

UNIVERSIDADE DE LISBOA
FACULDADE DE CIÊNCIAS
DEPARTAMENTO DE ESTATÍSTICA E INVESTIGAÇÃO
OPERACIONAL



FEIRA DOS MOMENTOS
Planeamento Experimental e Investigação de
Localização e Escala em Populações não Gaussianas

João Paulo Oliveira Martins

DOUTORAMENTO EM ESTATÍSTICA E INVESTIGAÇÃO
OPERACIONAL

Probabilidades e Estatística

2008

UNIVERSIDADE DE LISBOA
FACULDADE DE CIÊNCIAS
DEPARTAMENTO DE ESTATÍSTICA E INVESTIGAÇÃO
OPERACIONAL



FEIRA DOS MOMENTOS
Planeamento Experimental e Investigação de
Localização e Escala em Populações não Gaussianas

João Paulo Oliveira Martins

Tese orientada por:

Professor Doutor Dinis Duarte Ferreira Pestana

Professora Doutora Sandra Maria Freitas Mendonça

**DOUTORAMENTO EM ESTATÍSTICA E INVESTIGAÇÃO
OPERACIONAL**

Probabilidades e Estatística

2008

Agradecimentos

Aos meus orientadores Professor Doutor Dinis Pestana e Professora Doutora Sandra Mendonça, pelo tempo e disponibilidade demonstrados para me orientarem nesta "última" etapa do percurso académico. Ao longo do tempo em que decorreu a elaboração deste trabalho reforçou-se uma relação que não foi apenas profissional mas também de amizade. Espero que a relação que tem sido tão profícua até ao momento continue a prolongar-se no tempo.

Ao Instituto Politécnico de Leiria pelas condições criadas para que a realização deste trabalho fosse possível e aos meus colegas com quem me apraz trabalhar.

Agradeço igualmente ao Centro de Estatística e Aplicações da Universidade de Lisboa pelas facilidades concedidas em termos de participação em eventos e pelos meios computacionais e documentais disponibilizados. Também agradeço à Fundação para a Ciência e Tecnologia pelo apoio financeiro prestado através de uma bolsa de doutoramento que me foi concedida.

Por fim, mas nem por isso menos importante, os meus agradecimentos vão para a minha família. Os meus pais e as minhas irmãs que são muito importantes naquilo que hoje sou e à Sofia que como ninguém ficou ciente das exigências de um trabalho de investigação. Agradeço também aos meus amigos que me apoiaram durante esta etapa pois só um bem estar de espírito possibilita um bom desempenho profissional.

A todos a minha gratidão.

João Paulo Oliveira Martins

Vila Nova de Famalicão, Novembro de 2008

Resumo

A primeira tarefa levada a cabo no âmbito deste trabalho consistiu no estudo da estimação dos coeficientes de um modelo de regressão polinomial por um planeamento óptimo. Os desenvolvimentos clássicos relativos a planeamentos discriminantes óptimos e planeamentos robustos óptimos serviram de inspiração para a definição de planeamentos mistos óptimos, que têm em conta quer a estimação do grau da regressão bem como dos seus coeficientes, sendo quase óptimos no que se refere aos critérios discriminante e robusto. O caso dos planeamentos mistos óptimos até grau 4 é caracterizado detalhadamente, e a investigação computacional mostra que a perda de eficiência comparativamente com os planeamentos discriminantes óptimos e robustos óptimos é inferior a 2%, enquanto a perda de eficiência dos planeamentos robustos óptimos comparada com os planeamentos discriminantes óptimos, ou vice-versa, pode atingir os 15%. A teoria dos momentos canónicos serve de suporte à apresentação dos resultados relativos a planeamentos óptimos.

Discutem-se as truncaturas da série de Taylor sugeridas pela aplicação do método delta. São apresentadas algumas extensões e aplica-se o método à avaliação da variância da soma de n variáveis aleatórias eventualmente correlacionadas, um problema com aplicações ao nível da Química. Ainda neste contexto, são consideradas as transformações estabilizadoras da variância apresentado-se uma extensão da definição para variáveis aleatórias univariadas com n parâmetros desconhecidos. O caso multivariado é também abordado e são dadas algumas pistas para compreender as condições para a existência ou não dessas transformações.

O método delta é também usado para estimar os primeiros quatro momentos da estatística de Student T_n . Mostramos que o comportamento da distribuição de T_n é próximo do comportamento de uma distribuição do tipo IV do sistema de Pearson.

Finalizamos, enfatizando o papel da assimetria na atracção e repulsão da média amostral e variância amostral.

Palavras-chave: planeamentos óptimos, momentos canónicos, método delta, transformações estabilizadoras da variância, estatística de Student, assimetria.

Abstract

The first aim of our work has been to investigate optimal designs to estimate coefficients in a polynomial regression model. Classical developments on optimal discriminant designs and on optimal robust designs inspired our definition of optimal mixed designs, which take into account the regression degree and the regression coefficients estimation simultaneously, being quasi-optimal in regards both to discriminant and to robustness criteria. The case of optimal mixed designs in polynomial regression up to degree 4 is thoroughly described, and intensive computational investigation shows that their loss of efficiency when compared to optimal discriminant or optimal robust designs is bounded by 2%, while the loss of efficiency of optimal robust when compared to discriminant designs, or vice-versa, can be as high as 15%. Canonical moments theory is used throughout to make a neat presentation of results on the optimality of designs.

The Taylor series truncation which is at the core of the delta method is discussed. Some extensions are presented and the method is applied to evaluate the standard deviation of the n eventually correlated random variables, a problem with interesting applications in Chemistry. Still in this context, stabilizing variance transformations are analysed and an extension of this concept to univariate random variables with n unknown parameters is presented. We also deal with the multivariate case and some hints are given to understand conditions for the existence of such transformations.

The delta method is also applied to estimate the first four moments of the Student t -statistic T_n . We show that T_n distribution behaviour is very close to the behaviour of some type IV distribution of the Pearson system. We finish emphasizing the asymmetry role in the attraction and repulsion of the sample mean and sample variance.

Keywords: optimal designs, canonical moments, delta method, stabilizing variance transformations, Student t -statistic, asymmetry.

Conteúdo

Introdução	3
1 Momentos	6
1.1 Momentos ordinários e cumulantes	6
1.2 Algumas propriedades de matrizes	11
1.3 Momentos Canônicos	17
1.3.1 Variáveis aleatórias com suporte não limitado	18
2 Planeamento de experiências	25
2.1 Planeamentos ótimos	25
2.2 Planeamentos D_1 -ótimo e D -ótimo	31
2.3 Planeamentos robustos de modelos e discriminantes	36
3 Transformações e o método delta	51
3.1 Porquê transformar?	51
3.2 Algumas transformações de variáveis aleatórias	53
3.3 Derivadas e séries de Taylor	55
3.4 Método delta	57
3.5 Extensões do método delta	61
3.6 Transformações estabilizadoras da variância	67
3.6.1 Definições	67
3.6.2 Caso univariado com n parâmetros desconhecidos	78
3.6.3 Caso multivariado	80
3.7 Uma aplicação interdisciplinar	91
3.7.1 Alguns conceitos químicos	92
3.7.2 Desvio padrão da soma de 2 variáveis aleatórias	93
3.7.3 <i>Clusters</i> de partículas	96

4	Inferência utilizando a estatística T_n	100
4.1	Leis estáveis para somas	100
4.2	Divisibilidade infinita	102
4.3	Leis limites da estatística $S_n(p)$	107
4.4	Momentos da estatística T_n	109
4.4.1	$E(T_n^2)$	110
4.4.2	$E(T_n^3)$	113
4.4.3	$E(T_n^4)$	116
4.4.4	Alguns resultados de simulação	121
4.5	Sistema de distribuições de Pearson	126
4.6	Distribuições tipo IV	129
4.7	Distribuição de T_n	133
4.8	Gaussiana generalizada, uma alternativa?	142
4.9	Expansões de Edgeworth	148
4.10	O papel da (as)simetria	152
	Bibliografia	155
	Anexos	163
	A. Dos momentos ordinários aos momentos canônicos	163
	B. Planeamentos mistos Θ_0^π -ótimos	175

Introdução

No prefácio do eminente tratado de Gnedenko e Kolmogorov (1954) sobre somas de variáveis aleatórias independentes está claramente expresso que nos resultados limite reside a essência da contribuição da estatística para a construção do conhecimento. De facto, a importância dos resultados assintóticos não se limita ao fascínio que exercem por irem à essência das coisas, permitindo-nos apreciar uma estrutura regular subjacente ao aparente caos de tudo aquilo que nos rodeia, como são o respaldo de muitas das aproximações a que constantemente recorremos no nosso trabalho profissional. Recorde-se o que outro génio da Estatística, Tukey, afirmou com a sua habitual eloquência: “mais vale uma solução aproximada de um problema claramente formulado, do que uma solução exacta de um problema formulado aproximadamente”, (tradução um pouco livre, obviamente).

Desde os primórdios do desenvolvimento da Estatística Matemática, as expansões em série (Gram-Chalier, Edgeworth) exibiram a fecundidade do uso de momentos; embora de interpretação menos imediata, os cumulantes de Thiele-Fisher são ainda mais versáteis neste contexto, e desde meados do século passado que foram explorados outros parâmetros, como os momentos canónicos e outros parâmetros que lhes estão intimamente associados e que porventura têm um papel mais intrínseco ainda na caracterização de medidas com suporte limitado. Fornecer uma panorâmica sobre a natureza destas quantidades e a sua utilidade na caracterização de uma função distribuição é o objectivo de uma parte do primeiro capítulo desta dissertação.

Os momentos canónicos terão um papel crucial no estudo de planeamento de experiências com suporte num subconjunto de um intervalo fechado. No capítulo 2, através da maximização de determinantes de Hankel, via momentos canónicos, apresentamos um novo tipo de planeamentos que designamos por planeamentos mistos. Assumindo que estamos em condições de utilizar regressão polinomial

de grau, no máximo, m podemos com a recolha de uma única amostra, estimar o grau da regressão bem como todos os parâmetros do modelo de uma forma quase óptima.

Pensamos que um título — porventura ambicioso — que melhor transmitisse o cerne das nossas preocupações deveria incluir uma referência a resultados assintóticos, ou a métodos de os obter; em particular, o método delta, que assenta na truncatura da expansão em série de Taylor, é um clássico na investigação de parâmetros de funções de variáveis aleatórias, mas também uma ferramenta imprescindível quando se pretende transformar dados de forma a regularizá-los em alguma acepção precisa, por exemplo por forma a estabilizar a variância. Iniciamos o capítulo 3 dando enfoque ao tópico, que nos parece central e merecer maior relevo no ensino da Estatística, das transformações de dados que viabilizem análises profícuas. Continuamos apresentando algumas extensões ao método delta, nomeadamente, analisando a convergência de aproximações resultantes de truncaturas que envolvam diversos termos. Neste contexto, aparecem as transformações estabilizadoras da variância como meio fundamental na obtenção de variáveis fulcrais — ferramenta primordial da Inferência Estatística. Terminamos, apresentado um exemplo de aplicação do método delta discutindo a variância da soma de k variáveis aleatórias. Este problema parece assumir importância a nível da Química (e não só)

A parte final da nossa investigação incidiu sobre o papel da simetria e do peso das caudas — caracterizáveis através dos 3.º e 4.º cumulantes padronizados, respectivamente — na qualidade das aproximações. Estas duas quantidades foram especialmente relevantes no capítulo 4 em que é apresentado um estudo sobre a função de distribuição da estatística de Student para uma parente não gaussiana. No decurso desta análise, que recorre ao sistema de distribuições que Pearson (1985) apresentou, sobressai o papel do achatamento e principalmente da assimetria na avaliação dos primeiros quatro momentos ordinários. Verificamos também que a

atração/repulsão da média amostral e da variância amostral está associada ao sinal do 3.º cumulante (ou do coeficiente de assimetria) e lançamos algumas conjecturas para uma futura investigação.

Em suma, a escolha do título “A Feira dos Momentos”, pretende transmitir desde logo que, ainda que haja uma linha condutora, os resultados assintóticos e os métodos que usamos para os deduzir recorrem a um variado espectro de valores esperados e funções de valores esperados, que com alguma latitude de linguagem podemos apodar de momentos e de transformações integrais.

1 Momentos

1.1 Momentos ordinários e cumulantes

Seja X uma variável aleatória com função de distribuição denotada por F_X .

Definição 1.1 Dado $k \in \mathbb{N}_0$ designa-se por **momento de ordem k da variável aleatória X** , representando-se por μ'_k ou por $E(X^k)$, o integral

$$\mu'_k = E(X^k) = \int_{\mathbb{R}} x^k dF_X(x)$$

se este for absolutamente convergente, isto é, se

$$\int_{\mathbb{R}} |x|^k dF_X(x) < \infty.$$

A definição de momento (ordinário) de uma variável aleatória também pode ser estendido a uma transformação de uma variável aleatória. Assim, se $g : \mathbb{R} \rightarrow \mathbb{R}$ for uma função Borel-mensurável então o momento de ordem k da variável aleatória $Y = g(X)$ é

$$E(Y^k) = \int_{\mathbb{R}} (g(x))^k dF_X(x)$$

caso

$$\int_{\mathbb{R}} |g(x)|^k dF_X(x) < \infty.$$

Tendo em conta a Definição 1.1, o momento de ordem 0 é igual a 1 e o momento de ordem 1 corresponde ao valor esperado, que denotaremos por μ . Note-se que nem todas as variáveis aleatórias admitem momentos de todas as ordens e que se uma variável aleatória admite o momento de ordem k então admite os momentos de todas as ordens inferiores a k . A existência de momentos depende do comportamento das caudas da função de distribuição. De facto,

$$|x|^\alpha P(X > x) \xrightarrow{x \rightarrow +\infty} 0 \iff \exists E(|X|^\alpha).$$

Até ao final deste trabalho, assumir-se-á a existência de todos os momentos referidos em cada ocasião.

Definição 1.2 *Sejam $k \in \mathbb{N}_0$ e X uma variável aleatória com função de distribuição F_X . Designa-se por **momento centrado de ordem k da variável aleatória X** , representando-se por μ_k ou por $E[(X - \mu)^k]$, o integral*

$$\mu_k = E[(X - \mu)^k] = \int_{\mathbb{R}} (x - \mu)^k dF_X(x)$$

se este for absolutamente convergente.

(Mais geralmente, podemos definir momentos centrados em A , $E[(X - A)^k]$; se $A = \mu$, temos os momentos centrados, se $A = 0$, os momentos; note-se que, uma vez que se pode exprimir $E[(X - \mu)^k]$ como função dos momentos μ'_n , com $n = 0, \dots, k$, o momento centrado de ordem k existe se, e só se, existir o correspondente momento de ordem k , pois como já anteriormente observámos, a existência deste implica a existência dos de ordem inferior).

Neste caso, o momento centrado de ordem 0 é 1, o de ordem 1 é 0 e o de ordem 2 corresponde ao valor da variância σ^2 .

Além destes momentos teóricos, é possível definir momentos amostrais ou momentos empíricos. Estes são naturalmente definidos a partir da função de distribuição empírica.

Definição 1.3 *Sejam $k \in \mathbb{N}_0$ e $X_{1:n}, \dots, X_{n:n}$ uma amostra aleatória de elementos distintos ordenados de forma crescente. Designa-se por **função de distribuição empírica** a função*

$$F_n(x) = \begin{cases} 0 & , x < X_{1:n} \\ \frac{k}{n} & , X_{k:n} \leq x < X_{k+1:n} \\ 1 & , x \geq X_{n:n} \end{cases}$$

No caso de existirem elementos iguais a definição anterior é facilmente generalizada.

Existem dois coeficientes importantes neste trabalho e de grande relevância ao nível da inferência estatística, nomeadamente, na escolha de um modelo que se adequa aos dados. Essas estatísticas são o coeficiente de assimetria e o coeficiente de achatamento ou curtose, que dão informações acerca da assimetria (e por consequência do peso relativo das caudas, isto é, de $F_X(-x) = P(X \leq -x)$ e de $1 - F_X(x) = P(X > x)$) e sobre a proximidade dos dados em relação ao valor médio, respectivamente.

Definição 1.4 *Dada uma variável aleatória X , designa-se por **coeficiente de assimetria (skewness) da variável aleatória X** , representando-se por $\gamma_1(X)$, o quociente*

$$\gamma_1(X) = \frac{\mu_3}{\sigma_X^3} = \frac{E[(X - \mu)^3]}{\sigma_X^3}.$$

Note-se que γ_1 pode assumir qualquer valor real. No caso de uma variável aleatória simétrica temos $\gamma_1 = 0$ pois, nessa situação, todos os momentos centrados de ordem ímpar são nulos. Um valor de γ_1 positivo implica caudas mais pesadas à esquerda acontecendo o inverso para um valor de γ_1 negativo.

Definição 1.5 *Dada uma variável aleatória X , designa-se por **achatamento ou curtose da variável aleatória X** , representando-se por $\gamma_2(X)$,*

$$\gamma_2(X) = \frac{\mu_4}{\sigma_X^4} - 3 = \frac{E[(X - \mu)^4]}{\sigma_X^4} - 3.$$

A subtracção de 3 na expressão da curtose torna o valor da expressão anterior nulo para uma variável aleatória Gaussiana padrão. No caso de $Z \sim \text{Gau}(0, 1)$ tem-se $E(Z^4) = 3$ o que implica $\gamma_2(Z) = 0$.

Em relação à curtose, assumindo sem perda de generalidade que $\mu = 0$,

$$\begin{aligned} \gamma_2(X) &= \frac{E(X^4)}{\sigma_X^4} - 3 \\ &= \frac{E(X^4)}{[E(X^2)]^2} - 3 \\ &= \frac{E(Y^2)}{[E(Y)]^2} - 3 \\ &\geq 1 - 3 = -2 \end{aligned}$$

pois $E(Y^2) \geq [E(Y)]^2 > 0$ caso existam os momentos em causa. Portanto, a curtose assume apenas valores no intervalo $[-2, +\infty[$ sendo nula no caso de distribuição parente Gaussiana padrão. No caso de variáveis aleatórias unimodais estandarizadas, se $\gamma_2 > 0$ a função tem um “pico” superior ao da gaussiana padrão (note-se que uma diminuição da variância traduz-se numa “constricção” da função densidade em torno da respectiva média). Caso contrário o “pico” da função densidade de probabilidade é inferior ao da gaussiana padrão.

A partir das definições anteriores é possível estabelecer algumas funções importantes.

Definição 1.6 *Dada uma variável aleatória contínua X com função de distribuição F_X designa-se por **função geradora de momentos da variável aleatória X** , representando-se por $M_X(t)$, a função*

$$M_X(t) = \int_{\mathbb{R}} \exp(tx) dF_X(x).$$

A designação de função geradora de momentos advém do facto de a imagem da derivada de ordem n dessa função no ponto 0 corresponder ao momento de ordem n de X uma vez que

$$\begin{aligned} \forall n \in \mathbb{N}_0, M_X^{(n)}(t) &= \frac{d}{dt^n} \int_{\mathbb{R}} e^{tx} dF_X(x) \\ &= \int_{\mathbb{R}} \frac{d}{dt^n} e^{tx} dF_X(x) \\ &= \int_{\mathbb{R}} x^n e^{tx} dF_X(x) \end{aligned}$$

admitindo que estamos em condições de permutar derivação e integração. Assim,

$$M_X^{(n)}(0) = \int_{\mathbb{R}} x^n dF_X(x) = E(X^n), \quad n \in \mathbb{N}_0.$$

Definição 1.7 Dada uma variável aleatória contínua X com função de distribuição F_X designa-se por **função geradora de cumulantes**, representando-se por $K_X(t)$, a função

$$K_X(t) = \ln M_X(t).$$

Definição 1.8 Considerando a expansão em série de MacLaurin da função geradora de cumulantes

$$K_X(t) = \sum_{i=0}^{+\infty} K_X^{(i)}(0) \frac{t^i}{i!} = \sum_{i=0}^{+\infty} k_i \frac{t^i}{i!},$$

k_i é designado por **cumulante de ordem i** .

O termo cumulante, também designado por semi-invariante, foi apresentado pela primeira vez por Thiele (1903) (cf. David e Edwards, 2001).

Podem estabelecer-se relações entre os momentos, momentos centrados e os cumulantes, usando a relação

$$K_X(t) = \ln M_X(t) = \ln \left(1 + \sum_{n=1}^{\infty} \frac{E(X^n) t^n}{n!} \right).$$

Uma vez que, para $x \in (-1, 1]$,

$$f(x) = \ln(1+x) = \sum_{n=1}^{\infty} (-1)^{n+1} (n-1)! \frac{x^n}{n!} = \sum_{n=1}^{\infty} (-1)^{n+1} \frac{x^n}{n}$$

tem-se que

$$K_X(t) = \sum_{j=1}^{\infty} \frac{(-1)^{j+1}}{j} \left(\sum_{n=1}^{\infty} \frac{E(X^n) t^n}{n!} \right)^j.$$

A expressão anterior permite relacionar momentos e cumulantes. As relações para os primeiros valores de k encontram-se abaixo sendo a continuação apenas um trabalho rotineiro de paciência que é dispensável prosseguir, podendo os interessados consultar Kendall e Stuart (1977, p. 73).

$$k_1 = E(X) = \mu$$

$$k_2 = E(X^2) - [E(X)]^2 = \sigma^2$$

$$k_3 = E(X^3) - 3\mu E(X^2) + 2[E(X)]^3 = E[(X - \mu)^3]$$

$$k_4 = E(X^4) - 3[E(X^2)]^2 - 4\mu E(X^3) + 12\mu^2 E(X^2) - 6\mu^4$$

$$= E[(X - \mu)^4] - 3E[(X - \mu)^2]^2 = E[(X - \mu)^4] - 3\sigma^4.$$

1.2 Algumas propriedades de matrizes

Antes de avançarmos para o estudo de uma alternativa aos momentos ordinários necessitamos de invocar algumas propriedades relativas a matrizes. As matrizes consideradas serão do tipo $A \in \mathbf{M}_p(\mathbb{R})$, isto é, matrizes quadradas de ordem p com valores reais. Inicialmente apresentaremos algumas matrizes especiais como é o caso das matrizes simétricas e ortogonais e algumas das suas propriedades. A parte final será dedicada à análise das matrizes na sua faceta de representantes de transformações lineares.

Começemos por analisar o subconjunto $\text{Sym}(k)$ das matrizes de ordem k , simétricas e invertíveis, com valores reais. Dada uma matriz $A \in \text{Sym}(k)$ podemos decompô-la num produto de matrizes utilizando os valores próprios $\lambda_1, \dots, \lambda_k$ (não necessariamente distintos) e os respectivos vectores próprios z_1, \dots, z_k :

$$A = \sum_{j \leq k} \lambda_j z_j z_j^T = Z^T D Z.$$

onde Z^T é uma matriz formada pelos vectores próprios ortonormados e D é uma matriz diagonal com os valores próprios de A .

Dois subconjuntos de $\text{Sym}(k)$ terão grande relevância. Trata-se do conjunto das matrizes definidas positivas $\text{PD}(k)$ e do conjunto das matrizes definidas não negativas $\text{NND}(k)$ que se podem definir da seguinte forma

$A \in \text{NND}(k)$ se, e só se, $A \in \text{Sym}(k)$ e $x^T A x \geq 0$ para qualquer $x \in \mathbb{R}^k$;

$A \in \text{PD}(k)$ se, e só se, $A \in \text{Sym}(k)$ e $x^T A x > 0$ para qualquer $x \in \mathbb{R}^k$.

Uma definição alternativa desses dois subconjuntos pode ser apresentada tendo em conta o menor valor próprio λ_{\min} da matriz A

$A \in \text{NND}(k)$ se, e só se, $A \in \text{Sym}(k)$ e $\lambda_{\min} \geq 0$;

$A \in \text{PD}(k)$ se, e só se, $A \in \text{Sym}(k)$ e $\lambda_{\min} > 0$.

No terceiro capítulo será importante o conceito de raiz quadrada.

Definição 1.9 Designa-se por **raiz quadrada** de uma matriz A , qualquer matriz R que satisfaça pelo menos uma das seguintes condições:

$$R R = R^2 = A \text{ ou } R^T R = A. \quad (1)$$

No caso de matrizes simétricas, definidas positivas e não diagonais cada uma das condições anteriores permite encontrar raízes quadradas distintas. Computacionalmente, diversos programas permitem encontrar raízes de uma matriz. Em

particular, no *software* MatLab os comandos “`sqrtm(A)`” e “`chol(A)`” permitem, respectivamente, obter soluções que satisfazem a primeira e segunda condição de (1). Em particular, o comando “`chol(A)`” permite obter a segunda decomposição em que R uma matriz triangular superior. Esta decomposição existe sempre para matrizes simétricas e definidas positivas e é conhecida ao nível da Análise Numérica por decomposição de Choleski. Mais detalhes sobre o *software* MatLab, extensivamente utilizado nesta dissertação, podem ser encontrados em Hanselman e Littlefield (2005).

Também no terceiro capítulo serão relevantes as matrizes ortogonais, isto é, matrizes invertíveis cuja inversa é igual à sua transposta. Os resultados que se seguem enunciam algumas das suas propriedades.

Lema 1.1 *Uma matriz ortogonal A é uma matriz cujas colunas (e linhas) constituem uma base ortonormada de \mathbb{R}^p .*

Demonstração Basta notar que se A é ortogonal então $AA^T = I$ e portanto os vectores linhas (e colunas) são normados e ortogonais. ■

O lema anterior sugere, segundo Beauregard e Fraleigh (1990), a designação de matriz ortonormada em substituição da habitual designação de matriz ortogonal.

Os próximos teoremas caracterizam uma matriz ortogonal A de ordem 2 tal que o módulo do determinante é unitário, isto é, $|\det(A)| = 1$.

Teorema 1.1 *Se as entradas de uma matriz ortogonal $A(t)$ de ordem 2 são funções contínuas de t e $\det[A(t)] = 1$ então A pode ser escrita como*

$$A(t) = \begin{pmatrix} \cos \phi(t) & \sin \phi(t) \\ -\sin \phi(t) & \cos \phi(t) \end{pmatrix}$$

onde ϕ é uma função contínua.

Demonstração Seja $A(t) = \begin{pmatrix} a(t) & b(t) \\ c(t) & d(t) \end{pmatrix}$ uma matriz ortogonal com determinante 1. Então $A^{-1}(t) = \begin{pmatrix} d(t) & -b(t) \\ -c(t) & a(t) \end{pmatrix}$ e portanto $a(t) = d(t)$ e $b(t) = -c(t)$ pois $A^{-1}(t) = [A(t)]^T$.

Pelo Lema 1.1, $[a(t)]^2 + [b(t)]^2 = 1$ e $[c(t)]^2 + [d(t)]^2 = 1$. Donde, o valor de cada entrada da matriz A varia entre -1 e 1 e pode ser representado por $\pm \cos \phi(t)$ ou $\pm \sin \phi(t)$ para $\phi(t)$ contínua (Courant, 1988). Além disso, em cada linha (ou coluna) existe apenas uma entrada que é do tipo $\pm \cos \phi(t)$ e outra do tipo $\pm \sin \phi(t)$. É indiferente escolher para a entrada (1,1) da matriz $\pm \cos \phi(t)$ ou $\pm \sin \phi(t)$ pois $\cos(t) = \sin\left(\frac{\pi}{2} - t\right)$ e $\cos(t) = \cos(-t)$. ■

Note-se que a matriz $A(t)$ constitui uma rotação de $\phi(t)$ radianos em torno da origem no sentido anti-horário (Oliveira, 1997).

Teorema 1.2 *Seja $A(t)$ uma matriz de ordem 2 cujas entradas são funções ϕ contínuas em t . Se $\det[A(t)] = -1$ então A pode ser escrita como*

$$A(t) = \begin{pmatrix} \cos \phi(t) & \sin \phi(t) \\ \sin \phi(t) & -\cos \phi(t) \end{pmatrix}.$$

Demonstração Idêntica à demonstração anterior. ■

Teorema 1.3 *Dadas duas matrizes invertíveis $A, B \in \mathbf{M}_p(\mathbb{R})$ e suas transpostas A^T e B^T tal que $A^T A = B^T B$ então a matriz $(A^T)^{-1} B^T$ é uma matriz ortogonal.*

Demonstração Para mostrar que $(A^T)^{-1} B^T$ é uma matriz ortogonal basta recorrer à definição de matriz ortogonal e verificar que

$$(A^T)^{-1} B^T \left[(A^T)^{-1} B^T \right]^T = I_p$$

onde I_p é a matriz identidade de ordem p . De facto,

$$\begin{aligned} & (A^T)^{-1} B^T \left[(A^T)^{-1} B^T \right]^T \\ &= (A^T)^{-1} B^T B \left[(A^T)^{-1} \right]^T \\ &= (A^T)^{-1} A^T A A^{-1} \\ &= I_p \end{aligned}$$

como queríamos demonstrar. ■

Para terminar apresenta-se um resultado que será utilizado no terceiro capítulo. A demonstração pode ser encontrada em Courant (1988).

Teorema 1.4 *Uma matriz $A \in \mathbf{M}_p(\mathbb{R})$ diferenciável é uma matriz jacobiana de uma transformação se, e só se, para cada linha i da matriz*

$$\frac{\partial a_{ij}}{\partial t_k} = \frac{\partial a_{ik}}{\partial t_j}$$

para quaisquer $k, j \in \{1, \dots, p\}$.

Os conceitos de matriz e transformação encontram-se associados pelo que determinados conceitos associados a transformações também se aplicam às matrizes. Assim, uma matriz de valores reais com ordem $n \times k$, $B \in \mathbb{R}^{n \times k}$, pode ser identificada com uma aplicação linear que a cada $x \in \mathbb{R}^k$ faz corresponder $Bx \in \mathbb{R}^n$. Nesse caso tem interesse definir os subespaços **Imagem de B**

$$\text{Im}(B) = \{Bx | x \in \mathbb{R}^k\}$$

e **Núcleo de B**

$$\text{N}(B) = \{x \in \mathbb{R}^k | Bx = 0\}.$$

Considerando o subconjunto $\text{NND}(k)$, sejam $A \in \text{NND}(k)$ e U uma raiz quadrada de A tal que

$$A = UU^T$$

então $\text{Im}(A) = \text{Im}(U)$.

Além disso, prova-se que

$$\text{Im}(X^T A X) = \text{Im}(X^T A)$$

pois se $\text{Im}(A) = \text{Im}(U)$ então

$$\text{Im}(X^T A) = \text{Im}(X^T U) = \text{Im}(X^T A X) \subseteq \text{Im}(X^T A)$$

para qualquer matriz $X \in M_{k \times n}(\mathbb{R})$.

A inversa de uma matriz nem sempre existe pelo que apresentamos a seguinte generalização.

Definição 1.10 *Dada uma matriz $A \in \mathbb{R}^{n \times k}$, qualquer matriz $G \in \mathbb{R}^{n \times k}$ que verifique $AGA = A$ é dita uma **matriz inversa generalizada** de A . O conjunto das matrizes inversas generalizadas de A é representado por A^- .*

Obviamente que se A for uma matriz invertível então A^- é um conjunto singular.

Lema 1.2 *Considerem-se as matrizes $A \in \mathbb{R}^{n \times k}$, $B \in \mathbb{R}^{n \times s}$ e uma matriz $G \in \mathbb{R}^{s \times n}$ inversa generalizada de B então*

$$\text{Im}(A) \subseteq \text{Im}(B) \text{ se, e só se, } BGA = A.$$

Se $\text{Im}(A) \subseteq \text{Im}(B)$ e $B \in \text{NND}(n)$ então o produto $A^T G A$ não depende da escolha de G , é definido não negativo e tem a mesma característica de A .

Demonstração Consultar Pukelsheim (1993, pp. 17-18). ■

1.3 Momentos Canônicos

As limitações de uma sucessão de momentos ordinários $\{\mu'_k\}_{k \in \mathbb{N}}$, caso existam momentos de todas as ordens, para caracterizar a medida associada θ estão amplamente discutidas na literatura. Uma simples transformação de localização e escala de uma medida θ gera uma nova sucessão de momentos completamente distinta da inicial. De igual forma, certas propriedades como a simetria não são imediatamente evidentes a partir dos termos da sucessão de momentos embora, nesse caso, tal seja superável com o recurso aos momentos centrados. Exemplificamos a seguir.

Exemplo 1.1 *Considerem-se duas variáveis aleatórias uniformes $X \sim U(-1, 1)$ e $Y \sim U(0, 1)$. A estas variáveis correspondem, respectivamente, as sucessões de momentos*

$$\mathbf{c} = 0, \frac{1}{3}, 0, \frac{1}{5} \dots \quad e \quad \mathbf{c}^* = \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \dots$$

Se é evidente que a variável aleatória X é simétrica a partir da respectiva sucessão de momentos \mathbf{c} o mesmo já não acontece com a variável Y . Por outro lado, se se considerar uma simples transformação da variável aleatória Y , $Z = Y + 1$, a sucessão de momentos correspondente é

$$\mathbf{c}^{**} = \frac{3}{2}, \frac{7}{3}, \frac{15}{4}, \frac{31}{5}, \dots$$

que é completamente distinta de \mathbf{c}^ .*

Considere-se uma medida de probabilidade θ com suporte em $[a, b]$ e k primeiros momentos c_1, \dots, c_k . Skibinsky (1967) analisou os possíveis valores para c_{k+1} . Verifica-se que c_{k+1} pode tomar qualquer valor de um intervalo $[c_{k+1}^-, c_{k+1}^+]$ onde os extremos do intervalo correspondem aos valores apresentados na Definição 4.11 da Secção A dos Anexos. De modo a localizar cada momento ordinário no intervalo

de valores admissíveis dados os momentos de ordem inferior definem-se os **momentos canônicos**

$$p_k = p_k(\mathbf{c}) = \frac{c_k - c_k^-}{c_k^+ - c_k^-}$$

com $c_k^+ > c_k^-$. Associados aos momentos canônicos estão as quantidades $q_k = 1 - p_k$.

O cálculo dos momentos canônicos pode ser feito directamente a partir dos momentos ordinários uma vez que

$$p_k = \frac{\underline{H}_k \overline{H}_{k-2}}{\underline{H}_{k-1} \overline{H}_{k-1}}$$

onde \underline{H}_i e \overline{H}_i são os determinantes de Hankel

$$\underline{H}_{2k} = \begin{vmatrix} c_0 & \cdots & c_k \\ \vdots & \ddots & \vdots \\ c_k & \cdots & c_{2k} \end{vmatrix} \quad \overline{H}_{2k} = \begin{vmatrix} c_1 - c_2 & \cdots & c_k - c_{k+1} \\ \vdots & \ddots & \vdots \\ c_k - c_{k+1} & \cdots & c_{2k-1} - c_{2k} \end{vmatrix}$$

$$\underline{H}_{2k+1} = \begin{vmatrix} c_1 & \cdots & c_{k+1} \\ \vdots & \ddots & \vdots \\ c_{k+1} & \cdots & c_{2k+1} \end{vmatrix} \quad \overline{H}_{2k+1} = \begin{vmatrix} c_0 - c_1 & \cdots & c_k - c_{k+1} \\ \vdots & \ddots & \vdots \\ c_k - c_{k+1} & \cdots & c_{2k} - c_{2k+1} \end{vmatrix}$$

Estes momentos têm a vantagem de serem invariantes relativamente a transformações lineares da medida de probabilidade associada. Assim, iremos assumir que o suporte das medidas de probabilidade em causa é $[-1, 1]$ caso nada se refira.

O cálculo dos momentos canônicos pode ser feito rapidamente utilizando o algoritmo Q-D. Mais detalhes podem ser encontrados na Secção A dos Anexos.

1.3.1 Variáveis aleatórias com suporte não limitado

Os momentos canônicos são apresentados por Dette e Studden (1997) para variáveis de suporte num subconjunto de um intervalo fechado. Os momentos

canônicos estão intimamente associados aos zeros de polinômios ortogonais mônicos que satisfazem a relação de recorrência:

$$\begin{aligned} P_0(x) &= 1; \\ P_1(x) &= x - \zeta_1; \\ P_{m+1}(x) &= (x - \zeta_{2m} - \zeta_{2m+1}) P_m(x) - \zeta_{2m-1} \zeta_{2m} P_{m-1}(x) \text{ para } m \geq 1. \end{aligned} \tag{2}$$

onde

$$\zeta_0 = 1 \text{ e } \zeta_m = p_m(1 - p_{m-1}) = p_m q_{m-1}, \quad n \geq 1. \tag{3}$$

Os polinômios ortogonais relativamente a uma medida μ satisfazem uma relação de recorrência do tipo

$$P_{n+1}(x) = (A_n x + B_n) P_n(x) - C_n P_{n-1}(x), \quad n = 0, 1, 2, \dots \tag{4}$$

com $P_{-1}(x) = 0$ e $A_{n-1} A_n C_n > 0$ independentemente do suporte da medida θ ser ou não limitado.

A questão que se coloca é saber se usando (2) e (4) podemos definir uma sucessão de momentos canônicos com a aplicação e interesse dos momentos já definidos para medidas com suporte num conjunto não limitado. Infelizmente tal não se irá verificar pelo que designaremos as constantes obtidas por **pseudomomentos canônicos**.

Comecemos pelo caso de uma variável aleatória gaussiana ou normal, padrão, $X \sim \text{Gau}(0, 1)$, com função densidade

$$\theta(x) = \frac{\exp(-x^2/2)}{\sqrt{2\pi}} I_{\mathbb{R}}(x).$$

Os polinômios de Hermite $H(x)$ são ortogonais relativamente à medida θ . Esses

polinômios na forma mónica definem-se recursivamente como:

$$\begin{aligned} H_0(x) &= 1; \\ H_1(x) &= x; \\ H_{n+1}(x) &= xH_n(x) - nH_{n-1}(x) \text{ para } n \geq 1. \end{aligned} \tag{5}$$

Os primeiros termos da sucessão de polinômios $(H_n)_{n \in \mathbb{N}_0}$ são

$$\begin{aligned} H_0(x) &= 1, \\ H_1(x) &= x, \\ H_2(x) &= x^2 - 1, \\ H_3(x) &= x^3 - 3x. \end{aligned}$$

Usando a última igualdade de (2) e (5) conclui-se que as quantidades ζ_m são dadas por:

$$\begin{aligned} \zeta_1 &= 1, \quad \zeta_2 = 1, \\ \zeta_3 &= -1, \quad \zeta_4 = -2, \\ \zeta_5 &= 2, \quad \zeta_6 = \frac{3}{2}, \\ \zeta_7 &= -\frac{3}{2} \text{ e } \zeta_8 = -\frac{4 \times 2}{3}. \end{aligned}$$

Mais geralmente, para $m \geq 2$,

$$\zeta_{2m} = -\zeta_{2m+1} = (-1)^{m+1} \frac{m \times (m-2) \times \dots \times m^*}{(m-1) \times (m-3) \times \dots \times (m-1)^*}$$

onde

$$m^* = \begin{cases} 2, & \text{se } m \text{ par} \\ 1, & \text{se } m \text{ ímpar} \end{cases}.$$

Portanto, para $m \geq 1$, $\zeta_m = p_m^* (1 - p_{m-1}^*)$, ou seja,

$$p_m^* = \frac{\zeta_m}{1 - p_{m-1}^*} \tag{6}$$

onde $p_0^* = 0$ e p_m^* é o pseudomomento canônico de ordem m .

Os primeiros n pseudomomentos canônicos podem ser determinados executando o *script*

```

zeta=[1 1 -1]
for j=2:n
  if mod(j,2)<0.5
    a=2:2:j
    A=prod(a)
    b=1:2:(j-1)
    B=prod(b)
    zeta(2*j)=A/B
  else
    a=1:2:j
    A=prod(a)
    b=2:2:(j-1)
    B=prod(b)
    zeta(2*j)=A/B
  end
  zeta(2*j+1)=-zeta(2*j)
end
p(1)=zeta(1)
for j=2:(2*n+1)
  p(j)=zeta(j)/(1-p(j-1))
end

```

Os primeiros pseudomomentos canônicos p_i^* encontram-se na tabela abaixo.

i	p_i^*
1	1
2	—
3	—
4	—
5	—

Tabela 1 – Pseudomomentos canônicos da variável aleatória gaussiana padrão

De facto, apenas existe o pseudomomento canónico de ordem 1 porque o denominador de (6) anula-se para $m = 2$ o que contrasta com a existência de momentos canónicos de qualquer ordem no caso de variáveis aleatórias com suporte limitado (ver Definição 4.12 da Secção A dos Anexos).

Vamos considerar outra função densidade de suporte não limitado mas para a qual seja possível definir os pseudomomentos canónicos de qualquer ordem. É o caso da densidade de uma variável aleatória $X \sim \text{Gama}(\alpha)$:

$$\theta(x) = \frac{x^{\alpha-1} \exp(-x)}{\Gamma(\alpha)} I_{(0,+\infty)}(x), \quad \alpha > -1.$$

Os polinómios ortogonais relativamente à medida θ são os polinómios de Laguerre generalizados $L^{(\alpha)}(x)$. Definem-se, na forma mónica, recursivamente como

$$\begin{aligned} L_0^{(\alpha)}(x) &= 1, \\ L_1^{(\alpha)}(x) &= x - \alpha - 1, \\ L_{m+1}^{(\alpha)}(x) &= (x - 2m - 1 - \alpha) L_m^{(\alpha)}(x) + (m + \alpha) L_{m-1}^{(\alpha)}(x), \quad \text{para } m \geq 1. \end{aligned} \tag{7}$$

Da última igualdade de (2) e de (7) obtém-se

$$\begin{cases} \zeta_{2m} = -\frac{m + \alpha}{\zeta_{2m-1}} \\ \zeta_{2m+1} = 2m + 1 + \alpha - \zeta_{2m} \end{cases}$$

Como $\zeta_0 = 1$, os valores de ζ_m ficam imediatamente estabelecidos. Recorrendo a (6) ficam definidos os pseudomomentos canónicos. Um *script* computacional que se pode utilizar para a sua determinação é:

```

zeta(1)=a
zeta(2)=(1+a)/a
zeta(3)=3+a-zeta(2)
for j=2:n
    zeta(2*j)=(j+a)/zeta(2*j-1)
    zeta(2*j+1)=2*j+1+a-zeta(2*j)
end
p(1)=zeta(1)
for j=2:(2*n+1)
    p(j)=zeta(j)/(1-p(j-1))
end
    
```

Por exemplo, para $\alpha = 3$ os primeiros pseudomomentos canónicos são

i	p_i^*	i	p_i^*	i	p_i^*	i	p_i^*
1	3	6	0.0867	11	15.4130	16	0.0323
2	0.6667	7	11.7057	12	0.0427	17	21.2785
3	22.000	8	0.0612	13	17.3558	18	0.0287
4	0.0325	9	13.4791	14	0.0368	19	23.2510
5	8.9732	10	0.0507	15	19.3125	20	0.0259

Tabela 2 – Pseudomomentos canónicos da variável aleatória Gama(3)

O facto de a distribuição parente ser simétrica não se reflecte nestas quantidades, tal como acontece com os momentos canónicos definidos anteriormente. Neste caso, todos os momentos de ordem ímpar pertencentes ao intervalo $(0, 1)$ são iguais a 0.5 (ver Teorema 4.12 da Secção A dos Anexos).

No Exemplo 4.5 da Secção A dos Anexos mostramos que

$$p_2 = \frac{c_2 - c_1^2}{c_1(1 - c_1)}.$$

Esta relação não se verifica para estes pseudomomentos pelo que não podem ser estabelecidas as mesmas relações entre momentos ordinários e momentos canônicos apresentadas para variáveis aleatórias com suporte limitado.

2 Planejamento de experiências

2.1 Planeamentos óptimos

Um modelo linear é um modelo da forma

$$\mathbf{Y} = X^T \boldsymbol{\theta} + \boldsymbol{\varepsilon} \quad (8)$$

onde $\boldsymbol{\theta} = (\theta_1, \dots, \theta_k)^T$ é um vector de parâmetros desconhecidos, $X = \mathbf{f}(x) = (f_1(x), \dots, f_k(x))$ é uma matriz $n \times k$ chamada **matriz do planeamento** que depende dos pontos de observação x_1, \dots, x_n e $\boldsymbol{\varepsilon} = (\varepsilon_1, \dots, \varepsilon_n)$ corresponde ao erro de valor médio nulo e variância constante σ^2 (homocedasticidade). Geralmente também se assume que $\mathbf{Y} \sim N(X^T \boldsymbol{\theta}, \sigma^2 I_n)$.

O **espaço do planeamento**, o conjunto de todos os pontos possíveis onde podem ser feitas as observações, será representado por \mathcal{X} . As colunas da matriz X são do tipo com f_1, \dots, f_k contínuas e linearmente independentes.

O problema que geralmente se coloca é estimar o vector de parâmetros $\boldsymbol{\theta}$ e a variância σ^2 a partir dos valores observados $\mathbf{Y} = (Y_1, \dots, Y_n)^T$. Aqui iremos simplesmente considerar os estimadores lineares centrados de $\boldsymbol{\theta}$ que denotaremos por

$$\widehat{\boldsymbol{\theta}}_L = L \mathbf{Y} \quad (9)$$

onde L é uma matriz de ordem $k \times n$.

A fim de comparar diferentes estimadores lineares centrados de $\boldsymbol{\theta}$ iremos introduzir a noção de ordem de Loewner.

Definição 2.1 *Dadas duas matrizes simétricas de ordem k , $A, B \in \text{Sym}(k)$, diz-se que:*

1. $A \geq B$ com respeito à ordem de Loewner se, e só se, $A - B \in \text{NND}(k)$;
2. $A > B$ com respeito à ordem de Loewner se, e só se, $A - B \in \text{PD}(k)$.

A definição anterior permite considerar o estimador linear centrado, conhecido como estimador de Gauss-Markov,

$$\widehat{\boldsymbol{\theta}}^{GM} = (X^T X)^{-1} X^T Y$$

que é o **melhor estimador linear centrado** (BLUE) com respeito à ordenação de Loewner, isto é,

$$\sigma^2 (X^T X)^{-1} = D(\widehat{\boldsymbol{\theta}}^{GM}) \leq D(\widehat{\boldsymbol{\theta}}_L)$$

para todos os estimadores lineares centrados $\widehat{\boldsymbol{\theta}}_L$ de $\boldsymbol{\theta}$ (Dette e Studden, 1997, p. 131) e onde $D(A)$ é a matriz de dispersão de A .

Uma extensão do resultado anterior pode ser feita para a situação de conhecer não directamente o vector de parâmetros $\boldsymbol{\theta}$, mas antes uma combinação linear desses parâmetros digamos $z_j \boldsymbol{\theta}$ onde $z_j \in \mathbb{R}^k$ ($j = 1, \dots, s$). Designaremos a matriz resultante $\mathbf{z}\boldsymbol{\theta}$ (de ordem $k \times s$) por K . O BLUE para $K^T \boldsymbol{\theta}$ existe se $\text{Im}(K) \subseteq \text{Im}(X^T)$. Nesse caso, o BLUE é dado por,

$$\widehat{\boldsymbol{\theta}}^K = K^T (X^T X)^- X^T \mathbf{Y} \quad (10)$$

onde $(X^T X)^-$ é a matriz inversa generalizada de $X^T X$. A respectiva matriz de dispersão é

$$D(\widehat{\boldsymbol{\theta}}^K) = \sigma^2 K^T (X^T X)^- K.$$

Note-se que pelo Lema 1.2 a matriz de dispersão $D(\widehat{\boldsymbol{\theta}}^K)$ não depende da escolha da matriz inversa generalizada de $X^T X$.

Exemplo 2.1 Se $\mathbf{f}(x) = (1, x, \dots, x^m)^T$ é um vector de monómios então o modelo linear em (8) reduz-se à regressão polinomial simples de grau m , isto é,

$$Y = \sum_{j=0}^m \theta_{j+1} x^j + \varepsilon. \quad (11)$$

A matriz X terá característica n se os pontos de observação x_1, \dots, x_n forem todos distintos. O estimador BLUE de $K^T \boldsymbol{\theta}$ ($K \in \mathbb{R}^{(m+1) \times s}$) é dado por (10) onde

$$\frac{1}{n} X^T X = \begin{bmatrix} 1 & c_1 & c_2 & \cdots & c_m \\ c_1 & c_2 & c_2 & \cdots & c_{m+1} \\ \vdots & \vdots & \vdots & & \vdots \\ c_m & c_{m+1} & c_{m+2} & \cdots & c_{2m} \end{bmatrix}$$

e $c_j = \frac{1}{n} \sum_{i=1}^n x_i^j, j = 0, \dots, 2m.$

Considerando o modelo linear (8) com as hipóteses usuais vamos considerar o teste

$$H_0 : K^T \boldsymbol{\theta} = 0$$

onde $K \in \mathbb{R}^{k \times s}$ é uma matriz com característica $s \leq k$. Se $n > \text{car}(X)$, onde $\text{car}(X)$ é a característica da matriz X , o teste F para testar H_0 rejeita essa hipótese para valores grandes da estatística

$$F = \frac{n - \text{car}(X)}{\text{car}(K)} \cdot \frac{\left(\widehat{\boldsymbol{\theta}}^K\right)^T \left(K^T (X^T X)^{-} K\right)^{-} \widehat{\boldsymbol{\theta}}^K}{Y^T (I_n - X (X^T X)^{-} X^T) Y}.$$

F tem distribuição não central F de Fisher-Snedecor com graus de liberdade iguais a $\text{car}(K)$ e $n - \text{car}(X)$. O parâmetro de não centralidade é

$$\frac{1}{\sigma^2} (K^T \boldsymbol{\theta})^T \left(K^T (X^T X)^{-} K\right)^{-} (K^T \boldsymbol{\theta}).$$

Se o parâmetro de não centralidade aumentar então a potência do teste também aumenta.

Exemplo 2.2 Considerando o modelo (11) poderemos estar interessados em determinar se a regressão polinomial mais apropriada é de grau m ou $m - 1$, isto é, testar

$$H_0 : K^T \boldsymbol{\theta} = \theta_m = 0.$$

onde $K = e_m = (0, \dots, 0, 1)^T$. Assumindo que existem pelo menos $m+1$ observações feitas em pontos distintos, por (10), $\widehat{\theta}_m^{e_m} = e_m^T \widehat{\theta}^{GM}$ portanto a estatística de teste é ($n > m + 1$),

$$F = \frac{\left(\widehat{\theta}_m^{e_m}\right)^2 \left(e_m^T (X^T X)^{-1} e_m\right)^{-1}}{\frac{1}{n-m-1} \mathbf{Y}^T (I_n - X (X^T X)^{-1} X^T) \mathbf{Y}}$$

que tem distribuição $F_{(1, n-m-1)}$ com parâmetro de não centralidade

$$\frac{1}{\sigma^2} \theta_m^2 \left(e_m^T (X^T X)^{-1} e_m\right)^{-1}.$$

Como verificado anteriormente, F aumenta com o aumento do parâmetro de não centralidade pelo que o teste pode ser melhorado diminuindo a quantidade

$$e_m^T (X^T X)^{-1} e_m$$

que depende da escolha de x_1, \dots, x_n .

O exemplo anterior mostra como é importante ponderar em que pontos deve ser recolhida a amostra. Considere-se uma amostra x_1, \dots, x_n em que existem apenas x_1, \dots, x_l pontos distintos. Iremos representar o número de vezes que cada ponto aparece na amostra através da **matriz do planeamento exacto**:

$$\xi_{(n)} = \begin{pmatrix} x_1 & \dots & x_l \\ \frac{n_1}{n} & \dots & \frac{n_l}{n} \end{pmatrix}.$$

Assim,

$$\begin{aligned} X^T X &= \sum_{j=1}^n f(x_j) f^T(x_j) = n \sum_{j=1}^l \frac{n_j}{n} f(x_j) f^T(x_j) \\ &= n \int_{\mathcal{X}} f(x) f^T(x) d\xi_{(n)}(x) = nM(\xi_{(n)}), \end{aligned}$$

em que $M(\xi_{(n)})$ é chamada **matriz de informação do planeamento**. A matriz de informação $X^T X$ está directamente relacionada com a matriz de covariâncias

$\sigma^2 M^{-}(\xi_{(n)})/n$ das estimativas dos coeficientes do modelo linear (8) obtidas pelo método dos mínimos quadrados. $M^{-}(\xi_{(n)})$ representa a inversa generalizada da matriz de informação.

Seja $K \in \mathbb{R}^{k \times s}$ com $s \leq k$ e considere-se o problema de estimar $K^T \boldsymbol{\theta}$. O BLUE, nesse caso, é

$$\widehat{\boldsymbol{\theta}}^K = K^T (X^T X)^{-} X^T \mathbf{Y} = \frac{1}{n} K^T M^{-}(\xi_{(n)}) X^T \mathbf{Y}.$$

A matriz de dispersão é dada por

$$D(\widehat{\boldsymbol{\theta}}_K) = \frac{\sigma^2}{n} K^T M^{-}(\xi_{(n)}) K \quad (12)$$

que depende do planeamento $\xi_{(n)}$. Portanto, na prática, o problema consiste em encontrar um planeamento que minimize um funcional dessa matriz uma vez que a minimização com respeito à ordenação de Loewner é, em geral, muito complicada. A dificuldade em encontrar uma solução óptima com pesos múltiplos de $1/n$ leva à consideração de **planeamentos (aproximados)** definidos como sendo qualquer medida de probabilidade com suporte nos **pontos** x_1, \dots, x_s do espaço de planeamento \mathcal{X} e **pesos** w_1, \dots, w_s respectivamente. A utilização deste conceito mais geral de planeamento tem a vantagem facilitar uma maximização óptima de (12), (Fedorov, 1972). Na prática, utiliza-se depois um planeamento exacto que seja próximo do planeamento aproximado óptimo.

Existem diversos critérios para determinar planeamentos óptimos, isto é, planeamentos que maximizem algum funcional da matriz de informação do planeamento $\xi_{(n)}$ para estimar $K^T \boldsymbol{\theta}$,

$$C_K(M(\xi_{(n)})) = (K^T M^{-}(\xi_{(n)}) K)^{-1}, \quad (13)$$

uma vez que é difícil maximizar (13) directamente. O procedimento habitual é considerar as funções ϕ_p de Kiefer (Kiefer e Studden, 1976).

Definição 2.2 Seja $p \in [-\infty, 1]$ e considere-se

$\phi_p : \{C : C \in \text{PD}(s)\} \longrightarrow \mathbb{R}$ definida por

$$\phi_p(C) = \begin{cases} \left(\frac{1}{s} \text{traço}(C^p)\right)^{1/p} & \text{se } p \neq -\infty, 0 \\ (\det C)^{1/s} & \text{se } p = 0 \\ \lambda_{\min}(C) & \text{se } p = -\infty \end{cases}$$

e $\phi_p : \{C : C \in \text{NND}(s)\} \longrightarrow \mathbb{R}$ definida por

$$\phi_p(C) = \begin{cases} \left(\frac{1}{s} \text{traço}(C^p)\right)^{1/p} & \text{se } p \in (0, 1] \\ 0 & \text{se } p = [-\infty, 0] \end{cases}.$$

Um **planeamento** ϕ_p -**óptimo** é um planeamento que maximiza $\phi_p(C_K M(\xi_{(n)}))$.

Em particular, quando $p = 0$, um planeamento $\xi_{(n)}$ é chamado **D-óptimo** quando maximiza

$$\phi_0(C_K(M(\xi_{(n)}))) = (\det(K^T M^-(\xi_{(n)}) K))^{-1/s}.$$

Se $K = I_k$ então o planeamento D-óptimo maximiza $\det(M(\xi_{(n)}))^{-1/s}$. Se $K = (0, \dots, 0, 1)^T$ então o planeamento óptimo é designado por **D₁-óptimo**.

Para avaliar a qualidade de um planeamento na estimação de $K^T \theta$ utiliza-se o conceito de **eficiência** que a seguir se define. Dois planeamentos também podem ser comparados quanto à sua eficiência.

Definição 2.3 A **eficiência** ϕ_p de um planeamento ξ para estimar $K^T \theta$ é definida pelo quociente

$$\text{ef}_p(\xi) = \frac{\phi_p(C_K(M(\xi)))}{\sup_{\eta} (\phi_p(C_K(M(\eta))))}.$$

A **eficiência** ϕ_p de um planeamento ξ_1 relativamente a um planeamento ξ_2 é dado por

$$r_p(\xi_1, \xi_2) = \frac{\phi_p(C_K(M(\xi_1)))}{\phi_p(C_K(M(\xi_2)))}.$$

2.2 Planeamentos D_1 -óptimo e D -óptimo

Até ao final deste capítulo vamos considerar que existem condições para assumir que $\mathbf{f}(x) = (1, x, \dots, x^m)^T$. Assim, o nosso modelo linear será a regressão polinomial simples de grau m , isto é,

$$Y = \sum_{j=0}^m \theta_j x^j + \varepsilon. \quad (14)$$

com $x \in [-1, 1]$.

Uma das nossas tarefas será encontrar o planeamento aproximadamente óptimo para estimar o parâmetro $\boldsymbol{\theta}$. A outra corresponde a lidar com o problema de estimar o coeficiente do termo de maior grau θ_m (planeamento D_1 -óptimo). Iremos começar pelo segundo caso.

Considere-se $K = e_m = (0, \dots, 0, 1)^T \in \mathbb{R}^{m+1}$. Como vimos anteriormente a matriz de informação de $K^T \boldsymbol{\theta} = \theta_m$ é dada por

$$C_{e_m}(M(\xi_{(n)})) = (e_m^T M^{-1}(\xi_{(n)}) e_m)^{-1} = \frac{|M_m(\xi)|}{|M_{m-1}(\xi)|} \quad (15)$$

(Dette e Studden, 1997, p. 150) onde

$$M_m(\xi) = \int_{-1}^1 \mathbf{f}_m(x) \mathbf{f}_m^T(x) d\xi_{(n)}(x) = (c_{i+j})_{i,j=0}^m$$

Dette e Studden (1997) mostram que o determinante da matriz de informação está relacionado com um dos determinantes de Hankel apresentados anteriormente

$$M_m(\xi_{(n)}) = 2^{m(m+1)} \underline{H}_{2m}. \quad (16)$$

Por sua vez o quociente de determinantes de Hankel (15) pode ser escrito como função dos momentos canónicos p_1, \dots, p_{2m} . Maximizando essa expressão em termos dos momentos canónicos obtém-se o planeamento D_1 -óptimo apresentado no seguinte teorema.

Teorema 2.1 *O planeamento D_1 -óptimo ξ^{D_1} relativamente à regressão polinomial de grau m tem peso igual a $1/m$ nos zeros do polinómio de Chebyshev de segunda espécie $U_{m-1}(x)$ e peso $1/2m$ nos pontos -1 e 1 . Se, em vez de considerarmos que $x \in [-1, 1]$, admitirmos que $x \in [a, b]$ o planeamento D_1 -óptimo é dado por*

$$\xi_{[a,b]}(\{x\}) = \xi \left(\left\{ \frac{2x - b - a}{b - a} \right\} \right).$$

Para os graus um a quatro os pontos de observação devem ser os seguintes

m	Pontos
2	$-1, 0, 1$
3	$-1, -0.5, 0.5, 1$
4	$-1, -0.70711, 0, 0.70711, 1$
5	$-1, -0.80902, -0.30901, 0.30901, 0.80902, 1$

Tabela 3 – Pontos de observação de um planeamento D_1 -óptimo

A matriz de informação de um planeamento D_1 -óptimo ξ^{D_1} é dado por (Dette e Studden, 1997, p. 151)

$$C_{e_m}(M(\xi^D)) = \left(\frac{1}{2}\right)^{2m-2}.$$

Assim, tendo em conta (15), a **D_1 -eficiência de um planeamento ξ na estimação de θ_m** é dada por

$$\text{eff}_m^{D_1}(\xi) = \frac{C_{e_m}(M(\xi))}{C_{e_m}(M(\xi^{D_1}))} = \frac{2^{2m-2} |M_m(\xi)|}{|M_{m-1}(\xi)|}. \quad (17)$$

O planeamento D -óptimo é obtido à custa da maximização de uma função ϕ_0 de Kiefer da matriz de informação do planeamento relativo à regressão polinomial de grau m . Por (16) o determinante da matriz de informação pode ser escrito à custa de determinantes de Hankel.

A partir dessa expressão, Dette e Studden (1997, pp. 149 e 150) escrevem a matriz de informação anterior em termos dos momentos canónicos e maximizam a expressão resultante obtendo o planeamento D -óptimo descrito no próximo teorema.

Teorema 2.2 *O planeamento D -óptimo ξ^D para o parâmetro θ relativamente à regressão polinomial de grau m tem pesos iguais nos $m + 1$ zeros do polinómio $(x^2 - 1) L'_m(x)$ onde L'_m é a derivada do r -ésimo polinómio de Legendre.*

Para os graus um a quatro os pontos de observação devem ser os seguintes

m	Pontos
1	-1, 1
2	-1, 0, 1
3	-1, -0.447214, 0.447214, 1
4	-1, -0.654654, 0.654654, 1

Tabela 4 – Pontos de observação de um planeamento D -óptimo

A **D -eficiência de um planeamento ξ na estimação de $\theta = (\theta_1, \dots, \theta_m)$** é dada por

$$\text{eff}_m^D(\xi) = \left(\frac{|M_m(\xi)|}{|M_m(\xi^D)|} \right)^{1/(m+1)}.$$

onde o determinante da matriz de informação de um planeamento D -óptimo ξ^D é igual a (Dette e Studden, 1997, p. 150)

$$\det [M_m(\xi^D)]^{1/(m+1)} = \left(\frac{m}{2m-1} \right)^m \prod_{j=2}^m \left(\frac{(m-j+1)^2}{(2(m-j)+1)(2(m-j)+3)} \right)^{m+1-j}.$$

Exemplo 2.3 *Vamos considerar a regressão polinomial de grau dois definida em $\mathcal{X} = [-1, 1]$ e investigar se o termo de segundo grau é relevante no modelo*

($K = e_2 = (0, 0, 1)^T$). A matriz (invertível) de um planeamento exacto $\xi_{(n)}$ de uma amostra de dimensão n é da forma

$$M(\xi_{(n)}) = \begin{bmatrix} 1 & c_1 & c_2 \\ c_1 & c_2 & c_3 \\ c_2 & c_3 & c_4 \end{bmatrix}.$$

Para estimar $\theta_2 = e_2\theta$ é necessário que o planeamento tenha pelo menos 3 pontos distintos. Por (10), o estimador BLUE para $\theta_2 = e_2^T\theta$ é proporcional a

$$\begin{aligned} [C_K(M(\xi_{(n)}))]^{-1} &= (K^T M^{-1}(\xi_{(n)}) K) \\ &= e_2^T M^{-1}(\xi_{(n)}) e_2 \\ &= \frac{c_2 - c_1^2}{\det(M(\xi_{(n)}))} \in \mathbb{R}^+. \end{aligned} \quad (18)$$

Krafft e Schaefer (1995) conseguiram minimizar a quantidade $[C_K(M(\xi_{(n)}))]^{-1}$, e consequentemente minimizar a variância do estimador BLUE para θ_2 , no conjunto de todos os planeamentos com matrizes de informação invertíveis obtendo

$$\xi_{(n)}^* = \begin{cases} \begin{pmatrix} -1 & 0 & 1 \\ 0.25 & 0.5 & 0.25 \end{pmatrix}, & n = 4p \\ \begin{pmatrix} -1 & 0 & 1 \\ \frac{p}{4p+1} & \frac{2p+1}{4p+1} & \frac{p}{4p+1} \end{pmatrix}, & n = 4p+1 \\ \begin{pmatrix} -1 & \pm x_0(n) & 1 \\ \frac{p}{4p+2} & \frac{2p+1}{4p+2} & \frac{p+1}{4p+2} \end{pmatrix}, & n = 4p+2 \\ \begin{pmatrix} -1 & 0 & 1 \\ \frac{p+1}{4p+3} & \frac{2p+1}{4p+3} & \frac{p+1}{4p+3} \end{pmatrix}, & n = 4p+3 \end{cases}$$

onde, no caso $n = 4p + 2$, temos 2 soluções consoante se escolha $x_0(n)$ ou $-x_0(n)$.

$x_0(n)$ é a raiz real do polinómio

$$n^2x^3 - 3nx^2 + (n^2 - 2)x - n.$$

A obtenção de planeamentos exactos como no exemplo anterior é muito complicada. Usando a teoria dos momentos canónicos podemos obter planeamentos quase óptimos com maior facilidade. Assim, o Teorema 2.1 garante que o planeamento aproximado óptimo coincide com o planeamento exacto óptimo para $n = 4p$. Se n não for múltiplo de quatro deve-se utilizar o planeamento exacto “mais próximo”

$$\tilde{\xi}_{(n)}^* = \begin{pmatrix} -1 & 0 & 1 \\ \frac{n_0}{n} & 1 - \frac{2n_0}{n} & \frac{n_0}{n} \end{pmatrix}$$

onde n_0 é o inteiro mais próximo de $n/4$. Assim, o planeamento obtido através do Teorema 2.1 coincide com o planeamento óptimo do exemplo anterior para $n \neq 4p + 2$. Nesse caso, podemos calcular o rácio entre os valores de $C_{\theta_2}(M(\xi))$ do planeamento aproximado e do óptimo a fim de avaliar as discrepâncias. Os resultados obtidos sumarizam-se na seguinte tabela.

p	1	2	3	4	5
n	6	10	14	18	22
$x_0(p)$	0.0707	0.0408	0.0289	0.0224	0.0183
eficiência relativa	0.9327	0.9759	0.9877	0.9925	0.9950

Tabela 5 – Eficiência relativa entre o planeamento óptimo e quase óptimo

Note-se que para $p \geq 4$ a eficiência relativa já supera 99%.

2.3 Planeamentos robustos de modelos e discriminantes

Nas secções anteriores assumimos que o modelo linear (8) é conhecido e tratámos do caso particular da regressão polinomial de grau m , nomeadamente no que concerne à estimação do coeficiente do termo de maior grau e à estimação do vector de parâmetros θ . Iremos continuar a lidar com a regressão polinomial mas assumindo apenas que o grau é, no máximo, m .

Nesse contexto colocam-se dois problemas ao investigador:

- determinar o grau da regressão polinomial a utilizar;
- estimar todos os coeficientes relevantes da regressão polinomial.

Este problema surge com grande relevância uma vez que o planeamento D_1 -óptimo para a regressão polinomial de grau um utiliza dois pontos o que é obviamente extremamente ineficiente para estimar os parâmetros de uma regressão polinomial de grau dois. Mesmo que se utilize o planeamento D_1 -óptimo de um grau mais elevado verifica-se que este pode ser pouco eficiente se o grau mais adequado for inferior. A tabela seguinte mostra a eficiência do planeamento óptimo de grau quatro, $\xi_4^{D_1}$, para a regressão polinomial de grau $m = 1, 2, 3, 4$.

m	$\text{eff}_m(\xi_4^{D_1})$
1	1
2	0.5
3	0.5
4	0.5

Tabela 6 – Eficiência do planeamento D_1 -óptimo $\xi_4^{D_1}$

Anderson (1962) sugere o seguinte critério de decisão. Para determinar qual dos modelos da família

$$h_l(x) = \sum_{i=0}^l \theta_{li} x^i = \boldsymbol{\theta}_l^T \mathbf{f}_l(x), \quad l = 1, \dots, m$$

é mais apropriado considere-se o conjunto

$$\mathcal{F}_m = \{h_l(x) \mid l = 1, \dots, m\}$$

de todos os modelos de regressão polinomial até grau m . Para uma sequência de níveis de significância $\alpha_1, \dots, \alpha_m$ não decrescentes teste-se a hipótese

$$H_0 : \theta_{jj} = 0 \tag{19}$$

a um nível de significância α_j e escolha-se o maior inteiro para o qual se rejeite a hipótese nula. Este método apresenta diversas propriedades óptimas. Para Anderson (1962) a probabilidade de escolher um grau mais elevado que o desejado não depende da matriz de planeamento $X_l = (x_i^j)_{i=1, \dots, n}^{j=0, \dots, l}$ desde que esta seja invertível. Este método conduz à introdução do conceito de planeamentos discriminantes e planeamentos robustos que se segue.

Vimos anteriormente que para testar a hipótese (19) poderíamos utilizar o teste F . Referimos na altura que a distribuição da estatística de teste depende do parâmetro de não centralidade

$$\delta_l^2(\xi) = \frac{\theta_{ll}^2}{\sigma^2} \left(e_l^T (X_l^T X_l)^{-1} e_l \right)^{-1}.$$

Mais uma vez o que está em causa é maximizar as quantidades anteriores para $l = 1, \dots, m$. Assim, as quantidades

$$\text{eff}_l^{D_1}(\xi) = \frac{\delta_l^2(\xi)}{\sup_{\eta} \delta_l^2(\eta)} = 2^{2l-2} \frac{|M_l(\xi)|}{|M_{l-1}(\xi)|}$$

devem ser o maior possíveis. Porém, não é possível uma maximização simultânea de todas elas. A alternativa é maximizar uma média ponderada dessas quantidades em que as ponderações $\boldsymbol{\pi} = (\pi_1, \dots, \pi_m)$ ($\pi_i > 0$ e $\sum_{i=1}^m \pi_i = 1$) correspondem à confiança do investigador no facto do polinómio a utilizar mais adequado ser de grau $1, \dots, m$ respectivamente.

Definição 2.4 Um planeamento $\xi_{0,\boldsymbol{\pi}}$ é designado por **planeamento discriminante $\Psi_0^\boldsymbol{\pi}$ -óptimo para a classe \mathcal{F}_m** com respeito ao vector $\boldsymbol{\pi}$ se, e só se, maximiza a média geométrica ponderada

$$\Psi_0^\boldsymbol{\pi}(\xi) = \prod_{l=1}^m (\text{eff}_l^{D_1}(\xi))^{\pi_l} = \prod_{l=1}^m 2^{2l-2} \frac{|M_l(\xi)|}{|M_{l-1}(\xi)|}.$$

O caso $\boldsymbol{\pi} = (0, \dots, 0, 1)$ corresponde ao planeamento D_1 -óptimo.

O próximo teorema permite identificar em termos dos momentos canónicos quais os planeamentos discriminantes $\Psi_0^\boldsymbol{\pi}$ -óptimos.

Teorema 2.3 O planeamento discriminante $\Psi_0^\boldsymbol{\pi}$ -óptimo para a classe \mathcal{F}_m é univocamente determinado pelos momentos canónicos

$$\begin{aligned} p_{2i} &= \frac{\sigma_i}{\sigma_i + \sigma_{i+1}}, \quad i = 1, \dots, m-1 \\ p_{2i-1} &= 0.5, \quad i = 1, \dots, m \\ p_{2m} &= 1 \end{aligned}$$

onde $\sigma_i = \sum_{l=i}^m \beta_l$, $i = 1, \dots, m$.

Demonstração Consultar Lau e Studden (1985). ■

Para identificar a medida de probabilidade associada aos momentos canónicos até $m = 4$ referidos no teorema anterior iremos apresentar como teorema o Corolário 4.2.2 do livro de Dette e Studden (1997).

Teorema 2.4 *A medida de probabilidade no intervalo $[-1, 1]$ correspondente à sequência de momentos canónicos:*

- $\left(\frac{1}{2}, p_2, \frac{1}{2}, 1\right)$ tem pesos $p_2/2, q_2, p_2/2$ nos pontos $-1, 0, 1$, respectivamente;
- $\left(\frac{1}{2}, p_2, \frac{1}{2}, p_4, \frac{1}{2}, 1\right)$ tem pesos $a, \frac{1}{2}-a, \frac{1}{2}-a, a$ nos pontos $-1, -\sqrt{p_2q_4}, \sqrt{p_2q_4}, 1$, respectivamente, onde $a = \frac{p_2p_4}{2(1-p_2q_4)}$;
- $\left(\frac{1}{2}, p_2, \frac{1}{2}, p_4, \frac{1}{2}, p_6, \frac{1}{2}, 1\right)$ tem pesos $a, b, 1-2a-2b, b, a$ nos pontos $-1, -\sqrt{p_2q_4+p_4q_6}, 0, \sqrt{p_2q_4+p_4q_6}, 1$, respectivamente, onde $a = \frac{p_2p_4p_6}{2(q_2q_4+p_4p_6)}$ e $b = \frac{p_2q_2q_4}{2((p_2q_4+p_4p_6)+(q_2q_4+p_4p_6))}$.

Exemplo 2.4 *Os planeamentos discriminantes Ψ_0^π -óptimos para as classes $\mathcal{F}_2, \mathcal{F}_3$ e \mathcal{F}_4 , admitindo que as ponderações do vector π são iguais, podem ser obtidos a partir dos Teoremas 2.3 e 2.4 e resumidos na seguinte tabela (utilizando a notação de ambos os teoremas). Os pesos nos pontos de suporte do planeamento são iguais a $1/m$.*

m	π	Pontos	$\text{eff}_1^{D_1}/\text{eff}_2^{D_1}/\text{eff}_3^{D_1}/\text{eff}_4^{D_1}(\xi)$
2	$\left(\frac{1}{2}, \frac{1}{2}\right)$	$-1, 0, 1$	0.817/1/ - /-
3	$\left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right)$	$-1, -0.4472, 0, 4472, 1$	0.600/0.640/0.853/-
4	$\left(\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}\right)$	$-1, -0.6547, 0, 0.6547, 1$	0.571/0.588/0.627/0.836

Tabela 7 – Planeamentos discriminantes Ψ_0^π -óptimos

É curioso observar que os planeamentos discriminantes Ψ_0^π -óptimos para a classe \mathcal{F}_m e os planeamentos D-óptimos para uma regressão polinomial de grau m coincidem para um vector π “uniforme”.

Comparando as eficiências destes planeamentos discriminantes Ψ_0^π -óptimos com a eficiência do planeamento D_1 -óptimo para a regressão polinomial de grau quatro apresentada anteriormente verifica-se que a perda de eficiência no caso de grau quatro é compensada com um aumento de eficiência no caso do modelo de regressão ter um grau inferior.

De forma semelhante à utilizada para determinar os planeamentos discriminantes Ψ_0^π -óptimos podemos encontrar planeamentos que maximizem a média geométrica ponderada das D-eficiências até um determinado grau m . São os designados planeamentos robustos do modelo que são razoavelmente eficientes para um grupo de possíveis graus da regressão polinomial subjacente.

Definição 2.5 *Dado um vector $\boldsymbol{\pi} = (\pi_1, \dots, \pi_m)$ em que $\sum_{l=1}^m \pi_l = 1$ e $\pi_i > 0$, um planeamento $\xi_{0,\boldsymbol{\pi}}$ é designado por **planeamento robusto do modelo Ξ_0^π -óptimo para a classe \mathcal{F}_m** com respeito ao vector $\boldsymbol{\pi}$ se, e só se, maximiza a média geométrica ponderada*

$$\Xi_0^\pi(\xi) = \prod_{l=1}^m (\text{eff}_l^D(\xi))^{\pi_l} = \prod_{l=1}^m \left(\frac{|M_l(\xi)|}{|M_l(\xi^D)|} \right)^{\pi_l/(l+1)}.$$

Um teorema apresentado por Dette e Studden (1995) permite encontrar os momentos canónicos que caracterizam univocamente um planeamento robusto do modelo Ξ_0^π -óptimo.

Teorema 2.5 *O planeamento robusto do modelo Ξ_0^π -óptimo para a classe \mathcal{F}_m com respeito ao vector $\boldsymbol{\pi}$ possui momentos canónicos*

$$\begin{aligned} p_{2i-1} &= \frac{1}{2}, \quad i = 1, \dots, m \\ p_{2i} &= \frac{\sigma_i}{\sigma_i + \sigma_{i+1}}, \quad i = 1, \dots, m-1 \\ p_{2m} &= 1 \end{aligned}$$

onde $\sigma_i = \sum_{l=i}^m (l+1-i) \pi_l / (l+1)$.

Tal como anteriormente, para caracterizar as medidas de probabilidade associadas aos momentos canónicos definidos no teorema anterior temos de aplicar o Teorema 2.4.

Exemplo 2.5 *Os planeamentos robustos do modelo Ξ_0^π -óptimos para as classes \mathcal{F}_2 , \mathcal{F}_3 e \mathcal{F}_4 admitindo que as ponderações do vector π são iguais encontram-se resumidos nas seguintes tabelas*

m	Pontos	Pesos respectivos	$\text{eff}_1^D/\text{eff}_2^D/\text{eff}_3^D/\text{eff}_4^D(\xi)$
2	-1, 0, 1	0.389, 0.222, 0, 389	0.881/0.968/ - / -
3	-1, -0.401, 0.401, 1	0.319, 0.181, 0.181, 0.319	0.835/0.914/0.954/ -
4	-1, -0.605, 0, 0.605, 1	0.271, 0.152, 0.153, 0.152, 0.271	0.809/0.883/0.927/0.949

Tabela 8 – Planeamentos robustos do modelo Ξ_0^π -óptimos

A avaliação do modelo de regressão polinomial mais parcimonioso para descrever um fenómeno pode então ser feito em duas etapas admitindo que o grau na regressão polinomial não ultrapassa um determinado valor predeterminado. Em geral, não se recorre a graus muito elevados pois os polinómios de grau elevado podem ser muito instáveis. Por exemplo, considerar o grau quatro como limite pode ser uma escolha razoável para a generalidade das situações.

Porém, razões de ordem económica, de tempo ou outras apelam frequentemente a um processo mais célere. A nossa sugestão é que para um grau máximo pré-fixado m se combine o planeamento discriminante Ψ_0^π -óptimo para a classe F_m com o planeamento robusto do modelo Ξ_0^π -óptimo para a mesma classe.

Assim, na nossa opinião, deve ser utilizada uma combinação dos dois critérios,

isto é, à maximização da média ponderada

$$\begin{aligned} \Theta_0^\pi &= \prod_{l=1}^m (\text{eff}_l^{D_1}(\xi))^{\pi_l} \prod_{j=1}^m (\text{eff}_j^D(\xi))^{\pi_j} \\ &= \prod_{l=1}^m \left(2^{2^{l-2}} \frac{|M_l(\xi)|}{|M_{l-1}(\xi)|} \right)^{\pi_l} \prod_{j=1}^m \left(\frac{M_j(\xi)}{M_j(\xi_j^D)} \right)^{\pi_j/(j+1)} \end{aligned} \quad (20)$$

Definição 2.6 Dado um vector $\pi = (\pi_1, \dots, \pi_m)$ em que $\sum_{l=1}^m \pi_l = 1$ e $\pi_i > 0$, um planeamento $\xi_{0,\pi}$ é designado por **planeamento misto Θ_0^π -óptimo para a classe F_m** com respeito ao vector π se, e só se, maximiza a média geométrica ponderada Θ_0^π definida em (20).

Uma listagem dos planeamentos mistos Θ_0^π -óptimos para $m = 2, 3, 4$ encontra-se nos Anexos. Vamos comparar a eficiência entre os 3 tipos de planeamento considerados. Para esses três valores de m vamos calcular os valores de Ψ_0^π e Ξ_0^π definidos anteriormente.

Tabela 9 – Valores de Ψ_0^π para $m = 2$ e $\pi = (a, 1 - a)$.

a	$\Psi_0^\pi(D)$	$\Psi_0^\pi(R)$	$\Psi_0^\pi(M)$	$100 \times [\Psi_0^\pi(R) - \Psi_0^\pi(D)]$	$100 \times [\Psi_0^\pi(M) - \Psi_0^\pi(D)]$
0.05	0.967	0.866	0.955	-10.076	-1.110
0.10	0.935	0.844	0.926	-9.118	-0.995
0.15	0.907	0.824	0.898	-8.233	-0.890
0.20	0.880	0.806	0.872	-7.414	-0.795
0.25	0.856	0.790	0.849	-6.657	-0.707
0.30	0.834	0.775	0.828	-5.958	-0.627
0.35	0.814	0.761	0.809	-5.312	-0.554
0.40	0.797	0.750	0.792	-4.714	-0.487
0.45	0.782	0.741	0.778	-4.162	-0.426
0.50	0.770	0.733	0.766	-3.650	-0.370
0.55	0.760	0.728	0.757	-3.177	-0.320

(continua na página seguinte)

2.3 Planeamentos robustos de modelos e discriminantes

Tabela 9 – Valores de Ψ_0^π para $m = 2$ e $\pi = (a, 1 - a)$.

0.60	0.753	0.726	0.751	-2.739	-0.273
0.65	0.750	0.727	0.748	-2.331	-0.230
0.70	0.751	0.731	0.749	-1.951	-0.191
0.75	0.757	0.741	0.755	-1.596	-0.155
0.80	0.768	0.756	0.767	-1.262	-0.121
0.85	0.789	0.779	0.788	-0.943	-0.090
0.90	0.822	0.815	0.821	-0.635	-0.060
0.95	0.877	0.873	0.876	-0.328	-0.031

a	$\Xi_0^\pi(D)$	$\Xi_0^\pi(R)$	$\Xi_0^\pi(M)$	$100 \times [\Xi_0^\pi(D) - \Xi_0^\pi(R)]$	$100 \times [\Xi_0^\pi(M) - \Xi_0^\pi(R)]$
0.05	0.939	0.990	0.967	-5.084	-2.343
0.10	0.934	0.981	0.959	-4.679	-2.161
0.15	0.929	0.972	0.952	-4.293	-1.986
0.20	0.924	0.963	0.945	-3.923	-1.819
0.25	0.919	0.955	0.938	-3.570	-1.658
0.30	0.915	0.947	0.932	-3.233	-1.505
0.35	0.911	0.941	0.927	-2.912	-1.359
0.40	0.908	0.934	0.922	-2.607	-1.219
0.45	0.906	0.929	0.918	-2.317	-1.086
0.50	0.904	0.924	0.914	-2.042	-0.959
0.55	0.902	0.920	0.912	-1.781	-0.839
0.60	0.902	0.917	0.910	-1.534	-0.724
0.65	0.903	0.916	0.910	-1.301	-0.616
0.70	0.905	0.916	0.911	-1.081	-0.513
0.75	0.909	0.917	0.913	-0.873	-0.416
0.80	0.914	0.921	0.918	-0.678	-0.324
0.85	0.923	0.928	0.926	-0.495	-0.237
0.90	0.937	0.940	0.938	-0.321	-0.154
0.95	0.957	0.959	0.958	-0.158	-0.076

Tabela 10 – Valores de Ξ_0^π para $m = 2$ e $\pi = (a, 1 - a)$.

2.3 Planeamentos robustos de modelos e discriminantes

a	b	$\Psi_0^\pi(D)$	$\Psi_0^\pi(R)$	$\Psi_0^\pi(M)$	$100 \times [\Psi_0^\pi(R) - \Psi_0^\pi(D)]$	$100 \times [\Psi_0^\pi(M) - \Psi_0^\pi(D)]$
0.1	0.1	0.875	0.772	0.866	-10.305	-0.976
0.1	0.2	0.825	0.736	0.816	-8.885	-0.862
0.1	0.3	0.783	0.705	0.775	-7.788	-0.778
0.1	0.4	0.751	0.681	0.744	-6.989	-0.721
0.1	0.5	0.730	0.665	0.723	-6.478	-0.688
0.1	0.6	0.722	0.659	0.715	-6.266	-0.683
0.1	0.7	0.733	0.669	0.726	-6.405	-0.708
0.1	0.8	0.776	0.706	0.769	-7.058	-0.781
0.2	0.1	0.824	0.741	0.817	-8.285	-0.758
0.2	0.2	0.777	0.707	0.771	-7.049	-0.665
0.2	0.3	0.740	0.679	0.734	-6.133	-0.600
0.2	0.4	0.714	0.659	0.708	-5.518	-0.561
0.2	0.5	0.700	0.648	0.695	-5.206	-0.548
0.2	0.6	0.704	0.652	0.699	-5.232	-0.564
0.2	0.7	0.739	0.681	0.733	-5.724	-0.621
0.3	0.1	0.781	0.716	0.776	-6.556	-0.580
0.3	0.2	0.738	0.684	0.733	-5.485	-0.504
0.3	0.3	0.706	0.659	0.702	-4.741	-0.456
0.3	0.4	0.686	0.643	0.682	-4.314	-0.435
0.3	0.5	0.683	0.641	0.679	-4.227	-0.442
0.3	0.6	0.709	0.663	0.704	-4.577	-0.487
0.4	0.1	0.747	0.696	0.743	-5.077	-0.435
0.4	0.2	0.708	0.666	0.704	-4.157	-0.373
0.4	0.3	0.681	0.645	0.678	-3.587	-0.342
0.4	0.4	0.671	0.637	0.667	-3.374	-0.340
0.4	0.5	0.687	0.651	0.683	-3.591	-0.373
0.5	0.1	0.722	0.684	0.719	-3.809	-0.316
0.5	0.2	0.687	0.657	0.685	-3.038	-0.269
0.5	0.3	0.668	0.642	0.666	-2.667	-0.255
0.5	0.4	0.675	0.647	0.672	-2.745	-0.278
0.6	0.1	0.708	0.681	0.706	-2.718	-0.219
0.6	0.2	0.679	0.657	0.677	-2.116	-0.188
0.6	0.3	0.674	0.654	0.672	-2.024	-0.198
0.7	0.1	0.708	0.690	0.706	-1.775	-0.141
0.7	0.2	0.689	0.675	0.688	-1.424	-0.131
0.8	0.1	0.729	0.719	0.728	-0.982	-0.080
1/3	1/3	0.689	0.648	0.685	-4.174	-0.407

Tabela 11 – Valores de Ψ_0^π para $m = 3$ e $\pi = (a, b, 1 - a - b)$.

2.3 Planeamentos robustos de modelos e discriminantes

a	b	$\Xi_0^\pi(D)$	$\Xi_0^\pi(R)$	$\Xi_0^\pi(M)$	$100 \times [\Xi_0^\pi(D) - \Xi_0^\pi(R)]$	$100 \times [\Xi_0^\pi(M) - \Xi_0^\pi(R)]$
0.1	0.1	0.913	0.962	0.938	-4.996	-2.433
0.1	0.2	0.905	0.951	0.929	-4.597	-2.208
0.1	0.3	0.898	0.941	0.920	-4.288	-2.029
0.1	0.4	0.892	0.932	0.914	-4.070	-1.898
0.1	0.5	0.887	0.927	0.909	-3.945	-1.817
0.1	0.6	0.885	0.924	0.907	-3.920	-1.790
0.1	0.7	0.887	0.927	0.909	-4.005	-1.823
0.1	0.8	0.897	0.939	0.920	-4.221	-1.929
0.2	0.1	0.900	0.941	0.921	-4.076	-2.005
0.2	0.2	0.893	0.930	0.912	-3.713	-1.797
0.2	0.3	0.887	0.922	0.905	-3.449	-1.641
0.2	0.4	0.882	0.915	0.900	-3.288	-1.538
0.2	0.5	0.880	0.912	0.897	-3.234	-1.492
0.2	0.6	0.881	0.914	0.898	-3.296	-1.510
0.2	0.7	0.889	0.924	0.908	-3.494	-1.603
0.3	0.1	0.890	0.922	0.906	-3.258	-1.620
0.3	0.2	0.884	0.913	0.899	-2.930	-1.429
0.3	0.3	0.879	0.906	0.893	-2.716	-1.297
0.3	0.4	0.875	0.902	0.889	-2.621	-1.227
0.3	0.5	0.876	0.902	0.890	-2.654	-1.226
0.3	0.6	0.882	0.911	0.898	-2.831	-1.305
0.4	0.1	0.881	0.906	0.894	-2.534	-1.273
0.4	0.2	0.876	0.899	0.888	-2.245	-1.101
0.4	0.3	0.873	0.894	0.884	-2.090	-0.999
0.4	0.4	0.872	0.893	0.883	-2.080	-0.973
0.4	0.5	0.878	0.900	0.890	-2.229	-1.034
0.5	0.1	0.875	0.894	0.884	-1.898	-0.963
0.5	0.2	0.872	0.888	0.880	-1.654	-0.814
0.5	0.3	0.871	0.887	0.879	-1.579	-0.752
0.5	0.4	0.875	0.892	0.884	-1.688	-0.789
0.6	0.1	0.873	0.887	0.880	-1.344	-0.687
0.6	0.2	0.872	0.884	0.878	-1.164	-0.571
0.6	0.3	0.876	0.888	0.882	-1.209	-0.572
0.7	0.1	0.877	0.886	0.882	-0.868	-0.446
0.7	0.2	0.881	0.889	0.885	-0.798	-0.385
0.8	0.1	0.892	0.897	0.895	-0.482	-0.244
1/3	1/3	0.875	0.900	0.888	-2.458	-1.166

Tabela 12 – Valores de Ξ_0^π para $m = 3$ e $\pi = (a, b, 1 - a - b)$.

2.3 Planeamentos robustos de modelos e discriminantes

Tabela 13 – Valores de Ψ_0^π para $m = 4$ e $\pi = (a, b, c, 1 - a - b - c)$.

a	b	c	$\Psi_0^\pi(D)$	$\Psi_0^\pi(R)$	$\Psi_0^\pi(M)$	$100 \times [\Psi_0^\pi(R) - \Psi_0^\pi(D)]$	$100 \times [\Psi_0^\pi(M) - \Psi_0^\pi(D)]$
0.1	0.1	0.1	0.820	0.722	0.811	-9.740	-0.814
0.1	0.1	0.2	0.773	0.688	0.766	-8.483	-0.741
0.1	0.1	0.3	0.736	0.660	0.729	-7.586	-0.693
0.1	0.1	0.4	0.710	0.640	0.703	-7.035	-0.668
0.1	0.1	0.5	0.696	0.628	0.690	-6.839	-0.668
0.1	0.1	0.6	0.700	0.630	0.693	-7.059	-0.696
0.1	0.1	0.7	0.735	0.656	0.727	-7.876	-0.768
0.1	0.2	0.1	0.772	0.691	0.765	-8.138	-0.698
0.1	0.2	0.2	0.730	0.659	0.724	-7.068	-0.639
0.1	0.2	0.3	0.698	0.634	0.692	-6.369	-0.607
0.1	0.2	0.4	0.679	0.618	0.673	-6.043	-0.599
0.1	0.2	0.5	0.676	0.614	0.669	-6.132	-0.619
0.1	0.2	0.6	0.701	0.633	0.694	-6.787	-0.681
0.1	0.3	0.1	0.734	0.665	0.728	-6.914	-0.620
0.1	0.3	0.2	0.696	0.636	0.690	-6.013	-0.574
0.1	0.3	0.3	0.669	0.614	0.664	-5.519	-0.555
0.1	0.3	0.4	0.659	0.604	0.653	-5.459	-0.566
0.1	0.3	0.5	0.675	0.616	0.669	-5.960	-0.617
0.1	0.4	0.1	0.705	0.644	0.699	-6.039	-0.575
0.1	0.4	0.2	0.671	0.618	0.665	-5.306	-0.540
0.1	0.4	0.3	0.652	0.602	0.647	-5.053	-0.538
0.1	0.4	0.4	0.659	0.605	0.653	-5.387	-0.577
0.1	0.5	0.1	0.686	0.631	0.681	-5.500	-0.559
0.1	0.5	0.2	0.658	0.608	0.652	-4.962	-0.537
0.1	0.5	0.3	0.653	0.602	0.647	-5.079	-0.561
0.1	0.6	0.1	0.681	0.627	0.675	-5.312	-0.574
0.1	0.6	0.2	0.662	0.612	0.657	-5.088	-0.573
0.1	0.7	0.1	0.695	0.640	0.689	-5.571	-0.625
0.2	0.1	0.1	0.772	0.694	0.766	-7.742	-0.622
0.2	0.1	0.2	0.730	0.663	0.724	-6.673	-0.565
0.2	0.1	0.3	0.698	0.638	0.692	-5.970	-0.532
0.2	0.1	0.4	0.678	0.622	0.673	-5.632	-0.524
0.2	0.1	0.5	0.675	0.618	0.670	-5.698	-0.541
0.2	0.1	0.6	0.700	0.637	0.694	-6.309	-0.596
0.2	0.2	0.1	0.729	0.665	0.723	-6.330	-0.525
0.2	0.2	0.2	0.691	0.636	0.686	-5.448	-0.483
0.2	0.2	0.3	0.664	0.615	0.660	-4.959	-0.466
0.2	0.2	0.4	0.654	0.605	0.649	-4.888	-0.477
0.2	0.2	0.5	0.670	0.617	0.665	-5.352	-0.523
0.2	0.3	0.1	0.695	0.642	0.690	-5.301	-0.467
0.2	0.3	0.2	0.661	0.615	0.657	-4.598	-0.437
0.2	0.3	0.3	0.643	0.599	0.638	-4.354	-0.437
0.2	0.3	0.4	0.649	0.602	0.644	-4.663	-0.474
0.2	0.4	0.1	0.671	0.625	0.667	-4.632	-0.441

(continua na página seguinte)

2.3 Planeamentos robustos de modelos e discriminantes

Tabela 13 – Valores de Ψ_0^π para $m = 4$ e $\pi = (a, b, c, 1 - a - b - c)$.

0.2	0.4	0.2	0.643	0.602	0.639	-4.131	-0.424
0.2	0.4	0.3	0.639	0.596	0.634	-4.244	-0.449
0.2	0.5	0.1	0.660	0.617	0.656	-4.330	-0.447
0.2	0.5	0.2	0.642	0.601	0.638	-4.135	-0.451
0.2	0.6	0.1	0.668	0.624	0.663	-4.468	-0.489
0.3	0.1	0.1	0.732	0.672	0.728	-6.028	-0.466
0.3	0.1	0.2	0.694	0.643	0.690	-5.135	-0.423
0.3	0.1	0.3	0.668	0.621	0.664	-4.630	-0.406
0.3	0.1	0.4	0.657	0.612	0.653	-4.536	-0.413
0.3	0.1	0.5	0.674	0.624	0.669	-4.961	-0.454
0.3	0.2	0.1	0.693	0.645	0.689	-4.790	-0.387
0.3	0.2	0.2	0.659	0.618	0.656	-4.094	-0.359
0.3	0.2	0.3	0.641	0.603	0.637	-3.844	-0.359
0.3	0.2	0.4	0.647	0.606	0.643	-4.124	-0.393
0.3	0.3	0.1	0.664	0.624	0.660	-3.952	-0.347
0.3	0.3	0.2	0.636	0.601	0.633	-3.471	-0.333
0.3	0.3	0.3	0.632	0.596	0.628	-3.573	-0.357
0.3	0.4	0.1	0.647	0.612	0.643	-3.511	-0.342
0.3	0.4	0.2	0.630	0.596	0.626	-3.336	-0.348
0.3	0.5	0.1	0.648	0.613	0.645	-3.523	-0.374
0.4	0.1	0.1	0.701	0.655	0.698	-4.557	-0.341
0.4	0.1	0.2	0.667	0.629	0.664	-3.838	-0.311
0.4	0.1	0.3	0.649	0.613	0.645	-3.560	-0.308
0.4	0.1	0.4	0.655	0.617	0.652	-3.802	-0.337
0.4	0.2	0.1	0.665	0.631	0.663	-3.484	-0.277
0.4	0.2	0.2	0.638	0.608	0.635	-3.000	-0.265
0.4	0.2	0.3	0.633	0.603	0.630	-3.080	-0.286
0.4	0.3	0.1	0.642	0.614	0.640	-2.859	-0.257
0.4	0.3	0.2	0.625	0.598	0.622	-2.690	-0.265
0.4	0.4	0.1	0.637	0.609	0.634	-2.724	-0.278
0.5	0.1	0.1	0.679	0.646	0.676	-3.289	-0.240
0.5	0.1	0.2	0.650	0.623	0.648	-2.769	-0.224
0.5	0.1	0.3	0.646	0.618	0.644	-2.809	-0.240
0.5	0.2	0.1	0.648	0.624	0.646	-2.398	-0.193
0.5	0.2	0.2	0.630	0.608	0.628	-2.215	-0.201
0.5	0.3	0.1	0.634	0.613	0.632	-2.067	-0.198
0.6	0.1	0.1	0.667	0.645	0.666	-2.200	-0.159
0.6	0.1	0.2	0.650	0.630	0.648	-1.969	-0.161
0.6	0.2	0.1	0.644	0.628	0.643	-1.570	-0.136
0.7	0.1	0.1	0.672	0.659	0.671	-1.305	-0.098
0.25	0.25	0.25	0.648	0.606	0.644	-4.137	-0.391

2.3 Planeamentos robustos de modelos e discriminantes

Tabela 14 – Valores de Ξ_0^π para $m = 4$ e $\pi = (a, b, c, 1 - a - b - c)$.

a	b	c	$\Xi_0^\pi(D)$	$\Xi_0^\pi(R)$	$\Xi_0^\pi(M)$	$100 \times [\Xi_0^\pi(D) - \Xi_0^\pi(R)]$	$100 \times [\Xi_0^\pi(M) - \Xi_0^\pi(R)]$
0.1	0.1	0.1	0.901	0.947	0.924	-4.527	-2.284
0.1	0.1	0.2	0.896	0.938	0.917	-4.274	-2.114
0.1	0.1	0.3	0.890	0.931	0.911	-4.113	-1.998
0.1	0.1	0.4	0.886	0.926	0.907	-4.048	-1.939
0.1	0.1	0.5	0.883	0.923	0.904	-4.084	-1.941
0.1	0.1	0.6	0.882	0.924	0.904	-4.229	-2.010
0.1	0.1	0.7	0.887	0.932	0.910	-4.503	-2.158
0.1	0.2	0.1	0.892	0.932	0.912	-4.024	-1.998
0.1	0.2	0.2	0.887	0.925	0.906	-3.811	-1.849
0.1	0.2	0.3	0.882	0.919	0.901	-3.705	-1.762
0.1	0.2	0.4	0.878	0.916	0.898	-3.711	-1.743
0.1	0.2	0.5	0.877	0.916	0.898	-3.839	-1.797
0.1	0.2	0.6	0.881	0.922	0.903	-4.106	-1.937
0.1	0.3	0.1	0.883	0.920	0.902	-3.662	-1.779
0.1	0.3	0.2	0.878	0.913	0.897	-3.492	-1.653
0.1	0.3	0.3	0.875	0.909	0.893	-3.450	-1.603
0.1	0.3	0.4	0.873	0.909	0.892	-3.546	-1.634
0.1	0.3	0.5	0.876	0.914	0.896	-3.798	-1.762
0.1	0.4	0.1	0.875	0.910	0.893	-3.439	-1.628
0.1	0.4	0.2	0.872	0.905	0.890	-3.319	-1.531
0.1	0.4	0.3	0.870	0.903	0.888	-3.361	-1.529
0.1	0.4	0.4	0.872	0.908	0.891	-3.585	-1.636
0.1	0.5	0.1	0.869	0.903	0.887	-3.355	-1.549
0.1	0.5	0.2	0.868	0.901	0.886	-3.302	-1.491
0.1	0.5	0.3	0.869	0.904	0.888	-3.472	-1.564
0.1	0.6	0.1	0.867	0.901	0.886	-3.412	-1.546
0.1	0.6	0.2	0.869	0.904	0.888	-3.477	-1.556
0.1	0.7	0.1	0.872	0.908	0.892	-3.636	-1.635
0.2	0.1	0.1	0.888	0.924	0.906	-3.630	-1.854
0.2	0.1	0.2	0.882	0.917	0.900	-3.407	-1.702
0.2	0.1	0.3	0.878	0.911	0.895	-3.288	-1.610
0.2	0.1	0.4	0.874	0.907	0.891	-3.278	-1.583
0.2	0.1	0.5	0.873	0.907	0.891	-3.385	-1.627
0.2	0.1	0.6	0.877	0.913	0.895	-3.626	-1.754
0.2	0.2	0.1	0.879	0.911	0.895	-3.171	-1.590
0.2	0.2	0.2	0.875	0.905	0.890	-2.996	-1.462
0.2	0.2	0.3	0.871	0.900	0.886	-2.945	-1.407
0.2	0.2	0.4	0.869	0.899	0.885	-3.028	-1.432
0.2	0.2	0.5	0.872	0.904	0.889	-3.261	-1.549
0.2	0.3	0.1	0.872	0.900	0.886	-2.870	-1.400
0.2	0.3	0.2	0.868	0.896	0.883	-2.749	-1.302
0.2	0.3	0.3	0.866	0.894	0.881	-2.785	-1.296
0.2	0.3	0.4	0.868	0.898	0.884	-2.996	-1.395
0.2	0.4	0.1	0.866	0.893	0.880	-2.727	-1.288

(continua na página seguinte)

2.3 Planeamentos robustos de modelos e discriminantes

Tabela 14 – Valores de Ξ_0^π para $m = 4$ e $\pi = (a, b, c, 1 - a - b - c)$.

0.2	0.4	0.2	0.864	0.891	0.878	-2.677	-1.230
0.2	0.4	0.3	0.865	0.893	0.880	-2.841	-1.299
0.2	0.5	0.1	0.863	0.890	0.878	-2.743	-1.259
0.2	0.5	0.2	0.864	0.892	0.879	-2.809	-1.269
0.2	0.6	0.1	0.866	0.896	0.882	-2.940	-1.330
0.3	0.1	0.1	0.876	0.904	0.890	-2.842	-1.469
0.3	0.1	0.2	0.871	0.898	0.884	-2.651	-1.336
0.3	0.1	0.3	0.867	0.893	0.881	-2.583	-1.273
0.3	0.1	0.4	0.866	0.892	0.879	-2.643	-1.287
0.3	0.1	0.5	0.868	0.897	0.883	-2.848	-1.391
0.3	0.2	0.1	0.869	0.893	0.881	-2.429	-1.227
0.3	0.2	0.2	0.865	0.888	0.877	-2.300	-1.125
0.3	0.2	0.3	0.863	0.886	0.875	-2.322	-1.112
0.3	0.2	0.4	0.864	0.889	0.877	-2.513	-1.202
0.3	0.3	0.1	0.863	0.885	0.874	-2.198	-1.071
0.3	0.3	0.2	0.861	0.882	0.872	-2.145	-1.011
0.3	0.3	0.3	0.861	0.884	0.874	-2.298	-1.074
0.3	0.4	0.1	0.860	0.881	0.871	-2.152	-1.008
0.3	0.4	0.2	0.860	0.882	0.872	-2.218	-1.016
0.3	0.5	0.1	0.862	0.885	0.874	-2.309	-1.053
0.4	0.1	0.1	0.866	0.888	0.876	-2.153	-1.125
0.4	0.1	0.2	0.862	0.882	0.872	-2.003	-1.015
0.4	0.1	0.3	0.860	0.880	0.870	-2.001	-0.991
0.4	0.1	0.4	0.862	0.883	0.873	-2.161	-1.066
0.4	0.2	0.1	0.861	0.879	0.870	-1.793	-0.909
0.4	0.2	0.2	0.859	0.876	0.867	-1.728	-0.843
0.4	0.2	0.3	0.859	0.878	0.869	-1.862	-0.896
0.4	0.3	0.1	0.858	0.874	0.866	-1.652	-0.797
0.4	0.3	0.2	0.858	0.875	0.867	-1.711	-0.802
0.4	0.4	0.1	0.859	0.877	0.869	-1.749	-0.808
0.5	0.1	0.1	0.860	0.875	0.867	-1.555	-0.819
0.5	0.1	0.2	0.857	0.872	0.865	-1.463	-0.742
0.5	0.1	0.3	0.858	0.874	0.866	-1.563	-0.779
0.5	0.2	0.1	0.857	0.870	0.864	-1.266	-0.638
0.5	0.2	0.2	0.857	0.870	0.864	-1.309	-0.634
0.5	0.3	0.1	0.859	0.871	0.865	-1.265	-0.597
0.6	0.1	0.1	0.859	0.869	0.863	-1.045	-0.551
0.6	0.1	0.2	0.859	0.869	0.864	-1.050	-0.531
0.6	0.2	0.1	0.860	0.869	0.865	-0.878	-0.430
0.7	0.1	0.1	0.865	0.872	0.868	-0.637	-0.330
0.25	0.25	0.25	0.865	0.890	0.878	-2.503	-1.190

O planeamento misto Θ_0^π -óptimo perde comparativamente com o planeamento discriminante Ψ_0^π -óptimo, no máximo, de 1% de eficiência na estimação do grau do polinómio a utilizar admitindo que faz sentido utilizar regressão polinomial até ao grau m . É assim um planeamento com um bom comportamento nessa área. Comparando o planeamento óptimo na estimação dos parâmetros do modelo de regressão polinomial com o planeamento misto Θ_0^π -óptimo verifica-se que, em geral, a perda de eficiência não ultrapassa os 2%. Quando se analisa a perda de eficiência dos planeamentos discriminantes Ψ_0^π -óptimos na estimação dos parâmetros do modelo e dos planeamentos robustos do modelo Ξ_0^π -óptimos na estimação do coeficiente do termo de maior grau relevante verifica-se que essa perda pode ultrapassar os 10%! Assim, os planeamentos mistos Θ_0^π -óptimos apresentam-se como alternativas válidas à utilização de dois planeamentos na medida em que permite, numa só etapa, estimar o coeficiente do termo de maior grau relevante no modelo e simultaneamente estimar todos os coeficientes do modelo de forma quase óptima.

3 Transformações e o método delta

3.1 Porquê transformar?

Todos os resultados estatísticos e, mais geralmente, de toda a ciência, são formulados com base em premissas ou hipóteses. É exemplo disso, no estudo das distribuições, o facto da distribuição da variável fulcral $\frac{(n-1)S_C^2}{\sigma^2}$ ser a distribuição qui-quadrado com $n-1$ graus de liberdade caso a população parente apresente um comportamento gaussiano. Ou a vasta panóplia de resultados da regressão linear que se encontra geralmente alicerçada em pressupostos que envolvem a normalidade, homocedasticidade e independência dos dados.

O problema de aplicação de toda a teoria construída é que, com uma frequência maior que o desejável, os dados reais obtidos não “obedecem” às hipóteses de que partem os resultados teóricos. A realidade apresenta uma diversidade muito superior aos comportamentos actualmente descritos pela Estatística. Perante este cenário, o investigador pode ter uma de duas atitudes:

- para cada caso desenvolver uma bateria de resultados que permita uma boa interpretação dos dados;
- transformar os dados de forma a que se enquadrem dentro das premissas já estudadas.

Perante estas duas possibilidades, parece preferível optar, se possível, pela segunda. O primeiro cenário implica um processo hercúleo de tentar adequar todas as situações reais a uma base teórica específica que pode redundar numa tarefa morosa e incapaz de, em tempo útil, dar resposta às necessidades. No segundo cenário, encara-se a situação pela positiva, na medida em que a inexistência de uma base

teórica específica não é necessariamente uma dificuldade. A teoria existente pode ser suficiente para dar resposta aos problemas.

A transformação de dados é uma forma de “adaptar” os dados recolhidos ao conhecimento existente. A unidade escolhida para a re-expressão dos dados deve, sempre que possível, permitir uma fácil interpretação dos dados transformados (Box e Cox, 1964).

Embora os pressupostos para a análise de dados resultantes da observação e medição do investigador sejam diversificados são geralmente cinco as premissas que devem ser observadas para uma análise mais facilitada. Estes são:

- aditividade dos efeitos;
- homocedasticidade;
- independência;
- linearidade;
- simetria, e preferencialmente, um comportamento gaussiano dos dados.

Felizmente, não é necessário utilizar cinco transformações para que os dados verifiquem essas condições. Tukey (1957) afirma que frequentemente ao aplicar uma transformação com um objectivo atingimos os outros objectivos simultaneamente.

Hoaglin *et al.* (1992) explora num par de páginas o que designa de coincidência sortuda entre uma redução de dispersão dos dados (atingindo uma situação próxima de homocedasticidade) simultaneamente com a simetria dos dados. Aponta como factor mais provável dessa “boa sorte” o facto de a maior parte dos dados ter no zero uma fronteira natural dos valores estatísticos na maior parte de medições e contagens. Esse limite à esquerda e a inexistência de um limite à direita são a causa

de uma menor dispersão dos dados e de uma assimetria positiva. Transformações usuais como a logaritimização tendem a comprimir os dados de forma não linear sendo maior essa compressão para os valores mais elevados o que terá como consequência uma maior simetria dos dados transformados.

3.2 Algumas transformações de variáveis aleatórias

A utilização de transformações de variáveis aleatórias encontra suporte no facto de as densidades da variável aleatória inicial e da variável aleatória transformada estarem relacionadas.

Teorema 3.1 *Seja X uma variável aleatória contínua com função densidade f_X e seja g uma função diferenciável em \mathbb{R} tal que g' tem sinal constante e g é invertível no suporte de X . Então $Y = g(X)$ é uma variável aleatória contínua com função densidade de probabilidade f_Y dada por*

$$\begin{aligned} f_Y(x) &= f_X(g^{-1}(x)) \left| \frac{d}{dx}(g^{-1}(x)) \right| \\ &= f_X(g^{-1}(x)) \left| \frac{1}{g'(g^{-1}(x))} \right|. \end{aligned}$$

Uma transformação muito comum é a transformação que altera os parâmetros de localização e de escala mantendo a forma, isto é

$$\begin{aligned} g: \mathbb{R} &\rightarrow \mathbb{R} \\ x &\mapsto g(x) = \frac{x - \mu}{\sigma} \end{aligned} \tag{21}$$

com $\sigma > 0$ e $\mu \in \mathbb{R}$. Este tipo de **transformações lineares** ou de localização e escala têm subjacente a transformação das variáveis em estudo em variáveis do mesmo tipo (de Kintchine). Estas transformações foram muito utilizadas na modelação de

dados pela gaussiana padrão até ser dado o devido enfoque à assimetria e ao peso das caudas como discutiremos na secção seguinte.

Para alterar a forma de uma distribuição utilizam-se frequentemente **transformações potência** para variáveis aleatórias com suporte positivo. Na prática, quando temos uma amostra aleatória (X_1, \dots, X_n) , podemos previamente considerar a transformação

$$(X_1, \dots, X_n) \rightarrow (X_1 - X_{1:n}, \dots, X_n - X_{1:n})$$

onde $X_{1:n}$ é o mínimo da amostra.

As transformações potência constituem uma família caracterizada por

$$g^\lambda : D \subseteq \mathbb{R}^+ \rightarrow \mathbb{R}$$

$$x \mapsto g^\lambda(x) = \frac{x^\lambda - 1}{\lambda}$$

com $\lambda \in \mathbb{R}$, considerando que o caso $\lambda = 0$ corresponde a

$$\lim_{\lambda \rightarrow 0} \frac{x^\lambda - 1}{\lambda} = \lim_{\lambda \rightarrow 0} \frac{x^\lambda \ln x}{1} = \ln x.$$

Esta família de transformações altera, para $\lambda \neq 1$, a forma da distribuição de uma variável aleatória. É muitas vezes nesta família de funções que se procura um elemento que permita transformar os dados de forma a que se adequem às premissas enumeradas anteriormente. Hoaglin, Mosteller e Tukey (1983) discutem, em pormenor, a relevância destas transformações dando pistas para uma boa escolha do parâmetro λ .

Um outro tipo de transformações amplamente estudadas que estão na base dos métodos de Monte Carlo (Fishman, 2005) têm por base o teorema da transformação uniformizante:

Teorema 3.2 *Seja X uma variável aleatória com distribuição F_X estritamente crescente. Então $Y = F_X(X)$ tem distribuição uniforme, $U(0, 1)$.*

Não deixa de ser interessante que uma variável aleatória com uma função densidade tão simples esteja na base da simulação. Este resultado aponta um método para a criação de números pseudo-aleatórios para qualquer distribuição a partir de números pseudo-aleatórios de uma distribuição uniforme $(0, 1)$ pois

$$X \stackrel{d}{=} F^{-1}(Y).$$

O problema resume-se a encontrar uma expressão para F^{-1} . Caso F não seja injectiva pode-se considerar a função inversa generalizada de F (Gut, 1991).

3.3 Derivadas e séries de Taylor

O método delta que abordaremos nas secções seguintes utiliza truncaturas da série de Taylor pelo que neste momento iremos apresentar algumas definições e teoremas relativos às séries de Taylor.

Definição 3.1 *Considere-se uma função $f : U \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$. Se U for aberto diz-se que f é de **classe** C^k em U , ou abreviadamente $f \in C^k(U)$, se todas as derivadas parciais de ordem k de f existirem e forem contínuas em U . No caso de f admitir as derivadas de todas as ordens utiliza-se a notação $f \in C^\infty$.*

A existência e continuidade das derivadas parciais de ordem k implica a existência e continuidade das derivadas de ordem inferior a k . No decurso desta dissertação, dados $n \in \mathbb{N}$ e $f : U \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$, a derivada parcial $\frac{\partial f}{\partial x_i}(x_1, \dots, x_n)$ será representada apenas por $\frac{\partial f}{\partial x_i}$ quando tal representação não deixar dúvidas sobre o ponto (x_1, \dots, x_n) de \mathbb{R}^n em que estamos a calculá-la.

Definição 3.2 *Dados uma função real de variável real f tal que $f \in C^\infty(U)$ e um ponto interior a de U , então o polinómio*

$$T_n(x) = \sum_{i=0}^n \frac{f^{(i)}(a)}{i!} (x-a)^i$$

é o polinómio de Taylor de ordem n da função f no ponto a .

$$T(x) = \sum_{i=0}^{\infty} \frac{f^{(i)}(a)}{i!} (x-a)^i$$

é a série de Taylor da função f no ponto a .

Em particular, se $a = 0$ a série anterior toma o nome de **série de MacLaurin**.

Definição 3.3 *Dados uma função real de variável real f tal que $f \in C^n(U)$ e um ponto interior a de U , a diferença*

$$f(x) - T_n(x) = R_n(x)$$

é o resto de Taylor de ordem n da função f no ponto c .

O teorema abaixo apresenta uma expressão conhecida para o resto de Taylor que permite majorar o erro obtido pelas aproximações por polinómios de Taylor.

Teorema 3.3 *Dados uma função real de variável real f tal que $f \in C^{n+1}(U)$ e um ponto interior a de U então o resto de Taylor de ordem n da função f nesse ponto é dado por*

$$R_n(x) = \int f^{(n+1)}(t) \frac{(x-t)^n}{n!} dt \text{ em que } t \in]\min(x, a); \max(x, a)[.$$

A definição de polinómio e de série de Taylor podem ser alargadas a uma função com suporte num subconjunto U de \mathbb{R}^k .

Definição 3.4 Dados uma função $f : U \subseteq \mathbb{R}^k \rightarrow \mathbb{R}$ tal que $f \in C^\infty(U)$ e um ponto interior $\mathbf{a} = (\mathbf{a}_1, \dots, \mathbf{a}_k)$ de U então, o polinómio

$$T_n(x_1, \dots, x_k) = \sum_{i=0}^n \frac{1}{i!} \left\{ \left[\sum_{j=1}^k (x_j - a_j) \frac{\partial}{\partial x'_j} \right]^i f(x'_1, \dots, x'_k) \right\}_{x'_1=a_1, \dots, x'_k=a_k}$$

é o *polinómio de Taylor de ordem n da função f no ponto \mathbf{a}* .

$$T(x_1, \dots, x_k) = \sum_{i=0}^{\infty} \frac{1}{i!} \left\{ \left[\sum_{j=1}^k (x_j - a_j) \frac{\partial}{\partial x'_j} \right]^i f(x'_1, \dots, x'_k) \right\}_{x'_1=a_1, \dots, x'_k=a_k}$$

é a *série de Taylor em torno de \mathbf{a}* .

3.4 Método delta

O método delta consiste, em traços gerais, em truncar a expansão em série de Taylor de uma função com vista à obtenção de aproximações, nomeadamente, para os momentos de uma estatística de interesse. A prática habitual consiste em truncar a expansão após a primeira derivada. No caso de esta se anular no ponto em que é efectuada a expansão recorre-se ao termo correspondente à segunda derivada. Para demonstrar o principal teorema que suporta este método é necessário apresentar parte do enunciado do teorema de Slutsky.

Teorema 3.4 *Sejam $\{X_n\}_{n \in \mathbb{N}}$ e $\{Y_n\}_{n \in \mathbb{N}}$ duas sucessões de variáveis aleatórias e $c \in \mathbb{R}$ tais que $X_n \xrightarrow[n \rightarrow \infty]{d} X$ e $Y_n \xrightarrow[n \rightarrow \infty]{p} c$ então:*

1. $X_n + Y_n \xrightarrow[n \rightarrow \infty]{d} X + c;$
2. $X_n Y_n \xrightarrow[n \rightarrow \infty]{d} cX.$

A demonstração deste teorema pode ser consultada em diversa bibliografia. Passamos agora ao teorema principal que suporta a aplicação do método delta.

Teorema 3.5 *Dada $\{X_n\}_{n \in \mathbb{N}}$ uma sucessão de variáveis aleatórias tal que*

$$\sqrt{n}(X_n - \theta) \xrightarrow[n \rightarrow \infty]{d} Z \sim \text{Gau}(0, \sigma),$$

uma função f e θ fixo tal que $f'(\theta)$ existe e é diferente de zero, então

$$\sqrt{n}[f(X_n) - f(\theta)] \xrightarrow[n \rightarrow \infty]{d} f'(\theta) Z \sim \text{Gau}(0, \sigma |f'(\theta)|).$$

Demonstração A expansão de $f(X_n)$ em série de Taylor de ordem 1 em torno de θ é

$$f(X_n) \stackrel{d}{=} f(\theta) + f'(\theta)(X_n - \theta) + R_1(X_n)$$

onde $R_1(X_n) \xrightarrow[n \rightarrow \infty]{p} 0$ quando $X_n \xrightarrow[n \rightarrow \infty]{p} \theta$.

Logo, as seguintes expressões são assintoticamente iguais em distribuição:

$$\sqrt{n}[f(X_n) - f(\theta)] \approx \sqrt{n}f'(\theta)(X_n - \theta). \quad (22)$$

Recorrendo ao Teorema de Slutsky é possível afirmar que

$$\sqrt{n}f'(\theta)(X_n - \theta) \xrightarrow[n \rightarrow \infty]{d} f'(\theta) Z \sim \text{Gau}(0, \sigma |f'(\theta)|)$$

donde, utilizando a aproximação (22) obtém-se o resultado pretendido. ■

As hipóteses do teorema podem ser relaxadas, nomeadamente, a constante normalizadora não tem de ser necessariamente \sqrt{n} (Bishop *et al*, 1975).

Há duas extensões do método delta bastante importantes. Uma lida com a situação em que não é verificada a hipótese $f'(\theta) \neq 0$. A outra relaciona-se com o caso em que f é uma função multivariada. Uma panorâmica sobre o tema pode ser encontrada em Chandra (1999).

O primeiro caso é suportado pelo teorema que se segue e que é conhecido como método delta de segunda ordem.

Teorema 3.6 Dadas $\{X_n\}_{n \in \mathbb{N}}$ uma sucessão de variáveis aleatórias tal que

$$\sqrt{n}(X_n - \theta) \xrightarrow[n \rightarrow \infty]{d} Z \sim \text{Gau}(0, \sigma),$$

uma função f e θ fixo tal que $f'(\theta) = 0$ e $f''(\theta)$ existe e é diferente de zero, então

$$n[f(X_n) - f(\theta)] \xrightarrow[n \rightarrow \infty]{d} \sigma^2 \frac{f''(\theta)}{2} K,$$

onde $K \sim \chi_1^2$.

Demonstração A expansão de $f(X_n)$ em série de Taylor de ordem 2 em torno de θ é

$$f(X_n) \stackrel{d}{=} f(\theta) + f'(\theta)(X_n - \theta) + \frac{f''(\theta)}{2}(X_n - \theta)^2 + R_2(X_n)$$

onde $R_2(X_n) \xrightarrow[n \rightarrow \infty]{p} 0$ quando $X_n \xrightarrow[n \rightarrow \infty]{p} \theta$.

Como $f'(\theta) = 0$,

$$f(X_n) - f(\theta) \approx \frac{f''(\theta)}{2}(X_n - \theta)^2, \quad (23)$$

em distribuição. Ora o quadrado de uma variável aleatória com distribuição $\text{Gau}(0, 1)$ é uma variável aleatória com distribuição χ_1^2 , donde

$$n \frac{(X_n - \theta)^2}{\sigma^2} \sim \chi_1^2.$$

Portanto, tal como na demonstração anterior, o Teorema de Slutsky garante que

$$\frac{nf''(\theta)}{2}(X_n - \theta)^2 \xrightarrow[n \rightarrow \infty]{d} \sigma^2 \frac{f''(\theta)}{2} K$$

onde $K \sim \chi_1^2$.

Logo pela aproximação (23) obtém-se o resultado pretendido. ■

Em algumas situações poderá ocorrer que $f'(\theta)$ bem como $f''(\theta)$ sejam simultaneamente nulas pelo que a aproximação $f(X_n) - f(\theta)$ deverá ser feita pela parcela da série de Taylor correspondente à primeira derivada não nula em θ . Obtém-se uma

convergência em distribuição para uma potência de uma variável aleatória gaussiana padrão de expoente igual à ordem dessa derivada.

O tratamento do caso de uma função dependente de duas ou mais variáveis é estabelecido pelo próximo teorema. Utilizaremos a notação $\mathbf{X}_n = (X_{n1}, \dots, X_{np})$ para representar um vector aleatório de dimensão p e a representação $\mu = (\mu_1, \dots, \mu_p)$ para um vector de parâmetros da mesma dimensão.

Teorema 3.7 *Seja \mathbf{X}_n um vector aleatório, μ vector de parâmetros com valores em $D^p \subseteq \mathbb{R}^p$ e $f : D^p \rightarrow \mathbb{R}^p$ uma função diferenciável. Se*

$$\sqrt{n} (\mathbf{X}_n^T - \mu^T) \xrightarrow[n \rightarrow \infty]{d} \mathbf{Z} \sim \text{Gau}(\mathbf{0}, W)$$

em que $\mathbf{0}$ é o vector nulo e W é a matriz das covariâncias, então

$$\sqrt{n} \left[[f(\mathbf{X}_n)]^T - [f(\mu)]^T \right] \xrightarrow[n \rightarrow \infty]{d} \mathbf{Z}^* \sim \text{Gau} \left(\mathbf{0}, \left(\frac{\partial f}{\partial \mu} \right) W \left(\frac{\partial f}{\partial \mu} \right)^T \right)$$

Demonstração Pela definição (3.4)

$$f_i(\mathbf{X}_n) \stackrel{d}{=} f_i(X_{n1}, \dots, X_{np}) \stackrel{d}{=} f_i(\mu_1, \dots, \mu_p) + \sum_{j=1}^p \frac{\partial f_i}{\partial \mu_j} (X_{nj} - \mu_j) + R_1(\mathbf{X}_n)$$

onde $R_1(\mathbf{X}_n) \rightarrow 0$ em probabilidade, quando $(X_{n1}, \dots, X_{np}) \rightarrow (\mu_1, \dots, \mu_p)$ em probabilidade. Em notação matricial pode-se representar

$$[f(\mathbf{X}_n)]^T \stackrel{d}{=} [f(\mu)]^T + (\mathbf{X}_n - \mu) \left(\frac{\partial f}{\partial \mu} \right)^T + R_1(\mathbf{X}_n).$$

Donde,

$$\sqrt{n} \left[[f(\mathbf{X}_n)]^T - [f(\mu)]^T \right] \approx \sqrt{n} (\mathbf{X}_n - \mu) \left(\frac{\partial f}{\partial \mu} \right)^T,$$

em distribuição. Portanto, por uma generalização do teorema de Slutsky,

$$\sqrt{n} \left[[f(\mathbf{X}_n)]^T - [f(\mu)]^T \right] \xrightarrow[n \rightarrow \infty]{d} \mathbf{Z}^* \sim \text{Gau} \left(\mathbf{0}, \left(\frac{\partial f}{\partial \mu} \right) W \left(\frac{\partial f}{\partial \mu} \right)^T \right)$$

como se queria demonstrar. ■

Mais detalhes sobre o método delta aplicado a funções multivariadas podem ser encontrados em Lehmann e Casella (2001).

3.5 Extensões do método delta

Nesta secção analisamos o caso em que na truncatura da série de Taylor utilizamos um número maior de termos. Nas condições do Teorema 3.6 mas admitindo que $f'(\theta) \neq 0$ o que estará em jogo é a soma de duas variáveis aleatórias dependentes. Porém, trata-se de uma situação de dependência relativamente simples de analisar. Além disso, utilizando os momentos ordinários de ordem mais baixa da população parente, avaliaremos o papel da dispersão, assimetria e curtose na qualidade das aproximações obtidas por esta via. Por uma questão de simplificação iremos assumir no decurso desta secção que estamos perante variáveis aleatórias absolutamente contínuas. Sancetta (2007) fornece um enquadramento para os problemas de convergência fraca que aqui serão abordados.

Tal como ilustrado para o método delta na sua forma tradicional começemos por enunciar o teorema que suporta a utilização dos termos correspondentes à primeira e à segunda derivada.

Teorema 3.8 *Nas condições do Teorema 3.5 mas admitindo que $f'(\theta) \neq 0$, temos*

$$\sqrt{n}[f(X_n) - f(\theta)] \approx |a| Z_n,$$

em distribuição, onde $a = \frac{f''(\theta)}{2}$ e Z_n é uma variável aleatória cuja função densidade é dada por

$$f_{Z_n}(z) = \sqrt{n} \frac{\exp(-nr_1^2/2\sigma^2) + \exp(-nr_2^2/2\sigma^2)}{\sigma \sqrt{2\pi(b^2 + 4z)}} I_{[b^2/4, +\infty)}(z),$$

$$\text{onde } r_1 = \frac{-b - \sqrt{b^2 + 4w}}{2}, r_2 = \frac{-b + \sqrt{b^2 + 4w}}{2} \text{ e } b = \frac{2f'(\theta)}{f''(\theta)}.$$

Demonstração Considere-se a aproximação para $f(X_n) - f(\theta)$ dada pelo polinómio de Taylor de segunda ordem

$$\begin{aligned} f(X_n) - f(\theta) &\approx f'(\theta)(X_n - \theta) + \frac{f''(\theta)}{2}(X_n - \theta)^2 \\ &= \frac{f''(\theta)}{2} [(X_n - \theta)^2 + b(X_n - \theta)] \\ &= a(Y_n^2 + bY_n), \end{aligned} \tag{24}$$

onde $Y_n = X_n - \theta$, $a = f''(\theta)/2$ e $b = 2f'(\theta)/f''(\theta)$. Em seguida, consideraremos a variável aleatória $W_n = Y_n^2 + bY_n$ onde, por hipótese, $b \neq 0$. O suporte de W_n está contido em $\left[-\frac{b^2}{4}, +\infty\right[$ que é o contradomínio de $f(x) = bx + x^2$ e

$$F_{W_n}(w) = P(Y_n^2 + bY_n - w \leq 0).$$

Identificando $Y_n^2 + bY_n - w$ a $g(Y_n)$ onde $g(x) = x^2 + bx - w$ e atendendo às raízes reais da função, caso existam, designamo-las por r_1 e r_2 onde $r_1 \leq r_2$, $r_1 = \frac{-b - \sqrt{b^2 + 4w}}{2}$ e $r_2 = \frac{-b + \sqrt{b^2 + 4w}}{2}$, temos

$$F_{W_n}(w) = P(r_1 \leq Y_n \leq r_2) = [F_{Y_n}(r_2) - F_{Y_n}(r_1)] I_{[-b^2/4, +\infty[}.$$

Assim, a função densidade de probabilidade de W_n é dada por:

$$f_{W_n}(w) = \frac{f_{Y_n}(r_1) + f_{Y_n}(r_2)}{\sqrt{b^2 + 4w}} I_{[-b^2/4, +\infty[}(w).$$

Pelo Teorema 3.5 sabemos que $Y_n \approx \text{Gau}(0, \sigma/\sqrt{n})$ donde resulta finalmente a função densidade de probabilidade de $Z_n = a(Y_n^2 + bY_n) = aW_n$

$$f_{Z_n}(z) = \frac{\phi\left(\sqrt{n}\frac{r_1}{a\sigma}\right) + \phi\left(\sqrt{n}\frac{r_2}{a\sigma}\right)}{\sigma|a|\sqrt{b^2 + 4z}} I_{[b^2/4, +\infty[}(z).$$

onde ϕ é a função densidade de probabilidade de uma variável aleatória gaussiana padrão. Então

$$f_{Z_n}(z) = \sqrt{n} \frac{\exp\left(-\frac{nr_1^2}{2(a\sigma)^2}\right) + \exp\left(-\frac{nr_2^2}{2(a\sigma)^2}\right)}{\sigma |a| \sqrt{2\pi(b^2 + 4z)}} I_{[b^2/4, +\infty[}(z)$$

pelo que o resultado é agora imediato. ■

Investigamos agora em que condições a convergência sugerida pelo teorema anterior é mais rápida. Nas condições do Teorema 3.8

$$f(X_n) - f(\theta) \approx f'(\theta)(X_n - \theta) + \frac{f''(\theta)}{2}(X_n - \theta)^2.$$

A aproximação assintótica apresentada corresponde à soma de duas variáveis aleatórias claramente dependentes. A variância dessa soma é dada por

$$\begin{aligned} & \text{var}\left(f'(\theta)(X_n - \theta) + \frac{f''(\theta)}{2}(X_n - \theta)^2\right) \\ &= \text{var}(f'(\theta)(X_n - \theta)) + \text{var}\left(\frac{f''(\theta)}{2}(X_n - \theta)^2\right) \\ &+ 2\text{cov}\left(f'(\theta)(X_n - \theta), \frac{f''(\theta)}{2}(X_n - \theta)^2\right). \end{aligned}$$

Quando $n \rightarrow +\infty$, a última expressão é assintoticamente igual, em distribuição, a ($a = f'(\theta)$ e $b = f''(\theta)$)

$$a^2 \frac{\sigma^2}{n} + \frac{b^2 \sigma^4}{4n^2} \text{var}(Z^2) + ab \left[\text{cov}(X_n, X_n^2) - 2\theta \left(\frac{\sigma}{\sqrt{n}}\right)^2 \right]$$

onde $Z \sim \text{Gau}(0, \sigma)$, se estamos nas condições do Teorema 3.5, logo $Z/\sigma \sim \text{Gau}(0, 1)$ pelo que $Z^2/\sigma^2 \sim \chi_1^2$ e $\text{var}(Z^2) = \sigma^4 \text{var}(Z^2/\sigma^2) = 2\sigma^4$. Para n suficientemente grande

$$\begin{aligned} \text{cov}(X_n, X_n^2) &\approx E[(X_n - \theta)(X_n^2 - E(X_n^2))] \\ &\approx E(X_n^3) - \left(\frac{\sigma}{\sqrt{n}}\right)^2 \theta - \theta^3, \end{aligned} \tag{25}$$

em distribuição.

Relativamente ao terceiro momento centrado de X_n temos

$$E[(X_n - \theta)^3] \approx E(X_n^3) - 3\theta \left(\frac{\sigma}{\sqrt{n}}\right)^2 - \theta^3 \quad (26)$$

Utilizando a definição de coeficiente de assimetria $\gamma_1(X)$, podemos escrever o limite assintótico do termo $E(X_n^3)$ de (25) como função da assimetria:

$$E(X^3) = \sigma^3 \gamma_1(X) + 3\theta\sigma^2 + \theta^3, \quad (27)$$

onde X é uma variável aleatória com valor médio θ e variância σ^2 .

Então, por (25) e (27),

$$\text{cov}(X_n, X_n^2) \approx \left(\frac{\sigma}{\sqrt{n}}\right)^3 \gamma_1(X_n) + 2\theta \left(\frac{\sigma}{\sqrt{n}}\right)^2.$$

Finalmente, obtemos o limite assintótico da variância em estudo como sendo

$$[f'(\theta)]^2 \frac{\sigma^2}{n} + \frac{[f''(\theta)]^2 \sigma^8}{2n^2} + f'(\theta) f''(\theta) \frac{\sigma^3}{n^{3/2}} \gamma_1(X_n). \quad (28)$$

O terceiro termo de (28) é nulo caso X_n seja uma variável aleatória simétrica. Portanto, a simetria favorece a convergência subjacente ao Teorema 3.8. Tal não é certamente uma surpresa pois a distribuição gaussiana é a única com cumulantes todos nulos a partir da ordem 3 e no método delta (e suas extensões) admitimos *a priori* que o comportamento assintótico de X_n é aproximadamente gaussiano e portanto simétrico.

Estas aproximações consideradas a partir de truncaturas da série de Taylor são muitas vezes utilizadas para determinar aproximações a momentos de transformações de variáveis aleatórias. Habitualmente, os momentos de maior interesse numa distribuição são os quatro primeiros pelo que vamos estender o raciocínio utilizado na demonstração anterior para a truncatura da série de Taylor após a terceira e quarta derivadas seguindo Pestana e Martins (2007). No primeiro caso,

admitindo que as derivadas consideradas não se anulam em θ , estamos a considerar a aproximação

$$\begin{aligned} f(X_n) - f(\theta) &\approx \sum_{i=1}^3 f^{(i)}(\theta) \frac{(X_n - \theta)^i}{i!} \\ &= \frac{f'''(\theta)}{3!} ((X_n - \theta)^3 + b(X_n - \theta)^2 + c(X_n - \theta)) \\ &= a(Y_n^3 + bY_n^2 + cY_n). \end{aligned}$$

onde $Y_n = X_n - \theta$, $a = f'''(\theta)/3!$, $b = 3f''(\theta)/f'''(\theta)$ e $c = 6f'(\theta)/f'''(\theta)$. Tal como anteriormente, iremos considerar a função de distribuição da variável aleatória

$$W_n = Y_n^3 + bY_n^2 + cY_n$$

onde $b, c \neq 0$. O suporte de W_n está contido e até pode ser igual a \mathbb{R} uma vez que $\lim_{x \rightarrow \pm\infty} f(x) = \pm\infty$ onde $f(x) = x^3 + bx^2 + cx$:

$$\begin{aligned} F_{W_n}(w) &= P(Y_n^3 + bY_n^2 + cY_n \leq w) \\ &= P(Y_n^3 + bY_n^2 + cY_n - w \leq 0). \end{aligned}$$

Identifiquemos $Y_n^3 + bY_n^2 + cY_n - w$ com $g(Y_n)$ onde $g(x) = x^3 + bx^2 + cx - w$, cujas raízes reais podem ser, em número, uma, duas ou três. Assim, há que considerar três casos:

N.º raízes	Raízes	F_{W_n}
1	r	$F_{Y_n}(r)$
2	uma raiz r e uma raiz dupla r^*	$F_{Y_n}(\max(r, r^*))$
3	$r_1 < r_2 < r_3$	$F_{Y_n}(r_1) + F_{Y_n}(r_3) - F_{Y_n}(r_2)$

Tabela 15 – F_{W_n} para um polinómio de grau três

F_{Y_n} é assintoticamente igual a uma função de distribuição de uma variável aleatória gaussiana com valor médio nulo e desvio padrão $\frac{\sigma}{\sqrt{n}}$.

No segundo caso, pretendemos analisar a aproximação

$$\begin{aligned}
 f(X_n) - f(\theta) &\approx \sum_{i=1}^4 \frac{f^{(i)}(\theta)}{i!} (X_n - \theta)^i \\
 &= \frac{f^{(4)}(\theta)}{4!} ((X_n - \theta)^4 + b(X_n - \theta)^3) \\
 &\quad + \frac{f^{(4)}(\theta)}{4!} (c(X_n - \theta)^2 + d(X_n - \theta)) \\
 &= a(Y_n^4 + bY_n^3 + cY_n^2 + dY_n),
 \end{aligned}$$

onde $Y_n = X_n - \theta$, $a = f^{(4)}(\theta)/4!$, $b = 4f'''(\theta)/f^{(4)}(\theta)$, $c = 12f''(\theta)/f^{(4)}(\theta)$ e $d = 24f'(\theta)/f^{(4)}(\theta)$.

Portanto, é necessário calcular a função de distribuição da variável aleatória

$$W_n = Y_n^4 + bY_n^3 + cY_n^2 + dY_n$$

onde, por hipótese, $b, c, d \neq 0$:

$$F_{W_n}(w) = P(Y_n^4 + bY_n^3 + cY_n^2 + dY_n - w \leq 0).$$

Identificando $W_n - w$ com $g(W_n)$ onde $g(x) = x^3 + bx^2 + cx - w$, o número de raízes reais de g varia entre zero e quatro. Analisando cada um dos casos a função de distribuição de W_n é dada pela seguinte tabela.

N.º raízes	Raízes	F_{W_n}
0	—	0
1	r	0
2	2 raízes duplas	0
	raízes simples $r_1 < r_2$	$F_{Y_n}(r_2) - F_{Y_n}(r_1)$
3	raízes simples $r_1 < r_2$	$F_{Y_n}(r_2) - F_{Y_n}(r_1)$
4	$r_1 < r_2 < r_3 < r_4$	$F_{Y_n}(r_2) + F_{Y_n}(r_4) - F_{Y_n}(r_1) - F_{Y_n}(r_3)$

Tabela 16 – F_{W_n} para um polinómio de grau quatro

3.6 Transformações estabilizadoras da variância

A utilização de distribuições de variáveis fulcrais a nível da Inferência Estatística está geralmente condicionada ao conhecimento da variância. Quando a variância é desconhecida, que é a situação mais comum, é usual aplicar-se uma transformação estabilizadora da variância à variável de interesse, o que permite obter uma variável fulcral cuja variância já não depende de parâmetros desconhecidos.

Essa não dependência de parâmetros desconhecidos pode ser encarada apenas como uma propriedade assintótica o que conduz a duas definições de transformação estabilizadora da variância.

Se o processo de determinação dessas transformações “assintóticas” é bem conhecido para o caso univariado com um parâmetro desconhecido tal já não é verdade para outras situações. Nesta secção propomos um método de obtenção de transformações estabilizadoras da variância no caso univariado com mais de um parâmetro desconhecido terminando com um exemplo de aplicação relativo à distribuição gama.

3.6.1 Definições

Começaremos por abordar o conceito de transformação estabilizadora da variância da forma heurística como é habitualmente apresentado por autores como Bishop *et al* (1975) entre outros. Em seguida, será feita a apresentação clássica que pode ser encontrada em Bartlett (1947) e Hoaglin, Mosteller e Tukey (1983) que tem por base expansões em série de Taylor.

Da construção heurística desse conceito retira-se imediatamente que, em geral, a existência dessa transformação está garantida embora possa conduzir a resultados com expressões analíticas de grande complexidade e com reduzida aplicação prática. O caso univariado com n parâmetros desconhecidos não é, em geral, abordado na

literatura. Apresentamos um método de obtenção de transformações estabilizadoras da variância para essa situação e um exemplo de como este método pode ser útil na obtenção de intervalos de confiança.

A extensão deste conceito a vectores aleatórios é possível, mas a existência deixa de estar garantida, embora seja possível assegurar a unicidade do resultado, caso exista, a menos de uma transformação linear. Holland (1973) apresenta uma condição necessária e suficiente para a existência dessa transformação no caso de pares aleatórios, que tem algumas limitações que adiante discutiremos seguindo a abordagem de Pestana e Martins (2007).

Nas condições do Teorema 3.5 (método delta) podemos afirmar que

$$\sqrt{n} [f(X_n) - f(\theta)] \xrightarrow[n \rightarrow \infty]{d} Z \sim \text{Gau}(0, \sigma |f'(\theta)|).$$

Ora, o desvio padrão σ depende habitualmente de θ . Assim, podemos escrever $\sigma = \sigma(\theta)$. Se θ for desconhecido a aplicação prática do resultado estabelecido por esse teorema fica limitada.

Porém, é possível encontrar uma transformação da sucessão de variáveis aleatórias $\{X_n\}_{n \in \mathbb{N}}$ de tal forma que o desvio padrão da variável resultante da aplicação dessa transformação seja assintoticamente independente de θ . Uma transformação desse tipo é usualmente referida como **transformação estabilizadora da variância**. Será esta a definição considerada nas secções que se seguem.

Definição 3.5 *Dada uma sucessão de variáveis aleatórias $\{X_n\}_{n \in \mathbb{N}}$ tal que*

$$\sqrt{n} (X_n - \theta) \xrightarrow[n \rightarrow \infty]{d} Z \sim \text{Gau}(0, \sigma),$$

*diz-se que $g : \mathbb{R} \rightarrow \mathbb{R}$ é uma **transformação estabilizadora da variância** de $\{X_n\}_{n \in \mathbb{N}}$ se $a_n (g(X_n) - g(\theta))$ tem distribuição assintótica $\text{Gau}(0, c)$ onde $\{a_n\}_{n \in \mathbb{N}}$ é uma sucessão de números reais e $c > 0$ não depende de θ .*

Holland (1973) sugere a designação mais sugestiva de “transformação assintoticamente estabilizadora da variância” dado que as distribuições envolvidas são assintóticas. A sucessão a_n de interesse é geralmente $a_n = \sqrt{n}$ pois a velocidade de convergência garantida pelo Teorema Limite Central é da ordem $n^{-1/2}$.

Admitindo que $g'(\theta) > 0$ e escrevendo, sem perda de generalidade, atendendo ao Teorema 3.5, $c = \sigma(\theta)g'(\theta)$, pode-se obter uma expressão para g utilizando a igualdade anterior:

$$\begin{aligned} c &= \sigma(\theta)g'(\theta) \\ \Leftrightarrow g'(\theta) &= \frac{c}{\sigma(\theta)} \\ \Leftrightarrow g(\theta_1) &= g(\theta_0) + c \int_{\theta_0}^{\theta_1} \frac{d\theta}{\sigma(\theta)} \end{aligned}$$

para quaisquer θ_0 e θ_1 pertencentes ao domínio de g .

Portanto, a existência de uma transformação estabilizadora da variância fica garantida. Além disso, qualquer transformação linear de uma transformação estabilizadora da variância também o é. Assim, se g é uma transformação estabilizadora da variância então $f(x) = ag(x - b)$ para quaisquer $a \in \mathbb{R} \setminus \{0\}$ e $b \in \mathbb{R}$ também é, como facilmente se verifica.

Estas transformações vão permitir testar hipóteses relativas a θ bem como o cálculo de intervalos de confiança para θ . Para testar $H_0 : \theta = \theta_0$ versus $H_1 : \theta \neq \theta_0$, utilizando o Teorema 3.5, deve rejeitar-se H_0 se $\frac{\sqrt{n}|g(X_n) - g(\theta)|}{c} > z_{\alpha/2}$ onde $P(Z > z_{\alpha/2}) = \frac{\alpha}{2}$ e $Z \sim \text{Gau}(0, 1)$ (os intervalos com região de rejeição unilateral são determinados de forma análoga). Por “inversão” da região de não rejeição da hipótese H_0 obtém-se um intervalo com aproximadamente $(1 - \alpha) \times 100\%$ de confiança para θ :

$$\left] g(X_n) - z_{\alpha/2} \frac{c}{\sqrt{n}}, g(X_n) + z_{\alpha/2} \frac{c}{\sqrt{n}} \right[. \quad (29)$$

Vejamos, em seguida, alguns exemplos.

Exemplo 3.1 Considere-se uma sucessão $\{X_n\}_{n \in \mathbb{N}}$ de variáveis aleatórias independentes com distribuição Poisson e valor médio λ , $X \sim P(\lambda)$. Então, a partir do Teorema Limite Central é imediato que

$$\sqrt{n}(\bar{X}_n - \lambda) \xrightarrow[n \rightarrow \infty]{d} Z_1 \sim \text{Gau}(0, \sqrt{\lambda}).$$

Calculando uma transformação estabilizadora da variância para a distribuição assintótica de \bar{X}_n e tomando $c = \frac{1}{2}$ resulta

$$g(\lambda) = c \int_0^\lambda \frac{dv}{\sqrt{v}} = \sqrt{\lambda}.$$

Logo, pelo Teorema 3.5,

$$\sqrt{n}(\sqrt{\bar{X}_n} - \sqrt{\lambda}) \xrightarrow[n \rightarrow \infty]{d} Z_2 \sim \text{Gau}\left(0, \frac{1}{2}\right).$$

Anscombe (1948) apresenta a transformação

$$f(x) = \sqrt{x + 3/8}$$

como alternativa mais eficiente que a transformação raiz quadrada apresentada no exemplo anterior uma vez que minimiza a variância da variável transformada. No entanto, a distribuição limite mantém-se a mesma. O autor também analisa o caso da distribuição Binomial e distribuição Binomial Negativa. Em relação à distribuição Binomial, Pitman (1993) utiliza a assimetria da distribuição para melhorar a velocidade de convergência.

Vamos analisar um exemplo semelhante ao exemplo anterior na medida em que o valor médio e a variância coincidem mas em que a população parente é gaussiana, isto é,

$$X \sim \text{Gau}(\theta, \sqrt{\theta})$$

para $\theta > 0$ que tem por função densidade de probabilidade

$$f_{\theta}(x) = \frac{1}{\sqrt{2\pi\theta}} \exp \left[-\frac{1}{2} \left(\frac{x - \theta}{\sqrt{\theta}} \right)^2 \right] I_{\mathbb{R}}(x).$$

Exemplo 3.2 Dada uma amostra $\underline{x} = (x_1, \dots, x_n)$ extraída de uma população gaussiana de valor médio e variância igual a θ , a respectiva função de verosimilhança é dada por

$$\mathcal{L}(\theta|\underline{x}) = \prod_{k=1}^n \frac{1}{\sqrt{2\pi\theta}} \exp \left[-\frac{1}{2} \left(\frac{x_k - \theta}{\sqrt{\theta}} \right)^2 \right]$$

donde

$$\ln \mathcal{L}(\theta|\underline{x}) = \sum_{k=1}^n \left[-\ln \sqrt{2\pi\theta} - \frac{1}{2} \left(\frac{x_k - \theta}{\sqrt{\theta}} \right)^2 \right]$$

Derivando em ordem a θ obtemos

$$\begin{aligned} \frac{\partial}{\partial \theta} \ln \mathcal{L}(\theta|\underline{x}) &= \sum_{k=1}^n \left[-\ln \sqrt{2\pi\theta} - \frac{1}{2} \left(\frac{x_k - \theta}{\sqrt{\theta}} \right)^2 \right] \\ &= \sum_{k=1}^n \left(-\frac{1}{2\theta} + \frac{x_k - \theta}{\sqrt{\theta}} \cdot \frac{x_k + \theta}{2\theta\sqrt{\theta}} \right) \\ &= -\frac{n}{2\theta} + \sum_{k=1}^n \frac{x_k^2 - \theta^2}{2\theta^2}. \end{aligned}$$

Igualando a zero a última expressão obtemos o estimador de máxima verosimilhança

$$\begin{aligned} \frac{n}{2\theta} &= \sum_{k=1}^n \frac{x_k^2 - \theta^2}{2\theta^2} \\ \Leftrightarrow n\theta &= \sum_{k=1}^n (x_k^2 - \theta^2) \\ \Leftrightarrow \theta &= \frac{1}{n} \sum_{k=1}^n x_k^2 - \theta^2 \\ \Leftrightarrow \theta^2 + \theta - T &= 0 \end{aligned}$$

onde $T = \frac{1}{n} \sum_{k=1}^n x_k^2$. Logo, o estimador de máxima verosimilhança é

$$\hat{\theta} = \frac{\sqrt{1+T} - 1}{2}$$

pois $\theta > 0$ e a estatística T é suficiente mínima.

A matriz de informação de Fisher é igual a

$$\begin{aligned} \mathcal{I}(\theta) &= \mathbb{E} \left[\left(\frac{\partial}{\partial \theta} \ln \mathcal{L}(\theta|X) \right)^2 \right] \\ &= \mathbb{E} \left[\left(-\frac{1}{2\theta} + \frac{X^2 - \theta^2}{2\theta^2} \right)^2 \right] \\ &= \mathbb{E} \left[\frac{1}{4\theta^2} - \frac{1}{\theta} \cdot \frac{X^2 - \theta^2}{2\theta^2} + \left(\frac{X^2 - \theta^2}{2\theta^2} \right)^2 \right] \\ &= \frac{1}{4\theta^2} - \frac{\theta}{2\theta^3} + \frac{3 + 4\theta}{4\theta^4} \\ &= \frac{2\theta + 1}{2\theta^2}. \end{aligned}$$

Logo,

$$\sqrt{n} (\hat{\theta} - \theta) \xrightarrow[n \rightarrow \infty]{d} Z^* \sim \text{Gau} \left(0, \theta \sqrt{\frac{2}{2\theta + 1}} \right).$$

Calculando a transformação estabilizadora da variância ($c = 1$)

$$g(\theta) = \int_a^\theta \sqrt{\frac{2\theta + 1}{2}} \cdot \frac{d\theta}{\theta} + g(a)$$

com a ajuda do software Mathematica 6.0 verifica-se que $g = h \circ j$ onde

$$\begin{aligned} h(x) &= x - \frac{1}{2} \ln \left(\frac{x+1}{x-1} \right) \\ &= x - \arg \coth x \end{aligned}$$

em que $\arg \coth(\cdot)$ é a inversa da função cotangente hiperbólica e

$$j(x) = \sqrt{2x + 1}$$

Logo,

$$\sqrt{n} \left(g \left(\hat{\theta} \right) - g \left(\theta \right) \right) \xrightarrow[n \rightarrow \infty]{d} Z \sim \text{Gau} (0, 1).$$

Estes dois exemplos evidenciam que a expressão de uma transformação estabilizadora da variância pode ser simples como no primeiro caso ou complexa como no segundo caso.

Os dois próximos dois resultados generalizam os resultados presentes em Chandra (1999). São relativos à distribuição assintótica do desvio padrão amostral S_n e da variância amostral S_n^2 . A autora estabelece uma aproximação assintótica para a distribuição de $\frac{n-1}{n} S_n^2$ onde

$$S_n^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2$$

no caso de população parente gaussiana. Aqui, generalizaremos esse resultado para o caso de uma população parente genérica. O resultado será enunciado para a estatística S_n^2 , assintoticamente, equivalente a $\frac{n-1}{n} S_n^2$.

Teorema 3.9 *Considere-se $\{X_n\}_{n \in \mathbb{N}}$ uma sucessão de variáveis aleatórias independentes e identicamente distribuídas a X onde $\sigma^2 = \text{var}(X) < \infty$. Então*

$$\sqrt{n} \left(\ln S_n^2 - \ln \sigma^2 \right) \xrightarrow[n \rightarrow \infty]{d} Z_1 \sim \text{Gau} \left(0; \sqrt{\gamma_2(X) + 2} \right).$$

Demonstração Segundo Rocha e Martins (2005),

$$E(S_n^2) = \sigma^2$$

e

$$\text{var}(S_n^2) = \left(\frac{\gamma_2(X)}{n} + \frac{2}{n-1} \right) \sigma^4$$

donde, pelo Teorema Limite Central, $\sqrt{n}(S_n^2 - \sigma^2)$ tem distribuição assintótica

$$\text{Gau} \left(0; \sqrt{\gamma_2(X) + 2\sigma^2} \right).$$

Admitindo que $\gamma_2(X)$ é conhecido tomemos $\theta = \sigma^2$. A transformação estabilizadora da variância para $c = \sqrt{\gamma_2(X) + 2}$ é

$$g(\theta) = c \int_1^\theta \frac{1}{\sqrt{\gamma_2(X) + 2v}} dv + g(1) = \ln \theta.$$

Logo, pelo Teorema 3.5,

$$\sqrt{n} (\ln S_n^2 - \ln \sigma^2) \xrightarrow[n \rightarrow \infty]{d} Z_1 \sim \text{Gau} \left(0; \sqrt{\gamma_2(X) + 2} \right),$$

como queríamos demonstrar. ■

O resultado anterior é equivalente ao apresentado por Chandra (1999) para populações gaussianas em que $\gamma_2(X) = 0$ e que conduz a uma variância assintótica igual a dois. A convergência assintótica prescrita pelo teorema anterior para os quantis mais elevados pode ser muito lenta mesmo para o caso da população parente seguir uma distribuição gaussiana.

Considere-se uma amostra aleatória X_1, \dots, X_n identicamente distribuída a $Z \sim \text{Gau}(0, 1)$. Utilizando o software MatLab 7, podemos obter intervalos com 95% de confiança $]L, U[$ para determinados quantis elevados da estatística S_n^2 e para diversos valores de n . Naturalmente, as estimativas obtidas a partir do Teorema 3.9 são dadas por $\sqrt{2}z_{\alpha/2}$ onde $P(Z > z_{\alpha/2}) = \alpha/2$.

Os resultados obtidos para os quantis 0.90, 0.95, 0.975 e 0.99 são apresentados nos gráficos que se seguem.

3.6 Transformações estabilizadoras da variância

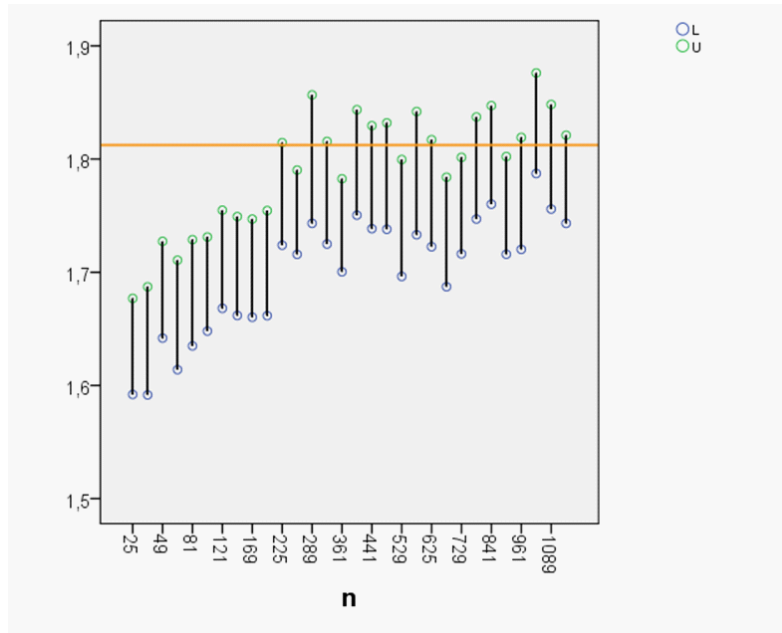


Gráfico 1 – Intervalos de confiança para o quantil 0.90

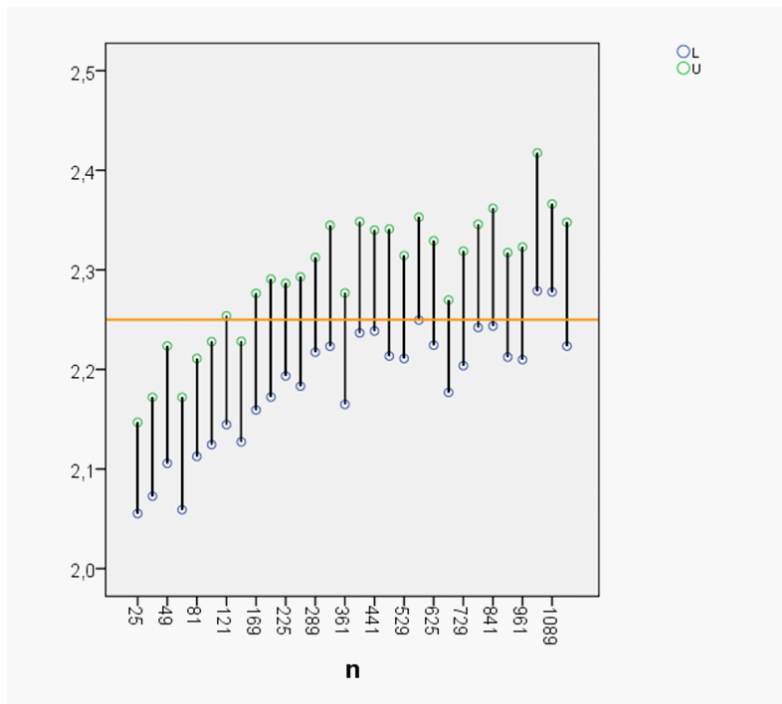


Gráfico 2 – Intervalos de confiança para o quantil 0.95

3.6 Transformações estabilizadoras da variância

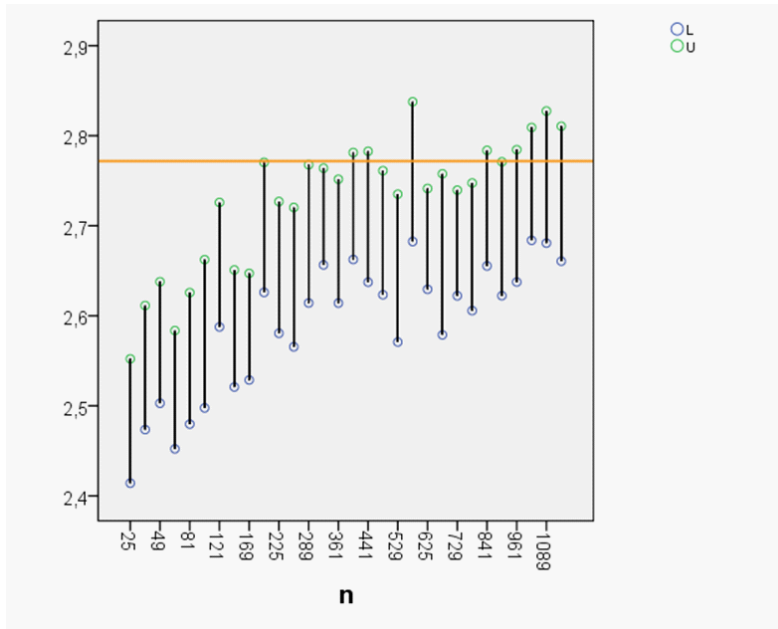


Gráfico 3 – Intervalos de confiança para o quantil 0.975

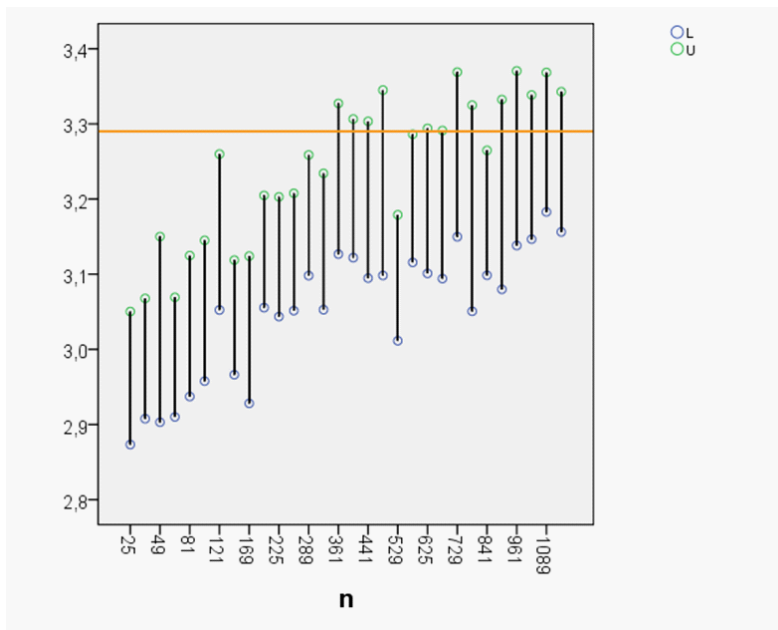


Gráfico 4 – Intervalos de confiança para o quantil 0.99

A recta horizontal apresentada em cada gráfico corresponde à estimativa dada pelo Teorema 3.9. É preciso avançar até amostras de dimensão na ordem das três centenas para que as estimativas pertençam aos intervalos de confiança $]L, U[$.

Rocha e Martins (2005) encontraram aproximações gerais para a esperança e variância de S_n aplicando o método delta:

$$E(S_n) \approx \sigma - \frac{\sigma}{8} \left(\frac{\gamma_2(X)}{n} + \frac{2}{n-1} \right) \quad (30)$$

e

$$\text{var}(S_n) \approx \frac{\sigma^2}{4} \left(\frac{\gamma_2(X)}{n} + \frac{2}{n-1} \right). \quad (31)$$

Estes dois resultados permitem apresentar um corolário relativo a transformações estabilizadoras da variância do desvio padrão amostral.

Corolário 3.1 *Considere-se $\{X_n\}_{n \in \mathbb{N}}$ uma sucessão de variáveis aleatórias independentes e identicamente distribuídas a X onde $\sigma^2 = \text{var}(X) < \infty$. Então*

$$\sqrt{n}(\ln S_n - \ln \sigma) \xrightarrow[n \rightarrow \infty]{d} Z_1 \curvearrowright \text{Gau} \left(0, \frac{\sqrt{\gamma_2(X) + 2}}{2} \right).$$

Demonstração O resultado é uma consequência de (30) e (31) ou então pode ser deduzido a partir do Teorema 3.9. ■

Como referido anteriormente, o conceito de transformação estabilizadora da variância também pode ser aplicado a uma única variável aleatória. Esta visão corresponde à definição clássica do conceito.

Definição 3.6 *Dada uma variável aleatória X cuja variância depende de um parâmetro desconhecido θ diz-se que $g : \mathbb{R} \rightarrow \mathbb{R}$ é uma **transformação estabilizadora da variância** de X se $\text{var}(g(X))$ é constante como função de θ .*

A definição clássica aplica-se apenas a uma variável aleatória ao passo que a definição heurística aplica-se a uma sucessão de variáveis aleatórias. Vejamos um exemplo similar ao Exemplo 3.1.

Exemplo 3.3 Consideremos uma variável aleatória $X \sim P(\lambda)$. Pretende-se encontrar uma transformação g tal que $c^2 = E[g(X^2)] - [E(g(X))]^2$. Ora, se g for infinitamente derivável, então aplicando a definição de valor esperado à expressão anterior,

$$c^2 = \sum_{i=0}^{\infty} (g(i))^2 \frac{\exp(-\lambda) \lambda^i}{i!} - \left(\sum_{i=0}^{\infty} g(i) \frac{\exp(-\lambda) \lambda^i}{i!} \right)^2$$

$$\Leftrightarrow c^2 = \sum_{i=0}^{\infty} (g(i))^2 \frac{\lambda^i \sum_{j=0}^{\infty} \frac{(-\lambda)^j}{j!}}{i!} - \left(\sum_{i=0}^{\infty} g(i) \frac{\lambda^i \sum_{j=0}^{\infty} \frac{(-\lambda)^j}{j!}}{i!} \right)^2$$

Dos somatórios do membro direito da igualdade anterior apenas duas parcelas não dependem de λ pelo que

$$c^2 = [g(0)]^2 - [g(0)]^2 = 0$$

Logo, $\text{var}(g(X)) = 0$ pelo que g é constante e portanto não existe transformação estabilizadora da variância na sua versão clássica.

Ao considerar uma variável aleatória em vez de uma sucessão de variáveis aleatórias nas condições definidas anteriormente a existência de uma transformação estabilizadora da variância deixa de estar garantida. Daqui em diante, invocaremos apenas a definição heurística de transformação estabilizadora da variância (Definição 3.5).

3.6.2 Caso univariado com n parâmetros desconhecidos

No caso de distribuições assintóticas cuja variância dependa de n parâmetros desconhecidos, o processo anterior permite apenas estabelecer uma variância assintótica dependente de n parâmetros. Porém, pode-se considerar um parâmetro

de cada vez e em cada passo determinar uma transformação estabilizadora da variância relativamente ao parâmetro fixado. A transformação estabilizadora da variância do vector de parâmetros θ corresponderá à composição de n transformações

$$g(\theta) = g_n \circ \dots \circ g_1(\theta).$$

Claro que encontrar essa função composta depende da complexidade das primitivas envolvidas. Além disso, é necessário ter em atenção o domínio dessa função composta. Exemplificamos este método exibindo intervalos de confiança para o valor médio de uma distribuição gama. O exemplo pode servir de guia para a resolução de problemas de maior dificuldade algébrica. Aqui, apresenta-se um caso onde esta metodologia sugerida pode ser aplicada.

Exemplo 3.4 *Consideremos uma variável aleatória $X \sim \text{Gama}(\alpha, \beta)$ com função densidade*

$$f_X(x) = \frac{x^{\alpha-1} \exp\left(-\frac{x}{\beta}\right)}{\beta^\alpha \Gamma(\alpha)}, \quad x > 0, \alpha > 0, \beta > 0.$$

Verifica-se facilmente que $E(X) = \alpha\beta$ e $\text{var}(X) = \alpha\beta^2$.

Considerando uma sucessão de variáveis aleatórias $\{X_n\}_{n \in \mathbb{N}}$ independentes e identicamente distribuídas a X conclui-se, pelo Teorema Limite Central, que

$$\sqrt{n}(\bar{X}_n - \alpha\beta) \xrightarrow[n \rightarrow \infty]{d} Z_1 \sim \text{Gau}(0, \sqrt{\alpha}\beta).$$

Começando por considerar o parâmetro β temos que ($c_1 = \sqrt{\alpha}$)

$$g_1(\beta) = \sqrt{\alpha} \int_1^\beta \frac{1}{\sqrt{\alpha}v} dv + g_1(1) = \ln \beta.$$

Donde

$$\sqrt{n}(g_1(\bar{X}_n) - g_1(\alpha\beta)) \xrightarrow[n \rightarrow \infty]{d} Z_2 \sim \text{Gau}(0, \sqrt{\alpha}).$$

Em relação a $\sqrt{\alpha}$ tem de proceder-se ao seguinte cálculo ($c_2 = 1/2$)

$$g_2(\alpha) = \frac{1}{2} \int_1^\alpha \frac{1}{\sqrt{v}} dv + g_2(1) = \sqrt{\alpha}$$

Logo,

$$\sqrt{n} (g_2 \circ g_1(\bar{X}_n) - g_2 \circ g_1(\alpha\beta)) \xrightarrow[n \rightarrow \infty]{d} Z_3 \frown \text{Gau} \left(0, \frac{1}{2}\right).$$

Portanto,

$$\sqrt{n} \left(\sqrt{\ln(\bar{X}_n)} - \sqrt{\ln(\alpha\beta)} \right) \xrightarrow[n \rightarrow \infty]{d} Z_3 \frown \text{Gau} \left(0, \frac{1}{2}\right). \quad (32)$$

A partir deste resultado é possível estabelecer intervalos de confiança e realizar testes de hipóteses relativamente ao valor médio $\alpha\beta$ da distribuição gama tendo em conta (29).

Obviamente, a aproximação é válida para $\alpha\beta > 1$ devido ao domínio da função composta em causa.

A composição de funções não é, em geral, comutativa pelo que se se começar por considerar o parâmetro α da variância assintótica obtém-se

$$\sqrt{n} \left(\ln \sqrt{\bar{X}_n} - \ln \sqrt{\alpha\beta} \right) \xrightarrow[n \rightarrow \infty]{d} Z_3 \frown \text{Gau}(0, 2).$$

Curiosamente, este resultado aplica-se para quaisquer valores de α e β ao contrário do resultado (32).

3.6.3 Caso multivariado

Anteriormente, foram consideradas transformações estabilizadoras da variância para variáveis aleatórias unidimensionais. Porém, o conceito pode ser estendido a um vector aleatório $\mathbf{X}_p = (X_1, \dots, X_p)$ e a um parâmetro desconhecido $\theta = (\theta_1, \dots, \theta_p)$ que varia em $A \subseteq \mathbb{R}^p$. A definição, neste caso, tem por motivação o Teorema 3.7.

Definição 3.7 Dada uma sucessão de vectores aleatórios $\{\mathbf{X}_{p_n}\}_{n \in \mathbb{N}}$ diz-se que $g : A \rightarrow \mathbb{R}^p$ é uma **transformação estabilizadora da covariância** de \mathbf{X}_{p_n} se $a_n (g(\mathbf{X}_{p_n}) - g(\theta))$ tem distribuição assintótica $\text{Gau}(0, W)$ onde $\{a_n\}_{n \in \mathbb{N}}$ é uma sucessão de números reais e W é a matriz das covariâncias (que não depende de θ).

A matriz W é simétrica e definida positiva e portanto invertível. Anteriormente, através de cálculos simples, obtivemos um método expedito de cálculo da transformação estabilizadora da variância para uma dada sucessão de variáveis aleatórias. No caso multivariado o processo complica-se uma vez que é necessário operar com matrizes.

Assim, dada uma sucessão de vectores aleatórios \mathbf{X}_{p_n} tal que

$$\sqrt{n} (\mathbf{X}_{p_n} - \theta) \xrightarrow[n \rightarrow \infty]{d} \mathbf{Z} \sim \text{Gau}(\mathbf{0}, W)$$

tem-se que $g : A \rightarrow \mathbb{R}^p$ é uma transformação estabilizadora da covariância caso a matriz jacobiana da transformação g , $\left(\frac{\partial g}{\partial \theta}\right)$, goze da propriedade

$$\left(\frac{\partial g}{\partial \theta}\right)^T \left(\frac{\partial g}{\partial \theta}\right) = W^{-1}.$$

A transformação g em causa é única a menos de uma matriz real multiplicativa e de outra aditiva como estabelece o seguinte teorema.

Teorema 3.10 Se f e g são transformações estabilizadoras da covariância então $g(\theta) = Mf(\theta) + c$ onde $g(\theta)$ e $f(\theta)$ são, respectivamente, as representações matriciais das transformações g e f relativamente a uma mesma base, c é um vector e M é uma matriz ortogonal.

Demonstração Como referido anteriormente $\left(\frac{\partial g}{\partial \theta}\right)^T \left(\frac{\partial g}{\partial \theta}\right) = W^{-1}$. Logo se f e g são funções estabilizadoras da covariância tem-se

$$\left(\frac{\partial g}{\partial \theta}\right)^T \left(\frac{\partial g}{\partial \theta}\right) = \left(\frac{\partial f}{\partial \theta}\right)^T \left(\frac{\partial f}{\partial \theta}\right) = W^{-1}$$

pelo que

$$\begin{aligned} \left(\frac{\partial g}{\partial \theta}\right) &= \left[\left(\frac{\partial g}{\partial \theta}\right)^T\right]^{-1} \left(\frac{\partial f}{\partial \theta}\right)^T \left(\frac{\partial f}{\partial \theta}\right) \\ \Leftrightarrow \left(\frac{\partial g}{\partial \theta}\right) &= M(\theta) \left(\frac{\partial f}{\partial \theta}\right) \end{aligned}$$

onde $M(\theta) = \left[\left(\frac{\partial g}{\partial \theta}\right)^T\right]^{-1} \left(\frac{\partial f}{\partial \theta}\right)^T = \left(\frac{\partial g}{\partial \theta}\right) \left(\frac{\partial f}{\partial \theta}\right)^{-1}$ é uma matriz ortogonal pelo Teorema 1.3. Logo, aplicando o teorema da derivada da função inversa e o teorema da derivada da função composta e tendo em conta que f e g são injectivas

$$M(\theta) = \left(\frac{\partial g}{\partial \theta}\right) \left(\frac{\partial f}{\partial \theta}\right)^{-1} = \left(\frac{\partial g}{\partial \theta}\right) \left(\frac{\partial \theta}{\partial f}\right) = \left(\frac{\partial g}{\partial f}\right).$$

$M(\theta)$ não depende de θ pelo que pode ser identificada apenas por M . Assim, integrando a expressão $M(\theta) = \left(\frac{\partial g}{\partial f}\right)$ obtém-se

$$g(\theta) = Mf(\theta) + c$$

como se pretendia. ■

Ao contrário do caso univariado, a existência de transformações estabilizadoras da variância nem sempre está garantida no caso multivariado. Sabe-se que, caso exista uma transformação g nessas condições,

$$\left(\frac{\partial g}{\partial \theta}\right)^T \left(\frac{\partial g}{\partial \theta}\right) = W^{-1}.$$

Como W^{-1} é simétrica e definida positiva apresenta mais de uma raiz quadrada. Seja R uma raiz quadrada de W^{-1} tal que $R^T R = \left(\frac{\partial g}{\partial \theta}\right)^T \left(\frac{\partial g}{\partial \theta}\right)$. Pelo Teorema 1.3

$$\left(\frac{\partial g}{\partial \theta}\right) = \Gamma(\theta) R \tag{33}$$

onde $\Gamma(\theta)$ é uma matriz ortogonal. Portanto, a existência de uma transformação estabilizadora da covariância é equivalente à existência de uma matriz ortogonal $\Gamma(\theta)$ tal que $A(\theta) = \Gamma(\theta)R$ seja uma matriz jacobiana de uma transformação injectiva. Ora, uma condição necessária e suficiente para a existência de uma matriz jacobiana diferenciável foi estabelecida pelo Teorema 1.4.

Holland (1973) apresenta uma condição relativamente complicada para a existência de transformações estabilizadoras da covariância no caso bivariado que se baseia na igualdade anterior. Usando o Teorema 1.1 e assumindo, sem perda de generalidade, que a distribuição assintótica de $g(\mathbf{X}_n)$ tem por matriz de covariâncias a matriz identidade, a matriz ortogonal $\Gamma(\theta)$ pode ser representada por

$$\Gamma(\theta) = \begin{pmatrix} \cos \phi(t) & \sin \phi(t) \\ -\sin \phi(t) & \cos \phi(t) \end{pmatrix} \quad (34)$$

no caso de determinante unitário. O tratamento do caso $\det(\Gamma(\theta)) = -1$ é semelhante.

Aqui, apresenta-se esse resultado obtido por Holland (1973) seguido da respectiva demonstração realizada com maior detalhe.

Teorema 3.11 *Uma matriz ortogonal $\Gamma(\theta)$ existe tal que $\left(\frac{\partial g}{\partial \theta}\right) = \Gamma(\theta)R$ para uma função $g : D^2 \rightarrow \mathbb{R}^2$ se, e só se, os elementos de R satisfazem a equação diferencial*

$$\begin{aligned} & \frac{\partial}{\partial \theta_2} \left[\frac{1}{\det(R)} \left(r_{11} \left(\frac{\partial r_{11}}{\partial \theta_2} - \frac{\partial r_{12}}{\partial \theta_1} \right) + r_{21} \left(\frac{\partial r_{21}}{\partial \theta_2} - \frac{\partial r_{22}}{\partial \theta_1} \right) \right) \right] \\ & = \frac{\partial}{\partial \theta_1} \left[\frac{1}{\det(R)} \left(r_{12} \left(\frac{\partial r_{11}}{\partial \theta_2} - \frac{\partial r_{12}}{\partial \theta_1} \right) + r_{22} \left(\frac{\partial r_{21}}{\partial \theta_2} - \frac{\partial r_{22}}{\partial \theta_1} \right) \right) \right]. \end{aligned}$$

No caso de $\Gamma(\theta)$ existir então $\phi(t)$ pode ser determinada a partir da resolução do sistema

$$\begin{pmatrix} \frac{\partial \phi}{\partial \theta_1} \\ \frac{\partial \phi}{\partial \theta_2} \end{pmatrix} = \frac{R^T}{\det(R)} \begin{pmatrix} \frac{\partial r_{11}}{\partial \theta_2} - \frac{\partial r_{12}}{\partial \theta_1} \\ \frac{\partial r_{21}}{\partial \theta_2} - \frac{\partial r_{22}}{\partial \theta_1} \end{pmatrix}.$$

Demonstração Como referido anteriormente, se a transformação estabilizadora da covariância existir, a matriz $B(\theta) = \Gamma(\theta)R$ é a matriz jacobiana de g para determinada matriz $\Gamma(\theta)$.

A condição necessária e suficiente para verificar se uma dada matriz é uma matriz jacobiana diferenciável de uma função é

$$\begin{cases} \frac{\partial b_{11}}{\partial \theta_2} = \frac{\partial b_{12}}{\partial \theta_1} \\ \frac{\partial b_{21}}{\partial \theta_2} = \frac{\partial b_{22}}{\partial \theta_1} \end{cases}.$$

Ora,

$$\begin{aligned} B(\theta) = \Gamma(\theta)R &= \begin{pmatrix} \cos \phi(t) & \sin \phi(t) \\ -\sin \phi(t) & \cos \phi(t) \end{pmatrix} \begin{pmatrix} r_{11} & r_{12} \\ r_{21} & r_{22} \end{pmatrix} \\ &= \begin{pmatrix} r_{11} \cos \phi(t) + r_{21} \sin \phi(t) & r_{12} \cos \phi(t) + r_{22} \sin \phi(t) \\ -r_{11} \sin \phi(t) + r_{21} \cos \phi(t) & -r_{12} \sin \phi(t) + r_{22} \cos \phi(t) \end{pmatrix}. \end{aligned}$$

Portanto,

$$\begin{cases} \frac{\partial b_{11}}{\partial \theta_2} = \frac{\partial b_{12}}{\partial \theta_1} \\ \frac{\partial b_{21}}{\partial \theta_2} = \frac{\partial b_{22}}{\partial \theta_1} \end{cases} \Leftrightarrow \begin{cases} \frac{\partial (r_{11} \cos \phi(t) + r_{21} \sin \phi(t))}{\partial \theta_2} = \frac{\partial (r_{12} \cos \phi(t) + r_{22} \sin \phi(t))}{\partial \theta_1} \\ \frac{\partial (-r_{11} \sin \phi(t) + r_{21} \cos \phi(t))}{\partial \theta_2} = \frac{\partial (-r_{12} \sin \phi(t) + r_{22} \cos \phi(t))}{\partial \theta_1} \end{cases}.$$

Calculando as derivadas parciais

$$\left\{ \begin{array}{l} \frac{\partial r_{11}}{\partial \theta_2} \cos \phi(t) - r_{11} \sin \phi(t) \frac{\partial \phi}{\partial \theta_2} + \frac{\partial r_{21}}{\partial \theta_2} \sin \phi(t) + r_{21} \cos \phi(t) \frac{\partial \phi}{\partial \theta_2} \\ = \frac{\partial r_{12}}{\partial \theta_1} \cos \phi(t) - r_{12} \sin \phi(t) \frac{\partial \phi}{\partial \theta_1} + \frac{\partial r_{22}}{\partial \theta_1} \sin \phi(t) + r_{22} \cos \phi(t) \frac{\partial \phi}{\partial \theta_1} \\ - \frac{\partial r_{11}}{\partial \theta_2} \sin \phi(t) - r_{11} \cos \phi(t) \frac{\partial \phi}{\partial \theta_2} + \frac{\partial r_{21}}{\partial \theta_2} \cos \phi(t) - r_{21} \sin \phi(t) \frac{\partial \phi}{\partial \theta_2} \\ = - \frac{\partial r_{12}}{\partial \theta_1} \sin \phi(t) - r_{12} \cos \phi(t) \frac{\partial \phi}{\partial \theta_1} + \frac{\partial r_{22}}{\partial \theta_1} \cos \phi(t) - r_{22} \sin \phi(t) \frac{\partial \phi}{\partial \theta_1} \end{array} \right.$$

Rearranjando os termos das duas igualdades

$$\left\{ \begin{array}{l} \frac{\partial r_{11}}{\partial \theta_2} \cos \phi(t) - \frac{\partial r_{12}}{\partial \theta_1} \cos \phi(t) + \frac{\partial r_{21}}{\partial \theta_2} \sin \phi(t) - \frac{\partial r_{22}}{\partial \theta_1} \sin \phi(t) \\ = r_{11} \sin \phi(t) \frac{\partial \phi}{\partial \theta_2} - r_{12} \sin \phi(t) \frac{\partial \phi}{\partial \theta_1} - r_{21} \cos \phi(t) \frac{\partial \phi}{\partial \theta_2} + r_{22} \cos \phi(t) \frac{\partial \phi}{\partial \theta_1} \\ - \frac{\partial r_{11}}{\partial \theta_2} \sin \phi(t) + \frac{\partial r_{12}}{\partial \theta_1} \sin \phi(t) + \frac{\partial r_{21}}{\partial \theta_2} \cos \phi(t) - \frac{\partial r_{22}}{\partial \theta_1} \cos \phi(t) \\ = r_{11} \cos \phi(t) \frac{\partial \phi}{\partial \theta_2} - r_{12} \cos \phi(t) \frac{\partial \phi}{\partial \theta_1} + r_{21} \sin \phi(t) \frac{\partial \phi}{\partial \theta_2} - r_{22} \sin \phi(t) \frac{\partial \phi}{\partial \theta_1} \end{array} \right.$$

Utilizando a notação matricial

$$\begin{aligned} & \begin{pmatrix} \cos \phi(t) & \sin \phi(t) \\ -\sin \phi(t) & \cos \phi(t) \end{pmatrix} \begin{pmatrix} \frac{\partial r_{11}}{\partial \theta_2} - \frac{\partial r_{12}}{\partial \theta_1} \\ \frac{\partial r_{21}}{\partial \theta_2}(t) - \frac{\partial r_{22}}{\partial \theta_1} \end{pmatrix} \\ &= \begin{pmatrix} \cos \phi(t) & \sin \phi(t) \\ -\sin \phi(t) & \cos \phi(t) \end{pmatrix} \begin{pmatrix} r_{22} & -r_{21} \\ -r_{12} & r_{11} \end{pmatrix} \begin{pmatrix} \frac{\partial \phi}{\partial \theta_1} \\ \frac{\partial \phi}{\partial \theta_2} \end{pmatrix}. \end{aligned}$$

Portanto,

$$\Gamma(\theta) \begin{pmatrix} \frac{\partial r_{11}}{\partial \theta_2} - \frac{\partial r_{12}}{\partial \theta_1} \\ \frac{\partial r_{21}}{\partial \theta_2} - \frac{\partial r_{22}}{\partial \theta_1} \end{pmatrix} = \Gamma(\theta) \det(R) (R^T)^{-1} \begin{pmatrix} \frac{\partial \phi}{\partial \theta_1} \\ \frac{\partial \phi}{\partial \theta_2} \end{pmatrix}$$

que é equivalente a

$$\begin{pmatrix} \frac{\partial \phi}{\partial \theta_1} \\ \frac{\partial \phi}{\partial \theta_2} \end{pmatrix} = \frac{R^T}{\det(R)} \begin{pmatrix} \frac{\partial r_{11}}{\partial \theta_2} - \frac{\partial r_{12}}{\partial \theta_1} \\ \frac{\partial r_{21}}{\partial \theta_2} - \frac{\partial r_{22}}{\partial \theta_1} \end{pmatrix}.$$

Assim, a existência da matriz $\Gamma(\theta)$ depende da existência de ϕ . Aplicando novamente o Teorema 1.4 obtém-se

$$\frac{\partial}{\partial \theta_2} \left(\frac{\partial \phi}{\partial \theta_1} \right) = \frac{\partial}{\partial \theta_1} \left(\frac{\partial \phi}{\partial \theta_2} \right) \quad (35)$$

Ou seja,

$$\begin{aligned} & \frac{\partial}{\partial \theta_2} \left[\frac{1}{\det(R)} \left(r_{11} \left(\frac{\partial r_{11}}{\partial \theta_2} - \frac{\partial r_{12}}{\partial \theta_1} \right) + r_{21} \left(\frac{\partial r_{21}}{\partial \theta_2} - \frac{\partial r_{22}}{\partial \theta_1} \right) \right) \right] \\ &= \frac{\partial}{\partial \theta_1} \left[\frac{1}{\det(R)} \left(r_{12} \left(\frac{\partial r_{11}}{\partial \theta_2} - \frac{\partial r_{12}}{\partial \theta_1} \right) + r_{22} \left(\frac{\partial r_{21}}{\partial \theta_2} - \frac{\partial r_{22}}{\partial \theta_1} \right) \right) \right] \end{aligned}$$

como se queria demonstrar. ■

A verificação da existência de transformações estabilizadoras da covariância pode-se revelar bastante morosa. Em algumas situações especiais os cálculos podem ser obviados. Se R é uma matriz diagonal então as duas igualdades referidas no enunciado do teorema anterior podem ser simplificadas.

Corolário 3.2 *Se R é uma matriz diagonal, uma matriz ortogonal $\Gamma(\theta)$ existe tal que $\left(\frac{\partial g}{\partial \theta}\right) = \Gamma(\theta) R$ para uma função $g : D^2 \rightarrow \mathbb{R}^2$ se, e só se,*

$$\frac{\partial}{\partial \theta_1} \left(\frac{1}{r_{11}} \frac{\partial r_{22}}{\partial \theta_1} \right) + \frac{\partial}{\partial \theta_2} \left(\frac{1}{r_{22}} \frac{\partial r_{11}}{\partial \theta_2} \right) = 0.$$

No caso de $\Gamma(\theta)$ existir então $\phi(t)$ pode ser determinada a partir da resolução do sistema

$$\begin{pmatrix} \frac{\partial \phi}{\partial \theta_1} \\ \frac{\partial \phi}{\partial \theta_2} \end{pmatrix} = \begin{pmatrix} \frac{1}{r_{22}} \frac{\partial r_{11}}{\partial \theta_2} \\ -\frac{1}{r_{11}} \frac{\partial r_{22}}{\partial \theta_1} \end{pmatrix}.$$

Demonstração O resultado é imediato substituindo, na equação diferencial do Teorema 3.11, r_{12} e r_{21} por zero. ■

Uma conclusão imediata do Corolário 3.2 é que no caso de duas variáveis aleatórias independentes X e Y , em que a variância de X é constante como função do parâmetro θ_2 e a variância de Y é constante como função do parâmetro θ_1 , existe transformação estabilizadora da covariância.

Outra situação particularmente simples acontece quando existe um único parâmetro desconhecido e a covariância entre as variáveis envolvidas coincide com a variância de uma delas. É o caso de $X \sim P(\lambda)$ e $Y|_{X=x} \sim P(x)$ ou de $X \sim \text{Rayleigh}(\alpha)$ e $Y|_{X=x} \sim \text{Exp}(x)$. Essas situações enquadram-se no seguinte corolário.

Corolário 3.3 *Se R é uma matriz tal que $r_{11} = r_{12}$ e $\theta_1 = \theta_2$ então uma matriz ortogonal $\Gamma(\theta)$ existe tal que $\left(\frac{\partial g}{\partial \theta}\right) = \Gamma(\theta) R$ para uma função $g : D^2 \rightarrow \mathbb{R}^2$ se, e só se,*

$$r_{11} = c \frac{\partial r_{22}}{\partial \theta_1}.$$

Demonstração Pode-se assumir, sem perda de generalidade, que a matriz R é triangular superior pelo que $r_{21} = 0$. Por hipótese $r_{11} = r_{12}$ e $\theta_1 = \theta_2$ donde,

substituindo na equação diferencial do Teorema 3.11 obtém-se,

$$\begin{aligned} & \frac{\partial}{\partial \theta_1} \left[\frac{1}{r_{11}r_{22}} \left(r_{11} \left(\frac{\partial r_{11}}{\partial \theta_1} - \frac{\partial r_{11}}{\partial \theta_1} \right) + r_{21} \left(0 - \frac{\partial r_{22}}{\partial \theta_1} \right) \right) \right] \\ &= \frac{\partial}{\partial \theta_1} \left[\frac{1}{r_{11}r_{22}} \left(r_{12} \left(\frac{\partial r_{11}}{\partial \theta_1} - \frac{\partial r_{11}}{\partial \theta_1} \right) + r_{22} \left(0 - \frac{\partial r_{22}}{\partial \theta_1} \right) \right) \right] \end{aligned}$$

Simplificando,

$$0 = \frac{\partial}{\partial \theta_1} \left(-\frac{1}{r_{11}} \frac{\partial r_{22}}{\partial \theta_1} \right).$$

Logo existe transformação estabilizadora da covariância se, e só se, $-\frac{1}{r_{11}} \frac{\partial r_{22}}{\partial \theta_1}$ for constante $k = k(\theta_1)$ como função de θ_1 , isto é,

$$-\frac{1}{r_{11}} \frac{\partial r_{22}}{\partial \theta_1} = k$$

Tomando $c = -\frac{1}{k}$ o resultado é imediato. ■

Exemplo 3.5 *A independência entre a média amostral e variância amostral é uma característica da gaussiana bem como constitui uma forma de definir essa distribuição dado que essa independência só se verifica essa situação. Por outro lado, nesse caso, (Rocha e Martins, 2005),*

$$\text{var}(\bar{X}_n) = \frac{\sigma^2}{n} \text{ e } \text{var}(S_n^2) = \frac{2\sigma^4}{n-1}$$

donde, por uma extensão do Teorema Limite Central,

$$\sqrt{n} \left[(\bar{X}, S^2)^T - (\mu, \sigma^2)^T \right] \xrightarrow[n \rightarrow \infty]{d} \mathbf{Z} \sim \text{Gau}(\mathbf{0}, W)$$

onde $W = \begin{pmatrix} \sigma^2 & 0 \\ 0 & 2\sigma^4 \end{pmatrix}$.

Uma raiz quadrada de W^{-1} é $R = \begin{pmatrix} \frac{1}{\sigma} & 0 \\ 0 & \frac{1}{\sqrt{2}\sigma^2} \end{pmatrix}$. Tomemos $\theta_1 = \mu$ e $\theta_2 = \sigma^2$.

Pelo Corolário 3.2, uma transformação estabilizadora da covariância existe se, e só

se,

$$\frac{\partial}{\partial \theta_1} \left(\sqrt{\theta_2} \times 0 \right) + \frac{\partial}{\partial \theta_2} \left(\sqrt{2\theta_2} \times \left(-\frac{1}{2\theta_2^{3/2}} \right) \right) = 0 \quad (36)$$

o que não se verifica pois o resultado da soma anterior é

$$\frac{\partial}{\partial \theta_1} \left(\frac{1}{\sqrt{\theta_2}} \times 0 \right) - \frac{\partial}{\partial \theta_2} \left(\sqrt{2\theta_2} \times \frac{1}{2\theta_2^{3/2}} \right) = 0 - \frac{\partial}{\partial \theta_2} \left(\frac{1}{\sqrt{2\theta_2}} \right) \neq 0$$

Logo, não existe uma transformação estabilizadora da covariância para o par aleatório (\bar{X}_n, S_n^2) no caso de distribuição parente gaussiana.

Pares Aleatórios	Matriz Covariâncias	
(\bar{X}, S^2) em pop. gaussianas	$\begin{pmatrix} \theta_2 & 0 \\ 0 & 2\theta_2^2 \end{pmatrix}, \begin{cases} \theta_1 = \mu \\ \theta_2 = \sigma^2 \end{cases}$	N
(\bar{X}, S^2) caso geral	$\begin{pmatrix} \theta_2 & \gamma_1 \theta_2^{1.5} \\ \gamma_1 \theta_2^{1.5} & (\gamma_2 + 2) \theta_2^2 \end{pmatrix}, \begin{cases} \theta_1 = \mu \\ \theta_2 = \sigma^2 \end{cases}$	N
$X \sim P(\lambda)$ e $Y _{X=x} \sim \text{Gama}(x, \delta)$	$\begin{pmatrix} \theta_1 & \theta_2 \\ \theta_2 & 2\theta_2\theta_1^{-1} \end{pmatrix}, \begin{cases} \theta_1 = \lambda \\ \theta_2 = \delta\lambda \end{cases}$	S
$X \sim P(\lambda)$ e $Y _{X=x} \sim P(x)$	$\begin{pmatrix} \theta_1 & \theta_1 \\ \theta_1 & 2\theta_1 \end{pmatrix}, \begin{cases} \theta_1 = \lambda \\ \theta_2 = \lambda \end{cases}$	N
$X \sim \text{Ge}(p)$ e $Y _{X=x} \sim B(x, p)$	$\begin{pmatrix} \frac{1 - \theta_1^{-1}}{\theta_1^{-2}} & \frac{1 - \theta_1^{-1}}{\theta_1^{-1}} \\ \frac{1 - \theta_1^{-1}}{\theta_1^{-1}} & 2 - 2\theta_1^{-1} \end{pmatrix}, \begin{cases} \theta_1 = \frac{1}{p} \\ \theta_2 = 1 \end{cases}$	N
$X \sim P(\lambda)$ e $Y _{X=x} \sim \text{Gama}(\alpha, x)$	$\begin{pmatrix} \theta_1 & \theta_2 \\ \theta_2 & \theta_1^{-1}\theta_2^2 + \theta_2 + \theta_1\theta_2 \end{pmatrix}, \begin{cases} \theta_1 = \lambda \\ \theta_2 = \alpha\lambda \end{cases}$	N
$X \sim P(\lambda)$ e $Y _{X=x} \sim B(x, p)$	$\begin{pmatrix} \theta_1 & \theta_2 \\ \theta_2 & \theta_2 \end{pmatrix}, \begin{cases} \theta_1 = \lambda \\ \theta_2 = p\lambda \end{cases}$	S
$X \sim \text{Exp}(\lambda)$ e $Y _{X=x} \sim \text{Gau}(a, \sqrt{x})$	$\begin{pmatrix} \theta_1^2 & 0 \\ 0 & \theta_1 \end{pmatrix}, \begin{cases} \theta_1 = \lambda \\ \theta_2 = a \end{cases}$	N
$X \sim \text{Exp}(\lambda)$ e $Y _{X=x} \sim \text{Gau}(x, \sqrt{a})$	$\begin{pmatrix} \theta_1^2 & 0 \\ 0 & \theta_1^2 + \theta_2 \end{pmatrix}, \begin{cases} \theta_1 = \lambda \\ \theta_2 = a \end{cases}$	N
$X \sim P(\lambda)$ e $Y _{X=x} \sim U(0, \lambda)$	$\begin{pmatrix} \theta_1 & \theta_2 \\ \theta_2 & \frac{\theta_2^2 + \theta_1}{3} \end{pmatrix}, \begin{cases} \theta_1 = \lambda \\ \theta_2 = \frac{\lambda}{2} \end{cases}$	N
$X \sim P(\lambda)$ e $Y _{X=x} \sim t_{(n)} + x$	$\begin{pmatrix} \theta_1 & \theta_1 \\ \theta_1 & 3 \end{pmatrix}, \begin{cases} \theta_1 = \lambda \\ \theta_2 = \sigma^2, n \geq 3 \end{cases}$	N

Tabela 17 – Matriz de covariâncias de alguns pares aleatórios

A tabela anterior indica para alguns pares aleatórios as respectivas matrizes de covariância e se existe ou não (S ou N) transformação estabilizadora da covariância.

Um dos problemas de aplicação do Teorema 3.11 é escrever a matriz R como função dos parâmetros θ_1 e θ_2 o que poderá nem sequer ser possível. Mas, nos casos em que é possível aplicar o teorema verifica-se que a existência de uma transformação estabilizadora da covariância parece ser a excepção e não a regra. O facto de serem raros os casos em que a existência se verifica dificulta a percepção de qual a relação que deve existir entre as entradas da matriz de covariâncias W para que exista uma transformação estabilizadora da covariância. Note-se que essa transformação não existe para alguns casos em que:

- a matriz de covariâncias depende de um único parâmetro;
- a matriz de covariâncias depende dos dois parâmetros;
- a covariância entre as variáveis é nula;
- a covariância depende de um único parâmetro
- as variâncias dependem de um único parâmetro;
- as variâncias dependem dos dois parâmetros.

Os exemplos

$$X \sim P(\lambda) \text{ e } Y|_{X=x} \sim \text{Gama}(x, \delta)$$

e

$$X \sim P(\lambda) \text{ e } Y|_{X=x} \sim \text{Gama}(\alpha, x)$$

são semelhantes na medida em que a matriz de covariâncias depende dos dois parâmetros, a covariância depende de um único parâmetro, a variância da variável aleatória X depende de um parâmetro e a variância da variável aleatória Y depende

dos dois parâmetros nos dois casos. Porém, para o primeiro caso existe transformação estabilizadora da covariância e no segundo caso não. Será que as variâncias não podem envolver expressões “aditivas” dos parâmetros?

Alargar o método proposto por Holland para vectores aleatórios de dimensão superior a dois é uma tarefa complexa uma vez que o método assenta na representação (34) de uma matriz ortogonal de dimensão dois. Uma representação desse tipo para uma ordem superior depara-se com cálculos de grande complexidade.

É de esperar que a existência de transformações estabilizadoras da covariância para vectores aleatórios de dimensão superior a dois seja também excepcional. Song (1982), usando uma demonstração por redução ao absurdo, mostra que para o caso relativamente simples da distribuição multinomial não existe tal transformação, seja qual for o número de parâmetros considerado.

3.7 Uma aplicação interdisciplinar

Um trabalho recente de Pfeiffer e Heremans (2005) apresenta uma justificação, que parece conter alguns lapsos, partindo de uma restrição porventura excessiva de correlação unitária, para uma fórmula empírica usada em misturas de líquidos,

$$\sqrt{\frac{\theta}{M}}\beta = \varphi_A \sqrt{\frac{\theta_A}{M_A}}\beta_A + \varphi_B \sqrt{\frac{\theta_B}{M_B}}\beta_B \quad (37)$$

onde $\varphi_A = \frac{V_A}{V}$ e $\varphi_B = \frac{V_B}{V}$ são as fracções de volume da mistura e θ é a densidade da mistura. Argumentos sofisticados de Termodinâmica permitem então “traduzir” grandezas macroscópicas em termos de grandezas microscópicas e vice-versa.

Sem nos pronunciarmos sobre a validade química desta justificação de uma prestigiosa fórmula usada, até ao presente, com bases meramente empíricas, abordamos a questão geral de aproximar o desvio padrão de uma soma (e, obvia-

mente, se $\rho_{X,Y} = -1$ então $\sigma_{X+Y} = |\sigma_X - \sigma_Y|$. De facto, se $\rho_{X,Y} = 0$ então $\text{var}(X + Y) = \text{var}(X) + \text{var}(Y)$ e se $\rho_{X,Y} = 1$ então $\sigma_{X+Y} = \sigma_X + \sigma_Y$. Fora estes casos, as regras para obtenção de uma expressão manejável para σ_{X+Y} podem ser obtidas recorrendo à expressão formal $\sigma_Y = k\sigma_X$, $k \geq 1$ (sem perda de generalidade), e transformando assim o problema numa aproximação da função $f(k) = \sqrt{k^2 + 2\rho k + 1}$ usando a fórmula de Taylor.

Apresentamos expressões formais seguindo Pestana e Martins (2007), que generalizam as de Pfeiffer e Heremans (2005), sem qualquer pretensão sobre a validade termodinâmica dos resultados.

3.7.1 Alguns conceitos químicos

Apesar de não pretendermos fazer uma transposição para a área da Química dos resultados obtidos, optámos por listar algumas fórmulas químicas que permitem compreender a motivação subjacente ao estudo do desvio padrão da soma de n variáveis aleatórias e compreender os conceitos envolvidos em (37).

A densidade θ de um composto puro é igual ao quociente entre a massa m desse composto e o respectivo volume V

$$\theta = \frac{m}{V}. \quad (38)$$

A compressão isotérmica pode ser calculada através do quociente

$$\beta_T = \frac{\text{var}(\delta V)}{RTV} \quad (39)$$

em que T é a temperatura do sistema, R é uma constante, V é o volume e $\text{var}(\delta V)$ é a variância das flutuações de volume do sistema. A variância das flutuações de volume do sistema pode ser decomposta na soma das flutuações de volume a nível das partículas do sistema. Essa soma aplicada a (37) e, conseqüentemente, a (39)

reflecte a importância, a nível químico, do estudo de aproximações para o desvio padrão da soma de variáveis aleatórias.

3.7.2 Desvio padrão da soma de 2 variáveis aleatórias

Nesta secção, procedemos a uma análise genérica sobre a variância e desvio padrão da soma de duas variáveis aleatórias.

A variância da soma de duas variáveis aleatórias X e Y é, caso existam $\text{var}(X)$ e $\text{var}(Y)$, igual a

$$\text{var}(X + Y) = \text{var}(X) + \text{var}(Y) + 2\text{cov}(X, Y). \quad (40)$$

Considerando $\rho = \rho_{X,Y} = \frac{\text{cov}(X, Y)}{\sigma_X \sigma_Y}$ a expressão anterior pode ser reescrita como

$$\text{var}(X + Y) = \text{var}(X) + \text{var}(Y) + 2\rho\sigma_X\sigma_Y. \quad (41)$$

No caso em que $\rho = 0$, obtém-se $\text{var}(X + Y) = \text{var}(X) + \text{var}(Y)$.

Se $|\rho| = 1$ resulta que

$$\begin{aligned} \text{var}(X + Y) &= \text{var}(X) + \text{var}(Y) + 2\rho\sigma_X\sigma_Y \\ &= \sigma_X^2 + \sigma_Y^2 + 2\rho\sigma_X\sigma_Y \\ &= (\sigma_X + \rho\sigma_Y)^2. \end{aligned} \quad (42)$$

Logo, no caso $\rho = 1$, $\sigma_{X+Y} = \sigma_X + \sigma_Y$ e no caso $\rho = -1$, $\sigma_{X+Y} = |\sigma_X - \sigma_Y|$.

Em geral, assumindo, sem perda de generalidade, que $\sigma_Y \approx k\sigma_X$, $k \geq 1$ tem-se que,

$$\begin{aligned} \text{var}(X + Y) &= \text{var}(X) + \text{var}(Y) + 2\rho\sigma_X\sigma_Y \\ &= \sigma_X^2 + \sigma_Y^2 + 2\rho\sigma_X\sigma_Y \\ &\approx \sigma_X^2 + k^2\sigma_X^2 + 2\rho k\sigma_X^2 \\ &= \sigma_X^2 (k^2 + 2\rho k + 1). \end{aligned} \quad (43)$$

Logo,

$$\sigma_{X+Y} = \sigma_X \sqrt{k^2 + 2\rho k + 1}. \quad (44)$$

A expressão (44) para o desvio padrão pode ser identificada com a função:

$$f(k) = \sqrt{k^2 + 2\rho k + 1}. \quad (45)$$

Esta pode ser aproximada pelo polinómio de Taylor de primeira ordem. Para tal é necessário encontrar a primeira derivada de f

$$f'(k) = \frac{2k + 2\rho}{2\sqrt{k^2 + 2\rho k + 1}} = \frac{k + \rho}{\sqrt{k^2 + 2\rho k + 1}}. \quad (46)$$

O valor de f' no ponto $k = 1$ é

$$f'(1) = \frac{1 + \rho}{\sqrt{1^2 + 2\rho + 1}} = \frac{1 + \rho}{\sqrt{2 + 2\rho}} = \sqrt{\frac{1 + \rho}{2}}. \quad (47)$$

Portanto usando a aproximação dada pelo polinómio de Taylor de ordem 1 (em torno de $k = 1$) para (45) obtém-se

$$f(k) = \sqrt{2 + 2\rho} + \frac{\sqrt{1 + \rho}}{\sqrt{2}} (k - 1) + O(k - 1)^2. \quad (48)$$

Pelo que a expressão (44) pode ser aproximada por

$$\begin{aligned} \sigma_{X+Y} &\approx \sigma_X \left(\sqrt{2 + 2\rho} + \frac{\sqrt{1 + \rho}}{\sqrt{2}} (k - 1) \right) \\ &= \sqrt{2 + 2\rho} \sigma_X + \frac{\sqrt{1 + \rho}}{\sqrt{2}} \sigma_Y - \frac{\sqrt{1 + \rho}}{\sqrt{2}} \sigma_X \\ &= \left(\sqrt{2 + 2\rho} - \frac{\sqrt{1 + \rho}}{\sqrt{2}} \right) \sigma_X + \frac{\sqrt{1 + \rho}}{\sqrt{2}} \sigma_Y \\ &= \frac{\sqrt{1 + \rho}}{\sqrt{2}} \sigma_X + \frac{\sqrt{1 + \rho}}{\sqrt{2}} \sigma_Y. \end{aligned} \quad (49)$$

O contributo de cada desvio padrão é o mesmo para o desvio padrão da soma.

No caso da aproximação ser feita à custa do polinómio de Taylor de segunda ordem é necessário calcular a expressão da segunda derivada de f :

$$\begin{aligned}
 f''(k) &= \frac{\sqrt{k^2 + 2\rho k + 1} - \frac{2k + 2\rho}{2\sqrt{k^2 + 2\rho k + 1}}(k + \rho)}{k^2 + 2\rho k + 1} \\
 &= \frac{\frac{k^2 + 2\rho k + 1 - (k + \rho)^2}{\sqrt{k^2 + 2\rho k + 1}}}{k^2 + 2\rho k + 1} \\
 &= \frac{1 - \rho^2}{(k^2 + 2\rho k + 1)^{3/2}}
 \end{aligned} \tag{50}$$

No ponto $k = 1$, a segunda derivada toma o seguinte valor

$$f''(1) = \frac{1 - \rho^2}{(1^2 + 2\rho + 1)^{1.5}} = \frac{(1 - \rho)(1 + \rho)}{2\sqrt{2}(\rho + 1)^{1.5}} = \frac{1 - \rho}{2\sqrt{2\rho + 2}}. \tag{51}$$

Logo,

$$f(k) = \sqrt{2 + 2\rho} + \frac{\sqrt{1 + \rho}}{\sqrt{2}}(k - 1) + \frac{1 - \rho}{2!2\sqrt{2\rho + 2}}(k - 1)^2 + O(k - 1)^3. \tag{52}$$

Substituindo em (44),

$$\begin{aligned}
 \sigma_{X+Y} &\approx \sigma_X \left(\sqrt{2 + 2\rho} + \frac{\sqrt{1 + \rho}}{\sqrt{2}}(k - 1) + \frac{1 - \rho}{2!2\sqrt{2\rho + 2}}(k - 1)^2 \right) \\
 &= \frac{\sqrt{1 + \rho}}{\sqrt{2}}\sigma_X + \frac{\sqrt{1 + \rho}}{\sqrt{2}}\sigma_Y + \frac{1 - \rho}{4\sqrt{2 + 2\rho}}(k\sigma_Y - 2\sigma_Y + \sigma_X) \\
 &= \left(\frac{\sqrt{1 + \rho}}{\sqrt{2}} + \frac{1 - \rho}{4\sqrt{2 + 2\rho}} \right) \sigma_X \\
 &+ \left(\frac{\sqrt{1 + \rho}}{\sqrt{2}} + \frac{k(1 - \rho)}{4\sqrt{2 + 2\rho}} - \frac{1 - \rho}{2\sqrt{2 + 2\rho}} \right) \sigma_Y \\
 &= \left(\frac{4(1 + \rho)}{4\sqrt{2 + 2\rho}} + \frac{1 - \rho}{4\sqrt{2 + 2\rho}} \right) \sigma_X \\
 &+ \left[\frac{\sqrt{1 + \rho}}{\sqrt{2}} + \frac{k - k\rho - 2 + 2\rho}{4\sqrt{2 + 2\rho}} \right] \sigma_Y \\
 &= \frac{5 + 3\rho}{4\sqrt{2 + 2\rho}}\sigma_X + \frac{4(1 + \rho) - 2 + 2\rho}{4\sqrt{2 + 2\rho}}\sigma_Y + \frac{k(1 - \rho)}{4\sqrt{2 + 2\rho}}\sigma_Y \\
 &= \frac{5 + 3\rho}{4\sqrt{2 + 2\rho}}\sigma_X + \frac{1 + 3\rho}{2\sqrt{2 + 2\rho}}\sigma_Y + \frac{k(1 - \rho)}{4\sqrt{2 + 2\rho}}\sigma_Y.
 \end{aligned} \tag{53}$$

O contributo de cada um dos desvios padrões é semelhante a menos da última parcela que é aproximadamente nula para $\rho \approx 1$.

3.7.3 Clusters de partículas

Em Termodinâmica, e em outras áreas da Ciência, a previsão de propriedades macroscópicas é feita a partir do conhecimento das propriedades microscópicas dos materiais em causa. Para conhecer a covariância entre os valores de uma propriedade macroscópica quantitativa P de uma mistura de partículas de dois tipos, A e B , pode-se utilizar o valor da correlação a nível microscópico entre os valores dessa propriedade das diferentes partículas. Admitamos então que estamos perante um sistema com n_A e n_B partículas do tipo A e B , respectivamente, e seja $n = n_A + n_B$. A covariância total do sistema, cov_T , é igual à soma das covariâncias entre as n partículas

$$\text{cov}_T = \sum_{a,b=1}^n \text{cov}(P_a, P_b). \quad (54)$$

A expressão anterior envolve n^2 parcelas que podem ser agrupadas em 3 somatórios relativos à soma das covariâncias entre partículas do tipo A , do tipo B e de ambas. Assumindo que o valor de P para partículas do mesmo tipo é identicamente distribuído, a expressão (54) pode ser escrita como

$$\text{cov}_T = n_A^2 \text{var}(P_A) + n_B^2 \text{var}(P_B) + 2n_A n_B \text{cov}P_A, P_B.$$

Inicialmente começaremos por analisar o caso em que a correlação entre o valor de P das diferentes partículas é unitária para depois estudar o que se passa se a correlação for apenas próxima de 1.

Correlação unitária Admitindo que P_A e P_B estão completamente correlacionados pode-se afirmar, sem perda de generalidade, que $P_A = kP_B$ em que $k \geq 1$ donde,

$$\begin{aligned} \text{cov}_T &= k^2 n_A^2 \text{var}(P_B) + n_B^2 \text{var}(P_B) + 2kn_A n_B \text{cov}(P_B, P_B) \\ &= k^2 n_A \text{var}(P_B) + n_B^2 \text{var}(P_B) + 2kn_A n_B \text{var}(P_B). \end{aligned} \quad (55)$$

Pelo que,

$$\sqrt{\text{cov}_T} = \sqrt{\text{var}(P_B)} \sqrt{k^2 n_A^2 + n_B^2 + 2kn_A n_B} = \sqrt{\text{var}(P_B)} f(k), \quad (56)$$

onde $f(k) = \sqrt{k^2 n_A^2 + n_B^2 + 2kn_A n_B}$.

Calculando a primeira derivada de f

$$f'(k) = \frac{1}{2} \frac{2n_A^2 k + 2n_A n_B}{\sqrt{k^2 n_A^2 + n_B^2 + 2kn_A n_B}} = \frac{n_A^2 k + n_A n_B}{\sqrt{k^2 n_A^2 + n_B^2 + 2kn_A n_B}}. \quad (57)$$

Pelo que, aproximando a expressão analítica de f pelo polinómio de Taylor de primeira ordem calculado em $k = 1$,

$$\begin{aligned} f(k) &\approx \sqrt{n_A^2 + n_B^2 + 2n_A n_B} + \frac{n_A^2 + n_A n_B}{\sqrt{n_A + n_B^2 + 2n_A n_B}} (k - 1) \\ &= \frac{(n_A^2 + n_A n_B) k + n_B^2 + n_A n_B}{\sqrt{n_A + n_B^2 + 2n_A n_B}}. \end{aligned} \quad (58)$$

Substituindo em (56),

$$\begin{aligned} \sqrt{\text{cov}_T} &\approx \sqrt{\text{var}(P_B)} \frac{(n_A^2 + n_A n_B) k + n_B^2 + n_A n_B}{\sqrt{n_A + n_B^2 + 2n_A n_B}} \\ &= \frac{n_A^2 + n_A n_B}{\sqrt{n_A + n_B^2 + 2n_A n_B}} \sqrt{\text{var}(P_A)} \\ &\quad + \frac{n_B^2 + n_A n_B}{\sqrt{n_A + n_B^2 + 2n_A n_B}} \sqrt{\text{var}(P_B)}. \end{aligned} \quad (59)$$

Note-se que, para o caso de uma mistura pura ($n_A = 0$ ou $n_B = 0$) a expressão (59) faz sentido pois obtém-se

$$\sqrt{\text{cov}_T} = n_B \sqrt{\text{var}(P_B)} \quad \text{ou} \quad \sqrt{\text{cov}_T} = n_A \sqrt{\text{var}(P_A)},$$

respectivamente.

Correlação quase unitária Para o caso em que a correlação não é 1 mas é próxima de 1 podemos afirmar que $P_A \approx kP_j$ em que $k \geq 1$ donde,

$$\begin{aligned} \text{cov}_T &\approx k^2 n_A^2 \text{var}(P_B) + n_B^2 \text{var}(P_B) + 2k\rho n_A n_B \text{var}(P_B) \\ &= \text{var}(P_j) (k^2 n_A^2 + n_B^2 + 2k\rho n_A n_B). \end{aligned} \quad (60)$$

Pelo que,

$$\sqrt{\text{cov}_T} \approx \sqrt{\text{var}(P_j)} \sqrt{k^2 n_A^2 + n_B^2 + 2k\rho n_A n_B} = \sqrt{\text{var}(P_j)} f(k). \quad (61)$$

onde $f(k) = \sqrt{k^2 n_A^2 + n_B^2 + 2k\rho n_A n_B}$.

Então, a primeira derivada de f é

$$f'(k) = \frac{1}{2} \frac{2n_A^2 k + 2\rho n_A n_B}{\sqrt{k^2 n_A^2 + n_B^2 + 2k\rho n_A n_B}} = \frac{n_A^2 k + \rho n_A n_B}{\sqrt{k^2 n_A^2 + n_B^2 + 2k\rho n_A n_B}}. \quad (62)$$

Pelo que aproximando f pelo polinómio de Taylor de ordem 1 em $k = 1$,

$$\begin{aligned} f(k) &\approx \sqrt{n_A^2 + n_B^2 + 2\rho n_A n_B} + \frac{n_A^2 + \rho n_A n_B}{\sqrt{n_A^2 + n_B^2 + 2\rho n_A n_B}} (k - 1) \\ &= \frac{n_A^2 + n_B^2 + 2\rho n_A n_B + (n_A^2 + \rho n_A n_B)(k - 1)}{\sqrt{n_A^2 + n_B^2 + 2\rho n_A n_B}} \\ &= \frac{n_B^2 + \rho n_A n_B + k(n_A^2 + \rho n_A n_B)}{\sqrt{n_A^2 + n_B^2 + 2\rho n_A n_B}}. \end{aligned} \quad (63)$$

Substituindo em (61),

$$\begin{aligned}
 \sqrt{\text{cov}_T} &\approx \sqrt{\text{var}(P_j)} \frac{n_B^2 + \rho n_A n_B + k(kn_A^2 + \rho n_A n_B)}{\sqrt{n_A^2 + n_B^2 + 2\rho n_A n_B}} \\
 &= \sqrt{\text{var}(P_A)} \frac{n_A^2 + \rho n_A n_B}{\sqrt{n_A^2 + n_B^2 + 2\rho n_A n_B}} \\
 &\quad + \sqrt{\text{var}(P_j)} \frac{n_B^2 + \rho n_A n_B}{\sqrt{n_A^2 + n_B^2 + 2\rho n_A n_B}}.
 \end{aligned}$$

Note-se que, quando a correlação é unitária, obtém-se novamente a expressão (59). Verifica-se, mais uma vez, que o peso de cada uma das variâncias é idêntico, o que suporta a ideia de que a aproximação obtida tem interesse.

4 Inferência utilizando a estatística T_n

4.1 Leis estáveis para somas

Consideremos uma sucessão de variáveis aleatórias X_1, \dots, X_n independentes e identicamente distribuídas a X . Um dos principais problemas na identificação do limite de somas de variáveis aleatórias identicamente distribuídas foi conhecer o limite de somas convenientemente normalizadas

$$S_n = \frac{X_1 + \dots + X_n - A_n}{B_n} \quad (64)$$

para determinadas constantes $A_n \in \mathbb{R}$ e $B_n > 0$ ($n \in \mathbb{N}$). A constante B_n tende a conferir um valor médio (ou, caso este não exista, uma mediana) a S_n ao passo que a constante A_n limita a dispersão. A relação (64) serve de base à seguinte definição.

Definição 4.1 *Considere-se uma sucessão de variáveis aleatórias X_1, \dots, X_n independentes e identicamente distribuídas a X . Caso existam constantes normalizadoras A_n e $B_n > 0$ ($n \in \mathbb{N}$) tais que*

$$\frac{S_n - A_n}{B_n} \xrightarrow[n \rightarrow \infty]{d} Y \quad (65)$$

onde $S_n = X_1 + \dots + X_n$ e Y é uma variável aleatória não degenerada, diz-se que X pertence ao **domínio de atracção da variável aleatória Y** e representa-se por $X \in \mathcal{D}(Y)$.

Dada uma variável aleatória Y que seja limite de (65), a soma $S_n^* = \sum_{i=1}^n Y_i$ de variáveis aleatórias independentes e identicamente distribuídas a Y verifica

$$\frac{S_n^* - A_n^*}{B_n^*} \rightarrow Y$$

para determinadas constantes normalizadoras A_n^* e $B_n^* > 0$. Daí a designação de **leis**

estáveis (para somas) para as leis limite de (65). A caracterização dos domínios de atracção das leis estáveis para somas deve-se ao trabalho pioneiro de Gnedenko (1940).

Das leis estáveis, a mais conhecida é a gaussiana pois o Teorema Limite Central, na sua versão mais simples, garante que a soma de variáveis independentes e identicamente distribuídas com segundo momento finito está no domínio de atracção da gaussiana.

Feller (1971) mostra que as constantes B_n de (65) são do tipo $n^{1/\alpha} L(n)$ onde L é uma função de variação lenta no sentido de Karamata, isto é,

$$\frac{L(tx)}{L(x)} \xrightarrow{x \rightarrow \infty} K \in [0, \infty)$$

onde $t > 0$. A constante α toma valores no intervalo $(0, 2]$ e é designada por **expoente característico da lei estável** Y e caracteriza essa mesma lei permitindo a identificação da lei estável por Y_α .

A função característica de uma variável aleatória estável é dada por

$$\varphi_X(t) = \exp \left\{ iat - c |t|^\alpha \left[1 + i\beta \frac{t}{|t|} w(t, \alpha) \right] \right\}$$

onde a é a localização, $c > 0$ é a escala, α é o expoente característico, $\beta \in [-1, 1]$ é o coeficiente de assimetria e

$$w(t, \alpha) = \begin{cases} \tan \frac{\pi\alpha}{2}, & \alpha \in (0, 1) \cup (1, 2] \\ \frac{2}{\pi} \ln |t|, & \alpha = 1 \end{cases}.$$

Apenas se consegue inverter a função característica para $\alpha = 2$ (Gaussiana), $\alpha = 1$ e $\beta = 0$ (Cauchy) e $\alpha = 0.5$ e $\beta = 1$ (Lévy).

Para verificar se uma variável aleatória pertence ao domínio de atracção de uma lei estável Y_α tem-se habitualmente em conta o “peso” das caudas. Nesse sentido, são comumente conhecidos os seguintes teoremas.

Teorema 4.1 $X \in D(Y_2)$ se, e só se, o seu segundo momento truncado

$$V(x) = \int_{-x}^x u^2 dF(u)$$

for uma função de variação lenta.

Teorema 4.2 $X \in D(Y_\alpha)$, $\alpha \in (0, 2)$ se, e só se,

1. $\frac{1 - F_X(x)}{F_X(-x)} \xrightarrow{x \rightarrow \infty} K \in [0, \infty]$
2. $\frac{F_X(-tx) + 1 - F_X(tx)}{F_X(-x) + 1 - F_X(x)} \xrightarrow{x \rightarrow \infty} t^{-\alpha}$

A condição 2 do Teorema 4.2 é equivalente à soma das caudas $P(|X| > x)$ ser uma função de variação regular com expoente $-\alpha$.

4.2 Divisibilidade infinita

Começamos esta secção por analisar algumas transformadas integrais: transformada de Laplace e função característica. Estas duas transformadas são muito úteis no estudo da soma de variáveis aleatórias independentes uma vez que a transformada da soma corresponde ao produto dessas transformadas integrais. Apresentamos o conceito de função completamente monótona que está associado à identificação de distribuições infinitamente divisíveis. Será esta associação que permitirá mais adiante mostrar que pelo menos parte da família gaussiana generalizada é infinitamente divisível.

Definição 4.2 Se F é uma função de distribuição com suporte não negativo então a **transformada de Laplace** de F é dada por ($\lambda \geq 0$)

$$L(\lambda) = E(\exp(-\lambda X)) = \int_0^\infty \exp(-\lambda x) dF(x).$$

Uma transformada de Laplace determina univocamente uma função de distribuição. Porém, aplica-se apenas a distribuições com suporte contido em \mathbb{R}_0^+ .

Definição 4.3 *A função característica de uma variável aleatória X ou, mais propriamente, da função de distribuição F_X é*

$$\varphi_X(t) = E(\exp(itX)) = \int_{\mathbb{R}} \exp(itx) dF(x)$$

Entre as propriedades mais relevantes das funções características, destacam-se as seguintes:

- está definida para qualquer $t \in \mathbb{R}$;
- é uniformemente contínua em \mathbb{R} ;
- se $Y = aX + b$ então $\varphi_Y(t) = \exp(itb) \varphi_X(at)$
- se X e Y são variáveis aleatórias independentes então $\varphi_{X+Y} = \varphi_X \varphi_Y$.

A última dessas propriedades torna a função característica numa ferramenta útil para lidar com somas de variáveis aleatórias independentes, como já tinha sido referido. A transformada de Laplace goza de propriedades semelhantes à exceção da primeira pois

$$\varphi_X(t) = L_X(-it).$$

O cálculo de uma função característica pode-se revelar bastante difícil pois envolve um integral definido no plano complexo. Associadas a estas transformadas integrais estão as funções completamente monótonas.

Definição 4.4 *Uma função $f \in C^\infty$ com domínio contido em \mathbb{R}_0^+ é **completamente monótona** quando*

$$(-1)^n f^{(n)}(x) \geq 0$$

para $x \in \mathbb{R}_0^+$.

O enfoque dado a este tipo de funções fica a dever-se a um teorema de Bernstein como refere Feller (1966). Aqui apresentamos um enunciado equivalente apresentado por este último autor.

Teorema 4.3 *Uma função $f \in C^\infty$ com domínio contido em \mathbb{R}_0^+ é completamente monótona se, e só se, é da forma*

$$f(\lambda) = \int_0^\infty \exp(-\lambda x) dF(x)$$

onde F é uma função não decrescente.

Uma função completamente monótona é, a menos de uma constante multiplicativa, a transformada de Laplace de uma variável aleatória de suporte contido em \mathbb{R}_0^+ . Vejamos duas propriedades destas funções (não é apresentada demonstração).

Teorema 4.4 *Se f e g são completamente monótonas então o produto fg é também uma função completamente monótona.*

Teorema 4.5 *Se f é completamente monótona, g é uma função não negativa e g' é completamente monótona então $f \circ g$ é completamente monótona.*

A gênese do estudo das “somadas infinitésimas” deve-se a Paul Lévy (Pestana e Velosa, 2008). O autor interessou-se pela questão de conhecer que tipo de distribuições poderiam ser o limite de somas de parcelas infinitamente pequenas e independentes. Introduziu, desse modo, o conceito de divisibilidade infinita que se segue.

Definição 4.5 *Uma função de distribuição F_X é **infinitamente divisível** quando, para qualquer $n \in \mathbb{N}$,*

$$S_n = X_1 + \dots + X_n \stackrel{d}{=} X$$

onde X_1, \dots, X_n são independentes e identicamente distribuídas a X .

Em termos de funções características a definição anterior é equivalente à seguinte definição.

Definição 4.6 *Uma função característica φ é **infinitamente divisível** quando, para qualquer $n \in \mathbb{N}$, existe uma função característica φ_n tal que*

$$\varphi_n^n = \varphi.$$

Se φ for infinitamente divisível então φ não pode ter zeros reais (Pestana e Velosa, 2008, p. 832). Nesse caso φ admite a representação

$$\varphi = \exp(\psi).$$

A função ψ é muitas vezes referidas como 2.^a função característica.

Khintchine (1937) mostrou que uma função de distribuição é infinitamente divisível se, e só se, o logaritmo da sua função característica é da forma

$$\psi(t) = \ln \varphi(t) = i\gamma t + \int_{\mathbb{R}} A(u, t) \frac{1 + u^2}{u^2} dG(u),$$

onde

$$A(u, t) = \exp(iut) - 1 - \frac{itu}{1 + u^2}, \tag{66}$$

$\gamma \in \mathbb{R}$, G é uma função não decrescente de variação limitada tal que $G(-\infty) = 0$ e o integral quando $u = 0$ é igual a $-t^2/2$.

Lévy (1937) apresentou a seguinte fórmula equivalente:

$$\psi(t) = i\gamma t - \frac{\sigma^2 t^2}{2} + \int_{-\infty}^0 A(u, t) dM(u) + \int_0^{+\infty} A(u, t) dN(u)$$

onde A é dado por (66), M e N são funções não decrescentes em $(-\infty, 0)$ e $(0, +\infty)$, respectivamente, verificam $M(-\infty) = N(+\infty) = 0$ e para qualquer $\varepsilon > 0$,

$$\int_{-\varepsilon}^0 u^2 dM(u) + \int_0^{\varepsilon} u^2 dN(u) < \infty.$$

Ambas as fórmulas são uma generalização de um resultado obtido por Kolmogorov para variáveis aleatórias infinitamente divisíveis de variância finita:

$$\psi(t) = i\gamma t - \frac{t^2}{2}K(0) + \int_{\mathbb{R} \setminus \{0\}} \frac{\exp(itu) - 1 - itu}{u^2} K(u) du \quad (67)$$

onde $\gamma \in \mathbb{R}$ e K é uma função não decrescente de variação limitada com $K(-\infty) = 0$.

Exemplo 4.1 Consideremos o caso de uma variável aleatória gaussiana X de valor médio μ e desvio padrão σ . A função característica de X é dada por

$$\varphi(t) = \exp\left(\mu it - \frac{(t\sigma)^2}{2}\right).$$

Ora,

$$\ln \varphi(t) = \mu it - \frac{(t\sigma)^2}{2}$$

que é da forma (67) para $\gamma = \mu$, $K(u) = \sigma^2 I_{\{0\}}(x)$.

O grande resultado sobre divisibilidade infinita foi apresentado por Finetti (1930).

Teorema 4.6 X é uma variável aleatória infinitamente divisível se, e só se, X for o limite fraco de uma sucessão de variáveis aleatórias Poisson compostas.

Uma variável aleatória Poisson composta é uma soma de variáveis aleatórias independentes e identicamente distribuídas em que o número de parcelas é uma variável aleatória com distribuição Poisson independente das parcelas, isto é,

$$X = \sum_{k=1}^N Y_k,$$

onde as variáveis aleatórias Y_k são identicamente distribuídas e independentes entre si e de $N \sim P(\lambda)$. Verifica-se facilmente que

$$\varphi_X(t) = \exp(\lambda(\varphi(t) - 1))$$

onde φ é a função característica de Y_k .

4.3 Leis limites da estatística $S_n(p)$

Consideremos a estatística

$$S_n(p) = \frac{\sum_{i=1}^n X_i}{\left(\sum_{i=1}^n |X_i|^p\right)^{1/p}}$$

onde as variáveis aleatórias X_1, \dots, X_n são independentes e identicamente distribuídas e $0 < p < \infty$. A definição da estatística $S_n(p)$ pode ser estendida a $p = \infty$:

$$S_n(\infty) = \frac{\sum_{i=1}^n X_i}{|X_{\max}|}$$

onde $X_{\max} = \max(X_1, \dots, X_n)$.

Darling (1952) obteve a função característica de $S_n(\infty)$ para variáveis aleatórias pertencentes a um domínio de atracção estável. Utilizando uma metodologia distinta, Logan, Mallows, Rice e Shepp (1973) obtiveram a função de distribuição de $S_n(p)$ para os restantes valores de $p < \infty$ assumindo também que as variáveis aleatórias X_i pertencem a um domínio de atracção estável. Nesse mesmo artigo conjecturam que o limite de $S_n(p)$ só poderá ser não degenerado com suporte em dois pontos se essa condição se verificar. Mason e Zinn (2005) demonstram essa conjectura para o caso de variáveis aleatórias simétricas.

O caso $p = 2$ merece especial atenção pois a distribuição de $S_n(2)$ relaciona-se com a distribuição de

$$T_n = \sqrt{n} \frac{\bar{X}_n - \mu}{S_n}$$

da seguinte forma:

$$\begin{aligned}
 S_n(2) \left(\frac{n-1}{n - S_n^2(2)} \right)^{1/2} &= \frac{\sum_{i=1}^n X_i}{\left(\sum_{i=1}^n |X_i|^2 \right)^{1/2}} \times \frac{\sqrt{n-1}}{\left(n - \frac{\left(\sum_{i=1}^n X_i \right)^2}{\sum_{i=1}^n X_i^2} \right)^{1/2}} \\
 &= \frac{\sqrt{n-1} \sum_{i=1}^n X_i}{\sqrt{n \sum_{i=1}^n X_i^2 - \left(\sum_{i=1}^n X_i \right)^2}} \tag{68} \\
 &= \sqrt{\frac{n-1}{n}} \frac{\sum_{i=1}^n X_i}{\sqrt{\sum_{i=1}^n X_i^2 - n \bar{X}_n^2}} \\
 &= \sqrt{n} \frac{\bar{X}_n}{S_n} = T_n.
 \end{aligned}$$

Tendo em conta a relação (68), Efron (1969) conclui que a distribuição limite de $S_n(2)$ e de T_n , caso existam, são coincidentes.

Aplicando o Teorema Limite Central e a Lei dos Grandes Números facilmente se verifica que a distribuição limite $S_n(2)$ é a distribuição gaussiana padrão quando $E(X_i) = 0$ e $\text{var}(X_i) < \infty$. Mais, essas hipóteses podem ser relaxadas mantendo-se a gaussiana como distribuição limite. Pestana e Velosa (2008) fornecem uma panorâmica sobre o assunto.

Giné, Götze e Mason (1997) demonstram uma das conjecturas do artigo de Logan, Mallows, Rice e Shepp (1973). Os autores concluem que se o limite de $S_n(2)$ é não degenerado então as variáveis aleatórias X_i pertencem ao domínio de atracção da lei gaussiana e $E(X_i) = 0$. Chistyakov e Götze (2004) generalizam o resultado anterior mostrando que o limite de $S_n(2)$ é uma função de distribuição G tal que $G(\{-1, 1\}) \neq 1$ se, e só se, para cada $i = 1, \dots, n$:

- X_i pertence ao domínio de atracção de uma lei estável de expoente característico α , $X_i \in D(\alpha)$;
- $E(X_i) = 0$ se $1 < \alpha \leq 2$;
- se $\alpha = 1$ então X_i pertence ao domínio de atracção da lei Cauchy.

4.4 Momentos da estatística T_n

No que se segue iremos considerar uma variável aleatória X com valor médio $E(X) = \mu$, variância $\text{var}(X) = \sigma^2$ e momentos finitos até, pelo menos, à ordem seis. Dada uma amostra aleatória X_1, \dots, X_n extraída da população parente X consideremos a estatística

$$T_n = \sqrt{n} \frac{\bar{X}_n - \mu}{S_n}.$$

Rocha e Martins (2005) obtêm uma expressão assintótica para o valor médio de T_n ,

$$E(T_n) \approx -\frac{\gamma_1(X)}{2\sqrt{n}}, \quad (69)$$

onde $\gamma_1(X) = \frac{E[(X-\mu)^3]}{\sigma^3}$ é o coeficiente de simetria de X . De (69) resulta que o valor médio da estatística de Student tem sinal contrário do coeficiente de assimetria da população parente em causa.

Nesta secção propomos estender a metodologia adoptada por Rocha e Martins (2005) ao cálculo dos momentos ordinários de ordem dois a quatro da estatística T_n . O procedimento utilizado envolveu a aplicação do método delta, nomeadamente, a expansão em série de Taylor de uma função apropriada e cálculo dos valores esperados dessa expansão.

4.4.1 $E(T_n^2)$

Consideremos a função bvariada

$$f(x, y) = \left(\frac{x - \mu}{\sqrt{y}} \right)^2 = \frac{(x - \mu)^2}{y}$$

de domínio $\mathbb{R} \times \mathbb{R}^+$ que será identificada com $\left(\frac{\bar{X} - \mu}{\sqrt{S^2}} \right)^2$. Vamos calcular o correspondente polinómio de Taylor de ordem 4 no ponto (μ, σ^2) . Para tal, necessitamos de todas as derivadas parciais de f de quarta ordem não nulas em (μ, σ^2) . Nessas condições temos (omitindo o ponto em que é avaliada a derivada):

$$\frac{\partial^2 f}{\partial x^2}(x, y) = \frac{2}{\sigma^2}, \quad \frac{\partial^3 f}{\partial x^2 \partial y}(x, y) = -\frac{2}{\sigma^4} \quad \text{e} \quad \frac{\partial^4 f}{\partial x^2 \partial y^2}(x, y) = \frac{4}{\sigma^6}.$$

Portanto,

$$\begin{aligned} E(T_n^2) &= E \left[\left(\frac{\bar{X} - \mu}{\sqrt{S^2/n}} \right)^2 \right] = nE[f(\bar{X}, S^2)] \\ &\approx n \left\{ \frac{1}{2!} \cdot \frac{2}{\sigma^2} \text{var}(\bar{X}) - \frac{3}{3!} \cdot \frac{2}{\sigma^4} E[(\bar{X} - \mu)^2 (S^2 - \sigma^2)] \right\} \\ &\quad + n \frac{6}{4!} \cdot \frac{4}{\sigma^6} E[(\bar{X} - \mu)^2 (S^2 - \sigma^2)^2] \\ &= \frac{n}{\sigma^2} \text{var}(\bar{X}) - \frac{n}{\sigma^4} E[(\bar{X} - \mu)^2 (S^2 - \sigma^2)] \\ &\quad + \frac{n}{\sigma^6} E[(\bar{X} - \mu)^2 (S^2 - \sigma^2)^2] \tag{70} \\ &= n \left\{ \frac{1}{\sigma^2} \cdot \frac{\sigma^2}{n} - \frac{1}{\sigma^4} E[(\bar{X} - \mu)^2 (S^2 - \sigma^2)] \right\} \\ &\quad + \frac{n}{\sigma^6} E[(\bar{X} - \mu)^2 (S^2 - \sigma^2)^2] \\ &= 1 - \frac{n}{\sigma^4} E[(\bar{X} - \mu)^2 (S^2 - \sigma^2)] + \frac{n}{\sigma^6} E[(\bar{X} - \mu)^2 (S^2 - \sigma^2)^2]. \end{aligned}$$

Para prosseguir necessitamos de conhecer uma expressão para $E[(\bar{X} - \mu)^2 (S^2 - \sigma^2)]$ e para $E[(\bar{X} - \mu)^2 (S^2 - \sigma^2)^2]$.

Em relação ao primeiro caso

$$\mathbb{E} \left[(\bar{X} - \mu)^2 (S^2 - \sigma^2) \right] = \mathbb{E} \left(\bar{X}^2 S^2 \right) - 2\mu \mathbb{E} \left(\bar{X} S^2 \right) + \mu^2 \sigma^2 - \frac{\sigma^4}{n}. \quad (71)$$

É fácil verificar que

$$\mathbb{E} \left(\bar{X} S^2 \right) = \frac{\gamma_1(X) \sigma^3}{n} + \mu \sigma^2 \quad (72)$$

e

$$\begin{aligned} \mathbb{E} \left(\bar{X}^2 S^2 \right) &= \frac{1}{n^2} \left[\mathbb{E} (X^4) + 2(n-2) \mu \mathbb{E} (X^3) + (n-3) (\mathbb{E} (X^2))^2 \right] \\ &+ \frac{1}{n^2} \left[(n-6)(n-2) \mu^2 \mathbb{E} (X^2) - (n-2)(n-3) \mu^4 \right]. \end{aligned} \quad (73)$$

Substituindo em (71)

$$\begin{aligned} \mathbb{E} \left[(\bar{X} - \mu)^2 (S^2 - \sigma^2) \right] &= \frac{1}{n^2} \left[\mathbb{E} (X^4) + 2(n-2) \mu \mathbb{E} (X^3) \right] \\ &+ \frac{1}{n^2} \left[(n-3) (\mathbb{E} (X^2))^2 + (n-6)(n-2) \mu^2 \mathbb{E} (X^2) \right] \\ &- \frac{1}{n^2} (n-2)(n-3) \mu^4 \\ &- 2\mu \left(\frac{\gamma_1(X) \sigma^3}{n} + \mu \sigma^2 \right) + \mu^2 \sigma^2 - \frac{\sigma^4}{n} \\ &= \frac{\mathbb{E} (X^4) - 4\mu \mathbb{E} (X^3) - 3\sigma^4 + 6\mu^2 \sigma^2 + 3\mu^4}{n^2} \\ &= \frac{\sigma^4 \gamma_2(X)}{n^2}. \end{aligned}$$

Em relação ao segundo caso

$$\begin{aligned} \mathbb{E} \left[(\bar{X} - \mu)^2 (S^2 - \sigma^2)^2 \right] &= \mathbb{E} \left(\bar{X}^2 S^4 \right) - 2\sigma^2 \mathbb{E} \left(\bar{X}^2 S^2 \right) + \sigma^4 \mathbb{E} \left(\bar{X}^2 \right) \\ &- 2\mu \mathbb{E} \left(\bar{X} S^4 \right) + 4\mu \sigma^2 \mathbb{E} \left(\bar{X} S^2 \right) + \mu^2 \mathbb{E} \left(S^4 \right) - 3\mu^2 \sigma^4. \end{aligned} \quad (74)$$

Portanto, necessitamos de conhecer

$$\mathbb{E} (S^4) = \frac{\sigma^4 (\gamma_2(X) + 3)}{n} + \frac{n^2 - 2n + 3}{n(n-1)} \sigma^4, \quad (75)$$

$$\begin{aligned}
\mathbb{E}(\bar{X}S^4) &= \frac{\mathbb{E}(X^5)}{n^2} + \frac{(n-5)\mu\mathbb{E}(X^4)}{n^2} \\
&+ \frac{(2n^2 - 8n + 10)(\sigma^2 + \mu^2)\mathbb{E}(X^3)}{n^2(n-1)} \\
&+ \frac{((n-2)n^2 - 6n(n-2) + 15(n-2))\mu(\sigma^2 + \mu^2)^2}{n^2(n-1)} \\
&+ \frac{(-6n(n-2) + 10(n-2))\mu^2\mathbb{E}(X^3)}{n^2(n-1)} \\
&+ \frac{(-2n(n-2)(n-3) + 10(n-2)(n-3))\mu^3(\sigma^2 + \mu^2)}{n^2(n-1)} \\
&+ \frac{(n-2)(n-3)(n-4)\mu^5}{n^2(n-1)}
\end{aligned} \tag{76}$$

e

$$\begin{aligned}
\mathbb{E}(\bar{X}^2S^4) &= \frac{\mathbb{E}(X^6)}{n^3} + \frac{2(n-3)\mu\mathbb{E}(X^5)}{n^3} + \frac{(2n^2 - 8n + 10)(\mathbb{E}(X^3))^2}{n^3(n-1)} \\
&+ \frac{(3n^2 - 14n + 15)(\sigma^2 + \mu^2)\mathbb{E}(X^4)}{n^3(n-1)} \\
&+ \frac{(n-2)(n^2 - 6n + 15)(\mathbb{E}(X^2))^3}{n^3(n-1)} \\
&+ \frac{(n-2)(n^2 - 12n + 15)\mu^2\mathbb{E}(X^4)}{n^3(n-1)} \\
&+ \frac{(n-2)(4n^2 - 32n + 60)\mu\mathbb{E}(X^2)\mathbb{E}(X^3)}{n^3(n-1)} \\
&+ \frac{(n-2)(n-3)(n^2 - 12n + 45)\mu^2(\mathbb{E}(X^2))^2}{n^3(n-1)} \\
&+ \frac{(n-2)(n-3)(-8n + 20)\mu^3\mathbb{E}(X^3)}{n^3(n-1)} \\
&+ \frac{(n-2)(n-3)(n-4)(-2n + 15)\mu^4\mathbb{E}(X^2)}{n^3(n-1)} \\
&+ \frac{(n-2)(n-3)(n-4)(n-5)\mu^6}{n^3(n-1)}.
\end{aligned} \tag{77}$$

Tendo em conta (72), (73) e os resultados anteriores (74) fica igual a

$$\mathbb{E}\left[(\bar{X} - \mu)^2(S^2 - \sigma^2)^2\right] = \sigma^6 \frac{\gamma_2(X) + 2\gamma_1^2(X) + 2}{n(n-1)} + O(n^{-3}).$$

Continuando o desenvolvimento de (70) resulta que

$$\begin{aligned}
E(T_n^2) &\approx 1 - \frac{n}{\sigma^4} E\left[(\bar{X} - \mu)^2 (S^2 - \sigma^2)\right] + \frac{n}{\sigma^6} E\left[(\bar{X} - \mu)^2 (S^2 - \sigma^2)^2\right] \\
&= 1 - \frac{n}{\sigma^4} \frac{\sigma^4 \gamma_2(X)}{n^2} + \frac{n}{\sigma^6} \frac{\sigma^6 \gamma_2(X) + 2\gamma_1^2(X) + 2}{n(n-1)} + O(n^{-2}) \\
&= 1 - \frac{\gamma_2(X)}{n} + \frac{\gamma_2(X) + 2\gamma_1^2(X) + 2}{(n-1)} + O(n^{-2}) \\
&= 1 - \frac{\gamma_2(X)}{n} + \frac{\gamma_2(X) + 2\gamma_1^2(X) + 2}{(n-1)} + O(n^{-2}) \\
&= 1 + 2\frac{\gamma_1^2(X) + 1}{n} + \frac{\gamma_2(X)}{n(n-1)} + O(n^{-2}) \\
&= 1 + 2\frac{\gamma_1^2(X) + 1}{n} + O(n^{-2}).
\end{aligned} \tag{78}$$

Do valor obtido na aproximação resulta que $E(T^2)$ converge, quando $n \rightarrow \infty$, para um como seria de esperar. O mais interessante é que a convergência é feita pelo lado direito, isto é, por valores acima de um. Esta mesma conclusão é verificada pelas simulações realizadas mais adiante.

4.4.2 $E(T_n^3)$

Para o cálculo de $E(T_n^3)$ o procedimento é semelhante ao anterior aumentando apenas a quantidade de cálculos envolvidos. Desta vez, a função bivariada a considerar naturalmente será

$$f(x, y) = \left(\frac{x - \mu}{\sqrt{y}}\right)^3 = \frac{(x - \mu)^3}{\sqrt{y^3}}.$$

Iremos considerar o polinómio de Taylor de ordem três no ponto (μ, σ^2) . As derivadas parciais envolvidas na aproximação são:

$$\frac{\partial^3 f}{\partial x^3}(x, y) = \frac{6}{\sigma^3} \quad \text{e} \quad \frac{\partial^4 f}{\partial x^3 \partial y}(x, y) = -\frac{9}{\sigma^5}.$$

Assim,

$$\begin{aligned}
 E(T_n^3) &= E\left[\left(\frac{\bar{X} - \mu}{\sqrt{S^2/n}}\right)^3\right] = n^{3/2}E[f(\bar{X}, S^2)] \\
 &\approx \frac{n^{3/2}}{3!}E\left[\frac{6}{\sigma^3}(\bar{X} - \mu)^3 - \frac{9}{\sigma^5}(\bar{X} - \mu)^3(S^2 - \sigma^2)\right] \\
 &= \frac{n^{3/2}}{\sigma^3}E(\bar{X} - \mu)^3 - \frac{3n^{3/2}}{2\sigma^5}E[(\bar{X} - \mu)^3(S^2 - \sigma^2)] \\
 &= \frac{n^{3/2}\sigma^3}{\sigma^3}\gamma_1(\bar{X}) - \frac{3n^{3/2}}{2\sigma^5}E[(\bar{X} - \mu)^3(S^2 - \sigma^2)].
 \end{aligned} \tag{79}$$

Para determinar $\gamma_1(\bar{X})$ em função dos momentos populacionais podemos recorrer à função geradora de cumulantes de \bar{X} , $K_{\bar{X}} = \ln M_{\bar{X}}$ onde $M_{\bar{X}}$ é a função geradora de momentos de \bar{X} . Ora,

$$\begin{aligned}
 M_{\bar{X}}(t) &= E(\exp(t\bar{X})) \\
 &= E\left(\exp\left(\frac{t}{n}\sum_{i=1}^n X_i\right)\right) \\
 &= \prod_{i=1}^n E\left(\exp\left(\frac{t}{n}X_i\right)\right) \\
 &= \prod_{i=1}^n M_{X_i}\left(\frac{t}{n}\right) \\
 &= \left[M_X\left(\frac{t}{n}\right)\right]^n
 \end{aligned} \tag{80}$$

pelo que

$$K_{\bar{X}}(t) = \ln \left[M^n \left(\frac{t}{n} \right) \right] = n \ln \left[M \left(\frac{t}{n} \right) \right] = nK_X \left(\frac{t}{n} \right). \tag{81}$$

Logo,

$$\gamma_1(\bar{X}) = \frac{E[(\bar{X} - \mu)^3]}{\sigma_{\bar{X}}^3} = \frac{nk_3}{(nk_2)^{1.5}} = \frac{k_3}{\sqrt{nk_2^{1.5}}} = \frac{\gamma_1(X)}{\sqrt{n}}$$

onde k_2 e k_3 são, respectivamente, o segundo e terceiro cumulante de X .

Em relação a $E[(\bar{X} - \mu)^3(S^2 - \sigma^2)]$ podemos escrever como sendo igual a

$$\begin{aligned} \mathbb{E} \left[(\bar{X} - \mu)^3 (S^2 - \sigma^2) \right] &= \mathbb{E} \left(\bar{X}^3 S^2 \right) - 3\mu \mathbb{E} \left(\bar{X}^2 S^2 \right) + 3\mu^2 \mathbb{E} \left(\bar{X} S^2 \right) \\ &\quad - \sigma^2 \mathbb{E} \left(\bar{X}^3 \right) + 3\mu\sigma^2 \left(\frac{\sigma^2}{n} + \mu^2 \right) - 3\mu^3 \sigma^2. \end{aligned} \quad (82)$$

Ora,

$$\mathbb{E} \left(\bar{X}^3 \right) = \frac{(n-1)(n+1)\mu^3 + 3(n-1)\mu\sigma^2 + \mathbb{E}(X^3)}{n^2} \quad (83)$$

e

$$\begin{aligned} \mathbb{E} \left(\bar{X}^3 S^2 \right) &= \frac{3\mu \mathbb{E}(X^4) + 4(\sigma^2 + \mu^2) \mathbb{E}(X^3) + (3n-16)\mu^2 \mathbb{E}(X^3)}{n^2} \\ &\quad + \frac{(3n-21)\mu \mathbb{E}(X^2)^2 + (n^2 - 15n + 56)\mu^3 \mathbb{E}(X^2)}{n^2} \\ &\quad - \frac{(n^2 - 9n + 26)\mu^5}{n^2} + O(n^{-3}). \end{aligned} \quad (84)$$

Logo, por (72), (73), (83) e (84) resulta que

$$\mathbb{E} \left[(\bar{X} - \mu)^3 (S^2 - \sigma^2) \right] = 3\sigma^5 \frac{\gamma_1(X)}{n^2} + O(n^{-3}).$$

Substituindo em (79) obtém-se

$$\begin{aligned} \mathbb{E}(T_n^3) &\approx n^{3/2} \frac{\sigma_{\bar{X}}^3}{\sigma^3} \gamma_1(\bar{X}) - \frac{3n^{3/2}}{2\sigma^5} \left[\mathbb{E} \left[(\bar{X} - \mu)^3 (S^2 - \sigma^2) \right] \right] \\ &= \frac{\gamma_1(X)}{\sqrt{n}} - \frac{9}{2} \frac{\gamma_1(X)}{\sqrt{n}} + O(n^{-3/2}) \\ &= -\frac{7}{2\sqrt{n}} \gamma_1(X) + O(n^{-3/2}). \end{aligned} \quad (85)$$

O sinal da assimetria da estatística de Student é contrário ao sinal da assimetria da respectiva distribuição parente. Podemos ainda calcular o terceiro momento centrado de T_n .

$$\begin{aligned} \mathbb{E}(T_n - \mu)^3 &\approx -\frac{7}{2\sqrt{n}} \gamma_1(X) - 3 \left(-\frac{\gamma_1(X)}{2\sqrt{n}} \right) \left(1 + 2 \frac{\gamma_1^2(X) + 1}{n} \right) \\ &\quad - 3 \left(-\frac{\gamma_1(X)}{2\sqrt{n}} \right)^3 + O(n^{-3/2}) \\ &= -\frac{2\gamma_1(X)}{\sqrt{n}} + O(n^{-3/2}) \end{aligned}$$

4.4.3 $E(T_n^4)$

No cálculo de uma aproximação para este último valor esperado teremos em conta a função

$$f(x, y) = \left(\frac{x - \mu}{\sqrt{y}} \right)^4 = \frac{(x - \mu)^4}{y^2}.$$

Neste caso, iremos utilizar o polinómio de Taylor de ordem seis. Necessitamos, para avançar com os cálculos, as seguintes derivadas parciais avaliadas no ponto (μ, σ^2) :

$$\frac{\partial^4 f}{\partial x^4}(x, y) = \frac{4!}{\sigma^4}, \quad \frac{\partial^5 f}{\partial x^4 \partial y}(x, y) = -\frac{48}{\sigma^6} \quad \text{e} \quad \frac{\partial^6 f}{\partial x^4 \partial y^2}(x, y) = \frac{144}{\sigma^8}.$$

Assim,

$$\begin{aligned} E(T_n^4) &= E \left[\left(\frac{\bar{X} - \mu}{\sqrt{S^2/n}} \right)^4 \right] = n^2 E[f(\bar{X}, S^2)] \\ &\approx n^2 \left[\frac{1}{4!} \frac{4!}{\sigma^4} E(\bar{X} - \mu)^4 - \frac{5 \times 48}{5! \sigma^6} E((\bar{X} - \mu)^4 (S^2 - \sigma^2)) \right] \\ &+ n^2 \left[\frac{15 \times 144}{6! \sigma^8} E((\bar{X} - \mu)^4 (S^2 - \sigma^2)) \right] \quad (86) \\ &= 3 + \gamma_2(\bar{X}) - \frac{2n^2}{\sigma^6} E((\bar{X} - \mu)^4 (S^2 - \sigma^2)) \\ &+ \frac{3n^2}{\sigma^8} E((\bar{X} - \mu)^4 (S^2 - \sigma^2)). \end{aligned}$$

Ora, torna-se necessário determinar o valor de $\gamma_2(\bar{X})$, $E((\bar{X} - \mu)^4 (S^2 - \sigma^2))$ e $E((\bar{X} - \mu)^4 (S^2 - \sigma^2))$ em função dos momentos ordinários. Começando por $\gamma_2(\bar{X})$, basta recordar as igualdades (80) e (81) para concluir que

$$\gamma_2(\bar{X}) = \frac{E[(\bar{X} - \mu)^4]}{\sigma_{\bar{X}}^4} - 3 = \frac{nk_4}{(nk_2)^2} = \frac{\gamma_2(X)}{n} \quad (87)$$

onde k_2 e k_4 são, respectivamente, o segundo e quarto cumulante.

$E((\bar{X} - \mu)^4 (S^2 - \sigma^2))$ é igual a

$$\begin{aligned} & \mathbb{E} \left(\bar{X}^4 S^2 - 4\mu \bar{X}^3 S^2 + 6\mu^2 \bar{X}^2 S^2 - 4\mu^3 \bar{X} S^2 \right) \\ & + \mathbb{E} \left(-\sigma^2 \bar{X}^4 + 4\mu \sigma^2 \bar{X}^3 - 6\mu^2 \sigma^2 \bar{X}^2 + 4\mu^3 \sigma^2 \bar{X} \right). \end{aligned} \quad (88)$$

Ora,

$$\begin{aligned} \mathbb{E} \left(\bar{X}^4 \right) &= \frac{\mathbb{E} (X^4) + 4(n-1)\mu \mathbb{E} (X^3) + 3(n-1)(\sigma^2 + \mu^2)^2}{n^3} \\ &+ \frac{6(n-1)(n-2)\mu^2(\sigma^2 + \mu^2) + (n-1)(n-2)(n-3)\mu^4}{n^3} \end{aligned}$$

e

$$\begin{aligned} \mathbb{E} \left(\bar{X}^4 S^2 \right) &= \frac{4\mu \mathbb{E} (X^5) + 7\sigma^2 \mathbb{E} (X^4) + (6n-20)\mu^2 \mathbb{E} (X^4)}{n^3} \\ &+ \frac{4(\mathbb{E} (X^3))^2 + (16n-92)\mu \sigma^2 \mathbb{E} (X^3)}{n^3} \\ &+ \frac{(4n^2 - 24n + 32)\mu^3 \mathbb{E} (X^3) + (n^3 - 12n^2 + 20n + 45)\mu^4 \sigma^2}{n^3} \\ &+ \frac{(6n^2 - 66n + 198)\sigma^4 \mu^2 + (3n - 21)\sigma^6}{n^3} \\ &+ \frac{(-4n^2 + 18n - 20)\mu^6}{n^3} + O(n^{-4}). \end{aligned}$$

Donde, (88) é igual a

$$\mathbb{E} \left((\bar{X} - \mu)^4 (S^2 - \sigma^2) \right) = \frac{6\gamma_2(X) + 4\gamma_1^2(X)}{n^3} \sigma^6 + O(n^{-4}). \quad (89)$$

Por fim,

$$\begin{aligned} \mathbb{E} \left((\bar{X} - \mu)^4 (S^2 - \sigma^2)^2 \right) &= \mathbb{E} \left(\bar{X}^4 S^4 \right) + \mu^4 \mathbb{E} (S^4) + \sigma^4 \mathbb{E} \left(\bar{X}^4 \right) \\ &- 4\mu^3 \mathbb{E} (\bar{X} S^4) - 4\mu \mathbb{E} \left(\bar{X}^3 S^4 \right) - 4\mu \sigma^4 \mathbb{E} \left(\bar{X}^3 \right) \\ &- 2\sigma^2 \mathbb{E} \left(\bar{X}^4 S^2 \right) + 6\mu^2 \mathbb{E} \left(\bar{X}^2 S^4 \right) + 6\mu^2 \sigma^4 \mathbb{E} \left(\bar{X}^2 \right) \\ &+ 8\mu^3 \sigma^2 \mathbb{E} (\bar{X} S^2) + 8\mu \sigma^2 \mathbb{E} \left(\bar{X}^3 S^2 \right) \\ &- 5\mu^4 \sigma^4 - 12\mu^2 \sigma^2 \mathbb{E} \left(\bar{X}^2 S^2 \right) \end{aligned} \quad (90)$$

Para conseguir simplificar a expressão (90) é necessário calcular os seguintes momentos:

$$\begin{aligned} \mathbb{E}(\overline{X}^4) &= \frac{4\mu\mathbb{E}(X^3) + 3\sigma^4 - 12\sigma^2\mu^2 - 4\mu^4 + 6\mu^2n\sigma^2 + \mu^4n^2}{n^2} \\ &\quad + \frac{d - 4\mu c + 6\mu^2\sigma^2 + 3\mu^4 - 3\sigma^4}{n^3}, \end{aligned}$$

$$\begin{aligned} \mathbb{E}(\overline{X}^4 S^2) &= \frac{4\mu\mathbb{E}(X^5) + 7\sigma^2\mathbb{E}(X^4) + (6n - 20)\mu^2\mathbb{E}(X^4)}{n^3} \\ &\quad + \frac{4(\mathbb{E}(X^3))^2 + (16n - 92)\mu\sigma^2\mathbb{E}(X^3) + (4n^2 - 24n + 32)\mu^3\mathbb{E}(X^6)}{n^3} \\ &\quad + \frac{(n^3 - 12n^2 + 20n + 45)\mu^4\sigma^2 + (6n^2 - 66n + 198)\sigma^4\mu^2}{n^3} \\ &\quad + \frac{(3n - 21)\sigma^6 + (-4n^2 + 18n - 20)\mu^6}{n^3} + O(n^{-4}), \end{aligned}$$

$$\begin{aligned} \mathbb{E}(\overline{X}^4 S^4) &= \frac{6\mu^2\mathbb{E}(X^6) + 20\mu(\sigma^2 + \mu^2)\mathbb{E}(X^5) + 28\mu\mathbb{E}(X^3)\mathbb{E}(X^4)}{n^2(n-1)} \\ &\quad + \frac{16(\sigma^2 + \mu^2)^2\mathbb{E}(X^4) + 20(\sigma^2 + \mu^2)(\mathbb{E}(X^3))^2 + (-60 + 4n)\mu^3\mathbb{E}(X^5)}{n^2(n-1)} \\ &\quad + \frac{(-300 + 18n)\mu^2(\sigma^2 + \mu^2)\mathbb{E}(X^4) + (12n - 180)\mu^2(\mathbb{E}(X^3))^2}{n^2(n-1)} \\ &\quad + \frac{(28n - 440)\mu(\sigma^2 + \mu^2)^2\mathbb{E}(X^3) + (3n - 45)(\sigma^2 + \mu^2)^4}{n^2(n-1)} \\ &\quad + \frac{(n^2 - 39n + 366)\mu^4\mathbb{E}(X^4) + (8n^2 - 232n + 2208)\mu^3(\sigma^2 + \mu^2)\mathbb{E}(X^3)}{n^2(n-1)} \\ &\quad + \frac{(6n^2 - 144n + 1386)\mu^2(\sigma^2 + \mu^2)^3 + (-12n^2 + 224n - 1636)\mu^5\mathbb{E}(X^3)}{n^2(n-1)} \\ &\quad + \frac{(n^3 - 44n^2 + 701n - 5224)\mu^4(\sigma^2 + \mu^2)^2}{n^2(n-1)} \\ &\quad + \frac{(-2n^3 + 68n^2 - 870n + 5500)\mu^6(\sigma^2 + \mu^2)}{n^2(n-1)} \\ &\quad + \frac{(n^3 - 27n^2 + 295n - 1665)\mu^8}{n^2(n-1)} + O(n^{-4}) \end{aligned}$$

e

$$\begin{aligned}
 E\left(\overline{X}^3 S^4\right) &= \frac{3\mu E(X^6)}{n^3} + \frac{5(\sigma^2 + \mu^2) E(X^5) + 7E(X^3) E(X^4) + (3n - 26) \mu^2 E(X^5)}{n^2(n-1)} \\
 &+ \frac{(9n - 88) \mu (\sigma^2 + \mu^2) (X^4) + (6n - 52) \mu (E(X^3))^2}{n^2(n-1)} \\
 &+ \frac{(7n - 64) (\sigma^2 + \mu^2)^2 E(X^3) + (3n^2 - 45n + 273) \mu (\sigma^2 + \mu^2)^3}{n^2(n-1)} \\
 &+ \frac{(n^2 - 25n + 141) \mu^3 E(X^4) + (6n^2 - 110n + 646) \mu^2 (\sigma^2 + \mu^2) E(X^3)}{n^2(n-1)} \\
 &+ \frac{(n^3 - 29n^2 + 311n - 1489) \mu^3 (\sigma^2 + \mu^2)^2}{n^2(n-1)} \\
 &+ \frac{(-10n^2 + 125n - 575) \mu^4 E(X^3)}{n^2(n-1)} \\
 &+ \frac{(-2n^3 + 49n^2 - 436n + 1799) \mu^5 (\sigma^2 + \mu^2)}{n^2(n-1)} \\
 &+ \frac{(n^3 - 20n^2 + 155n - 580) \mu^7}{n^2(n-1)} + O(n^{-4}).
 \end{aligned}$$

Com mais estes quatro resultados estamos em condições de determinar $E\left((\overline{X} - \mu)^4 (S^2 - \sigma^2)^2\right)$ obtendo-se

$$E\left((\overline{X} - \mu)^4 (S^2 - \sigma^2)^2\right) = \frac{12\sigma^8 \gamma_1^2(X) + 3\sigma^8 \gamma_2(X) + 6\sigma^8}{n^2(n-1)} + O(n^{-3}). \quad (91)$$

Assim, utilizando (87), (89) e (91) em (86) obtemos

$$\begin{aligned}
 E(T_n^4) &\approx 3 + \gamma_2(\overline{X}) - \frac{2n^2}{\sigma^6} E\left((\overline{X} - \mu)^4 (S^2 - \sigma^2)^2\right) \\
 &- \frac{2}{\sigma^6} \left(\frac{6\gamma_2(X) + 4\gamma_1^2(X)}{n} \sigma^6 \right) \\
 &= 3 + \frac{\gamma_2(X)}{n} - \frac{2}{\sigma^6} \left(\frac{6\gamma_2(X) + 4\gamma_1^2(X)}{n} \sigma^6 \right) \\
 &+ \frac{3}{\sigma^8} \left(\frac{12\sigma^8 \gamma_1^2(X)}{(n-1)} + \frac{3\sigma^8 \gamma_2(X)}{(n-1)} + \frac{6\sigma^8}{(n-1)} \right) + O(n^{-2}).
 \end{aligned}$$

Logo,

$$\begin{aligned}
 E(T_n^4) &= 3 + \frac{\gamma_2(X)}{n} - 4 \left(\frac{3\gamma_2(X) + 2\gamma_1^2(X)}{n} \right) \\
 &\quad + \frac{9}{n-1} (4\gamma_1^2(X) + \gamma_2(X) + 2) + O(n^{-2}) \\
 &= 3 + 2 \frac{14\gamma_1^2(X) - \gamma_2(X) + 9}{n-1} + O(n^{-2}).
 \end{aligned} \tag{92}$$

O momento ordinário de quarta ordem da estatística T_n de Student depende sobretudo da assimetria da distribuição parente, mais propriamente, do quadrado desse valor. Assim, o sinal da assimetria não tem influência no valor de $E(T_n^4)$. O que realmente interessa é a intensidade da assimetria. A não ser para casos de elevadíssimo achatamento, a expressão (92) permite concluir que o momento de quarta ordem é, em geral, superior ao valor limite três. Tal como no caso momento de ordem dois, a aproximação a três é feita por valores à direita. As simulações apresentadas adiante confirmam esta observação.

Para terminar iremos calcular uma expressão para o momento centrado de quarta ordem de T_n . A importância do quadrado da assimetria nos momentos ordinários de ordem par reflecte-se agora na aproximação a $E(T_n - \mu)^4$:

$$\begin{aligned}
 E(T_n - \mu)^4 &\approx 3 + 2 \frac{14\gamma_1^2(X) - \gamma_2(X) + 9}{n-1} - 4 \left(-\frac{\gamma_1(X)}{2\sqrt{n}} \right) \left(-\frac{7\gamma_1(X)}{2\sqrt{n}} \right) \\
 &\quad + 6 \left(-\frac{\gamma_1(X)}{2\sqrt{n}} \right)^2 \left(1 + 2 \frac{\gamma_1^2(X) + 1}{n} \right) - 3 \left(-\frac{\gamma_1(X)}{2\sqrt{n}} \right)^4 \\
 &= 3 + \frac{-4\gamma_2(X) + 45\gamma_1^2(X) + 36}{2(n-1)} + O(n^{-2}).
 \end{aligned}$$

4.4.4 Alguns resultados de simulação

Para determinar se as aproximações obtidas estariam de acordo com estimativas obtidas por simulação efectuou-se um trabalho de simulação de 10000 amostras de diferentes dimensões n de distribuições parentes F_X distintas. Depois, compararam-se os valores obtidos por simulação com as estimativas obtidas anteriormente para os primeiros quatro momentos ordinários de T_n . Recordando os resultados anteriores, as estimativas a utilizar serão:

Momento de T_n	Estimativa
$E(T_n)$	$-\frac{\gamma_1(X)}{2\sqrt{n}}$
$E(T_n^2)$	$1 + 2\frac{\gamma_1^2(X) + 1}{n}$
$E(T_n^3)$	$-\frac{7\gamma_1(X)}{2\sqrt{n}}$
$E(T_n^4)$	$3 + 2\frac{14\gamma_1^2(X) - \gamma_2(X) + 9}{n - 1}$

Tabela 18 – Estimativas para quatro momentos de T_n

As tabelas seguintes comparam os resultados estimados com os resultados obtidos por simulação.

X	n	$E(T_n)$		$E(T_n^2)$	
		Estim.	Simul.	Estim.	Simul.
$Y - 1$ onde $Y \sim \text{Poisson}(1)$	25	-0.1	-0.1041	1.16	1.1606
$-Y$ onde $Y \sim \text{Exp}(1)$	25	0.2	0.2214	1.4	1.3966
	400	0.05	0.0522	1.025	1.0197
Uniforme(0, 1)	25	0	0.0151	1.08	1.0672
	36	0	0.0104	1.0556	1.0607
Beta(1/3, 2/3)	25	-0.0667	-0.0807	1.1156	1.1281
	20	-0.0745	-0.0837	1.1444	1.1521
Beta(2/3, 1/3)	25	0.0667	0.0762	1.1156	1.1239
	81	0.0370	0.0354	1.0357	1.0071
Laplace(1)	25	0	-0.0076	1.08	1.0695
Binomial(10, 0.1)	20	-0.0943	-0.1126	1.1711	1.2016
$\chi_{(1)}^2$	20	-0.3162	-0.4030	1.9	2.1673
	81	-0.1571	-0.1649	1.2222	1.2316
	400	-0.0707	-0.0860	1.045	1.0532

Tabela 19 – Valores estimados *versus* valores simulados para $E(T_n)$ e $E(T_n^2)$

X	n	$E(T_n^3)$		$E(T_n^4)$	
		Estim.	Simul.	Estim.	Simul.
$Y - 1$ onde $Y \sim \text{Poisson}(1)$	25	-0.7	-0.9691	4.8333	5.7531
$-Y$ onde $Y \sim \text{Exp}(1)$	25	1.4	2.3731	7.9167	10.9118
	400	0.35	0.3046	3.2957	3.1950
Uniforme(0, 1)	25	0	0.0750	3.85	4.0035
	36	0	0.0102	3.5829	3.7474
Beta(1/3, 2/3)	25	-0.4667	-0.7619	4.3519	5.3808
	20	-0.5217	-0.8835	4.7076	5.8624
Beta(2/3, 1/3)	25	0.4667	0.6602	4.3519	4.7313
	81	0.2593	0.2443	3.4056	3.1589
Laplace(1)	25	0	-0.0463	3.5	3.4904
Binomial(10, 0.1)	20	-0.6599	-0.9722	4.9415	5.7709
$\chi_{(1)}^2$	20	-2.2136	-7.0806	14.4736	42.7626
	81	-1.0999	-1.5932	5.725	7.4493
	400	-0.4950	-0.6370	3.5464	3.7802

Tabela 20 – Valores estimados *versus* valores simulados para $E(T_n^3)$ e $E(T_n^4)$

A seguir é apresentado para cada um dos momentos ordinários de T_n em análise um gráfico de dispersão comparando os valores estimados com os valores obtidos por simulação.

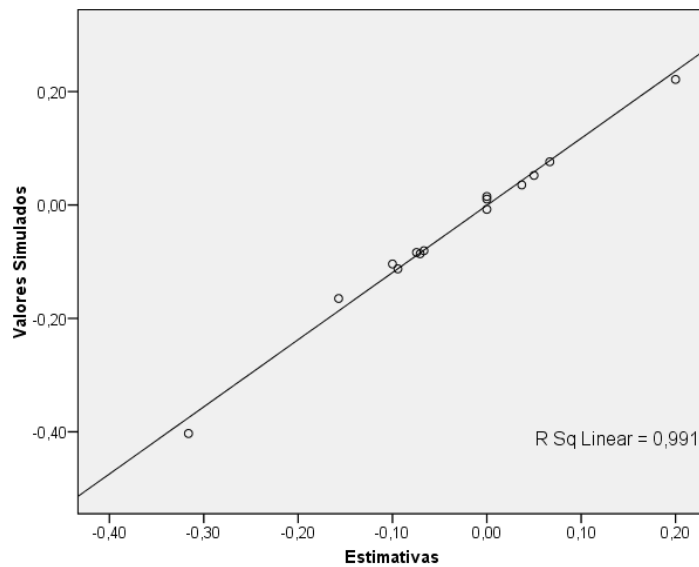


Gráfico 5 – Valores estimados *versus* valores simulados de $E(T_n)$

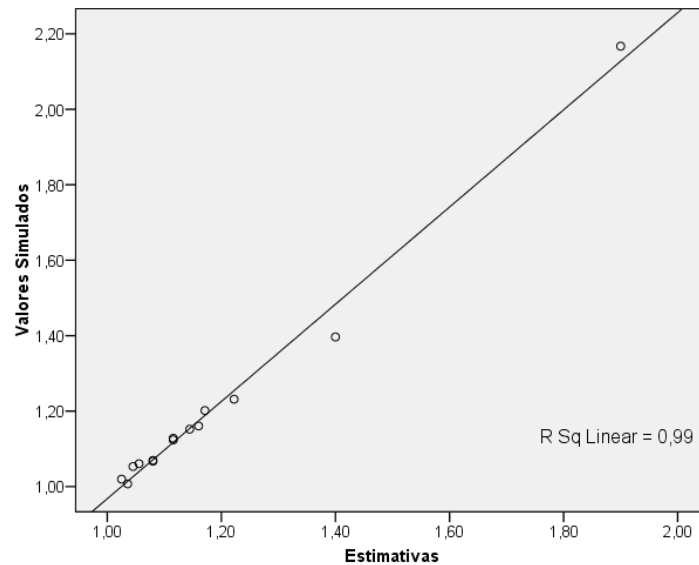


Gráfico 6 – Valores estimados *versus* valores simulados de $E(T_n^2)$

No Gráfico 6 temos duas observações muito influentes. Porém retirando a observação mais influente ou até essas duas observações o coeficiente de determinação mantém-se elevado, isto é, acima dos 0,97.

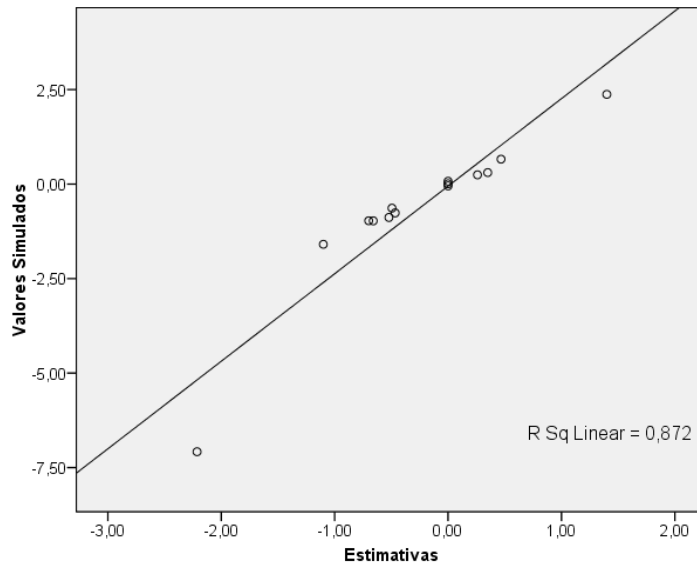


Gráfico 7 – Valores estimados *versus* valores simulados de $E(T_n^3)$

Retirando a observação mais influente do Gráfico 7 o coeficiente de determinação sobe imediatamente para 0.985.

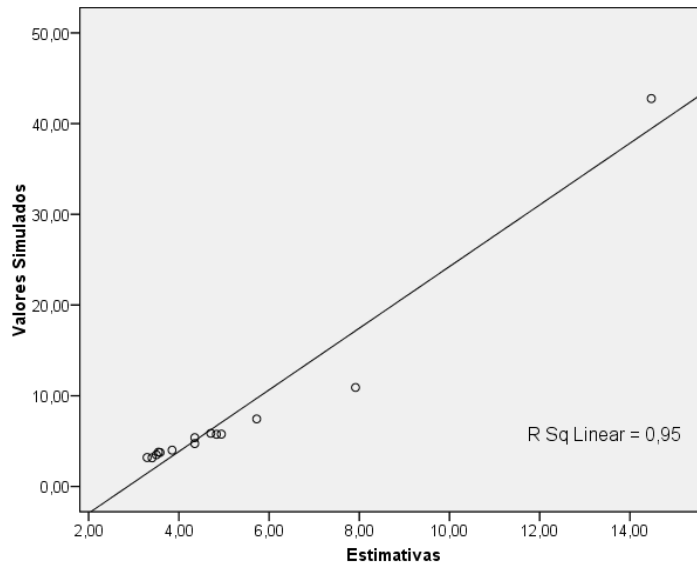


Gráfico 8 – Valores estimados *versus* valores simulados de $E(T_n^4)$

Tal como no Gráfico 7, se retirarmos a observação mais influente do Gráfico 8 o coeficiente de determinação aumenta. Neste caso, o coeficiente de determinação aumenta para 0.991.

Como seria de esperar estimativas e simulações estão altamente correlacionadas e observa-se, tal como seria desejável, a ausência de um padrão de sobreestimação ou subestimação. No que concerne aos momentos de ordem ímpar, as aproximações obtidas são aceitáveis na medida em que estimativas e valores obtidos por simulação coincidem em sinal. As expressões (78) e (92) para $E(T^2)$ e $E(T^4)$ indicavam que a aproximação aos valores limite um e três, respectivamente, seria feita por valores à direita. Tal é verificado pelos resultados da simulação.

4.5 Sistema de distribuições de Pearson

Na análise de dados, para além de transformar, pretendemos frequentemente ajustar um modelo distribucional. Assim, um dos grandes problemas ao nível da Estatística é encontrar famílias de distribuições que se possam ajustar a uma diversidade de dados tão vasta quanto possível. Um grupo de distribuições nessas condições é habitualmente designado de **sistema de distribuições**.

Inicialmente o ajustamento foi feito utilizando a distribuição gaussiana variando apenas a localização (valor médio) e a escala (desvio padrão). Porém, era descurada uma componente muito importante que é a assimetria. Como defende Efron (1969), a simetria é uma propriedade mais importante do que habitualmente se julga.

Os trabalhos nesta área têm revelado que os três primeiros momentos e algum conhecimento acerca das caudas são determinantes na identificação de uma distribuição que se ajuste de forma conveniente a um conjunto de dados. Um dos sistemas de distribuições mais populares é o sistema de distribuições de Pearson estabelecido por Karl Pearson (1895). Salsburg (2002) que no seu livro *The Lady*

Tasting Tea procede a uma resenha dos eventos mais importantes na história da Estatística tem um comentário curioso em relação à criação do sistema de distribuições de Pearson. O autor defende que a grande criação de Pearson é a assimetria. Os modelos gaussianos, dependentes unicamente de uma localização e de uma escala, eram muito limitados em termos de modelação de dados reais. A introdução do terceiro e quarto momentos nos modelos permitiu captar mais informação da amostra.

Esse sistema de distribuições consiste no conjunto de funções densidade de probabilidade f que satisfazem a equação diferencial:

$$(\ln f_X(x))' = -\frac{x+a}{b_0+b_1x+b_2x^2} \quad (93)$$

onde a, b_0, b_1 e b_2 são parâmetros da distribuição.

Como referem Andreev, Kanto e Malo (2005) uma das propriedades do sistema de distribuições de Pearson que o torna tão utilizado é a relação directa entre os quatro parâmetros da distribuição e os quatro primeiros momentos centrados μ_1, μ_2, μ_3 e μ_4 . Essa relação pode ser obtida multiplicando ambos os membros de (93) por x^r obtendo-se

$$x^r (b_0 + b_1x + b_2x^2) f'(x) + x^r (x+a) f(x) = 0.$$

Integrando por partes a expressão anterior e admitindo que $x^r f(x) \rightarrow 0$ quando $x \rightarrow \pm\infty$ para $r \leq 5$ temos que

$$-rb_0\mu'_{r-1} - (r+1)b_1\mu'_r - (r+2)b_2\mu'_{r+1} + \mu'_{r+1} + a\mu'_r = 0.$$

Substituindo na equação anterior a constante r por 0, 1, 2 e 3 obtemos o sistema (admitindo $\mu'_{-1} = 0, \mu'_0 = 1$ e $\mu'_1 = 0$)

$$\begin{cases} -b_1 + a = 0 \\ -b_0 - 3b_2\mu'_2 + \mu'_2 = 0 \\ -3b_1\mu'_2 - 4b_2\mu'_3 + \mu'_3 + a\mu'_2 = 0 \\ -3b_0\mu'_2 - 4b_1\mu'_3 - 5b_2\mu'_4 + \mu'_4 + a\mu'_3 = 0 \end{cases} \quad (94)$$

Resolvendo o sistema em ordem a a, b_0, b_1 e b_2 resulta que

$$\begin{cases} b_0 = \frac{3\mu_3^2 - 4\mu_2\mu_4}{18\mu_2^3 + 12\mu_3^2 - 10\mu_2\mu_4} \mu_2 \\ b_1 = a = -\frac{3\mu_2^2 + \mu_4}{18\mu_2^3 + 12\mu_3^2 - 10\mu_2\mu_4} \mu_3 \\ b_2 = \frac{6\mu_2^3 + 3\mu_3^2 - 2\mu_2\mu_4}{18\mu_2^3 + 12\mu_3^2 - 10\mu_2\mu_4} \end{cases}$$

uma vez que da assunção $\mu'_1 = 0$ resulta que $\mu_r = \mu'_r$ para $r \geq 1$.

A solução de (94) também pode ser escrita utilizando os coeficientes de assimetria e achatamento:

$$\begin{cases} b_0 = \frac{3\gamma_1^2(X) - 4(\gamma_2(X) + 3)}{18 + 12\gamma_1^2(X) - 10(\gamma_2(X) + 3)} \sigma^2 \\ b_1 = a = -\frac{6 + \gamma_2(X)}{18 + 12\gamma_1^2(X) - 10(\gamma_2(X) + 3)} \gamma_1(X) \sigma \\ b_2 = \frac{6 + 3\gamma_1^2(X) - 2(\gamma_2(X) + 3)}{18 + 12\gamma_1^2(X) - 10(\gamma_2(X) + 3)} \end{cases} \quad (95)$$

Johnson, Kotz e Balakrishnan (1994) fornecem uma panorâmica mais abrangente sobre o tema. Porém, convém notar que a expressão apresentada pelos autores para o coeficiente b_1 contém uma imprecisão. Onde deveria estar $\gamma_1(X)$ aparece $\sqrt{\beta_1}$ em que $\beta_1 = \gamma_1^2(X)$, o que naturalmente não é equivalente. Os autores agrupam as distribuições do sistema de Pearson em sete tipos seguindo a metodologia de Pearson que considera os diferentes polinómios em denominador de (93), nomeadamente, a natureza das suas raízes. Resumidamente estes tipos são ($k = b_1^2 / (4b_0b_2)$):

Tipo	Característica
I	$k < 0$
II	$\gamma_1(X) = 0$ e $\gamma_2(X) < 0$
III	$2\gamma_2(X) - 3\gamma_1^2(X) = 0$
IV	$0 < k < 1$
V	$k = 1$
VI	$k > 1$
VII	$\gamma_1(X) = 0$ e $\gamma_2(X) > 0$

Tabela 21 – Tipos de distribuições do sistema de Pearson

Convém referir que registaram-se já diversas tentativas de estender o sistema (93), nomeadamente, através da substituição do numerador por polinómios de grau dois ou três ou do denominador pelo polinómio $c_0 + c_1x^2 + c_2x^4$. Tal não tem conduzido ou a modelos parcimoniosos ou a modelos mais abrangentes, pelo que o sistema apresentado por Pearson continua actual e com grande aplicabilidade.

Importa mencionar também o papel das expansões assintóticas na aproximação de uma função de distribuição. São exemplo as expansões de Edgeworth (1905) que constituíram um progresso notável em relação às séries de Gram-Charlier e que abordaremos na Secção 4.9.

4.6 Distribuições tipo IV

No que se segue terão um interesse particular as distribuições tipo IV do sistema de Pearson abordadas na Secção 4.5. Como referido anteriormente as funções densidade f que são do tipo IV são as soluções da equação diferencial (93) que verificam $0 < b_1^2/(4b_0b_2) < 1$.

Ora,

$$0 < \frac{b_1^2}{4b_0b_2} < 1 \Leftrightarrow b_1^2 - 4b_0b_2 < 0 \quad (96)$$

o que significa que o polinómio $b_0 + b_1x + b_2x^2$ não admite raízes reais. Logo, podemos identificar esse polinómio com

$$B_0 + b_2(x + B_1)^2$$

onde $B_0 = b_0 - \frac{b_1^2}{4b_2}$ e $B_1 = \frac{b_1}{2b_2}$. Substituindo em (93) fica

$$\begin{aligned} (\ln f_X(x))' &= -\frac{x + a}{B_0 + b_2(x + B_1)^2} \\ &= -\frac{(x + B_1) + (a - B_1)}{B_0 + b_2(x + B_1)^2} \\ &= -\frac{1}{\sqrt{b_2B_0}} \frac{\sqrt{b_2/B_0}(x + B_1) + \sqrt{b_2/B_0}(a - B_1)}{1 + \left(\frac{x + B_1}{\sqrt{B_0/b_2}}\right)^2} \end{aligned}$$

onde as raízes $\sqrt{b_2B_0}$ e $\sqrt{b_2/B_0}$ estão bem definidas por (96).

Resolvendo em ordem a f obtém-se

$$f(x) = K [B_0 + b_2(x + B_1)^2]^{-(2b_2)^{-1}} \exp\left(-\frac{a - B_1}{\sqrt{b_2B_0}} \arctan \frac{x + B_1}{\sqrt{B_0/b_2}}\right) \quad (97)$$

onde K é uma constante de normalização.

A reparametrização mais usual é obtida da forma que a seguir se resume. Consideremos

$$y = x + B_1$$

e

$$\alpha = \sqrt{\frac{B_0}{b_2}}.$$

Aplicando ao polinómio $b_0 + b_1x + b_2x^2 = B_0 + b_2(x + B_1)^2$ fica

$$b_2(y^2 + \alpha^2).$$

Assim, (93) é equivalente a

$$(\ln f_X(y - B_1))' = -\frac{y - B_1 + a}{b_2(y^2 + \alpha^2)}.$$

Resolvendo a equação diferencial as soluções são da forma

$$f(y - B_1) \propto \exp\left(-\frac{1}{b_2} \int_{\mathbb{R}} \frac{y - B_1 + a}{y^2 + \alpha^2} dy\right).$$

A solução do integral é

$$f(y - B_1) \propto \exp\left(-\frac{1}{2b_2} \ln\left(1 + \frac{y^2}{\alpha^2}\right) + \frac{\ln \alpha}{2b_2} + \frac{b_1 - 2b_2a}{2b_2^2\alpha} \arctan \frac{y}{\alpha} + C\right)$$

onde $C \in \mathbb{R}$. Utilizando as substituições

$$m = \frac{1}{2b_2}$$

e

$$\nu = -\frac{b_1 - 2b_2a}{2b_2^2\alpha}$$

obtemos como solução

$$f(y - B_1) \propto \left(1 + \frac{y^2}{\alpha^2}\right)^{-m} \exp\left(-\nu \arctan \frac{y}{\alpha}\right).$$

A solução, devidamente normalizada e com um parâmetro de escala λ é do tipo ($\nu \neq 0$)

$$f(x) = \frac{\left|\frac{\Gamma\left(m + \frac{\nu i}{2}\right)}{\Gamma(m)}\right|^2}{\alpha \text{Beta}(m - 0.5, 0.5)} \left(1 + \left(\frac{x - \lambda}{\alpha}\right)^2\right)^{-m} \exp\left(-\nu \arctan \frac{x - \lambda}{\alpha}\right) \quad (98)$$

onde $\text{Beta}(a, b)$ é a função beta definida como ($a, b > 0$)

$$\text{Beta}(a, b) = \int_0^1 x^{a-1} (1-t)^{b-1} dx = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$$

em que $\Gamma(\cdot)$ é a função gama. Para (98) estar bem definido é necessário assumir que $m > 1/2$. Habitualmente assume-se igualmente que $\alpha > 0$ pois a função anterior permanece invariante quando se troca simultaneamente o sinal de α e de ν (Heinrich, 2004). O caso $\nu = 0$ corresponde às distribuições de tipo VII. Nesse caso, obtemos a distribuição t-Student com m graus de liberdade.

Nenhuma das distribuições habitualmente utilizadas em Estatística é do tipo IV. Uma das dificuldades em utilizar uma distribuição de tipo IV é a inexistência de uma expressão analítica fechada para a função de distribuição. Vários investigadores têm-se dedicado à obtenção de aproximações à distribuição tipo IV, nomeadamente, Bouver (1973) que determinou diversos valores de funções de distribuição de tipo IV. Davis e Stephens (1983) construíram um algoritmo para determinação de alguns quantis da distribuição para diversos valores de assimetria e achatamento que pode ser aplicado a qualquer um dos tipos de distribuições do sistema de Pearson.

Woodward (1976) e Skates (1993) determinaram expressões aproximadas para uma função de distribuição tipo IV com um comportamento particularmente interessante ao nível das caudas. No primeiro caso, o resultado obtido a partir da função densidade (98) é:

$$F(x) \simeq 1 - k_1 \alpha^2 \left(1 + \left(\frac{x - \lambda}{\alpha} \right)^2 \right)^{-m+1} \frac{\exp(-\nu \arctan((x - \lambda)/\alpha))}{2m(x - \lambda) + \alpha\nu}$$

onde k_1 é uma constante de normalização.

Skates (1993) utiliza a distribuição gaussiana padrão Φ , para obter várias aproximações para uma função de distribuição de tipo IV. Mais recentemente, a partir da função densidade (98), Heinrich (2004) apresenta uma expressão exacta para a respectiva função de distribuição que depende da função confluyente hipergeométrica $H(a, b, c, z)$

$$F(y) = \frac{K\alpha(1+y^2)^{-m}}{2m-1} \exp(-\nu \arctan y) (i-y) H\left(1, \frac{m+\nu i}{2}, 2m, \frac{1}{1-iy}\right),$$

onde K é uma constante de normalização e $y = (x - \lambda) / \alpha$. O autor discute nesse artigo a convergência da série inerente ao recurso da função hipergeométrica bem a utilização da fórmula fora da região de convergência.

Stuart e Ord (1994) analisam os momentos das distribuições tipo IV. O valor médio de uma distribuição tipo IV existe para $m > 1$ e é dado por

$$\mu'_1 = \lambda - \frac{\alpha\nu}{2(m-1)}.$$

Os momentos centrados de ordem 2, 3 e 4 são dados por

$$\begin{aligned}\mu_2 &= \frac{\alpha^2}{r^2(r-1)}(r^2 + \nu^2), \quad m > 3/2 \\ \mu_3 &= -\frac{4\alpha^3\nu(r^2 + \nu^2)}{r^3(r-1)(r-2)}, \quad m > 2 \\ \mu_4 &= \frac{3\alpha^4(r^2 + \nu^2)[(r+6)(r^2 + \nu^2) - 8r^2]}{r^4(r-1)(r-2)(r-3)}, \quad m > 5/2\end{aligned}$$

onde $r = 2(m-1)$.

Os autores também apresentam uma expressão de recorrência para os momentos centrados de uma distribuição tipo IV. A fórmula de recorrência é dada por ($\mu_0 = 1$ e $\mu_1 = 0$):

$$\mu_n = \frac{\alpha(n-1)}{r^2[r-(n-1)]} [-2\nu r \mu_{n-1} + a(r^2 + \nu^2) \mu_{n-2}], \quad n \geq 2.$$

4.7 Distribuição de T_n

As aproximações obtidas para os quatro primeiros momentos de T_n permitem utilizar agora o sistema de Pearson para aproximar a distribuição de T_n . Essas aproximações dependem apenas do coeficiente de assimetria $\gamma_1(X)$ e do coeficiente de achatamento $\gamma_2(X)$ da população parente e da dimensão amostral n . Considerando diversos valores para essas quantidades podemos obter estimativas para o

coeficiente de assimetria $\gamma_1(T_{n,X})$ e coeficiente de achatamento $\gamma_2(T_{n,X})$ da estatística de Student T_n para diferentes valores de n utilizando os resultados da Secção 4.4. Atendendo a (95) podemos avaliar o valor de $k = b_1^2 / (4b_0b_2)$ a partir das estimativas para $\gamma_1(T_{n,X})$ e $\gamma_2(T_{n,X})$.

Os resultados das simulações indicam que o valor n não é importante na determinação do tipo de distribuição do sistema de Pearson que melhor se ajusta à distribuição de T_n . Também se verifica que o comportamento de T_n é aproximadamente do tipo IV para um grande número de distribuições parente, nomeadamente, para as distribuições mais utilizadas em Estatística. De forma grosseira podemos afirmar que a aproximação aplica-se a distribuições parente com assimetria, em módulo, e achatamento não superior a três. Uma aproximação dessa região encontra-se delimitada no gráfico que se segue:

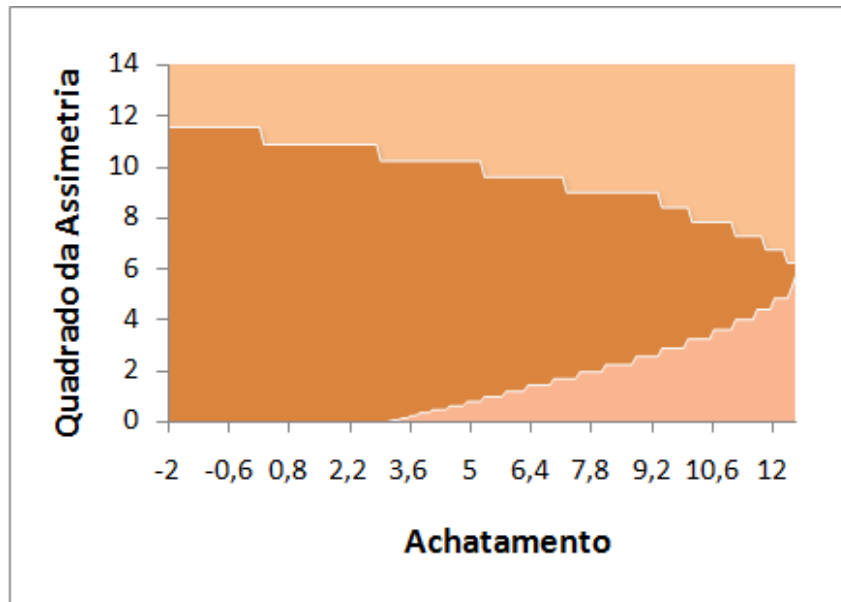


Gráfico 9 – Região onde é aplicável a aproximação por uma distribuição tipo IV

Comparamos agora, para diversas distribuições, as estimativas apresentadas, as estimativas obtidas a partir da distribuição $t_{(n-1)}$ e os valores obtidos por simulação

para os quantis 0.90, 0.95, 0.975 e 0.99 de T_n . Para a simulação de quantis de probabilidade χ_p de uma estatística T a partir de uma amostra simulada x_1, \dots, x_N sabe-se que (Devroye (1986))

$$P(\chi_p \in [X_{r:n}, X_{s:n}]) \approx 1 - \alpha$$

onde $r = \lceil Np - \sqrt{Np(1-p)}z_{1-\alpha/2} \rceil$ e $s = \lceil Np + \sqrt{Np(1-p)}z_{1-\alpha/2} \rceil$ em que $[x]$ é a parte inteira de x e $z_{1-\alpha/2}$ é o quantil de probabilidade $1 - \alpha$.

Para amostras de dimensão $N = 10000$ de variáveis aleatórias contínuas, os quantis 0.90, 0.95, 0.975 e 0.99 encontram-se com aproximadamente 95% de confiança, respectivamente, nos intervalos

$$[T_{8941:10000}, T_{9058:10000}]$$

$$[T_{9457:10000}, T_{9542:10000}]$$

$$[T_{9719:10000}, T_{9780:10000}]$$

$$[T_{9880:10000}, T_{9919:10000}]$$

Para encontrar o membro da família de distribuições tipo IV que verifica as estimativas obtidas para os quatro primeiros momentos de T_n utilizaremos as relações apresentadas por Stuart e Ord (1994) entre os parâmetros apresentados em (98) e o valor médio μ , a variância μ_2 , o coeficiente de assimetria $\gamma_1(X)$ e o coeficiente de achatamento $\gamma_2(X)$ da variável aleatória cuja distribuição se pretende aproximar:

$$m = \frac{3(\gamma_2(X) - \gamma_1^2(X) - 1)}{2\gamma_2(X) - 3\gamma_1^2(X) - 6} + 1,$$

$$\nu = -\frac{2(m-1)(m-2)\gamma_1(X)}{\sqrt{4(2m-3) - \gamma_1^2(X)(m-2)^2}},$$

$$\alpha = \frac{\sqrt{\mu_2(4(2m-3) - \gamma_1^2(X)(m-2)^2)}}{2}$$

e

$$\lambda = \mu - \frac{(m-2)\gamma_1(X)\sqrt{\mu_2}}{2}.$$

No caso dos momentos populacionais não serem conhecidos, é possível obter estimativas utilizando o método dos momentos. A utilização do método da máxima verossimilhança fica dependente da maximização da função de verossimilhança

$$-m \sum_{i=1}^n \ln \left(1 + \left(\frac{x_i - \lambda}{\alpha} \right)^2 \right) - \nu \sum_{i=1}^n \arctan \left(\frac{x_i - \lambda}{\alpha} \right) + nk \quad (99)$$

onde n é a dimensão da amostra x_1, \dots, x_n e k é a constante de normalização

$$\frac{\left| \frac{\Gamma \left(m + \frac{\nu}{2} i \right)}{\Gamma(m)} \right|^2}{\alpha \text{Beta}(m-0.5, 0.5)}.$$

A maximização de (99) é apenas possível numericamente.

As tabelas que se seguem apresentam os resultados obtidos. L_i e U_i representam, respectivamente, o limite inferior e o limite superior do intervalo de 95% de confiança para o quantil i de T_n tendo em conta a distribuição parente apresentada.

X = Y - 1 onde Y ~ Poisson(1) e n = 25					
Quantis	Estim. $t_{(n-1)}$	Estim. tipo IV	Simul.	L_i	U_i
0.90	1.3178	1.220	1.1836	1.1407	1.1932
0.95	1.7109	1.578	1.5541	1.4964	1.5718
0.975	2.0639	1.893	1.8516	1.8091	1.9005
0.99	2.4922	2.271	2.1768	2.1158	2.2433

$X \sim \text{Binomial}(10, 0.5)$ e $n = 20$					
Quantis	Estim. $t_{(n-1)}$	Estim. tipo IV	Simul.	L	U
0.90	1.3277	1.325	1.3366	1.3084	1.3708
0.95	1.7291	1.719	1.7097	1.6743	1.7562
0.975	2.0930	2.072	2.0959	2.0366	2.1213
0.99	2.5395	2.500	2.4846	2.4291	2.6020

$X \sim \text{Binomial}(10, 0.1)$ e $n = 20$					
Quantis	Estim. $t_{(n-1)}$	Estim. tipo IV	Simul.	L_i	U_i
0.90	1.3277	1.232	1.1888	1.1567	1.2280
0.95	1.7291	1.597	1.5519	1.5062	1.5841
0.975	2.0930	1.920	1.8718	1.7983	1.9165
0.99	2.5395	2.306	2.2380	2.1794	2.3333

$X = -Y$ onde $Y \sim \text{Exp}(1)$ e $n = 25$					
Quantis	Estim. $t_{(n-1)}$	Estim. tipo IV	Simul.	L	U
0.90	1.3163	1.680	1.7159	1.6727	1.7779
0.95	1.7081	2.191	2.3147	2.2394	2.3770
0.975	2.0595	2.678	2.8962	2.7972	3.0109
0.99	2.4851	3.304	3.7374	3.5290	3.9473
$n = 400$					
0.90	1.284	1.361	1.3608	1.3242	1.3945
0.95	1.649	1.764	1.7507	1.7137	1.7820
0.975	1.966	2.123	2.0743	2.0171	2.1204
0.99	2.336	2.552	2.4917	2.4031	2.6000

$X \sim U(0,1)$ e $n = 25$					
Quantis	Estim. $t_{(n-1)}$	Estim. tipo IV	Simul.	L_i	U_i
0.90	1.3163	1.311	1.2991	1.2675	1.3361
0.95	1.7081	1.702	1.7203	1.6782	1.7768
0.975	2.0595	2.054	2.0869	2.0256	2.1543
0.99	2.4851	2.480	2.5456	2.4491	2.6205
$n = 36$					
0.90	1.3062	1.302	1.2742	1.2369	1.3081
0.95	1.6896	1.685	1.6622	1.6130	1.7098
0.975	2.0301	2.027	1.9722	1.9204	2.0366
0.99	2.4377	2.436	2.3472	2.2990	2.4057

$X \sim \text{Beta}(1/3, 2/3)$ e $n = 20$					
Quantis	Estim. $t_{(n-1)}$	Estim. tipo IV	Simul.	L	U
0.90	1.3277	1.240	1.1926	1.1563	1.2271
0.95	1.7291	1.611	1.5490	1.5083	1.5907
0.975	2.0930	1.942	1.8673	1.8100	1.9164
0.99	2.5395	2.343	2.2738	2.1878	2.3962
$n = 25$					
0.90	1.3178	1.238	1.2136	1.1864	1.2468
0.95	1.7109	1.602	1.5750	1.5397	1.6126
0.975	2.0639	1.926	1.9047	1.8409	1.9538
0.99	2.4922	2.314	2.2978	2.2361	2.4170

$X \sim \text{Beta}(2/3, 1/3)$ e $n = 25$					
Quantis	Estim. $t_{(n-1)}$	Estim. tipo IV	Simul.	L	U
0.90	1.3178	1.410	1.410	1.3732	1.4530
0.95	1.7109	1.844	1.8722	1.8155	1.9374
0.975	2.0639	2.244	2.3214	2.2483	2.3935
0.99	2.4922	2.742	2.8565	2.7423	3.0640
$n = 81$					
0.90	1.2922	1.345	1.3382	1.3099	1.3726
0.95	1.6641	1.744	1.7142	1.664	1.7516
0.975	1.9901	2.101	2.0472	1.9963	2.1176
0.99	2.3739	2.530	2.4889	2.388	2.5509

$X \sim \text{Beta}(1/2, 1/2)$ e $n = 20$					
Quantis	Estim. $t_{(n-1)}$	Estim. tipo IV	Simul.	L_i	U_i
0.90	1.3277	1.318	1.3005	1.2535	1.3416
0.95	1.7291	1.716	1.6938	1.6575	1.7386
0.975	2.0930	2.076	2.0506	1.9956	2.1286
0.99	2.5395	2.518	2.5419	2.4292	2.6318
$n = 25$					
0.90	1.3178	1.310	1.3156	1.2807	1.3479
0.95	1.7109	1.702	1.7228	1.6656	1.7759
0.975	2.0639	2.055	2.0798	2.0160	2.1394
0.99	2.4922	2.484	2.5638	2.4631	2.7564

$X \sim \text{Beta}(1/4, 3/4)$ e $n = 20$					
Quantis	Estim. $t_{(n-1)}$	Estim. tipo IV	Simul.	L_i	U_i
0.90	1.3277	1.214	1.1733	1.1398	1.1960
0.95	1.7291	1.585	1.4908	1.4616	1.5392
0.975	2.0930	1.916	1.7941	1.7479	1.8344
0.99	2.5395	2.317	2.0976	2.0530	2.2031
$n = 25$					
0.90	1.3178	1.210	1.1894	1.1593	1.2240
0.95	1.7109	1.571	1.5060	1.4631	1.5432
0.975	2.0639	1.891	1.7814	1.7332	1.8241
0.99	2.4922	2.274	2.1117	2.0387	2.1938
$n = 30$					
0.90	1.3114	1.209	1.1681	1.1384	1.2005
0.95	1.6991	1.564	1.4863	1.4521	1.5214
0.975	2.0452	1.876	1.7735	1.7246	1.8220
0.99	2.4620	2.25	2.1733	2.1011	2.2300
$n = 36$					
0.90	1.3062	1.208	1.1817	1.1463	1.2187
0.95	1.6896	1.569	1.4992	1.4698	1.5338
0.975	2.0301	1.866	1.7603	1.7145	1.8175
0.99	2.4377	2.231	2.1045	2.0402	2.1831
$n = 49$					
0.90	1.299	1.210	1.2067	1.1789	1.2378
0.95	1.677	1.556	1.5393	1.4993	1.5750
0.975	2.011	1.857	1.8028	1.7758	1.8519
0.99	2.407	2.210	2.1616	2.1012	2.2499

$X \sim \text{Laplace}(1)$ e $n = 25$					
Quantis	Estim. $t_{(n-1)}$	Estim. tipo IV	Simul.	L_i	U_i
0.90	1.3178	1.332	1.3069	1.2737	1.3423
0.95	1.7109	1.709	1.6744	1.6354	1.7177
0.975	2.0639	2.037	1.9901	1.9551	2.0549
0.99	2.4922	2.418	2.4288	2.3400	2.4789

$X \sim \chi^2_{(1)}$ e $n = 20$					
Quantis	Estim. $t_{(n-1)}$	Estim. tipo IV	Simul.	L_i	U_i
0.90	1.3277	1.317	1.0437	1.0207	1.0653
0.95	1.7291	1.805	1.2865	1.2598	1.3177
0.975	2.0930	2.256	1.5301	1.4941	1.5697
0.99	2.5395	2.805	1.8141	1.7725	1.8747
$n = 81$					
0.90	1.2922	1.173	1.1123	1.0855	1.1407
0.95	1.6641	1.517	1.4102	1.3791	1.4467
0.975	1.9901	1.815	1.6436	1.6137	1.6828
0.99	2.3739	2.164	1.9620	1.9186	2.0174
$n = 400$					
0.90	1.284	1.200	1.1639	1.1325	1.1894
0.95	1.649	1.529	1.4792	1.4391	1.5156
0.975	1.966	1.809	1.7595	1.7201	1.7944
0.99	2.336	2.127	2.0558	2.0094	2.1232

Tabelas 22 a 33 – Estimativas para os quantis de T_n para diversas distribuições

Designando por e_i as estimativas obtidas para os quantis de T_n podemos medir a qualidade dessas estimativas utilizando as quantidades

$$d_i = \begin{cases} \frac{L_i - e_i}{U_i - L_i}, & |e_i - L_i| \leq |e_i - U_i| \\ \frac{e_i - U_i}{U_i - L_i}, & |e_i - L_i| > |e_i - U_i| \end{cases}$$

onde $i = 0.90, 0.95, 0.975, 0.99$. Para as distribuições anteriores a média das quantidades d_i no caso da aproximação pela distribuição $t_{(n-1)}$ é 1.43. No caso da utilização de uma distribuição tipo IV a média é 0.42. Como esperado, o comportamento das distribuições tipo IV revela-se muito melhor na aproximação aos quantis mais elevados de T_n .

4.8 Gaussiana generalizada, uma alternativa?

Uma alternativa à utilização do sistema de distribuições de Pearson poderia ser o recurso a uma família conhecida de distribuições da qual a distribuição gaussiana fosse um membro. Esta metodologia é natural na medida em que o Teorema Limite Central conjuntamente com o teorema de Slutsky garantem a convergência de T_n para a gaussiana, quando $n \rightarrow \infty$, na generalidade das situações.

Tendo por base este raciocínio uma candidata natural é a família gaussiana generalizada que se define da forma que se segue.

Definição 4.7 *Considere-se uma variável aleatória $X_{\beta,\mu,\sigma}$ com função densidade*

$$f(x|\mu, \sigma, \beta) = \frac{\exp\left[-\frac{1}{2}\left|\frac{x - \mu}{\sigma}\right|^{\frac{2}{\beta+1}}\right]}{\frac{\beta + 3}{2} \frac{\sigma}{2} \Gamma\left(\frac{\beta + 3}{2}\right)} I_{\mathbb{R}} \quad (100)$$

em que $\beta > -1, \mu \in \mathbb{R}$ e $\sigma > 0$. $X_{\beta, \mu, \sigma}$ pertence à **família gaussiana generalizada** se f é a sua função densidade de probabilidade. Nesse caso, utiliza-se a notação $X_{\beta, \mu, \sigma} \sim \text{Gau}_{\beta}(\mu, \sigma)$.

Em particular, esta família engloba a distribuição gaussiana ($\beta = 0$) e a distribuição Laplace ($\beta = 1$). Esta família é infinitamente divisível pelo menos para $\beta \in [0, 1]$ como mostramos a seguir.

Lema 4.1 *A versão simetrizada X^* de uma variável aleatória X com suporte contido em \mathbb{R}_0^+ tem função característica $\psi = \text{Re}(\varphi_X)$.*

Demonstração Seja ψ a função característica da versão simetrizada X^* de uma variável aleatória X então

$$\begin{aligned}\psi(t) &= \text{E}(\exp(itX^*)) \\ &= \text{E}[\cos(tX^*) + i \sin(tX^*)].\end{aligned}$$

Como X^* é simétrica tem-se que

$$\begin{aligned}\psi(t) &= \text{E}(\cos(tX^*)) \\ &= \int_{-\infty}^0 \cos(tx) \frac{dF_X(x)}{2} + \int_0^{+\infty} \cos(tx) \frac{dF_X(x)}{2} \\ &= \int_0^{+\infty} \cos(-tx) \frac{dF_X(-x)}{2} + \int_0^{+\infty} \cos(tx) \frac{dF_X(x)}{2} \\ &= \int_0^{+\infty} \cos(tx) dF_X(x) \\ &= \text{Re}(\varphi_X(t)).\end{aligned}$$

como queríamos mostrar.

■

Teorema 4.7 *A família de funções (100) é infinitamente divisível para $\beta \in [0, 1]$.*

Demonstração A função

$$f(x) = \exp(-x) I_{\mathbb{R}_0^+}(x)$$

é completamente monótona pois

$$f^{(n)}(x) = (-1)^{n+1} \exp(-x) I_{\mathbb{R}_0^+}(x).$$

Dado o parâmetro $\beta \in (0, 1)$, a função

$$g(x) = x^\beta I_{\mathbb{R}_0^+}(x)$$

é não negativa e tem derivada

$$h(x) = g'(x) = \beta x^{\beta-1} I_{\mathbb{R}_0^+}(x)$$

que é completamente monótona pois

$$h^{(n)}(x) = \beta(\beta-1)\dots(\beta-n)x^{\beta-n-1} I_{\mathbb{R}_0^+}(x).$$

Logo, pelo Teorema 4.5, $l = f \circ h$ é completamente monótona. Pelo Teorema 4.3, l é da forma

$$l(x) = \int_0^\infty \exp(-xu) dF(u).$$

em que F é uma função não decrescente. Portanto l é, a menos de uma constante multiplicativa, a transformada de Laplace de uma variável aleatória X . A função característica dessa variável aleatória é, ignorando a constante multiplicativa, dada por

$$\varphi_X(t) = L_X(-it) = \int_0^\infty \exp(itu) dF(u).$$

Se φ for infinitamente divisível pela Definição 4.6 existe, para cada $n \in \mathbb{N}$, uma função característica φ_n tal que

$$\varphi = \varphi_n^n.$$

Pelo Lema 4.1, a parte real da função característica de uma variável aleatória com suporte contido em \mathbb{R}_0^+ é também uma função característica (da correspondente variável aleatória simetrizada). Além disso, prova-se que ψ é infinitamente divisível se φ for infinitamente divisível. Aplicando estes resultados a uma variável aleatória $X_{\mu,\sigma,\beta}$ com função densidade

$$f(x|\mu, \sigma, \beta) = \frac{\exp\left[-\frac{1}{2}\left|\frac{x-\mu}{\sigma}\right|^{\frac{2}{\beta+1}}\right]}{\frac{\beta+3}{2} \frac{\sigma}{2} \Gamma\left(\frac{\beta+3}{2}\right)} I_{\mathbb{R}}(x)$$

com $\beta > -1$, $\mu \in \mathbb{R}$ e $\sigma > 0$ conclui-se que, para $\beta \in (0, 1)$, X é infinitamente divisível. ■

O primeiro momento ordinário de $X_{\beta,\mu,\sigma}$ corresponde ao parâmetro μ como se verifica a seguir.

$$\begin{aligned} E(X_{\beta,\mu,\sigma}) &= \int_{\mathbb{R}} x \frac{\exp\left[-\frac{1}{2}\left|\frac{x-\mu}{\sigma}\right|^{\frac{2}{\beta+1}}\right]}{\frac{\beta+3}{2} \frac{\sigma}{2} \Gamma\left(\frac{\beta+3}{2}\right)} dx \\ &= \int_{-\infty}^{\mu} x \frac{\exp\left[-\frac{1}{2}\left(-\frac{x-\mu}{\sigma}\right)^{\frac{2}{\beta+1}}\right]}{\frac{\beta+3}{2} \frac{\sigma}{2} \Gamma\left(\frac{\beta+3}{2}\right)} dx + \int_{\mu}^{+\infty} x \frac{\exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^{\frac{2}{\beta+1}}\right]}{\frac{\beta+3}{2} \frac{\sigma}{2} \Gamma\left(\frac{\beta+3}{2}\right)} dx \end{aligned}$$

Efectuando as mudanças de variável $y = \frac{1}{2}\left(-\frac{x-\mu}{\sigma}\right)^{\frac{2}{\beta+1}}$ e $y = \frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^{\frac{2}{\beta+1}}$ no primeiro e segundo integral, respectivamente, obtém-se:

$$\begin{aligned} E(X_{\beta,\mu,\sigma}) &= \mu + \frac{\frac{\beta+1}{2}}{\Gamma\left(\frac{\beta+3}{2}\right)} \left\{ \int_{+\infty}^0 y \frac{\beta-1}{2} \exp(-y) dy + \int_0^{+\infty} y \frac{\beta-1}{2} \exp(-y) dy \right\} \\ &= \mu. \end{aligned}$$

Os momentos centrados em torno de μ de ordem ímpar são todos nulos pois a função densidade é simétrica em torno de μ . Os momentos ordinários de ordem par, quando $\mu = 0$ e $\sigma = 1$, são dados por ($k \in \mathbb{N}$):

$$E(X_{\beta,0,1}^{2k}) = \int_{\mathbb{R}} x^{2k} \frac{\exp\left[-\frac{1}{2}|x|^{\frac{2}{\beta+1}}\right]}{2^{\frac{\beta+3}{2}} \Gamma\left(\frac{\beta+3}{2}\right)} dx.$$

Como a função integranda é par

$$E(X_{\beta,0,1}^{2k}) = 2 \int_0^{+\infty} x^{2k} \frac{\exp\left[-\frac{1}{2}x^{\frac{2}{\beta+1}}\right]}{2^{\frac{\beta+3}{2}} \Gamma\left(\frac{\beta+3}{2}\right)} dx$$

Efectuando a mudança de variável $y = \frac{1}{2}x^{\frac{2}{1+\beta}}$

$$\begin{aligned} E(X_{\beta,0,1}^{2k}) &= 2 \int_0^{+\infty} (2y)^{k(\beta+1)} \frac{\exp(-y)}{2^{\frac{\beta+3}{2}} \Gamma\left(\frac{\beta+3}{2}\right)} 2^{\frac{\beta-1}{2}} (1+\beta) y^{\frac{\beta-1}{2}} dy \\ &= \frac{2 \times 2^{k(\beta+1)} \times 2^{\frac{\beta-1}{2}} (\beta+1)}{2^{\frac{\beta+3}{2}} \Gamma\left(\frac{\beta+3}{2}\right)} \int_0^{+\infty} y^{k(1+\beta)} \exp(-y) y^{\frac{\beta-1}{2}} dy \end{aligned}$$

$$\begin{aligned}
 &= \frac{2^{k(\beta+1)} \times (\beta+1)}{2\Gamma\left(\frac{\beta+3}{2}\right)} \int_0^{+\infty} y^{k(\beta+1)+\frac{\beta-1}{2}} \exp(-y) dy \\
 &= 2^{k(\beta+1)} \frac{\Gamma\left(k(\beta+1) + \frac{\beta+1}{2}\right)}{\Gamma\left(\frac{\beta+1}{2}\right)} \\
 &= 2^{k(\beta+1)} \frac{\Gamma\left(\left(k + \frac{1}{2}\right)\beta + 1\right)}{\Gamma\left(\frac{\beta+1}{2}\right)}
 \end{aligned}$$

Assim,

$$\text{var}(X_{\beta,\mu,\sigma}) = \sigma^2 \text{var}(X_{\beta,0,1}) = \sigma^2 2^{\beta+1} \frac{\Gamma\left(\frac{3(\beta+1)}{2}\right)}{\Gamma\left(\frac{\beta+1}{2}\right)}.$$

e

$$\begin{aligned}
 \gamma_2(X_{\beta,\mu,\sigma}) &= \frac{\mathbb{E}\left[(X_{\beta,\mu,\sigma} - \mu_{X_{\beta,\mu,\sigma}})^4\right]}{\sigma_{X_{\beta,\mu,\sigma}}^4} - 3 = \frac{\mathbb{E}\left[(\sigma X_{\beta,0,1} + \mu - \mu)^4\right]}{[\mathbb{E}(X_{\beta,\mu,\sigma} - \mu)^2]^2} - 3 \\
 &= \frac{\sigma^4 \mathbb{E}(X_{\beta,0,1}^4)}{[\mathbb{E}(\sigma X_{\beta,0,1} + \mu - \mu)^2]^2} - 3 = \frac{\mathbb{E}(X_{\beta,0,1}^4)}{[\mathbb{E}(X_{\beta,0,1}^2)]^2} - 3 \\
 &= \frac{\Gamma\left(2(\beta+1) + \frac{\beta+1}{2}\right)}{\left[\Gamma\left(\beta+1 + \frac{\beta+1}{2}\right)\right]^2} - 3 \\
 &\quad \frac{\Gamma\left(\frac{\beta+1}{2}\right)}{\Gamma\left(\frac{\beta+1}{2}\right)} \\
 &= \frac{\Gamma\left(\frac{\beta+1}{2}\right) \Gamma\left(5\frac{\beta+1}{2}\right)}{\left[\Gamma\left(3\frac{\beta+1}{2}\right)\right]^2} - 3.
 \end{aligned}$$

A aplicação desta família no contexto da aproximação da distribuição da estatística de Student T_n fica limitada pelo facto de todos os membros da família terem

função densidade simétrica em torno de μ . Também as expressões dos momentos de ordem dois e quatro estão muito relacionadas o que impede a desejável variabilidade em termos de valores possíveis para o par variância e achatamento.

Embora sejam relevantes as desvantagens apontadas, há que notar que o achatamento dos membros da família gaussiana generalizada está compreendido entre $] -1.2, +\infty[$ o que engloba um largo espectro de valores que o achatamento pode tomar.

4.9 Expansões de Edgeworth

A série de Edgeworth permite escrever a função densidade de probabilidade de uma variável contínua g à custa de uma função densidade conhecida f . Geralmente, a função f utilizada é a função densidade da Gaussiana padrão Φ . No nosso caso, iremos optar por utilizar uma densidade tipo IV.

Consideremos as funções características φ e ψ das distribuições cujas funções densidade são dadas por f e g , respectivamente. Da definição de cumulantes resulta a identidade

$$\psi(t) = \exp \left[\sum_{r=1}^{\infty} (k_r - \gamma_r) \frac{(it)^r}{r!} \right] \varphi(t). \quad (101)$$

A expansão da função exponencial anterior em série de Taylor permite obter expansões para a função característica ψ com base na função característica (conhecida) φ .

A ordenação dos termos da expansão de acordo com os sucessivos termos da expansão em série de Taylor conduz a uma **série de Gram-Charlier**. A assunção de que a diferença entre os cumulantes de ordem r é da ordem

$$k_r - \gamma_r = O(n^{1-r/2}), \quad r \geq 2,$$

permite ordenar as ordens de acordo com a sua relevância e assim obter a chamada **expansão de Edgeworth**. Partindo de uma condição semelhante a (101)

$$g(t) = \exp \left[\sum_{r=1}^{\infty} (k_r - \gamma_r) \left\{ \frac{(-D)^r}{r!} \right\} \right] f(t) \quad (102)$$

obtêm-se expansões semelhantes para a função densidade/função de distribuição a partir de uma função densidade/função de distribuição conhecida. O operador

$$\exp \left[\sum_{r=1}^{\infty} (k_r - \gamma_r) \left\{ \frac{(-D)^r}{r!} \right\} \right]$$

deve ser encarado como a expansão

$$\sum_{i=0}^{\infty} \frac{\left[\sum_{r=1}^{\infty} (k_r - \gamma_r) \left\{ \frac{(-D)^r}{r!} \right\} \right]^i}{i!}$$

aplicado a f onde $D^j f = \frac{\partial^j f}{\partial x^j}$ (Johnson, Kotz e Balakrishnan (1994, p. 26)).

A expansão de Edgeworth para a densidade g_X obtida a partir de (102), sendo a função densidade f uma função densidade tipo IV, f_{IV} , de uma variável aleatória estandardizada no sentido habitual (com valor médio nulo e desvio padrão unitário), é, a menos de um termo de ordem $O(n^{-3/2})$

$$g_X(x) = f_{IV}(x) - \frac{1}{6}\gamma_1(X) F_{IV}^{(3)}(x) + \frac{1}{24}\gamma_2(X) F_{IV}^{(4)}(x) + \frac{\gamma_1^2(X)}{72} F_{IV}^{(6)}(x). \quad (103)$$

Para a função distribuição a expansão é, a menos de um termo de ordem $O(n^{-3/2})$

$$\int_{-\infty}^x g_X(t) dt = F_{IV}(x) - \frac{1}{6}\gamma_1(X) F_{IV}^{(2)}(x) + \frac{1}{24}\gamma_2(X) F_{IV}^{(3)}(x) + \frac{\gamma_1^2(X)}{72} F_{IV}^{(5)}(x).$$

Para obter estimativas para os quantis de T_n com base na aproximação anterior necessitamos de conhecer as derivadas de uma função densidade tipo IV. Uma alternativa é calcular os polinómios ortogonais associados à densidade.

Caso estes formem uma sucessão de Appell (família que inclui, entre outros, os polinómios de Hermite, os de Bernoulli e os de Euler), isto é, sejam uma sucessão de polinómios $\{p_n\}_{n \in \mathbb{N}}$ tal que

$$\frac{d}{dx} p_n(x) = n p_{n-1}(x)$$

as derivadas estão directamente a eles associadas (Chihara, 1978).

Alguns cálculos permitem obter os primeiros polinómios ortogonais com respeito a uma densidade de tipo IV para a parametrização (98) tal que o valor médio da respectiva variável aleatória é zero ($r = 2(m - 1)$).

$$Q_0(x) = 1$$

$$Q_1(x) = x$$

$$Q_2(x) = x^2 + 4\alpha \frac{\nu}{(r-2)r} x - \alpha^2 \frac{r^2 + \nu^2}{r^2(r-1)}$$

$$Q_3(x) = x^3 + 12\alpha \frac{\nu}{(r-4)r} x^2 - 3 \frac{(r^3 - 4r^2 + r\nu^2 - 12\nu^2)\alpha^2}{(r^2 - 7r + 12)r^2} x - 8\alpha^3 \nu \frac{r^2 + \nu^2}{(r^2 - 6r + 8)r^3}$$

$$Q_4(x) = x^4 + 24 \frac{\alpha\nu}{r(r-6)} x^3 + 6\alpha^2 \frac{6r^2 - r^3 + 30\nu^2 - r\nu^2}{r^2(30 - 11r + r^2)} x^2$$

$$+ 8\alpha^3 \frac{60\nu^3 + 36r^2\nu - 7r\nu^3 - 7r^3\nu}{r^3(74r - 120 - 15r^2 + r^3)} x$$

$$+ 3\alpha^4 \frac{r^5 - 6r^4 - 30\nu^4 + r\nu^4 - 36r^2\nu^2 + 2r^3\nu^2}{r^4(63r - 90 - 14r^2 + r^3)}$$

Se consideramos a variável estandardizada no sentido habitual obtemos polinómios com coeficientes com uma aparência ainda menos agradável. Tal acontece pois, nesse caso, a relação entre α e as constantes r (ou m) e ν é não linear:

$$\alpha = \sqrt{\frac{r^2(r-1)}{(r^2 + \nu^2)}}$$

E, a constante α é que determina a variância da distribuição.

Infelizmente, a sucessão dos polinómios ortogonais relativamente a uma função densidade de tipo IV não é uma sucessão de Appell. Assim, a necessidade de avaliar

as derivadas de (103) torna o cálculo de quantis demorado e não conduz a resultados muito úteis.

Exemplo 4.2 *Consideremos uma variável aleatória $Y = X - 1$ onde $X \sim P(\lambda)$. O quadro seguinte apresenta as estimativas dadas por (103) para a função de distribuição $F_{T_{25}}$ nos pontos que constituem os limites inferior (L_i) e superior (U_i) dos intervalos a 95% de confiança obtidos na Secção 4.7 para quatro quantis de T_{25} .*

Quantis	L_i	$F_{T_{25}}$	U_i	$F_{T_{25}}$
0.90	1.1407	0.8862	1.1932	0.9007
0.95	1.4964	0.9641	1.5718	0.9747
0.975	1.8091	0.9969	1.9005	1.0017
0.99	2.1158	1.0075	2.2433	1.0084

Tabela 34 – Alguns valores de $F_{T_{25}}$ obtidos a partir de (103)

Uma vez que lidamos com expressões assintóticas não é surpreendente obter estimativas para valores de uma função de distribuição que não se enquadram no intervalo $[0, 1]$. As estimativas obtidas no exemplo anterior são inferiores em qualidade às estimativas obtidas com o recurso às distribuições tipo IV.

Uma possível alternativa às expansões de Edgeworth são as expansões de Edgeworth diferidas. Nesse caso, a sua utilização é complicada pois envolve a resolução de equações do tipo

$$\exp K(x) = \lambda$$

em ordem a x , onde K é um polinómio de coeficientes reais de grau 4.

4.10 O papel da (as)simetria

É notório, em todo este capítulo, o papel relevante que o coeficiente de assimetria ou mais vulgarmente a assimetria desempenha na aproximação à distribuição e aos quantis mais elevados de T_n . O módulo deste coeficiente revelou-se muito importante na proximidade da distribuição de T_n ao quase inevitável limite gaussiano, desde que verificadas as condições discutidas anteriormente. Do sinal da assimetria depende também a relação entre as distribuições tipo IV e a distribuição $t_{(n-1)}$ como verificamos a seguir.

Utilizando as aproximações obtidas para os quatro primeiros momentos de T_n na Secção 4.4 podemos avaliar a diferença entre uma função de distribuição de tipo IV, F_{IV} , associada a uma variável aleatória estandardizada que verifica esses quatro momentos e a função de distribuição t-Student com $n - 1$ graus de liberdade, G_{n-1} , avaliando a diferença entre as respectivas expansões de Edgeworth de *kernel* gaussiano. Assim, atendendo aos resultados obtidos na Secção 4.4, a função de distribuição de tipo IV que verifica as aproximações obtidas para os quatro primeiros momentos da estatística T_n é igual a

$$\begin{aligned}
 F_{IV}(x) = & -\frac{1}{6} \left(\frac{-\frac{2\gamma_1(X)}{\sqrt{n}}}{\left(1 + 2\frac{\gamma_1^2(X) + 1}{n}\right)^{3/2}} \right) \Phi^{(2)}(x) \\
 & + \frac{1}{24} \left(\frac{3 + \frac{-4\gamma_2(X) + 45\gamma_1^2(X) + 36}{2(n-1)}}{\left(1 + 2\frac{\gamma_1^2(X) + 1}{n}\right)^2} \right) \Phi^{(3)}(x) \\
 & + \frac{\left(-\frac{2\gamma_1(X)}{\sqrt{n}}\right)^2}{\left(1 + 2\frac{\gamma_1^2(X) + 1}{n}\right)^3} \Phi^{(5)}(x) + O(n^{-3/2})
 \end{aligned} \tag{104}$$

Para escrever a expansão de Edgeworth de *kernel* gaussiano de G_{n-1} é necessário recordar que se $X \sim t_{(n)}$ então

$$E(X) = E(X^3) = 0, \quad E(X^2) = \frac{n}{n-2} \text{ e } E(X^4) = \frac{3n^2}{8+n(n-6)}$$

para $n > 4$. Portanto,

$$G_{n-1}(x) = \frac{1}{24} \left(\frac{6(n-2)}{n^2-6n+8} \right) \Phi^{(3)}(x) + O(n^{-3/2})$$

Logo, a diferença entre $F_{IV} - G_{n-1}$ é dada por

$$\begin{aligned} & -\frac{1}{6} \left(-\frac{2\gamma_1(X)}{\sqrt{n}} \right) \Phi^{(2)}(x) \\ & + \frac{1}{24} \left(\frac{-4\gamma_2(X) + 45\gamma_1^2(X) + 36}{2(n-1)} - \frac{6(n-2)}{n^2-6n+8} \right) \Phi^{(3)}(x) \\ & + \frac{\left(-\frac{2\gamma_1(X)}{\sqrt{n}} \right)^2}{72} \Phi^{(5)}(x) + O(n^{-3/2}) \\ & = \frac{\gamma_1(X)}{3\sqrt{n}} \Phi^{(2)}(x) + O(n^{-1}) \end{aligned}$$

onde Φ é a função de distribuição de uma variável aleatória gaussiana padrão.

Se $\gamma_1(X) > 0$ então $F_{IV}(x) - G_{t_{(n-1)}}(x) > 0$ e se $\gamma_1(X) < 0$ então $F_{IV}(x) - G_{t_{(n-1)}}(x) < 0$. Este comportamento dependente da assimetria poderá explicar a diferença em termos de qualidade nas aproximações aos quantis da estatística T_n dadas pelas duas distribuições. Aliás, nos exemplos analisados na Secção 4.7 as duas inequações anteriores foram sempre verificadas.

Um outro resultado interessante que esta investigação desvendou foi a relação entre a assimetria e o conceito de atracção/repulsão apresentado por Pestana e Velosa (2008) na última edição do seu livro.

Definição 4.8 Dadas duas variáveis aleatórias contínuas X e Y , diz-se que X e Y *se atraem* caso

$$f_{(X,Y)}(x, y) > f_X(x) f_Y(y)$$

onde f_X , f_Y são, respectivamente as funções densidade de X e Y e $f_{(X,Y)}$ é a função densidade conjunta de X e Y . Da mesma forma, X e Y *se repelem* caso

$$f_{(X,Y)}(x, y) < f_X(x) f_Y(y).$$

É, nesta altura, interessante recordar o resultado (72):

$$E(\bar{X}_n S_n^2) = \frac{\gamma_1(X) \sigma^3}{n} + \mu \sigma^2.$$

Portanto,

$$\begin{aligned} \text{cov}(\bar{X}_n, S_n^2) &= E(\bar{X}_n S_n^2) - E(\bar{X}_n) E(S_n^2) \\ &= \frac{\gamma_1(X) \sigma^3}{n} + \mu \sigma^2 - \mu \sigma^2 \\ &= \frac{\gamma_1(X) \sigma^3}{n}. \end{aligned} \tag{105}$$

Assim, se $\gamma_1(X) > 0$ então $\text{cov}(\bar{X}_n, S_n^2) > 0$. Além disso, conjecturamos que \bar{X}_n e S_n^2 se atraem. Se $\gamma_1(X) < 0$ então $\text{cov}(\bar{X}_n, S_n^2) < 0$ e conjecturamos que \bar{X}_n e S_n^2 se repelem. Portanto, a diferença entre a função densidade de probabilidade conjunta $f_{(X,Y)}(x, y)$ e o produto das marginais $f_X(x) f_Y(y)$, ou medida dessa diferença, pode permitir avaliar o grau de dependência entre as variáveis aleatórias X e Y .

Note-se que a partir de (105) conclui-se que se uma população for simétrica (não gaussiana) então a média e a variância de uma amostra extraída dessa população são dependentes mas não correlacionadas.

Bibliografia

Anderson, T. (1962). The choice of the degree of a polynomial regression as a multiple decision problem. *Annals of Mathematical Statistics*, Vol. 33, 255-265.

Andreev, A., Kanto, A., Malo, P. (2005). *Simple Approach for Distribution Selection in the Pearson System*, Helsinki School of Economics, Helsinki.

Anscombe, F. (1948). The transformation of poisson, binomial and negative-binomial data. *Biometrika*, Vol. 35, 246-254.

Barndorff-Nielsen, O., and Cox, D. (1989). *Asymptotic Techniques for Use in Statistics*, Chapman and Hall, London.

Bartlett, M. (1947). Square-root transformations in analysis of variance. *Journal of the Royal Statistical Society*, Vol. 3, 68-78.

Beauregard, R. A., and Fraleigh, J. B. (1990). *Linear Algebra*, 2nd ed., Addison-Wesley Publishing Company, New York.

Bishop, Y., Fienberg, S., Holland, P. (1975). *Discrete Multivariate Analysis: Theory and Practice*, MIT Press, Cambridge.

Bouver, H. (1973). *Table of the cumulative standardized Pearson type IV distribution function*, Themis Report N. 28, Technical Report N. 100, Department of Statistics and Computer Science, The University of Georgia, Athens.

Box, G., and Cox, D. (1964). An Analysis of Transformations. *Journal of Royal Statistical Society*, Vol. B26, 211-252.

Casella, G., and Berger, R. L. (2002). *Statistical Inference*, 2nd edition, Duxbury, Pacific Grover.

Chandra, T. (1999). *A First Course in Asymptotic Theory of Statistics*, Narosa, New Delhi.

Chihara, T. (1978). *An Introduction to Orthogonal Polynomials*, Gordon and Breach, New York.

Chistyakov, G., and Götze, F. (2004). Limit distributions of studentized means. *The Annals of Probability*, Vol. 32, 28-77.

Cramér, H. (1946). *Mathematical Methods of Statistics*, Princeton University Press, Princeton.

Courant, R. (1988). *Differential and Integral Calculus*, Vol. 2. Interscience Publishers, New York.

Curtiss, J. (1943). On transformations used in the analysis of variance. *Annals of Mathematical Statistics*, Vol. 14, 107-132.

Dahlquist, G., e Björk, A. (1974). *Numerical Methods*, Prentice Hall, New Jersey.

Darling, D. (1952). The influence of the maximum term in the addition of independent random variables. *Transactions of the American Mathematical Society*, Vol. 73, 95-107.

Darmois, D. (1951). Sur une propriété caractéristique de la loi de probabilité de Laplace. *C. R. Acad. Sci. Paris*, 232, 1999-2000.

David, H., and Edwards, A. W. F. (2001). *Annotated Readings in the History of Statistics*, Springer-Verlag, New York.

Davis, C., and Stephens, M. (1983). Approximate percentage points using Pearson curves. *Applied Statistics*, Vol. 32, 322-327.

Dette, H., Heiligers, B., Studden, W. (1995). Minimax designs in linear regression models, *The Annals of Statistics*, Vol. 23, 30-40.

Dette, H., and Studden, W. (1995). Optimal designs for polynomial regression when the degree of the polynomial is not known. *Statistica Sinica*, Vol. 5, 459-474.

Dette, H., and Studden, W. (1997). *The Theory of Canonical Moments with Applications in Statistics, Probability and Analysis*. Wiley-Interscience Publication, New York.

Devroye, L. (1986). *Non-Uniform Random Variate Generation*, Springer-Verlag, New York.

Devroye, L. (1996). Random variate generation in one line of code. *Proceedings of the 28th Conference on Winter Simulation*, 265-272.

Edgeworth, F. (1905). The law of error. *Proc. Cambridge Philos. Soc.*, Vol. 20, 36-66.

Efron, B. (1969). Students t -test under symmetry conditions. *Journal of American Statistical Society*, Vol. 64, 1278-1302.

Fedorov, V. (1972). *Theory of Optimal Experiments*, Academic Press, New York.

Feller, W. (1971). *An Introduction to Probability Theory and Its Applications*, Vol. 2. Wiley, New York.

Finetti, B. (1930). Le funzione caratteristiche di legge istantanea, *Rend. Ac. Lenci.*, Vol. 12, 278-282.

Fishman, G. (1973). *Concepts and Methods in Discrete Event Digital Simulation*, Wiley-Interscience, New York.

Fishman, G. (2005). *A First Course in Monte Carlo Methods*, Duxbury Press.

Gilbert, N. (1989). *Biometrical Interpretation – Making Sense of Statistics in Biology*, 2nd edition, Oxford University Press, Oxford.

Giné, E., Götze, F., Mason, D. (1997). When is the Student t -statistic asymptotically standard normal?. *The Annals of Probability*, Vol. 25, 1514-1531.

Gnedenko, B. (1940). On the theory of domains of attraction of stable laws. *Uchemye Zapiski, Moskov Gos. Univ.*, Vol. 30, 61-72.

Gnedenko, B., and Kolmogorov, A. (1968). *Limit Distributions for Sums of Independent Random Variables*. Addison-Wesley, Reading.

Gut, A. (1991). *An Intermediate Course in Probability*, Springer, New York.

Hanselman, D., and Littlefield, B. (2005). *Mastering MATLAB 7*, Prentice Hall, Maine.

Heinrich, J. (2004). A guide to the Pearson type IV distribution, www-cdf.fnal.gov/publications/cdf6820_pearson4.pdf.

Hoaglin, D., Mosteller, F., Tukey, J. (1993). *Análise Exploratória de Dados. Técnicas Robustas – Um Guia*, Edições Salamandra, Lisboa.

Holland, P. (1973). Covariance stabilizing transformations. *The Annals of Statistics*, Vol. 14, 84-92.

Johnson, N., Kotz, S., Balakrishnan, N. (1994). *Continuous Univariate Distributions*, Vol. I, 2nd edition, Wiley Interscience Publication, New York.

Karlin, S., and Studden, W. (1966). *Tchebycheff Systems: with Applications in Analysis and Statistics*, Interscience Publishers, New York.

Karlin, S., and Studden, W. (1966). Optimal experimental designs. *The Annals of Mathematical Statistics*, Vol. 37, 783-815.

Kendall, M., and Stuart, A. (1977). *The Advanced Theory of Statistics, Vol. I: Distribution Theory*, 4th edition, Charles Griffin, London.

Khintchine. A. (1937). Deduction nouvelle d'une formule de M. Paul Lévy. *Bull. Univ. d'État Moskou, Sér Internat.*, Sec. A 1, No. 1, 1-5.

Kiefer, J., and Studden, W. (1976). Optimal designs for large degree polynomial regression. *The Annals of Statistics*, Vol. 4, 1113-1123.

Klingenberg, W. (1983). *A Course in Differential Geometry*, Springer-Verlag, New York.

Kraft, O., and Schaefer, M. (1995). Exact Elfving-minimax designs for quadratic regression. *Statistics Sinica*, Vol. 5, 475-485.

Krein, M., and Nudelman, A. (1977). The Markov moment problem and extremal problems. *Translations of Mathematical Monographs*, Vol. 50, AMS, Providence.

Lau, T. (1993). Higher-order Kappa-type statistics for a dichotomous attribute in multiple ratings. *Biometrics*, Vol. 49, 535-542.

Lau, T., and Studden, W. (1985). Optimal designs for trigonometric and polynomial regression using canonical moments. *The Annals of Statistics*, Vol. 13, 383-394.

Lehmann, E., and Casella, G. (2001). *Theory of Point Estimation*, Springer-Verlag, New York.

Lévy, P. (1937). *Théorie de l'Addition des Variables Aléatoires*. Gauthier-Villars, Paris.

Logan, B., Mallows, C., Rice, S., Shepp, L. (1973). Limit distributions of self-normalized sums. *The Annals of Probability*, Vol. 1, 788-809.

Martins, J., e Mendonça, S. (2008). Transformações Estabilizadoras da Variância. *Actas do XV Congresso Anual da Sociedade Portuguesa de Estatística*, Edições SPE, Lisboa, 343-350.

Mason, D., and Zinn, J. (2005). When does a randomly weighted self-normalized sum converge in distribution?. *Electronic Communications in Probability*, Vol. 10, 70-81.

Morris, C. (1983). Natural exponential families with quadratic variance functions. *The Annals of Statistics*, Vol. 2, 515-529.

Murty, V., and Studden, W. (1972). Optimal designs for estimating the slope of a polynomial regression. *Journal of the American Statistical Association*, Vol. 67, 869-873.

Oliveira, A. (1997). *Transformações Geométricas*. Universidade Aberta, Lisboa.

Pestana, D., e Velosa, S. (2008). *Introdução à Probabilidade e à Estatística*, 3.^a edição. Fundação Calouste Gulbenkian, Lisboa.

Pestana, D., e Martins, J. (2007). Tem Rho!. *Actas do XIV Congresso Anual da Sociedade Portuguesa de Estatística*, Edições SPE, Lisboa. 519-526.

Pestana, D., and Martins, J. (2007). Extensions of the delta method. *Proceedings of the 56th Congress of the International Statistical Institute* (electronic publication).

Pearson, K. (1895). Memoir on skew variation in homogeneous material. *Philosophical Transactions of the Royal Society*, A186, 323-414.

Pitman, J. (1993). *Probability*, Springer-Verlag, New York.

Pfeiffer, H., and Heremans, K. (2005). The sound velocity in ideal liquid mixtures from thermal volume fluctuations, *ChemPhysChem*, Vol. 6, 697-705.

Pukelsheim, F. (1993). *Optimal Design of Experiments*, Wiley-Interscience, New York.

Rocha, J., e Martins, J. (2005). Estatística de Student com parente não gaussiana. *Actas do XII Congresso Anual da Sociedade Portuguesa de Estatística*, Edições SPE, Lisboa, 651-656.

Salsburg, D. (2002). *The Lady Tasting Tea: How Statistics Revolutionized Science in the Twentieth Century*, Holtzbrinck Publishers, New York.

Sancetta, A. (2007). Weak convergence of laws on \mathbb{R}^K with common marginals, *Journal of Theoretical Probability*, Vol. 20, 371-380.

Skates, S. (1993). On secant approximations to cumulative distribution functions. *Biometrika*, Vol. 80, 223-235.

Skibinski, M. (1967). The range of the $(n+1)$ th moment for distributions on $[0,1]$. *Journal of Applied Probability*, Vol. 4, 1753-1764.

Skitovich, V. (1954). Linear forms of independent random variables and the normal distribution. *Izv. Akad. Nauk SSSR*, Vol. 18, 185-200 (traduzido em *Selected Translations in Mathematical Statistics and Probability*, Vol. 2, American Mathematical Society, Rhode Island, 211-228).

Song, C. (1982). Covariance stabilizing transformations and a conjecture of Holland. *The Annals of Statistics*, Vol. 10, 313-315.

Stuart, A., e Ord, J. (1994). *Kendall's Advanced Theory of Statistics*, Vol. 1, 6th edition, Oxford University Press, Oxford.

Studden, W. (1980). D_s -optimal designs for polynomial regression using continued fractions. *Annals of Statistics*, Vol. 8, 1132-1141.

Student (1908). On the probable error of the mean. *Biometrika*, Vol. 6, 1-25.

Tukey, J. (1955). On the comparative anatomy of transformations. *The Annals of Mathematical Statistics*, Vol. 28, 602-632.

Woodward, W. (1976). Approximation of Pearson type IV tail probabilities. *Journal of the American Statistical Association*, Vol. 71, 513-514.

Zhuoqiong, H., Studden, W., Sun, D. (1996). Optimal designs for rational models. *The Annals of Statistics*, Vol. 24, 2128-2147.

Anexos

A. Dos momentos ordinários aos momentos canónicos

Uma sucessão infinita de momentos de uma medida θ será representada por $\mathbf{c}(\theta) = (c_1(\theta), c_2(\theta), \dots)$. A sucessão dos primeiros k momentos será representada por $\mathbf{c}_k(\theta) = (c_1(\theta), c_2(\theta), \dots, c_k(\theta))$.

Se a medida considerada for uma medida de Dirac concentrada num ponto x , δ_x , a sucessão de momentos é x, x^2, x^3, x^4, \dots . As medidas de probabilidade consideradas nesta secção têm o intervalo $[0, 1]$ como suporte máximo a menos que seja referido algo em contrário. Feitas estas considerações iremos agora abordar o conceito de momentos canónicos.

O conjunto de todas as sucessões de momentos definidas a partir do conjunto \mathcal{P} de todas as medidas de probabilidade com suporte contido num intervalo $[a, b]$ dá lugar à seguinte definição.

Definição 4.9 *O espaço dos momentos associado a \mathcal{P} , $M = M[a, b]$, é dado por $M = \{\mathbf{c}(\theta) \mid \theta \in \mathcal{P}\}$. Se considerarmos a projecção de M em \mathbb{R}^k obtém-se o espaço de momentos de ordem k , $M_k = \{\mathbf{c}_k(\theta) \mid \theta \in \mathcal{P}\}$.*

Em particular, no caso de uma medida de Dirac δ_x , obtém-se o elemento de M dado por (x, x^2, x^3, \dots) .

De forma análoga podemos definir o conjunto de medidas de probabilidade que verificam determinada sucessão de momentos. Assim, dado $\mathbf{c}_k = (c_1, c_2, \dots, c_k)$ fixo, representar-se-á o conjunto de medidas de probabilidade associado a \mathbf{c}_k por

$$\mathcal{P}(\mathbf{c}_k) = \{\theta \in \mathcal{P} \mid c_n = c_n(\theta), n = 1, \dots, k\}.$$

A partir das sucessões de momentos x, x^2, x^3, \dots de todas as possíveis medidas de Dirac δ_x , com $x \in [0, 1]$, é possível caracterizar os elementos de qualquer espaço $M_k = M_k[0, 1]$, onde $k \in \mathbb{N}$.

Teorema 4.8 M_k é o menor conjunto convexo C_k que contém o conjunto $\{\mathbf{c}_k(\delta_x), x \in [0, 1]\}$.

Demonstração Obviamente que $C_k \subseteq M_k$. Seja $\mathbf{c}_k^0 = (c_1^0, \dots, c_k^0)$ um ponto tal que $\mathbf{c}_k^0 \in M_k$ e $\mathbf{c}_k^0 \notin C_k$.

Como C_k é fechado, existe um hiperplano que separa C_k e \mathbf{c}_k^0 , isto é, existe um vector $\mathbf{a}_k = (a_1, \dots, a_k)$ não nulo e uma constante b tal que

$$\sum_{i=1}^k a_i c_i^0 < b \text{ e } \sum_{i=1}^k a_i c_i \geq b \quad (106)$$

para qualquer $\mathbf{c}_k = (c_1, \dots, c_k) \in C_k$. Em particular, dado $\mathbf{c}_k = \mathbf{c}_k(\delta_x) \in C_k$ verifica-se que

$$\sum_{i=1}^k a_i x^i \geq b \quad (107)$$

para $x \in [0, 1]$.

Dada uma medida de probabilidade μ_0 com k primeiros momentos dados por \mathbf{c}_k^0 conclui-se que

$$\begin{aligned} \int_0^1 \left(\sum_{i=1}^k a_i x^i \right) d\mu_0(x) &\geq b \\ \Leftrightarrow \sum_{i=1}^k a_i \int_0^1 x^i d\mu_0(x) &\geq b \\ \Leftrightarrow \sum_{i=1}^k a_i c_i^0 &\geq b \end{aligned}$$

por (107) o que contradiz (106). ■

Corolário 4.1 Cada $\mathbf{c}_k \in M_k$ pode ser escrito como

$$\mathbf{c}_k = \sum_{i=1}^r \lambda_i \mathbf{c}_k(\delta_{x_i}) \quad (108)$$

onde $x_i \in [0, 1]$, $\lambda_i > 0$, $\sum_{i=1}^r \lambda_i = 1$ e $r \leq k + 1$.

Definição 4.10 A equação (108) é chamada uma **representação de \mathbf{c}_k** . Os valores x_i são chamados os **pontos de suporte** da representação de \mathbf{c}_k e os valores λ_i são designados por **pesos**. O **índice de uma representação** $\sum_{i=1}^r \lambda_i \mathbf{c}_k(\delta_{x_i})$ é dado pela diferença entre r e metade do número de pontos do suporte pertencentes ao conjunto $\{0, 1\}$. Designa-se por **índice de um ponto** $\mathbf{c}_k \in M_k$, $I(\mathbf{c}_k)$, ao mínimo dos índices de todas as representações possíveis de \mathbf{c}_k .

Exemplo 4.3 Pelo Teorema 4.8, $M_2 = M_2[0, 1]$ é o menor conjunto convexo que contém $\{(x, x^2) | x \in [0, 1]\}$. Portanto M_2 é o subconjunto de \mathbb{R}^2 representado abaixo.

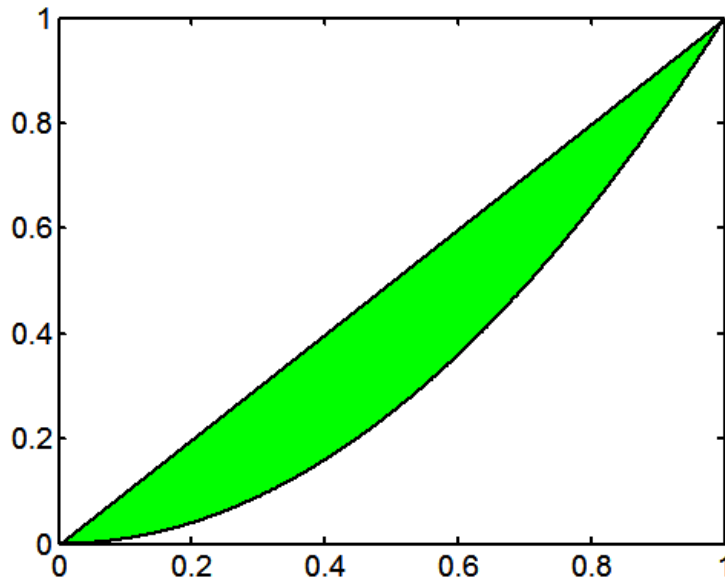


Gráfico 10 – Representação gráfica de M_2

Uma representação de $\mathbf{c}_2 = \left(\frac{1}{2}, \frac{5}{16}\right)$ pode ser

$$\mathbf{c}_2 = \frac{1}{2}\mathbf{c}_2(\delta_{1/4}) + \frac{1}{2}\mathbf{c}_2(\delta_{3/4}) = \frac{1}{2}\left(\frac{1}{4}, \left(\frac{1}{4}\right)^2\right) + \frac{1}{2}\left(\frac{3}{4}, \left(\frac{3}{4}\right)^2\right)$$

com índice $2 - 0/2 = 2$. Porém é possível representar igualmente \mathbf{c}_2 por

$$\mathbf{c}_2 = \frac{1}{5}\mathbf{c}_2(\delta_0) + \frac{4}{5}\mathbf{c}_2(\delta_{5/8})$$

com índice $2 - 1/2 = 1.5$. Não existem representações com índice inferior pois para ter índice 1 teria de ser do tipo

$$\mathbf{c}_2 = a_1\mathbf{c}_2(\delta_0) + a_2\mathbf{c}_2(\delta_1) = a_1(0, 0) + a_2(1, 1)$$

com $a_1 + a_2 = 1$ o que é impossível. Logo o índice de \mathbf{c}_2 é 1.5.

No que segue será usada a notação ∂M_k para a fronteira de M_k e $\text{Int}M_k$ para o interior de M_k .

Teorema 4.9 *São equivalentes as afirmações*

1. $\mathbf{c}_k \in \partial M_k$.
2. $I(\mathbf{c}_k) \leq \frac{k}{2}$.
3. $\mathcal{P}(\mathbf{c}_k)$ é um conjunto singular.

Uma demonstração encontra-se em Dette e Studden (1997, p.7). A partir deste teorema conclui-se imediatamente que, no exemplo anterior, $\mathbf{c}_2 \in \text{Int}M_2$ pois $I(\mathbf{c}_2) = 1.5$.

Do teorema resulta também que se $\mathbf{c}_k \in \partial M_k$ então existe apenas uma medida θ que corresponde a \mathbf{c}_k o que determina todos os valores c_n ($n > k$).

Prova-se que o espaço M_k é compacto pelo que é possível apresentar a seguinte definição.

Definição 4.11 O máximo $c_{k+1}^+ = \max_{\theta \in \mathcal{P}(\mathbf{c}_k)} c_{k+1}(\theta)$ é designado por **momento máximo** de ordem $k+1$ relativamente ao conjunto de medidas de probabilidade $\mathcal{P}(\mathbf{c}_k)$. De igual forma, o mínimo $c_{k+1}^- = \min_{\theta \in \mathcal{P}(\mathbf{c}_k)} c_{k+1}(\theta)$ é designado por **momento mínimo** de ordem $k+1$ relativamente a $\mathcal{P}(\mathbf{c}_k)$, $(k \in 1, 2, \dots)$.

O momento máximo e mínimo de ordem um correspondem, respectivamente, a $c_1^+ = \max_{\theta \in \mathcal{P}} c_1(\theta)$ e $c_1^- = \min_{\theta \in \mathcal{P}} c_1(\theta)$ onde \mathcal{P} é o conjunto de medidas de probabilidade com suporte no intervalo $[a, b]$.

Pelos Teoremas 4.8 e 4.9, se $\mathbf{c}_k \in \text{Int}M_k$ então $\mathcal{P}(\mathbf{c}_k)$ é infinito o que garante, nesse caso, que $c_{k+1}^+ - c_{k+1}^- > 0$.

Exemplo 4.4 Consideremos $c_1 \in [0, 1]$. O momento máximo de ordem 2 dado c_1 é $c_2^+ = c_1$ e o momento mínimo de ordem 2 é $c_2^- = c_1^2$.

Teorema 4.10 Cada $\mathbf{c}_k \in \text{Int}M_k$ tem duas representações de índice $\frac{k+1}{2}$:

$$\sigma_k^+ = (\mathbf{c}_k, c_{k+1}^+) \text{ e } \sigma_k^- = (\mathbf{c}_k, c_{k+1}^-).$$

Demonstração Como $\sigma_k^+, \sigma_k^- \in \partial M_k$ e $\mathbf{c}_k \in \text{Int}M_k$ por hipótese, pelo Teorema 4.9 o índice das representações de σ_k^+ e σ_k^- é, no máximo, $\frac{k+1}{2}$ e, por outro lado, $I(\mathbf{c}_k) > \frac{k}{2}$. Logo σ_k^+ e σ_k^- têm índice exactamente igual a $\frac{k+1}{2}$. ■

As representações σ_k^+ e σ_k^- têm o nome de **representação principal superior e inferior** respectivamente.

Definição 4.12 Sejam $\mathbf{c} \in M$, $N = N(\mathbf{c}) = \min \{n \in \mathbb{N} | \mathbf{c}_n \in \partial M_n\}$ se existir, caso contrário defina-se $N(\mathbf{c}) = \infty$. Chama-se momento canónico de ordem k ($k \leq N$) a

$$p_k = p_k(\mathbf{c}) = \frac{c_k - c_k^-}{c_k^+ - c_k^-}.$$

Naturalmente se $N = N(\mathbf{c}) < \infty$ então $P_A \in (0, 1)$ para $1 \leq i < N$ e $p_N \in \{0, 1\}$. Esta definição cria uma ordem relativa entre as diversas sucessões de momentos na medida em que cada momento canônico de ordem k corresponde ao quociente entre a distância entre o momento ordinário de ordem k e o valor mínimo possível para o momento de ordem k dados c_1, \dots, c_{k-1} e a amplitude entre os momentos ordinários máximo e mínimo possíveis dados c_1, \dots, c_{k-1} .

Exemplo 4.5 *Continuando o Exemplo 4.3 verifica-se facilmente que $c_1^- = 0$, $c_1^+ = 1$, $c_2^- = c_1^2$ e $c_2^+ = c_1$ para $c_1 \in (0, 1)$. Aplicando a Definição 4.12, $p_1 = c_1$ e $p_2 = \frac{c_2 - c_1^2}{c_1(1 - c_1)}$. De modo inverso, $c_1 = p_1$ e $c_2 = p_1(p_1 + p_2(1 - p_1))$.*

O momentos canônicos serão de extrema utilidade na identificação da medida de probabilidade correspondente a uma sucessão de momentos. Feller (1966) garante que se a função for definida num intervalo fechado, então, dada uma sucessão de momentos \mathbf{c} , existe uma única medida de probabilidade θ associada. Portanto, nesse caso, $p_k(\mathbf{c})$ pode ser identificada com $p_k(\theta)$.

Teorema 4.11 *Os momentos canônicos são invariantes relativamente a transformações lineares da medida de probabilidade θ .*

Detle e Studden (1997) justificam o teorema anterior apresentando expressões que permitem escrever os momentos associados a uma medida de probabilidade definida no intervalo $[0, 1]$ com os momentos associados a uma medida de probabilidade semelhante mas definida no intervalo $[a, b]$. Designando cada uma das sucessões de momentos associadas a essas medidas por $\{c_k\}_{k \in \mathbb{N}}$ e por $\{b_k\}_{k \in \mathbb{N}}$, respectivamente:

$$(b_n - b_n^-) = (b - a)^n (c_n - c_n^-) \quad \text{e} \quad (b_n^+ - b_n^-) = (b - a)^n (c_n^+ - c_n^-).$$

Se a transformação da medida de probabilidade for do tipo $y = 1 - x$ as conclusões mantêm-se para n par, isto é,

$$b_n - b_n^- = c_n - c_n^- \quad \text{e} \quad b_n^+ - b_n^- = c_n^+ - c_n^-.$$

Mas para n ímpar

$$b_n^+ - b_n = c_n - c_n^- \text{ e } b_n^+ - b_n^- = c_n^+ - c_n^-.$$

Estas relações permitem estabelecer o seguinte teorema.

Teorema 4.12 *Seja θ^* a medida de probabilidade obtida pela transformação $y = 1 - x$ de uma medida de probabilidade θ . Para $1 \leq N \leq N(\mathbf{c})$,*

$$p_k^* = \begin{cases} p_k & , \text{ se } k \text{ par} \\ q_k = 1 - p_k & , \text{ se } k \text{ ímpar} \end{cases}.$$

Obviamente, se para qualquer k ímpar $p_k = \frac{1}{2}$, então θ é simétrica.

Vamos agora caracterizar os momentos canónicos a partir dos momentos usuais. Como será necessário trabalhar com polinómios não negativos em $[0, 1]$ iremos identificar um polinómio de grau n nessas condições pelo $(n + 1)$ -uplo \mathbf{a} formado pelos seus coeficientes. O conjunto de todos esses $(n + 1)$ -uplos será representado por P_{n+1} .

Teorema 4.13

1. $\mathbf{c}_k \in M_k$ se, e só se, $\sum_{i=0}^k a_i c_i \geq 0$ quando $\mathbf{a} \in P_{k+1}$.
2. $\mathbf{c}_k \in \text{Int}M_k$ se, e só se, $\sum_{i=0}^k a_i c_i > 0$ quando $\mathbf{a} \in P_{k+1}$ e $\mathbf{a} \neq 0$.

Demonstração Note-se que $c_0 = 1$. A demonstração é semelhante à demonstração do Teorema 4.8. Mais detalhes podem ser obtidos em Dette e Studden (1997, p. 17). ■

Teorema 4.14 *P é um polinómio não negativo de grau não superior a n definido no intervalo [0, 1] se, e só se,*

$$P(x) = \begin{cases} P_m^2(x) + x(1-x)Q_{m-1}^2(x) & \text{se } n = 2m \\ xP_m^2(x) + (1-x)Q_{m-1}^2(x) & \text{se } n = 2m + 1 \end{cases}$$

onde P_m e Q_m são polinómios de grau não superior a m .

Demonstração Consultar Krein e Nudelman (1977). ■

A caracterização dos polinómios dada pelo teorema anterior será útil na aplicação do Teorema 4.13 permitindo-nos verificar se uma dada sucessão é uma sucessão de momentos. Na literatura este problema é conhecido como o problema dos momentos de Hausdorff.

Dados $\mathbf{c}_k \in M_k$ e um polinómio $P(x) = \sum_{i=0}^k a_i x^i$ defina-se

$$L(P) = a_0 + \sum_{i=1}^k a_i c_i.$$

Se $k = 2m$ os Teoremas 4.13 e 4.14 implicam que $\mathbf{c}_k \in M_k$ se, e só se, $L(P_m^2)$ e $L(x(1-x)Q_{m-1}^2)$ são não negativos para quaisquer polinómios P_m e Q_{m-1} . Se $\mathbf{a} = (a_0, a_1, \dots, a_m)$ são os coeficientes do polinómio P_m segue-se que

$$L(P_m^2) = \sum_{i=0}^m \sum_{j=0}^m a_i a_j c_{i+j} \geq 0.$$

Portanto a matriz $[c_{i+j}]$ em que a entrada (i, j) é c_{i+j} com $i, j \in \{0, 1, \dots, m\}$ é definida não negativa. Um raciocínio semelhante aplicado a $L(x(1-x)Q_{m-1}^2)$ revela que a matriz $[c_{i+j+1} - c_{i+j+2}]_{i,j=0}^{m-1}$ é definida não negativa. De igual forma $\mathbf{c}_k \in M_k$ no caso de $k = 2m + 1$ se as matrizes $[c_{i+j+1}]_{i,j=0}^m$ e $[c_{i+j} - c_{i+j+1}]_{i,j=0}^m$ são definidas não negativas. Os determinantes destas matrizes são respectivamente

os **determinantes de Hankel** \underline{H}_{2m} , \overline{H}_{2m} , \underline{H}_{2m+1} e \overline{H}_{2m+1} . Genericamente, as matrizes de Hankel são matrizes do tipo

$$\begin{pmatrix} a_1 & \cdots & a_n \\ a_2 & \cdots & a_{n+1} \\ \vdots & \ddots & \vdots \\ a_m & \cdots & a_{m+n} \end{pmatrix}.$$

Teorema 4.15

1. $\mathbf{c}_k \in M_k$ se, e só se, $\underline{H}_i, \overline{H}_i$ são não negativos para $i \in \{1, \dots, n\}$.
2. $\mathbf{c}_k \in \text{Int}M_k$ se, e só se, $\underline{H}_i, \overline{H}_i$ são positivos para $i \in \{1, \dots, n\}$.

A partir dos determinantes de Hankel podemos determinar os momentos canônicos (Dette e Studden (1997)).

Teorema 4.16 Para todo $1 \leq n \leq N(\mathbf{c})$,

$$p_n = \frac{\underline{H}_n \overline{H}_{n-2}}{\underline{H}_{n-1} \overline{H}_{n-1}} \text{ e } q_n = 1 - p_n = \frac{\underline{H}_{n-2} \overline{H}_n}{\underline{H}_{n-1} \overline{H}_{n-1}}.$$

A aplicação de um algoritmo, sobejamente conhecido como algoritmo Q-D, que tem como suporte o Teorema 4.17 apresentado adiante, para a determinação dos momentos canônicos tem como *output* os termos da sucessão $(\zeta_n)_{n \in \mathbb{N}_0}$ que a seguir se apresenta:

$$\zeta_0 = 1, \zeta_1 = p_1 \text{ e } \zeta_n = q_{n-1} p_n = \frac{\underline{H}_n \underline{H}_{n-3}}{\underline{H}_{n-1} \underline{H}_{n-2}}.$$

para $k > 1$.

O algoritmo Q-D envolve o cálculo das quantidades

$$q_m^{(n)} = \frac{H_m^{(n+1)} H_{m-1}^{(n)}}{H_m^{(n)} H_{m-1}^{(n+1)}} \quad (m \geq 1, n \geq 0) \tag{109}$$

e

$$e_m^{(n)} = \frac{H_{m+1}^{(n)} H_{m-1}^{(n+1)}}{H_m^{(n)} H_m^{(n+1)}} \quad (m, n \geq 0) \quad (110)$$

em que $H_m^{(n)}$ é o determinante de Hankel:

$$\begin{vmatrix} c_n & c_{n+1} & \cdots & c_{n+m-1} \\ c_{n+1} & c_{n+2} & \cdots & c_{n+m} \\ \vdots & \vdots & & \vdots \\ c_{n+m-1} & c_{n+m} & \cdots & c_{n+2m-2} \end{vmatrix}$$

com $m, n \geq 0$ e $H_{-1}^{(n)} = 0$. Em particular, $H_0^{(n)} = 1$ e $q_1^{(n)} = \frac{c_{n+1}}{c_n}$.

Ora as quantidades $q_m^{(0)}$ e $e_m^{(0)}$ correspondem respectivamente a ζ_{2m-1} e ζ_{2m} determinando portanto univocamente os momentos canônicos p_m .

No cálculo dessas quantidades, para obter os k primeiros momentos canônicos é necessário calcular $\frac{k^2 + k}{2}$ quantidades. Calculam-se as quantidades pela seguinte ordem:

$$\begin{aligned} & q_1^{(0)}, \dots, q_1^{(k-1)} \\ & e_1^{(0)}, \dots, e_1^{(k-2)} \\ & q_2^{(0)}, \dots, q_2^{(k-3)} \\ & e_2^{(0)}, \dots, e_2^{(k-4)} \\ & \vdots \end{aligned}$$

Feita esta breve introdução estamos em condições de apresentar o algoritmo Q-D para o cálculo dos momentos canônicos correspondentes a um vector de momentos pertencente a $\text{Int}M$. A demonstração consiste apenas na manipulação algébrica das quantidades em jogo.

Teorema 4.17 Considerem-se as quantidades $q_m^{(n)}$ e $e_m^{(n)}$ definidas em (109) e (110).

Para $m \geq 1$ e $n \geq 0$,

$$e_m^{(n)} = q_m^{(n+1)} - q_m^{(n)} + e_{m-1}^{(n+1)} \quad e \quad q_{m+1}^{(n)} = \frac{e_m^{(n+1)} q_m^{(n+1)}}{e_m^{(n)}}.$$

Executando um pequeno *script* baseado no algoritmo anterior que tem como único *input* um vector coluna c com os momentos de uma variável aleatória obtêm-se os respectivos momentos canônicos:

```

k=size(c)
k=k(1)
for i=1:k
    c(k+2-i)=c(k-i+1)
end
c(1)=1
k=size(c)
k=k(1)
c(k+1)=0
a=3
Q=zeros(k+1,k+1)
for i=2:k+1
    Q(i,2)=c(i)/c(i-1)
end
while a<k+1
    for i=a:k
        Q(i,a)=Q(i,a-1)-Q(i-1,a-1)+Q(i-1,a-2)
    end
    for j=a+1:k
        Q(j,a+1)=Q(j,a)*Q(j-1,a-1)/Q(j-1,a)
    end
    a=a+2
end
z=diag(Q)
z=z(2:1:k)
p(1)=z(1)
for i=2:k-1
    p(i,1)=z(i)/(1-p(i-1))
end

```

O *output* p é o vector coluna dos momentos canônicos. Vejamos um exemplo.

Exemplo 4.6 Considere-se uma variável aleatória com distribuição uniforme $X \sim U(0, 1)$. A tabela abaixo compara os vinte primeiros momentos canônicos com os momentos ordinários. Estes últimos são dados por $E(X^i) = \frac{1}{i+1}$.

i	p_i	$E(X^i)$	i	p_i	$E(X^i)$
1	0.5000	0.5	12	0.4615	0.0769
2	0.3333	0.3333	13	0.5000	0.0714
3	0.5000	0.25	14	0.4667	0.0666
4	0.4000	0.2	15	0.5000	0.0625
5	0.5000	0.1667	16	0.4706	0.0588
6	0.4286	0.1429	17	0.5000	0.0555
7	0.5000	0.1250	18	0.4737	0.0526
8	0.4444	0.1111	19	0.4999	0.05
9	0.5000	0.1	20	0.4766	0.0476
10	0.5500	0.0909	21	0.4986	0.0455
11	0.5000	0.0833	22	0.4790	0.0435

Tabela 35 – Momentos canônicos e ordinários da distribuição $U(0, 1)$

Os momentos canônicos de ordem ímpar são iguais a 0.5 como seria de esperar pelo Teorema 4.12. A partir da ordem vinte o algoritmo torna-se instável tendo a iteração vinte e dois um erro da ordem 10^{-2} ao passo que na iteração vinte era da ordem 10^{-4} . Tal situação aponta para um mau condicionamento do algoritmo na acumulação de erros de cálculo.

B. Planeamentos mistos Θ_0^π -óptimos

a	peso em ± 1	peso em 0
0.05	0.2835	0.4330
0.10	0.2895	0.4211
0.15	0.2958	0.4084
0.20	0.3025	0.3951
0.25	0.3095	0.3810
0.30	0.3170	0.3660
0.35	0.3249	0.3502
0.40	0.3333	0.3333
0.45	0.3423	0.3154
0.50	0.3519	0.2963
0.55	0.3621	0.2759
0.60	0.3730	0.2540
0.65	0.3848	0.2305
0.70	0.3974	0.2051
0.75	0.4111	0.1778
0.80	0.4259	0.1481
0.85	0.4420	0.1159
0.90	0.4596	0.0808
0.95	0.4788	0.0423

Tabela 36 – Planeamento misto Θ_0^π -óptimo, $m = 2$, $\pi = (a, 1 - a)$

a	b	t	peso em ± 1	peso em $\pm t$
0.1	0.1	0.4911	0.2119	0.2881
0.1	0.2	0.4748	0.2190	0.2810
0.1	0.3	0.4553	0.2267	0.2733
0.1	0.4	0.4315	0.2350	0.2650
0.1	0.5	0.4019	0.2440	0.2560
0.1	0.6	0.3635	0.2538	0.2462
0.1	0.7	0.3112	0.2646	0.2354
0.1	0.8	0.2318	0.2764	0.2236
0.2	0.1	0.5001	0.2256	0.2744
0.2	0.2	0.4808	0.2338	0.2662
0.2	0.3	0.4569	0.2428	0.2572
0.2	0.4	0.4269	0.2525	0.2475
0.2	0.5	0.3876	0.2632	0.2368
0.2	0.6	0.3333	0.2750	0.2250
0.2	0.7	0.2496	0.2880	0.2120
0.3	0.1	0.5095	0.2415	0.2585
0.3	0.2	0.4858	0.2513	0.2487
0.3	0.3	0.4556	0.2619	0.2381
0.3	0.4	0.4155	0.2736	0.2264
0.3	0.5	0.3593	0.2865	0.2135
0.3	0.6	0.2709	0.3009	0.1991
0.4	0.1	0.5190	0.2606	0.2394
0.4	0.2	0.4889	0.2722	0.2278
0.4	0.3	0.4483	0.2851	0.2149
0.4	0.4	0.3902	0.2994	0.2006
0.4	0.5	0.2967	0.3154	0.1846
0.5	0.1	0.5281	0.2836	0.2164
0.5	0.2	0.4875	0.2978	0.2022
0.5	0.3	0.4279	0.3137	0.1863
0.5	0.4	0.3289	0.3316	0.1684
0.6	0.1	0.5351	0.3121	0.1879
0.6	0.2	0.4748	0.3299	0.1701
0.6	0.3	0.3705	0.3500	0.1500
0.3	0.4	0.5353	0.3481	0.1519
0.3	0.5	0.4267	0.3710	0.1290
0.3	0.6	0.5085	0.3953	0.1047
1/3	1/3	0.4407	0.2731	0.2269

Tabela 37 – Planeamento misto Θ_0^π -óptimo, $m = 3$, $\pi = (a, b, 1 - a - b)$

Tabela 38 – Planeamento misto Θ_0^π -óptimo, $m = 4$, $\pi = (a, b, c, 1 - a - b - c)$

a	b	c	t	peso em ± 1	peso em $\pm t$	peso em 0
0.1	0.1	0.1	0.6973	0.1717	0.2177	0.2210
0.1	0.1	0.2	0.6836	0.1764	0.2176	0.2119
0.1	0.1	0.3	0.6673	0.1814	0.2184	0.2004
0.1	0.1	0.4	0.6474	0.1867	0.2206	0.1853
0.1	0.1	0.5	0.6228	0.1924	0.2252	0.1647
0.1	0.1	0.6	0.5913	0.1985	0.2341	0.1349
0.1	0.1	0.7	0.5495	0.2050	0.2513	0.0874
0.1	0.2	0.1	0.6937	0.1811	0.2036	0.2307
0.1	0.2	0.2	0.6765	0.1864	0.2036	0.2200
0.1	0.2	0.3	0.6553	0.1920	0.2051	0.2057
0.1	0.2	0.4	0.6284	0.1981	0.2091	0.1856
0.1	0.2	0.5	0.5933	0.2046	0.2178	0.1553
0.1	0.2	0.6	0.5453	0.2115	0.2363	0.1043
0.1	0.3	0.1	0.6886	0.1917	0.1873	0.2420
0.1	0.3	0.2	0.6661	0.1977	0.1879	0.2289
0.1	0.3	0.3	0.6370	0.2042	0.1908	0.2100
0.1	0.3	0.4	0.5979	0.2111	0.1988	0.1802
0.1	0.3	0.5	0.5423	0.2186	0.2182	0.1265
0.1	0.4	0.1	0.6811	0.2038	0.1686	0.2552
0.1	0.4	0.2	0.6499	0.2107	0.1701	0.2383
0.1	0.4	0.3	0.6065	0.2182	0.1766	0.2104
0.1	0.4	0.4	0.5417	0.2262	0.1959	0.1557
0.1	0.5	0.1	0.6692	0.2177	0.1469	0.2707
0.1	0.5	0.2	0.6217	0.2258	0.1509	0.2466
0.1	0.5	0.3	0.5462	0.2345	0.1683	0.1943
0.1	0.6	0.1	0.6481	0.2340	0.1216	0.2887
0.1	0.6	0.2	0.5610	0.2435	0.1342	0.2447
0.1	0.7	0.1	0.5981	0.2533	0.0935	0.3064
0.2	0.1	0.1	0.7031	0.1851	0.2128	0.2041
0.2	0.1	0.2	0.6869	0.1907	0.2124	0.1937
0.2	0.1	0.3	0.6670	0.1967	0.2132	0.1801
0.2	0.1	0.4	0.6419	0.2032	0.2162	0.1613
0.2	0.1	0.5	0.6092	0.2101	0.2231	0.1335
0.2	0.1	0.6	0.5647	0.2175	0.2384	0.0881
0.2	0.2	0.1	0.6978	0.1964	0.1965	0.2143
0.2	0.2	0.2	0.6766	0.2028	0.1964	0.2017
0.2	0.2	0.3	0.6492	0.2097	0.1985	0.1837
0.2	0.2	0.4	0.6125	0.2171	0.2049	0.1559
0.2	0.2	0.5	0.5607	0.2251	0.2212	0.1074
0.2	0.3	0.1	0.6900	0.2093	0.1775	0.2265
0.2	0.3	0.2	0.6604	0.2167	0.1783	0.2101
0.2	0.3	0.3	0.6195	0.2247	0.1835	0.1836
0.2	0.3	0.4	0.5586	0.2333	0.2001	0.1332
0.2	0.4	0.1	0.6778	0.2242	0.1552	0.2411
0.2	0.4	0.2	0.6325	0.2329	0.1584	0.2175

(continua na página seguinte)

Tabela 38 – Planeamento misto Θ_0^π -óptimo, $m = 4$, $\pi = (a, b, c, 1 - a - b - c)$

0.2	0.4	0.3	0.5610	0.2423	0.1737	0.1681
0.2	0.5	0.1	0.6562	0.2418	0.1291	0.2582
0.2	0.5	0.2	0.5728	0.2520	0.1404	0.2151
0.2	0.6	0.1	0.6058	0.2627	0.0997	0.2753
0.3	0.1	0.1	0.7089	0.2014	0.2058	0.1857
0.3	0.1	0.2	0.6891	0.2082	0.2049	0.1737
0.3	0.1	0.3	0.6637	0.2156	0.2060	0.1569
0.3	0.1	0.4	0.6297	0.2236	0.2107	0.1316
0.3	0.1	0.5	0.5820	0.2322	0.2234	0.0888
0.3	0.2	0.1	0.7009	0.2152	0.1865	0.1966
0.3	0.2	0.2	0.6732	0.2231	0.1864	0.1809
0.3	0.2	0.3	0.6349	0.2317	0.1902	0.1562
0.3	0.2	0.4	0.5785	0.2411	0.2035	0.1109
0.3	0.3	0.1	0.6884	0.2313	0.1639	0.2098
0.3	0.3	0.2	0.6457	0.2406	0.1658	0.1872
0.3	0.3	0.3	0.5787	0.2508	0.1784	0.1416
0.3	0.4	0.1	0.6663	0.2503	0.1369	0.2256
0.3	0.4	0.2	0.5872	0.2614	0.1464	0.1844
0.3	0.5	0.1	0.6155	0.2731	0.1061	0.2417
0.4	0.1	0.1	0.7144	0.2216	0.1957	0.1656
0.4	0.1	0.2	0.6888	0.2301	0.1944	0.1510
0.4	0.1	0.3	0.6536	0.2394	0.1963	0.1287
0.4	0.1	0.4	0.6022	0.2495	0.2059	0.0892
0.4	0.2	0.1	0.7016	0.2389	0.1726	0.1771
0.4	0.2	0.2	0.6619	0.2490	0.1730	0.1561
0.4	0.2	0.3	0.6001	0.2601	0.1823	0.1152
0.4	0.3	0.1	0.6791	0.2595	0.1449	0.1912
0.4	0.3	0.2	0.6051	0.2717	0.1520	0.1527
0.4	0.4	0.1	0.6281	0.2845	0.1126	0.2057
0.5	0.1	0.1	0.7182	0.2472	0.1811	0.1434
0.5	0.1	0.2	0.6821	0.2582	0.1796	0.1245
0.5	0.1	0.3	0.6263	0.2703	0.1850	0.0894
0.5	0.2	0.1	0.6956	0.2697	0.1527	0.1552
0.5	0.2	0.2	0.6276	0.2830	0.1566	0.1207
0.5	0.3	0.1	0.6445	0.2972	0.1190	0.1676
0.6	0.1	0.1	0.7172	0.2810	0.1598	0.1184
0.6	0.1	0.2	0.6566	0.2957	0.1597	0.0892
0.6	0.2	0.1	0.6667	0.3114	0.1246	0.1280
0.7	0.1	0.1	0.6978	0.3274	0.1285	0.0882
0.25	0.25	0.25	0.6484	0.2239	0.1839	0.1845

Índice

- BLUE, 26, 27, 29, 34
- coeficiente
- de achatamento, 8, 9, 133–135
 - de assimetria, 8, 64, 101, 109, 133–135, 152
- cumulante, 11
- cumulante de ordem i , 10
- decomposição de Choleski, 13
- determinante de Hankel, 3, 18, 31, 32, 171, 172
- distribuição
- de tipo IV, 129, 132–135, 142, 148–150
 - gaussiana generalizada, 102, 142, 148
- domínio de atracção, 100, 101, 107, 108
- eficiência
- D, 33, 40
 - D_1 , 32
 - ϕ_p , 30
 - relativa, 31, 35
- estimador de Gauss-Markov, 26
- expansão de Edgeworth, 129, 148, 149, 151–153
- função
- completamente monótona, 102–104, 144
 - ϕ_p de Kiefer, 29
 - geradora de
 - cumulantes, 10
 - momentos, 9, 10
- infinitamente divisível, 102, 104–106, 143–145
- método delta, 55, 57, 58, 61, 64, 68, 77, 109
- matriz
- de dispersão, 26, 29
 - de informação, 28, 29, 31–33
 - de informação de Fisher, 72
 - do planeamento, 25, 28, 37
 - inversa generalizada, 16, 26
- momento
- canónico, 18, 19, 31, 33, 35, 38–41, 167–169, 171
 - centrado de ordem k , 7
 - de ordem k , 6
- ordem de Loewner, 25, 26, 29

planeamento

aproximado, 29, 35

D-óptimo, 30, 32, 33, 39

D₁-óptimo, 30–32, 36, 38

discriminante, 38, 41, 50

misto, 42, 50, 175–177

ϕ_p -óptimo, 30

robusto, 40, 41, 50

polinómio de Taylor, 56, 57

polinómios ortogonais

de Hermite, 19, 150

de Laguerre generalizados, 22

pseudomomentos canónicos, 19, 21–23

raiz quadrada de uma matriz, 12, 82, 88

sistema de Pearson, 126–129, 132–134,
142

teorema de Slutsky, 57–60, 142

transformação

de Laplace, 103

estabilizadora

da covariância, 81

da covariância, 81, 83, 84, 86–91

da variância, 67–70, 72–74, 77–79

potência, 54