidade, tem uma definição distinta, correspondendo à transformada de Fourier de uma variável aleatória, optamos por utilizar unicamente o nome de função indicatriz de um conjunto para evitar qualquer possível confusão.

Pretendemos, desta forma, salientar duas ideias. Em primeiro lugar que, como o conjunto ternário de Cantor ilustra, é muito mais fácil encontrar borelianos em $\mathbb{R}$ que o contrário, pois é preciso muita imaginação para encontrarmos algum conjunto que não o seja. Por outro lado, é preciso ter o devido cuidado, pois nem todos os conjuntos são mensuráveis, o que pode originar patologias como as que verificamos com a escada do diabo, pelo facto de usarmos, para definir esta função, o conjunto de Cantor que não é mensurável.

Regressando às variáveis aleatórias, foquemos ainda que o espaço de probabilidade $(\mathbb{R}, \mathcal{B}, \mathbb{P}_X)$, obtido pela aplicação da função X ao espaço $(\Omega, \mathcal{F}, \mathbb{P})$, obedece aos axiomas da probabilidade. Utilizando a definição de função mensurável, que nos garante que $\forall \mathbf{A} \in \mathcal{B} \Rightarrow f^{-1}(\mathbf{A}) \in \mathcal{F}$, podemos demonstrar que a classe $\mathcal{B}$ e a função $\mathbb{P}_X$ obedecem aos axiomas propostos por Kolmogoroff.

Assim, utilizando uma variável aleatória X , a medida $\mathbb{P}_X$ associada aos borelianos (a medida $\mathbb{P}$, do espaço inicial $(\Omega, \mathcal{F}, \mathbb{P})$, transposta para $\mathbb{R}$) estará definida para quase todos os conjunto de $\mathbb{R}$. Porém, apesar de $\mathbb{R}$ ser um conjunto com o qual estamos mais familiarizados, continuamos a operar com funções de conjuntos. Mas, como referimos, dado que a classe $\mathcal{B}$ pode ser gerada pelas semi-rectas $] - \infty, x]$, $\forall x \in \mathbb{R}$, a função de distribuição $F_X(x) = \mathbb{P}_X(] - \infty, x])$ é uma função usual, cujo argumento x é um ponto. A função de distribuição F_X contém toda a informação do espaço de probabilidade $(\mathbb{R}, \mathcal{B}, \mathbb{P}_X)$ e é uma função cujo argumento é um ponto, razões pelas quais se tornou num dos principais conceitos da Teoria da Probabilidade.

Definição 10.2 (Função de Distribuição). Seja X uma variável aleatória associada ao espaço de probabilidade $(\Omega, \mathcal{A}, \mathbb{P})$. Chama-se função de distribuição da variável aleatória X à função $F_X : \mathbb{R} \rightarrow \mathbb{R}$ tal que

$$F_X(x) = \mathbb{P}_X(] - \infty, x]) = \mathbb{P}(\{\omega \in \Omega : X(\omega) \leq x\}). \quad (10.40)$$

Refira-se que Kolmogoroff (1933) define a função de distribuição através de $F_X(x) = \mathbb{P}(X < x)$, porém, como o raciocínio é análogo ao que aqui expomos optamos por utilizar a definição usual $F_X(x) = \mathbb{P}(X \leq x)$. Pelo facto de esta função ser fundamental em toda a construção da Teoria da Probabilidade vamos apresentar as suas principais características. Consideremos um espaço de probabilidade $(\Omega, \mathcal{A}, \mathbb{P})$, ao qual aplicamos uma variável aleatória X , através da qual se obtém o espaço de probabilidade $(\mathbb{R}, \mathcal{B}, \mathbb{P}_X)$ caracterizado pela função de distribuição $F_X(x) = \mathbb{P}_X([-\infty, x])$. Esta função verifica as seguintes propriedades:

1. $\mathbb{P}_X([a, b]) = F_X(b) - F_X(a)$, para $a < b$, pois pela aditividade de $\mathbb{P}_X$ obtemos

$$\mathbb{P}_X([-\infty, b]) = \mathbb{P}_X([-\infty, a] \cup]a, b]) = \mathbb{P}_X([-\infty, a]) + \mathbb{P}_X(]a, b])^{(18)}.$$

2. F_X é monótona não decrescente — $F(b) \geq F(a)$ para $b \geq a$. Esta propriedade é consequência imediata da propriedade anterior, uma vez que $F_X(b) - F_X(a) = \mathbb{P}_X([a, b]) \geq 0$;
3. Limites de F_X : $\lim_{x \rightarrow -\infty} F_X(x) = 0$ e $\lim_{x \rightarrow +\infty} F_X(x) = 1$. Considerando uma sequência $a_1 > a_2 > \dots > a_n > \dots \searrow -\infty$, então os conjuntos definidos por $\mathbf{A}_i =]-\infty, a_i]$ serão decrescentes verificando $\mathbf{A} = \bigcap_{i=1}^{\infty} \mathbf{A}_i = \emptyset$. Desta forma podemos concluir que $\lim_{i \rightarrow \infty} F_X(a_i) = \lim_{i \rightarrow \infty} \mathbb{P}_X(\mathbf{A}_i) = \mathbb{P}_X(\emptyset) = 0$. De forma análoga,

⁽¹⁸⁾ Notemos que esta característica está na base das Distribuições, pois considerando o espaço de medida $(\Omega, \mathcal{F}, \mu)$, onde μ é uma medida finita, denomina-se F por função de distribuição da medida μ se e só se $\mu([a, b]) = F(b) - F(a)$ para quaisquer $a \leq b \in \mathbb{R}$. Recordemos, por exemplo, que a fórmula de Barrow é definida por

$$\mu([a, b]) = F(b) - F(a) = \int_a^b F' \, d\mu.$$

definindo uma sequência $b_1 < b_2 < \cdots < b_n < \cdots \nearrow +\infty$, os conjunto $\mathbf{B}_i =]-\infty, b_i]$ serão crescentes, sendo o seu limite $\mathbf{B} = \bigcap_{i=1}^{\infty} \mathbf{B}_i = \mathbb{R}$. Assim deduzimos que $\lim_{i \rightarrow \infty} F_X(b_i) = \lim_{i \rightarrow \infty} \mathbb{P}_X(\mathbf{B}_i) = \mathbb{P}_X(\mathbb{R}) = 1$.

4. F_X é limitada — $0 \leq F_x \leq 1, \forall x \in \mathbb{R}$ (consequência das duas propriedades precedentes).
5. F_X é contínua à direita — $\lim_{x \rightarrow x_0^+} F_X(x) = F_X(x_0)$. Se definirmos uma sequência $b_1 > b_2 > \cdots > b_n > \cdots \searrow a$, os acontecimentos $\mathbf{B}_i =]a, b_i]$ (decrecentes) e $\mathbf{B} = \bigcap_{i=1}^{\infty} \mathbf{B}_i = \emptyset$ teremos, pela propriedade 1, $\lim_{i \rightarrow \infty} F_X(b_i) - F_X(a) = \lim_{i \rightarrow \infty} \mathbb{P}_X(\mathbf{B}_i) = \mathbb{P}_X(\emptyset) = 0$.
6. $\lim_{x \rightarrow x_0^-} F_X(x) = F_X(x_0) - \mathbb{P}_X(\{x_0\})$. Se definirmos uma sequência $a_1 < a_2 < \cdots < a_n < \cdots \nearrow b$, os acontecimentos $\mathbf{A}_i =]a_i, b]$ (decrecentes) e $\mathbf{A} = \bigcap_{i=1}^{\infty} \mathbf{A}_i = \{b\}$ teremos, pela propriedade 1, $\lim_{i \rightarrow \infty} F_X(b) - F_X(a_i) = \lim_{i \rightarrow \infty} \mathbb{P}_X(\mathbf{A}_i) = \mathbb{P}_X(\{b\})$. Assim, a função será contínua (à esquerda) se $\mathbb{P}_X(\{b\}) = 0$ (isto é, $\mathbb{P}_X(\{x_0\}) = 0$ na fórmula inicial). Os pontos de descontinuidade da função de distribuição correspondem a pontos com probabilidade não nula $\mathbb{P}_X(\{b\}) \neq 0$.

Salientemos, também, que qualquer função F que satisfaça as propriedades 2, 3 e 5 determina um espaço de probabilidade $(\mathbb{R}, \mathcal{B}, \mathbb{P})$, isto é, pode ser considerada uma função de distribuição de um espaço de probabilidade.

Deste modo, em vez de utilizarmos o espaço de probabilidade, podemos associar a cada experiência aleatória uma variável aleatória X que é caracterizada pela sua função de distribuição — $F_X(x)$, que pode ser escrita através do integral

$$F_X(x) = \mathbb{P}_X(X \leq x) = \int_{-\infty}^x dF_X(t), \forall x \in \mathbb{R}. \quad (10.41)$$

Para garantir a representação integral de F_X temos que recorrer ao Teo-

rema de Radon-Nikodym baseado na noção de medidas absolutamente contínuas.

Definição 10.3 (Continuidade Absoluta). Sejam μ e ν duas medidas definidas em $\mathcal{F}$. Dizemos que μ é absolutamente contínua em relação a ν e representamos por $\mu \ll \nu$ se e só se

$$\forall \mathbf{A} \in \mathcal{F} : \nu(\mathbf{A}) = 0 \Rightarrow \mu(\mathbf{A}) = 0. \quad (10.42)$$

Desta forma uma medida μ é absolutamente contínua em relação à medida ν se todos os conjuntos de $\mathcal{F}$ com medida ν nula também tiverem medida μ nula. Com esta noção chegamos a um dos teoremas mais importantes na construção de Kolmogoroff, o Teorema de Radon-Nikodym.

Teorema 10.6 (Teorema de Radon-Nikodym). *Num espaço mensurável $(\mathbf{X}, \mathcal{F})$ seja μ uma medida σ -finita. Seja ν uma medida finita, absolutamente contínua com respeito a μ , $\nu \ll \mu$, então existe uma função $\mathcal{F}$ -mensurável $f : \mathbf{X} \rightarrow [0, +\infty]$ tal que*

$$\nu(\mathbf{A}) = \int_{\mathbf{A}} f \, d\mu, \quad (10.43)$$

para qualquer $\mathbf{A} \in \mathcal{F}$, onde $f = \frac{d\nu}{d\mu}$ é a derivada de Radon-Nikodym (ou densidade de Radon-Nikodym) de ν em ordem à medida μ . Se existir outra função g , tal que $\nu(\mathbf{A}) = \int_{\mathbf{A}} g \, d\mu$ para qualquer $\mathbf{A} \in \mathcal{F}$, então $\mu(f \neq g) = 0$ (a função f é única excepto num conjunto de medida μ nula).

Consoante as especificidades da função de distribuição F_X da variável aleatória X , teremos funções densidade com características distintas. Deste modo, habitualmente as variáveis aleatórias são divididas em discretas, (absolutamente) contínuas ou mistas.

Se a função de distribuição F_X for constante, excepto num conjunto constituído por um número finito ou infinito numerável de pontos, a variável

aleatória será discreta e teremos

$$\mathbb{P}_X(\mathbf{A}) = \int_{\mathbf{A}} f_X(x_i) \, d\# = \int_{\mathbf{A}} dF_X(t), \quad (10.44)$$

onde $f_X(x_i)$ denomina-se por função massa de probabilidade da variável aleatória X , verificando-se $f_X(x_i) = \mathbb{P}_X(x_i)$. Neste caso $f_X = \frac{d\mathbb{P}_X}{d\#}$, isto é, a função massa de probabilidade da variável aleatória X corresponde à densidade de Radon-Nikodym, sendo a derivada em ordem à medida de contagem. Assim, a função de probabilidade também pode ser vista como uma função densidade (de Radon-Nikodym). O suporte da medida $\mathbb{P}_X$ será definido por $\mathbb{S}_X = \{x_i \in \mathbb{R} : \mathbb{P}_X(x_i) > 0\}$, que será constituído por um número finito ou infinito numerável de pontos. Desta forma teremos $\sum_{x_i \in \mathbb{S}_X} f_X(x_i) = 1$ e, para cada acontecimento $\mathbf{A} \in \mathcal{F}$, se verificará $\mathbb{P}_X(\mathbf{A}) = \sum_{x_i \in \mathbf{A}} f_X(x_i)$.

Se a função de distribuição $F_X(x)$ da variável aleatória X é contínua em $\mathbb{R}$, então a função densidade $f_X(x)$ será dada por $f_X(x) = \frac{d\mathbb{P}_X}{d\lambda}$, onde λ representa a medida de Lebesgue, isto é, a função densidade corresponde à derivada de Radon-Nikodym em ordem à medida de Lebesgue. Assim teremos

$$f_X(x) = \frac{d}{dx} F_X(x), \quad (10.45)$$

para todos os valores $x \in \mathbb{R}$, excepto, quando muito, um conjunto numerável de pontos. Neste caso denominamos a variável aleatória por absolutamente contínua e podemos expressar a probabilidade $\mathbb{P}_X(\mathbf{A})$ para qualquer boreliano $\mathbf{A} \in \mathcal{B}$ através de

$$\mathbb{P}_X(\mathbf{A}) = \int_{\mathbf{A}} f_X(x) \, dx \quad (10.46)$$

e obrigatoriamente teremos

$$\mathbb{P}_X(\mathbb{R}) = \int_{\mathbb{R}} f_X(x) \, dx = 1. \quad (10.47)$$

Nestes casos o suporte de $\mathbb{P}_X$ pode ser definido por

$$\mathbb{S}_X = \{x \in \mathbb{R} : \mathbb{P}_X[x - \varepsilon < X(\omega) < x + \varepsilon] > 0, \forall \varepsilon > 0\}, \quad (10.48)$$

onde a função densidade será positiva, logo, podemos igualmente definir o suporte através de

$$\mathbb{S}_X = \{x \in \mathbb{R} : f_X(x) > 0\}. \quad (10.49)$$

Deste modo, $\mathbb{S}_X$ representa os valores de $\mathbb{R}$ que a variável aleatória X pode assumir, apesar de terem probabilidade nula de ocorrerem.

Há ainda as variáveis mistas que, como o próprio nome indica, serão constituídas por pontos com probabilidade não nula, mas cuja soma das probabilidades é distinta da unidade, e por pontos com probabilidade nula mas que podem ocorrer. Neste caso a função de distribuição não será contínua em $\mathbb{R}$ (pois terá *saltos* nos pontos que assumem uma probabilidade não nula), nem será constante excepto num número finito ou infinito numerável de pontos. Nesta situação podemos dividir a análise da função densidade nos pontos com probabilidade não nula e nos intervalos onde a função de distribuição é contínua mas não constante. Para efectuarmos esta divisão é necessário recorrermos à noção de medidas singulares e ao Teorema da Decomposição de Lebesgue.

Definição 10.4 (Medidas singulares). Sejam μ e ν duas medidas. Dizemos que as medidas μ e ν são singulares, representando por $\mu \perp \nu$, se e só se μ tem suporte $\mathbb{S}_\mu$ num conjunto de medida ν nula ($\nu(\mathbb{S}_\mu) = 0$) e ν tem suporte $\mathbb{S}_\nu$ num conjunto de medida μ nula ($\mu(\mathbb{S}_\nu) = 0$).

Deste modo, o objectivo é dividir a análise das variáveis mistas em duas medidas singulares, uma que caracterize a sua parte discreta e outra que caracterize a sua parte contínua. Esta divisão é possível através do Teorema da Decomposição de Lebesgue.

Teorema 10.7 (Teorema da Decomposição de Lebesgue). *Sejam μ e ν duas medidas definidas em $(X, \mathcal{F})$. Então existe um único par de medidas (ν_c, ν_d)*

definidas em $(X, \mathcal{F})$ tais que seja possível a decomposição $\nu = \nu_c + \nu_d$, com $\nu_c \ll \mu$ e $\nu_d \perp \mu$.

O Teorema da Decomposição de Lebesgue permite-nos decompor uma medida ν em duas medidas singulares. Assim, quando estamos perante funções de distribuição que não se encaixam nas características das discretas ou absolutamente contínuas, podemos, pelo Teorema de Decomposição de Lebesgue, decompor a medida $\mathbb{P}_X$ em duas medidas singulares, uma discreta e outra contínua, para caracterizar a variável.

Doravante não vamos fazer distinção entre variáveis aleatórias discretas, contínuas ou mistas e denominá-las-emos sempre função densidade (mesmo que se trate de uma função de probabilidade de uma variável aleatória discreta).

Salientemos (uma vez mais) que, com estas definições, se torna bastante mais atractivo modelar um fenómeno aleatório procurando a função de distribuição que o caracterize, do que tentar operar no espaço de probabilidade inicial. Nesta transposição a variável aleatória e a função de distribuição desempenham um papel fundamental, sendo, por esta razão, dois conceitos de importância capital na Teoria da Probabilidade. Deste modo, podemos caracterizar o espaço de probabilidade $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathbb{P}_X)$ recorrendo unicamente a uma função de distribuição F_X .

Pacheco d'Amorim efectua um raciocínio semelhante com o seu conceito ponto imagem, apesar de não ter disponíveis os resultados da Teoria da Medida que Kolmogoroff tinha vinte anos depois e, consequentemente, não ter fundamentado convenientemente a sua construção (conforme comentamos no capítulo 6). Contudo, pensamos que na construção do ponto imagem de Pacheco d'Amorim, caso tivesse sido fundamentada com os resultados utilizados por Kolmogoroff, teríamos o espaço de probabilidade $(\Omega, \mathcal{F}, \mathbb{P})$ associado à

escolha à sorte do elemento livre x , caracterizado pela equipossibilidade, e o espaço de probabilidade $(\mathbb{R}, \mathcal{B}, \mathbb{P}_y)$, associado ao ponto imagem y , caracterizado por uma lei de possibilidade π_y . A lei de possibilidade π_y corresponde à derivada de Radon-Nikodym da medida de Lebesgue ou da medida de contagem (consoante estejamos a trabalhar com lançamentos ou tiragens) ou a ambas, caso seja um misto entre lançamentos e tiragens, onde teríamos que decompor π_y nas duas situações. A transformação do ponto livre para o ponto imagem, onde utilizamos $x = f(y)$, seria descrita através de $x = F_y^{-1}(y)$, com F_y^{-1} representando a função inversa (ou inversa generalizada para as situações onde existam pontos com probabilidade não nula) da função F_y definida por $F_y(y_0) = \mathbb{P}_y(y \leq y_0)$. Por este motivo consideramos que Pacheco d'Amorim não esteve longe de alcançar a definição de conceitos, actualmente essenciais na Teoria da Probabilidade, ainda desconhecidos em 1914.

10.2.4 Esperança Matemática e Probabilidade Condi- onada

Kolmogoroff define a esperança matemática de uma variável aleatória, que é uma função mensurável, pelo integral de Lebesgue dessa função.

Definição 10.5 (Esperança Matemática). Seja $(\Omega, \mathcal{F}, \mathbb{P})$ um espaço de probabilidade e X uma variável aleatória definida em $\mathcal{F}$. A **Esperança Matemática** da variável aleatória X é definida como sendo o integral de Lebesgue da função X em ordem à medida $\mathbb{P}$, isto é,

$$\mathbb{E}(X) = \int_{\Omega} X(\omega) \, d\mathbb{P}(\omega) \quad (10.50)$$

ou, de igual forma, pelo integral de Stieltjes

$$\mathbb{E}(X) = \int_{\mathbb{R}} x \, dF_X(x), \quad (10.51)$$

onde F_X representa a função de distribuição da variável aleatória X definida por $F_X = \mathbb{P}_X (]-\infty, x])$. O valor esperado da variável aleatória X existe se e só se

$$\int_{\mathbb{R}} |x| \, dF_X(x) < +\infty. \quad (10.52)$$

Através das propriedades do integral podemos demonstrar diversas propriedades do valor esperado, tais como: $|\mathbb{E}(X)| \leq \mathbb{E}(|X|)$; $\inf(X) \leq \mathbb{E}(X) \leq \sup(X)$; se $X \leq Y$, então $\mathbb{E}(X) \leq \mathbb{E}(Y)$; $\mathbb{E}(a + bX + cY) = a + b\mathbb{E}(X) + c\mathbb{E}(Y)$ (para a, b, c constantes reais); se a série $\sum_{i=1}^{\infty} \mathbb{E}(X_i)$ converge, então $\mathbb{E}(\sum_{i=1}^{+\infty} X_i) = \sum_{i=1}^{\infty} \mathbb{E}(X_i)$. Muitas outras propriedades essenciais em diferentes áreas da Teoria da Probabilidade se poderiam apontar.

Em relação à probabilidade condicional, Kolmogoroff define, para o caso finito, a probabilidade do acontecimento $\mathbf{A}$ condicionada a $\mathbf{B}$, considerando que $\mathbb{P}(\mathbf{B}) \neq 0$, da forma usual, isto é, através de⁽¹⁹⁾

$$\mathbb{P}_{\mathbf{B}}(\mathbf{A}) = \mathbb{P}(\mathbf{A}|\mathbf{B}) = \frac{\mathbb{P}(\mathbf{A} \cap \mathbf{B})}{\mathbb{P}(\mathbf{B})}. \quad (10.53)$$

Assim, se $\mathbf{B}$ é um acontecimento de probabilidade não nula definido no espaço de probabilidade $(\Omega, \mathcal{A}, \mathbb{P})$, tal que $\mathbb{P}(\mathbf{B}) > 0$, então $\mathbb{P}_{\mathbf{B}}(\mathbf{A}) = \mathbb{P}(\mathbf{A}|\mathbf{B})$, definida pela expressão (10.53), é uma medida de probabilidade definida no mesmo espaço mensurável $(\Omega, \mathcal{A})$, isto é, $(\Omega, \mathcal{A}, \mathbb{P}_{\mathbf{B}})$ é um espaço de probabilidade (satisfaz os axiomas para o caso finito).

No caso geral, Kolmogoroff (1933, p. 47–48) utiliza outros argumentos que expomos recorrendo à apresentação original e a notações mais correntes. Seja $(\Omega, \mathcal{A}, \mathbb{P})$ um espaço de probabilidade, $\mathbf{B} \in \mathcal{A}$ um acontecimento e U uma variável aleatória. Kolmogoroff procura a função (variável aleatória)

⁽¹⁹⁾ Embora Kolmogoroff, na sua obra, representasse a probabilidade condicionada através de $\mathbb{P}_{\mathbf{B}}(\mathbf{A})$, vamos utilizar a notação corrente que foi introduzida por Harold Jeffreys em 1931.

$\mathbb{P}_U(\mathbf{B}) = \pi(U(\omega), \mathbf{B})$ tal que $\omega \mapsto \pi(U(\omega), \mathbf{B})$, que denomina por probabilidade condicional de $\mathbf{B}$ sabendo U e que deverá ser tal que, para qualquer acontecimento $\mathbf{A} \in \mathcal{B}$ com $\mathbb{P}(U(\omega) \in \mathbf{A}) > 0$, verifica⁽²⁰⁾

$$\mathbb{P}(\mathbf{B}|U \in \mathbf{A}) = \int_{\Omega} \pi(U(\omega), \mathbf{B}) \, d\mathbb{P}(\cdot|U \in \mathbf{A}). \quad (10.54)$$

Se, para qualquer $\mathbf{A} \in \mathcal{B}$ definirmos $Q_{\mathbf{B}}(\mathbf{A}) = \mathbb{P}(\mathbf{B} \cap U^{-1}(\mathbf{A}))$ e se representarmos por $\mathbb{P}_U$ a medida de U definida por $\mathbb{P}_U(\mathbf{A}) = \mathbb{P}(U^{-1}(\mathbf{A}))$, então $\mathbb{P}_U(\mathbf{A}) = 0 \Rightarrow Q_{\mathbf{B}}(\mathbf{A}) = 0$ (isto é, a medida $Q_{\mathbf{B}}$ é absolutamente contínua em relação a $\mathbb{P}_U$ — $Q_{\mathbf{B}} \ll \mathbb{P}_U$) e pelo Teorema de Radon-Nikodym (Teorema 10.6) existe uma única função $f_{\mathbf{B}}$ (excepto num conjunto de probabilidade nula) tal que para qualquer boreliano $\mathbf{A}$ teremos

$$Q_{\mathbf{B}}(\mathbf{A}) = \int_{\mathbf{A}} f_{\mathbf{B}} \, d\mathbb{P}_U = \int_{\mathbb{R}} \mathbb{I}_{\mathbf{A}} f_{\mathbf{B}} \, d\mathbb{P}_U = \int_{\Omega} \mathbb{I}_{U \in \mathbf{A}} f_{\mathbf{B}} \circ U \, d\mathbb{P} \quad (10.55)$$

e, consequentemente,

$$\mathbb{P}(\mathbf{B}|U \in \mathbf{A}) = \int_{\Omega} f_{\mathbf{B}} \circ U \, d\mathbb{P}(\cdot|U \in \mathbf{A}), \quad (10.56)$$

sendo a variável aleatória $\pi(U(\omega), \mathbf{B})$ definida por $\pi(U(\omega), \mathbf{B}) = f_{\mathbf{B}} \circ U$.

De uma forma mais simplificada, podemos visualizar esta construção do seguinte modo. Consideremos um espaço de probabilidade $(\Omega, \mathcal{A}, \mathbb{P})$ e uma variável aleatória X não negativa tal que $\mathbb{E}(X) = 1$. Se definirmos a função de conjuntos $\mathbb{Q}$ em $\mathcal{A}$ através de

$$\mathbb{Q}(\mathbf{A}) = \mathbb{E}(\mathbb{I}_{\mathbf{A}} X), \quad \forall \mathbf{A} \in \mathcal{A}, \quad (10.57)$$

facilmente concluímos que $(\Omega, \mathcal{A}, \mathbb{Q})$ é um espaço de probabilidade pois

$$\mathbb{Q}(\mathbf{A}) = \mathbb{E}(\mathbb{I}_{\mathbf{A}} X) \geq 0 \quad (10.58)$$

⁽²⁰⁾ As variáveis aleatórias, como referimos, são funções $\Omega \mapsto \mathbb{R}$. Nós representamos por $U(\omega) \in \mathbf{A}$, ou simplesmente por $U \in \mathbf{A}$, os conjuntos de Ω , pertencentes a $\mathcal{A}$, definidos por $\{\omega : U(\omega) \in \mathbf{A}\}$ que Kolmogoroff representou por $\{u \subset \mathbf{A}\}$.

como consequência de $X \geq 0$,

$$\mathbb{Q}(\Omega) = \mathbb{E}(\mathbb{I}_\Omega X) = \mathbb{E}(X) = 1, \quad (10.59)$$

e, considerando uma sucessão de acontecimentos $\{\mathbf{A}_i\}_{i \in \mathbb{N}}$ disjuntos dois a dois, teremos

$$\begin{aligned} \mathbb{Q}\left(\bigcup_{i=1}^{\infty} \mathbf{A}_i\right) &= \mathbb{E}\left(\mathbb{I}_{\bigcup_{i=1}^{\infty} \mathbf{A}_i} X\right) = \mathbb{E}\left(\sum_{i=1}^{\infty} \mathbb{I}_{\mathbf{A}_i} X\right) = \\ &= \sum_{i=1}^{\infty} \mathbb{E}(\mathbb{I}_{\mathbf{A}_i} X) = \sum_{i=1}^{\infty} \mathbb{Q}(\mathbf{A}_i). \end{aligned} \quad (10.60)$$

Notemos agora que, quando $\mathbb{P}(\mathbf{A}) = 0$, teremos também $\mathbb{Q}(\mathbf{A}) = 0$, uma vez que

$$\mathbb{Q}(\mathbf{A}) = \mathbb{E}(\mathbb{I}_{\mathbf{A}} X) = \int_{\Omega} \mathbb{I}_{\mathbf{A}} X \, d\mathbb{P} = 0, \quad (10.61)$$

dado que $\mathbb{P}(\mathbf{A}) = 0$. Assim, a medida $\mathbb{Q}$ é absolutamente contínua em relação à medida $\mathbb{P}$, i.e. $\mathbb{Q} \ll \mathbb{P}$. Então pelo Teorema de Radon-Nikodym podemos concluir que a medida $\mathbb{Q}$ que verifica (10.57) é única (quase certamente).

Em suma, através do Teorema de Radon-Nikodym, demonstrado em 1930, Kolmogoroff garantiu que existe uma única função densidade (excepto num conjunto com probabilidade nula) que caracteriza a probabilidade condicionada. Desta forma podemos enunciar este resultado, apresentado por Kolmogoroff, da forma que passamos a reconstruir.

Teorema 10.8. *Considerando um espaço de probabilidade $(\Omega, \mathcal{A}, \mathbb{P})$ e um acontecimento fixo $\mathbf{B} \in \mathcal{A}$ tal que $\mathbb{P}(\mathbf{B}) > 0$, então definindo uma medida de probabilidade condicional a $\mathbf{B}$, $\mathbb{P}_{\mathbf{B}}(\mathbf{A}) = \mathbb{P}(\mathbf{A}|\mathbf{B})$, $\forall \mathbf{A} \in \mathcal{A}$, também $(\Omega, \mathcal{A}, \mathbb{P}_{\mathbf{B}})$ é um espaço de probabilidade.*

A probabilidade condicionada está, a partir deste momento, matematicamente definida, não surgindo da observação da realidade, isto é, sem depender de qualquer contextualização, nomeadamente de uma ideia sequencial,

nem estando restrita à situação em que o condicionado é obrigatoriamente um subconjunto do condicionante (como na construção de Diogo Pacheco d'Amorim). Por outro lado, como referimos previamente, com o recurso ao Teorema de Radon-Nikodym, Kolmogoroff consegue definir a probabilidade condicionada de forma a garantir a unicidade de uma única medida definida no espaço condicionado e, consequentemente, resolver o actualmente denominado paradoxo de Borel-Kolmogoroff, que apresentamos na secção 4.6.

Refira-se que a probabilidade condicionada é, de certa forma, mais geral que a probabilidade absoluta (não condicionada), pois esta última pode ser vista como um caso particular da probabilidade condicionada considerando $\mathbb{P}(\mathbf{A}) = \mathbb{P}(\mathbf{A}|\Omega)$. Por este motivo existem axiomáticas que preferem basear-se nesta probabilidade para a construção da teoria, tal como a de Pacheco d'Amorim (1914) ou a proposta de Rényi (1955), baseada em espaços de probabilidade condicional.

10.2.5 Espaços de dimensão superior e independência

No caso de pretendermos analisar uma experiência $\mathcal{E}$ que consiste em efectuar conjuntamente n sub-experiências aleatórias $\mathcal{E}_1, \mathcal{E}_2, \dots, \mathcal{E}_n$, onde cada experiência $\mathcal{E}_i$ tem associada o espaço de probabilidade $(\Omega_{(i)}, \mathcal{A}_{(i)}, \mathbb{P}_{(i)})$, $i = 1, \dots, n$, teremos de utilizar como espaço de probabilidade o espaço produto (tarefa nada fácil, mesmo no caso mais simples de independência, isto é, de ortogonalidade dos espaços). Deste modo, para a experiência $\mathcal{E}$, podemos representar por $\Omega = \prod_{i=1}^n \Omega_{(i)}$ o espaço de resultado, por $\mathcal{A}$ a σ -álgebra induzida e, neste contexto, pretendemos encontrar (se existir) a função $\mathbb{P}$ tal que $(\Omega, \mathcal{A}, \mathbb{P})$ seja um espaço de probabilidade consistente com as medidas de probabilidade dos espaços marginais $(\Omega_{(i)}, \mathcal{A}_{(i)}, \mathbb{P}_{(i)})$. Kolmogoroff analisa este problema para as situações que $\Omega_i = \mathbb{R}$ e $\mathcal{A}_{(i)} = \mathcal{B}(\mathbb{R})$ (os borelianos) recorrendo às funções de distribuição associadas. Deste modo, se trabalharmos

ao nível das funções de distribuição, teremos, no caso n -dimensional,

$$F_{(X_1, \dots, X_n)}(x_1, \dots, x_n) = \mathbb{P}_{(X_1, \dots, X_n)}(X_1 \leq x_1 \wedge \dots \wedge X_n \leq x_n), \quad (10.62)$$

e correspondente função densidade conjunta $f_{(X_1, \dots, X_n)}(x_1, \dots, x_n)$. Esta função está relacionada com as funções densidade marginais $f_{X_i}(x_i)$, para $i = 1, \dots, n$, associadas aos espaços $(\mathbb{R}, \mathcal{B}, \mathbb{P}_{X_i})$ obtidos pela aplicação da função mensurável X_i ao espaço $(\Omega_{(i)}, \mathcal{A}_{(i)}, \mathbb{P}_{(i)})$. Contudo, esta relação é definida pela forma de dependência entre as variáveis X_i , razão pela qual conceitos como o de independência e o de permutabilidade assumem enorme importância na Teoria da Probabilidade.

A função de distribuição conjunta verifica diversas características, semelhantes às do caso unidimensional, tais como ser não decrescente e contínua à direita em cada variável e obedecendo aos seguintes limites:

$$\lim_{x_i \rightarrow -\infty} F_{(X_1, \dots, X_n)}(x_1, \dots, x_n) = 0, \quad \forall i; \quad (10.63)$$

e

$$\lim_{x_1 \rightarrow \infty, \dots, x_n \rightarrow \infty} F_{(X_1, \dots, X_n)}(x_1, \dots, x_n) = 1. \quad (10.64)$$

Um dos pontos mais importantes dos fundamentos de Kolmogoroff é a sua análise dos espaços de dimensão infinita. Com efeito, Kolmogoroff define as condições que a função de distribuição de um espaço de dimensão infinita deverá verificar de forma a ser consistente com um espaço de medida de dimensão infinita, o que actualmente se denomina por Teorema da Consistência de Kolmogoroff, mas o autor designou por Teorema Fundamental.

Teorema 10.9 (Teorema da Consistência de Kolmogoroff). *Todos os sistemas de funções de distribuição $F_{X_1, \dots, X_n}$ satisfazendo*

$$F_{(X_{i_1}, \dots, X_{i_n})}(x_{i_1}, \dots, x_{i_n}) = F_{(X_1, \dots, X_n)}(x_1, \dots, x_n) \quad (10.65)$$

onde $\{i_1, i_2, \dots, i_n\}$ representa uma permutação do conjunto $\{1, 2, \dots, n\}$ e

$$F_{(X_1, \dots, X_k)}(x_1, \dots, x_k) = F_{(X_1, \dots, X_n)}(x_1, \dots, x_k, +\infty, \dots, +\infty) \quad (10.66)$$

com $k < n$, definem uma função de probabilidade $\mathbb{P}(\mathbf{A})$ em $\mathcal{A}^n$ que satisfaz os axiomas. Esta função de probabilidade $\mathbb{P}(\mathbf{A})$ pode ser estendida (pelo Teorema da Extensão de Carathéodory) à σ -álgebra $\mathcal{B}(\mathcal{A}^n)$.

Segundo Shafer e Vovk (2006) a análise de espaços de dimensão infinita já tinham sido analisados previamente por Daniell, num artigo de 1919, mas crêem que Kolmogoroff desconhecesse o resultado. Contudo, o Teorema da Consistência de Kolmogoroff representa um dos resultados mais importantes inseridos nos seus fundamentos, por ser essencial, por exemplo, nos resultados limites e nos processos estocásticos contínuos.

A independência é um dos conceitos fundamentais do Cálculo das Probabilidades, como podemos inferir das seguintes palavras de Kolmogoroff (1933).

“The mere introduction of the above concepts, therefore, would not be sufficient to produce a basis for development of a large new theory.

Historically, the independence of experiments and random variables represents the very mathematical concept that has given the theory of probability its peculiar stamp.”

[Kolmogoroff 1933, p. 8]

Desta forma, para Kolmogoroff, a noção de independência é responsável pelo poder que os métodos probabilísticos possuem dentro da Matemática pura.

Definição 10.6 (Acontecimentos independentes). Seja $(\Omega, \mathcal{A}, \mathbb{P})$ um espaço de probabilidade.

1. Diz-se que os acontecimentos $\mathbf{A}_i$, $i = 1, \dots, n$, com $\mathbf{A}_i \in \mathcal{A}$, são mutuamente independentes se, para todo o subconjunto $\{i_1, \dots, i_k\}$ de $\{1, \dots, n\}$, verificar-se

$$\mathbb{P} \left(\bigcap_{j=1}^k \mathbf{A}_{i_j} \right) = \prod_{j=1}^k \mathbb{P}(\mathbf{A}_{i_j}), \quad (10.67)$$

para $k = 2, \dots, n$;

2. Seja $\{\mathbf{A}_i\}_{i=1}^{+\infty}$ uma sucessão de acontecimentos de $\mathcal{A}$. Diz-se que os acontecimentos $\{\mathbf{A}_i\}$ são independentes se e só se para todo $k \in \mathbb{N}$ e para todo subconjunto $\{i_1, \dots, i_k\}$ de $\mathbb{N}$ verificar-se

$$\mathbb{P} \left(\bigcap_{j=1}^k \mathbf{A}_{i_j} \right) = \prod_{j=1}^k \mathbb{P}(\mathbf{A}_{i_j}). \quad (10.68)$$

Deste modo, para que num conjunto de n acontecimentos estes sejam independentes, não é suficiente que os acontecimentos sejam independentes dois a dois, i.e. $\mathbb{P}(\mathbf{A}_i \cap \mathbf{A}_j) = \mathbb{P}(\mathbf{A}_i) \mathbb{P}(\mathbf{A}_j)$, mas que tenham de verificar todas as $2^n - n - 1$ equações presentes em (10.67), como ilustra o seguinte exemplo que Kolmogoroff (1933) refere ser originário de Bernstein.

Exemplo 10.3 (Exemplo de Bernstein). Consideremos uma experiência aleatória $\mathcal{E}$ com espaço de resultados $\Omega = \{\mathbf{A}_1, \mathbf{A}_2, \mathbf{A}_3, \mathbf{A}_4\}$ com probabilidades associadas $\mathbb{P}(\mathbf{A}_i) = \frac{1}{4}$. Definam-se os acontecimentos $\mathbf{B}_1 = \{\mathbf{A}_1, \mathbf{A}_2\}$, $\mathbf{B}_2 = \{\mathbf{A}_1, \mathbf{A}_3\}$ e $\mathbf{B}_3 = \{\mathbf{A}_1, \mathbf{A}_4\}$. Facilmente se determina $\mathbb{P}(\mathbf{B}_1 \cap \mathbf{B}_2) = \mathbb{P}(\mathbf{B}_1 \cap \mathbf{B}_3) = \mathbb{P}(\mathbf{B}_2 \cap \mathbf{B}_3) = \frac{1}{4} = \left(\frac{1}{2}\right)^2$. Contudo, teremos $\mathbb{P}(\mathbf{B}_1 \cap \mathbf{B}_2 \cap \mathbf{B}_3) = \frac{1}{4} \neq \left(\frac{1}{2}\right)^3$. ♦

Além da independência dos acontecimentos, podemos falar em independência de σ -álgebras e independência de variáveis aleatórias, estando estes conceitos relacionados. Considerando um espaço de probabilidade $(\Omega, \mathcal{A}, \mathbb{P})$, os acontecimentos $\mathbf{A}_1 \in \mathcal{A}$ e $\mathbf{A}_2 \in \mathcal{A}$ são independentes se e só se $\mathbb{P}(\mathbf{A}_1 \cap \mathbf{A}_2) = \mathbb{P}(\mathbf{A}_1) \mathbb{P}(\mathbf{A}_2)$. Duas sub σ -álgebras $\mathcal{A}_1$ e $\mathcal{A}_2$ de $\mathcal{A}$ (i.e.,

duas σ -álgebras contidas em $\mathcal{A}$ são independentes se e só se $\forall \mathbf{A}_1 \in \mathcal{A}_1$ e para $\forall \mathbf{A}_2 \in \mathcal{A}_2$ verificar-se $\mathbb{P}(\mathbf{A}_1 \cap \mathbf{A}_2) = \mathbb{P}(\mathbf{A}_1) \mathbb{P}(\mathbf{A}_2)$. Assim, as duas sub σ -álgebras $\mathcal{A}_1$ e $\mathcal{A}_2$ são independentes se todos os acontecimentos de $\mathcal{A}_1$ forem independentes de todos os acontecimentos de $\mathcal{A}_2$. Duas variáveis aleatórias X_1 e X_2 são independentes se e só se as σ -álgebras $\mathcal{A}_1$ e $\mathcal{A}_2$ forem independentes, onde $\mathcal{A}_1$ é a σ -álgebra gerada pelos conjuntos $\{X_1 \leq x\}$ e $\mathcal{A}_2$ é a σ -álgebra gerada pelos conjuntos $\{X_2 \leq x\}$. Desta forma, se as variáveis aleatórias X_1 e X_2 são independentes, teremos $F_{(X_1, X_2)}(x_1, x_2) = F_{(X_1)}(x_1) F_{(X_2)}(x_2)$ para qualquer par $(x_1, x_2) \in \mathbb{R}^2$. Generalizando este resultado, as variáveis aleatórias $X_1, X_2, \dots, X_n$ são independentes se e só se as σ -álgebras $\mathcal{A}_1, \dots, \mathcal{A}_n$ forem independentes, onde $\mathcal{A}_i$ é a σ -álgebra gerada pelos conjuntos $\{X_i \leq x\}$. Nestas situações teremos

$$F_{(X_1, \dots, X_n)}(x_1, \dots, x_n) = \prod_{i=1}^n F_{(X_i)}(x_i), \quad (10.69)$$

para qualquer ponto $(x_1, \dots, x_n) \in \mathbb{R}^n$, resultado este que é fundamental nas aplicações da Teoria da Probabilidade.

Na construção de Pacheco d'Amorim, no seu capítulo PONTO IMAGEM, nomeadamente quando o autor demonstra um resultado semelhante ao Teorema de Fubinni recorrendo à função de possibilidade condicionada, existe a preocupação de lidar com a função de possibilidade conjunta. Contudo, após este resultado, o autor não explora mais esta ideia na sua obra. O conceito de independência também é fundamental na construção de Pacheco d'Amorim. Esta importância é, desde logo, salientada pela clara distinção entre selecções em classes (regiões) compostas e selecções em complexos de classes (regiões). Nas aplicações, onde recorre às Leis de Bernoulli e análogas, as observações provêm obrigatoriamente de experiência nas mesmas circunstâncias, logo, independentes. Porém, Pacheco d'Amorim não apresenta uma definição clara de independência entre as experiências ou entre os resultados da mesma experiência.

10.2.6 Aplicabilidade

Na secção 10.1, com o objectivo de salientar a dificuldade que existia, na época em que Pacheco d'Amorim apresentou a sua tese de doutoramento, de fundamentar a ligação entre a Probabilidade e a Estatística, expusemos esta mesma ideia. A axiomática de Kolmogoroff fornece toda a estrutura matemática, mas a probabilidade é unicamente vista como um conceito matemático, uma medida, pois qualquer função que satisfaça os axiomas é uma medida de probabilidade. Assim sendo, esta construção não permite determinar as probabilidades associadas a experiências aleatórias elementares, possibilitando apenas determinar relações entre as probabilidades dos diversos acontecimentos consoante as suas características (independência, incompatibilidade, complementaridade, entre outras). Desta forma, a axiomática de Kolmogoroff permite manipular as probabilidades de vários acontecimentos, mas não determiná-las e/ou interpretá-las. Porém, Kolmogoroff não ambiciona apresentar uma axiomática que explique a aplicabilidade deste conceito (Estatística), referindo na sua obra que, com este objectivo, existem outras propostas de axiomatização da probabilidade, referindo von Mises (1919) e Bernstein (1917)⁽²¹⁾.

(21) A visão de von Mises será desenvolvida na secção 10.3. Uma das primeiras tentativas de axiomatização da Probabilidade foi proposta por Sergei Natanovich Bernstein (1880-1968) que, em 1917, num artigo publicado em russo (a nossa apresentação da axiomática de Bernstein baseia-se em Maistrov (1974), sendo esta axiomática referida por diversas ocasiões nos fundamentos de Kolmogoroff (1933)), propõe uma axiomática para a probabilidade, baseada na comparação qualitativa dos acontecimentos de acordo com as suas probabilidades, constituída por dois axiomas. Representando por Ω o acontecimento certo, por $\emptyset$ o acontecimento impossível, por $\succ$ “é mais provável que” e por $\cong$ “é tão provável como”, os dois axiomas propostos por Bernstein em 1917 são:

1. Se $\mathbf{A} \neq \Omega$ então $\Omega \succ \mathbf{A}$ e se $\mathbf{A} \neq \emptyset$ então $\mathbf{A} \succ \emptyset$.
2. Se $\mathbf{A} \cong \mathbf{A}_1$, $\mathbf{B} \cong \mathbf{B}_1$ e $(\mathbf{A} \cap \mathbf{B}) = (\mathbf{A}_1 \cap \mathbf{B}_1) = \emptyset$, então $(\mathbf{A} \cup \mathbf{B}) \cong (\mathbf{A}_1 \cup \mathbf{B}_1)$.

“There are other postulational systems of the theory of probability, particularly those in which the concept of probability is not treated as one of the basic concepts, but is itself expressed by means of other concepts. (For example R. von Mises and S. Bernstein)”

[Kolmogoroff 1933, p. 2]

Com base na obra de Richard von Mises, Kolmogoroff expõe umas pequenas notas sobre a ligação da probabilidade aos dados reais, provenientes de uma qualquer experiência aleatória. Estas notas estão inseridas no primeiro capítulo referente a espaços finitos, pois Kolmogoroff considera que os espaços infinitos são unicamente idealizações da realidade *“to which nothing corresponds in the outside world”*.

A ligação entre a Teoria da Probabilidade e o mundo real, para Kolmogoroff, deverá ser efectuada através de uma experiência aleatória que pode ser repetida um qualquer número de vezes supondo-se sob o mesmo conjunto de condições. Representemos por $\Omega = \{\omega_1, \dots, \omega_n\}$ o conjunto de todas as possibilidades de resultado da experiência onde alguns resultados podem, em geral, não ocorrer, ainda que sejam incluídos no espaço-amostra Ω todos os resultados que nos pareçam *a priori* possíveis. Seja $\mathbf{A}$ um conjunto desses resultados, com $\mathbf{A} \subset \Omega$. Assim, sob determinadas condições (que Kolmogoroff não explicita na sua obra), podemos assumir que é associado ao acontecimento $\mathbf{A}$ o número real $\mathbb{P}(\mathbf{A})$ com as seguintes características:

1. É quase certo que, se as condições da experiência são repetidas um grande número de vezes, seja n , então se m é o número de ocorrências

Maistrov (1974) afirma que Bernstein apresenta, num livro em russo editado em 1927, uma axiomática detalhada para a teoria da probabilidade constituída por três axiomas: o da comparação de probabilidades, o dos acontecimentos incompatíveis e o da combinação de acontecimentos. Maistrov refere ainda que Glivenko, em 1939, demonstra a equivalência da axiomática de Bernstein com a de Kolmogoroff.

do acontecimento $\mathbf{A}$, o rácio $\frac{m}{n}$ vai diferir muito pouco de $\mathbb{P}(\mathbf{A})$, isto é, as frequências relativas do acontecimento $\mathbf{A}$ vão diferir pouco da sua probabilidade;

2. Se $\mathbb{P}(\mathbf{A})$ é muito pequeno, podemos estar praticamente certos de que, se efectuarmos a experiência uma vez, o acontecimento $\mathbf{A}$ não ocorrerá (princípio de Cournot).

Komogoroff salienta que as frequências relativas f , resultantes de efectuarmos uma experiência aleatória $\mathcal{E}$ repetidas vezes, obedecem aos axiomas da probabilidade finita (recordamos que Kolmogoroff expõe estas ideias de aplicabilidade da probabilidade unicamente quando enunciou os axiomas para experiências com um número finito de resultados). Assim, as frequência relativas $f_{\mathbf{A}}$ de qualquer acontecimento $\mathbf{A}$ verificarão $0 \leq f_{\mathbf{A}} = \frac{n_{\mathbf{A}}}{n} \leq 1$ e, se $\mathbf{A}$ e $\mathbf{B}$ forem incompatíveis, teremos $f_{\mathbf{A} \cup \mathbf{B}} = f_{\mathbf{A}} + f_{\mathbf{B}}$, razão pela qual o autor considera o axioma da aditividade apropriado para modelar a probabilidade.

Das ideia propostas por Kolmogoroff para a ligação entre o mundo real e a probabilidade podemos inferir que, quando desconhecemos as probabilidades associadas a uma experiência aleatória, poderemos unicamente tentar obter valores aproximados recorrendo aos resultados de uma série de experiências aleatórias idênticas. Por outro lado, no que se refere às aplicações, Kolmogoroff considera importante o princípio de Cournot, de que podemos ignorar, na prática, os acontecimento com medida de probabilidade pequena. Contudo, Kolmogoroff salienta que a segunda ideia apresentada não implica que o acontecimento não aconteça, pois podemos ter acontecimentos de medida nula (que denominou por acontecimentos praticamente impossíveis) que, numa série suficientemente longa poderão ocorrer. Assim, neste caso, em relação ao primeiro princípio teríamos que as frequências relativas seriam muito próximas da probabilidade, por exemplo $\frac{1}{n}$, quando efectuamos um número

n elevado de experiências. Existem habitualmente duas formas de enunciar o princípio de Cournot. O princípio de Cournot forte afirma que um acontecimento com baixa (ou nula) probabilidade pode ser retirado da análise, pois não irá acontecer se efectuarmos uma única prova. O forma fraca do princípio refere que os acontecimentos com fraca ou nula probabilidade vão raramente ocorrer numa sequência de provas. Notemos que Kolmogoroff refere as duas formas do princípio. Shafer e Vovk (2006) apresentam o princípio de Cournot mais detalhadamente.

10.2.7 Comentário aos fundamentos de Kolmogoroff e a versão condicional de Rényi

Frequentemente é dito que Kolmogoroff revolucionou a Teoria da Probabilidade pelo facto de ter tratado a probabilidade como uma medida. Contudo, esta não é a opinião de Kolmogoroff, pois este considera que não apresenta originalidade na sua obra no que concerne à apresentação do conceito de probabilidade como uma medida, afirmando que apenas pretende colmatar a inexistência de uma exposição completa de toda a sua construção em forma de monografia, o que Kolmogoroff conseguiu, e de forma bastante sucinta, se pensarmos que este livro revolucionário é constituído por apenas 84 páginas⁽²²⁾.

“While a conception of probability theory based on the above general viewpoints has been current for some time among certain mathematicians, there was lacking a complete exposition of the whole system, free of the extraneous complications.”

⁽²²⁾ Considerando a segunda edição inglesa de 1956, que inclui uma bibliografia suplementar de obras consideradas importantes publicadas durante os 23 anos que separam a edição original desta versão.

[Kolmogoff 1933, preface, p. v]

Esta opinião é convergente com a opinião de Paul Lévy (cf. Shafer e Vovk (2005, p. 55)).

“Starting in 1924, I gradually became accustomed to the idea that one should not consider merely what I called the true probability laws. I tried to extend a true law. I arrived at the idea, arbitrary as it might be, of a law defined in a certain Borel field. I did not think of saying to myself that this was the correct foundation for the probability calculus; I did not have the idea of publishing so simple an idea. Then, one day, I received Kolmogorov’s tract on the foundations of the probability calculus. I realized what a chance I had lost. But it was too late. When would I ever be able to figure out which of my ideas merited being published?.”

[Lévy 1970, p. 66–67]

Desta forma, no que respeita à axiomática, é-nos difícil definir qual a originalidade de Kolmogoroff, apesar de sabermos que diversos autores, tais como Maurice Fréchet, Paul Lévy, Harold Crámer, entre outros, já utilizavam as propriedades da medida de probabilidade, provavelmente sem terem presente um conjunto organizado de axiomas como Kolmogoroff (1933) propõe.

Porém, os fundamentos de Kolmogoroff não se restringem à apresentação dos axiomas. O que esta obra apresenta como novidade é a construção de probabilidades em espaços de dimensão infinita com o tratamento das funções de distribuição em espaços de dimensão infinita, essencial para as demonstrações dos teoremas limites, através do seu Teorema da Consistência, a diferenciação e a integração do valor esperado em ordem a um parâmetro (que não referimos neste trabalho) e, especialmente, a definição de probabilidade condicionada e de valor esperado condicionado que, nos casos em

que o espaço-amostra é infinito, recorre ao Teorema de Radon-Nikodym, resolvendo o paradoxo de Borel (que também é actualmente denominado por paradoxo de Borel-Kolmogoroff, por ter sido supostamente criado por Borel (1909), apesar de aparecer já na obra de Bertrand (1888), e resolvido por Kolmogoroff).

Caso Pacheco d'Amorim, quando escreveu a sua tese de doutoramento, tivesse disponíveis os resultados da Teoria da Medida para fundamentar a sua construção, para estar em concordância com a sua visão de probabilidade, a sua formalização seria mais semelhante à axiomática de Rényi (1955), pois para Pacheco d'Amorim uma probabilidade é sempre uma probabilidade condicionada. A partir desta ideia Rényi (1955) apresenta uma axiomática semelhante à de Kolmogoroff, mas o seu ponto de partida, o seu conceito primitivo, é a probabilidade condicionada. Rényi (1970), apesar de considerar a axiomática de Kolmogoroff fundamental por ter fornecido a base matemática para o desenvolvimento da Teoria da Probabilidade, justifica a necessidade de introduzir uma nova axiomática pelo facto de os fundamentos de Kolmogoroff não permitirem trabalhar com distribuições ilimitadas (medidas ilimitadas) e estas serem frequentes na Física (Mecânica Quântica) ou em processos estocásticos e cadeias de Markov. Como exemplo refere a utilização da distribuição uniforme definida em todo o espaço euclidiano, ou a escolha à sorte de um número natural onde todos os números possuem a mesma possibilidade de ocorrer. Estas situações não são permitidas pela axiomática de Kolmogoroff. Assim, o seu objectivo é utilizar a probabilidade condicionada, de forma que seja possível utilizar as distribuições referidas. Rényi (1970, p. 70) menciona que a ideia desta axiomática é do próprio Kolmogoroff, apesar deste não ter publicado nada sobre o assunto.

Enunciemos, então, a axiomática de Rényi. De forma análoga à apresentação da axiomática de Kolmogoroff, representemos por Ω o espaço dos

resultados (espaço dos acontecimentos elementares), por $\mathcal{A}$ (espaço dos acontecimentos) uma σ -álgebra gerada por subconjuntos de Ω e $\mathcal{A}'$ uma classe não vazia de conjuntos tal que $\mathcal{A}' \subset \mathcal{A}$. Consideremos uma função de conjuntos $\mathbb{P}(\mathbf{A}|\mathbf{B})$ com $\mathbf{A} \in \mathcal{A}$ e $\mathbf{B} \in \mathcal{A}'$, que é denominada por **probabilidade do acontecimento $\mathbf{A}$ condicionada ao acontecimento $\mathbf{B}$** que satisfaz os seguintes axiomas:

1. $\mathbb{P}(\mathbf{A}|\mathbf{B}) \geq 0$ e $\mathbb{P}(\mathbf{B}|\mathbf{B}) = 1$;
2. Para qualquer acontecimento fixo $\mathbf{B} \in \mathcal{A}'$, $\mathbb{P}(\mathbf{A}|\mathbf{B})$ será uma função de $\mathbf{A} \in \mathcal{A}$ que é uma medida em $(\Omega, \mathcal{A})$, pelo que, se $\{\mathbf{A}_i\}_{i=1}^{\infty}$ for uma sucessão de acontecimentos de $\mathcal{A}$ incompatíveis dois a dois ($\mathbf{A}_i \cap \mathbf{A}_j = \emptyset, \forall i \neq j$), então

$$\mathbb{P}\left(\bigcup_{i=1}^{+\infty} \mathbf{A}_i \middle| \mathbf{B}\right) = \sum_{i=1}^{+\infty} \mathbb{P}(\mathbf{A}_i|\mathbf{B}); \quad (10.70)$$

3. Sejam $\mathbf{A} \in \mathcal{A}$, $\mathbf{B} \in \mathcal{A}'$ e $\mathbf{C} \in \mathcal{A}'$ três acontecimentos tais que $\mathbf{B} \subset \mathbf{C}$ e $\mathbb{P}(\mathbf{B}|\mathbf{C}) > 0$, então

$$\mathbb{P}(\mathbf{A}|\mathbf{B}) = \frac{\mathbb{P}(\mathbf{A} \cap \mathbf{B}|\mathbf{C})}{\mathbb{P}(\mathbf{B}|\mathbf{C})}. \quad (10.71)$$

Rényi (1970) salienta que, caso se verifiquem simultaneamente os três axiomas, então $[\Omega, \mathcal{A}, \mathcal{A}', \mathbb{P}(\mathbf{A}|\mathbf{B})]$ será um **espaço de probabilidade condicional**.

Se considerarmos um espaço de probabilidade de Kolmogoroff $(\Omega, \mathcal{A}, \mathbb{P}^*)$ e representarmos por $\mathcal{A}^*$ a colecção de todos os conjuntos de $\mathcal{A}$ com probabilidade não nula ($\mathbb{P}^*(\mathbf{B}) > 0$), então $[\Omega, \mathcal{A}, \mathcal{A}^*, \mathbb{P}^*(\mathbf{A}|\mathbf{B})]$ será um espaço de probabilidade condicionado (gerado pelo espaço de Kolmogoroff $(\Omega, \mathcal{A}, \mathbb{P}^*)$) definindo

$$\mathbb{P}^*(\mathbf{A}|\mathbf{B}) = \frac{\mathbb{P}^*(\mathbf{A} \cap \mathbf{B})}{\mathbb{P}^*(\mathbf{B})}, \mathbf{A} \in \mathcal{A}, \mathbf{B} \in \mathcal{A}^*. \quad (10.72)$$

Todavia, o espaço de probabilidade condicional de Rényi não é obrigatoriamente o mesmo que um espaço de probabilidade condicionada de Kolmogoroff, pois na construção de Rényi a classe $\mathcal{A}'$ não será constituída por todos os acontecimentos de probabilidade não nula e poderá conter acontecimentos com probabilidade nula (ao contrário da construção que vimos para o espaço de probabilidade condicionada de Kolmogoroff). Desta forma teremos em geral $\mathcal{A}' \neq \mathcal{A}^*$. Rényi (1970), nos exemplos que apresenta, salienta as diferenças entre estas duas construções. Se Ω representar o espaço euclidiano n -dimensional, $\mathcal{A}$ a σ -álgebra contendo todos os conjuntos mensuráveis de Ω e f uma função mensurável não negativa definida em Ω , então podemos definir

$$\mathbb{P}(\mathbf{A}|\mathbf{B}) = \frac{\int_{\mathbf{A} \cap \mathbf{B}} f(\omega) \, d\omega}{\int_{\mathbf{B}} f(\omega) \, d\omega}. \quad (10.73)$$

Se $\int_{\Omega} f(\omega) \, d\omega < \infty$, então o espaço de probabilidade de Kolmogoroff estará bem definido, mas, se $\int_{\Omega} f(\omega) \, d\omega = \infty$, tal não acontece, pois não terá associado um espaço de probabilidade original $(\Omega, \mathcal{A}, \mathbb{P})$. Na construção de Rényi poderemos utilizar igualmente esta função, que verifica $\int_{\Omega} f(\omega) \, d\omega = \infty$, na construção do espaço de medida condicional. Por exemplo, definindo $f(\omega) = k$ para qualquer ω , com k uma constante real, obtemos a distribuição uniforme no espaço n -dimensional e, conseqüentemente,

$$\mathbb{P}(\mathbf{A}|\mathbf{B}) = \frac{\mu_n(\mathbf{A} \cap \mathbf{B})}{\mu_n(\mathbf{B})}, \quad (10.74)$$

onde μ_n representa a medida de Lebesgue n dimensional. De forma análoga, para o caso discreto, o espaço de probabilidade condicional de Kolmogoroff só pode ser gerado se $\sum_{i=1}^{\infty} p_i$ for convergente, não permitindo a definição de uma distribuição uniforme quando temos um número infinito numerável de resultados possíveis. Na construção de Rényi esta construção já é admissível, pois não é necessário recorrer ao espaço de probabilidade original $(\Omega, \mathcal{A}, \mathbb{P})$ para obtermos o espaço condicional. Rényi (1970) apresenta detalhadamente

a sua construção.

Na nossa opinião, a axiomática proposta por Rényi poderia servir de fundamentação para as ideias concebidas por Pacheco d'Amorim na sua tese de doutoramento. Por um lado, pela visão de Probabilidade de Pacheco d'Amorim, uma probabilidade deverá ser sempre condicionada, tal como na axiomática de Rényi (ao contrário da de Kolmogoroff). Em segundo lugar, a tentativa, pouco clara, de Pacheco d'Amorim (1914, p. 51–52) fundamentar a probabilidade em espaços ilimitados é efectuada no sentido de definir a distribuição uniforme (em qualquer lançamento à sorte todos os pontos possuem a mesma possibilidade) em regiões ilimitadas, principal razão apresentada por Rényi para a formação de uma nova axiomática distinta da de Kolmogoroff. Contudo, infelizmente, o apêndice onde Pacheco d'Amorim iria resolver este problema não foi efectivamente realizado, possivelmente por não ser de fácil resolução. Saliente-se, porém, que Pacheco d'Amorim, enquanto apresentava a sua proposta de construção da Teoria da Probabilidade, sentia a necessidade de justificar a probabilidade uniforme em espaços ilimitados, o que Rényi resolve 40 anos depois. Neste sentido, porventura, na formalização da concepção de Pacheco d'Amorim que efectuamos na secção 10.1 poderíamos ter recorrido aos espaços condicionais de Rényi, em substituição do recurso aos espaços de probabilidade de Kolmogoroff. Assim, poderíamos utilizar o espaço de probabilidade condicional $[\Omega, \mathcal{A}, \mathcal{A}', \mathbb{P}(\mathbf{A}|\mathbf{B})]$ para retratar o espaço inicial, onde é seleccionado à sorte o elemento x cuja escolha é caracterizada pela equipossibilidade num espaço limitado ou ilimitado, e o espaço de probabilidade condicional $[\mathbb{R}, \mathcal{B}, \mathcal{B}', \mathbb{P}_y(\mathbf{A}|\mathbf{B})]$ para o ponto imagem $y = f(x)$ que caracteriza o fenómeno que estamos a observar.

Se pretendermos modelar o mundo real com recurso à Teoria da Probabilidade, teremos de especificar completamente os três elementos de um espaço de probabilidade $(\Omega, \mathcal{A}, \mathbb{P})$, ou os quatro elementos de $[\Omega, \mathcal{A}, \mathcal{A}', \mathbb{P}(\mathbf{A}|\mathbf{B})]$, se

utilizarmos a probabilidade condicionada como primitiva, tal como Rényi e Pacheco d'Amorim. Para atingirmos este objectivo é essencial a existência de uma ligação entre os elementos do espaço de probabilidade e as observações que temos da realidade. Com este fim recorreremos muitas vezes a variáveis aleatórias, funções dos espaços primitivos, para operarmos em espaços $(\mathbb{R}, \mathcal{B}, \mathbb{P}_X)$, ou $[\mathbb{R}, \mathcal{B}, \mathcal{B}', \mathbb{P}_X(\mathbf{A}|\mathbf{B})]$, associados aos números reais. Estes espaços podem ser caracterizados por uma função real de variável real F_X denominada função de distribuição. Contudo, é necessária uma ligação entre as observações de um fenómeno aleatório e a função F_X . Esta questão não é unicamente matemática, pois é precisa uma teoria que explique como ligar o modelo probabilístico aos fenómenos reais. É esta ligação que está em questão quando debatemos as distintas interpretações de probabilidade. Por esta razão, vamos apresentar a visão frequentista de probabilidade de Richard von Mises na secção 10.3, baseada nos colectivos e em variáveis aleatórias independentes e identicamente distribuídas, e a interpretação bayesiana personalista de Bruno de Finetti, centrada no conceito de permutabilidade e nos Teoremas de Representação, na secção 10.4.

Antes de começarmos a análise das duas principais escolas de Estatística, refira-se que ao longo do século XX, existiram diversas discussões, por vezes bem extremadas, sobre a interpretação da probabilidade, seja ela frequentista, personalista, lógica ou qualquer outra. Contudo, é principalmente a partir do Problema de Behrens-Fisher⁽²³⁾ que as posições chegaram a extremos. Os problemas em inferência estatística anteriores tinham as mesmas

(23) O problema consiste em encontrar a distribuição da estatística que permita comparar a média de duas populações com distribuição gaussiana sem exigir homocedasticidade (igualdade das variâncias). A primeira proposta de solução surge nos trabalhos de Behrens em 1929. Velosa (2003) efectua uma investigação analítica detalhada deste problema sem entrar nas discussões filosóficas inerentes à interpretação de probabilidade.

soluções, independentemente da interpretação e hipóteses inerentes a cada escola. Assim, até à década de 1930, apesar da existência de diferentes concepções de probabilidade, os problemas acabavam por ter a mesma resposta. Na verdade, apesar das suas divergências, todas as interpretações convergiam para os mesmos resultados e, por conseguinte, não existiam grandes disputas entre as várias escolas. Este viver (quase) pacífico entre as diferentes escolas terminou com o surgir de soluções distintas no Problema de Behrens-Fisher, onde se concluiu que as distintas visões poderiam originar resultados diferenciados. Este problema, que fomentou a discussão entre as distintas interpretações de probabilidade, nomeadamente entre a frequentista, a bayesiana e a fiducial⁽²⁴⁾, é posterior à tese de doutoramento de Pacheco d'Amorim, tendo tornado o debate entre as diferentes escolas de probabilidade mais acesa do que nunca.

10.3 Os *Colectivos* de von Mises e a visão frequentista

“Since the first task of any scientific endeavor is a limitation of its scope, we limit our scope, roughly speaking, to the mathematical theory of repetitive events. We do not attempt to discuss in this book ideas and problems of subjective probability”.

[Richard von Mises 1964, p. 1]

A interpretação frequentista de probabilidade tem origem nos trabalhos de Venn, principalmente no seu livro *The Logic of Chance*, publicado em 1866, onde a probabilidade de um acontecimento é determinada observando

⁽²⁴⁾ Fisher (1930, 1935), indubitavelmente um dos grandes nomes da Estatística durante o século XX, introduz a (obscura) probabilidade fiducial para destacar a sua visão da interpretação bayesiana, em relação à qual tinha uma atitude muito crítica.

a frequência relativa desse acontecimento numa sucessão numerável de experiências aleatórias idênticas e independentes. Nesta perspectiva, considerando que efectuamos uma experiência aleatória n vezes onde o acontecimento $\mathbf{A}$ ocorreu $n_{\mathbf{A}}$ vezes, a sua frequência relativa $f_n(\mathbf{A})$ será o rácio $\frac{n_{\mathbf{A}}}{n}$ e, caso haja limite deste rácio quando n aumenta indefinidamente, então a probabilidade do acontecimento $\mathbf{A}$ é definida por:

$$\mathbb{P}(\mathbf{A}) = \lim_{n \rightarrow +\infty} \frac{n_{\mathbf{A}}}{n} = \lim_{n \rightarrow +\infty} f_n(\mathbf{A}). \quad (10.75)$$

Este limite não tem o significado habitual utilizado em análise matemática, que para um dado ε positivo se tenha $|f_n(\mathbf{A}) - \mathbb{P}(\mathbf{A})| < \varepsilon$ para qualquer n superior a determinado valor $n(\varepsilon)$ que depende de ε . O limite utilizado é o da convergência em probabilidade, que normalmente representamos por

$$f_n(\mathbf{A}) \xrightarrow{\mathbb{P}} \mathbb{P}(\mathbf{A}), \quad (10.76)$$

que significa que

$$\forall \varepsilon > 0 : \lim_{n \rightarrow +\infty} \mathbb{P}(|f_n(\mathbf{A}) - \mathbb{P}(\mathbf{A})| > \varepsilon) = 0, \quad (10.77)$$

ou seja, que para um dado ε positivo se tenha, para qualquer n superior a determinado valor $n(\varepsilon)$ que depende de ε , a probabilidade de ocorrer $|f_n(\mathbf{A}) - \mathbb{P}(\mathbf{A})| > \varepsilon$ na vizinhança de zero.

No entanto, a formalização desta visão começou somente com os trabalhos de Richard von Mises a partir de 1919. Refira-se que a interpretação de von Mises, que aqui expomos, não é a única visão frequencista de probabilidade. Por exemplo, Karl Raimund Popper (1902–1994) apresenta em 1959 a interpretação propensista de probabilidade, que é igualmente baseada nas frequências. Nesta visão os fenómenos aleatórios possuem uma propensão física (disposição, tendência) que a observação repetida dos fenómenos deverá evidenciar (lembrando, de certa forma, a construção de Pinto (1913) e as relações teleológicas que refere). Assim, a tendência para a estabilidade

dos resultados estatísticos, demonstrados nos teoremas limites e associados à interpretação frequencista de probabilidade, é uma manifestação das propensões que actuam como uma força que garante a estabilidade das frequências. Nesta visão, ao contrário do que é usual nas interpretações frequencistas, é possível associar uma probabilidade a uma experiência única, pela propensão de ocorrência de cada resultado.

10.3.1 A axiomática de Richard von Mises

O físico alemão Richard von Mises apresenta, a partir de 1919, uma formalização desta interpretação de probabilidade, que está na base da Estatística frequencista⁽²⁵⁾, com o recurso ao seu conceito de *colectivo*, considerando que a probabilidade só está definida em *colectivos*. Se, numa sucessão de experiências aleatórias, representarmos cada resultado do espaço-amostra por um número e considerarmos a sequência formada pelos resultados temos um *colectivo* no caso de se verificarem as seguintes duas condições:

1. *Existência de limite* — as frequências relativas de um colectivo existem, isto é, as frequências relativas f_{n_i} de cada dígito (resultado da experiência) convergem para um número p_{n_i} quando o número de experiência aumenta indefinidamente;
2. *Aleatoriedade* — as frequências relativas de qualquer sub-sequência da sequência original têm os mesmos limites p_{n_i} .

Nestas condições, p_{n_i} será, então, a probabilidade do acontecimento **A** associado ao dígito n_i . Deste modo, para von Mises, a probabilidade só existe

⁽²⁵⁾ A Estatística frequencista é habitualmente denominada por Estatística clássica, o que não se deverá confundir com a definição clássica (ou de Laplace) de probabilidade, que nada tem a ver com o assunto.

quando aplicada a *colectivos*. O primeiro postulado garante a convergência das frequências relativas, enquanto o segundo postulado (também denominado *princípio da exclusão do sistema de jogo*) garante a aleatoriedade dos resultados, sendo o conceito de *colectivo* de von Mises o fundamento da Estatística frequentista. Notemos que, ao considerarmos esta definição, as probabilidades não se aplicam a acontecimentos únicos, tais como quem vai ser o próximo campeão nacional de futebol, quem vai ganhar determinada corrida ou quem vai ganhar as próximas eleições legislativas em Portugal. Apenas se aplica a experiências aleatórias que podem ser repetidas quantas vezes sejam necessárias sob idênticas condições (tal como na construção de Pacheco d'Amorim). Contudo, não temos presentes as duas restrições da definição clássica de probabilidade, pois nesta definição o universo não tem de ser obrigatoriamente finito, nem constituído por resultados elementares equiprováveis.

10.3.2 O Cálculo das Probabilidades nos *colectivos*

“Tout le problème du calcul des probabilités consiste en ceci: Trouver la fonction de répartition dans le collectif déduit, étant données les fonctions de répartition dans les collectifs initiaux”

“la recherche des probabilités primordiales n'est ni un devoir ni un but de la théorie mathématique des probabilités.”

[Richard von Mises 1932, p. 148]

Richard von Mises⁽²⁶⁾ divide os problemas de Probabilidade em duas situações: quando temos um número de resultados finito (probabilidades aritméticas) e as probabilidades contínuas (probabilidades geométricas). Notemos que, mesmo no artigo de 1932, não se encontra a mínima referência de von

⁽²⁶⁾ A apresentação que aqui efectuamos é baseada sobretudo em von Mises (1932), obra publicada após a leccionação de uns cursos no *Institut Henri Poincaré*.

Mises ao caso infinito numerável. Utiliza, no entanto, conceitos que não estavam disponíveis em 1914. Por exemplo, define a função de distribuição (*fonction de répartition*) por $f(x) = \mathbb{P}(X \leq x)$, considerando que, nos casos em que a função $f(x)$ é derivável, então estaremos no caso da probabilidade geométrica, caso contrário teremos um número finito de pontos onde $f(x)$ cresce e estaremos no caso de probabilidade aritmética; utiliza igualmente como definição dos momentos de uma distribuição, o integral de Stieltjes

$$\mathbb{E}(X^p) = \int x^p df(x). \quad (10.78)$$

Para von Mises não é objectivo da Teoria da Probabilidade deduzir ou explicar a distribuição do colectivo original (por exemplo, através da equiprobabilidade ou de outra hipótese semelhante); esta informação deverá ser dada pelo problema, utilizando, por exemplo, considerações da Física teórica. Deste modo, o objectivo da Teoria da Probabilidade é deduzir as distribuições de um outro colectivo obtido através do inicial, considerando que, para tal, existem apenas quatro operações elementares (e combinações delas) para deduzir um novo colectivo através dos colectivos dados. O número de operações simples a efectuar pode ser finito (problema finito) ou infinito (problema infinito numerável), onde teremos, por exemplo, a Lei dos Grandes Números ou o Teorema Limite Central. Vamos, nesta secção, definir as quatro operações elementares de Richard von Mises.

Sejam $E = e_1, e_2, e_3, \dots$ os elementos do colectivo $\mathcal{C}$ (as experiências) e $X = x_1, x_2, x_3, \dots$ os correspondentes resultados (*caractère distinctif*).

Definição 10.7 (Primeira Operação: **Seleccção**). Se estabelecermos uma série $\alpha_1, \alpha_2, \alpha_3, \dots$ de números naturais crescentes e considerarmos os elementos $E' = e_{\alpha_1}, e_{\alpha_2}, e_{\alpha_3}, \dots$ com os seus resultados $X' = x_{\alpha_1}, x_{\alpha_2}, x_{\alpha_3}, \dots$ tem-se um novo colectivo $\mathcal{C}'$ caracterizado pela mesma distribuição.

Assim, a primeira operação, onde estamos a seleccionar um sub-sequência do primeiro colectivo, pelo segundo axioma, não altera a distribuição do colectivo.

Definição 10.8 (Segunda Operação: **Mistura**). Se em todos os elementos $e_1, e_2, e_3, \dots$ do colectivo dado $\mathcal{C}$ associarmos aos resultados $x_1, x_2, x_3, \dots$ novos caracteres $x'_1, x'_2, x'_3, \dots$, onde a cada resultado novo x'_j é associado um (ou mais) resultados do colectivo inicial $x_{j_1}, x_{j_2}, \dots$, obtém-se um novo colectivo $\mathcal{C}'$ onde $p'_j = \sum_i p_{j_i}$ (p' representam as probabilidades no novo colectivo e p as probabilidades no colectivo inicial). Esta operação corresponde à **Lei da Adição**.

Com esta nova operação podemos facilmente deduzir o Teorema das Probabilidades Totais ou Lei da Adição das Probabilidades, não dependendo a dedução de qualquer suposição de equiprobabilidade. Pelo contrário, através desta operação podemos deduzir a definição clássica de probabilidades para o caso de equiprobabilidade no colectivo inicial, pois

$$p'_j = \sum_i p_{j_i} = \frac{m'}{m}, \quad (10.79)$$

obtendo-se o número de casos favoráveis a dividir pelo número de casos possíveis, sendo então um teorema e não uma definição.

Definição 10.9 (Terceira Operação: **Partição**). Se aos elementos E do colectivo $\mathcal{C}$ retirarmos todos os elementos com um determinado resultado, obtém-se um novo colectivo $\mathcal{C}'$ onde $p'_j = \frac{p_j}{P}$ sendo P a probabilidade dos resultados não retirados no colectivo inicial, que corresponde à **Lei da Divisão**.

Com esta operação Richard von Mises define a probabilidade do novo colectivo $\mathcal{C}'$, obtido através do inicial $\mathcal{C}$, retirando alguns resultados. Desta

forma define a **Lei da Divisão** e obtém o conceito de probabilidade condicionada. A seguinte operação permite obter um novo colectivo $\mathcal{C}$ através de dois colectivos dados, $\mathcal{C}'$ e $\mathcal{C}''$.

Definição 10.10 (Quarta Operação: **Composição**). Sejam $E' = e'_1, e'_2, e'_3, \dots$ os elementos do colectivo $\mathcal{C}'$ e $X' = x'_1, x'_2, x'_3, \dots$ os correspondentes resultados e sejam $E'' = e''_1, e''_2, e''_3, \dots$ os elementos do colectivo $\mathcal{C}''$ e $X'' = x''_1, x''_2, x''_3, \dots$ os correspondentes resultados. O elemento e_1 do novo colectivo $\mathcal{C}$ é o ponto com coordenadas x'_1 e x''_1 num espaço a duas dimensões⁽²⁷⁾. Desta forma $E = e_1, e_2, e_3, \dots$ tem como resultados $X = x'_1 x''_1, x'_2 x''_2, x'_3 x''_3, \dots$. A probabilidade de uma combinação qualquer de x' e x'' é o produto da probabilidade de x' no colectivo $\mathcal{C}'$ e de x'' no colectivo $\mathcal{C}''$, o que corresponderá à **Lei da Multiplicação**. Nesta operação o autor salienta que não são necessariamente associados os mesmos elementos a cada resultado, isto é, podemos ter um colectivo $\mathcal{C}$ com n resultados x_i distintos com probabilidades p_i ($i = 1, \dots, n$) e a cada um destes resultados associar um novo colectivo $\mathcal{C}_i$ que não será igual para todos os x_i do colectivo inicial $\mathcal{C}$. Assim, nestas composições, nem sempre a composição é efectuada de igual forma para todos os resultados de um colectivo (nem sempre os colectivos são independentes).

Analisemos estas operações com um exemplo retirado de von Mises (1932), baseado na correspondência de 1654 entre Fermat e Pascal.

Problema 10.1. Qual a probabilidade de, em quatro lançamentos de um dado, sair pelo menos uma vez a face 6? E qual a probabilidade de, em 24 lançamentos de dois dados, obter pelo menos um doze?

SOLUÇÃO: Quando lançamos repetidamente um dado, temos um colectivo $\mathcal{C}$ com seis resultados $x_1, x_2, \dots, x_6$ possíveis distintos, onde x_i re-

⁽²⁷⁾ Ou, de forma geral, se os colectivos iniciais tiverem dimensão r' e r'' , o novo colectivo terá dimensão $r' + r''$.

apresenta a saída da face i . Como foi referido, esta teoria não justifica as probabilidades do colectivo inicial (nem pretende fazê-lo). Consideremos, então, seis probabilidades associadas $p_1, p_2, \dots, p_6$, tal que a sua soma seja igual à unidade. Podemos agora utilizar a operação de selecção e escolhemos quatro colectivos $\mathcal{C}_1, \mathcal{C}_2, \mathcal{C}_3, \mathcal{C}_4$ de $\mathcal{C}$, onde os elementos do colectivo $\mathcal{C}_i$ serão $e_i, e_{i+4}, \dots, e_{4n+i}, \dots$ de $\mathcal{C}$. Estes novos colectivos têm as mesmas probabilidades que o colectivo inicial $\mathcal{C}$. Depois aplicamos a operação de mistura aos quatro novos colectivos de forma a associarmos os resultados $x_1, x_2, \dots, x_5$ e obter unicamente dois resultados possíveis em cada um dos novos colectivos $\mathcal{C}'_1, \mathcal{C}'_2, \mathcal{C}'_3, \mathcal{C}'_4$, correspondentes a $x'_1 = \{x_1, x_2, x_3, x_4, x_5\}$ e $x'_2 = x_6$. Consequentemente, teremos $p'_1 = \sum_{i=1}^5 p_i$ e $p'_6 = p_6$. Finalmente aplicamos a propriedade da composição entre dois colectivos, primeiro entre $\mathcal{C}'_1$ e $\mathcal{C}'_2$, depois entre o resultante e $\mathcal{C}'_3$ e, por fim, entre o resultante da última composição e $\mathcal{C}'_4$. Com estas três composições vamos obter, uma vez que cada um dos colectivos componentes possui dois resultados distintos, 2^4 resultados possíveis correspondentes a pontos em quatro dimensões constituídos por x'_1 e x'_2 , sendo a probabilidade de cada ponto obtido pela multiplicação das probabilidades de cada uma das suas coordenadas. Uma vez que pretendemos todos os pontos menos o de coordenadas $\{x'_1, x'_1, x'_1, x'_1\}$ que tem probabilidade $p_1'^4 = (1 - p_6)^4$, a probabilidade procurada será dada por $\mathbb{P}_1 = 1 - (1 - p_6)^4$.

Richard von Mises salienta que, nesta análise, nunca foi invocado o princípio da razão insuficiente ou a equiprobabilidade. Contudo, se considerarmos que o dado lançado é um dado justo e, como tal, terá igual probabilidade de ocorrer cada uma das faces, teremos $p_i = \frac{1}{6}$ e $\mathbb{P}_1 = \frac{671}{1296} \approx 0.51775$.

Na resposta à segunda pergunta podemos aplicar raciocínio análogo. Neste caso principiamos a análise com dois colectivos $\mathcal{C}_1$ e $\mathcal{C}_2$, cada um com seis resultados possíveis distintos, sendo p_i a probabilidade de saída da face i no primeiro dado e q_i a probabilidade de saída da face i no segundo dado

(nada nos garante que as probabilidades associadas a estes dois colectivos sejam idênticas). Com a composição destes colectivos iríamos obter um novo colectivo $\mathcal{C}^*$ com duas coordenadas, cada uma referente a um dos dados. Uma vez que pretendemos analisar, em cada lançamento, se sai a totalidade de doze pontos, podemos aplicar a operação de mistura, definindo-se dois resultados distintos, quando saiem 12 pontos (em ambos os dados sai a face 6 virada para cima), que representamos por x_1^* , que terá probabilidade $p_1^* = p_6 q_6$, e os restantes casos são agrupados em x_0^* , que terá probabilidade $p_0^* = 1 - p_6 q_6$. Finalmente, aplicando a propriedade de selecção, obtemos 24 colectivos $\mathcal{C}_i^*$, com $i = 1, \dots, 24$, aos quais aplicamos a operação de composição para obtermos o colectivo $\mathcal{C}^{**}$ constituído por 2^{24} pontos com 24 coordenadas. Pela Lei da Multiplicação teremos $(1 - p_6 q_6)^{24}$ como probabilidade de não saírem doze pontos em qualquer lançamento e, consequentemente, a probabilidade pretendida será dada por $\mathbb{P}_2 = 1 - (1 - p_6 q_6)^{24}$. Se considerarmos dois dados equilibrados, com probabilidade igual a $\frac{1}{6}$ para cada face, teremos $\mathbb{P}_2 = 0.4914$. $\diamond$

Problema 10.2 (Problema das urnas). Consideremos m urnas distintas com bolas brancas e bolas pretas, sendo p_i a probabilidade de se retirar uma bola branca da urna i e $q_i = 1 - p_i$ a probabilidade de ser retirada uma bola preta da urna i . Se escolhermos aleatoriamente uma urna e dessa urna tirarmos uma bola que é branca, qual a probabilidade de a bola ter sido retirada da urna número i ?

SOLUÇÃO: Consideremos m colectivos $\mathcal{C}_i$, cada um com dois resultados possíveis, com probabilidades, respectivamente, p_i e q_i . Seja $\mathcal{C}$ um outro colectivo, associado à escolha da urna, com m resultados com probabilidades p'_j tais que $\sum_{j=1}^m p'_j = 1$. Desta forma teremos $m + 1$ colectivos iniciais. Começemos por determinar a probabilidade de, escolhendo à sorte uma urna e tirando à sorte uma bola, esta ser branca. Em primeiro lugar aplica-se a

operação de composição entre os m colectivos $\mathcal{C}_i$ e o colectivo $\mathcal{C}$, obtendo-se a probabilidade de ser obtida bola branca na urna i , que é dada por $p_i p'_i$. O colectivo obtido pela composição terá $2m$ resultados distintos definidos por $2m$ pontos com duas dimensões. Notemos que, se recorrermos à operação de mistura, deduzimos que a probabilidade de obtermos a bola branca é dada por $\mathbb{P}_{\mathbf{B}} = \sum_{i=1}^m p_i p'_i$ (que corresponde ao Teorema das Probabilidades Totais). Para obtermos a probabilidade pretendida, recorreremos à operação de partição, retirando os pontos que têm bola preta e, pela Lei da Divisão, a probabilidade de ser escolhida a urna i será dada por

$$\mathbb{P} = \frac{p_i p'_i}{p_1 p'_1 + p_2 p'_2 + \cdots + p_m p'_m}. \quad (10.80)$$

◇

Richard von Mises deduz assim a fórmula de Bayes, salientando que este resultado é independente da suposição de equiprobabilidade. Recordemos que Pacheco d'Amorim afirma que apresenta a primeira demonstração rigorosa do Teorema de Bayes, pelo facto de as demonstrações habituais serem apresentadas em obras onde a probabilidade só está definida para as situações onde se verifica a equiprobabilidade (definição clássica de probabilidade). Desta forma, Richard von Mises, ao salientar a independência deste resultado da suposição da existência de casos igualmente possíveis, parece ter a mesma posição de Pacheco d'Amorim ao discordar das tradicionais demonstrações deste resultado.

Salientemos, pelo descrito, que von Mises principia a sua construção com um colectivo com distribuição conhecida, sem recurso à Teoria da Probabilidade, pois, como foi assinalado, não pretende justificar a distribuição dos colectivos iniciais, e transforma-o, através das operações que concebe, noutro colectivo cuja densidade depende da densidade dos elementos do colectivo primitivo. Pacheco d'Amorim considera que o espaço primitivo é sempre proveniente de uma escolha aleatória, caracterizada por uma distribuição

uniforme, podendo estas escolhas dar origem a outras distribuições através do seu conceito de ponto imagem. Assim, as características que analisamos em fenómenos aleatórios muitas vezes não correspondem a pontos livres, mas antes a imagens desses pontos e, como tal, a sua distribuição depende da função de transformação e da região (classe) inicial. Contudo, nas aplicações, quando pretendemos inferir a distribuição de uma característica de um fenómeno aleatório, nenhum dos dois autores se preocupa com a distribuição primitiva, do colectivo ou escolha inicial, unicamente com a distribuição que dá origem aos dados observados.

10.3.3 Os Teoremas Fundamentais

Richard von Mises demonstra os teoremas limite recorrendo a conceitos mais actuais, que Pacheco d'Amorim não tinha disponíveis, tais com a função característica (que já tinha sido utilizada antes da tese de doutoramento de Pacheco d'Amorim, mas ainda não tinha tido os desenvolvimentos que decorreram durante a década de 1920, nomeadamente por Paul Lévy (1925)), que denomina por transformada de Laplace, ou a operação de convolução, que denomina por soma de dois colectivos. Deste modo von Mises deduz o Primeiro Teorema Fundamental do Cálculo de Probabilidades (que corresponde ao Teorema Limite Central) com referência ao trabalho de Lindeberg.

Para von Mises as Leis dos Grandes Números são uma consequência do Primeiro Teorema Fundamental, sendo unicamente um caso particular. A primeira Lei dos Grandes Números diz que a probabilidade de que a média aritmética da soma de n variáveis, cada uma com determinada distribuição, seja situada dentro de limites estreitos numa vizinhança da sua esperança matemática, tende para a unidade quando n aumenta indefinidamente, sendo suficiente supor que a soma dos desvios (variâncias) das n distribuições particulares cresce para infinito com n . A segunda Lei dos Grandes Números,

também consequência do Primeiro Teorema Fundamental, considera que, se os primeiros n elementos observados de um colectivo têm média aritmética dos seus resultados x , então a probabilidade, de que a esperança matemática do resultado esteja situada dentro de limites estreitos numa vizinhança de x , tende para a unidade e, para n suficientemente grande, a influência da probabilidade *a priori* vai desaparecendo.

O segundo Teorema Fundamental enunciado por Richard von Mises refere-se às conclusões *a posteriori*, isto é, às conclusões que poderemos inferir após a observação de n repetições de uma experiência. Desta forma, para n infinito, teremos o Segundo Teorema Fundamental que corresponde a uma generalização da fórmula de Bayes. Nesta dedução von Mises conclui que a probabilidade *a priori* desaparece quando determinamos o limite de n a aumentar indefinidamente, isto é, a probabilidade *a posteriori* unicamente depende das frequências relativas verificadas nas n experiências observadas. Assim, para von Mises, ao contrário das visões bayesianas, não é necessário utilizarmos uma probabilidade *a priori* para podermos inferir sobre as probabilidades de um colectivo.

Desta forma, para Richard von Mises, a ligação entre a probabilidade e o mundo real (onde considera que existem três tipos de problemas: as questões dos jogos de azar, a estatística geral e os seguros e as questões colocadas pela Física moderna) é unicamente possível através da observação repetida da mesma experiência e recorrendo aos resultados limites, nomeadamente ao Teorema Limite Central. Desta forma a sua visão é, nas ligações entre a probabilidade e o mundo real, muito semelhante à proposta de Pacheco d'Amorim.

10.3.4 Independência *versus* Estatística clássica

Se considerarmos uma amostra aleatória como uma sequência retirada de um *colectivo*, a primeira condição da axiomática de von Mises garante a homogeneidade distribucional (que as variáveis aleatórias que compõem a amostra tenham a mesma distribuição marginal) e a segunda condição garante a aleatoriedade dos resultados (independência entre as variáveis). Assim, nesta visão de von Mises, numa amostra as variáveis aleatórias são independentes e identicamente distribuídas, o que torna o tratamento matemático da função densidade conjunta trivial, pois pela independência

$$f(x_1, x_2, \dots, x_n | \theta) = \prod_{i=1}^n f_i(x_i | \theta), \quad (10.81)$$

e, como pela homogeneidade distribucional se tem $f_i(x_i | \theta) = f(x_i | \theta)$, podemos concluir que

$$\prod_{i=1}^n f_i(x_i | \theta) = \prod_{i=1}^n f(x_i | \theta) = f^n(x_i | \theta), \quad (10.82)$$

com $\theta \in \Theta$. Saliente-se que este modelo de probabilidade representa o modelo mais elementar, pois as hipóteses subjacentes, independência e homogeneidade distribucional, simplificam bastante a sua análise. Na Estatística clássica a base são as distribuições amostrais, as distribuições de uma função da amostra aleatória — $h(X_1, X_2, \dots, X_n)$. Os estimadores, estatísticas de testes e previsores são funções $h(X_1, X_2, \dots, X_n)$ da amostra aleatória $(X_1, X_2, \dots, X_n)$. Desta forma correspondem também a variáveis aleatórias. É fundamental, em toda a inferência na Estatística clássica, determinar estas distribuições amostrais. Para tal, a hipótese de que as variáveis aleatórias $(X_1, X_2, \dots, X_n)$ são independentes e identicamente distribuídas, consequência dos dois axiomas de Richard von Mises, é primordial.

Salientemos que, para Pacheco d'Amorim, as hipóteses de independência e homogeneidade distribucional também são fundamentais para a análise de

qualquer fenómeno. Assim, a base das aplicações de Pacheco d'Amorim é semelhante às hipóteses utilizadas por Richard von Mises na sua axiomática. No entanto, como destacamos anteriormente, é essencial nesta visão a dedução das distribuições das funções das observações $h(X_1, X_2, \dots, X_n)$ que são utilizadas para analisar o fenómeno. Pacheco d'Amorim, apesar de apresentar a ideia filosófica que está na construção dos testes de significância, não formaliza a sua construção, provavelmente porque, para alcançar tal objectivo, teria de recorrer a resultados sobre distribuições que não apresenta na sua obra.

Deste modo, quer Pacheco d'Amorim quer von Mises consideram que unicamente poderemos inferir sobre a distribuição da característica em estudo através da observação repetida dos fenómenos em situações idênticas, isto é, se considerarmos uma amostra aleatória $(X_1, \dots, X_n)$ constituída por n variáveis aleatórias independentes e identicamente distribuídas. Ambos os autores consideram ainda que para efectuar inferência não é necessário recorrer a nenhuma distribuição *a priori* dos parâmetros em análise. Por estas razões consideramos que as ideias de Diogo Pacheco d'Amorim implícitas na construção do Cálculo das Probabilidades e suas aplicações são, em muitos aspectos, semelhantes às de Richard von Mises.

10.4 A *Permutabilidade* de Bruno de Finetti e a visão bayesiana

“My thesis, paradoxically, and a little provocatively, but nonetheless genuinely, is simply this

PROBABILITY DOES NOT EXIST.

(...) probability too, if regarded as something endowed with some kind of objective existence, is no less a misleading misconception, an illusory attempt to exteriorize or materialize our true probabilistic beliefs."

[Bruno de Finetti 1974, p. x]

Na visão bayesiana personalista (ou, mais frequentemente, subjectiva) a probabilidade de uma certa hipótese ou acontecimento é vista como o nosso grau de credibilidade (*personal degrees of believe*) relativamente a essa hipótese ou proposição. Desta forma, na interpretação bayesiana as probabilidades são habitualmente aplicáveis a proposições e não a acontecimentos definidos numa σ -álgebra como na axiomática de Kolmogoroff. Nesta perspectiva não é obrigatório que haja apenas uma probabilidade para determinado acontecimento, podendo esta variar de indivíduo para indivíduo, isto é, perante a mesma quantidade de informação, indivíduos diferentes podem associar diferentes graus de credibilidade à mesma proposição⁽²⁸⁾, pois é possível associar qualquer grau a uma proposição, desde que os graus de credibilidade obedeçam a determinados critérios de consistência, tais como o princípio da coerência (ausência do *dutch book* onde há perca certa por parte de um dos apostadores). Por exemplo, se numa experiência aleatória com três possíveis resultados distintos (onde só ocorre um) a aposta de uma unidade monetária em cada acontecimento originar um ganho de duas unidades monetárias, então estamos perante uma situação onde pode ocorrer *dutch book*, pois, se apostarmos uma unidade em cada um dos três resultados, iremos ganhar duas unidades monetárias, seja qual for o resultado da experiência, e assim decerto perderemos dinheiro. Para a maioria dos subjectivistas, os graus de credibilidade são determinados pela disposição de um indivíduo para fazer

⁽²⁸⁾ Apesar de, em rigor, a informação disponível ser sempre diferente de indivíduo para indivíduo, pois todos possuímos experiências de vida distintas.

determinadas escolhas — *behavior criteria*, pelo que as probabilidades são determinadas pelos comportamentos em situações onde existe incerteza.

Diversos autores apresentaram uma visão subjectiva de probabilidade, tais como Frank Plumton Ramsey (1926)⁽²⁹⁾, Bruno de Finetti (1937), Savage (1954), entre muitos outros. Para Bruno de Finetti (1906 – 1985)⁽³⁰⁾, provavelmente o expoente máximo desta interpretação, a probabilidade é obrigatoriamente subjectiva, correspondendo ao grau de credibilidade do sujeito i , logo, condicionada à informação que o sujeito i possui, na ocorrência do acontecimento $\mathbf{A}$, que representaremos por $\mathbb{P}_i(\mathbf{A})$. Normalmente é omitido o índice i referente ao sujeito, utilizando-se $\mathbb{P}(\mathbf{A})$ em vez de $\mathbb{P}_i(\mathbf{A})$. Todavia, como referimos, esta visão considera que não existe nenhum argumento que garanta que todos os indivíduos, que tenham a mesma informação, atribuam as mesmas probabilidades, podendo assim haver $\mathbb{P}_i(\mathbf{A}) \neq \mathbb{P}_j(\mathbf{A})$ para $i \neq j$ (desde que, para cada um destes indivíduos, as suas probabilidades satisfaçam o princípio da coerência). Por esta razão a sua famosa frase —

⁽²⁹⁾ Para Frank Ramsey (1903–1930) o grau de credibilidade de uma relação é sempre mensurável, sendo os graus de credibilidade medidos pelos comportamentos dos agentes — *behavior criteria* — “*in many cases, I think, our judgment about the strength of our belief is really about how we should act in hypothetical circumstances*”. Para Ramsey, ao contrário de Bruno de Finetti, a utilização de apostas em dinheiro para medir os graus de credibilidade de uma pessoa é incorrecta — “*It is inexact partly because of the diminishing marginal utility of money, partly because the person may have a special eagerness or reluctance to bet, because he either enjoys or dislikes excitement or for any other reason, e.g. to make a book*”. Desta forma defende as apostas em *goods* e *bads* (utilidade) em vez de dinheiro. No entanto, apoia-se no conceito de *apostas* para quantificar uma probabilidade e introduz a noção de *consistência* (alguns autores usam o termo *coerência*), considerando que os graus de credibilidade de cada indivíduo devem ser coerentes, isto é, garantir a inexistência de contradições entre eles (obedecem, por exemplo, à propriedade transitiva).

⁽³⁰⁾ Cifarelli & Regazzini (1996) e Bernardo (1998) apresentam a vida e obra de Bruno de Finetti.

Probability does not exist — acaba por afirmar, de uma forma clara, que a probabilidade não tem uma existência física, não tem obrigatoriamente um valor objectivo, mas tão somente uma existência subjectiva, que depende do indivíduo. A visão de Bruno de Finetti é a base da Estatística bayesiana, nomeadamente o seu conceito de permutabilidade (*exchangeability*), que corresponde a acontecimentos que ocorrem numa sequência cuja ordem não afecta as suas probabilidades. Este conceito desempenha um papel capital nesta teoria, considerando que numa sequência de provas a ordem dos sucessos é irrelevante para a determinação das probabilidades.

10.4.1 Axiomática de Bruno de Finetti

Para Bruno de Finetti (1937) a probabilidade é um conceito primitivo que guia o comportamento humano na tomada de decisão sob incerteza. A teoria criada por de Finetti pode ser dividida em duas partes. Na primeira parte, referente à teoria axiomática da probabilidade qualitativa, de Finetti apresenta as regras (axiomas) a que a probabilidade deverá obedecer, utilizando os conceitos primitivos $\succ$ que representa “é mais provável que”, $\succeq$ “é pelo menos tão provável como” e $\cong$ que significa “é tão provável como”. Os axiomas propostos por de Finetti, baseados nestes conceitos primitivos, são cinco:

1. Para quaisquer acontecimentos $\mathbf{E}$ e $\mathbf{E}'$ verifica-se $\mathbf{E} \succeq \mathbf{E}'$ ou $\mathbf{E}' \succeq \mathbf{E}$ (se se verificarem ambas, isto é $\mathbf{E} \succeq \mathbf{E}'$ e $\mathbf{E}' \succeq \mathbf{E}$, então $\mathbf{E} \cong \mathbf{E}'$);
2. Se $\mathbf{E}$ não é o acontecimento certo nem o impossível então $\Omega \succ \mathbf{E} \succ \emptyset$;
3. $\succeq$ é transitivo, ou seja, se $\mathbf{E} \succeq \mathbf{E}'$ e $\mathbf{E}' \succeq \mathbf{E}''$ então $\mathbf{E} \succeq \mathbf{E}''$;
4. Se $\mathbf{E}_1$ e $\mathbf{E}_2$ são ambos incompatíveis com $\mathbf{E}$, $\mathbf{E} + \mathbf{E}_1 \succeq \mathbf{E} + \mathbf{E}_2$ se e só se $\mathbf{E}_1 \succeq \mathbf{E}_2$ e $\mathbf{E} + \mathbf{E}_1 \cong \mathbf{E} + \mathbf{E}_2$ se e só se $\mathbf{E}_1 \cong \mathbf{E}_2$;

5. Se $\mathbf{E}' \subset \mathbf{E}$ e $\mathbf{E}'' \subset \mathbf{E}$, então $\mathbf{E}' \succeq \mathbf{E}''$ se e só se $\mathbf{E}'|\mathbf{E} \succeq \mathbf{E}''|\mathbf{E}$, onde $\mathbf{E}'|\mathbf{E}$ representa $\mathbf{E}'$ condicionado a $\mathbf{E}$.

Desta forma Bruno de Finetti apresenta as operações que poderemos invocar para determinar as probabilidades. Notemos que, ao contrário de Kolmogoroff, Bruno de Finetti não se preocupa, nesta concepção, com probabilidades de uniões numeráveis.

Na segunda parte, de análise quantitativa, de Finetti define como medir uma probabilidade subjectiva, isto é, como transformar os nossos graus de incerteza na determinação de um número. Esta análise é feita através de apostas coerentes (*coherent bets*), onde não é permitido a um adversário que aposte contra nós a possibilidade de ganhar sempre (ganhar seja qual for o resultado). Refira-se que é através da coerência que Bruno de Finetti deduz os teoremas fundamentais do Cálculo das Probabilidades, tais como o Teorema da Probabilidade Total e o Teorema da Probabilidade Composta. É nesta ideia que vamos centrar a nossa atenção na próxima secção.

10.4.2 Definição de probabilidade e princípio da coerência

Para Bruno de Finetti toda a incerteza deve ser descrita através de probabilidades, podendo estas ser definidas pelo comportamento do indivíduo em situações de apostas. Deste modo, para de Finetti a probabilidade de uma proposição é definida da forma que passamos a expor.

Definição 10.11 (Probabilidade). Representando por $\mathbb{P}_i(\mathbf{A})$ a probabilidade que o indivíduo i atribui à proposição $\mathbf{A}$, então $S \times \mathbb{P}_i(\mathbf{A})$ é o preço máximo que o indivíduo i está disposto a apostar na ocorrência do acontecimento $\mathbf{A}$, onde pode ganhar um prémio de S unidades monetárias se, e somente se, ocorrer $\mathbf{A}$ (caso contrário perderá as $S \times \mathbb{P}_i(\mathbf{A})$ monetárias que apostou).

Particularizando esta definição para S igual a uma unidade monetária, a probabilidade $\mathbb{P}_i(\mathbf{A})$ é o preço máximo que o indivíduo i está disposto a apostar na ocorrência da hipótese $\mathbf{A}$, onde pode ganhar um prêmio unitário se, e somente se, ocorrer $\mathbf{A}$. Desta forma, caso $\mathbf{A}$ ocorra, o apostador ganhará $1 - \mathbb{P}_i(\mathbf{A})$ unidades monetárias, caso contrário perderá $\mathbb{P}_i(\mathbf{A})$ unidades monetárias⁽³¹⁾, razão pela qual se denomina por *vantagens* (*odds*) da proposição $\mathbf{A}$ ao quociente

$$\pi_i(\mathbf{A}) = \frac{\mathbb{P}_i(\mathbf{A})}{\mathbb{P}_i(\overline{\mathbf{A}})} = \frac{\mathbb{P}_i(\mathbf{A})}{1 - \mathbb{P}_i(\mathbf{A})}. \quad (10.83)$$

Deste modo, as *vantagens* ou *odds* do acontecimento $\mathbf{A}$ são definidas pelo quociente entre a probabilidade do acontecimento $\mathbf{A}$ e a probabilidade do seu complementar. Esta quantidade também permite medir o grau de incerteza num determinado acontecimento; por exemplo, se as *vantagens* do acontecimento $\mathbf{A}$ são iguais a $\frac{1}{3}$ significa, na interpretação bayesiana, que o grau de credibilidade do apostador em relação a o acontecimento $\mathbf{A}$ ocorrer é três vezes inferior em relação ao seu grau de credibilidade de que o acontecimento $\mathbf{A}$ não ocorrerá. De uma forma geral, se as vantagens a favor da proposição $\mathbf{A}$ são $\frac{p}{q}$, isto significa que a probabilidade associada à proposição $\mathbf{A}$ é p . Numa aposta, quando afirmamos que temos *vantagens* de $\frac{1}{3}$ de ocorrer $\mathbf{A}$, significa que estamos dispostos a apostar uma unidade monetária na ocorrência de $\mathbf{A}$, se ganharmos 3 unidades monetárias no caso de este acontecimento ocorrer (recebemos 4 unidades monetárias uma vez que pagamos uma unidade para entrar no jogo).

Para este autor, o princípio da coerência, que é definido pela ausência do *dutch book*, é suficiente para garantir todas as propriedades do Cálculo das Probabilidades. O *dutch book* (ou banca holandesa) consiste em, numa aposta, não haver perda garantida, isto é, a aposta ser feita de tal forma

⁽³¹⁾ Notemos que, conseqüentemente, o valor esperado do ganho nesta aposta é nulo.

que, seja qual for o resultado, ou o apostador ou a banca (que corresponde ao outro apostador) perca dinheiro. Deste modo, o princípio da coerência garante que numa sequência de apostas não exista um ganho garantido para a banca.

“C’est précisément cette condition de cohérence qui constitue le seul principe d’où l’on puisse déduire tout le calcul des probabilités: ce calcul apparaît alors comme l’ensemble des règles auxquelles l’évaluation subjective des probabilités de divers événements par un même individu doit être assujettie si l’on ne veut pas qu’il y ait entre elles une contradiction fondamentale.”

[Bruno de Finetti 1937, p. 7]

Vamos apresentar algumas propriedades que ilustram esta ideia de Bruno de Finetti, segundo as quais este princípio é suficiente para demonstrar as principais características da probabilidade.

Teorema 10.10. *Verifica-se o princípio da coerência se e só se*

$$0 \leq \mathbb{P}_i(\mathbf{A}) \leq 1. \quad (10.84)$$

Demonstração. Se apostar uma unidade na ocorrência de $\mathbf{A}$, cuja probabilidade representamos por $\mathbb{P}_i(\mathbf{A}) = p_i$, então ganharei $1 - p_i$ unidades monetárias se $\mathbf{A}$ ocorrer e ganharei $-p_i$ unidades monetárias se $\mathbf{A}$ não ocorrer (isto é, perderei p_i unidades monetárias). Para que não se verifique o *dutch book*, estas quantidades, $1 - p_i$ e $-p_i$, não podem assumir o mesmo sinal. Desta forma, se $p_i < 0$ teremos $1 - p_i > 0$ e $-p_i > 0$, logo, não se verificará o princípio da coerência uma vez que, se $\mathbf{A}_i$ ocorrer, ganho $1 - p_i$ unidades monetárias e, se $\mathbf{A}_i$ não ocorrer, ganho $|p_i|$ unidades monetárias, logo, existe perda certa para a banca. Se considerarmos $p_i > 1$, então $1 - p_i < 0$ e $-p_i < 0$, logo, verificar-se-á o *dutch book*, uma vez que, se $\mathbf{A}_i$ ocorrer, perco

$p_i - 1$ unidades monetárias e, se $\mathbf{A}_i$ não ocorrer, perco p_i unidades monetárias, logo, terei perda certa se apostar. Assim concluímos que unicamente não existirá perda certa se $0 \leq p_i \leq 1$, pois, se $\mathbf{A}_i$ ocorrer, ganho $1 - p_i > 0$ e, se $\mathbf{A}_i$ não ocorrer, ganho $-p_i < 0$ (isto é, perco p_i). $\square$

Teorema 10.11. *Verifica-se o princípio da coerência se e só se*

$$\mathbb{P}_i(\Omega) = 1. \quad (10.85)$$

Demonstração. Se apostarmos em Ω , como Ω ocorre sempre, ganhamos sempre. Então, para não haver perda certa para nós nem para a banca, o ganho tem que ser nulo, logo $\mathbb{P}_i(\Omega) - 1 = 0$, ou seja, $\mathbb{P}_i(\Omega) = 1$ (notemos que é a única situação em que $1 - p_i$ e $-p_i$ não são ambos positivos ou ambos negativos). $\square$

Teorema 10.12 (Aditividade). *Verifica-se o princípio da coerência se e só se*

$$\forall \mathbf{A}, \mathbf{B} \in \Omega : \mathbf{A} \cap \mathbf{B} = \emptyset \Rightarrow \mathbb{P}_i(\mathbf{A} \cup \mathbf{B}) = \mathbb{P}_i(\mathbf{A}) + \mathbb{P}_i(\mathbf{B}). \quad (10.86)$$

Demonstração. Consideremos apenas três acontecimentos do sistema de apostas: os acontecimentos $\mathbf{A}$ e $\mathbf{B}$, com $\mathbf{A} \cap \mathbf{B} = \emptyset$ e o acontecimento $\mathbf{C} = \mathbf{A} \cup \mathbf{B}$. Assim, podemos apostar $S_{\mathbf{A}}p_{\mathbf{A}}$ em $\mathbf{A}$, $S_{\mathbf{B}}p_{\mathbf{B}}$ em $\mathbf{B}$ e $S_{\mathbf{C}}p_{\mathbf{C}}$ em $\mathbf{C}$. Temos então as três seguintes possibilidades de ocorrência:

Ocorrência	Ganhos
$\mathbf{A} \cap \overline{\mathbf{B}}$	$G_{\mathbf{A} \cap \overline{\mathbf{B}}} = S_{\mathbf{A}} + S_{\mathbf{C}} - (S_{\mathbf{A}}p_{\mathbf{A}} + S_{\mathbf{B}}p_{\mathbf{B}} + S_{\mathbf{C}}p_{\mathbf{C}})$
$\overline{\mathbf{A}} \cap \mathbf{B}$	$G_{\overline{\mathbf{A}} \cap \mathbf{B}} = S_{\mathbf{B}} + S_{\mathbf{C}} - (S_{\mathbf{A}}p_{\mathbf{A}} + S_{\mathbf{B}}p_{\mathbf{B}} + S_{\mathbf{C}}p_{\mathbf{C}})$
$\overline{\mathbf{A}} \cap \overline{\mathbf{B}}$	$G_{\overline{\mathbf{A}} \cap \overline{\mathbf{B}}} = -(S_{\mathbf{A}}p_{\mathbf{A}} + S_{\mathbf{B}}p_{\mathbf{B}} + S_{\mathbf{C}}p_{\mathbf{C}})$

A única forma de não haver perda certa é a de os ganhos $G_{\mathbf{A} \cap \overline{\mathbf{B}}}$, $G_{\overline{\mathbf{A}} \cap \mathbf{B}}$ e $G_{\overline{\mathbf{A}} \cap \overline{\mathbf{B}}}$ não terem todos o mesmo sinal (nem todos positivos, nem todos negativos). Notemos que estamos perante um sistema de equações lineares que, caso

seja possível e determinado, significa que podemos fixar quaisquer valores para os ganhos $G_{\mathbf{A} \cap \overline{\mathbf{B}}}$, $G_{\overline{\mathbf{A}} \cap \mathbf{B}}$ e $G_{\overline{\mathbf{A}} \cap \overline{\mathbf{B}}}$ (por exemplo todos positivos), pois haverá sempre uma combinação de valores para $S_{\mathbf{A}}$, $S_{\mathbf{B}}$ e $S_{\mathbf{C}}$ para obtermos esses ganhos. Desta forma, o sistema de equações lineares não pode ser determinado. Como tal, o determinante do sistema tem que ser nulo, logo

$$\begin{vmatrix} 1 - p_{\mathbf{A}} & -p_{\mathbf{B}} & 1 - p_{\mathbf{C}} \\ -p_{\mathbf{A}} & 1 - p_{\mathbf{B}} & 1 - p_{\mathbf{C}} \\ -p_{\mathbf{A}} & -p_{\mathbf{B}} & -p_{\mathbf{C}} \end{vmatrix} = 0 \Rightarrow p_{\mathbf{C}} = p_{\mathbf{A}} + p_{\mathbf{B}}, \quad (10.87)$$

isto é,

$$\mathbb{P}_i(\mathbf{A} \cup \mathbf{B}) = \mathbb{P}_i(\mathbf{A}) + \mathbb{P}_i(\mathbf{B}). \quad (10.88)$$

□

Caso, no teorema anterior, não exigíssemos acontecimentos incompatíveis, $\mathbf{A} \cap \mathbf{B} = \emptyset$, então teríamos que incluir o resultado $\mathbf{D} = \mathbf{A} \cap \mathbf{B}$ com probabilidade $p_{\mathbf{D}}$ e ganho dado por

$$G_{\mathbf{A} \cap \mathbf{B}} = S_{\mathbf{A}} + S_{\mathbf{B}} + S_{\mathbf{C}} + S_{\mathbf{D}} - (S_{\mathbf{A}}p_{\mathbf{A}} + S_{\mathbf{B}}p_{\mathbf{B}} + S_{\mathbf{C}}p_{\mathbf{C}} + S_{\mathbf{D}}p_{\mathbf{D}}). \quad (10.89)$$

Seguindo raciocínio análogo, agora com uma matriz de dimensão 4, deduziríamos que

$$\mathbb{P}_i(\mathbf{A} \cup \mathbf{B}) = \mathbb{P}_i(\mathbf{A}) + \mathbb{P}_i(\mathbf{B}) - \mathbb{P}_i(\mathbf{A} \cap \mathbf{B}). \quad (10.90)$$

Teorema 10.13. *Seja $\mathbf{A}_1, \dots, \mathbf{A}_n$ um conjunto de acontecimento disjuntos onde um e um só deverá ocorrer e seja $p_i = \mathbb{P}_j(\mathbf{A}_i)$ os graus de credibilidade do indivíduo j na ocorrência do resultado $\mathbf{A}_i$ então, o facto de inexistência do dutch book é condição necessária e suficiente para que*

$$\sum_{i=1}^n p_i = 1. \quad (10.91)$$

Demonstração. Se representarmos os ganhos associados à ocorrência de $\mathbf{A}_i$ por G_i , então:

$$G_i = S_i - \sum_{i=1}^n p_i S_i. \quad (10.92)$$

Pelas mesmas razões invocadas na demonstração da aditividade, o determinante da matriz do sistema tem que ser nulo, logo

$$\begin{vmatrix} 1-p_1 & -p_2 & \cdots & -p_n \\ -p_1 & 1-p_2 & \cdots & -p_n \\ \cdots & \cdots & \cdots & \cdots \\ -p_1 & -p_2 & \cdots & 1-p_n \end{vmatrix} = 0 \Rightarrow 1 - \sum_{i=1}^n p_i = 0 \Rightarrow \sum_{i=1}^n p_i = 1. \quad (10.93)$$

□

Desta forma, de Finetti demonstrou que, se o conjunto de acontecimentos $\mathbf{A}_1, \dots, \mathbf{A}_n$ definem uma partição do espaço de resultados Ω e $p_i = \mathbb{P}_j(\mathbf{A}_i)$ representam os graus de credibilidade do indivíduo j na ocorrência do resultado $\mathbf{A}_i$, então a inexistência do *dutch book* é condição necessária e suficiente para que as probabilidades p_i constituam uma função de probabilidade ($\sum_{i=1}^n p_i = 1$ e $p_i \geq 0, \forall i$), deduzindo, também, os axiomas de Kolmogoroff no caso de o espaço de resultados ser finito. Desta forma, de Finetti considera que não se trata de axiomas, mas antes de proposições que surgem como consequência da sua definição de probabilidade.

Com esta definição de probabilidade, a aplicabilidade não se restringe a experiências aleatórias onde existe equiprobabilidade ou que possam ser repetidas indefinidamente sob as mesmas condições. Com base nela, qualquer experiência onde exista incerteza deve ser expressa através de uma distribuição de probabilidade. Assim, esta definição de probabilidade permite-nos aplicá-la a experiências únicas, não sendo necessário exigir a possibilidade de repetição infinita da experiência em condições idênticas.

10.4.3 Aditividade *versus* σ -aditividade

O princípio de coerência, segundo Bruno de Finetti, nada nos garante em relação à σ -aditividade. No entanto, este autor considera tal propriedade

desnecessária, tendo argumentado que a probabilidade não é σ -aditiva, pois, na prática, nunca trabalhamos com espaços infinitos. Apesar da complexidade de tratamento matemático que daqui pode advir, Bruno de Finetti opõe-se, portanto, à imposição de σ -aditiva na probabilidade. Salientemos que a hipótese de que a probabilidade é σ -aditiva simplifica bastante o tratamento de alguns problemas, o que leva, por exemplo, Fishburn (1986) a afirmar:

“I will not hesitate to invoke it (countable additivity) when its denial would create mathematical complexities of little interest to the topic at hand.”

[Fishburn 1986, p. 1]

Esta ideia parece ser partilhada pelo próprio Kolmogoroff que, como referimos previamente, afirma na sua obra que a inclusão da σ -aditividade e consequente recurso às σ -álgebras são apenas acontecimentos idealizados, sem nenhuma ligação com o mundo real, embora acabem por garantir a existência de conclusões contraditórias através da estrutura matemática que proporcionam.

Porque é que é exigida aditividade numerável? A aditividade finita é aceite por todos, pois, de facto, corresponde à nossa intuição. Se tivermos dois objectos com determinada medida, então se juntarmos os dois objectos a medida conjunta será igual à soma das medidas individuais. E, se assumirmos a aditividade para dois objectos, estamos a assumi-la para qualquer número finito de objectos. A aditividade não numerável é demasiado forte, bastando pensarmos que qualquer intervalo real fechado $[a, b]$, com $a < b$, pode ser obtido pela união não numerável de pontos $\{\omega\}$. Contudo, o seu comprimento é $b - a$ e o comprimento de $\{\omega\}$ é zero, isto é,

$$1 = \mu([0, 1]) \neq \sum_{\omega \in [0, 1]} \mu(\omega) = 0. \quad (10.94)$$

Perante isto, tal hipótese parece-nos inadequada. Assim, é de aceitação pacífica a aditividade finita e a não aditividade não numerável. A discussão entre as diferentes visões de probabilidade surge no que se refere à σ -aditividade, isto é, à aditividade infinita numerável. A σ -aditividade é aceite pela maioria, mas não por todos.

“Lorsque les événements considérés sont en nombre infini, notre définition n’introduit aucune difficulté nouvelle: $\mathbb{P}$ est une fonction de probabilité pour la classe infinie d’événements $\mathcal{E}$ lorsqu’elle l’est pour toute sous-classe finie de $\mathcal{E}$. Cette conclusion implique que le théorème des probabilités totales ne peut être étendu au cas d’une infinité, même dénombrable, d’événements.”

[Bruno de Finetti 1937, p. 13]

Bruno de Finetti, por exemplo, rejeita que a probabilidade tenha de ser σ -aditiva. Refira-se que Kolmogoroff salienta que a σ -aditividade é uma ferramenta útil na análise de muitos problemas na Teoria da Probabilidade, embora não a considere uma propriedade intrínseca ao mundo real, o que o leva a afirmar:

“We limit ourselves, arbitrarily, to only those models which satisfy Axiom VI. This limitation has been found expedient in researches of the most diverse sort.”

[Kolmogoroff 1933, p. 15]

De facto, a hipótese de σ -aditividade, em muitos problemas, facilita a sua análise, como na investigação dos teoremas de convergência que são fundamentais para a aplicabilidade desta teoria. Porém, esta hipótese adjacente à medida de probabilidade advém, segundo Kolmogoroff, da simplificação matemática que esta condição origina no tratamento matemático.

Não é a simplicidade ou complexidade da análise o único argumento que deve ser discutido a favor ou contra a σ -aditividade. De facto, a σ -aditividade

desempenha um papel importante nas demonstrações dos teoremas limites. No entanto, apesar da complexidade matemática acrescida pela não utilização da σ -aditividade, Chen (1977) fornece versões da Lei Forte dos Grande Números demonstradas recorrendo unicamente à aditividade. Existirá muito ainda por investigar neste campo, todavia, vamos prosseguir a discussão centrando-a unicamente nos fundamentos.

Um dos argumentos mais famosos de Bruno de Finetti é o problema “*guess the number*” ou a tentativa de adivinhar o número em que alguém está a pensar (escolha aleatória de um número natural). Este problema consiste em escolher à sorte um número natural. Assim, ao atribuírmos probabilidades subjectivas a cada número, devemos, em simultâneo, ter em consideração que não temos nenhuma razão para considerarmos que um número é mais (ou menos) provável do que qualquer outro. Por esta razão devemos atribuir a mesma probabilidade a cada número natural, seja p . Não existe, contudo, nenhum valor para p que satisfaça a σ -aditividade, ou seja, que verifique

$$\mathbb{P}\left(\bigcup_{i=1}^{+\infty}\{i\}\right) = \sum_{i=1}^{+\infty}\mathbb{P}(\{i\}) = \sum_{i=1}^{+\infty}p = 1. \quad (10.95)$$

Se $p > 0$, por mais pequeno que seja, a série é divergente e não se verificará a igualdade. Se $p = 0$ a série assume o valor zero e também não verificará a igualdade (neste caso todos os números teriam probabilidade nula, no entanto, a probabilidade de a escolha recair sobre um $n \in \mathbb{N}$ é 1). Como tal não é possível atribuir a distribuição uniforme aos números naturais no caso de a probabilidade ser considerada σ -aditiva⁽³²⁾. Salientemos, contudo, que, neste problema, se considerarmos $p > 0$, a probabilidade também não pode

⁽³²⁾ Esta afirmação é efectuada no contexto da axiomática de Kolmogoroff pois, conforme referimos, a axiomática proposta por Rényi (1955), baseada em espaços de probabilidade condicional, admite esta situação simultaneamente com a σ -aditividade inerente aos espaços de probabilidade.

ser aditiva, pois, se considerarmos que a probabilidade de qualquer número é p , então $\sum_{i=1}^n p > 1$ para qualquer n natural tal que $n > p^{-1}$. Assim, para considerarmos a distribuição uniforme em $\mathbb{N}$, teremos de assumir $p = 0$, isto é, todos os elementos terão probabilidade nula e a probabilidade não será σ -aditiva, pois

$$1 = \mathbb{P} \left(\bigcup_{i=1}^{\infty} \{i\} \right) \neq \sum_{i=1}^{+\infty} \mathbb{P}(\{i\}) = \sum_{i=1}^{+\infty} 0 = 0. \quad (10.96)$$

Desta forma, Bruno de Finetti argumenta que a σ -aditividade é uma restrição desnecessária, pois elimina algumas distribuições, tal como a uniforme em $\mathbb{N}$, que considera uma distribuição admissível, caso não seja impingida essa restrição. Notemos que Bruno de Finetti não está contra a existência de probabilidades (distribuições) σ -aditivas, mas opõe-se ao facto de todas as probabilidades terem de ser σ -aditivas, pois, pelos fundamentos que apresenta, esta condição não é necessária. Deste modo, uma vez que considera que a σ -aditividade não é implicada pelos seus fundamentos, nomeadamente o princípio de coerência e os ambientes de apostas, então não devemos impor esta restrição aos nossos modelos.

Diferentemente, Jon Williamson (1999) defende que, se aceitarmos o princípio da coerência ou a ausência do *dutch book* como fundamentos da probabilidade subjectiva, então temos que aceitar a σ -aditividade, pois esta é uma consequência do princípio. Desta forma, Williamson afirma que as probabilidades são coerentes se e só se são σ -aditivas.

Teorema 10.14 (Teorema de Williamson). *As probabilidades são coerentes se e só se satisfazerem a propriedade σ -aditiva, sob a hipótese de a quantidade de dinheiro apostada ser finita, isto é, $\sum_{i=1}^{+\infty} S_i < +\infty$.*

Demonstração. Consideremos que $\{a_i\}_{i=1}^{+\infty}$ representa os possíveis resultados de uma experiência aleatória (que, naturalmente, são mutuamente exclusivos

e exaustivos, isto é, o resultado da experiência será um e um só destes resultados) e sejam as respectivas probabilidades representadas por $p_i = \mathbb{P}(a_i)$ (que supomos $p_i > 0$) e as apostas por $S_i p_i$. Desta forma o ganho associado à ocorrência do resultado a_i é dado por

$$G_i = S_i - \sum_{i=1}^{+\infty} S_i p_i, \quad (10.97)$$

que é convergente, pelo facto de a quantidade de dinheiro apostada ser finita, pois $\sum_{i=1}^{+\infty} S_i p_i < \sum_{i=1}^{+\infty} S_i < +\infty$. Pretendemos provar que $G_k \geq 0$ para pelo menos um valor de k e $G_k \leq 0$ para pelo menos um valor de $k \iff \sum_{i=1}^{+\infty} p_i = 1$.

Supondo $\sum_{i=1}^{+\infty} p_i < 1$, então considerando $S_i = S$ constante e uma vez que $\sum_{i=1}^{+\infty} S_i p_i < \sum_{i=1}^{+\infty} S_i < \infty$ podemos concluir que

$$G_i = S_i - \sum_{i=1}^{+\infty} S_i p_i = S \left(1 - \sum_{i=1}^{+\infty} p_i \right) > 0, \quad (10.98)$$

pois $(1 - \sum_{i=1}^{+\infty} p_i) > 0$ pela hipótese $\sum_{i=1}^{+\infty} p_i < 1$. Desta forma o lucro é sempre positivo se apostarmos a mesma quantidade em todos os resultados e, conseqüentemente, se $\sum_{i=1}^{+\infty} p_i < 1$, poderá ocorrer o *dutch book*. Se supusermos $\sum_{i=1}^{+\infty} p_i > 1$, a conclusão será análoga, pois teremos um lucro sempre negativo. Desta forma, unicamente se $\sum_{i=1}^{+\infty} p_i = 1$ não se verificará o *dutch book*. Assim, supondo que $\sum_{i=1}^{+\infty} p_i = 1$, teremos então

$$\begin{aligned} \sum_{i=1}^{+\infty} G_i p_i &= \sum_{i=1}^{+\infty} p_i \left[S_i - \sum_{j=1}^{+\infty} S_j p_j \right] = \sum_{i=1}^{+\infty} p_i S_i - \sum_{i=1}^{+\infty} p_i \sum_{j=1}^{+\infty} S_j p_j = \\ &= \sum_{i=1}^{+\infty} p_i S_i - \sum_{j=1}^{+\infty} S_j p_j = 0. \end{aligned} \quad (10.99)$$

Para que se verifique a igualdade (10.99) existem duas possibilidades: ou $G_i = 0, \forall i$ (todos os resultados possíveis terão ganho nulo, o que acontecerá unicamente se $S_i = S$) ou, se existir algum $G_j > 0$, então terá de existir pelo menos um $G_k < 0$. Qualquer uma destas possibilidades verifica o princípio da coerência. $\square$

Porém este resultado não inclui todas as possibilidades. Por exemplo, não inclui a distribuição uniforme de Bruno de Finetti onde $\mathbb{P}(\{i\}) = 0$ para qualquer $i \in \mathbb{N}$, mas $\mathbb{P}(\mathbb{N}) = 1$ não se verificando a σ -aditividade.

A σ -aditividade continua ainda hoje a ser debatida, não havendo uma resposta definitiva sobre este assunto. Pacheco d'Amorim não analisa esta questão na sua tese pelo facto de, segundo o autor, a sua obra já estar demasiado extensa. Verifica-se, no entanto, que não é um problema de fácil resolução. Refira-se, por fim, que com a axiomática de Rényi podemos definir a distribuição uniforme em $\mathbb{N}$ pelo facto de trabalharmos em espaços de probabilidade condicional, uma vez que na construção deste espaço não existe o espaço de probabilidade intermédio (referente à probabilidade absoluta) onde seja exigido $\mathbb{P}(\Omega) = 1$. Assim podemos definir um espaço de probabilidade condicional de Rényi através de

$$\mathbb{P}(\mathbf{A}|\mathbf{B}) = \frac{\#(\mathbf{A} \cap \mathbf{B})}{\#(\mathbf{B})}, \quad (10.100)$$

onde $\#(\mathbf{B})$ pode assumir valores ilimitados. Pacheco d'Amorim, na sua obra, pretende definir estas situações, de distribuição uniforme em espaços ilimitados (discretos ou contínuos), mas não o efectuou.

10.4.4 Probabilidade Condicionada

Outro conceito fundamental na probabilidade bayesiana é o de probabilidade condicionada que Bruno de Finetti define recorrendo, uma vez mais, a ambientes de apostas e ao princípio da coerência.

Definição 10.12 (Probabilidade condicionada). Seja $S \times \mathbb{P}_i(\mathbf{A}|\mathbf{B})$ o preço máximo que o indivíduo i está disposto a apostar na ocorrência do acontecimento $\mathbf{A}$, onde pode ganhar um prémio de S unidades monetárias se, e somente se, ocorrer simultaneamente $\mathbf{A}$ e $\mathbf{B}$, sendo que, no caso de $\mathbf{B}$ não

ocorrer, a aposta deixa de fazer efeito, sendo o valor apostado $S \times \mathbb{P}_i(\mathbf{A}|\mathbf{B})$ devolvido.

Desta forma, se o acontecimento $\mathbf{B}$ não ocorrer, não haverá movimentação de dinheiro (ou, o que é equivalente, o dinheiro apostado é devolvido); caso o acontecimento $\mathbf{B}$ ocorra, o apostador paga o dinheiro da aposta, $S \times \mathbb{P}_i(\mathbf{A}|\mathbf{B})$ unidades monetárias, recebendo S unidades monetárias se e só se $\mathbf{A}$ ocorrer (ganhando, neste caso, $S \times (1 - \mathbb{P}_i(\mathbf{A}|\mathbf{B}))$ unidades monetárias).

Segundo de Finetti, podemos atribuir quaisquer valores às probabilidades $\mathbb{P}(\mathbf{A} \cap \mathbf{B})$, $\mathbb{P}(\mathbf{B})$ e $\mathbb{P}(\mathbf{A}|\mathbf{B})$ que, desde que sejamos coerentes (não exista o *dutch book*), satisfazem a regra da multiplicação (Teorema da Probabilidade Composta).

Teorema 10.15 (Teorema da Probabilidade Composta). *Verifica-se o princípio da coerência se e só se*

$$\mathbb{P}(\mathbf{A} \cap \mathbf{B}) = \mathbb{P}(\mathbf{A}|\mathbf{B})\mathbb{P}(\mathbf{B}), \quad (10.101)$$

propriedade denominada por Regra da Multiplicação.

Demonstração. Como as probabilidades que pretendemos relacionar são $\mathbb{P}(\mathbf{A} \cap \mathbf{B})$, $\mathbb{P}(\mathbf{A}|\mathbf{B})$ e $\mathbb{P}(\mathbf{B})$, representemos as apostas respectivas através de $S_{\mathbf{A} \cap \mathbf{B}}\mathbb{P}(\mathbf{A} \cap \mathbf{B})$, $S_{\mathbf{A}|\mathbf{B}}\mathbb{P}(\mathbf{A}|\mathbf{B})$ e $S_{\mathbf{B}}\mathbb{P}(\mathbf{B})$. Temos, desta forma, as três seguintes possibilidades de ocorrência:

1. Pode ocorrer $\mathbf{A} \cap \mathbf{B}$, onde teremos um ganho dado por

$$G_{\mathbf{A} \cap \mathbf{B}} = S_{\mathbf{A} \cap \mathbf{B}} + S_{\mathbf{A}|\mathbf{B}} + S_{\mathbf{B}} - (S_{\mathbf{A} \cap \mathbf{B}}\mathbb{P}(\mathbf{A} \cap \mathbf{B}) + S_{\mathbf{A}|\mathbf{B}}\mathbb{P}(\mathbf{A}|\mathbf{B}) + S_{\mathbf{B}}\mathbb{P}(\mathbf{B}));$$

2. Pode ocorrer $\bar{\mathbf{A}} \cap \mathbf{B}$, onde teremos um ganho dado por

$$G_{\bar{\mathbf{A}} \cap \mathbf{B}} = S_{\mathbf{B}} - (S_{\mathbf{A} \cap \mathbf{B}}\mathbb{P}(\mathbf{A} \cap \mathbf{B}) + S_{\mathbf{A}|\mathbf{B}}\mathbb{P}(\mathbf{A}|\mathbf{B}) + S_{\mathbf{B}}\mathbb{P}(\mathbf{B}));$$

3. Pode ocorrer $\overline{\mathbf{B}}$, onde teremos um ganho dado por

$$G_{\overline{\mathbf{B}}} = -(S_{\mathbf{A} \cap \mathbf{B}} \mathbb{P}(\mathbf{A} \cap \mathbf{B}) + S_{\mathbf{B}} \mathbb{P}(\mathbf{B})),$$

pois, nesta situação, a aposta $S_{\mathbf{A}|\mathbf{B}} \mathbb{P}(\mathbf{A}|\mathbf{B})$ é devolvida.

Uma vez que o determinante da matriz do sistema tem de ser nulo, caso contrário poderíamos escolher o montante ganho, teremos

$$\begin{vmatrix} 1 - \mathbb{P}(\mathbf{A} \cap \mathbf{B}) & 1 - \mathbb{P}(\mathbf{B}) & 1 - \mathbb{P}(\mathbf{A}|\mathbf{B}) \\ -\mathbb{P}(\mathbf{A} \cap \mathbf{B}) & 1 - \mathbb{P}(\mathbf{B}) & -\mathbb{P}(\mathbf{A}|\mathbf{B}) \\ -\mathbb{P}(\mathbf{A} \cap \mathbf{B}) & -\mathbb{P}(\mathbf{B}) & 0 \end{vmatrix} =$$

$$= \mathbb{P}(\mathbf{A} \cap \mathbf{B}) - \mathbb{P}(\mathbf{A}|\mathbf{B})\mathbb{P}(\mathbf{B}) = 0 \Leftrightarrow \mathbb{P}(\mathbf{A} \cap \mathbf{B}) = \mathbb{P}(\mathbf{A}|\mathbf{B})\mathbb{P}(\mathbf{B}).$$

□

Como consequência, a probabilidade condicionada $\mathbb{P}(\mathbf{A}|\mathbf{B})$ pode ser obtida através de

$$\mathbb{P}(\mathbf{A}|\mathbf{B}) = \frac{\mathbb{P}(\mathbf{A} \cap \mathbf{B})}{\mathbb{P}(\mathbf{B})}, \quad (10.102)$$

se $\mathbb{P}(\mathbf{B}) \neq 0$. O Teorema de Bayes, fundamental nesta visão da Probabilidade, pode ser visto como um corolário da regra da multiplicação e do Teorema das Probabilidades Totais.

Teorema 10.16 (Teorema de Bayes). *Consideremos $(\mathbf{A}_1, \dots, \mathbf{A}_n)$, com $n \in \mathbb{N}$, uma partição do espaço-amostra Ω , i.e. $\mathbb{P}(\mathbf{A}_i) > 0$, $\mathbf{A}_i \cap \mathbf{A}_j = \emptyset$ para qualquer $i \neq j$ e $\cup_{i=1}^n \mathbf{A}_i = \Omega$. Seja $\mathbf{B} \in \Omega$ um acontecimento qualquer tal que $\mathbb{P}(\mathbf{B}) \neq 0$, então*

$$\mathbb{P}(\mathbf{A}_i|\mathbf{B}) = \frac{\mathbb{P}(\mathbf{B}|\mathbf{A}_i) \mathbb{P}(\mathbf{A}_i)}{\sum_{j=1}^n \mathbb{P}(\mathbf{B}|\mathbf{A}_j) \mathbb{P}(\mathbf{A}_j)}, \quad i = 1, \dots, n. \quad (10.103)$$

Demonstração. Para $i = 1, \dots, n$ temos

$$\mathbb{P}(\mathbf{A}_i|\mathbf{B}) = \frac{\mathbb{P}(\mathbf{A}_i \cap \mathbf{B})}{\mathbb{P}(\mathbf{B})} = \frac{\mathbb{P}(\mathbf{B}|\mathbf{A}_i) \mathbb{P}(\mathbf{A}_i)}{\sum_{j=1}^n \mathbb{P}(\mathbf{B}|\mathbf{A}_j) \mathbb{P}(\mathbf{A}_j)}.$$

□

Notemos que a definição de probabilidade condicionada, apesar de dependente de ambiente de apostas, não é restrita às situações $\mathbf{B} \subset \mathbf{A}$ tal como na concepção de Pacheco d'Amorim, pois é possível a ocorrência de $\mathbf{A}$ quando $\mathbf{B}$ não acontece. Contudo, estas situações $(\mathbf{A} \cap \overline{\mathbf{B}})$ são, de certa forma, retiradas da análise, pois, sempre que $\mathbf{B}$ não ocorra, a aposta é anulada (ninguém ganha nada). Esta ideia corresponde à nossa ideia intuitiva de probabilidade condicionada, onde o novo universo, da probabilidade condicionada, corresponde às situações em que ocorre o condicionante $\mathbf{B}$ e, naturalmente, o acontecimento $\mathbf{A} \cap \overline{\mathbf{B}}$ não nos interessa nestes casos.

10.4.5 Permutabilidade e Teoremas de Representação

Bruno de Finetti considera que não pode haver independência entre as variáveis aleatórias com probabilidade constante e desconhecida, pois nesse caso não se aprenderia nada com a experiência, refutando desta forma a visão de von Mises anteriormente apresentada. Para tornarmos mais clara esta ideia, recordemos, de forma muito sucinta, o exemplo presente no artigo de Laplace (1774), considerado por alguns o primeiro artigo bayesiano⁽³³⁾.

Consideremos uma urna contendo uma infinidade de bolas (ou um número finito de bolas onde as bolas são retiradas com reposição) brancas e pretas com percentagem desconhecida, de onde retiramos $n + m$ bolas da urna, das quais n são brancas e m são pretas. Qual a probabilidade de que a próxima bola a ser retirada seja branca?

Seja X_i a variável aleatória que representa a i -ésima bola retirada, que assume o valor 1 no caso de a bola ser branca e assume o valor 0 caso seja preta — uma sucessão de provas de Bernoulli com probabilidade de sucesso

⁽³³⁾ Este exemplo foi analisado com mais pormenor na secção 3.7.

desconhecida. Pretende-se determinar o valor de

$$\mathbb{P} \left(X_{n+m+1} = 1 \left| \sum_{i=1}^{n+m} X_i = n \right. \right). \quad (10.104)$$

Se considerarmos que as variáveis aleatórias X_i são independentes, como na visão clássica (frequentista), então

$$\mathbb{P} \left(X_{n+m+1} = 1 \left| \sum_{i=1}^{n+m} X_i = n \right. \right) = \mathbb{P}(X_{n+m+1} = 1) = \mathbb{P}(X_1 = 1), \quad (10.105)$$

não havendo qualquer processo de aprendizagem, ou seja, o facto de já termos retirado $n + m$ bolas, das quais n são brancas, não altera a nossa visão sobre a probabilidade e, consequentemente, é impossível fazer inferência estatística nestas situações. Este facto é fortemente criticado pelos defensores da interpretação bayesiana de probabilidade. Se, no início, não sabemos qual a probabilidade de cada cor e se retirarmos 10^{10} bolas e todas elas forem brancas, continuaremos sem saber nada sobre a probabilidade de uma bola ser branca? Para os bayesianos, como a probabilidade de sucesso é desconhecida, o facto de saírem bolas brancas ou pretas altera o nosso grau de credibilidade sobre a cor que vai sair a seguir. Se em 10^{10} bolas retiradas todas forem brancas, teremos um forte grau de credibilidade de que a próxima bola a sair também seja branca. Formalizando esta ideia,

$$\mathbb{P} \left(X_{n+m+1} = 1 \left| \sum_{i=1}^{n+m} X_i = n \right. \right) = \frac{\mathbb{P} \left(X_{n+m+1} = 1, \sum_{i=1}^{n+m} X_i = n \right)}{\mathbb{P} \left(\sum_{i=1}^{n+m} X_i = n \right)} \quad (10.106)$$

onde, para a resolução da probabilidade pretendida, se recorre ao Teorema de Bayes. Esta interpretação de probabilidade denomina-se por bayesiana devido à utilização do Teorema de Bayes na actualização das probabilidades após a obtenção de nova informação. Assim, o Teorema de Bayes é utilizado para combinar a informação *a priori* $h(\theta)$ com a informação contida nos dados $f(x | \theta)$ para a obtenção da probabilidade a posteriori $h(\theta | x)$, através

de

$$h(\theta | x) = \frac{f(x | \theta) h(\theta)}{\int_{\Theta} f(x | \theta) h(\theta) d\theta}, \quad (10.107)$$

ou, muitas vezes, representando os dados (observações provenientes de uma amostra) por D , obtemos

$$p(\theta|D) \propto p(D|\theta) \cdot p(\theta). \quad (10.108)$$

Sempre que houver nova informação devemos recorrer uma vez mais ao mesmo resultado para actualizar as probabilidades. Nesta actualização temos a distribuição *a priori* sobre o parâmetro, ou vector de parâmetros θ . As distribuições *a priori* são definidas, segundo Bruno de Finetti, pela vontade de agir do indivíduo sobre alguma proposição.

Salientemos que na estatística clássica o parâmetro $\theta \in \Theta$ é um vector desconhecido mas fixo, não sendo assim possível atribuir probabilidades aos valores admissíveis de θ . Na estatística bayesiana, como toda a incerteza deve ser descrita através de probabilidades e $\theta \in \Theta$ é um vector desconhecido, então devemos quantificar a incerteza associada a θ através das distribuições de probabilidade. Não significa isto que θ varie ou que possa assumir vários valores pois θ continua a ser interpretado como um vector com valores fixos; no entanto, como é desconhecido, as probabilidades que associamos a θ espelham o nosso grau de credibilidade sobre os valores de Θ que consideramos possíveis para θ assumir com a informação de que dispomos.

Assim, nesta visão, existe a combinação da informação tida *a priori* (distribuições *a priori*) com a informação obtida por amostragem (funções de verosimilhança), através do Teorema de Bayes, para se obterem as distribuições *a posteriori* que são a base deste tipo de inferência. Este tipo de inferência (bayesiana) apresenta várias interpretações de probabilidade a que correspondem diferentes formas de determinar a distribuição *a priori*. Para os bayesianos personalistas, como Bruno de Finetti, as distribuições *a priori*

são definidas pela vontade de agir do indivíduo sobre alguma proposição, isto é, os graus de credibilidade são determinados pela disposição de um indivíduo para fazer determinada escolha (*behavior criteria*).

Como as variáveis não podem ser consideradas independentes para a existência de um processo de aprendizagem, Bruno de Finetti desenvolve um conceito de dependência que, por um lado, permita fazer inferência (haver aprendizagem com as experiências), por outro lado, que seja tão fraca quanto possível de forma a incluir os outros tipos de dependência. Naturalmente isto não significa que os bayesianos não aceitem o conceito de independência das variáveis aleatórias, pois, se estivermos perante provas de Bernoulli com probabilidades conhecidas, estas provas são consideradas independentes, mesmo pelos defensores desta teoria. Não concordam que, no caso de a probabilidade de sucesso ser desconhecida, as provas sejam independentes, pois, à medida que vão sendo efectuadas novas provas, os seus resultados fornecem-nos informação sobre a probabilidade p desconhecida e, conseqüentemente, os nossos graus de credibilidade sobre o valor de p são alterados. Desta forma, quando p não é conhecido, as provas não podem ser consideradas independentes.

O tipo de dependência que de Finetti criou denomina-se por permutabilidade (*exchangeability*), onde uma sequência de variáveis aleatórias se diz permutável se a sua distribuição conjunta for invariante para permutações das variáveis.

Definição 10.13 (Permutabilidade). Um conjunto finito de variáveis (quantidades) aleatórias $(X_1, X_2, \dots, X_n)$ diz-se permutável quando

$$f(x_1, x_2, \dots, x_n) = f(x_{\pi(1)}, x_{\pi(2)}, \dots, x_{\pi(n)}), \quad (10.109)$$

para qualquer permutação $\{\pi(1), \pi(2), \dots, \pi(n)\}$ do conjunto $\{1, 2, \dots, n\}$. Um conjunto infinito de variáveis (quantidades) aleatórias X_i , $i = 1, 2, \dots$ diz-se permutável se qualquer subsucessão finita for permutável.

As variáveis aleatórias independentes e identicamente distribuídas são permutáveis, mas as variáveis aleatórias permutáveis não são necessariamente independentes (apesar de serem identicamente distribuídas). Por exemplo, se considerarmos uma população dividida em dois grupos e recolhermos uma amostra sem reposição, a distribuição conjunta da amostra aleatória tem distribuição hipergeométrica, onde as variáveis que constituem a amostra são permutáveis (a sua ordem não irá alterar a distribuição conjunta), mas não são independentes. A importância deste tipo de dependência é destacado pelos Teoremas de Representação, que garantem que variáveis permutáveis são independentes condicionalmente a um valor fixo do parâmetro.

Teorema 10.17 (Teorema de Representação de Finetti). *Se $X_i, i = 1, 2, \dots$ é uma sucessão infinita de variáveis aleatórias binárias permutáveis, então qualquer que seja o inteiro n existe uma função de distribuição F_θ tal que*

$$f(x_1, x_2, \dots, x_n) = \int_0^1 \prod_{i=1}^n \theta^{x_i} (1 - \theta)^{1-x_i} dF_\theta(\theta), \quad (10.110)$$

onde

$$F_\theta(\theta) = \lim_{n \rightarrow \infty} \mathbb{P} \left(\frac{\sum_{i=1}^n X_i}{n} \leq \theta \right), \quad (10.111)$$

e

$$\theta = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n X_i. \quad (10.112)$$

Desta forma, a amostra $X_1, \dots, X_n$ comporta-se como se se tratasse de um conjunto de variáveis aleatórias independentes condicionalmente a θ , pois

$$f(x_1, x_2, \dots, x_n | \theta) = \prod_{i=1}^n f(x_i | \theta) = \prod_{i=1}^n \theta^{x_i} (1 - \theta)^{1-x_i}. \quad (10.113)$$

Se a distribuição de θ é degenerada no ponto θ_0 , então o Teorema de Representação de Finetti diz-nos que as variáveis aleatórias $(X_1, X_2, \dots, X_n)$ são

independentes e identicamente distribuídas com função massa de probabilidade $f(x_i) = \theta_0^{x_i} (1 - \theta_0)^{1-x_i}$.

Num caso mais geral, o Teorema da Representação diz que, se considerarmos que um conjunto de observações tem origem numa sequência permutável, então este conjunto de observações constitui uma amostra aleatória de um modelo de probabilidade $\{f(x_1, x_2, \dots, x_n | \theta), \theta \in \Theta\}$, que depende de um parâmetro θ (ou vector de parâmetros), que é definido como o limite, quando a amostra aumenta indefinidamente, de uma qualquer função das suas observações e que será caracterizado por uma distribuição $F_\theta(\theta)$. Desta forma teremos

$$\begin{aligned} f(x_1, x_2, \dots, x_n) &= \int_{\Theta} f(x_1, x_2, \dots, x_n | \theta) \, dF_\theta(\theta) = \\ &= \int_{\Theta} \prod_{i=1}^n f(x_i | \theta) \, dF_\theta(\theta). \end{aligned} \quad (10.114)$$

Todavia, dado que as funções de densidade marginal $f(x_i | \theta)$ dependem de θ , não podemos permutar o integral como o produto na expressão (10.114). Desta forma, quando estamos perante variáveis aleatórias permutáveis, a inferência depende da distribuição de θ , que corresponde à distribuição *a priori*. Estes teoremas de representação, conjuntamente com a noção de permutabilidade, são um pilar fundamental na visão bayesiana de probabilidade.

Saliente-se que os Teoremas da Representação são resultados da Teoria da Probabilidade, não existindo qualquer dúvida sobre a sua validade. Porém, estes exigem uma distribuição inicial e é esta exigência que constitui o cerne da questão. A sua existência separa bayesianos de não bayesianos. Mesmo entre os bayesianos, a forma como esta distribuição é obtida é o principal factor de divisão entre as distintas interpretações bayesianas.

Na secção 3.7 apresentamos a resolução de Pacheco d'Amorim de um problema semelhante ao habitualmente denominado Regra da Sucessão de

Laplace. Nessa resolução, conforme comentamos, Pacheco d'Amorim considerou que as variáveis só eram independentes quando consideradas condicionalmente a um valor fixo da proporção de bolas brancas (urna fixa). Contudo, esta suposição decorreu do facto de Pacheco d'Amorim considerar dois problemas idênticos: um onde escolhemos uma urna (de entre um conjunto de urnas com proporção de bolas brancas conhecida) e depois tiramos bolas sempre da mesma urna, onde é claro que a distribuição das cores das bolas retiradas depende da urna escolhida; e um outro problema onde só temos uma urna mas a percentagem de bolas brancas é desconhecida. Apesar de Pacheco d'Amorim ter resolvido este problema com o *modus operandi* da escola bayesiana, parecendo utilizar uma ideia semelhante à de permutabilidade, e de ter dado bastante importância ao Teorema de Bayes, quer no caso discreto quer no contínuo, pensamos que a visão de Pacheco d'Amorim está em conformidade com a visão frequencista e não com a interpretação bayesiana. Notemos, por exemplo, que, na CONCLUSÃO da sua tese, Pacheco d'Amorim não se preocupa com a utilização de qualquer tipo de distribuição *a priori*, fundamentando a sua inferência unicamente na informação que advém da observação repetida do fenómeno. Por este motivo enquadrámos a sua construção na visão frequencista de probabilidade, sendo, em alguns aspectos, como referimos ao longo da secção 10.3, muito semelhante à visão de Richard von Mises.

10.4.6 A visão lógica indutiva

A interpretação subjectiva ou personalista de probabilidade não é a única visão bayesiana de probabilidade, existe a interpretação bayesiana objectiva⁽³⁴⁾

⁽³⁴⁾ Bandyopadhyay (2005) distingue quatro versões de bayesianismo objectivo, desde a visão mais extremista à mais moderada (mais próxima da visão personalista).

ou lógica indutiva. Nesta visão a probabilidade representa uma relação lógica entre uma proposição e um nível de conhecimento, logo, será sempre uma probabilidade condicionada, pois, nesta interpretação, não faz sentido a existência das probabilidades absolutas (não condicionadas). Assim, uma proposição só assume uma probabilidade para o mesmo nível de conhecimento. Deste modo, dada uma proposição (hipótese) $\mathbf{H}$ e um nível de conhecimento $\mathbf{C}$ existe um e um só número real p tal que a probabilidade de $\mathbf{H}$ relativo a $\mathbf{C}$ é p . Nesta visão de probabilidade, $\mathbb{P}(\mathbf{A}|\mathbf{B})$ é o grau de credibilidade (*degree of belief*) sobre uma relação lógica entre uma hipótese (premissa) $\mathbf{B}$ e uma proposição (conclusão) $\mathbf{A}$, considerando que a probabilidade de um acontecimento depende necessariamente da informação que possuímos, ou seja, das hipóteses consideradas. Isto permite-nos que, nesta visão, encaremos a probabilidade como uma extensão da Lógica clássica, sendo a probabilidade uma medida de um grau de implicação de uma proposição pela informação disponível e, consequentemente, um grau de credibilidade objectivo, onde se verifica a característica da unicidade (para uma determinada quantidade de informação existe um único grau de implicação associado a cada proposição). É uma generalização da lógica tradicional, onde se quantificam os graus intermédios de implicação da proposição pela informação existente (quando a informação não permite concluir que a proposição é verdadeira ou que é falsa). Nesta visão destacam-se alguns autores tais como de Morgan (1847), G. Boole (1854), W. E. Johnson (1921), J. M. Keynes (1921), H. Jeffreys (1939) ou, mais recentemente, E. T. Jaynes (1996). Refira-se, no entanto, que os conceitos subjectivo e lógico de probabilidade representam, ambos, um grau de credibilidade numa proposição condicionado à informação disponível.

Keynes (1921) define a probabilidade como o grau de credibilidade (*degree of belief*) sobre uma relação lógica entre uma hipótese (premissa) h e uma proposição (conclusão) p , representando a probabilidade por p/h , pois

esta depende das hipóteses consideradas (sendo reavaliada caso as hipóteses sejam alteradas). Assim, segundo Keynes, para agentes com o mesmo nível de conhecimento a probabilidade de uma dada proposição é a mesma. Todavia, considera que a probabilidade de uma relação lógica nem sempre é mensurável, sendo unicamente possível determiná-la quando é aplicável o *princípio da indiferença* (nome que o autor dá ao *princípio da razão insuficiente*, presente na Teoria Clássica da Probabilidade), nem, por vezes, comparável (sendo impossível de analisar se o grau de credibilidade de uma proposição é igual, menor ou maior que o de outra proposição). Keynes, para as situações em que existe uma relação de probabilidade entre a conclusão p e uma premissa h apresenta um conjunto de axiomas através dos quais demonstra as principais propriedades da probabilidade, tal como a aditividade, a multiplicação, a independência e o princípio da inversão. Keynes divide os problemas de indução (inferência) em três classes, presentes nas situações seguintes: conhecendo as probabilidades, pretendemos inferir que frequências esperar; conhecendo as frequências estatísticas, desvendar o que podemos afirmar sobre as probabilidades (inverso do primeiro problema); conhecendo as frequências estatísticas de um conjunto de dados observado, descobrir o que podemos inferir sobre as frequências relativas de outros conjuntos de dados ainda não observados.

Na sua construção, Keynes salienta a importância do Teorema de Bernoulli, focando as suas condições, e elogia o trabalho da escola russa, com destaque para Chebycheff, considerando a escola Laplaciana obsoleta. Estabelece dois métodos directos para encontrar a probabilidade *a priori*: o uso do princípio da indiferença e a inversão do Teorema de Bernoulli. Considera, assim, que é possível inverter o Teorema de Bernoulli (que na sua opinião se deveria denominar por estabilidade das frequências estatísticas, em vez de Lei dos Grandes Números) sob determinadas condições. Deste modo, sa-

bendo as frequências relativas num conjunto de observações é possível inferir sobre as respectivas probabilidades, sendo essencial apreender as condições de utilização do Teorema de Bernoulli, entre as quais destaca a estabilidade.

Harold Jeffreys (1939)⁽³⁵⁾ apresenta uma visão semelhante à de Keynes, ainda que discorde deste por considerar que as probabilidades só são parcialmente ordenadas. Para Jeffreys a probabilidade é uma função com dois argumentos, é a probabilidade de p com os dados q e não apenas a probabilidade de p . Este autor, após apresentar as oito regras a que a indução deve obedecer, apresenta os seus axiomas e convenções⁽³⁶⁾ que a probabilidade deve verificar. Com estas regras deduz igualmente as principais propriedades da probabilidade. Nesta visão, tal como na de Keynes, o Teorema de Bayes, ou *princípio da probabilidade inversa*, assume importância capital, sendo utilizado $\mathbb{P}(q_r|pH) \propto \mathbb{P}(q_r|H) \mathbb{P}(p|q_rH)$ para efectuar inferência, onde $\mathbb{P}(q_r|H)$ é a probabilidade *a priori*, $\mathbb{P}(q_r|pH)$ a probabilidade *a posteriori* e $\mathbb{P}(p|q_rH)$ a *verosimilhança* (termo introduzido por Fisher). Para Jeffreys (1939) a distribuição *a priori* deverá ser escolhida de forma a obter resultados equivalentes (princípio da invariância), com qualquer transformação não singular dos parâmetros (transformações injectivas)⁽³⁷⁾. Com este método Jeffreys defende,

⁽³⁵⁾ Jeffreys, no prefácio da sua obra, refere que publicou juntamente com Wrinch dois artigos em 1919 e 1921 na *Philosophical Magazine* que antecederam a obra de Keynes. Refere ainda que as semelhanças entre as suas visões se devem ao facto de ambos terem sido alunos de W. E. Johnson.

⁽³⁶⁾ As convenções diferem dos axiomas por não serem necessários, isto é, são apenas regras introduzidas por conveniência, tendo a propriedade de que com outras regras seriam obtidos resultados equivalentes. Por exemplo, considerar que atribuímos um número maior a uma proposição que consideramos mais provável ou a probabilidade unitária à proposição certa.

⁽³⁷⁾ Seja $\beta = \psi(\theta)$ uma transformação injectiva do parâmetro θ que tem suporte num intervalo contínuo de valores. Se a distribuição de θ é uniforme, não é necessariamente

para a distribuição *a priori* do parâmetro θ , a utilização do princípio da razão insuficiente para Θ finito, o princípio da invariância sob transformações lineares para parâmetros de localização em que Θ é um intervalo limitado ou $\mathbb{R}$ (conduzindo à uniforme própria ou imprópria) e o princípio da invariância sob potências para parâmetros de escala em que $\Theta = \mathbb{R}^+$, de onde se obtém $h(\theta) \propto \theta^{-1} \mathbb{I}_{(0,+\infty)}(\theta)$.

Jaynes (1996), na sua obra inacabada, utiliza a imagem de um robot (contrariando desta forma a ideia personalista de probabilidade) onde são inseridos os dados para que ele, unicamente com as regras (*desiderata*) a que a probabilidade deverá obedecer, consiga efectuar inferência. Na sua exposição Jaynes salienta a importância dos Teoremas de Cox (1946) e de Shannon (1948). Jaynes (1996) começa por apresentar os *desiderata* da sua construção, considerando que os graus de plausibilidade (probabilidades) são representados por números reais que fazem correspondência qualitativa com o senso comum, sendo consistentes. Para Jaynes os graus de plausibilidade são consistentes se, para uma conclusão que possa ser obtida por mais de uma forma, as distintas possibilidades conduzirem ao mesmo resultado. Se o robot imaginário levar em consideração toda a informação que possui para a inferência (pois, sendo robot, não possui ideologia e, como tal, não ignora qualquer informação) e se o robot associar sempre os mesmos graus de plausibilidade para equivalentes estados de conhecimento, isto é, se em dois problemas o estado de conhecimento do robot for o mesmo, então os graus de plausibilidade associados serão os mesmos. O Teorema de Cox (1946), que assume papel capital na fundamentação de Jaynes, consiste em, unicamente recorrendo aos *desiderata*, demonstrar o Teorema da Multiplicação

$$\mathbb{P}(\mathbf{AB}|\mathbf{C}) = \mathbb{P}(\mathbf{A}|\mathbf{BC}) \mathbb{P}(\mathbf{B}|\mathbf{C}) = \mathbb{P}(\mathbf{B}|\mathbf{AC}) \mathbb{P}(\mathbf{A}|\mathbf{C}) \quad (10.115)$$

verdade que a distribuição de β também o seja, pois, sendo $h(\theta)$ a distribuição (*a priori*) de θ , tem-se que a distribuição (*a priori*) de β será $g(\psi) = h(\theta(\psi)) \left| \frac{d\theta}{d\psi} \right|$.

e o da Adição

$$\mathbb{P}(\mathbf{A}|\mathbf{B}) + \mathbb{P}(\overline{\mathbf{A}}|\mathbf{B}) = 1 \quad (10.116)$$

e consequentemente o Teorema de Bayes

$$\mathbb{P}(\mathbf{A}|\mathbf{BC}) = \mathbb{P}(\mathbf{A}|\mathbf{C}) \frac{\mathbb{P}(\mathbf{B}|\mathbf{AC})}{\mathbb{P}(\mathbf{B}|\mathbf{C})}. \quad (10.117)$$

Desta forma, tal como em qualquer visão bayesiana, teremos o Teorema de Bayes a relacionar a probabilidade *a priori* com a verosimilhança para determinarmos a probabilidade *a posteriori*. Na discussão da determinação da probabilidade *a priori*, Jaynes destaca o Teorema de Shannon (1948). Este teorema introduz o princípio da máxima entropia na probabilidade e pode ser considerado uma generalização do princípio da indiferença, uma vez que o princípio da máxima entropia fornece a distribuição uniforme quando temos um número finito de resultados sendo a única restrição do método a soma das probabilidades ser igual à unidade. A entropia é uma medida da desordem e imprevisibilidade de um sistema físico utilizada em Física (Termodinâmica) desde o final do século XIX. Quanto maior for a imprevisibilidade, maior é a entropia. Pretende-se, assim, que uma distribuição não informativa tenha a máxima entropia. Refira-se que a definição de entropia de uma distribuição, embora seja aceite no caso discreto, não é pacífica a sua fórmula no caso contínuo. Outro problema deste método advém da complexidade computacional que requer.

Além da visão bayesiana subjectiva, fundada nos princípios de Bruno de Finetti, e da visão bayesiana objectiva ou lógica, baseada nas ideias de Keynes e Jeffreys, há outras visões bayesianas de Probabilidade, tal como a definição bayesiana empírica, onde a informação *a priori* é estimada unicamente através de observações passadas (sendo os dados a única fonte de informação).

10.5 Comentário final à construção de Diogo Pacheco d'Amorim

*“Um pouco mais de sol - eu era brasa,
Um pouco mais de azul - eu era além.
Para atingir, faltou-me um golpe de asa ...
Se ao menos eu permanecesse alguém ...”*

[Mário de Sá Carneiro (1890–1916), Quase, 1913]

O principal objectivo da tese de doutoramento de Pacheco d'Amorim é apresentar os fundamentos do Cálculo das Probabilidades e propor uma ligação desta teoria com a Estatística. Na nossa opinião, a sua visão é extremamente avançada para a época em que escreveu a sua tese, mesmo tendo em consideração quer os erros que comete quer a falta de formalização de alguns conceitos que utiliza.

Na construção de Pacheco d'Amorim temos o espaço de probabilidade condicional $[\Omega, \mathcal{A}, \mathcal{A}', \mathbb{P}(\mathbf{A}|\mathbf{B})]$, onde a medida $\mathbb{P}(\mathbf{A}|\mathbf{B})$, pelo facto de ser consequência de uma escolha à sorte, é caracterizada pela equipossibilidade dos elementos de Ω . A apresentação de Pacheco d'Amorim é restrita a espaços de resultados Ω limitados de $\mathbb{R}^n$ (ou de $\mathbb{N}^n$ para o caso discreto), apesar de ter o intuito de estender as suas definições a escolhas efectuadas em espaços de resultados Ω ilimitados. Com o seu conceito de ponto imagem $y = f(x)$ podemos transpor a medida $\mathbb{P}(\mathbf{A}|\mathbf{B})$ para o espaço de probabilidade condicional $[\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), \mathcal{B}'(\mathbb{R}^n), \mathbb{P}_y(\mathbf{A}|\mathbf{B})]$, induzido pela transformação $y = f(x)$, sendo este espaço caracterizado pela função de possibilidade π_y . Pacheco d'Amorim não se preocupa com a devida fundamentação destes espaços de probabilidade condicional, embora seja relevante assinalar que, quando es-

creveu a sua tese de doutoramento, a Teoria da Medida ainda estava a dar os seus primeiros passos, não permitindo a justificação deste tipo de espaços tal como Rényi (1955) conseguiu quarenta anos depois.

Os teoremas limites, nomeadamente as Leis dos Grandes Números, são a base das suas aplicações. As condições destes teoremas exigem que as experiências sejam efectuadas em iguais condições, o que implica que a nossa amostra aleatória seja uma concretização de um conjunto de variáveis aleatórias independentes e identicamente distribuídas. Deste modo Pacheco d'Amorim considera que unicamente podemos inferir sobre as probabilidades dos possíveis resultados de uma experiência aleatória através da observação repetida da experiência em condições análogas, não efectuando qualquer referência à necessidade de recorrer a uma distribuição *a priori* para efectuar inferência. A sua visão da inferência estatística é, assim, semelhante à proposta por Richard von Mises.

Pelas razões expostas, apesar de a formalização aqui utilizada ser nossa (baseada nas ideias de Kolmogoroff (1933) e posteriormente de Rényi (1970)) e não do autor, pensamos que a visão de Pacheco d'Amorim é extremamente válida, mesmo na actualidade. Caso a sua obra tivesse usufruído de divulgação internacional, poderia ter sido posteriormente provida da devida fundamentação, recorrendo à Teoria da Medida, necessária para evitar qualquer patologia e dar à sua construção o rigor matemático exigível. Nesta suposição, a sua proposta de construção poderia ser hoje uma interpretação de probabilidade, incluída nas visões frequencistas, tal como os trabalhos de Richard von Mises ou Rényi, que estaria presente nas aplicações que efectuamos usualmente. Todo este mérito nos leva a considerar que Pacheco d'Amorim esteve a um “golpe de asa” de pertencer à lista das personalidades que contribuíram para o desenvolvimento da Teoria da Probabilidade (e suas aplicações) como uma área da Matemática, acabando com os parado-

xos que assombravam esta Teoria no início do século XX e conferindo-lhe o rigor necessário para ter actualmente uma importância capital em todas as áreas do conhecimento. Na lista de notáveis destacam-se nomes, quase todos referidos ao longo deste trabalho, tais como Bertrand, Poincaré, Borel, Bachelier, Lévy, Kolmogoroff, von Mises, de Finetti, Rényi, entre muitos outros. Causa-nos perplexidade que entre eles não seja apontado Pacheco d'Amorim e que não exista qualquer referência ao seu trabalho em obras internacionais.

“This branch of Mathematics (probability) is the only one, I believe, in which good writers frequently get results entirely erroneous.”

[Pierce, Charles Sanders (1839–1914)]

A tese de Pacheco d'Amorim contém alguns erros, tal como muitas outras obras dedicadas à Teoria da Probabilidade, mesmo escritas pelos mais notáveis matemáticos franceses, no início do século XX, e não chega a atingir o seu objectivo de obter uma construção que consiga tornear todos os paradoxos que assombravam esta teoria na época. Realce-se, todavia, que as suas ideias antecipam, em muitos aspectos, conceitos actualmente centrais na Teoria da Probabilidade e da Estatística. Em suma, a construção proposta por Pacheco d'Amorim não possui na formalização o rigor exigido, mas, no ponto de vista intuitivo, na sua abordagem filosófica à construção da Probabilidade e à sua ligação com a Estatística, antecipa diversas ideias que só virão a surgir, nas publicações internacionais, depois da sua tese de doutoramento e que, na actualidade, são primordiais. cremos, assim, dispor de motivos sólidos para consideramos que esta obra merece ocupar um lugar de destaque na História da Teoria da Probabilidade e da Estatística.

BIBLIOGRAFIA

Referências Bibliográficas

- ADAMS, M. & GUILLEMIN, V. (1996). *Measure Theory and Probability*, 2nd ed., Birkhäuser, Boston.
- ALCAYDE Y CARVAJAL, N. (1908). *Cálculo de Probabilidades*, Imprenta del Colegio de Huérfanos de la Guerra, Guadalajara.
- ALEXANDERSON, G. (2000) *The Random Walks of George Pólya*, The Mathematical Association of America, U.S.A..
- BACHELIER, L. (1900). Théorie de la Spéculation, *Annales Scientifiques de l'École Normale Supérieure* **3** t. 17, p. 21–86.
- BACHELIER, L. (1901). Théorie Mathématique du Jeu. *Annales Scientifiques de l'École Normale Supérieure* **3** t. 18, p. 143–209.
- BACHELIER, L. (1912). *Calcul des Probabilités*, Gauthier–Villars, Paris.
- BALAKRISHNAN, N. & NEVZOROV, V. (2003). *A Primer on Statistical Distributions*, John Wiley & Sons, New York.
- BANDYOPADHYAY, P. (2005). *Two Dogmas of Strong Objective Bayesianism*, Montana State University.
- BARBIER, M. (1860). Note sur le Problème de l'Aiguille et le Jeu du Joint Couvert, *Journal de Mathématiques Pures et Appliquées II* **5**, p. 273–286.
- BASHARIN, G.; LANGVILLE, A. & NAUMOV, V. (2004). The Life and Work of A. Markov, *Linear Algebra and its Applications* **386**, p. 3–26.

- BAYES, T. (1764). An Essay Toward Solving a Problem in the Doctrine of Chances, *Philosophical Transactions of the Royal Society of London*, **53**, p. 370–418.
- BENZI, M.; BENZI, M. & SENETA, E. (2007). Francesco Paolo Cantelli, *International Statistical Review* **75** n.º 2, p. 127–130.
- BERNARDO, J. & Smith, A. (1994). *Bayesian Theory*, John Wiley & Sons, England. (Reimpresso em 2001)
- BERNARDO, J. (1996). The Concept of Exchangeability and its Applications, *Far East J. Mathematical Sciences* **4**, p. 111–121.
- BERNARDO, J. (1998). Bruno de Finetti en la Estadística Contemporánea, *Historia de la Matemática en el Siglo XX* (S. Ríos, ed), Real Academia de Ciências, Madrid, p. 63–80.
- BERNARDO, J. (2006). Bayesian Statistics, *The New Palgrave Dictionary of Economics*, 2nd ed., S. N. Durlauf and L. E. Blume eds., New York.
- BERNOULLI, J. (1713). *Ars Conjectandi*, Basle. (Translated into English by E. D. Sylla, *The Art of Conjecturing*, The Johns Hopkins University press, 2006)
- BERNSTEIN, S. (1917). On the Axiomatic Foundation of the Theory of Probability, *Communications of the Kharkiv Mathematical Society* **15**, p. 209–274.
- BERTRAND, J. (1888). *Calcul des Probabilités*, Gauthier–Villars, Paris. (2^a ed. 1907)
- BILLINGSLEY, P. (1986). *Probability & Measure*, 2nd. ed., John Wiley & Sons, New York.
- BINGHAM, N. (2000). Studies in the History of Probability and Statistics XLVI. Measure into Probability: from Lebesgue to Kolmogorov, *Biometrika* **87** n.º 1, p. 145–156.

- BINGHAM, N. (2007). Regular Variation and Probability: The Early Years, *Journal of Computational and Applied Mathematics* **200**, p. 357–363.
- BOOLE, G. (1854). *An Investigation of the Laws of Thought: On which are Founded the Mathematical Theories of Logic and Probabilities*, Walton and Maberly, Cambridge.
- BOREL, É. (1909). *Éléments de la Théorie des Probabilités*, Éditions Albin Michel, Paris. (2^a ed. 1910)
- BOREL, É. (1909a). Les Probabilités Dénombrables et leurs Applications Arithmétiques, *Rendiconti del Circolo Matematico di Palermo* t. **27**, p. 247–276. (Reimpresso em Borel (1972), Vol II, p. 1055–1079)
- BOREL, É. (1914). *Le Hasard*, Librairie Félix Alcan, Paris.
- BOREL, É. (1924). “Apropos of a Treatise on Probability” em Kyburg & Smokler (1964) p. 45–60.
- BOREL, É. (1972). *Œuvres de Emile Borel*, four volumes, éditions du Centre National de la Recherche Scientifique, Paris.
- BOX, G. & MÜLLER, M. (1958). A Note on the Generation of Random Normal Deviates, *The Annals of Mathematical Statistics* **29** n.º 2, p. 610–611.
- BROGGI, U. (1907). *Traité des Assurances sur la Vie: Avec Développement sur le Calcul des Probabilités*, Librairie Scientifique A. Hermann, Paris. (Traduit de l’Italien par S. Lattès)
- BRU, B. (2005). Poisson, the Probability Calculus, and Public Education, *Journ@l Electronique d’Histoire des Probabilités et de la Statistique* **1**, n.º 2.
- BRU, B. (2006). Les Leçons de Calcul des Probabilités de Joseph Bertrand «Les Lois du Hasard», *Journ@l Electronique d’Histoire des Probabilités et de la Statistique* **2**, n.º 2.

- BUFFON, G. (1733). Geometrie, *Histoire de L'Académie Royale des Sciences*, p. 43–45.
- BUFFON, G. (1777). “Essai d’Arithmétique Morale” in *Suppléments à l’Histoire Naturelle Générale et Particulière* **4**, p. 46–123, Paris.
- LE CAM, L. (1986). The Central Limit Theorem Around 1935, *Statistical Science* **1** n.º 1, p. 78–96.
- CARVALHO, J. & GOMES, A. (1994). Diogo Pacheco d’Amorim: O Professor e o Cidadão, *Actas do II Congresso Anual da Sociedade Portuguesa de Estatística*, p. 49–54.
- CARVALLO, E. (1912). *Le Calcul des Probabilités et ses Applications*, Gauthier–Villars, Paris.
- CASTELNUOVO, G. (1919). *Calcolo delle Probabilità*, Albrighi, Segati & C., Roma.
- CHARPENTIER, É.; LESNE, A. & NIKOLSKI, N. (2007). *Kolmogorov’s Heritage in Mathematics*, Springer–Verlag, New York.
- CHEN, R. (1977). On Almost Sure Convergence in a Finitely Additive Setting, *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* **37**, p. 341–356.
- CHUNG, K. L. (2000). *Pólya Work in Probability*. Publicado como anexo de G. Alexanderson (2000).
- CIFARELLI, D. & REGAZZINI, E. (1996). De Finetti’s Contribution to Probability and Statistics, *Statistical Science* **11** n.º 4, p. 253–282.
- CONDORCET, M. (1785). *Essai sur l’Application de l’Analyse à la Probabilité des Décisions Rendues à la Pluralité des Voix*, Paris.
- COURNOT, A. A. (1838). Mémoire sur les Applications du Calcul des Chances à la Statistique Judiciaire, *Journal des Mathématiques Pures et Appliquées* **12** t. 3, p. 257–334.

- COURNOT, A. A. (1843). *Exposition de la Théorie des Chances et des Probabilités*, Librairie de L. Hachette, Paris.
- COURTAULT *et al.* (2000). Louis Bachelier, on the Centenary of *Théorie de la Spéculation*, *Mathematical Finance* **10** n.º 3, p. 341–353.
- COX, R. T. (1946). Probability, Frequency and Reasonable Expectation, *American Journal Physics* **14** n.º 1, p. 1–13.
- CRAMÉR, H. (1976). Half a Century with Probability Theory: Some Personal Recollections, *The Annals of Probability* **4**, p. 509–546.
- CZUBER, E. (1902). *Probabilités et Moyennes Géométriques*, Librairie Scientifique A. Hermann, Paris. (Tradução de Czuber (1884) *Geometrische Wahrscheinlichkeiten und Mittelwerte*, Leipzig, Verlag)
- DAVID, F. N. (1962). *Games, Gods and Gambling: a History of Probability and Statistical Ideas*, Charles Griffin & Co., London.
- DAVID, H. A. & EDWARDS, A. W. (2001). *Annotated Readings in the History of Statistics*, Springer-Verlag, New York.
- DROESBEKE, J. (2005). La Place de l’Enseignement dans la Vie et l’Oeuvre de Quételet, *Journ@l Electronique d’Histoire des Probabilités et de la Statistique* **1** n.º 2.
- DUDLEY, R. M. (2002). *Real Analysis and Probability*, Cambridge University Press, Cambridge.
- DURRET, R. (1996). *Probability: Theory and Examples*, 2nd ed., Duxbury Press, Belmont.
- EDWARDS, A. W. (1983). Pascal’s Problem: The ‘Gambler’s Ruin’, *International Statistical Review* **51** n.º 1, p. 73–79.
- FELLER, W. (1968). *An Introduction to Probability Theory and its Applications*, Vol. I, 3 ed., John Wiley & Sons, New York.

- FELLER, W. (1971). *An Introduction to Probability Theory and its Applications*, Vol. II, 3 ed., John Wiley & Sons, New York.
- DE FINETTI, B. (1937). La Prévision: ses Lois Logiques, ses Sources Subjectives, *Annales de l'Institut Henri Poincaré* **7** n.º 1, p. 1–68.
- DE FINETTI, B. (1974). *Theory of Probability*, Vol. 1, John Wiley & Sons, New York.
- FISHBURN, P. (1986). The Axioms of Subjective Probability, *Statistical Science* **1** n.º 3, p. 335–358.
- FISHER, R. A. (1922). On the Mathematical Foundations of Theoretical Statistics. *Philosophical Transactions of the Royal Society*, p. 309–368.
- FISHER, R. A. (1925). Theory of Statistical Estimation. *Proceedings of the Cambridge Philosophical Society* **22**, p. 700–725.
- FISHER, R. A. (1930). Inverse Probability. *Proceedings of the Cambridge Philosophical Society* **26**, p. 528–535.
- FISHER, R. A. (1935). The Fiducial Argument in Statistical Inference, *Annals of Eugenics* **6**, p. 391–398.
- FISHER, R. A. (1990). *Statistical Methods, Experimental Design, and Scientific Inference: A Re-issue of Statistical Methods for Research Workers, the Design of Experiments, and Statistical Methods and Scientific Inference*; edited by J. H. Bennett, with a foreword by F. Yates, Oxford: Oxford University Press. (includi *Statistical Methods for Research Workers*, *The Design of Experiments* e *Statistical Methods and Scientific Inference*)
- GALTON, F. (1885). Regression Towards Mediocrity in Hereditary Stature, *Journal Antropological Inst.* **15**.
- GNEDENKO, B. V. & KOLMOGOROV, A. N. (1954). *Limit Distributions for Sums of Independent Random Variables*, Addison-Wesley Publishing

- Company (translated from the Russian, annotated, and revised by Kai Lai Chung).
- GOURAUD, C. (1848). *Histoire des Calculs des Probabilités Depuis ses Origines Jusqu'à nos Jours*, Librairie D'Auguste Durand, Paris.
- GUIMARÃES, R. (1904). *Noções sobre Cálculo das Probabilidades, Theoria dos Erros e Método dos Mínimos Quadrados*, Biblioteca do Povo e das Escolas 223, Lisboa.
- HACKING, I. (1975). *The Emergence of Probability: a Philosophical Study of Early Ideas about Probability Introduction and Statistical Inference*, Cambridge University Press, Cambridge.
- HALD, A. (1998). *A History Mathematical Statistics from 1750 to 1930*, John Wiley & Sons, New York.
- HALD, A. (2003). *History of Probability and Statistics and their Applications Before 1750*, John Wiley & Sons, New Jersey.
- HEYDE, C. C. & SENETA, E. (2001). *Statisticians of the Centuries*, Springer-Verlag, New York.
- HILBERT, D. (1902). Mathematical Problems, *Bulletin of the American Mathematical Society* **8**, p. 437–479. (Reprint in Hilbert, D. (2000). *Bulletin of the American Mathematical Society* **37** n.º 4, p. 407–436.)
- HOLGATE, P. (1981). Studies in the History of Probability and Statistics XXXIX Buffon's Cycloid, *Biometrika* **68** n.º 3, p. 712–716.
- HUYGENS, C. (1657). *Libellus de Ratiociniis in Ludo Aleae*, translation by W. Browne in 1714.
- JAYNES, E. T. (1996). *Probability Theory: The Logic of Science*.
- JEFFREYS, H. (1931). *Scientific Inference*, Cambridge University Press, Cambridge. (3ª ed. de 1973)

- JEFFREYS, H. (1939). *Theory of Probability*, Oxford University Press, Oxford. (2^a ed. de 1948)
- JOHNSON, N. L.; KOTZ, S. & BALAKRISHNAN, N. (1994). *Continuous Univariate Distributions* Vol. 1, 2nd. ed., John Wiley & Sons, Canada.
- KAHANE, J. (1998). Le Mouvement Brownien: Un Essai sur les Origines de la Théorie Mathématique, *Matériaux pour l'Histoire des Mathématiques au XX^e Siècle, Actes du Colloque à la Mémoire de Jean Dieudonné (Nice, 1996)* **3**, p. 123–155. Société Mathématique de France.
- KEYNES, J. M. (1921). *A Treatise on Probability*, MacMillan and Company, London. (2^a ed. de 1943)
- KING, A. & READ, C. (1963). *Pathways to Probability: History of the Mathematics of Certainty and Chance*, Holt, Rinehart and Winston, New York.
- KOLMOGOROFF, A. N. (1933). *Grundbegriffe der Wahrscheinlichkeitstheorie*, Ergebnisse der Mathematik, Berlin. (Trad. Inglesa: Kolmogoroff, A. N. (1956). *Foundations of the Theory of Probability*, Chelsea, New York.)
- KYBURG, H. and SMOKLER, H. (1964). *Studies in Subjective Probability*, John Wiley & Sons, U. S. A.. (2^a ed. de 1980)
- LACROIX, S. F. (1816). *Traité Élémentaire du Calcul des Probabilités*, Bachelier, Paris. (2^a ed. de 1822)
- LAPLACE, P. S. (1774). Mémoire sur la Probabilité des Causes par les Événemens. (Trad. Inglesa: Stigler, S. M. (1986). Laplace's 1774 Memoir on Inverse Probability, *Statistical Science* **1** n.º 3, p. 359–378)
- LAPLACE, P. S. (1812). *Théorie Analytique des Probabilités*, Libraire pour les Mathématiques, Paris. (2^a ed. de 1814)
- LAURENT, H. (1908). *Statistique Mathématique*, Octave Doin, Paris.

- LEBESGUE, H. (1902). Intégrale, Longueur, Aire, *Annali di Matematica Pura ed Applicata* **7**, p. 231–359.
- LEHMANN, E. L. (1993). The Fisher, Neyman-Pearson Theories of Testing Hypotheses: One Theory or Two?, *Journal of the American Statistical Association* **88**, p. 1242–1249.
- LEHMANN, E. L. (1999). “Student” and Small-Sample Theory, *Statistical Science* **14**, p. 418–426.
- LEMOINE, E. (1872). Sur une Question de Probabilités, *Bulletin de la Société Mathématique de France* **1**, p. 39–40.
- LEMOINE, E. (1883). Quelques Questions de Probabilités Résolues Géométriquement, *Bulletin de la Société Mathématique de France* **11**, p. 13–25.
- LÉVY, P. (1925). *Calcul des Probabilités*, Gauthier-Villars, Paris.
- LÉVY, P. (1970). *Quelques Aspects de la Pensée d’un Mathématicien*, Blanchard, Paris.
- LOÈVE, M. (1973). Paul Lévy, 1886–1971, *The Annals of Probability* **1** n.º 1, p. 1–8.
- LOMNICKI, A. (1923). Nouveaux Fondements du Calcul des Probabilités (Définition de la Probabilité Fondée sur la Théorie des Ensembles). *Fundamenta Mathematicae* **4**, p. 34–71.
- LUKACS, E. (1970). *Characteristic Functions*, Charles Griffin & Company Limited, London.
- MACHADO, J. P. (1990). *Dicionário Etimológico da Língua Portuguesa*, Volume I, 6ª edição, Livros Horizonte.
- MAISTROV, L. E. (1974). *Probability Theory: a Historical Sketch*, Academic Press, New York.

- MARKOFF, A. (1913). *Démonstration du Second Théorème-limite du Calcul des Probabilités par la Méthode des Moments*, Imprimerie de L'Académie Impériale des Sciences, St.-Petersbourg.
- VON MISES, R. (1919). Grundlagen der Wahrscheinlichkeitsrechnung, *Mathematische Zeitschrift* **5**, p. 52–99.
- VON MISES, R. (1932). Théorie des Probabilités. Fondements et Applications, *Annales de l'Institut Henri Poincaré* **3** n.º 2, p. 137–190.
- VON MISES, R. (1938). Quelques Remarques sur les Fondements du Calcul des Probabilités, *Colloque D'Octobre 1937 sur la Théorie Des Probabilités*, M. Fréchet ed., p. 57–66.
- VON MISES, R. (1946). *Notes on Mathematical Theory of Probability and Statistics*, Harvard Graduate School of Engineering.
- VON MISES, R. (1964). *Mathematical Theory of Probability and Statistics*, Academic Press, New York.
- DE MOIVRE, A. (1718). *The Doctrine of Chances or, a Method of Calculating the Probabilities of Events in Play*, London. (3ª ed. de 1756)
- MONTESSUS DE BALLORE, R. (1908). *Leçons Élémentaires sur le Calcul des Probabilités*, Gauthier–Villars, Paris.
- MONTMORT, P. R. (1708). *Essay d'Analyse sur les Jeux de Hazard*, Jacques Quillau, Paris. (2ª ed. de 1713)
- DE MORGAN, A. (1847). *Formal Logic: or, The Calculus of Inference, Necessary and Probable*, Taylor and Walton, London.
- PACHECO D'AMORIM, D. (1914). *Elementos de Cálculo das Probabilidades*, Tese de doutoramento, Imprensa da Universidade de Coimbra, Coimbra. (A tradução para inglês, juntamente com original, editado por S. Mendonça, D. Pestana e R. Santos está disponível em <http://www.estg.ipleiria.pt/~rsantos>)

- PACHECO D'AMORIM, D. (1920). *Compêndio de Geometria*, Volume 1.^o — Classes 1.^a e 2.^a, França & Arménio, Coimbra.
- PACHECO D'AMORIM, D. (1956–57). *Cálculo das Probabilidades*, Imprensa da Universidade de Coimbra, Coimbra.
- PACHECO D'AMORIM, D. (2002). Oração de Sapientia Proferida na Abertura Solene da Universidade de Coimbra no Ano Lectivo de 1923–1924 Pelo Doutor Diogo Pacheco d'Amorim, in *Antologia de Textos Essenciais sobre a História da Matemática em Portugal*, Sociedade Portuguesa de Matemática, Lisboa.
- PACHECO D'AMORIM, D. (2004). *Compêndio de Geometria*, Sociedade Portuguesa de Matemática, Lisboa. (Fac-simile da obra “Compêndio de Geometria”, 9^a ed., Coimbra Editora, 1943)
- PAES, S. (1898). *Introdução à Teoria dos Erros das Observações*, Tese de Doutoramento, Faculdade de Matemática, Universidade de Coimbra.
- PASCAL, B. (1654). “Correspondance avec Fermat” in *Les Lettres de Blaise Pascal: Accompagnées de Lettres de ses Correspondants* publiées par Maurice Beaufreton (1922), Éditeur Scientifique, Paris.
- PAULINO, C.; TURKMAN, M. e MURTEIRA, B. (2003). *Estatística Bayesiana*, Fundação Calouste Gulbenkian, Lisboa.
- PEARSON, K. (1892). *The Grammar of Science*, Adam and Charles Black, London. (2^a ed. de 1900)
- PEARSON, K. (1895). Contributions to the Mathematical Theory of Evolution II. Skew Variations in Homogeneous Material, *Philosophical Transactions of the Royal Society of London*, Series A **186**, p. 343–414.
- PEARSON, K. (1900). On the Criterion that a Given System of Deviations from the Probable in the Case of Correlated System of Variables is such that it can be Reasonably Supposed to have Arisen from Random Sampling, *Philosophical Magazine* **50**, p. 157–175.

- PESTANA, D. D. (1994). Diogo Pacheco d'Amorim: Um Vulto Maior na História da Teoria das Probabilidades, *Actas do II Congresso Anual da Sociedade Portuguesa de Estatística*, p. 55–63.
- PESTANA, D. D. & VELOSA, S. F. (2001). Diogo Pacheco d'Amorim, *International Center for Mathematics Bulletin* **11**, p. 22–24.
- PESTANA, D. D. & VELOSA, S. F. (2002). *Introdução à Probabilidade e à Estatística*, Vol. 1, Fundação Calouste Gulbenkian, Lisboa. (2ª ed. de 2006)
- PINTO, J. F. (1913). *Noções de Calculo das Probabilidades para o Estabelecimento das Bases da Estatística*, Imprensa da Universidade, Coimbra. (publicado nos *Annaes da Academia Polytechnica do Porto* **8** (1913))
- VON PLATO, J. (1994). *Creating Modern Probability*, Cambridge University Press, Cambridge.
- POINCARÉ, H. (1896). *Calcul des Probabilités*, Gauthier–Villars, Paris. (2ª ed. de 1912)
- POINCARÉ, H. (1902). *La Science et l'Hypothèse*, Flammarion, Paris.
- POISSON, S. D. (1837). *Recherches sur la Probabilité des Jugements: en Matière Criminelle et en Matière Civile, Précédées des Règles Générales du Calcul des Probabilités*, Bachelier, Paris.
- PÓLYA, G. (1920). Über Den Zentralen Grenzwertsatz der Wahrscheinlichkeitsrechnung und Das Momentenproblem, *Math. Zeit.* **8**, p. 171–181.
- POPPER, K. R. (1959). The Propensity Interpretation of Probability, *The British Journal for the Philosophy of Science* **10**, p. 25–42.
- RAMSEY, F. P. (1926). “Truth and Probability” em Ramsey, F. P. (1931) *The Foundations of Mathematics and Other Logical Essays*, Cap. VII, p. 156–198, ed. R. Braithwaite, Kegan Paul, London.

- RAO, M. M. (1988). Paradoxes in Conditional Probability, *Journal of Multivariate Analysis* **27**, p. 434–446.
- REGAZZINI, E. (2005). Probability and Statistics in Italy During the First World War I: Cantelli and the Laws of Large Numbers, *Electronic Journ@l for History of Probability and Statistics* **1** n.º 1, p. 1–12.
- REIS, M. (1929). *Cálculo das Probabilidades*, Tese de doutoramento, Imprensa da Universidade de Coimbra, Coimbra.
- RÉNYI, A. (1955). On a New Axiomatic Theory of Probability, *Acta Math. Acad. Sci. Hung.* **6**, p. 285–335.
- RÉNYI, A. (1970). *Probability Theory*, North-Holland series in Applied Mathematics and Mechanics, Amesterdam.
- RODRIGUES, M. A. (direcção) (1992). *Memoria Professorum Vniversitatis Conimbrigensis 1772–1937*, Arquivo da Universidade de Coimbra, Coimbra.
- SANTOS, R. (2005). Probabilidade *Circa* 1914 e os *Elementos de Cálculo das Probabilidades* de Diogo Pacheco d’Amorim, *Boletim da Sociedade Portuguesa de Matemática* **53**, p. 127–144.
- SANTOS, R. (2007). Diogo Pacheco d’Amorim’s Probability Calculus, *Bulletin of the 56th Session of the International Statistical Institute, Proceedings*.
- SAVAGE, L. J. (1954). *The Foundations of Statistics*, John Wiley & Sons, Canada.
- SENETA, E.; PARSHALL, K. & JONGMANS, F. (2001). Nineteenth-Century Developments in Geometric Probability: J. J. Sylvester, M. W. Crofton, J.-É. Barbier, and J. Bertrand, *Archive History Exact Sciences* **55**, p. 501–524.

- SENETA, E. (2002). Karamata's Characterization Theorem, Feller, and Regular Variation in Probability Theory, *Publications de l'Institut Mathématique* **71**, p. 79–89.
- SHAFFER, G. & VOVK, V. (2005). The Origins and Legacy of Kolmogorov's Grundbegriffe, *The Game-Theoretic Probability and Finance Project Working Paper* N°. 4.
- SHAFFER, G. & VOVK, V. (2006). The Sources of Kolmogorov's Grundbegriffe, *Statistical Science* **21** n.º 1, p. 70–98.
- SHANNON, C. E. (1948). A Mathematical Theory of Communication, *The Bell System Technical Journal* **27**, p. 379–423, 623–659.
- SPANOS, A. (1999). *Probability Theory and Statistical Inference*, Cambridge University Press, Cambridge.
- STEINHAUS, H. (1923). Les Probabilités Dénombrables et leur Rapport à la Théorie de la Mesure, *Fundamenta Mathematicae* **4**, p. 286–310.
- STIGLER, S. M. (1978). Francis Ysidro Edgeworth, Statistician, *Journal of the Royal Statistical Society, Series A* **141** n.º 3, p. 287–322.
- STIGLER, S. M. (1980). Stigler's Law of Eponymy. *Transactions of the New York Academy of Sciences* **39**, p. 147–158.
- STIGLER, S. M. (1986). *The History of Statistics: The Measurement of Uncertainty Before 1900*, The Belknap Press of Harvard University Press, England.
- STIGLER, S. M. (1991). Stochastic Simulation in the Nineteenth Century, *Statistical Science* **6** n.º 1, p. 89–97.
- STUDENT (1908). The Probably Error of a Mean, *Biometrika* **6**, p. 1–25.
- SZÉKELY, G. J. (1986). *Paradoxes in Probability Theory and Mathematical Statistics*, D. Reidel Publishing Company, Budapest.

- TIAGO DE OLIVEIRA, J. (1991). PhDs in Portugal: 1898–1989, *Bulletin of the Institute of Mathematical Statistics* **20** n.º 6, reproduzido em Tiago de Oliveira (1995) p. 215–217.
- TIAGO DE OLIVEIRA, J. (1991a). Acaso, Determinismo e Indução, *Análise* **15**, reproduzido em Tiago de Oliveira (1995) p. 179–213.
- TIAGO DE OLIVEIRA, J. (1995). *Collected Works/Obras J. Tiago de Oliveira*, Volume II, Pendor, Évora.
- TODHUNTER, I. (1865). *A History of the Mathematical Theory of Probability from the Time of Pascal to that of Laplace*, Cambridge, London. (Reimpressão de 1949 da Chelsea Publishing Company)
- USPENSKY, J. V. (1937). *Introduction to Mathematical Probability*, McGraw–Hill, New York.
- VELOSA, S. F. (2003). *O Problema de Behrens-Fisher*, Escolar Editora, Lisboa.
- VENN, J. (1866). *The logic of chance*, MacMillan, London. (2ª ed. de 1876)
- VIANA, M. (2007). *Conjectura de Poincaré Geometria para entender o Universo*, slides da conferência efectuada em 12 de Dezembro na Fundação Gulbenkian, Lisboa. (<http://w3.impa.br/~viana>)
- WILLIAMSON, J. (1999). Countable Additivity and Subjective Probability, *The British Journal for the Philosophy of Science* **50**, p. 401–416.