

**FILTRO DE KALMAN
E
PREVISÃO ECONOMÉTRICA**

1. 5. 1. 1.
E.
1265-G. 34170

HB 139. C73 1987

21



**FILTRO DE KALMAN
E
PREVISÃO ECONOMÉTRICA**

por

Nuno Paulo de Sousa Arrobas Crato

Dissertação apresentada como requisito parcial para a obtenção do grau de Mestre em Métodos Matemáticos para Gestão de Empresas

pelo

Instituto Superior de Economia

da

Universidade Técnica de Lisboa

**Trabalho orientado pelo
Professor Doutor Daniel Muller**

a quem se agradece todo o cuidado posto na orientação deste estudo, reconhecendo, no entanto, que todos os erros, lacunas e insuficiências permanecem da exclusiva responsabilidade do seu autor.

Outubro de 1987

ÍNDICE

Introdução

primeira parte — operadores de previsão e projectores ortogonais

1. Estimação, previsão e filtros lineares
2. Optimalidade na estimação — o problema de Gauss
3. Processos gaussianos e não gaussianos
4. Filtros lineares e projectores ortogonais — solução de Wiener-Kolmogorov

segunda parte — previsão com filtro de Kalman e modelização em espaço de estados

5. Procedimentos de previsão recursiva
6. Algoritmo preditor de Kalman
7. Propriedades e extensões do filtro de Kalman
8. A previsão recursiva na visão bayesiana
9. Espaço de estados, problemas de identificação e estimação

terceira parte — problemas recentes

10. O filtro de Kalman como unificador de procedimentos
11. Alguns sucessos recentes

Conclusões

Referências

NOTAS ANEXAS

1. Processos estocásticos e sucessões cronológicas
2. Convergência em média quadrática
3. Funções de perda não pares
4. Variáveis aleatórias de segunda ordem
5. Variedades gaussianas e processos gaussianos
6. Séries de variáveis aleatórias
7. Pseudo-inversa
8. Processos puramente aleatórios — "ruído branco"
9. O espaço de Hilbert $L^2(\Omega)$
10. Processos lineares gerais, AR, MA, ARMA e ARIMA
11. Processos estacionários
12. Estudo de um caso: previsão inflacionária

Referências adicionais



Introdução

É objecto deste trabalho o estudo do Filtro de Kalman na óptica da sua aplicação à previsão econométrica.

O tema é vasto e tanto o ângulo de abordagem como os problemas em estudo sofreram uma delimitação prévia bem clara.

Optou-se por estudar apenas sucessões cronológicas — processos estocásticos discretos — pois é a sua análise a mais relevante na estatística e econometria. Optou-se também por focalizar todo o estudo no domínio tempo — prescindindo dos desenvolvimentos possíveis pela análise espectral — pois assim todos os resultados são obtidos de forma mais familiar à econometria estatística.

A análise foi directamente orientada para a modelização e previsão econométrica — o que significa que foram deliberadamente esquecidos problemas típicos da teoria do controlo e teoria dos sistemas, da análise do sinal passado e do alisamento de sucessões, assim como outros desenvolvimentos e algoritmos aplicáveis a campos científicos diversos.

A teoria de Kalman ficou assim delimitada ao terreno específico da previsão, aonde se tem revelado como susceptível das mais fecundas aplicações.

Na realidade, o interesse e actualidade do tema são bem revelados por uma sucessão recente de trabalhos econométricos e até pelo desenvolvimento de *software* comercial, aonde se redescobre a potencialidade e versatilidade da teoria de Kalman.

Trata-se essencialmente de uma *redescoberta*, pois desde o trabalho pioneiro de Rudolf Kalman, em 1960, cerca de vinte anos de desenvolvimento teórico se passaram sem que a econometria e a previsão económica praticamente tomassem contacto com essa nova abordagem da estimação estocástica.

Foi nos últimos anos que se assistiu a uma autêntica explosão de artigos e trabalhos econométricos que redescobriram resultados anteriores, os concretizara e desenvolvera em novos terrenos específicos.

Se esse facto, talvez ainda pouco conhecido entre nós, justifica cabalmente o interesse deste estudo, a realização do presente trabalho, tem origem em algumas lacunas essenciais.

Em primeiro lugar não se conhece uma exposição básica da teoria e do algoritmo de Kalman que esteja directamente orientada para o terreno econométrico.

Os textos base, que são ainda os de Jazwinski (1970) e Anderson e Moore (1979), "are lengthy and unintelligible to statisticians" (Meinhold e Singpurwalla, 1987): Em contrapartida, as exposições mais "pedagógicas" de Harvey (1981) ou dos bayesianos, focam pontos de vista particulares, de tal forma que, "the motivation behind and major applications of the algorithm cannot be fully appreciated" (Welch, 1987).

Essa limitação é ainda mais evidente na literatura nacional aonde, aparte uma exposição sintética de Neves (1984), apenas se conhecem algumas aplicações muito específicas (e.g. Lopes, 1985 e Tomé e Cunha, 1985).

Em segundo lugar, está por fazer um "survey" das aplicações econométricas recentes, resenha aonde se discutam de forma sistemática as potencialidades e limitações da modelização e do algoritmo de Kalman.

Essa lacuna é particularmente sentida por quem pretenda iniciar o estudo deste tema, que se vê defrontado com muitas referências dispersas, sendo obrigado, para a compreensão de problemas muitas vezes simples, ao confronto com uma vasta e não homogénea bibliografia.

Em terceiro lugar, e como consequência dessas limitações, a terminologia e a notação são altamente heterogéneas, tendo a diversidade de escolas destruído as tentativas iniciais de Kalman para uniformizar as designações.

São estes os três obstáculos básicos enfrentados por este trabalho, obstáculos que não se tem a veleidade de destruir mas que se tem a intenção de ajudar a remover.

Com este objectivo o trabalho desenvolve-se em três partes.

Na primeira parte redefinem-se os problema básicos de estimação estocástica na perspectiva que se julga ser a mais actual, rigorosa e fecunda: a perspectiva dos *operadores de projecção*. Nesse enquadramento teórico são encarados da mesma forma, e com grande generalidade, problemas diversos de estimação e previsão. A definição da *sucessão de inovações* aparece de forma directa e fica preparada a demonstração de Kalman, que ainda hoje se reconhece como sendo a mais fecunda.

Uma corroboração do interesse desta sequência foi fornecida pela publicação, já em fase adiantada da redacção deste trabalho, da obra *Time Series: Theory and Methods*, de Brockwell e Davis. Com outro objectivo e outra abordagem, estes autores percorrem um caminho paralelo, para um estudo muito geral da estimação de processos estocásticos.

Na segunda parte define-se o modelo de espaço de estados e apresentam-se os desenvolvimentos de Akaike, ao mesmo tempo que se deriva o filtro de Kalman e se discutem os processos recursivos de estimação.

Ainda nesta segunda parte algum relevo é dado a uma interpretação polémica da estimação recursiva, que é associada ao carácter dinâmico dos sistemas relembrando alguns resultados que nem sempre parecem ser correctamente entendidos.

Na terceira parte do trabalho é empreendido um "survey" de aplicações econométricas, tentando-se uma exposição sistemática da modelização de Kalman e referindo-se, em capítulo próprio, algumas simulações e estudos comparativos recentes.

Depois da exposição sistemática da teoria de Kalman, o "survey" aparece sobretudo com objectivos descritivos, pois a concretização econométrica da teoria geral não apresenta desenvolvimentos de vulto.

Um capítulo de conclusões dedica-se sobretudo à sistematização de problemas omissos e à abertura de pistas para uma investigação posterior.

Uma série de notas anexas, que não constitui parte integrante do trabalho, é incluída com o objectivo de referenciar alguns conceitos e definições básicas.

primeira parte

operadores de previsão e projectores ortogonais

Em sucessões cronológicas o problema da previsão pode ser visto como um caso particular do problema da estimação. Começar-se-á por definir este último e por apresentar, de forma directamente aplicável ao estudo consequente, um conjunto de resultados gerais hoje já clássicos.

Na primeira secção definir-se-á o problema da estimação e referir-se-á uma classe particular de estimadores: os filtros lineares.

Na segunda secção serão discutidos os critérios de optimalidade dos estimadores, definindo os métodos probabilísticos de máxima verosimilhança e de minimização de uma perda esperada. Esses métodos serão confrontados com procedimentos estatísticos não probabilísticos.

Em seguida serão apresentados os processos gaussianos e discutir-se-á a aplicabilidade dos resultados existentes.

Na secção final far-se-á referência aos preditores de Wiener-Kolmogorov e será estudada a sequência de inovações. A construção de um espaço de Hilbert de variáveis aleatórias de segunda ordem permitirá definir projectores ortogonais como operadores de previsão probabilística.

Todos estes conceitos serão largamente utilizados na dedução e estudo do Filtro de Kalman, a ser apresentado na parte seguinte deste trabalho.

1. Estimaco, previso e filtros lineares

Comear-se- por definir o problema da estimaco no quadro do estudo de sucesses cronolgicas aqui desenvolvido.

Sejam dadas duas sucesses de variveis, escalares ou vectoriais, indexadas pelo ´ndice t , discreto, que representa o tempo: $\{y_t\}$ e $\{x_t\}$ ⁽¹⁾. Admita-se que as variveis esto relacionadas por alguma funo conhecida, possivelmente dependente do ´ndice t , e que se lhe soma uma perturbao aleatria, escalar ou vectorial, que constitui a sucesso $\{v_t\}$, de valor esperado nulo.

$$y_t = h(x_t, t) + v_t \quad Ev_t = 0. \quad (1.1)$$

Para cada ´ndice t assume-se conhecido y_t e a forma de funo $h(., t)$. Tanto x_t como v_t representam variveis de valor desconhecido.

$\{y_t\}$ e $\{v_t\}$ so sucesses de variveis aleatrias indexadas pelo tempo. Representam pois processos estocsticos, neste caso de parmetro discreto, definindo SUCESSES CRONOLGICAS (V. Nota Anexa 1). A sucesso $\{x_t\}$ pode ser, ou no, formada por variveis aleatrias. Se tiver esse carcter estocstico constituir uma outra sucesso cronolgica.

Neste quadro adopta-se a definio seguinte.

O PROBLEMA DA ESTIMACO consiste na procura de um operador —ESTIMADOR— a ser aplicado sobre os elementos conhecidos de $\{y_t\}$, e que determine um valor considerado provavelmente prximo, por algum critrio de optimidade previamente formulado e para um dado ´ndice τ , do verdadeiro x_τ . Esse valor, a ESTIMATIVA de x_τ , representar-se- por \hat{x}_τ .

(1) Contrariamente  notaco usual em teoria das probabilidades no distinguir v.a. de uma sua realizaco, sendo ambas representadas por letras minsculas. As maisculas sero reservadas  matrizes ou a conjuntos (sequncias) de variveis. O tipo negro indica um vector, possivelmente de componente ´nico, e o tipo normal, indexado, um elemento de vector. A notaco $X(t)$ indica uma funo aleatria, em tempo contnuo ou discreto, podendo designar uma coleco de vectores aleatrios, mas x_t representa sempre um processo aleatrio discreto. O uso do smbolo $X(t)$ pode ser justificado pela pretenso de mostrar que o conceito tem uma interpretao mais lta do que a de uma simples sucesso cronolgica.

O problema da previsão não é mais do que um caso particular deste último. Dada a sequência das observações, sequência aleatória⁽²⁾ finita

$$Y_t = \{y_0, y_1, \dots, y_t\}, \quad (1.2)$$

ou infinita

$$Y_t = \{\dots, y_{t-n}, \dots, y_{t-1}, y_t\} \quad (1.3)$$

pretende-se uma estimativa de um valor futuro x_τ . Assim considerar-se-á que

PREVISÃO é uma estimativa de x_τ , designada por \hat{x}_τ , elaborada com o conhecimento da sequência das observações Y_t , sendo $t < \tau$.

Os operadores de estimação são muitas vezes denominados *filtros*, terminologia adoptada pela teoria dos processos estocásticos sob influência da teoria da informação, da engenharia do controlo e da electrónica.

A origem do termo será clara se designarmos $\{x_t\}$ como mensagem ou *signal* e se considerarmos como sua *recepção* a sequência Y_t , que é perturbada pela adição de um *ruído* $\{v_t\}$. O processo de estimação, obtendo $\{\hat{x}_t\}$, constituirá a recuperação aproximada do sinal, a partir das observações contaminadas pelo ruído.

Estará em causa, no dizer de Wiener, "The purification or filtering". "Very often the quantity which we really wish to observe is observable only after it has been in some way corrupted or altered by mixture, additively or not, with other time series. It is of importance to ascertain as nearly as possible in an appropriate sense, that is, in a statistical sense, what our data would have been like without the contamination of the other time series." (Wiener, 1949, p.9).

Neste contexto ter-se-á uma primeira definição de filtro⁽³⁾

FILTRO (estimador) é um operador que, aplicado sobre uma sequência de observações Y_t , no modelo (1.1), produz uma estimação de x_τ .

(2) Distinguir-se-á sequência de observações de sequência registada de observações (não aleatória).

(3) V. Wiener, 1949, pp.9-10 e 81-2, Koroliouk, 1983, p.253 e Priestley, 1981, pp.773-4 e 807.

Como o termo "filtro" designa também outros conceitos não exactamente coincidentes com o desta definição valerá a pena clarificar desde já as diversas acepções do termo.

No sentido mais geral, filtro é simplesmente um operador que transforma um processo estocástico num outro processo estocástico. Trata-se de um operador \mathcal{F} , aplicado sobre uma função aleatória de *input*, $Y(t)$, que a transforma na função aleatória de *output*, $Z(t)$.

Simbolicamente,

$$\mathcal{F}Y(t) = Z(t). \quad (1.4)$$

No quadro do modelo (1.1) qualquer operação sobre as observações é um filtro desde que transforme a sequência Y_t numa outra sequência.

Assim,⁽⁴⁾

FILTRO (de *input/output*) é um operador que, aplicado sobre a sequência de observações de $\{y_t\}$, produz uma outra sucessão cronológica $\{z_t\}$

Muito claramente, está aqui apenas referida a transformação de um processo aleatório num outro processo aleatório, não estando em causa propriedades particulares, como a da capacidade de estimação do sinal.

No caso relevante pretende-se uma estimativa \hat{x}_t , que seja função das observações Y_t . Se, em cada momento t , determinada forma funcional transformar Y_t em \hat{x}_t , o processo de estimação ou filtragem corresponde à transformação de $\{y_t\}$ em $\{\hat{x}_t\}$. Há pois um filtro, de *input/output*, que se aplica sobre $\{y_t\}$ e que consiste também num filtro, de *estimação*, que gera as estimativas \hat{x}_t .

Os dois conceitos estão ligados na acepção mais comum de filtro, que corresponde à definição que segue⁽⁵⁾.

FILTRO (*on line*) é um operador que, em cada t , transforma a sequência de observações Y_t numa estimativa para o valor de x_t no mesmo período: \hat{x}_t

Nesse sentido, que é o mais comum na teoria da estimação estocástica, filtro é um operador de geração de estimativas em "tempo real": $\mathcal{F}Y_t = \hat{x}_t$. Um processo estocástico observável

(4) V. Priestley, 1981, pp.285-7 e Anderson, 1971, p.399.

(5) V. Kalman, 1960, p.36 ou Jazwinski, 1970, pp.143-4, entre outros.

é transformado numa sucessão de estimativas que, sendo função de Y_t , são também elas estocásticas.

A sobrecarregar ainda mais a terminologia confrontam-se, do ponto de vista da relação temporal entre estimativas e observações, os conceitos de alisamento, filtragem e previsão. Conhecendo-se Y_t e pretendendo-se estimar x_τ , passará a falar-se de ALISAMENTO quando $\tau < t$, de FILTRAGEM quando $\tau = t$ e de PREVISÃO quando $\tau > t$. O conceito de estimação abrange evidentemente os três casos, e o *filtro on line* é precisamente um *operador de filtragem*.

O problema será melhor definido na secção 6., aonde serão feitas ainda algumas precisões. Refira-se agora o importante conceito de *filtro linear e invariante*⁽⁶⁾, aplicável ao termo em qualquer das acepções atrás definidas.

A propriedade de *invariância* temporal do filtro consistirá no facto de o mesmo *input* produzir sempre o mesmo *output*, qualquer que seja o momento de seu registo. Mais precisamente,

$$\begin{aligned} \text{FILTRO INVARIANTE é o operador de filtragem tal que,} \\ Y_{(t)} = Y_{(t+s)} \Rightarrow \mathcal{F}Y_{(t)} = \mathcal{F}Y_{(t+s)} \quad \forall s, \end{aligned} \quad (1.5)$$

o que significa que a regra de geração de estimativas se mantém constante no tempo.

Outro importante conceito é o de filtro linear, operador que transforma um *input* combinação linear de duas funções num *output* que é a mesma combinação linear dos *outputs* individuais.

$$\begin{aligned} \text{FILTRO LINEAR é o operador de filtragem tal que,} \\ \mathcal{F}[a_1 Y_{1(t)} + a_2 Y_{2(t)}] = a_1 \mathcal{F}Y_{1(t)} + a_2 \mathcal{F}Y_{2(t)} \end{aligned} \quad (1.6)$$

quaisquer que sejam as constantes a_1 e a_2 e as funções de *input* Y_1 e Y_2 .

É usual confundir filtro linear com *filtro afim* e vale a pena apresentar desde já a distinção.

(6) V. e.g. Parzen, 1962, p.104. Em certo sentido a invariância de um operador de filtragem está ligada à capacidade de reproduzir situações observáveis sem alteração de circunstâncias, o que é uma das características base das ciências experimentais. Os operadores invariantes constituem um grupo, denominado grupo de translação no tempo (Cf. Wiener, 1949, p.11).

FILTRO AFIM é o operador de filtragem tal que,

$$\mathcal{F}[a_1 Y_1(t) + a_2 Y_2(t)] = u(t) + a_1 \mathcal{F}Y_1(t) + a_2 \mathcal{F}Y_2(t), \quad (1.7)$$
em que $u(t)$ é função determinística, sendo a igualdade (1.7) válida para todas as constantes a_1 e a_2 e funções de input Y_1 e Y_2 .

O abuso de linguagem que consiste em confundir linearidade com afinidade é justificável por as atenções se centrarem sobre a parte aleatória do processo⁽⁷⁾. Mas o filtro linear é apenas um caso particular do filtro afim, correspondendo à situação $u(t) \equiv 0$.

Alguns exemplos ajudarão a mostrar a generalidade destes conceitos e a introduzir os problemas que seguem.

Exemplo 1.1 (Passeio aleatório)

Considere-se o processo estocástico $\{z_t\}$, com z_t inteiro, definido pelas equações:

$$z_{t+1} = z_t + w_t, \quad t = 0, 1, 2, \dots$$

$$Ew_t = \text{const}, \quad \text{cov}[w_k w_l] = 0 \quad \forall k \neq l$$

$$P[w_t = -1] + P[w_t = 0] + P[w_t = +1] = 1, \quad P[z_0 = 0] = 1$$

Reescrevendo o processo de acordo com a equação (1.1) ter-se-á $y_t = z_{t+1}$, $x_t = z_t$, $w_t = v_t$ e $h(x_t, t) = z_t$. Dada a sequência $Y_t = \{z_1, z_2, \dots, z_{t+1}\}$ e o conhecimento de $z_0 = 0$ pode estimar-se com probabilidade 1 a sequência $X_t = \{z_0, z_1, \dots, z_t\}$. O problema da previsão consiste em estabelecer uma estimativa para $x_{t+l} = z_{t+l}$. Sendo $l \geq 2$ essa estimativa será incerta e uma previsão aceitável é, por exemplo, a do valor médio condicionado pelo conhecimento de Y_t (que está concentrado no conhecimento de z_{t+1}):

$$\hat{z}_{t+l} = z_{t+1} + (l-1)Ew_t$$

Trata-se de um filtro preditivo invariante e afim (linear se $Ew_t = 0$), transformando $\{z_{t+1}\}$ em $\{\hat{z}_{t+l}\}$.

Em geral, $(l-1)Ew_t$ é não inteiro, pelo que tal previsão naturalmente apontará para valor não verificável. Surge como previsão alternativa a do valor modal. Definindo $m(l-1)$ como a moda de $\sum_{k=2}^l w_k$ ter-se-á outro filtro preditivo afim:

$$\hat{z}_{t+l} = z_{t+1} + m(l-1)$$

Admita-se agora que a distribuição de probabilidade de w_t é desconhecida e se pretende, precisamente, estimar o seu valor esperado. De acordo com a equação (1.1) o processo poderia ser modelizado

(7) V. Anderson e Moore, 1979, pp.92-8.

$$z_{t+1} - z_t = Ew_t + (w_t - Ew_t) \quad t = 1, 2, \dots$$

Ter-se-á agora $y_t = z_{t+1} - z_t$, $x_t = Ew_t$, $v_t = w_t - Ew_t$ e $h(x_t, t) = z_t$. Repare-se que a transformação $\{z_t\} \rightarrow \{y_t\}$ foi obtida através de um filtro linear e invariante (para $t \geq 1$). A estimação de $x_t = Ew_t$ pode ser realizada através do filtro (on line):

$$\hat{x}_t = \sum_1^t y_i / t,$$

filtro linear mas não invariante, que utiliza em cada momento a sequência de observações Y_t para gerar $\{\hat{x}_t\}$, sequência de estimativas centradas e convergentes para Ew_t .

Por razões diversas e largamente conhecidas, algumas das quais se espera vir a clarificar na secção 3., a classe dos estimadores lineares tem tal importância que sobre ela se centra a larga maioria dos estudos e resultados existentes.

No quadro do modelo (1.1) estimadores lineares são *filtros lineares* nas observações, portanto operadores do tipo,

$$\hat{x}_t = \mathcal{F}Y_t = \sum_i a_i y_i, \quad i \leq t \quad (1.8)$$

em que o somatório é finito ou se entende como o limite em média quadrática das somas parciais⁽⁸⁾.

Em particular *preditores lineares* serão operadores de ponderação das observações já registadas, operadores da forma (1.8) portanto.

Geometricamente isso significa que tanto os estimadores como os preditores lineares podem ser encarados como *vectores da variedade linear gerada pelas observações*: y_t . Tal conceito necessitará de ser precisado ao espaço particular das variáveis aleatórias em causa, o que envolve alguns problemas adicionais a estudar na secção 4. Caso o número de observações não seja finito é o *fecho da variedade linear* que deve ser considerado, incluindo aí tanto o conjunto de combinações lineares finitas como os possíveis limites (em média quadrática) dessas combinações lineares⁽⁹⁾.

(8) V. Nota Anexa 2

(9) V. Doob, 1953, p.75. A variedade linear gerada pela sequência $Y_t = \{y_t, y_{t-1} \dots\}$ será o conjunto \mathcal{Y}_t com elemento genérico y expressável na forma $y = \sum a_i y_i$ com $y_i \in Y_t$.

Trata-se de uma variedade linear pois $y_1, y_2 \in \mathcal{Y} \Rightarrow (a_1 y_1 + a_2 y_2) \in \mathcal{Y}$, com a_1 e a_2 constantes reais quaisquer. O fecho de \mathcal{Y} incluirá $y = l.i.m. y_n$ com $y_n \in \mathcal{Y}$.

Exemplo 1.2 (Efeito de Slutski-Yule)

Combinações lineares, mesmo finitas, podem alterar radicalmente as propriedades dos processos. Admita-se a sucessão estocástica $\{\varepsilon_t\}$ com $E\varepsilon_t = 0$ e $E\varepsilon_s\varepsilon_t = \delta_{st}\sigma_\varepsilon^2$ (δ_{st} para $s \neq t$, $\delta_{st} = 1$ para $s = t$). Construa-se o processo de médias móveis $u_t = \sum_1^k a_i\varepsilon_{t-i}$.

$\{u_t\}$ pertence à variedade gerada por $\{\varepsilon_t\}$. No entanto, enquanto as variáveis do primeiro processo são não correlacionadas o mesmo não se passa com o processo de médias móveis, pois

$$Eu_s u_t = \sum_1^{n-k} a_i a_{i-k} \sigma_\varepsilon^2 \quad (\text{Cf. Kendall, 1973, p.41}).$$

Em rigor dever-se-ão distinguir estimadores ou preditores lineares de estimadores ou preditores afins, que serão operadores do tipo

$$\hat{x}_\tau = u(\tau) + \sum_i a_i y_i, \quad (1.9)$$

de manuseabilidade mais complexa.

Em tal situação o problema pode ser directamente enfrentado em termos de geometria e álgebra afim, como o fazem, por exemplo, Kruskal ou Drygas⁽¹⁰⁾, e que o próprio Malinvaud, o grande impulsionador da linguagem geométrica na econometria estatística, reconhece como "mais elegante" e "menos restritivo", mas também como "menos acessível" (1980, p.158).

O problema também pode ser enfrentado reduzindo a estimação afim à estimação linear, por intermédio de alguns artifícios elementares que se revelam plenamente satisfatórios.

Uma solução típica em econometria é a introdução de uma variável degenerada, identicamente igual à unidade, de forma a criar uma dimensão adicional que resolva o problema do termo independente determinístico.

Outra solução consiste na transformação das variáveis aleatórias originais em variáveis centradas⁽¹¹⁾, o que permitirá trabalhar sempre com vectores colocados na origem.

(10) Kruskal, W. (1961), "The coordinate free approach to Gauss-Markov estimation and its application to missing and extra observations", in J. Neyman, Ed., *Proc. of the Fourth Berkeley Symp. on Math. Stat. and Prob.*, vol. 1, University of California Press.

Drygas, H. (1970), *The Coordinate-free Approach to Gauss-Markov Estimate*, Springer-Verlag, Heidelberg.

(11) Dado x definir-se-á a v.a. centrada $x' := x - Ex$.

Será essa a abordagem adoptada inicialmente neste trabalho, pelo que todos os raciocínios, excepto quando expressamente notado, serão conduzidos a partir de variáveis com esperança matemática nula⁽¹²⁾.

Exemplo 1.3

Admita-se que os processos $\{x_t\}$ e $\{y_t\}$ estão relacionados pela equação

$$(i) \quad y_t = ax_t + b + \varepsilon_t \quad \text{com} \quad Ex_t = E\varepsilon_t = 0$$

Medindo y_t é admissível a estimativa $\hat{x}_t = y_t/a - b/a$, que é um filtro afim (numa observação). Escrevendo equivalentemente a equação (i)

$$(ii) \quad y_t - b = y'_t = ax_t + \varepsilon_t$$

ter-se-á o filtro linear $\hat{x}_t = y'_t/a$. E considerando a variável $u_t \equiv 1$ e escrevendo $\begin{bmatrix} y_t & u_t \end{bmatrix} \begin{bmatrix} 1 \\ -b \end{bmatrix} = ax_t + \varepsilon_t$, com o vector de observações $\mathbf{y}_t = [y_t \ u_t]$, ter-se-á também filtro linear em \mathbf{y}_t : $\hat{x}_t = [y_t \ u_t] \begin{bmatrix} 1/a \\ -b/a \end{bmatrix}$.

Dê-se agora um exemplo de estimação linear pelo método dos mínimos quadrados ordinário (OLS, "ordinary least squares"). O exemplo mostra como modelo econométrico pode ser introduzido no esquema (1.1) e como os mínimos quadrados permitem determinar valores dos coeficientes a_i para o filtro (1.8).

Exemplo 1.4 (OLS)

Considere-se a observação de dois processos reais $\{y_t\}$ e $\{z_t\}$ e admita-se que estão relacionados pela equação seguinte, em que ε_t é v.a. de natureza desconhecida e \mathbf{b} vector linha k-dimensional.

$$y_t = \mathbf{b}z_t + \varepsilon_t$$

Reescrever o processo na forma da equação (1.1) tem como única dificuldade a interpretação de \mathbf{b} e de \mathbf{z}_t . A sucessão $\{\mathbf{x}_t\}$ considerar-se-á constante, com $\mathbf{x}_t = \mathbf{b}$, pois esses serão os valores desconhecidos. A função $h(\mathbf{x}_t, t)$ será evidentemente linear, os seus coeficientes serão as variáveis \mathbf{z}_t , parâmetros conhecidos da função.

Se se procurar \mathbf{b} como sendo o vector de coeficientes de recta de \mathbb{R}^{k+1} que mais perto se encontra dos pontos (y_t, \mathbf{z}_t) , no sentido da regressão pelos mínimos quadrados de y_t sobre \mathbf{z}_t , e se se proceder aos rearranjos matriciais normais em econometria (V. Johnston, 1984, pp.90 e 112-3), encontrar-se-á

$$\hat{\mathbf{b}} = (Z^T Z)^{-1} Z^T Y$$

(12) *Mais rigorosamente: esperança matemática não condicional nula.*

Significa isso que cada componente b_i do vector \mathbf{b} é estimado como combinação linear das observações y_1, y_2, \dots, y_t , sendo coeficientes de ponderação — coeficientes a_i do filtro — os valores dados por $(Z^T Z)^{-1} Z^T$.

Note-se, a propósito, como a proposta dos mínimos quadrados para os coeficientes a_i do filtro pode ser apresentada sem qualquer pressuposto probabilístico sobre o modelo em estudo⁽¹³⁾. Trata-se daquilo a que Jazwinski chama *procedimento estatístico sem fundamentação probabilística*. Na secção seguinte discutir-se-á o problema dos critérios probabilísticos que determinam a optimalidade dos coeficientes do filtro.

(13) V. Malinvaud, 1980, cap. 1 "Econometrics without stochastic models" e ainda Jazwinski, 1970, secção 5.3 "Statistical methods".

2. Optimalidade na estimação — o problema de Gauss

A formulação clara e a solução probabilística do problema da estimação surge pela primeira vez com Gauss e L egendre⁽¹⁾.

J  anteriormente se tinha sentido a necessidade de combinar, numa  nica estima o de par metros, o resultado de observa es m ltiplas n o exactamente consistentes com o resultado determin stico formulado. Ainda no s culo XVIII⁽²⁾ s o de destacar os trabalhos de Boscovich e Laplace, que formularam um m todo de estima o dos coeficientes de uma recta.

A pretens o de Laplace de que o seu m todo conduziria   estima o do *valor mais prov vel* dos par metros levaria Gauss a estudar o problema de forma probabil stica e a propor um outro princ pio,

“princ pio prefer vel ao de Laplace, de acordo com o qual a soma de todas as diferen as deve ser zero, e a soma das mesmas diferen as, todas somadas com sinal positivo, deve ser m nima. Pode ser mostrado que isto n o   admiss vel na base do c lculo das probabilidades pois leva   contradic es”. Carta de Gauss a Olbers, 24 de Mar o de 1807 (Cf. Plackett, 1972).

Gauss come a por estudar uma forma admiss vel para a distribui o dos erros e deduz a lei normal⁽³⁾. Postulando a normalidade dos erros num modelo semelhante ao da equa o (1.1) deduz o m todo dos m nimos quadrados ou, em geral, o da minimiza o “da forma de qualquer das pot ncias com expoente par”⁽⁴⁾.

  curioso verificar como a formula o de Gauss sobre o “valor mais prov vel” prefigura o m todo da m xima verosimilhan a, desenvolvido por R. A. Fisher em 1912. Inclusivamente, Gauss maximiza a fun o de densidade conjunta calculando o logaritmo dessa fun o (V. Sorenson, 1970, p.65).

- (1) L egendre publica os seus resultados em 1805 em *Nouvelles M thodes pour la D termination des Orbites des Com tes* e Gauss publica resultados fundamentalmente id nticos em 1809, na *Theoria Motus Corporum Coelestium*. H , no entanto, suficiente evid ncia de Gauss ter chegado  s mesmas conclus es muito antes, em 1794-5. Para uma discuss o do problema e um relato da pol mica entre Gauss e L egendre sobre a prioridade da descoberta ver Plackett, 1972.
- (2) V. C. Eisenhart (1961).
- (3) Para uma clara exposi o da dedu o de Gauss ver Borel e Deltheil, 1923, pp.150-7.
- (4) *Theoria Motus*, Cf. Plackett, 1972, p.242.

Facto igualmente interessante, Gauss tinha clara consciência de que era a assumpção da normalidade dos erros que garantia que o método dos mínimos quadrados forneceria a *moda da função de verosimilhança*.

“Quando a lei do erro é desconhecida é *impossível* encontrar o resultado mais provável. . . Chegamos ao método dos mínimos quadrados, independentemente da lei dos erros, quando o número de observações é suficientemente largo. Mas, com um número moderado de observações, fica-se completamente no escuro se a lei do erro é desconhecida. . .”. Carta a Olbers, 22 de Fevereiro de 1819, (Cf. Plackett, 1972).

Alguns anos mais tarde, em 1839, é novamente Gauss quem coloca o problema da estimação em outra das bases teóricas fundamentalmente admissíveis, a da *minimização de uma forma funcional dos erros*. Pode mesmo dizer-se que é Gauss, numa carta a Bessel datada de 28 de Fevereiro desse ano, que pela primeira vez prefigura o ponto de vista da teoria da decisão. Vale a pena citar um extrato completo dessa carta.

“é menos importante determinar o valor do parâmetro para o qual a probabilidade é maior, ainda que infinitamente pequena, do que aquele valor no qual nos apoiando jogamos o menos desvantajoso jogo; ora, se f_a denota a probabilidade de a incógnita x ter o valor a , então é menos importante que f_a tenha um máximo do que $\int f_x F(x - a) dx$, tomado sobre todos os valores possíveis de x , seja um mínimo, aonde para F é escolhida uma função que seja sempre positiva e que seja sempre crescente com argumentos crescentes, de uma forma apropriada. Escolher o quadrado para esta função é puramente arbitário e essa arbitriedade está na natureza do tema. Não fossem conhecidas as largas vantagens da escolha do quadrado e qualquer outra função poderia ser escolhida. . .”. (Cf. Plackett, 1972).

É altura de interromper este curto mas esmagador intróito histórico e de tentar definir, numa notação moderna e apropriada ao problema em causa, os critérios de optimalidade da estimação.

No modelo geral pressuposto pela equação (1.1)

$$y_t = h(x_t, t) + v_t,$$

o problema da estimação pode, em geral, ser abordado probabilisticamente por uma de duas vias: ou bem se pretende uma estimativa do valor mais provável de x_t ou então se tem em vista uma estimativa que minimize alguma função dos erros provavelmente cometidos.

Recorde-se que está disponível uma sequência de observações Y_t , e que se admite conhecida a forma funcional de $h(x_t, t)$.

Para adoptar um ponto de vista probabilístico é necessário estabelecer alguns pressupostos, e.g. uma forma definida para a distribuição das perturbações v_t . Chegando ao conhecimento da *distribuição de probabilidade condicional* de x_t dado Y_t , a qual se designará por

$$p(x_t, t|Y_t), \quad (2.1)$$

está construída uma base para a solução completa do problema. Isso acontece pois $p(x_t, t|Y_t)$ condensa toda a informação estatística sobre x_t que está contida nas observações registadas, e x_t inclui todas as incógnitas relevantes ⁽⁵⁾.

Similarmente,

$$p(x_\tau, t|Y_t), \quad \tau > t, \quad (2.2)$$

que é a *distribuição preditiva* de x_τ , feita no momento t e condicionada pelo conhecimento das observações, fornece uma base para a solução completa do problema preditivo.

Neste quadro probabilístico chamar-se-á ESTIMACÃO MODAL à maximização de (1.4), o que equivale a encontrar a moda dessa distribuição condicionada. De forma semelhante, e com referência a (1.5), se define PREVISÃO MODAL. Em ambos os casos se adopta a metodologia da máxima verosimilhança.

Vale a pena referir ainda a *estimativa de máxima verosimilhança conjunta*, que será a moda de

$$p(x_\tau, x_{\tau+1}, \dots, x_{\tau+k}; t|Y_t), \quad (2.3)$$

⁽⁵⁾ Este ponto de vista é compatível com o princípio de verosimilhança, V. Murteira, 1983, p.21. A justificação da aplicabilidade do princípio ao problema preditivo segue do raciocínio desenvolvido por Berger, 1985, p.32.

e que não coincide, em geral, com os valores obtidos a partir de (2.1). Ou seja, nada garante que as estimativas conjuntas (2.3), chamadas de *trajetória modal* (Jazwinski, 1970, p.156) contêm as modas das funções de verosimilhança particulares.

Uma outra abordagem probabilística é a adotada quando explicitamente se considera um critério de comparação dos erros das estimativas.

Seja \hat{x}_r uma estimativa ou previsão de x_r condicionada por Y_t . O erro

$$e := x_r - \hat{x}_r,$$

também denotado por \tilde{x}_r , é uma variável aleatória, pelo que se terá de introduzir alguma medida estatística.

A medida mais conhecida é a do *erro quadrático médio*, que se define por

$$EQM = E[(x_r - \hat{x}_r)^T (x_r - \hat{x}_r)], \quad (2.5)$$

mas, tal como afirmava Gauss, esta escolha é "puramente arbitrária", podendo-se substituir por qualquer outra "forma apropriada". Por outro lado, a convexidade desta função é "particularly disturbing" (Berger, 1985, p.61), penalizando demasiado fortemente os erros maiores.

Convirá então poder escolher outras classes de funções de penalização para medir estatisticamente os erros prováveis. O procedimento normal consiste em introduzir uma *função de perda* apropriada, $L(\cdot)$, e procurar minimizar a *perca esperada*:

$$E[L(e)] = E\{E[L(e)|Y_t]\}$$

Como é visível $E[\cdot]$ depende apenas de Y_t , não de \hat{x}_t , pelo que a minimização da perca esperada, EL , é equivalente à minimização de

$$E[L(e)|Y_t]. \quad (2.6)$$

Para a solução geral do problema conhecem-se resultados aplicáveis a duas grandes classes de funções de perca: funções pares e funções quadráticas.

Comece-se pelo primeiro tipo de funções iniciando o estudo com o caso escalar.

FUNÇÃO DE PERCA PAR⁽⁶⁾ é função de forma:

$$L(0) = 0, \quad (2.7')$$

$$e_1 \geq e_2 \geq 0 \Rightarrow L(e_1) \geq L(e_2) \geq 0, \quad (2.7'')$$

$$L(e) = L(-e). \quad (2.7''')$$

Um conhecido teorema de Sherman afirma que, se a distribuição da variável em causa, x_t , for "simétrica" em relação ao seu valor esperado, seja ele \bar{x}_t , e se for unimodal, então é esse valor \bar{x}_t que constitui a estimativa que minimiza a perca esperada.

Estando em causa a distribuição condicional $p(x_t|Y_t)$ e o valor esperado condicional $E[x_t|Y_t]$ que se representará igualmente por \bar{x}_t , os dados do problema não se alteram, pois a perca esperada é também condicionada pelos valores Y_t ⁽⁷⁾.

A "simetria" da distribuição em torno do seu valor esperado e a unimodalidade traduzem-se matematicamente pelas seguintes condições

(i) A função da distribuição $F(x)$ é SIMÉTRICA em torno do valor esperado \bar{x} :
$$F(x - \bar{x}) = 1 - F(\bar{x} - x), \quad (2.8)$$

(ii) A função da distribuição é CONVEXA para $x \leq \bar{x}$:

$$F[\lambda x_1 + (1 - \lambda)x_2] \leq \lambda F(x_1) + (1 - \lambda)F(x_2) \quad (2.9)$$

para $x_1, x_2 \leq \bar{x}$, $0 \leq \lambda \leq 1$.

Repare-se que (2.8) implica que a função de densidade da variável centrada, caso exista, seja par:

$$f_{x-\bar{x}}(x - \bar{x}) = f_{x-\bar{x}}(\bar{x} - x), \quad (2.10)$$

enquanto (2.9) implica que $F(x)$ seja contínua e diferenciável (V. Rudin, 1974, pp.62-3), o que acarreta a existência de $f(x)$ e o seu carácter monótono crescente no intervalo $]-\infty, \bar{x}]$. Estas propriedades implicam a "simetria" de $f(x)$, mais rigorosamente, a verificação de (2.10), e a sua unimodalidade.

No caso vectorial o problema pode ser estudado componente a componente, abordagem teoricamente justificável numa perspectiva decisional⁽⁸⁾.

(6) Pensa-se que a formulação de Jazwinski (1970, p.147) não está correcta. Segue-se de perto Kalman (1960, p.37). Ver a discussão na Nota Anexa 3.

(7) V. Doob, 1953, pp.35-7.

(8) Berger, 1985, pp.68-9.

Estabelecidas estas condições apresente-se o teorema de Sherman de forma directamente aplicável ao problema em estudo⁽⁹⁾.

Teorema 2.1 (Sherman)

Seja \mathbf{x}_t vector aleatório em que cada componente x_{it} tem função de distribuição condicionada que é simétrica em torno da média $\bar{x}_{it} = E[x_{it}|Y_t]$ e convexa para $x_{it} \leq \bar{x}_{it}$.

Seja $L(\cdot)$ função de perca par.

Então, o vector aleatório que minimiza, componente a componente, a perca esperada é o valor esperado condicional: $\mathbf{x}_t = E[\mathbf{x}_t|Y_t]$.

Demonstração: S. Sherman (1958), "Non-mean-square error criteria", *Trans. IRE Prof. Group Inf. Theory*, IT-4, pp.125-6, ou C.W.J. Granger (1969, p.206).

Os resultados deste teorema são suficientemente gerais para abranger grande número de casos com interesse. Ficam de fora as situações em que a função de perca não é par ou em que a distribuição condicional não é simétrica.

Passe-se agora à segunda grande classe de funções de perca, para as quais nenhum pressuposto sobre a forma da distribuição condicional é necessário.

Trata-se do caso conhecido dos estimadores de variância mínima, ou de erro quadrático médio mínimo, que se definirão de forma mais geral, embora mantendo essa designação. Não será necessário distinguir variáveis aleatórias reais de vectores aleatórios, pois a minimização é feita tanto componente a componente como aplicando a função de perca sobre o vector aleatório.

FUNÇÃO DE PERCA QUADRÁTICA é função da forma

$$L(\mathbf{e}) = \mathbf{e}^T \mathbf{S} \mathbf{e}, \quad (2.11)$$

em que \mathbf{S} é matriz semidefinida positiva.

A definição de *estimadores de variância mínima* generaliza-se àqueles que minimizem uma função de perca quadrática. O teorema que segue é suficientemente conhecido no caso simples de $L(\mathbf{e}) = \mathbf{e}^2$. Apresenta-se a demonstração no caso mais geral⁽¹⁰⁾ e apropriado ao estudo em causa.

⁽⁹⁾ Cf. Kalman, 1960, p.37.

⁽¹⁰⁾ V. Kalman, 1963, p.296.

Teorema 2.2

Seja \mathbf{x}_t v.a. de segunda ordem, $E\mathbf{x}_t\mathbf{x}_t^\top < \infty$,⁽¹¹⁾ com valor esperado condicional $\bar{\mathbf{x}}_t = E[\mathbf{x}_t|Y_t]$. Seja o estimador de \mathbf{x}_t , $\hat{\mathbf{x}}_t$, uma funcional em Y_t .

Então:

- (i) $\hat{\mathbf{x}}_t := \mathbf{x}_t$ é estimador centrado
- (ii) $\hat{\mathbf{x}}_t := \bar{\mathbf{x}}_t$ é estimador de variância mínima.

Demonstração: omitam-se os índices para simplificar a notação.

$$(i) E[\mathbf{x} - \bar{\mathbf{x}}] = E\mathbf{x} - E_y E_x[\mathbf{x}|Y] = 0$$

(ii) $EL = E\{E[(\mathbf{x} - \hat{\mathbf{x}})^\top S(\mathbf{x} - \hat{\mathbf{x}})|Y]\}$, aonde $\hat{\mathbf{x}}$ designa qualquer funcional de Y tal que $E\hat{\mathbf{x}}^\top \mathbf{x} < \infty$. Reescreva-se a expressão da perda esperada

$$EL(\mathbf{e}) = E\{E[(\mathbf{x} - \bar{\mathbf{x}} + \bar{\mathbf{x}} - \hat{\mathbf{x}})^\top S(\mathbf{x} - \bar{\mathbf{x}} + \bar{\mathbf{x}} - \hat{\mathbf{x}})|Y]\}$$

Como $\bar{\mathbf{x}}$ e $\hat{\mathbf{x}}$ são invariantes ao operador $E[-|Y]$

$$E[(\mathbf{x} - \bar{\mathbf{x}})^\top S(\bar{\mathbf{x}} - \hat{\mathbf{x}})|Y] = E[\mathbf{x} - \bar{\mathbf{x}}|Y]^\top S(\bar{\mathbf{x}} - \hat{\mathbf{x}}) = (\mathbf{x} - \bar{\mathbf{x}})^\top S(\bar{\mathbf{x}} - \hat{\mathbf{x}}) = 0$$

Por outro lado, $E[(\mathbf{x} - \bar{\mathbf{x}})^\top S(\mathbf{x} - \bar{\mathbf{x}})]$ é independente de $\hat{\mathbf{x}}$, pelo que

$$EL(\mathbf{e}) = \text{const} + E[(\bar{\mathbf{x}} - \hat{\mathbf{x}})^\top S(\bar{\mathbf{x}} - \hat{\mathbf{x}})]$$

Como S é semidefinida positiva a expressão é evidentemente minimizada com $\hat{\mathbf{x}} = \bar{\mathbf{x}}$.

As limitações e a generalidade dos teoremas 2.1 e 2.2 irão ser discutidas nas próximas secções. Na secção seguinte, em particular, argumentar-se-á a favor da generalidade dos resultados aqui introduzidos.

De qualquer forma, torna-se claro que o primeiro problema na procura de estimadores e preditores é a determinação da função de densidade condicional $p(\mathbf{x}_t, t|Y_t)$ ou de alguns dos seus parâmetros. Com esse conhecimento a estimativa óptima pode ser calculada, ou numa perspectiva modal ou para uma função de perda que seja introduzida.

Estando em causa a minimização de uma perda esperada o problema básico será a determinação do *valor esperado condicional*, estimador óptimo nos casos usualmente considerados.

Finalmente, interessará uma medida da qualidade da estimativa, medida usualmente fornecida pela **MATRIZ DE COVARIÂNCIA CONDICIONAL**

(11) V. Nota Anexa 4.

$$\Sigma_{r|t} = E[(\mathbf{x}_r - \hat{\mathbf{x}}_r)(\mathbf{x}_r - \hat{\mathbf{x}}_r)^\top | Y_t] \quad (2.12)$$

Através da matriz $\Sigma_{r|t}$ encontra-se uma medida da qualidade esperada da estimativa, o que é tão importante na abordagem probabilística como o próprio cálculo do valor estimado ou previsto:

$$E[(\mathbf{x}_r - \hat{\mathbf{x}}_r)^\top (\mathbf{x}_r - \hat{\mathbf{x}}_r)] = \text{tr} E \Sigma_{r|t} \quad (2.13)$$

O raciocínio desenvolvido nestes últimos parágrafos, e que toma como ponto de partida a determinação de $p(\mathbf{x}_r, t | Y_t)$ ou de alguns dos seus parâmetros, é o essencial daquilo a que se chamará *abordagem probabilística*.

3. Processos gaussianos e não gaussianos

Em processos gaussianos as condições de optimalidade dos estimadores ampliam-se consideravelmente, ou seja, os estimadores serão óptimos sob restrições mais fracas. Por outro lado, a assumpção da normalidade permite simplificar e generalizar um conjunto de deduções e resultados.

Seguindo a linha de pensamento proposta por Doob (1953, pp.71-8) organizar-se-á o estudo operando a distinção genérica entre processos gaussianos e não gaussianos. Assim se evitará futuramente, resultado a resultado, referir as condições de aplicabilidade a um e a outro tipo de processos, ficando tais condições à partida clarificadas.

Vale a pena começar por dedicar um pouco de atenção aos processos gaussianos. Alguns resultados gerais serão expressos sinteticamente na forma de lemas. As conclusões particulares que interessam para o problema em estudo serão apresentadas sob a forma de teoremas. Trata-se, na maioria dos casos, de proposições suficientemente conhecidas. A sua apresentação vale como passo na linha do raciocínio e também como base para referência em secções futuras. As demonstrações serão, muitas vezes, simplesmente delineadas.

Começar-se-á por definir processos gaussianos⁽¹⁾.

O processo estocástico $\{x_t\}$ é PROCESSO GAUSSIANO se qualquer conjunto finito de variáveis do processo seguir uma distribuição multinormal

A definição mantém-se se as variáveis do processo forem vectoriais. Nesse caso não só uma variável particular segue uma multinormal como o mesmo se passa com qualquer vector alargado, formado com um conjunto de variáveis em causa.

Recordam-se os seguintes resultados gerais da multinormal.

Lema 3.1

Se o vector n -dimensional x segue uma distribuição multinormal, com valor médio \bar{x} e matriz de covariância Σ , então,

(i) a distribuição de x fica completamente determinada pela definição do valor médio e da matriz de covariância;

(ii) sendo A matriz real $p \times m$, o vector p -dimensional $y = Ax$ segue uma multinormal com valor médio $A\bar{x}$ e matriz de covariância $A\Sigma A^T$;

(1) V. Nota Anexa 5.

(iii) em particular, qualquer distribuição marginal de um componente de \mathbf{x} , seja x_i , segue uma normal com valor médio \bar{x}_i e variância σ_{ii} , elemento diagonal da matriz Σ ;

(iv) Também qualquer sub-vector de \mathbf{x} , segue uma distribuição marginal multinormal;

(v) duas variáveis aleatórias x_i e x_j , componentes de \mathbf{x} , são independentes se e só se a sua covariância for nula, $\sigma_{ij} = 0$;

(vi) em geral, dois sub-vectores aleatórios \mathbf{x}_1 e \mathbf{x}_2 são independentes se e só se a sua matriz de covariância for diagonal por blocos, sendo $\Sigma_{12} = \Sigma_{21} = 0$.

Demonstração: V. Kendall e Stuart, 1987, vol I, §15.4 ou Rohatgi, 1976, p.234.

Admitir-se-á naturalmente

Teorema 3.1

Se o processo $\{y_t\}$ é gaussiano o resultado de qualquer filtro linear ou afim sobre ele aplicado é ainda um processo gaussiano.

Demonstração: No caso de o filtro ser finito, ou seja, resultar na combinação linear de um número finito de variáveis, a dedução é imediata pela aplicação de (ii) do lema precedente.

No caso infinito o teorema só é válido se $E[(\mathcal{F}Y_t)(\mathcal{F}Y_t)^T] < +\infty$ ⁽²⁾, caso em que todas as combinações lineares são gaussianas e o limite, existindo, é também gaussiano (V. Doob, 1953, pp.72-3 e 390).

É importante referir a forma da distribuição condicional de uma v.a. gaussiana.

(2) V. Nota Anexa 6.

Lema 3.2

Dado o vector aleatório $[x^T y^T]^T$ com valor médio e matriz de covariância representadas de acordo com essa partição

$$\begin{bmatrix} x \\ y \end{bmatrix} \sim N \left(\begin{bmatrix} \bar{x} \\ \bar{y} \end{bmatrix}, \begin{bmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{yx} & \Sigma_{yy} \end{bmatrix} \right)$$

a distribuição condicional $p(x|y)$ é multinormal, com valor médio e matriz de covariância, respectivamente

$$\bar{x} + \Sigma_{xy} \Sigma_{yy}^{-1} (y - \bar{y}), \quad \Sigma_{xx} - \Sigma_{xy} \Sigma_{yy}^{-1} \Sigma_{xy}^T. \quad (3.1)$$

inversamente, se a distribuição condicional for assim determinada, a distribuição conjunta é dada pela expressão anterior.

Demonstração: No pressuposto de Σ_{yy}^{-1} existir e de os vectores aleatórios terem dimensão finita veja-se e.g. Anderson, 1958, pp.28-9. Caso a assinalada inversa não exista, o teorema continua válido utilizando uma pseudo-inversa de Penrose⁽³⁾; veja-se Kalman, 1963, pp.381-3. Caso o vector não tenha dimensão finita o resultado mantém-se tomando x' e y' , sub-vectores finitos, arbitrários, de x e y ; veja-se Doob, 1953, p.390.

No caso finito ou, no caso infinito, admitindo a existência da v.a. em causa, deduzir-se-ão imediatamente os teoremas seguintes.

Teorema 3.2

Se os processos $\{y_t\}$ e $\{x_t\}$ são conjuntamente gaussianos a distribuição condicional $p(x_t, t|Y_t)$ é multinormal, com valor médio e covariância conforme (3.1.).

Teorema 3.3

Sendo conjuntamente gaussianos os processos $\{y_t\}$ e $\{x_t\}$ o estimador modal de máxima verosimilhança coincide com o estimador de perca mínima, desde que a função de perca seja par. Em particular, esse estimador único é o estimador de variância mínima.

(3) V. Nota Anexa 7.

Demonstração: Basta notar que a moda de $p(x_r, t|Y_t)$ coincide com o valor médio condicionado, uma vez que a distribuição é multinormal.

Significa isto que, no caso gaussiano, é indistinta a maximização da função de verosimilhança ou a escolha de uma função de perda par. O valor esperado condicionado é a estimativa óptima sob um largo leque de critérios de optimalidade. Por outro lado, e ainda como consequência do lema 3.2, a média de x_r condicionada por Y_t é função linear ou afim da sequência de observações. Então,

Teorema 3.4

Se os processos $\{x_t\}$ e $\{y_t\}$ são conjuntamente gaussianos o melhor estimador de x_r (modal, de variância mínima ou com qualquer função de perda par) e que é funcional nas observações Y_t é um filtro afim.

Corolário

O filtro afim do teorema 3.4 é filtro linear se $Ex_t = Ey_t = 0$.

Talvez seja elucidativo concretizar estes resultados num exemplo simples.

Exemplo 3.1

Considerem-se os processos $\{x_t\}, \{y_t\}$ com x_t e y_t conjuntamente binormais. Admita-se $Ex_t x_s = \delta_{ts} \sigma_x^2$, $Ey_t y_s = \delta_{ts} \sigma_y^2$, $Ex_t y_s = \delta_{ts} \sigma_{xy}$, $\forall t, s$. Além disso $Ex_t = \bar{x}$, $Ey_t = \bar{y}$.

Evidentemente,

$$E[x_t | y_t] = (\bar{x} - \frac{\sigma_{xy}}{\sigma_y^2} \bar{y}) + \frac{\sigma_{xy}}{\sigma_y^2} y_t$$

que é um filtro afim na observação concomitante y_t ; tem como termo independente, determinístico, a expressão entre parêntesis.

Centrem-se agora as variáveis fazendo $x'_t = x_t - \bar{x}$ e $y'_t = y_t - \bar{y}$. Ter-se-á $Ex'_t = Ey'_t = 0$ e o filtro será linear:

$$E[x'_t | Y'_t] = \frac{\sigma_{xy}}{\sigma_y^2} y'_t$$

Todos estes resultados destacam a especial importância dos processos gaussianos. Para a teoria da estimação e decisão essa importância é ainda maior pelo facto de ser possível estabelecer uma dualidade entre quaisquer processos de segunda ordem e processos gaussianos.

A possibilidade e sentido dessa correspondência tornar-se-á clara pelos teoremas seguintes.

Teorema 3.5

Com $t \in T$, T finito ou infinito, sejam dadas as funções

(i) $\mu(t)$ função qualquer com $t \in T$,

(ii) $\sigma(s, t)$ função com $s, t \in T$, tal que

$$\sigma(s, t) = \sigma(t, s)$$

e tal que a matriz quadrada $[\sigma(i, j)]$, com i e j pertencentes a qualquer subconjunto finito de T , seja semidefinida positiva.

Há sempre um processo gaussiano $\{x_t, t \in T\}$ satisfazendo

$$Ex_t = \mu(t)$$

$$Ex_s x_t - \mu(s)\mu(t) = \sigma(s, t)$$

Demonstração: V. Doob, 1953, pp.72-3.

Teorema 3.6

Seja $\{y_t\}$ um processo estocástico de segunda ordem: $Ey_t^2 < \infty$ $t \in T$

Há sempre um correspondente processo gaussiano, com o mesmo domínio do conjunto de parâmetros mas definido num outro espaço de probabilidade, cujas variáveis aleatórias $\{x_t\}$ satisfazem as equações

(i) $Ex_t = 0$.

(ii) $Ex_s x_t = Ey_s y_t$.

Demonstração: Basta notar que é sempre possível definir $r(s, t) = Ey_s y_t$ pelo que $r(s, t) = r(t, s)$; por outro lado, com qualquer sequência de números reais $a_1, a_2 \dots a_n$, $\sum_{s=1}^n \sum_{t=1}^n r(s, t) a_s a_t = \sum_{s=1}^n \sum_{t=1}^n Ey_s y_t \cdot a_s a_t = \sum_{s=1}^n E[y_s a_s]^2 \geq 0$, pelo que a matriz $[r(s, t)]$ é definida positiva. Acrescentando $\mu(t) \equiv 0$ as hipóteses do teorema 3.5 estão satisfeitas.

Corolário

Dado um processo estocástico de segunda ordem e a correspondente variedade fechada, há sempre uma variedade gaussiana fechada e uma transformação entre as duas variedades que, para cada t , é invariante face ao produto interno.

Demonstração: A extensão do teorema as combinações lineares finitas mantém obviamente (i) e (ii); o caso geral deriva da aplicação de limites, possível por se tratar de v.a. de segunda ordem.

Estes resultados justificam, rigorosamente, em processos de que se desconhece a lei probabilística, a prática de assumir o carácter gaussiano desses processos e de admitir a optimalidade das estimativas nas correspondentes condições mais restritas. Ao contrário do teorema de limite central, que não é válido em observações únicas ou em número limitado, trata-se de um raciocínio directamente aplicável a sucessões cronológicas.

Em concreto, sabe-se que o melhor filtro, em processos gaussianos, é um filtro linear ou afim, cuja computação envolve apenas os momentos até à segunda ordem. Por outro lado, com função de perda quadrática e um filtro linear ou afim, também estão apenas em causa as características até à segunda ordem. Há uma "distância esperada", $d(\hat{x}_t, x_t)$, que é redutível à variância do erro do estimador linear. Assim, se se passar da variedade linear à variedade gaussiana dual e se aí se procurar o estimador com menor perda esperada, ao se regressar à variedade original as características de segunda ordem obviamente mantêm-se, uma vez que apenas estiveram em causa, na variedade gaussiana, transformações lineares. Na variedade gaussiana os cálculos envolveram apenas as médias e matrizes de covariância — o correspondente resultado é válido para o processo dual deste, desde que o filtro se restrinja a ser da classe do operado na variedade gaussiana: linear ou afim.

Então, com essa restrição sobre o filtro, o estimador encontrado no processo gaussiano será também o melhor estimador, para a mesma função de perda, e relativo a todos os processos não gaussianos que tenham o mesmo valor médio e mesma covariância.

A verdadeira restrição encontra-se no carácter do filtro, *o papel desempenhado pela função de perda é secundário*. De facto, por simples inspecção dos teoremas 2.1 e 2.2 verifica-se que o estimador óptimo é sempre o valor esperado condicional. Significa isso, sob os pressupostos da distribuição no teorema 2.1, que o estimador óptimo para uma função de perda par é o estimador óptimo para qualquer outra função de perda par.

Em particular

Teorema 3.7

O estimador afim de variância mínima (3.1)

$$\bar{x} + \Sigma_{xy} \Sigma_{yy}^{-1} (y - \bar{y}),$$

é o melhor estimador afim para todos os processos de segunda ordem e para todas as funções de perda par, sob os pressupostos de distribuição estabelecidos no teorema 2.1

Demonstração: Segue das considerações precedentes (V. Kalman, 1963, p.297). O teorema 4.1 fornece prova equivalente.

Reconhece-se assim o infundamentado de muitos "blunt statements" (Kalman, 1963, p.297) sobre a dificuldade de manejo de funções de perda não quadráticas. Como o afirma Berger "in many statistical problems for which a loss symmetric... is suitable, the exact functional form of the loss is not crucial to the conclusion. Squared-error loss may then be a useful approximation". Em particular, quando a informação existente é relativamente precisa, a perda "real" pode ser bem aproximada por uma perda quadrática (Berger, 1985, p.61). Convirá então construir conceitos latos, cuja expressão estrita, no caso gaussiano, reproduza os estimadores de variância mínima, e que, em geral, permitam operar directamente em quaisquer processos, de distribuição não definida, de forma a chegar de maneira directa aos estimadores lineares correspondentes.

Nessa ordem de ideias o mais importante conceito é o de ESPERANÇA CONDICIONAL EM SENTIDO LATO⁽⁴⁾, que se denotará por $\hat{E}[x_r|Y_t]$, em que Y_t designa a sequência de variáveis condicionantes, e que se define como sendo \hat{x}_r pertencente à variedade gerada pelas observações: y_t , portanto da forma $\hat{x} = \Sigma a_i y_i$ e que minimiza $E[x_r - \hat{x}_r]^2$. Formalmente,

$$\hat{E}[x_r|Y_t] = \hat{x}_r : \min E[x_r - \hat{x}_r]^2, \quad \text{com } \hat{x}_r \in Y_t, \quad (3.2)$$

É pois admissível a notação $\hat{E}[x_r|y_t]$ também por vezes utilizada.

O operador $\hat{E}[-|-]$, definido em termos de variáveis centradas, é um filtro linear nas variáveis condicionantes, ou em outras que, equivalentemente, gerem y_t :

$$\hat{E}[x_r|Y_t] = \sum_i a_i^* y_i = FY_t, \quad (3.3)$$

sendo os coeficientes a_i^* os que resolvem o problema de minimização (3.2).

Se se abandonar o pressuposto de $\{y_t\}$ ser processo centrado, passando a verificar-se $Ey_t = \bar{y}_t$, a esperança condicional em sentido lato definir-se-á a partir das variáveis $y_t^* = y_t - \bar{y}_t$ que constituem a colecção Y_t^* :

$$\hat{E}[x_r|Y_t] := \hat{E}[x_r|Y_t^*] = \sum_i a_i^* (y_t - \bar{y}_t) \quad (3.4)$$

Se, além disso, $E x_r \neq 0$, tem todo o sentido definir o estimador

(4) V. Doob, 1953, p.155 ou Loève, 1963, p.462

$$\hat{x}_r = E x_r + \hat{E}[x_r|Y_t] = E x_r + \sum_i a_i^*(y_i - \bar{y}_i) \quad (3.5)$$

Será fácil provar que $E[-|-]$ e $\hat{E}[-|-]$ satisfazem equações funcionais idênticas (V. Doob, 1953, p.155). Nomeadamente,

Teorema 3.8

Sejam dados dois processos aleatórios de segunda ordem centrados, $\{x_t\}$ e $\{y_t\}$. Represente-se por Y uma colecção de variáveis y_i e por a uma constante real qualquer. São válidas, com probabilidade 1, as seguintes relações

$$(i) \quad \hat{E}[ax|Y] = a \cdot \hat{E}[x|Y] \quad (3.6)$$

$$(ii) \quad \hat{E}[x_1 + x_2|Y] = \hat{E}[x_1|Y] + \hat{E}[x_2|Y] \quad (3.7)$$

E, se $Y_1 \subset Y_2$

$$(iii) \quad \hat{E}[\hat{E}(x|Y_1)|Y_2] = \hat{E}[x|Y_2] \quad (3.8)$$

Finalmente, se $E y_1 y_2 = 0$, $\forall y_1 \in Y_1$ e $\forall y_2 \in Y_2$

$$(iv) \quad \hat{E}[x|Y_1, Y_2] = \hat{E}[x|Y_1] + \hat{E}[x|Y_2] \quad (3.9)$$

Demonstração: (i) e (ii) são triviais, seguem da linearidade do filtro (3.3), por exemplo. Para provar (iii) note-se que $x - \hat{E}[x|Y_1]$ é não correlacionado com qualquer elemento de Y_1 , pelo que o seu condicionamento lato por $Y_2 \subset Y_1$ é nulo: $\hat{E}[x - \hat{E}(x|Y_1)|Y_2] = 0$, o que é análogo a (iii). A prova de (iv) pode fazer-se directamente a partir de (3.1), V. Anderson e Moore, 1979, p.94, ou, alternativamente, notando o facto de $E y_{1i} y_{2j} = 0 \quad \forall y_{1i} \in Y_1, \forall y_{2j} \in Y_2$ implicar que o condicionamento lato, pela variedade conjuntamente gerada por Y_1, Y_2 , seja decomponível no problema de minimização da soma $E[x - \sum a_i y_{1i}]^2 + E[x - \sum b_j y_{2j}]^2$.

Com esta notação encontra-se uma dualidade entre a FUNÇÃO DE REGRESSÃO NO CASO GAUSSIANO,

$$\phi(Y_t) = E[x_t|Y_t], \quad (3.10)$$

e a função de regressão lata⁽⁵⁾

$$\hat{\phi}(Y_t) = \hat{E}[x_t|Y_t]. \quad (3.11)$$

e é ainda possível indicar fórmulas explícitas para essas funções. Em virtude da dualidade estabelecida, essas fórmulas terão expressão idêntica às existentes para o caso gaussiano — (3.1).

Mas vale a pena generalizar já estes resultados ao caso de processos constituídos por vectores aleatórios.

Com as óbvias modificações tudo se mantém válido. A única *nuance* conceptual consiste em entender o operador $\hat{E}[-| -]$ como aplicado a *cada componente* x_i de x , que será condicionado pela variedade gerada por *todos os componentes* dos condicionantes y . O problema (3.2) passa a escrever-se

$$\hat{E}[x_r|Y_t] = \hat{x}_r : \quad \min E[(x_r - \hat{x}_r)^\top (x_r - \hat{x}_r)] \quad (3.12)$$

com cada $x_i \in y_t$. Que a solução do problema componente a componente minimiza o erro quadrático indicado é facto evidente. Basta verificar que, se o faz componente a componente, também o fará no somatório de componentes não negativos a minimizar. As equações (3.3) a (3.11) são pois *generalizáveis a vectores aleatórios*.

É altura de aplicar estes conceitos num exemplo simples que servirá, simultaneamente, de introdução à abordagem geométrica desenvolvida na secção seguinte.

Exemplo 3.2 (Equações normais do OLS)

Considere-se a relação funcional determinística

$$x_t = z_t b$$

em que x_t é escalar, z_t vector-linha e b vector coluna de parâmetros, ambos k -dimensionais. Admita-se que as observações de x_t são feitas com um erro ε_t , de carácter desconhecido, registando-se na realidade $y_t = z_t b + \varepsilon_t$

(5) V. Priestley, 1981, p.74. Murteira (1979, vol I, p.224) denomina (3.10) como regressão de tipo I e (3.11) como regressão de tipo II (linear).

Agrupem-se as n observações no vector Y e na matriz Z .

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \dots \\ y_n \end{bmatrix} \quad Z = \begin{bmatrix} z_1 \\ z_2 \\ \dots \\ z_n \end{bmatrix}$$

Pretende-se uma estimativa para $X = [x_1 x_2 \dots x_n]^T$, problema que é equivalente ao de encontrar uma estimativa para b . Existindo os erros ε_t a equação matricial

$$X = Zb$$

é inconsistente, ou seja, X não pertence à variedade linear gerada pelas colunas de Z .

O problema pode ser resolvido de forma puramente determinística apenas com recurso à álgebra linear (Cf. Strang, 1976, pp.104-10). Uma linha de ataque consiste em procurar o vector da variedade gerada pelas colunas de Z , seja \hat{X} , que mais perto se encontra do vector Y . Ou seja, trata-se de procurar aquele $\hat{X} = Zb$ tal que o comprimento ou norma do vector $e = X - \hat{X}$ seja mínimo. Isso equivale, com o produto escalar, a norma e a distância euclídeana usuais, a exigir que e seja perpendicular à variedade gerada pelas colunas de Z , ou seja, a exigir que

$$\lambda Z \perp e$$

Tal condição é equivalente à seguinte

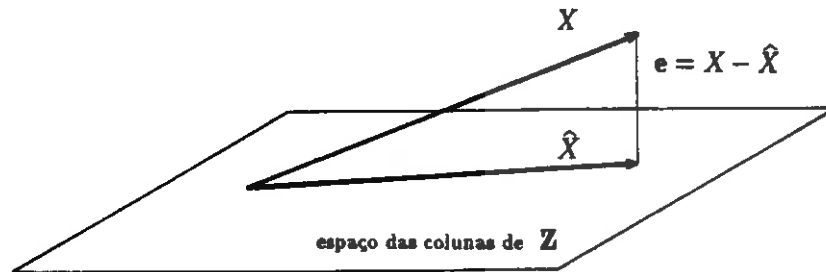
$$\lambda^T Z^T \cdot e = \lambda^T Z^T (X - Zb) = 0$$

Como λ é vector não-nulo arbitrário encontra-se

$$Z^T X = Z^T Zb$$

que é pura e simplesmente a equação normal do método dos mínimos quadrados. Admitindo que $r(Z) = k$

daí resultará $\hat{b} = (Z^T Z)^{-1} Z^T X$ e $\hat{X} = Zb$. A estimativa \hat{X} resulta da projecção ortogonal de X no espaço das colunas de Z .



Admita-se agora que y_t e ε_t são variáveis aleatórias de acordo com as hipóteses usuais no OLS. A estimativa $\hat{X} = \hat{E}[X|Y]$ correspondente, em processo gaussiano, a $\hat{X} = E[X|Y]$, é também o vector pertencente à variedade \mathcal{Y} que minimiza $E(e^T e) = E[(X - \hat{X})^T (X - \hat{X})]$. Pela aplicação directa do conhecido teorema de Gauss-Markov, chega-se igualmente às equações normais.

A terminar esta secção pensa-se ter deixado clarificadas as correspondências que são necessárias ao tratamento da estimação de forma muito geral. Daqui em diante não terá sentido proceder à distinção sistemática entre processos gaussianos e não gaussianos nem será necessário confrontar constantemente estimadores de perca quadrática mínima com estimadores *lineares* de perca quadrática mínima. Entende-se que os resultados são válidos, em sentido lato, restringindo a classe dos estimadores a filtros lineares ou afins, e que são válidos, em sentido estrito, no caso de os processos serem gaussianos.

4. Filtros lineares e projectores ortogonais – solução de Wiener-Kolmogorov

A determinação de forma funcional que permite obter a estimativa \hat{x}_r a partir de Y_t , ao ser expressa como a procura daquela variável que minimiza $EL(x_r - \hat{x}_r)$, terá uma interpretação geométrica clara se puder ser entendida como a procura do vector pertencente a um espaço funcional em Y_t , seja y_t , que minimiza uma distância esperada $d(x_r, Y_t)$.

Isso implica que $EL(x_r - \hat{x}_r)$ possa ser entendido como uma norma, ou distância, no espaço (x_r, y_t) . O conceito tornar-se-á operacional se essa norma for compatível com o produto escalar definido nesse espaço.

Não se conhecem estudos genéricos do problema mas apenas o estudo em determinados espaços lineares, com determinadas definições métricas e de produto escalar. Mas, como a forma específica da função de perda não é, na generalidade dos casos, relevante ou crucial para a estimação, o problema encontrar-se-á fundamentalmente resolvido se for possível introduzir conceitos métricos compatíveis com uma distância equivalente à perda quadrática esperada.

A construção teórica usual parte do espaço probabilizado em causa, Ω , para introduzir o espaço de Hilbert $L^2(\Omega)$. É $L^2(\Omega)$ constituído pelas VARIÁVEIS ALEATÓRIAS DE SEGUNDA ORDEM⁽¹⁾, ou seja, pelas v.a. reais tais que

$$Ex^2 < \infty, \quad (4.1)$$

ou, genericamente, pelos vectores aleatórios tais que $Exx^T < \infty$.

A função real PRODUTO ESCALAR é definida entre dois elementos de $L^2(\Omega)$ como

$$x|y = Exy, \quad (4.2)$$

e VARIÁVEIS ALEATÓRIAS ORTOGONAIS serão aquelas tais que $Exy = 0$. A partir do produto escalar constrói-se uma NORMA

$$\|x\| = E^{1/2}x^2, \quad (4.3)$$

(1) V. Nota Anexa 4.

e uma DISTÂNCIA

$$d(x, y) = \|x - y\| = E^{1/2}(x - y)^2. \quad (4.4)$$

A extensão ao caso vectorial é directa e também não representa dificuldade verificar que as funções reais introduzidas cumprem a axiomática usual. Algumas dificuldades que surgem pelo facto de as funções mensuráveis $x(\omega)$ poderem ter distância nula quando são idênticas apenas no sentido quase certo (q.c.) podem ser ultrapassados com a redefinição da igualdade⁽²⁾.

A definição de distância permite introduzir o conceito de CONVERGÊNCIA (em média quadrática)

$$l.i.m. x_n = x \quad \Leftrightarrow \quad \|x_n - x\| \rightarrow 0, \quad (4.5)$$

e a demonstração de que $L^2(\Omega)$ é um espaço completo é análoga à demonstração da completicidade do espaço de funções de quadrado integrável⁽³⁾.

Qualquer conjunto de n variáveis aleatórias mutuamente ortogonais gera um sub-espaço de $L^2(\Omega)$ n -dimensional. O estudo necessita pois de ser estendido a espaços de dimensão não limitada, sendo necessários os conceitos hilbertianos.

Como espaço munido de produto escalar, de dimensão não finita e completo na métrica definida pela sua norma, o espaço $L^2(\Omega)$ é um *Espaço de Hilbert*⁽⁴⁾.

Em $L^2(\Omega)$ pode ser definida uma SEQUÊNCIA ORTONORMAL de v.a. $\{y_i^*\}$

$$E y_i^* y_j^* = \delta_{ij}, \quad (4.6)$$

em que δ_{ij} é o usual delta de Kronecker⁽⁵⁾.

Um conjunto de v.a. que constitua uma *variedade linear* possuirá sempre uma sequência ortonormal que constituirá uma base particular dessa variedade. A variedade \mathcal{X} diz-se

(2) V. Nota Anexa 4

(3) Teorema de Riez-Fisher. V. e.g. Rudin, 1973, pp.292-3. V. Nota Anexa 9.

(4) V. Nota Anexa 9.

(5) $\delta_{ij} = 0$ para $i \neq j$, $\delta_{ij} = 1$ para $i = j$.

FECHADA SE

$$\{x_1, x_2, \dots, x_n, \dots\} \in \mathcal{X} \Rightarrow \text{l.i.m. } x_n \in \mathcal{X}. \quad (4.7)$$

Uma representação de $x \in \mathcal{X}$ pode ser obtida em termos dos COEFICIENTES DE FOURIER⁽⁶⁾ relativos a uma base $\{y_i^*\}$ particular, ortonormal, que se definirão da forma que segue.

Se $x = \sum a_i y_i^*$ os escalares $\{a_i\}$ serão os chamados coeficientes de Fourier e escrever-se-á

$$x \sim \sum_i a_i y_i^* \quad a_i = E x y_i^*. \quad (4.8)$$

A determinação de uma base ortonormal particular pode sempre ser efectuada num sub-espaço de $L^2(\Omega)$ por um processo de Schmidt, como é descrito por exemplo por Doob (1953, p.151) ou Loève (1963, p.459).

Admita-se daqui em diante que $\{y_i^*\}$ forma uma base ortonormal da variedade (fechada) $\mathcal{Y} \subset L^2(\Omega)$. Os resultados que seguem não dependem da base particular. Admita-se ainda que todas as variáveis têm esperança nula. A notação ficará simplificada, a ortogonalidade corresponderá à não correlação e o problema afim reduzir-se-á ao problema linear.

Em geral, qualquer v.a. $x \in L^2(\Omega)$ terá uma particular série de Fourier relativa a $\{y_i^*\}$

$$x \sim \sum_i a_i y_i^*$$

com os coeficientes $\{a_i\}$ univocamente determinados. Considere-se a v.a. $\hat{x} \in \mathcal{Y}$, com o mesmo valor dos coeficientes,

$$\hat{x} = \sum_i a_i y_i^*$$

Se $x \in \mathcal{Y}$ ter-se-á a identidade q.c. $\hat{x} = x$. Mas se $x \in (L^2 - \mathcal{Y})$ ter-se-á a identidade

$$x = \sum_i a_i y_i^* + (x - \hat{x}), \quad (4.9)$$

(6) V. e.g. Berberian, 1976, p.47.

o que mostra que os coeficientes de Fourier da v.a. $(x - \hat{x})$ são nulos, ou seja, que $(x - \hat{x})$ é ortogonal a \mathcal{Y} . Assim, qualquer elemento de $L^2(\Omega)$ pode ser decomposto, de forma única q.c., em $\hat{x} \in \mathcal{Y}$ e em $(x - \hat{x}) \perp \mathcal{Y}$.

Por outro lado, \hat{x} é a v.a. que minimiza $d(x, \mathcal{Y})$ ou seja, não há em \mathcal{Y} outra v.a. que esteja mais perto de x .

De facto, com qualquer outro, $\tilde{x} \in \mathcal{Y}$ verificar-se-á $(\hat{x} - \tilde{x}) \in \mathcal{Y}$ e $(\hat{x} - \tilde{x}) \perp (x - \hat{x})$. Então,

$$E\|x - \tilde{x}\|^2 = E\|x - \hat{x} + \hat{x} - \tilde{x}\|^2 = E\|x - \hat{x}\|^2 + E\|\hat{x} - \tilde{x}\|^2 \geq E\|x - \hat{x}\|^2, \quad (4.10)$$

transformando-se a desigualdade em igualdade quando $\hat{x} = \tilde{x}$ q.c.

Significa isto que $\hat{x} = \sum_i a_i y_i^*$ é o elemento de \mathcal{Y} que resolve o problema

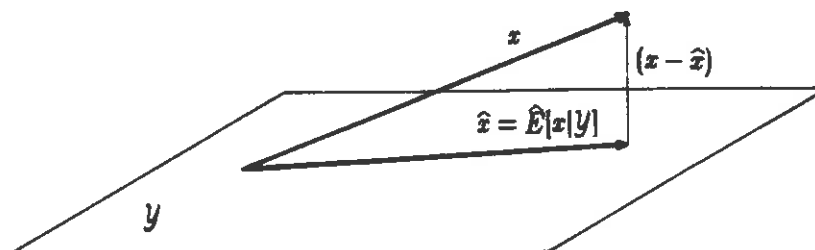
$$\inf \|x - \hat{x}\|^2 \equiv E\|x - \sum_i a_i y_i^*\|^2, \quad (4.11)$$

pelo que se trata do *valor esperado em sentido lato* de x condicionado a \mathcal{Y} :

$$\hat{x} = \hat{E}[x|\mathcal{Y}]$$

Em termos rigorosos e tendo em vista a equação (4.9), \hat{x} é designado, na teoria dos

espaços de Hilbert, como PROJECCÃO ORTOGONAL ⁽⁷⁾ de x em \mathcal{Y} . O conceito é de imediata intuição geométrica.



Tem então todo o sentido designar o operador $\hat{E}[-|\cdot]$ como OPERADOR DE PROJECCÃO ORTOGONAL, expressão de conteúdo equivalente à de esperança condicional em sentido lato e que, daqui em diante, será preferida.

As considerações precedentes conduzem directamente ao seguinte lema.

Lema 4.1 (Lema da Projecção)

Seja \mathcal{Y} variedade fechada sub-espaco de $L^2(\Omega)$.

Para cada $x \in L^2(\Omega)$ há uma única decomposição ortogonal, q.c.

(i) $x = \hat{x} + (x - \hat{x})$, $\hat{x} \in \mathcal{Y}$, $(x - \hat{x}) \perp \mathcal{Y}$.

Além disso essa decomposição é tal que

(ii) $E|x - \hat{x}|^2 = \inf E|x - \tilde{x}|^2$, com $\tilde{x} \in \mathcal{Y}$

Demonstração: segue das considerações precedentes. Demonstração sistemática encontra-se em Loève, 1963, pp.460-1

Corolário

$$\inf \|x - \tilde{x}\| = \|x - \hat{x}\| \Leftrightarrow (x - \hat{x}) \perp \mathcal{Y} \quad \forall \mathcal{Y} \in \mathcal{Y}$$

(7) Cf. Halmos, 1957, p.43

Demonstração: segue directamente do Lema, uma vez que a decomposição é única, tal que $(x - \hat{x}) \perp Y$ e dado (ii).

Como nota ao teorema repare-se que é considerada a variedade fechada Y . O objectivo é a extensão do raciocínio precedente a sub-espacos gerados por uma colecção não finita. Essa extensão, aliás, é relativamente directa (V. Doob, 1953, pp.153-4).

A introdução dos coeficientes de Fourier em $L^2(\Omega)$ pode ser vista como a criação de um espaço *isométrico e isomorfo* deste ⁽⁸⁾, no que se refere às propriedades à segunda ordem que são as que aqui interessam. A cada v.a. corresponde um vector de coeficientes reais, e o produto interno, a norma e a distância em $L^2(\Omega)$, bem como a ortogonalidade e a projecção aí introduzidas, podem ser interpretadas pelos conceitos correspondentes no espaço (hilbertiano) real dos vectores de coeficientes. Talvez este paralelo clarifique a intuição geométrica das operações em $L^2(\Omega)$. Para o problema da estimação ou previsão em causa, definido como a procura do filtro linear que minimiza a perca esperada, o Lema precedente como tem uma aplicação directa:

Teorema 4.1

Sejam $\{x_t\}$ e $\{y_t\}$ processos aleatórios de segunda ordem tais que $E x_t = E y_t = 0$. Verificada pelo menos uma das seguintes condições:

- (i) os processos são conjuntamente gaussianos, ou
- (ii) o filtro de estimação é linear em $\{y_t\}$ e a função perca é par, então a funcional em Y_t que é estimador óptimo de x_r é dada por $\hat{x}_r = \hat{E}[x_r | Y_t]$. (4.12)

Demonstração: Se os processos são gaussianos a estimação óptima é $E[x_r | Y_t]$, que corresponde à projecção. Estando em causa um filtro linear e função de perca quadrática a prova é semelhante à conduzida pela desigualdade (4.10): introduzindo a matriz s.d.p. S os produtos cruzados $E[(x - \hat{x})^T S (\hat{x} - \bar{x})]$ anulam-se pois $(x - \hat{x}) \perp (\hat{x} - \bar{x})$.

Sendo o melhor filtro linear para função de perca quadrática é também o melhor para qualquer função de perca par (V. Kalman, 1963, p.297).

Corolário (Equação de Wiener-Hopf)⁽⁹⁾

(8) V. Halmos, 1957, pp.29-30

(9) Trata-se da versão discreta da equação de Wiener-Hopf

Verificando-se pelo menos uma das condições do Teorema a optimalidade de \hat{x}_r é equivalente à condição

$$E x_r y_s^T = E \hat{x}_r y_s^T \quad \forall y_s \in y. \quad (4.13)$$

Demonstração: Basta notar que, se \hat{x}_r é a projecção em y então

$$E[(x_r - \hat{x}_r)y_s] = 0 \quad \forall y_s \in y.$$

No caso de os processos, centrados, $\{x_t\}$ e $\{y_t\}$ serem individual e mutuamente estacionários em covariância — $R_{xy}(t-s) = E x_t y_s^T = E x_{t+i} y_{s+i}^T$ — a equação de Wiener-Hopf permite simplificar a expressão dos coeficientes do filtro. Substitua-se $\hat{x}_r = \sum_i a_i y_{t-i}$ em (4.13). Daí resulta, pela continuidade do produto interno

$$E x_r y_s^T = E \left(\sum_i a_i y_{t-i} \right) y_s^T$$

$$R_{xy}(t-s) = \sum_i a_i R_{yy}(t-s-i), \quad i = 1, 2, \dots$$

o que é equivalente à expressão mais directa

$$R_{xy}(k) = \sum_i a_i R_{yy}(k-i), \quad i = 1, 2, \dots \quad (4.13')$$

Exemplo 4.1 (Processo univariável)

Admita-se que $E x_t = 0$ e que $\{x_t\}$ é um processo estacionário em covariância:

$$R(0) = 1$$

$$R(1) = \lambda / (1 + \lambda^2)$$

$$R(k) = 0, \quad |k| > 1$$

com $0 < \lambda < 1$. As equações (4.13') serão

$$\lambda / (1 + \lambda^2) = a_1 + \lambda / (1 + \lambda^2) \cdot a_2$$

$$0 = \lambda / (1 + \lambda^2) a_{k-1} + a_k + \lambda / (1 + \lambda^2) a_{k+1}, \quad k = 2, 3, \dots$$

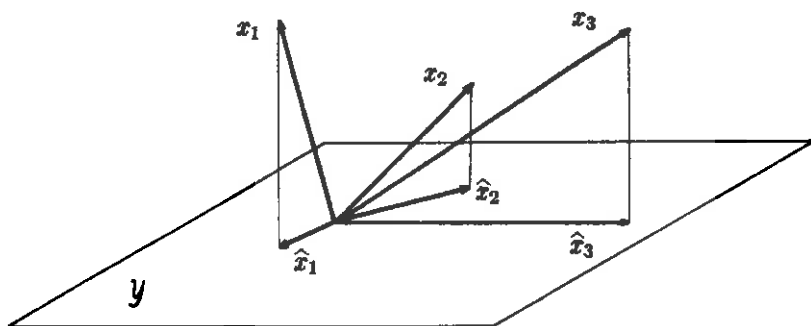
Verifica-se directamente que a solução é $a_k = -(-\lambda)^k$ e que o projector linear é

$$\hat{x}_t = +\lambda x_{t-1} - \lambda^2 x_{t-2} + \lambda^3 x_{t-3} - \dots$$

É interessante notar que, apesar de x_t e x_{t-2}, x_{t-3}, \dots serem não correlacionados, o preditor envolve todo o passado do processo. (Exemplo retirado de Karlin e Taylor, 1975, p.473)

Interpretem-se mais cuidadosamente as equações matriciais (4.12) e (4.13), removendo alguma possível ambiguidade.

A equação $\hat{x}_r = \hat{E}[x_r | Y_t]$ representa uma igualdade vectorial, interpretável termo a termo, mas a projecção é efectuada de *cada componente* sobre o espaço gerado pelos *vetores* y_t . O seguinte esquema, adaptado de Ruymgaart e Soong (1985, p.128), talvez seja elucidativo.



Quanto à versão da equação de Wiener-Hopf encontrada em (4.13), é uma igualdade matricial que está em causa. Admita-se que x tem k componentes e que y tem l . Está expressa uma igualdade entre duas matrizes de ordem $k \times l$. Significa isso que cada \hat{x}_i está condicionado por l equações

$$E x_{ir} y_{js} = E \hat{x}_{ir} y_{js} \quad j = 1, 2, \dots, l, \quad (4.14)$$

a serem por sua vez respeitadas para todo o y_s . Trata-se, na teoria da estimação linear em processos estocásticos, de equações que desempenham papel semelhante às equações normais do método dos mínimos quadrados.

Em concreto ter-se-á, para cada componente l do vector \hat{x}_r

$$\begin{aligned} \hat{x}_r^l &= a_0^l y_t^l + a_1^l y_{t-1}^l + a_2^l y_{t-2}^l + \dots + \\ &+ a_0^2 y_t^2 + a_1^2 y_{t-1}^2 + a_2^2 y_{t-2}^2 + \dots + \\ &+ \dots \\ &+ a_0^k y_t^k + a_1^k y_{t-1}^k + a_2^k y_{t-2}^k + \dots \end{aligned}$$

representando a_i^k o coeficiente ponderador de y_i^k , componente k do vector y_i , vector da observação no momento i .

Simplifique-se escrevendo

$$\hat{x}_r^l = \sum_i \sum_k a_i^k y_{i-1}^k$$

Sabe-se, pela equação de Wiener-Hopf,

$$E \left[\left(\sum_i \sum_k a_i^k y_{i-1}^k \right) \cdot y_0^j \right] = E[x_r^l \cdot y_0^j]$$

o que é representável por

$$\sum_i \sum_k a_i^k E[y_{i-1}^k \cdot y_0^j] = E[x_r^l \cdot y_0^j]$$

Simplificando drasticamente este conjunto de equações, e sem grandes preocupações de precisar a notação, encontra-se

$$A^l \cdot R_{yy}^l = R_{xy}^l$$

para cada componente l do vector x . Com um número finito de observações pode calcular-se

$$A^l = (R_{yy}^l)^+ \cdot R_{xy}^l$$

em que o símbolo $+$ representa uma inversa generalizada. A solução pode não ser única em termos dos coeficientes a_i^k , mas o lema da projecção garante que a sua introdução na fórmula de \hat{x}_r^l produz uma estimativa única. Encontra-se assim a expressão para os coeficientes da combinação linear que determina \hat{x}_r em termos de matrizes de covariâncias⁽¹⁰⁾ dos componentes de y_t, y_{t-1}, \dots e das matrizes de covariâncias cruzadas desses componentes com os de x_r .

(10) Continua-se a admitir que as variáveis têm valor esperado nulo

Por aqui se pode apreciar um pouco a dimensão do problema na falta de um algoritmo computacionalmente eficiente. A solução encontrada por Wiener em 1941 (Wiener, 1949) foi deduzida a partir do caso contínuo e admitia de alguns pressupostos restritivos (Cf. Jazwinski, 1970, p.45 e Priestley, 1981, pp.730, 762, 779). O problema limitava-se à separação de dois processos *aleatórios* e era considerado o caso limite de o número de observações passadas tender para infinito. As covariâncias $Ez_t x_s$ e $Ez_t y_s$ assumiam-se conhecidas, os processos eram estacionários e, limitação talvez mais grave, a solução do problema teria de ser completamente recalculada com a alteração do momento presente (t) ou do momento de estimação (τ). Pela mesma altura Kolmogorov⁽¹¹⁾, com uma distinta linha de ataque deduziu resultados fundamentalmente semelhantes⁽¹²⁾.

Da teoria de Wiener-Kolmogorov, para além da referência à solução pela igualdade de Wiener-Hopf e da consequente nota sobre as suas limitações práticas⁽¹³⁾, apenas interessam aqui alguns conceitos elementares necessários às secções seguintes. Aborde-se particularmente o conceito de *sequência de inovações* introduzido por Kolmogorov.

Continuando no caso discreto considere-se um único processo $\{x_t\}$, definido para $t \geq 0$, conhecendo-se Ex_t e $Ez_t x_s$. No momento zero a estimativa de x_1 será $\hat{E}[x_1|x_0]$ e definir-se-á a inovação registada em $t = 1$ como

$$\tilde{x}_1 = x_1 - \hat{E}[x_1|x_0]$$

Em geral, a estimativa feita no momento $t - 1$ será $\hat{E}[x_t|x_{t-1}x_{t-2}\dots x_0]$, ou $\hat{E}[x_t|X_{t-1}]$ em que X_{t-1} representa a sequência de observações.

A INOVAÇÃO⁽¹⁴⁾ representada pela observação de x_t será

$$\tilde{x}_t = x_t - \hat{E}[x_t|X_{t-1}]. \quad (4.15)$$

(11) Kolmogorov, A. (1941). "Interpolation und extrapolation von stationären Zufälligen Folgen", *Bull. Aca. Sci. (Nauk), U.R.S.S., Ser. Math.*, 5, 3-14.

(12) V. Wiener (1949, p.59, nota) sobre o paralelismo dos resultados. Para uma exposição introdutória V. Priestley, 1981, pp.730-779.

(13) V. Priestley, 1981, pp.730, 766 e 779 e Kalman, 1960, p.35.

(14) Anderson e Moore (1979, pp.100-4) distinguem *inovação*, definida por eles em termos do valor esperado, da *pseudo-inovação*, que definem através do projector. No quadro da teoria dos operadores lineares essa definição não parece consistente. Opta-se antes por seguir, entre outros, T.W. Anderson (1971, p.420). Assim, a inovação coincidirá com $x - E[x|-]$ apenas no caso gaussiano.

Para tornar consistente a definição assumam-se

$$\tilde{x}_0 = x_0 - Ex_0, \quad (4.16)$$

pressuposto necessário se o parâmetro t for limitado inferiormente.

A SEQUÊNCIA DE INOVAÇÕES

$$\tilde{X}_t = \{\tilde{x}_0, \tilde{x}_1, \dots, \tilde{x}_t\}$$

representa aquilo que há de *essencialmente novo* em cada observação, no sentido de que representa a parte de cada x_t observado que não é linearmente dependente, em $L^2(\Omega)$, das observações passadas. Em termos geométricos representa o componente de cada x_t que é ortogonal a X_{t-1} , sub-espaço gerado pelas observações passadas.

Concretizem-se estas notas e apresentem-se algumas propriedades notáveis da sequência \tilde{X}_t .

Teorema 4.2

Numa sequência de inovações $\{\tilde{x}_t\}$ verifica-se:

- (i) com o conhecimento dos momentos até à segunda ordem do processo $\{x_t\}$ é possível obter \tilde{x}_t a partir de X_t , tal como se pode obter x_t a partir de \tilde{X}_t ; as operações de transformação serão lineares se $Ex_t = 0 \quad \forall t$;
- (ii) $E\tilde{x}_t = 0 \quad \forall t$;
- (iii) $\tilde{x}_t \perp X_{t-1}$ ou seja, $E\tilde{x}_t x_t^T = 0 \quad \forall t < t$;
- (iv) $E\tilde{x}_t \tilde{x}_t^T = 0 \quad \forall t \neq t$;
- (v) $\hat{E}[x_t | X_{t-1}] = \hat{E}[x_t | \tilde{X}_{t-1}]$;
- (vi) $\tilde{x}_t = x_t - \hat{E}[x_t | X_{t-1}]$;
- (vii) $\hat{E}[x_t | X_{t-1}] = \sum_{i=0}^{t-1} \hat{E}[x_t | \tilde{x}_i]$

Demonstração: (i) é trivial pela observação de (4.15), em que \tilde{x}_t está expresso como combinação das observações x_t e das incluídas em X_{t-1} ; a relação inversa pode ser construída por indução; em (ii) note-se que $E\tilde{x}_t = Ex_t - E\hat{E}[x_t | X_{t-1}] = 0$; (iii) é uma extensão directa do lema da Projectção; (iv)

assuma-se $t > l$, \tilde{x}_t representa o componente de x_t que não é expresso como combinação linear de X_{t-1} , a qual inclui \tilde{x}_t ; a extensão ao caso geral é então directa; (v) resulta directamente de (i) pois $\hat{E}[x_t|X_{t-1}] = \hat{E}[x_t|X_{t-1}]$ e a variedade X_{t-1} , gerada pelas observações, inclui todas as combinações lineares destas, nomeadamente \tilde{X}_{t-1} ; (vi) é consequência directa do ponto anterior; (vii) resulta então de (v) e de (3.9).

Os diversos pontos deste teorema são fundamentais para a prática preditiva desenvolvida nas secções seguintes. Note-se, em particular, como é possível, na análise de qualquer processo, com qualquer valor esperado e qualquer covariância, chegar à sua decomposição numa sequência de inovações que é puramente aleatória — (ii) e (iv) — fornecendo uma base ortogonal para o preditor \hat{x}_t . Por outro lado destaque-se de novo

$$\hat{E}[x_t|X_{t-1}] = \hat{E}[x_t|\tilde{x}_{t-1}] + \hat{E}[x_t|\tilde{x}_{t-2}] + \dots + \hat{E}[x_t|\tilde{x}_{t-l}] + \dots, \quad (4.17)$$

que é expressão de cálculo muito mais directo do que $\hat{E}[x_t|X_{t-1}]$.

Sem outro objectivo que não seja o de fornecer uma primeira aplicação destes conceitos veja-se o exemplo seguinte

Exemplo 4.1 (Processo Autoregressivo)

Considere-se o processo $AR(1)$:

$$x_t - ax_{t-1} = \varepsilon_t \quad \text{com } |a| < 1, E\varepsilon_t = 0 \text{ e } E\varepsilon_t\varepsilon_s = \delta_{ts}\sigma_\varepsilon^2.$$

É visível que o componente essencialmente novo, não contido na variedade X_{t-1} , é ε_t , que constitui a sequência de inovações.

É conhecida a representação alternativa do processo $AR(1)$:

$$x_t = \varepsilon_t + a\varepsilon_{t-1} + a^2\varepsilon_{t-2} + \dots$$

Procure-se o preditor $\hat{x}_t = \hat{E}[x_t|X_{t-m}]$ aplicando directamente (4.17). Com $n > m$ ter-se-á obviamente $\hat{E}[x_t|\tilde{x}_{t-n}] = \hat{E}[x_t|\varepsilon_{t-n}] = a^n\varepsilon_{t-n}$ e onde se deduz

$$\hat{E}[x_t|X_{t-m}] = a^m\varepsilon_{t-m} + a^{m+1}\varepsilon_{t-m-1} + a^{m+2}\varepsilon_{t-m-2} + \dots$$

ou ainda $\hat{x}_t = a^m(\varepsilon_{t-m} + a\varepsilon_{t-m-1} + a^2\varepsilon_{t-m-2} + \dots)$, pelo que o preditor pode ser expresso na sequência de inovações ou, alternativamente, nas observações:

$$\hat{x}_t = a^m x_{t-m}$$

O exemplo mostra, de forma trivial, como o teorema 4.2 pode ser generalizado a preditores não só do momento imediato mas também de momentos $t + m$. O teorema seguinte generaliza estes resultados e aplica-os ao modelo (1.1) em dois importantes pontos.

Teorema 4.3

Sejam $\{y_t\}$ e $\{x_t\}$ dois processos aleatórios conjuntamente distribuídos. São válidas as seguintes relações:

- (i) $\hat{E}[x_\tau | Y_t] = \hat{E}[x_\tau | \tilde{Y}_t]$
- (ii) $\hat{E}[x_\tau | Y_t] = \hat{E}[x_\tau | \tilde{y}_t] + \hat{E}[x_\tau | \tilde{y}_{t-1}] + \dots$

Demonstração: A prova é em tudo idêntica à de (v) e (vii) do teorema anterior. Não é aqui obrigatória a condição $\tau > t$, atrás pressuposta para evitar trivialidades, uma vez que aí o processo observado coincidia com o processo estimado.

A terminar revejam-se as notas finais à secção anterior. Na realidade, a dedução dos diversos resultados aplicáveis a preditores lineares, que são projecções ou valores esperados condicionados em sentido lato, foi feita sem qualquer pressuposto restritivo do processo, apenas obrigado a pertencer a $L^2(\Omega)$. Os teoremas desta secção apontam condições de optimalidade para estimadores lineares e terão a versão estrita de indicarem condições de optimalidade geral para processos gaussianos. Procurar-se-á continuar a não restringir a classe dos processos em estudo, tendo em mente os resultados mais fortes no caso estrito de processos gaussianos.



segunda parte

previsão com Filtro de Kalman e modelização em espaço de estados

A solução teórica de um problema de estimação estocástica e a descoberta de algoritmos computacionalmente eficientes constituem partes indissociáveis da abordagem aplicada de um problema estatístico. Tal facto é particularmente visível na teoria da estimação e previsão desenvolvida por Kalman.

Muito mais do que a construção de um algoritmo os trabalhos de Kalman determinaram uma diferente abordagem da estimação estocástica e permitiram o estudo de sistemas dinâmicos, abrindo novas portas à previsão de processos aleatórios.

As secções que integram esta segunda parte procuram desenvolver a visão de um modelo estocástico dinâmico, mostrando como esse modelo é particularmente adaptado à previsão de sucessões cronológicas reais.

Na secção inicial introduzem-se os procedimentos de estimação recursivos, mostrando as suas vantagens computacionais na solução de problemas do tipo dos de Wiener-Kolmogorov.

Na secção seguinte, e no quadro dessa classe de procedimentos, deduz-se o célebre algoritmo construído por Rudolf Kalman em 1960 e usualmente designado por Filtro de Kalman.

É depois aprofundada a justificação teórica da estimação recursiva, mostrando como as vantagens computacionais desse tipo de algoritmos têm fundamento no carácter dinâmico dos sistemas.

Esta parte do trabalho tem ainda uma secção final em que se discutem potencialidades e limitações do Filtro de Kalman. Serão aí apresentados os modelos de espaço de estados, modelos que, em conjunto com o filtro, permitem uma nova abordagem dos sistemas estocásticos dinâmicos.

5. Procedimentos de previsão recursivos

Donald Knuth, na sua monumental obra sobre algoritmia e computação, define da seguinte forma uma relação de recorrência: "A rule which defines each element of a sequence in terms of preceding elements" (Knuth, 1973, p.630).

Em sucessões cronológicas uma análise recursiva representará o uso de uma técnica sequencial em que as estimativas são constantemente actualizadas e os dados são naturalmente trabalhados em ordem temporal.

A análise convencional de sucessões cronológicas, em contraste com esse tipo de procedimentos, envolve o processamento em bloco (*batch*) de um conjunto fixo de dados. Está em causa, habitualmente, uma *operação única* — como na solução das equações normais dos mínimos quadrados — ou *operações iterativas*, em que o mesmo conjunto fixo de dados é sucessivamente trabalhado por algoritmos de aproximação — como é exemplo a maximização da função de verosimilhança (V. e.g. Shader e Schmid, 1985).

Inicialmente, o interesse pelos procedimentos recursivos fundamentou-se apenas nas vantagens computacionais que esses métodos teriam, nomeadamente ao permitir a incorporação de novas observações sem obrigar a um outro processamento, em bloco, dos dados iniciais. Em 1821 Gauss, no contexto do método dos mínimos quadrados, deriva as fórmulas necessárias para a incorporação de uma observação adicional⁽¹⁾. Mas só em 1950 o problema volta a merecer atenção significativa, tendo Plackett (1950, pp.152-7) deduzido um algoritmo para adição de novo conjunto de observações⁽²⁾.

Como introdução à teoria de Kalman interessa aqui a dedução de algoritmos recursivos na estimação de processos estocásticos, pelo que os resultados serão apresentados nesse contexto.

Comece-se pelo caso finito e univariável, com uma sucessão cronológica única $\{x_t\}$, pretendendo-se uma estimativa para o valor da variável no momento τ , a partir do conhecimento de X_t e dos momentos cruzados $E x_s x_t$ e $E x_s x_\tau$, $s, t = 1, 2, \dots, t$. Recorde-se que o problema consiste em definir os coeficientes da projecção

$$\hat{x}_\tau = \hat{E}\{x_\tau | X_t\} = \sum_1^t a_i x_i \quad (5.1)$$

- (1) *Theoria combinationis observationum erroribus minimis abnoxiae, Werke, 4, Göttingen (Cf. Plackett, 1950).*
- (2) Tendo pois prioridade sobre Swerling, aparentemente ao contrário do apontado por Jazwinski, 1970, p.157: P. Swerling (1959) "First order error propagation in a stagewise smothing procedure for satellite observations" *J. Astronaut. Sci.*, 6, pp.46-52.

sendo esses coeficientes $\{a_i\}$ aqueles que minimizam $E[x_\tau - \hat{x}_\tau]^2$, com $\hat{x}_\tau \in \mathcal{X}_t$.

Aplicando o lema da projecção ou directamente a equação de Wiener-Hopf

$$E\hat{x}_\tau x_s = E x_\tau x_s \quad s = 1, 2, \dots, t \quad (5.2)$$

encontram-se as condições necessárias e suficientes para solução do problema: (5.1) e (5.2). Para a determinação dos coeficientes $\{a_i\}$ introduz-se a expressão (5.1) na equação de Wiener-Hopf, construindo-se o sistema

$$\begin{bmatrix} E x_1 x_1 & E x_1 x_2 & \dots & E x_1 x_t \\ E x_2 x_1 & \dots & & \\ \dots & & & \\ E x_t x_1 & \dots & & E x_t x_t \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \dots \\ a_t \end{bmatrix} = \begin{bmatrix} E x_1 x_\tau \\ E x_2 x_\tau \\ \dots \\ E x_t x_\tau \end{bmatrix} \quad (5.3)$$

Para cada $t \in \mathbb{N}$ o sistema tem sempre uma solução que é única se e só se a sequência X_t for constituída por variáveis linearmente independentes. No entanto, o lema da projecção garante que a estimativa resultante é única, qualquer que seja a determinação dos coeficientes $\{a_i\}$.

Sendo conhecido x_{t+1} e pretendendo-se actualizar a estimativa \hat{x}_τ o problema é semelhante e pode ser resolvido com o conhecimento adicional de $E x_s x_{t+1}$ e $E x_s x_\tau$, $s = 1, 2, \dots, t+1$.

A resolução do novo sistema (5.3) ampliado implicará o cálculo completo da nova sequência de coeficientes, a menos que se tenha previamente reduzido a matriz dos momentos cruzados à forma diagonal. É então conveniente começar por ortogonalizar a sequência de observações (Cf. Ruymgaart, 1985, p.115) usando, por exemplo, o processo de Gram-Schmidt.

Admita-se que as observações são independentes, o que é o caso geral. Não o sendo, a única alteração consistiria na redução da ordem da matriz $[E x_s x_t]$. Transforme-se X_t em X_t^* de tal forma que

$$E x_i^* x_j^* = \delta_{ij}$$

A resolução do sistema (5.3) fica extraordinariamente simplificada:

$$a_i^* = E x_i^* x_\tau \quad i = 1, 2, \dots, t \quad (5.4)$$

estando agora indentificados os coeficientes da projecção (5.1) com os coeficientes de Fourier referentes à base ortogonal do espaço X_t .

Como em cada momento a dimensão de X_t aumenta de uma unidade (excepto se $x_{t+1} \in X_t$), a ortonormalização permite construir um processo recursivo para o cálculo das novas estimativas. Designe-se por $\hat{x}_{r|t}$ a estimativa com o conhecimento de X_t . Ter-se-á

$$\begin{aligned}\hat{x}_{r|t} &= (Ex_r x_1^*)x_1^* + (Ex_r x_2^*)x_2^* + \dots + (Ex_r x_t^*)x_t^* \\ \hat{x}_{r|t+1} &= \hat{x}_{r|t} + (Ex_r x_{t+1}^*)x_{t+1}^*\end{aligned}$$

estando pois deduzido um procedimento recursivo de actualização das estimativas

$$\begin{aligned}\hat{x}_{r|0} &= 0 \\ \hat{x}_{r|t} &= \hat{x}_{r|t-1} + (Ex_r x_t^*)x_t^* \quad t = 1, 2, \dots\end{aligned} \tag{5.5}$$

A extensão ao caso vectorial é imediata, tal como o é a extensão ao caso não centrado ($Ex_t \neq 0$), bastando aplicar (5.5) às variáveis centradas. Também não apresenta quaisquer dificuldades a consideração de dois processos conjuntamente distribuídos, tal como no modelo (1.1):

$$\begin{aligned}\hat{x}_{r|0} &= 0 \\ \hat{x}_{r|t} &= \hat{x}_{r|t-1} + (Ex_r y_t^*)y_t^* \quad t = 1, 2, \dots\end{aligned} \tag{5.6}$$

Vê-se como a aplicação deste tipo de procedimentos permite uma drástica simplificação no tratamento do problema de Wiener-Kolmogorov que foi apresentado na secção anterior. Contudo, permanece o peso computacional da ortonormalização de X_t , pelo que será de indagar se a própria sequência de observações não fornece uma alternativa mais directa. A solução que aparece agora como óbvia consiste no tratamento da sequência de inovações que é, como se viu, directamente ortogonal.

Retorne-se ao caso finito e univariável, com uma sucessão cronológica única $\{x_t\}$. Pretende-se em cada momento t uma previsão para x_{t+1} , com o conhecimento de X_t , sequência de inovações, e dos momentos cruzados $Ex_t x_s$. O processo pode não ser estacionário em covariância mas admita-se $Ex_t = 0 \quad \forall t$. Defina-se o erro quadrático esperado da previsão de x_t com o conhecimento de X_{t-1} : $\|x_t - \hat{x}_t\|^2 = E\tilde{x}_t^2$. Nestas condições ter-se-á o algoritmo recursivo que se apresenta sobre a forma de teorema.

Teorema 5.1 (Algoritmo de inovações)

Com o processo estocástico de segunda ordem $\{x_t\}$, de valor esperado nulo e matriz de momentos cruzados $\{E x_t x_s\}$ não singular, os preditores em um passo $\hat{x}_{t+1} = \hat{E}[x_{t+1}|X_t]$ e os erros quadráticos médios $E\tilde{x}_{t+1}^2$ são recursivamente obtidos por

$$\begin{aligned} \hat{x}_1 &= 0 \\ \hat{x}_{t+1} &= \sum_{i=1}^t a_{i|t} \tilde{x}_{t+1-i} \quad t = 1, 2, \dots \end{aligned} \quad (5.7)$$

sendo $a_{i|t}$ os coeficientes da projecção $\hat{E}[x_{t+1}|X_t]$, elaborada em cada momento t , e que são obtidos por

$$\begin{aligned} E\tilde{x}_1^2 &= E x_1^2 \quad (5.8) \\ a_{t-k|t} &= (E\tilde{x}_{k+1}^2)^{-1} \left(E x_{t+1} x_{k+1} - \sum_{i=0}^{k-1} a_{k-i|k} a_{t-i|t} E\tilde{x}_{i+1}^2 \right) \\ E\tilde{x}_{t+1}^2 &= E x_{t+1}^2 - \sum_{i=0}^{t-1} a_{i+1|t}^2 E\tilde{x}_{i+1}^2 \quad t = 1, 2, \dots \end{aligned}$$

Demonstração: Como \tilde{X}_t é um conjunto ortogonal, o produto interno nos dois membros de (5.7) por \tilde{x}_{k+1} , $k = 0, 1, \dots, t-1$ encontra

$$E\hat{x}_{t+1} \tilde{x}_{k+1} = a_{t-k|t} E\tilde{x}_{k+1}^2$$

Como $E\tilde{x}_{t+1} \tilde{x}_{k+1} = 0$, $E[x_{t+1}(x_{k+1} - \hat{x}_{k+1})] = E[\hat{x}_{t+1}(x_{k+1} - \hat{x}_{k+1})]$ pelo que essa expressão permite encontrar

$$a_{t-k|t} = (E\tilde{x}_{k+1}^2)^{-1} \cdot E x_{t+1} \tilde{x}_{k+1}$$

Por outro lado, a representação (5.7), obviamente válida pelo teorema 4.2, permite exprimir \tilde{x}_{k+1} na forma $x_{k+1} - \sum a_{i|k} \tilde{x}_{k+1-i}$ a qual, substituída na expressão acima, encontra

$$a_{t-k|t} = (E\tilde{x}_{k+1}^2)^{-1} \left(Ex_{t+1}x_{k+1} - \sum_{i=0}^{k-1} a_{k-i|k} Ex_{t+1}\tilde{x}_{i+1} \right)$$

Introduzindo a já deduzida igualdade $Ex_{t+1}\tilde{x}_{i+1} = a_{t-i|t}E\tilde{x}_{i+1}^2$, encontra-se a equação que em (5.8) exprime $a_{t-k|t}$.

Resta derivar a fórmula recursiva de $E\tilde{x}_{t+1}^2$. Como $\tilde{x} = (x - \hat{x}) \perp \hat{x}$ a lei de Pitágoras⁽³⁾ permite escrever

$$E\tilde{x}_{t+1}^2 = \|x_{t+1}\|^2 - \|\hat{x}_{t+1}\|^2 = Ex_{t+1}^2 - \sum_{i=0}^{t-1} a_{t-i|t}^2 E\tilde{x}_{i+1}^2$$

O algoritmo que acabou de ser deduzido permite encontrar recursivamente os coeficientes e os erros esperados quadráticos, pela ordem:

$$\begin{aligned} &E\tilde{x}_1^2 \\ &a_{1|1}, \quad E\tilde{x}_2^2 \\ &a_{2|2}, \quad a_{1|2}, \quad E\tilde{x}_3^2 \\ &a_{3|3}, \quad a_{2|3}, \quad a_{1|3}, \quad E\tilde{x}_4^2 \\ &\dots \end{aligned}$$

Os coeficientes do desenvolvimento em série dos estimadores aparecem em termos da sequência das inovações

$$\hat{x}_t = \sum_{i=1}^{t-1} a_i \tilde{x}_{t-i} \quad (5.9)$$

expressão que é de uso extremamente simples e que pode em alguns casos ser ainda simplificada. O próprio processo pode ser agora representado, com os mesmos coeficientes e definindo $a_0 := 1$, em termos da sequência \tilde{X}_t :

$$x_t = \sum_{i=0}^{t-1} a_i \tilde{x}_{t-i} \quad (5.10)$$

(3) V. Nota Anexa 4

o que mostra uma forma de obter $\{x_t\}$ a partir de uma operação sobre $\{\tilde{x}_t\}$, tal como referido no teorema 4.2.

Exemplo 5.1 (Previsão de processo MA(1))

Considere-se o processo $x_t = \varepsilon_t + \theta\varepsilon_{t-1}$ em que $\{\varepsilon_t\}$ é processo puramente aleatório, com $E\varepsilon_t = 0$ e $E\varepsilon_t^2 = \sigma^2$. Como $E\tilde{x}_t^2 = (1 + \theta^2)\sigma^2$ $E x_t x_{t+1} = \theta\sigma^2$, anulando-se as covariâncias nos outros casos, é fácil encontrar, a partir de (5.8)

$$a_{i|t} = 0 \quad 2 \leq i \leq t$$

$$a_{1|t} = (E\tilde{x}_t^2)^{-1}\theta\sigma^2$$

$$E\tilde{x}_t^2 = (1 + \theta^2)\sigma^2$$

$$E\tilde{x}_{t+1}^2 = [1 + \theta^2 - (E\tilde{x}_t^2)^{-1}\theta^2\sigma^2]\sigma^2$$

A expressão do estimador aparece como

$$\hat{x}_{t+1} = \theta[1 + \theta^2 - (E\tilde{x}_t^2)^{-1}\theta^2\sigma^2]^{-1} \cdot \tilde{x}_t$$

Os erros das estimativas não são decrescentes com t e pode-se mostrar ainda que $\|\tilde{x}_t - \varepsilon_t\| \rightarrow 0$, $E\tilde{x}_t^2 \rightarrow \sigma^2$ e $a_{1|t} \rightarrow \theta$ se $|\theta| < 1$ ⁽⁴⁾.

Vale ainda a pena referir uma concretização de algoritmo de inovações aos processos ARMA⁽⁵⁾. Essa classe de processos será repetidamente abordada e convém introduzir desde já alguns conceitos e notações apropriadas.

Considere-se o modelo ARMA(p,q)

$$\alpha(B)x_t = \beta(B)\varepsilon_t, \quad E\varepsilon_t = 0 \quad E\varepsilon_t\varepsilon_s = \delta_{ts}\sigma^2 \quad (5.11)$$

aonde, como é usual, $\alpha(B)x_t = x_t + a_1x_{t-1} + \dots + a_px_{t-p}$ e $\beta(B)\varepsilon_t = \varepsilon_t + b_1\varepsilon_{t-1} + \dots + b_q\varepsilon_{t-q}$; convencie-se ainda $a_0 := 1$ e $b_0 := 1$.

Em vez de aplicar directamente o algoritmo de inovações ao processo (5.11) pode adoptar-se uma transformação proposta por Ansley (1979, p.61) que simplifica extraordinariamente a representação preditiva.

Considere-se o processo transformado

⁽⁴⁾ Sobre a aplicação deste algoritmo recursivo ver Brockwell e Davis (1987a).

⁽⁵⁾ V. Nota Anexa 10

$$z_t = \sigma^{-1} x_t \quad t = 1, 2, \dots, m \quad (5.12)$$

$$z_t = \sigma^{-1} \alpha(B)x_t \quad t > m$$

com $m = \max\{p, q\}$. É evidente que a variedade linear gerada pelas observações X_t coincide com a variedade gerada pelas observações Z_t pelo que os projectores ortogonais serão idênticos em X_t ou Z_t .

Aplicando o algoritmo ao processo $\{z_t\}$ obtem-se

$$\begin{aligned} \hat{z}_{t+1} &= \sum_{i=1}^t a_{i|t} \tilde{z}_{t+1-i} & 1 \leq t < m \\ \hat{z}_{t+1} &= \sum_{i=1}^q a_{i|t} \tilde{z}_{t+1-i} & m \leq t \end{aligned} \quad (5.13)$$

sendo os coeficientes $a_{i|t}$ e os erros quadráticos Ez_t^2 encontrados recursivamente pela aplicação de (5.8). O cálculo dos momentos cruzados $Ez_t x_t$ será feito de forma usual em processos ARMA (V. e.g. Box e Jenkins, 1970, p.74).

O facto notável em (5.13) é a anulação dos coeficientes $a_{i|t}$ quando $t \geq m$ e $i > q$. É uma consequência da aplicação em (5.8) do facto conhecido de $Ez_t x_t = 0$ quando $t > m$ e $|t - i| > q$ (*idem*, p.68).

Pela projecção em $X_{t-1} \equiv Z_{t-1}$ identificar-se-á

$$\begin{aligned} \hat{z}_t &= \sigma^{-1} \hat{x}_t & 1 \leq t \leq m \\ \hat{z}_t &= \sigma^{-1} (\hat{x}_t + a_1 x_{t-1} + \dots + a_p x_{t-p}) & t > m \\ \tilde{z}_t &= \sigma^{-1} \tilde{x}_t \end{aligned}$$

Procedendo às necessárias substituições encontra-se a expressão equivalente às equações (5.7) do algoritmo

$$\begin{aligned} \hat{x}_{t+1} &= \sum_{i=1}^t a_{i|t} \tilde{x}_{t+1-i} & 1 \leq t < m \\ \hat{x}_{t+1} &= (a_1 x_t + \dots + a_p x_{t-p}) + \sum_{i=1}^q a_{i|t} \tilde{x}_{t+1-i} & m \leq t \end{aligned} \quad (5.14)$$

com $E\hat{x}_{t+1}^2 = \sigma^2 E\hat{z}_{t+1}^2$ e com os coeficientes e os erros quadráticos determinados recursivamente como nas equações (5.8).

O grande interesse da representação de \hat{x}_{t+1} na forma (5.14) reside no facto de apenas necessitar, no máximo, de p observações e de q inovações, enquanto que a aplicação directa do algoritmo ao processo (5.11) requer a representação de \hat{x}_{t+1} em termos das t inovações precedentes.

6. Algoritmo preditor de Kalman

Os algoritmos recursivos apresentados na secção anterior, assim como os exemplos de aplicação da igualdade de Wiener-Hopf, tinham em comum o facto de serem lineares nas variáveis e de serem da dimensão, de parâmetros, finita. A abordagem de Box-Jenkins tem também essas características, explorando a linearidade e finitude dos modelos ARIMA. A teoria de Kalman é igualmente construída com modelos lineares e finitos, se bem que permita extensões a essa abordagem.

O primeiro traço distintivo de formulação de Kalman é o estudo dos processos estocásticos e das suas relações numa perspectiva *sistémica*. Considera-se a existência de *inputs*, aleatórios e determinísticos, que são transformados no interior do sistema, dando origem a um processo observado. Não se entrará aqui em pormenores, embora a concepção dos sistemas pareça fundamental na análise de problemas econométricos como os da identificação e seleção de modelos⁽¹⁾.

Segundo traço distintivo da teoria de Kalman é o carácter *dinâmico* do sistema⁽²⁾. Os parâmetros definidores das relações entre os processos e das suas próprias características de segunda ordem variam no tempo. Isso permite uma perspectiva de estudo muito mais geral do que a até aqui formulada.

Um terceiro traço característico da formulação de Kalman é o carácter *finito e linear* do sistema. As variáveis são de dimensão finita, tal como os parâmetros e as coordenadas definidoras do processo. A linearidade é atributo essencial à manuseabilidade matemática e universalidade de propriedades do modelo (Kalman, 1982, p.175).

O sistema, Σ , é definido por três matrizes de parâmetros descritivos, $\Sigma := (F, G, H)$, em que as duas primeiras condicionam a evolução do *estado do sistema*, sintetizado no vector aleatório x_t , e a terceira explica a transformação determinística sofrida pelo sinal. O *output* externamente mensurável concentra-se no vector aleatório y_t enquanto os *inputs* puramente aleatórios são determinados pelos vectores w_t e v_t .

- (1) Para uma descrição teórica da teoria dos sistemas ver Kalman et al. (1969) e para uma abordagem sistémica contrastada com a perspectiva econométrica ver Kalman (1982).
- (2) Aspecto compatível com o carácter invariante do sistema, em sentido paralelo ao definido na secção 1.

No caso discreto o modelo fica descrito pelas equações

$$\mathbf{x}_{t+1} = F_t \mathbf{x}_t + G_t \mathbf{w}_t, \quad (6.1')$$

$$y_t = H_t \mathbf{x}_t + v_t, \quad (6.1'')$$

sendo a primeira a EQUACÃO DE TRANSIÇÃO, que condiciona a mudança do estado, e a segunda a EQUACÃO DE MEDIDA, que determina a perturbação que o *output* observado sofre.

A equação de medida não é mais do que a concretização ao caso linear do modelo (1.1), mas agora o processo é generalizado admitindo a existência de uma variabilidade determinada na evolução de $\{\mathbf{x}_t\}$.

Um conjunto de exemplos já fornecidos (1.1, 1.3, 1.4) constitui concretização deste modelo no caso particular e restritivo de $F_t \equiv I$.

A introdução das duas equações que constituem o sistema permite uma muito maior generalidade, pois não só se estudará o problema da filtragem ou recuperação do sinal como ainda se permitirá a abordagem de um processo aleatório $\{\mathbf{x}_t\}$ em evolução. O caso estacionário e o não estacionário⁽³⁾ são ambos encarados da mesma forma, sem as usuais operações de diferenciação ou transformação, para obrigar à estacionaridade do processo estocástico.

Sem outras dificuldades que não as computacionais, o modelo pode ser ampliado para a consideração de processos aleatórios $\{\mathbf{w}_t\}$ e $\{v_t\}$ auto e cruzadamente correlacionados. Pode ainda ser adicionada uma componente determinística u_t . Na forma mais geral o sistema define-se pelas equações

$$\mathbf{x}_{t+1} = F_t \mathbf{x}_t + G_t \mathbf{w}_t + \Gamma_t u_t \quad (6.2')$$

$$y_t = H_t \mathbf{x}_t + v_t \quad (6.2'')$$

As características sistêmicas, dinâmicas e finitas da formulação de Kalman ficam bem claras com os modelos (6.1) ou (6.2).

(3) V. Nota Anexa II.

Dar-se-ão em seguida alguns exemplos que permitirão apreciar a generalidade das situações tratadas pelos sistemas apresentados.

Exemplo 6.1 (Modelo de regressão adaptativa)

Cooley e Prescott (1973) introduziram o modelo:

$$\begin{aligned} y_t &= \alpha_t + \beta x_t + u_t \\ \alpha_t &= \alpha_{t-1} + v_t \end{aligned}$$

e deduziram os estimadores de máxima verosimilhança para $\beta, \sigma_u^2, \sigma_v^2$ e α_0 , com os usuais pressupostos sobre as perturbações aleatórias.

No quadro do modelo (6.1) y_t e x_t constituem valores conhecidos e $\{\alpha_t\}$, tal como β , podem ser encarados como o processo a estimar. A equação de transição será

$$\begin{bmatrix} \alpha_t \\ \beta \end{bmatrix} = \begin{bmatrix} \alpha_{t-1} \\ \beta \end{bmatrix} + \begin{bmatrix} v_t \\ 0 \end{bmatrix}$$

e a equação de medida corresponde a

$$y_t = [1 \quad x_t] \begin{bmatrix} \alpha_t \\ \beta \end{bmatrix} + u_t$$

Exemplo 6.2 (Crescimento exponencial)

Considere-se um modelo de crescimento geométrico com taxa aleatória

$$\begin{aligned} y_t &= r_t y_{t-1} + v_t \\ r_t &= r_{t-1} + u_t \end{aligned}$$

Em cada momento t y_t constitui a observação, y_{t-1} a matriz H_t , e r_t a variável de estado, que se pretende estimar para determinar a taxa de crescimento $(1 - r_t)$. Outras formulações equivalentes são possíveis, residindo o interesse deste exemplo em mostrar como o modelo linear (6.1) é adaptado a processos com componente determinístico, de evolução não linear.

Exemplo 6.3 (Expectativas adaptativas)

Considere-se um modelo em que os preços P_t são formados com base no custo de produção C_{t-1} e na expectativa sobre a inflação futura. Os preços são fixados no momento $t-1$, com base nos custos de produção no mesmo período e na inflação implicitamente verificada: i_{t-2} . Adiciona-se-lhe uma componente aleatória, uma vez que a inflação implícita se considera conhecida apenas aproximadamente⁽⁴⁾

$$P_t = C_{t-1} + \alpha i_{t-2} P_{t-1} + v_t \quad 0 < \alpha < 1$$

Por sua vez os custos são formados por um componente interno f , fabricado em $t-2$ e por um componente importado no momento $t-1$ e de custo u_t conhecido:

$$C_t = \lambda(1 + i_{t-2})f + (1 - \lambda)u_{t-1} + w_{1,t} \quad 0 < \lambda < 1$$

A taxa de inflação implícita evolui em "passeio aleatório"

$$i_t = i_{t-1} + w_{2,t}$$

O estado do sistema, não observado, pode sintetizar-se no vector $[C_t \quad i_{t-1}]^T$ e o componente determinístico será $[\lambda f \quad (1 - \lambda)u_{t-1}]^T$. Uma formulação possível para as equações de transição e de medida será a seguinte

$$\underbrace{\begin{bmatrix} C_t \\ i_{t-1} \end{bmatrix}}_{x_{t+1}} = \underbrace{\begin{bmatrix} 0 & \lambda f \\ 0 & 1 \end{bmatrix}}_{F_t} \underbrace{\begin{bmatrix} C_{t-1} \\ i_{t-2} \end{bmatrix}}_{x_t} + \underbrace{\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}}_{G_t} \underbrace{\begin{bmatrix} w_{1,t} \\ w_{2,t-1} \end{bmatrix}}_{w_t} + \underbrace{\begin{bmatrix} \lambda & (1 - \lambda) \\ 0 & 0 \end{bmatrix}}_{\Gamma_t} \underbrace{\begin{bmatrix} f \\ u_{t-1} \end{bmatrix}}_{u_t}$$

$$\underbrace{P_t}_{y_t} = \underbrace{[1 \quad \alpha P_{t-1}]}_{H_t} \begin{bmatrix} C_{t-1} \\ i_{t-2} \end{bmatrix} + v_t$$

Apesar de estarem em causa, em cada momento, variáveis respeitantes a três diferentes instantes, se conseguiu-se caracterizar a evolução e medida do sistema em relação a t e $t+1$, unicamente.

Exemplo 6.4 (Mudanças estruturais)

Com o mesmo problema do exemplo anterior admita-se que, a partir de um momento conhecido de mudança estrutural, t_m , a taxa de inflação interna passa a sofrer um aumento médio γ pelo que o peso do

⁽⁴⁾ V. e.g. Johnston, 1984, p.349

componente importado, cujo preço se mantém, cresce na proporção $(1 + \gamma)$ e o peso do componente fabricado desce $(1 - \gamma)$ para minorar o agravamento do custo final.

As equações de transição alteram-se

$$C_t = (1 - \gamma)\lambda(1 + i_{t-2})f + (1 + \gamma)(1 - \lambda)u_{t-1} + w_{1,t}$$

$$i_t = (1 + \gamma)i_{t-1} + w_{2,t}$$

No entanto, o sistema manter-se-á, bastando proceder às seguintes definições

$$F_t = \begin{bmatrix} 0 & \lambda f \\ 0 & 1 \end{bmatrix}, \quad t < t_m; \quad F_t = \begin{bmatrix} 0 & (1 - \gamma)\lambda f \\ 0 & (1 - \gamma) \end{bmatrix}, \quad t \geq t_m$$

$$G_t = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad \forall t$$

$$\Gamma_t = \begin{bmatrix} \lambda & (1 - \lambda) \\ 0 & 0 \end{bmatrix}, \quad t < t_m; \quad \Gamma_t = \begin{bmatrix} \lambda(1 - \gamma) & (1 - \lambda)(1 + \gamma) \\ 0 & 0 \end{bmatrix}, \quad t \geq t_m$$

Reconhecida, embora ainda superficialmente, a generalidade de situações modelizáveis segundo o esquema de Kalman, importa definir o problema de estimação que se defrontará de seguida.

Como anteriormente, está em causa a recuperação aproximada de um sinal — variável de estado x_t — a partir da sua observação perturbada por um ruído aleatório. Mantém-se o critério de optimalidade e a condição de o estimador ser um filtro linear ou afim. Simplesmente, o problema agora encontra uma formulação mais geral e dever-se-á abandonar o pressuposto de as variáveis terem esperança matemática nula. Significa isso que, em geral, são estimadores afins que estão em causa. Uma solução possível, já esboçada no exemplo 3.1, e que tem a vantagem de manter a coerência e uniformidade da abordagem, consiste em aumentar de uma dimensão a variedade a que pertence \hat{x}_t com uma variável aleatória degenerada (por exemplo identicamente igual à unidade). Definir-se-á então o *estimador afim* como estimador linear nessa nova variedade.

PROJECTOR AFIM de x_t na variedade gerada por Y_t é o estimador

$$\hat{x}_t = \hat{E}[x_t | Y_t, 1] := \hat{E}[x_t | Y_t] + E x_t \quad (6.3)$$

Mantem-se obviamente a propriedade de \hat{x}_t ser o estimador dos mínimos quadrados — condicionado agora à afinidade em Y_t , ou à linearidade em $\{Y_t, 1\}$, o que é equivalente. As propriedades do operador $\hat{E}[-|-]$ mantêm-se, com as adaptações óbvias.

Vale a pena referir que, em processos gaussianos,

$$\hat{E}[x_t | Y_t, 1] = E[x_t | Y_t] \quad (6.4)$$

assim como é importante notar os seguintes corolários dos teoremas 3.9 e 4.2, deduzidos por simples introdução de (6.3) nas expressões (iii) e (vii) respectivas:

$$\hat{E}[x | y, 1] = Ex + \Sigma_{xy} \Sigma_{yy}^{-1} (y - Ey), \quad (6.5)$$

e, se $y_s \perp y_t$, $s = 0, 1, \dots, t$

$$\hat{E}[x | y_0, y_1; 1] = \hat{E}[x | y_0, 1] + \hat{E}[x | y_1, 1] - Ex \quad (6.6)$$

...

$$\hat{E}[x | y_0, y_1, \dots, y_t; 1] = \sum_{k=0}^t \hat{E}[x | y_k, 1] - t \cdot Ex$$

Como anteriormente o problema da estimação de x_τ com base em Y_t terá diferentes significados conforme $\tau > t$, $\tau = t$ ou $\tau < t$. O primeiro problema é o da *previsão*, o segundo o da *filtragem*, e o terceiro o do *alisamento* de x_τ . Em todos os casos o algoritmo é o *Filtro de Kalman*, entendendo-se aqui o termo "filtro" como sinónimo de estimador estocástico. O estudo subsequente concentrar-se-á sobre a previsão.

Problema 6.1 (Previsão em um passo com algoritmo recursivo em sistema com *inputs* aleatórios e determinísticos)

São dadas as equações de transição e de medida

$$x_{t+1} = F_t x_t + G_t w_t + \Gamma_t u_t \quad (6.2')$$

$$y_t = H_t x_t + v_t \quad (6.2'')$$

com

- x_t ($n \times 1$) vector aleatório de estado, desconhecido
- F_t ($n \times n$) matriz de transição, conhecida
- G_t ($n \times l$) matriz conhecida
- w_t ($l \times 1$) vector aleatório de "ruído", desconhecido
- Γ_t ($n \times k$) matriz conhecida

u_t ($k \times 1$) vector de *input* determinístico, conhecido
 y_t ($m \times 1$) vector aleatório observável
 H_t ($m \times n$) matriz conhecida
 v_t ($m \times 1$) vector aleatório de "ruído", desconhecido

assume-se que todas as variáveis aleatórias são de segunda ordem e que os erros constituem "ruído branco":

$$\begin{aligned}
 E v_t &= E w_t = 0 & E v_s v_t^T &= \delta_{st} R_s & E w_s w_t^T &= \delta_{st} Q_s \\
 E v_s w_t^T &= \delta_{st} S_s & & & & \forall t, s, t \geq t_0
 \end{aligned}$$

São ainda dadas condições iniciais, por convenção em $t_0 = 0$,

$$E x_0 = \bar{x}_0; \quad E x_0 x_0^T = P_0 \quad x_0 \perp v_s, \quad x_0 \perp w_s \quad \forall s \geq 0$$

com o conhecimento da sequência de observações $Y_t = \{y_0, y_1, \dots, y_t\}$, pretende-se a estimativa recursiva $\hat{x}_{t+1|t}$ cujo erro $\tilde{x}_{t+1} = x_{t+1} - \hat{x}_{t+1|t}$ tenha covariância $\Sigma_{t+1|t} = E \tilde{x}_{t+1} \tilde{x}_{t+1}^T$ mínima, com a restrição de $\hat{x}_{t+1|t}$ ser um estimador afim em Y_t .

Para a solução do problema seguir-se-á de perto Kalman (1960), o que corresponde a versões apontadas subsidiariamente por Anderson e Moore (1976, secção 5.4) e Jazwinski (1970, pp.200-5). De facto, estes últimos, assim como a maioria dos autores, apresentam as deduções primárias em forma distinta da original. Reconhece-se, no entanto, que "the most authoritative derivation is given in the original papers by Kalman and Kalman and Bucy" (Priestley, 1981, p.807).

Determine-se primeiro a equação recursiva para $\hat{x}_{t+1|t}$. Os índices serão omitidos sempre que disso não resultar ambiguidade.

Aplicando (4.17) e (6.6),

$$\hat{E}[x_{t+1}|Y_t, 1] = \hat{E}[x_{t+1}|\tilde{Y}_t, 1] = \hat{E}[x_{t+1}|\tilde{y}_t, 1] + \hat{E}[x_{t+1}|\tilde{Y}_{t-1}, 1] - E x_{t+1} \quad (6.7)$$

Por (6.5):

$$\hat{E}[x_{t+1}|\tilde{y}_t, 1] = E x_{t+1} + \Sigma_{x\tilde{y}} \cdot \Sigma_{\tilde{y}\tilde{y}}^{-1} \cdot \tilde{y}_t$$

Para determinar a covariância entre x_{t+1} e \tilde{y}_t ($\Sigma_{x\tilde{y}}$) observa-se que, operando o projector em ambos os lados da equação de medida e subtraindo o resultado a essa equação, ter-se-á $\tilde{y} = H\tilde{x} + v$; substituindo x_{t+1} pela sua expressão explícita ter-se-á

$$\begin{aligned}\Sigma_{xy} &= \Sigma(Fx_t + Gw_t + \Gamma u_t, H\tilde{x}_t + v_t) \\ &= E[(Fx_t + Gw_t - FEx_t)(\tilde{x}_t^T H^T + v_t^T)] \\ &= E[Fx_t \tilde{x}_t^T H] + GS_t\end{aligned}$$

aonde se faz uso do facto de $Ew = E\tilde{x} = Ev = 0$, de Γu ser não estocástico e de x e v , assim como w e \tilde{x} , serem não contemporaneamente correlacionados. Finalmente, como $x_t = \hat{x}_{t|t-1} + \tilde{x}_t$

$$\begin{aligned}\Sigma_{xy} &= F(E\tilde{x}\tilde{x}^T + E\tilde{x}\tilde{x}^T)H + GS \\ &= F\Sigma_{t|t-1}H + GS_t\end{aligned}$$

desaparecendo $E\tilde{x}\tilde{x}^T$ uma vez que $\tilde{x} \perp \mathcal{Y}$ e $\tilde{x} \in \mathcal{Y}$. Para calcular $\Sigma_{\tilde{y}\tilde{y}}$ faça-se uso da já deduzida expressão de \tilde{y} . Uma vez que \tilde{x}_t e v_t são não correlacionados

$$\Sigma_{\tilde{y}\tilde{y}} = H\Sigma_{t|t-1}H^T + R_t$$

Para simplificar a expressão das matrizes de covariâncias defina-se $K_t := \Sigma_{xy} \cdot \Sigma_{\tilde{y}\tilde{y}}^{-1}$, matriz chamada *ganho de Kalman*

$$K_t = (F_t \Sigma_{t|t-1} H_t + G_t S_t)(H_t \Sigma_{t|t-1} H_t^T + R_t)^{-1} \quad (6.8)$$

Está agora calculada a primeira parcela do último membro de (6.7).

Quanto a $\hat{E}[x_{t+1} | \tilde{Y}_{t-1}, 1]$, a não correlação de w_t e \tilde{y}_{t-1} e o facto de $\Gamma_t u_t$ ser não estocástico permite deduzir

$$\begin{aligned}\hat{E}[x_{t+1} | \tilde{Y}_{t-1}, 1] &= \hat{E}[Fx_t + Gw_t + \Gamma u_t | \tilde{Y}_{t-1}, 1] \\ &= F\hat{x}_{t|t-1} + \Gamma u_t\end{aligned}$$

Substituindo as diversas expressões em (6.7) encontra-se a *equação recursiva de previsão em um passo* de \hat{x}_t :

$$\hat{x}_{t+1|t} = F_t \hat{x}_{t|t-1} + \Gamma_t u_t + K_t (y_t - H_t \hat{x}_{t|t-1}) \quad (6.9)$$

O processo iniciar-se-á com a estimativa prévia às observações

$$\hat{x}_{0|-1} = \bar{x}_0 \quad (6.10)$$

A matriz K_t utiliza as matrizes de covariâncias dos erros, que aliás são essenciais como medida estatística da qualidade esperada da previsão. Deduzir-se-ão de seguida as fórmulas recursivas para o seu cômputo.

Subtraindo o estimador a ambos os membros da equação de transição e substituindo y_t pela sua expressão em x_t e v_t , encontra-se

$$\hat{x}_{t+1} = (F_t - K_t H_t) \hat{x}_t + G_t w_t - K_t v_t$$

Como $E\hat{x}_t = Ew_t = Ev_t = 0$ e \hat{x}_t é não correlacionado contemporaneamente com os outros dois vectores aleatórios $\hat{x}_t \perp v_t$, pelo que a covariância de \hat{x}_{t+1} é dada por

$$\begin{aligned} \Sigma_{t+1|t} = & (F_t - K_t H_t) \Sigma_{t|t-1} (F_t - K_t H_t)^T + G_t Q_t G_t^T + \\ & + K_t R_t K_t^T - G_t S_t K_t^T - K_t S_t^T G_t^T \end{aligned} \quad (6.11)$$

ou, alternativamente, substituindo K_t pela sua expressão, encontra-se a chamada *equação matricial de Riccati* (discreta):

$$\begin{aligned} \Sigma_{t+1|t} = & F_t \Sigma_{t|t-1} F_t^T - (F_t \Sigma_{t|t-1} H_t^T + G_t S_t) (H_t \Sigma_{t|t-1} H_t^T + R_t)^{-1} \times \\ & \times (F_t \Sigma_{t|t-1} H_t + G_t S_t)^T + G_t Q_t G_t^T \end{aligned} \quad (6.11')$$

O início do processo far-se-á com

$$\Sigma_{0|-1} = P_0 \quad (6.12)$$

A escolha entre (6.11) e (6.11') é sobretudo problema computacional. Em geral (6.11) é preferível, pois tem como base a soma de duas expressões que envolvem matrizes semidefinidas positivas, o que tende a preservar esse carácter de $\Sigma_{t+1|t}$, mesmo face a erros de arredondamento⁽⁵⁾.

Está agora deduzido o algoritmo recursivo que se apresentará sob a forma de teorema.

(5) V. Anderson e Moore, 1979, pp.109-11

Teorema 6.1 (Filtro de Kalman — problema 6.1)

Nas condições do problema 6.1 o filtro afim de variância mínima que fornece as estimativas $\hat{x}_{t+1|t}$ com o conhecimento de Y_t é dado pelo processo recursivo seguinte

$$(i) \quad \hat{x}_{t+1|t} = F_t \hat{x}_{t|t-1} + \Gamma_t \mathbf{u}_t + K_t (\mathbf{y}_t - H_t \hat{x}_{t|t-1}) \quad (6.9)$$

$$\hat{x}_{0|-1} = \bar{x}_0 \quad (6.10)$$

$$(ii) \quad \Sigma_{t+1|t} = (F_t - K_t H_t) \Sigma_{t|t-1} (F_t - K_t H_t)^T + G_t Q_t G_t^T + K_t R_t K_t^T - G_t^T S_t K_t S_t^T G_t^T \quad (6.11)$$

$$\Sigma_{0|-1} = P_0 \quad (6.12)$$

sendo

$$K_t := (F_t \Sigma_{t|t-1} H_t + G_t S_t) (H_t \Sigma_{t|t-1} H_t + R_t)^{-1} \quad (6.8)$$

Demonstração: segue das considerações precedentes.

A solução do problema 6.1 produz estimadores preditivos em um passo e o algoritmo de Kalman é adaptável a um conjunto de outros problemas, nomeadamente os de filtragem e alisamento⁽⁶⁾. Interessa sobretudo entender a abordagem a problemas de previsão em mais do que um passo

Problema 6.2 (Previsão em n passos)

Nas condições do problema 6.1 pretende-se a estimação $\hat{x}_{t+n|t}$ com o conhecimento de Y_t .

Em n passos o sistema determinará n transições por intermédio da equação (6.2'). Defina-se

$$\Phi_{k,l} := F_{k-1} F_{k-2} \dots F_l, \quad k > l \quad (6.13)$$

$$\Phi_{k,k} := I$$

⁽⁶⁾ Para o estudo dos diversos problemas de filtragem ver sobretudo Jazwinski, 1970, caps. 7 e 8 e para problemas de alisamento ver Anderson e Moore, 1979, cap.7 e Brown, 1983, cap.8.

$$\Phi_{l,k}^{-1} := \Phi_{k,l}$$

O estado do sistema no momento $t+n$, com $n \geq 1$, expresso em função do momento $t+1$, de que já se conhece um estimador, será dado por

$$x_{t+n} = \Phi_{t+n,t+1} x_{t+1} + \sum_{i=t+1}^{t+n-1} \Phi_{t+n,i} (G_i w_i + \Gamma_i u_i)$$

como os "ruídos" w_i aqui considerados são não correlacionados com Y_t , a projecção ortogonal de x_{t+n} em $\{y_t, 1\}$ é dada por

$$\hat{x}_{t+n|t} = \Phi_{t+n,t+1} \hat{x}_{t+1|t} + \sum_{i=t+1}^{t+n-1} \Phi_{t+n,i} \Gamma_i u_i \quad (6.14)$$

Para obter a covariância do erro associado

$$x_{t+n} - \hat{x}_{t+n|t} = \Phi_{t+n,t+1} (x_{t+1} - \hat{x}_{t+1|t}) + \sum_{i=t+1}^{t+n-1} \Phi_{t+n,i} G_i w_i$$

basta reconhecer que $(x_{t+1} - \hat{x}_{t+1|t})$ e os w_i incluídos na expressão não estão correlacionados para obter

$$\Sigma_{t+n|t} = \Phi_{t+n,t+1} \Sigma_{t+1|t} \Phi_{t+n,t+1}^T + \sum_{i=t+1}^{t+n-1} \Phi_{t+n,i} G_i Q_i G_i^T \Phi_{t+n,i}^T \quad (6.15)$$

Teorema 6.2 (Filtro de Kalman — problema 6.2)

Nas condições do problema 6.2, tendo-se obtido $\hat{x}_{t+1|t}$ e $\Sigma_{t+1|t}$, o preditor afim de variância mínima de $\hat{x}_{t+n|t}$, $n > 1$, é dado por

$$\hat{x}_{t+n|t} = \Phi_{t+n,t+1} \hat{x}_{t+1|t} + \sum_{i=t+1}^{t+n-1} \Phi_{t+n,i} \Gamma_i u_i \quad (6.14)$$

tendo o respectivo erro a matriz de covariância:

$$\begin{aligned} \Sigma_{t+n|t} = & \Phi_{t+n,t+1} \Sigma_{t+1|t} \Phi_{t+n,t+1}^T + \\ & + \sum_{i=t+1}^{t+n-1} \Phi_{t+n,i} G_i Q_i G_i^T \Phi_{t+n,i}^T \end{aligned} \quad (6.15)$$

Demonstração: segue das considerações precedentes.

Vale a pena considerar com mais pormenor alguns problemas relacionados com o filtro que acaba de ser apresentado. Para não sobrecarregar a notação e facilitar o raciocínio, assim como para centrar as atenções no problema mais comum, faça-se primeiro uma simplificação nos pressupostos do modelo estocástico.

Problema 6.3 (Previsão com algoritmo recursivo em sistema estocástico invariante sem inputs determinísticos)

São dadas as equações de transição e de medida

$$x_{t+1} = Fx_t + Gw_t \quad (6.16')$$

$$y_t = Hx_t + v_t \quad (6.16'')$$

com

- x_t ($n \times 1$) vector aleatório de estado, desconhecido
- F ($n \times n$) matriz de transição constante, conhecida
- G ($n \times l$) matriz constante conhecida
- w_t ($l \times 1$) vector aleatório de "ruído", desconhecido
- y_t ($m \times 1$) vector aleatório observável
- H ($m \times n$) matriz constante, conhecida
- v_t ($m \times 1$) vector aleatório de "ruído", desconhecido

assume-se que todas as variáveis aleatórias são de segunda ordem e que os erros constituem "ruído branco":

$$E\mathbf{v}_t = E\mathbf{w}_t = \mathbf{0} \quad E\mathbf{v}_t \mathbf{v}_t^T = \delta_{st} R \quad E\mathbf{w}_t \mathbf{w}_t^T = \delta_{st} Q$$

$$E\mathbf{v}_t \mathbf{w}_l^T = \mathbf{0} \quad \forall t, s, l, \geq t_0$$

São dadas as condições iniciais em $t_0 = 0$

$$E\mathbf{x}_0 = \bar{\mathbf{x}}_0 \quad E\mathbf{x}_0 \mathbf{x}_0^T = P_0 \quad \mathbf{x}_0 \perp \mathbf{v}_s, \quad \mathbf{x}_0 \perp \mathbf{w}_s \quad \forall s \geq 0$$

com o conhecimento da sequência de observações Y_t pretende-se a estimativa recursiva $\hat{\mathbf{x}}_{t+1|t}$ e, posteriormente o preditor $\hat{\mathbf{x}}_{t+n|t}$, $n > 1$, ambos com erro de covariância $\Sigma_{-|t}$ mínima, com a restrição de $\hat{\mathbf{x}}$ ser um estimador ou preditor afim em Y_t .

Resumam-se as simplificações no modelo que alteram os pressupostos dos problemas anteriores. O sistema $\Sigma := (F, G, H)$ passou a ser *invariante*, o que corresponde ao carácter fixo das matrizes que o definem; o *input* determinístico desapareceu; os processos puramente aleatórios $\{\mathbf{w}_t\}$ e $\{\mathbf{v}_t\}$ consideram-se agora não correlacionados.

Teorema 6.3 (Filtro de Kalman — problema 6.3)

Nas condições do problema 6.3 o filtro afim de variância mínima que fornece as estimativas $\hat{\mathbf{x}}_{t+1|t}$ com o conhecimento de Y_t é dado pelo seguinte processo recursivo

$$(i) \quad \hat{\mathbf{x}}_{t+1|t} = F\hat{\mathbf{x}}_{t|t-1} + K_t(\mathbf{y}_t - H\hat{\mathbf{x}}_{t|t-1}) \quad (6.17)$$

$$\hat{\mathbf{x}}_{0|-1} = \bar{\mathbf{x}}_0 \quad (6.18)$$

$$(ii) \quad \Sigma_{t+1|t} = (F - K_t H)\Sigma_{t|t-1}(F - K_t H)^T + GQG^T + K_t R K_t^T \quad (6.19)$$

$$\Sigma_{0|-1} = P_0 \quad (6.20)$$

com

$$K_t := (F\Sigma_{t|t-1}H)^T (H\Sigma_{t|t-1}H + R)^{-1} \quad (6.21)$$

obtido $\hat{\mathbf{x}}_{t+1|t}$ e não considerando ou desconhecendo \mathbf{y}_τ , $\tau > t$, ter-se-á o preditor

$$(iii) \quad \hat{\mathbf{x}}_{t+n|t} = F^{n-1} \hat{\mathbf{x}}_{t+1|t} \quad (6.22)$$

com matriz de covariância do erro

$$(iv) \quad \Sigma_{t+n|t} = F^{n-1}\Sigma_{t+1|t}(F^{n-1})^T + \sum_{i=1}^{n-1} F^i G Q (F^i G)^T \quad (6.23)$$

Demonstração: resulta directamente da aplicação dos teoremas anteriores, na situação particular de constância de matrizes F, G, H , de inexistência de Γu_t e de $S_t \equiv 0$.

7. Propriedades e extensões do Filtro de Kalman

Algumas propriedades do Filtro de Kalman são evidentes na sua própria demonstração. Refira-se de novo o carácter CENTRADO das estimativas e o facto de elas serem LINEARES ou AFINS nas observações e, nessa classe, serem de VARIÂNCIA MÍNIMA. Algumas outras propriedades do algoritmo serão mais facilmente perceptíveis se o encararmos como *sistema*, exterior àquele que é descrito pelo modelo (6.1).

O sistema linear dinâmico (6.1) tem como *input* os processos $\{v_t\}$ e $\{w_t\}$ e como *output* o processo $\{y_t\}$. O Filtro de Kalman é um sistema que tem como *input* as observações e como *output* a sequência de estimativas $\{\hat{x}_t\}$. Tal como o sistema (6.1) o filtro é linear (afim), funciona em tempo discreto e é de dimensão finita. O reparo parece trivial mas o facto é que é a descrição adoptada para o sistema assim como o carácter recursivo do filtro que permitem a verificação dessas características. O filtro de Wiener, por exemplo, assume que as observações a computar têm início em $t \rightarrow -\infty$.

O sistema linear dinâmico é por vezes chamado de *representação markoviana* uma vez que $\{x_t\}$ é um processo de Markov. De facto, para todo o $l > 0$

$$p(x_t | x_{t-l}, x_{t-l-1}, \dots) = p(x_t | x_{t-l})$$

A verificação é imediata exprimindo x_t em termos de x_{t-l}

$$x_t = F_{t-1}F_{t-2} \dots F_{t-l} x_{t-l} + G_{t-1}w_{t-1} + G_{t-2}w_{t-2} + \dots + G_{t-l}w_{t-l}$$

Ou, introduzindo o operador de transição $\Phi_{t,l}$ — ver (6.13),

$$x_t = \Phi_{t,l}x_{t-l} + \sum G_{t-i}w_{t-i}$$

fica claro que os w_{t-i} ($0 \leq i \leq l$), que são independentes de w_{t-k} ($k > l$), concentram a parte aleatória da evolução $x_t \rightarrow x_t$; a parte não aleatória é expressa pela matriz de transição $\Phi_{t,l}$, pelo que o condicionamento por x_{t-l} fica perfeitamente determinado, sendo irrelevantes as variáveis x_{t-l-k} .

No entanto, nada garante a propriedade markoviana de $\{y_t\}$. Este último processo pode ser expresso em função de um único x_{t-l} mas não é, em geral, redutível a função de um único dos seus valores passados $\{y_{t-1}, y_{t-2}, \dots\}$. Por outro lado, é verificável que

a correlação entre y_t e y_{t-1} não se anula necessariamente com $|H| > 1$ (1). Na secção 9 mostrar-se-á precisamente como um processo $\{y_t\}$, não markoviano, pode ser expresso através de um sistema $\Sigma := (F, G, H)$, decompondo a sua evolução em processo markoviano $\{x_t\}$, a determinar, em equações de transição e de medida, com matrizes não estocásticas, e no "ruído branco" de *input*.

A extensão do Filtro de Kalman ao caso contínuo, desenvolvida em 1961 por Kalman e Bucy(2), mantém, em tudo o que é essencial, as características do sistema. Sem sequer abordar essa formulação, que é sobretudo aplicável à engenharia do controlo e não à estatística ou à econometria, refira-se apenas que o sistema contínuo é descrito pelas equações diferenciais estocásticas

$$\frac{dx(t)}{dt} = F(t)x(t) + G(t)w(t) \quad (7.1')$$

$$\frac{dy(t)}{dt} = H(t)x(t) + v(t) \quad (7.1'')$$

seguinte-se a convenção de representar por $z(t)$ variáveis de processo em tempo contínuo.

Em qualquer caso o filtro pressupõe uma definição de $\Sigma := (F, G, H)$ e um conjunto inicial de condições: x_0 e P_0 . A essas condições iniciais some-se o facto — revelado por simples inspecção das equações (6.18) e (6.19) — de $\Sigma_{t|t-1}$ e K_t serem calculáveis independentemente de Y_t . Isso significa que a sequência de matrizes de covariância de erro e a sucessão de ganhos de Kalman *podem ser calculadas antes do início do processo*. Mais importante ainda: nenhuma particular sequência de observações poderá, mais do que outra, eliminar a incerteza pressuposta no modelo. Essa incerteza é definida no momento inicial, pelo que a determinação de $\Sigma_{0| -1}$ é problema, nesse sentido, *externo* ao filtro. Isso não significa, contudo, que a precisão "real" das estimativas não possa ser crescente com as observações acumuladas e revelar-se superior à dada por $\Sigma_{t|t-1}$.

O Filtro de Kalman é um filtro *invariante* — ver (1.5) — no sentido em que as regras de geração das estimativas não iniciais — equações (6.17), (6.19) e (6.21) — não dependem de t . No entanto, mesmo no caso de o sistema ser linear invariante, como no problema 6.3, em que os *inputs* e *outputs* são estacionários, o mesmo não se passa com o *output* do filtro.

(1) V. Anderson e Moore, 1979, p.18.

(2) O trabalho pioneiro é Kalman e Bucy (1961). Para uma exposição mais descritiva ver Jazwinski (1970) ou Brown (1983). Um tratamento moderno e rigoroso encontra-se em Ruyngaert e Soong (1985)

Uma vez que $\Sigma_{t|t-1}$ e K_t não são constantes o mesmo acontecerá com o processo $\{\hat{x}_t\}$. Poderá, quando muito, registrar-se uma estacionaridade assintótica (V. e.g. Anderson e Moore, 1979, pp.80-5).

O filtro que foi deduzido é um preditor em um passo. Poder-se-ia, de forma equivalente, ter deduzido um operador de estimação *on line*, resolvendo o problema da filtragem. Obter-se-ia, em vez de (6.17), a expressão⁽³⁾

$$\hat{x}_{t|t} = \hat{x}_{t|t-1} + K_t(y_t - H\hat{x}_{t|t-1}) \quad (7.2)$$

com

$$\hat{x}_{t|t-1} = F\hat{x}_{t-1|t-1}$$

Ou ainda, uma vez que $H\hat{x}_{t|t-1} = \hat{y}_{t|t-1}$, em notação óbvia,

$$\hat{x}_{t|t} = \hat{x}_{t|t-1} + K_t\tilde{y}_t \quad (7.3)$$

Também o preditor em um passo pode ser reduzido a forma semelhante

$$\hat{x}_{t+1|t} = \hat{x}_{t+1|t-1} + K_t\tilde{y}_t \quad (7.4)$$

Tanto (7.3) como (7.4) são expressões de uma notável simplicidade e que resumem uma característica simples e fundamental da estimação ou previsão recursiva, com filtros lineares ou afins: a nova estimativa, depois de conhecida a última observação, é uma *combinação linear da estimativa que seria feita sem esse conhecimento e da inovação entretanto registrada*. No caso em apreço o ponderador da inovação é o ganho de Kalman, mas as expressões (7.3) e (7.4) podem evidentemente ter a forma alternativa

$$\hat{x}_{t|t} = (I - K_t H)\hat{x}_{t|t-1} + K_t y_t \quad (7.3')$$

$$\hat{x}_{t+1|t} = (F - K_t H)\hat{x}_{t|t-1} + K_t y_t \quad (7.4')$$

(3) V. a dedução em Anderson e Moore, 1979, p.40.

ponderação agora expressa em termos das estimativas e das observações. Na secção seguinte a importância conceptual e prática destas formulações será mais profundamente debatida.

Refira-se o caso particular dos processos gaussianos. O filtro passará a ser *estimador de variância mínima* e não apenas estimador *afim* de variância mínima, uma vez que, nesse caso, os dois conceitos coincidem, como foi amplamente referido. Toda a demonstração do teorema (6.1) e conseqüente é então feita substituindo o operador $\hat{E}[x_{t|t-1}|Y_t, 1]$ pelo operador $E[x_{t|t-1}|Y_t]$, como aliás é prática (restritiva) em todos os textos clássicos que houve acesso, com excepção do de Kalman. É importante notar que o filtro dá então a esperança condicional e a covariância condicional, pelo que fornece, de forma perfeitamente definida, processo para *atualização de toda a distribuição de probabilidade condicional* de x_t . Em contraste, no caso não gaussiano, $\hat{x}_{t|t-1}$ não é o valor esperado condicional nem $\Sigma_{t|t-1}$ é a matriz de covariância condicionada. O primeiro é apenas o estimador afim de variância mínima, enquanto o segundo é a matriz de covariância, não condicional, do erro.

A demonstração que foi efectuada é pois perfeitamente geral. A abordagem de Kolmogorov da seqüência de inovações, seguida de um procedimento recursivo baseado na teoria dos projectores ortogonais em espaços de Hilbert, parece ser o método mais rigoroso e simultaneamente o mais fecundo. É o que começa hoje a ser reconhecido⁽⁴⁾. Além da demonstração restringida ao caso gaussiano (e.g. Jazwinski, 1970 e Anderson e Moore, 1979) há um sem número de outras demonstrações do filtro, fazendo uso das mais variadas técnicas. Vale a pena delinear muito sinteticamente algumas delas e chamar desde já a atenção para a dedução bayesiana que será efectuada na secção seguinte. É que o facto de o filtro poder ser reencontrado através de um sem número de caminhos, técnicas e pressupostos alternativos, é demonstrador das extensões variadas que possui.

MÍNIMOS QUADRADOS RECURSIVOS. É este talvez o procedimento com prioridade. Com efeito Swerling, em 1959 (*op.cit.*), deduz um conjunto de equações recursivas com base na teoria dos mínimos quadrados, podendo o seu algoritmo ser interpretado como prefigurando o Filtro de Kalman (Cf. Sorenson, 1970). No contexto econométrico o filtro foi rederivado por Duncan e Horn em 1972. Uma exposição do ponto de vista de Swerling e de seus seguidores encontra-se em Jazwinski (1970, pp.205-7). Uma descrição simples da dedução de Duncan e Horn encontra-se em Harvey (1981a, pp.104-10).

(4) No seguimento de Kalman (1960) os trabalhos base são os de Kailath (1968 e 1970), que generalizam o chamado "innovations approach". Mais recentemente veja-se Young, 1975, p.169, todo o desenvolvimento, no caso contínuo, de Ruymgaart e Soong, 1985, e ainda Brockwell e Davis, 1987 a.

Considere-se o modelo estudado por Duncan e Horn (1972), embora adaptando a notação e partindo do momento inicial.

$$y = Hx + v, \quad Ev = 0, \quad Evv^T = \sigma^2 \Omega$$

O estimador $\hat{x} = (H^T \Omega^{-1} H)^{-1} H^T \Omega^{-1} y$ não será rigorosamente o estimador de Aitken⁽⁵⁾ — pois x pode ser estocástico — mas paraleliza-o, na medida em que se demonstra que o erro da sua aplicação tem os momentos

$$E\hat{x} = 0 \quad E\hat{x}\hat{x}^T = \sigma^2 (H^T \Omega^{-1} H)^{-1}$$

Admita-se que existe informação *a priori* sobre x , condensada em \hat{x}_0 e tal que $E\hat{x}_0 = 0$ e $E\hat{x}_0\hat{x}_0^T := \sigma^2 P_0$. A nova estimação de x , fazendo uso do modelo de Aitken na equação alargada

$$\begin{bmatrix} \hat{x}_0 \\ y_1 \end{bmatrix} = \begin{bmatrix} I \\ H \end{bmatrix} x + \begin{bmatrix} \tilde{x}_0 \\ v_1 \end{bmatrix} \quad (7.5)$$

fornece

$$\hat{x}_{1|0} = (P_0^{-1} + H^T \Omega^{-1} H)^{-1} (P_0^{-1} \hat{x}_0 + H^T \Omega^{-1} y_1)$$

Se x fosse não estocástico estar-se-ia perante um caso particular do processo de *estimação mista* de Theil e Goldberg⁽⁶⁾. É importante frizar que *não* é esse geralmente o caso.

Introduzindo a equação de transição $x_t = Fx_{t-1} + Gw_t$ Duncan e Horn procuram a filtragem $\hat{x}_{t|t}$ (e não o preditor em um passo $\hat{x}_{t|t-1}$). Partindo de $\hat{x}_{1|0} = Fx_{0|0}$ derivam automaticamente

$$\Sigma_{1|0} = F\Sigma_{0|0}F^T + GQG^T$$

No momento 1 o estimador de Aitken fornece

$$\hat{x}_{1|1} = (\Sigma_{1|0}^{-1} + H^T \Omega^{-1} H)^{-1} (\Sigma_{1|0}^{-1} \hat{x}_{1|0} + H^T \Omega^{-1} y_1) \quad (7.6)$$

(5) V. e.g. Johnston, 1984, pp.287-93

(6) H. Theil A.S. Goldberg, 1961, pp.65-78.

o que inicia o processo recursivo para obter as estimativas $\hat{x}_{t|t}$ expressas em $\hat{x}_{t-1|t-1}$, ou pela equação de transição, expressas em $\hat{x}_{t|t-1}$. Utilizando o lema da inversão matricial (a ser demonstrado na secção seguinte) simplifica-se depois a fórmula recursiva para $\hat{x}_{t|t}$ e deduz-se a recursão para $\Sigma_{t|t}$. Não é importante referir pormenorizadamente esses passos. No caso geral, $t > 1$, o procedimento é exactamente o mesmo, simplesmente a equação alargada passará a incluir Y_t , o que apenas sobrecarrega a notação.

ESTIMADOR LINEAR DE VARIÂNCIA MÍNIMA. Esta abordagem parece ter sido iniciada por Battin em 1962⁽⁷⁾ e corresponde a uma linha de ataque extremamente directa. Uma exposição encontrar-se-á em Jazwinski (1970) e Sage e Melsa (1973) e também em Neves (1984).

Sabendo-se que $\hat{x}_{t|t-1} = Fx_{t-1|t-1}$ e que $\Sigma_{t|t-1} = F\Sigma_{t-1|t-1}F + GQG^T$ procura-se o estimador recursivo linear em $\hat{x}_{t|t-1}$ e y_t . Por conveniência exprime-se essa combinação linear, de *filtragem*, na forma

$$\hat{x}_{t|t} = \hat{x}_{t|t-1} + K_t[y_t - H\hat{x}_{t|t-1}] \quad (7.7)$$

e procura-se K_t que minimiza a *variância* somada dos erros

$$E\tilde{x}_t^T \tilde{x}_t = \text{tr}\Sigma_{t|t}, \quad \tilde{x}_t = x_t - \hat{x}_{t|t}$$

Substituindo na equação recursiva e tendo em conta a restrição $E\tilde{x}_t = 0$, após algumas manipulações algébricas encontra-se uma expressão com constante, K_t^1 :

$$K_t = \Sigma_{t|t-1}H^T(H\Sigma_{t|t-1}H^T + R)^{-1} + K_t^1$$

A minimização de $\text{tr}\Sigma_{t|t}$ fornece uma parcela em K_t^1 : $K_t^1(H\Sigma_{t|t-1}H^T + R)K_t^1$. Como a expressão entre parêntesis é matriz semidefinida positiva (envolve matrizes de covariâncias) a minimização do seu traço é feita com $K_t^1 = 0$, o que permite imediatamente encontrar a expressão de ganho de Kalman. A forma recursiva para $\Sigma_{t|t}$ é então deduzida por cálculo directo sobre a equação que fornece $\hat{x}_{t|t}$.

Tal como no caso dos mínimos quadrados recursivos esteve em causa a filtragem. A previsão em um passo é directa pela aplicação da matriz de transição: $\hat{x}_{t+1|t} = F\hat{x}_{t|t}$.

(7) R. H. Battin, (1962), "Statistical optimizing navigation procedure for space flight", *ARS J.*, pp.1681-96.

A terminar esta secção faça-se apenas uma breve referência a um largo campo teórico que ficou por abordar: o da teoria dos sistemas lineares. Aí se incluem problemas como os da estabilidade ou instabilidade dos processos dinâmicos, os da convergência ou divergência do filtro e da observabilidade e controlabilidade dos processos⁽⁸⁾.

Muito rapidamente, e apenas como nota à importância do problema, refiram-se os conceitos de divergência por modelização incorrecta e o da observabilidade.

Em caso de incorrecta modelização o filtro pode divergir, ou seja, pode a assumida matriz de covariância $\Sigma_{t|t-1}$ permanecer limitada, enquanto a verdadeira covariância dos erros aumenta em relação a $\Sigma_{t|t-1}$, possivelmente sem limite.

Relembre-se que o processo $\{x_t\}$ é não observável e que a matriz de covariâncias é apenas computada no pressuposto de adequabilidade da definição $\Sigma := (F, G, H)$ ao processo modelizado e de correcção das condições iniciais x_0 e P_0 .

Significa isso que nada garante a igualdade

$$\Sigma_{t+1|t} = E\bar{x}_{t+1}\bar{x}_{t+1}^T$$

pois o que é conhecido não é a sequência $\{\bar{x}_t := x_t - \hat{x}_t\}$ mas apenas uma sucessão de *presumíveis inovações*, por exemplo do tipo $\bar{x}_t^p := \hat{x}_{t|t} - \hat{x}_{t|t-1}$.

⁽⁸⁾ V. sobretudo Kalman (1963) e também Jazwinski (1970, pp.231-255) e Anderson e Moore (1979, pp.62-88).

Exemplo 7.1 (Divergência do filtro)

Considere-se o processo

$$x_t = x_{t-1} + w_t$$

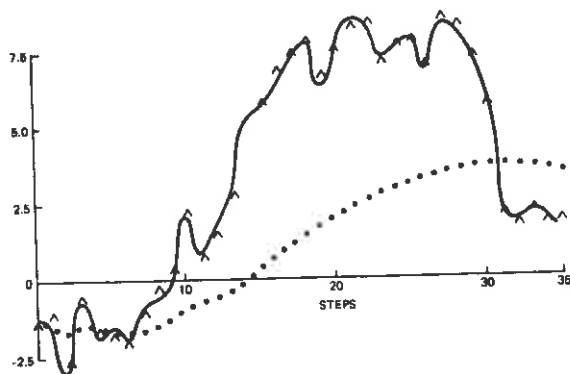
$$y_t = x_t + v_t$$

em que $\{w_t\}$ e $\{v_t\}$ são processos puramente aleatórios não correlacionados. Admita-se uma determinada realização deste sistema estocástico procurando-se estimar o seu estado pelo modelo

$$\hat{x}_t = \hat{x}_{t-1}$$

$$\hat{y}_t = \hat{x}_t + v_t$$

É admissível que as estimativas possam conter erros crescentes. Na figura seguinte mostra-se o resultado de uma simulação empreendida por Brown (1983, p.235) assinalando-se as estimativas do último modelo por pontos. As estimativas empreendidas com um filtro baseado no modelo efectivamente simulado estão representadas por chapéus (^) e o processo está descrito pela linha contínua.



Uma solução que parece óbvia para evitar erros de divergência por incorrecta modelização consiste em aumentar o número de variáveis para prever possíveis oscilações do estado. No exemplo anterior, poder-se-ia, eventualmente, admitir que o nível do processo $\{x_t\}$ não fosse constante, assim como se poderia pensar que ele poderia ter um crescimento linear, quadrático, etc. Essa formulação geral talvez impedisse erros de divergência.

Infelizmente o problema é bem mais complexo. O incremento das variáveis observadas pode conduzir a *sistemas não observáveis* ⁽⁹⁾.

⁽⁹⁾ Para uma definição rigorosa do conceito de observabilidade ver Jazwinski, 1970, pp.231-2.

Se esses sistemas são instáveis os erros de estimação também serão não estáveis (Brown, 1983, p.237). Por tudo isto se intui a extraordinária importância da teoria dos sistemas e a sua relação com o problema de identificação em econometria⁽¹⁰⁾.

Exemplo 7.2 (Desvios sistemáticos de estimação)

Admita-se o sistema

$$y_t = x_t - x_{t-1} + v_t$$

$$x_t = a + x_{t-1} + w_t$$

com a condição inicial $\hat{x}_{0|-1} = \bar{x}_0$ e $\Sigma_{0|-1} = P_0$. No caso geral $x_0 \neq \bar{x}_0$, e como as observações apenas medem o crescimento absoluto de x_t é natural que as estimativas do valor da variável continuem sempre enviesadas.

Encerre-se aqui este parêntesis e retorne-se à interpretação do filtro como algoritmo preditor.

(10) V. uma discussão do problema em Kalman (1982).

8. A previsão recursiva na visão bayesiana

Até ao momento a justificação da previsão recursiva tem sido apenas de ordem computacional. Quer se trate de encontrar um algoritmo para resolver o sistema de Wiener-Hopf quer se trate da determinação de um processo simples de incorporação de novas observações, o que tem estado em causa tem sido a simplicidade da computação. No filtro de Kalman já não é exactamente isso que se passa mas o aprofundamento da justificação da recursividade só agora vai ser feito.

Em certo sentido a recursividade da estimação está indissolivelmente ligada ao modelo adoptado para o sistema estocástico. O problema é em geral tão incompreendido que vale a pena dedicar-lhe um pouco de atenção, mesmo correndo o risco de apresentar resultados aparentemente triviais.

Considere-se, como exemplificação inicial, o caso simples de um processo $\{x_t\}$, estacionário em média, para o qual se pretendem preditores em um passo, centrados

$$\{x_t\}: \quad E x_t = a \quad E x_t x_s < \infty \quad (8.1')$$

$$\hat{x}_{t|t-1}: \quad E \hat{x}_{t|t-1} = E x_t \quad (8.1'')$$

Restrinja-se a classe dos preditores, de forma a serem filtros lineares e invariantes, expressáveis na forma recursiva em um passo

$$\hat{x}_{t+1|t} = \alpha_t x_t + \beta_t \hat{x}_{t|t-1}, \quad (8.1''')$$

Com estas simples condições mostrar-se-á de seguida que existe uma relação linear entre α_t e β_t e que, além disso, esses coeficientes são constantes.

Como se pretende, qualquer que seja o valor esperado do processo, que $E \hat{x}_{t+1|t} = E x_{t+1} = E x_t$, e como o mesmo se deverá passar com $E \hat{x}_{t|t-1}$, basta operar a esperança matemática nos dois membros de (8.1''') para concluir que $\alpha_t + \beta_t = 1$, necessariamente. Substituindo agora $\hat{x}_{t|t-1}$ pela sua expressão recursiva e fazendo o mesmo, sucessivamente, para os preditores antecedentes, encontra-se

$$\hat{x}_{t+1|t} = (1 - \beta_t) x_t + \beta_t (1 - \beta_{t-1}) x_{t-1} + \beta_t \beta_{t-1} (1 - \beta_{t-2}) x_{t-2} + \dots \quad (8.2')$$

$$\hat{x}_{t|t-1} = (1 - \beta_{t-1})x_{t-1} + \beta_{t-1}(1 - \beta_{t-2})x_{t-2} + \beta_{t-1}\beta_{t-2}(1 - \beta_{t-3})x_{t-3} + \dots \quad (8.2'')$$

Como se admitiu que o filtro era invariante, obrigatoriamente $\beta_t = \beta_{t-1} = \beta_{t-2} = \dots = \beta$, pois terão de ser válidas as igualdades

$$\begin{aligned} (1 - \beta_t) &= (1 - \beta_{t-1}) \\ \beta_t(1 - \beta_{t-1}) &= \beta_{t-1}(1 - \beta_{t-2}) \\ \beta_t\beta_{t-1}(1 - \beta_{t-2}) &= \beta_{t-1}\beta_{t-2}(1 - \beta_{t-3}) \\ &\dots \end{aligned}$$

Por outro lado, como as estimativas são centradas, a soma dos ponderadores em (8.2'') terá de igualar a unidade. Pretendendo-se, por extensão, contemplar o caso de o filtro não ser finito, isso implica que

$$\sum_{i=0}^{\infty} (1 - \beta)\beta^i = 1$$

Ora a série só converge com $|\beta| < 1$. Também não se poderia garantir que as variáveis aleatórias $\hat{x}_{t|t-1}$, expressas na forma (8.2), ficassem definidas, em média quadrática, se $|\beta| \geq 1$ ⁽¹⁾.

A forma explícita das estimativas será pois dada pelo filtro

$$\hat{x}_{t+1|t} = (1 - \beta)x_t + \beta(1 - \beta)x_{t-1} + \beta^2(1 - \beta)x_{t-2} + \dots \quad (8.3)$$

e a forma recursiva pela equação

$$\hat{x}_{t+1|t} = (1 - \beta)x_t + \beta\hat{x}_{t|t-1} \quad (8.4)$$

No caso de $0 < \beta < 1$ estas equações são a base da conhecida técnica do ALISAMENTO EXPONENCIAL (Brown, 1963).

(1) Admita-se $E x_t x_t = \delta_{tt} \sigma^2$. Como o processo $\{x_t\}$ é de segunda ordem a série $(1 - \beta) \sum \beta^i x_{t-i}$ converge, em m.q., se a série de potências $\sum \beta^{2i} < +\infty$. A condição suficiente é então dada por $|\beta| < 1$. V. Nota Anexa 6.

A dedução que acaba de ser feita mostra que essa técnica pode ser apresentada sob um fundamento absolutamente geral. Vale a pena destacar esse resultado.

Teorema 8.1 (Alisamento exponencial)

Com o processo de segunda ordem $\{x_t\}$, estacionário em média, um filtro linear, invariante, expressável em *forma recursiva*, linear em um passo, extensível a número ilimitado de variáveis, e que produza estimativas centradas é, *necessariamente*, um filtro em que o peso das observações decresce geometricamente à medida que estas se distanciam do momento de previsão.

Demonstração: Segue das considerações precedentes. Se se admite que o filtro é sempre finito — o que é diferente de admitir um momento de inicialização do processo recursivo — a conclusão já não é válida; mas também, nesse caso, a forma recursiva não será invariante.

A demonstração do teorema não oferece dificuldades, sobretudo depois dos trabalhos de Brown, Holt, Winters e outros sobre a técnica de alisamento exponencial. No entanto, tanto quanto se conhece, este não se encontra explicitamente formulado. Reputa-se que o teorema tem uma importância teórica qualitativa, pois mostra a existência de uma relação estreita entre a forma recursiva dos estimadores e o peso atribuível às observações. A relação entre essa forma dos estimadores e o tipo de modelos subjacente não é agora difícil.

Com efeito, e para continuar o estudo apenas pelo caso mais simples, de modelo sem tendência, repare-se que a formulação (8.1') é compatível com diferentes tipos de processos. Confrontem-se dois casos.

Em primeiro lugar o MODELO ECONOMETRICO CLÁSSICO, que terá a forma simplificada

$$x_t = a + \varepsilon_t \quad E\varepsilon_t = 0, \quad E\varepsilon_s \varepsilon_t = \delta_{st} \sigma^2 \quad (8.5)$$

e, em segundo lugar, o chamado MODELO DE MUTH (1960, p.302), que terá a forma

$$\begin{aligned} x_t &= \mu_t + \varepsilon_t & E\varepsilon_t &= E\eta_t = E\varepsilon_t \eta_t = 0 \\ \mu_t &= \mu_{t-1} + \eta_t & E\varepsilon_s \varepsilon_t &= k E\eta_s \eta_t = \delta_{st} \sigma^2 \end{aligned} \quad (8.6)$$

com a condição inicial $\mu_0 = a$.

No caso do modelo econométrico o preditor centrado, linear e de variância mínima é fornecido pela equação

$$\hat{x}_{t+1|t} = \sum_{i=1}^t x_i/t, \quad (8.7)$$

como é sobejamente conhecido. Tratando-se de um filtro com pesos homogêneos, $1/t$, isso significa, em particular, que um filtro com pesos diferenciados *não fornece* preditores de variância mínima para o modelo (8.5).

Em contraste, Muth (1960, pp.303-4) demonstrou que o filtro ótimo (de variância mínima), para o modelo em que o nível do processo evolui segundo "passeio aleatório" — modelo (8.6) — é o preditor de alisamento exponencial e não o OLS.

Há pois uma correspondência entre o peso ótimo das observações e o modelo subjacente.

O problema pode ser imediatamente intuído se se verificar que a covariância entre x_{t+1} e x_{t-i} é constante no modelo econométrico⁽²⁾ e decrescente no modelo de Muth. No primeiro caso qualquer observação fornece um dado de valor informativo equivalente para a estimação de $a = Ex_{t+1}$, enquanto no segundo a informação fornecida com x_{t-i} tem valor decrescente com i ($i > 0$).

Reduzindo ambos os casos a forma tratável pelo método dos mínimos quadrados ter-se-á, para o modelo econométrico

$$\begin{aligned} & \dots \\ x_{t-n} &= a + \varepsilon_{t-n} \\ & \dots \\ x_{t-1} &= a + \varepsilon_{t-1} \\ x_t &= a + \varepsilon_t \end{aligned} \quad (8.8)$$

e, transformando o modelo de Muth de forma adaptável à estimação de $\mu_{t+1} = Ex_{t+1}$,

(2) Aliás nula

$$x_{t-n} = \mu_{t+1} - \eta_t - \eta_{t-1} - \dots - \eta_{t-n} + \varepsilon_{t-n}$$

...

$$x_{t-1} = \mu_{t+1} - \eta_{t+1} - \eta_t + \varepsilon_t$$

$$x_t = \mu_{t+1} - \eta_{t+1} + \varepsilon_t$$

No primeiro caso a matriz de covariância correspondente às observações é escalar: $\Omega = \sigma_\varepsilon^2 I$. No segundo caso essa matriz terá a forma de $\Omega = \sigma_\varepsilon^2 I + \sigma_\eta^2 Q$, com Q dada por

$$Q = \begin{bmatrix} n & n-1 & \dots & 2 & 1 \\ n-1 & n-1 & \dots & 2 & 1 \\ \dots & & & & \\ 1 & 1 & \dots & 1 & 1 \end{bmatrix} \quad (8.10)$$

À estimação de α aplicar-se-á o OLS enquanto à estimação de μ_{t+1} se aplicará o modelo de Aitken (V. e.g. Johnston, 1984, p.291) o qual, como é óbvio face a Q , ponderará mais fortemente as últimas observações.

Não vale a pena prosseguir e apresentar a demonstração de Muth. Parece já clara a correspondência entre o alisamento exponencial e dada classe de modelos, que não é extensível aos esquemas econométricos tradicionais.

O que se disse para modelos sem tendência estender-se-á sem dificuldade aos casos em que a tendência linear evolua segundo "passeio aleatório" — MODELO DE THEIL E WAGE⁽³⁾:

$$\begin{aligned} x_t &= \mu_t + \varepsilon_t \\ \mu_t &= \mu_{t-1} + \beta_{t-1} + \eta_t \\ \beta_{t-1} &= \beta_{t-2} + \xi_{t-1} \end{aligned} \quad (8.11)$$

ou ainda para modelos mais complexos. No caso (8.11), por exemplo, será aplicável uma técnica conhecida como alisamento exponencial duplo (V. e.g. Gilchrist, 1976, pp.72 e segs.), mas o princípio de base é semelhante.

Na generalidade dos casos é possível adaptar esses modelos a uma forma do tipo da (8.9) e, em seguida, aplicar-lhe o modelo de Aitken. É nisso que consiste o chamado *método*

(3) V. Theil e Wage (1964), "Some observations on adaptive forecasting", *Management Science*, 6, pp.324-42.

dos mínimos quadrados ponderados ⁽⁴⁾ que dá um peso às observações decrescente com a sua antiguidade. Essa ponderação pode, aliás, ser obtida directamente a partir dos coeficientes β de alisamento. A técnica, aplicada sem pressupor um modelo compatível, é um exemplo de procedimento estatístico sem fundamentação probabilística, conceito a que se fez referência no final da primeira secção.

Estranhamente, a questão da correspondência entre modelos estocásticos e técnicas de previsão é muitas vezes esquecida, encontrando-se mesmo alguns erros⁽⁵⁾. A eficácia de métodos como os do alisamento deve ser entendida como indicando que o fenómeno em estudo deveria ser modelizado de acordo com um esquema que compreenda a variabilidade dos parâmetros.

De acordo com o resultado expresso no teorema 8.1 pode-se agora raciocinar em caminho inverso ao usual. Em lugar de comparar empiricamente processos preditivos e optar por filtros recursivos apenas por se concluir que eles permitem melhores resultados práticos, deve-se ir mais longe e concluir que essa melhoria prática significa que o processo, a ser adequadamente modelizado, deverá contemplar uma evolução estocástica.

O raciocínio habitual em previsão, de considerar que as observações mais recentes devem ser mais fortemente ponderadas, pois elas transportam informação estatisticamente mais significativa, terá de ser interpretado como obrigando à opção por modelos teóricos estocásticos. E, se o melhor filtro recursivo é precisamente o que pondera as observações de forma decrescente com a sua antiguidade (teorema 8.1), isso significa que *o sucesso da metodologia recursiva está indissoluvelmente ligado a modelos estocásticos nos parâmetros*⁽⁶⁾. Por aqui se vê que o Filtro de Kalman representa muito mais do que um algoritmo de previsão.

Abandone-se agora o caso particular de um simples processo estocástico e retorne-se ao modelo de Kalman. A equação preditiva é, de certa forma, uma generalização da fórmula de alisamento exponencial e, tal como esta, pode ser expressa em duas formas alternativas. Recorde-se (7.4) e (7.4') e reescrevam-se essas expressões do modo seguinte

(4) V. Montgomery e Johnson, 1976, pp.76-96.

(5) Montgomery e Johnson (1976, p.76) cometem expressamente a incorência de adoptar mínimos quadrados ponderados no contexto de um modelo econométrico homoescedástico. Mesmo Brown (1963, pp.97, 101) não parece nada claro neste problema. Uma discussão rigorosa das técnicas óptimas de alisamento, adaptáveis aos diversos tipos de processos, pode encontrar-se em Granger e Newbold, 1986, pp.172-5.

(6) Por parâmetros entendem-se aqui os usuais parâmetros descritivos da econometria: nível, tendência, taxa de crescimento, etc. . .

$$\hat{x}_{t+1|t} = F\hat{x}_{t|t-1} + K_t\tilde{y}_t \quad (8.12)$$

$$\hat{x}_{t+1|t} = (F - K_tH)\hat{x}_{t|t-1} + Ky_t \quad (8.13)$$

A primeira expressão é por vezes chamada *fórmula da correção do erro*, e a segunda a *fórmula de recorrência* (Gilchrist, 1976, p.43). Ambas terão uma interpretação bayesiana evidente, pois exprimem um processo de aprendizagem em que as estimativas são sequencialmente corrigidas pela incorporação de novas observações. Essa interpretação, para ser rigorosamente tratada, implica uma definição da natureza estatística do sistema. Admite-se, habitualmente, que o sistema seja gaussiano, o que significa que todos os seus vectores aleatórios seguem uma multinormal.

Neste quadro é possível resolver o problema de encontrar a distribuição a *posteriori* de x_{t+1} incorporando a nova observação y_t . O fundamento da abordagem bayesiana está no cálculo dessa distribuição a *posteriori* que é, como se sabe, proporcional ao produto da *verosimilhança* pela distribuição a *priori* (7). O cálculo recursivo das previsões torna-se equivalente à aplicação da fórmula

$$p(x_{t+1}|Y_t) \propto p(y_t|x_{t+1}, Y_{t-1}) \times p(x_{t+1}|Y_{t-1}) \quad (8.14)$$

A demonstração bayesiana é mais directa se se encarar o problema da filtragem. A previsão será imediata fazendo $E\hat{x}_{t+1|t} = F\hat{x}_{t|t}$ e $\Sigma_{t+1|t} = F\Sigma_{t|t}F^T + GQG^T$. Também é vantajoso utilizar as inovações em lugar das observações, e tal procedimento não levanta problemas pois estas são reconstrutíveis a partir daquelas. A fórmula bayesiana será então reescrita na forma

$$p(x_t|\tilde{y}_t, Y_{t-1}) \propto p(\tilde{y}_t|x_t, Y_{t-1}) \times p(x_t|Y_{t-1}) \quad (8.15)$$

(7) V. e.g. Murteira, 1983, pp.30-2.

A proporcionalidade é representada pelo símbolo \propto .

Teorema 8.2

Em sistema gaussiano, o filtro de Kalman forma as equações recursivas que permitem actualizar a distribuição *a posteriori* de x_t a partir da sua distribuição *a priori* e da *verossimilhança*, condicionadas todas pelo registo das observações Y_t .

Demonstração: Será apenas delineada⁽⁸⁾ e, por simplicidade, referida ao problema 6.3.

Decomponha-se a solução do problema nos seus vários passos.

- (i) No momento $t-1$ ter-se-á a distribuição *a posteriori* para x_{t-1} , perfeitamente determinada pela média $\hat{x}_{t-1|t-1}$ e pela variância $\Sigma_{t-1|t-1}$. Como o sistema é gaussiano é a distribuição condicional que está em causa (V. secção 6). Essa distribuição é posterior ao conhecimento de y_{t-1} .
- (ii) Calcula-se então a distribuição *a priori* para x_t , que terá valor esperado $\hat{x}_{t|t-1} = F\hat{x}_{t-1|t-1}$ e covariância $\Sigma_{t|t-1} = F\Sigma_{t-1|t-1}F^T + GQG^T$.
- (iii) Calcula-se também a *verossimilhança do erro* ($\tilde{y}_t|x_t, Y_{t-1}$); o seu valor esperado será $y_t - Hx_t = H(x_{t-1} - F\hat{x}_{t-1|t-1})$ e a sua covariância $H(F\Sigma_{t-1|t-1}F^T + GQG^T)H^T + R$.
- (iv) O cálculo da distribuição *a posteriori* de x_t é grandemente simplificado utilizando o lema 3.2. O vector multinormal $[x_t \ \tilde{y}_t]^T$, condicionado por Y_{t-1} , terá como valor médio $[F\hat{x}_{t-1|t-1} \ 0]^T$ e como matriz de covariância

$$\begin{bmatrix} F\Sigma_{t-1|t-1}F^T + GQG^T & (F\Sigma_{t-1|t-1}F^T + GQG^T)H^T \\ H(F\Sigma_{t-1|t-1}F^T + GQG^T) & H(F\Sigma_{t-1|t-1}F^T + GQG^T)H^T + R \end{bmatrix}$$

pelo que a aplicação do referido lema permite concluir que a distribuição *a posteriori* de $(x_t|\tilde{y}_t, Y_{t-1}) \equiv (x_t|Y_t)$ tem como média

$$F\hat{x}_{t-1|t-1} + K_t^* \tilde{y}_t \tag{8.15}$$

e como covariância

$$\begin{aligned} \Sigma_{t|t} &= F\Sigma_{t-1|t-1}F^T + GQG^T - K_t^* H(F\Sigma_{t-1|t-1}F^T + GQG^T) = \\ &= (F - K_t^* H)(\Sigma_{t-1|t-1}F^T + GQG^T) \end{aligned} \tag{8.16}$$

⁽⁸⁾ V. Meinhold e Singpurwalla, 1983.

em que K_t^* é definido por

$$K_t^* := (F\Sigma_{t-1|t-1}F^T + GQG^T)H^T (R + HF\Sigma_{t-1|t-1}F^T H^T + HGQG^T H^T)^{-1} \quad (8.17)$$

O Filtro de Kalman pode pois enquadrar-se numa visão bayesiana da estimação e previsão. Ele pode ser visto como um processo de formar uma primeira estimativa sobre o "estado da natureza" — x_t — e de adicionar uma correcção a essa estimativa, feita após o conhecimento da nova observação — y_t . A correcção é determinada pelo sucesso que a estimativa *a priori* teve na previsão da observação ⁽⁹⁾.

Um outro procedimento bayesiano, que consiste em formar uma primeira *estimativa difusa*, ou "não informativa", quando o conhecimento das variáveis de estado é muito insuficiente (Berger, 1985, pp.82 e segs.), encontra também aplicação directa no filtro de Kalman. Para isso será necessário procurar expressões alternativas às usuais no filtro, pois uma "prior" difusa corresponderá à situação de $\Sigma_{0|-1}$ ser não limitada, e isso não é comportável pelas equações de recorrência já deduzidas. O caminho consistirá em trabalhar com as matrizes inversas de Σ , chamadas essas **MATRIZES INFORMATIVAS** ⁽¹⁰⁾, de forma a que as equações do filtro possam admitir $\Sigma_{0|-1}^{-1} = 0$.

Será necessária a utilização do seguinte lema.

Lema 8.1 (Lema da inversão matricial)

Sejam A, B e C matrizes, respectivamente de ordem $(n \times n)$, $(p \times p)$ e $(n \times p)$. Ter-se-ão os seguintes resultados sempre que as inversas assinaladas existam

$$(I + ACB^{-1}C^T)^{-1}A = (A^{-1} + CB^{-1}C^T)^{-1} = A - AC(C^T AC + B)^{-1}C^T A \quad (8.18)$$

$$(I + ACB^{-1}C^T)^{-1}ACB^{-1} = (A^{-1} + CB^{-1}C^T)^{-1}CB^{-1} = AC(C^T AC + B)^{-1} \quad (8.19)$$

⁽⁹⁾ Foi aliás esta correspondência do filtro com a metodologia bayesiana que levou Harrison e Stevens, no seu trabalho pioneiro sobre previsão econométrica pelo filtro de Kalman (1971, 1976), a designar o seu método como "Bayesian Forecasting". É curioso que, no seguimento do debate que seguiu à apresentação do seu trabalho, tenham sentido necessidade de responder: "however, we are not religious bayesians". V. Harrison e Stevens, 1976, pp.205, 231 e 241.

⁽¹⁰⁾ V. Jazwinski, 1970, p.231 e Anderson e Moore, 1979, p.139.

Demonstração: A primeira igualdade de (8.18) é imediata; a segunda é verificada directamente fazendo o produto

$$[I + ACB^{-1}C^T][I - AC(C^T AC + B)^{-1}C^T] = I$$

As equações (8.19) resultam da anterior pre-multiplicando-as por CB^{-1} e aplicando a identidade $C^T ACB^{-1} = (C^T AC + B)B^{-1} - I$. Veja-se em Jazwinski, 1970, pp.261-2, alguns resultados complementares.

A aplicação do lema da inversão matricial permite encontrar forma alternativa para as equações do filtro de Kalman. As novas fórmulas recorrentes serão deduzidas no quadro do problema 6.3, não tendo dificuldade a sua extensão aos outros casos. Apresentar-se-ão tanto as equações de filtragem como as de previsão em um passo⁽¹¹⁾.

A aplicação de (8.18) à identidade

$$\Sigma_{t|t} = \Sigma_{t|t-1} - \Sigma_{t|t-1}H^T(H\Sigma_{t|t-1}H^T + R)^{-1}H\Sigma_{t|t-1}$$

encontra imediatamente

$$\Sigma_{t|t}^{-1} = \Sigma_{t|t-1}^{-1} + H^T R^{-1} H$$

Para a dedução seguinte convém definir

$$A_t := (F^{-1})\Sigma_{t|t}^{-1}F^{-1}$$

$$B_t := AG(G^T A_t G + Q^{-1})^{-1}$$

Aplicando de novo (8.18) à identidade

$$\Sigma_{t+1|t} = F\Sigma_{t|t}F^T + GQG^T$$

encontra-se

(11) Cf. Anderson e Moore, 1979, pp.139-40.

$$\begin{aligned}\Sigma_{t+1|t}^{-1} &= (A_t^{-1} + GQG^T)^{-1} \\ &= (I - A_tG[G^T A_tG + Q^{-1}]^{-1}G^T)A_t \\ &= (I - B_tG^T)A_t\end{aligned}$$

Utilizando (8.19) pode expressar-se o ganho de Kalman em termos das matrizes informativas

$$\begin{aligned}K_t &= F\Sigma_{t|t-1}H^T(H\Sigma_{t|t-1}H^T + R)^{-1} \\ &= F\Sigma_{t|t}^{-1}H^T R^{-1}\end{aligned}$$

Para encontrar as estimativas do estado é útil definir os vectores

$$\begin{aligned}\hat{\mathbf{a}}_{t|t-1} &:= \Sigma_{t|t-1}^{-1}\hat{\mathbf{x}}_{t|t-1} \\ \hat{\mathbf{a}}_{t|t} &:= \Sigma_{t|t}^{-1}\hat{\mathbf{x}}_{t|t}\end{aligned}$$

reencontrando-se as variáveis de estado pela solução numérica das equações, sem necessidade de proceder às inversões matriciais. A aplicação das fórmulas recorrentes para as covariâncias permite agora escrever

$$\begin{aligned}\hat{\mathbf{a}}_{t+1|t} &= (I - B_tG^T)A_tF\hat{\mathbf{x}}_{t|t} \\ &= (I - B_tG^T)(F^{-1})^T\Sigma_{t|t}^{-1}\hat{\mathbf{x}}_{t|t} \\ &= (I - B_tG^T)(F^{-1})^T\hat{\mathbf{a}}_{t|t}\end{aligned}$$

A equação de medida-actualização pode também ser transformada

$$\begin{aligned}\hat{\mathbf{x}}_{t|t} &= \hat{\mathbf{x}}_{t|t-1} + \Sigma_{t|t-1}H^T(H\Sigma_{t|t-1}H^T + R)^{-1}(\mathbf{y}_t - H\hat{\mathbf{x}}_{t|t-1}) \\ \hat{\mathbf{a}}_{t|t} &= \hat{\mathbf{a}}_{t|t-1} + H^T R^{-1}\mathbf{y}_t\end{aligned}$$

Sintetize-se a forma alternativa do filtro

Teorema 8.3 (Filtro informativo)

Nas condições do problema 6.3 as equações recursivas do filtro de Kalman, tanto para a previsão em um passo como para a filtragem, podem ser expressas em termos das matrizes informativas Σ^{-1} resultando nas seguintes equações

$$\hat{\mathbf{a}}_{t|t} = \hat{\mathbf{a}}_{t|t-1} + H^T R^{-1} \mathbf{y}_t \quad (8.20)$$

$$\hat{\mathbf{a}}_{t+1|t} = (I - B_t G^T) (F^{-1})^T \hat{\mathbf{a}}_{t|t} \quad (8.21)$$

$$\hat{\mathbf{a}}_{0|-1} = \Sigma_{0|-1}^{-1} \hat{\mathbf{x}}_{0|-1} \quad (8.22)$$

$$\Sigma_{t|t}^{-1} = \Sigma_{t|t-1}^{-1} + H^T R^{-1} H \quad (8.23)$$

$$\Sigma_{t+1|t}^{-1} = (I - B_t G^T) A_t \quad (8.24)$$

$$\Sigma_{0|-1}^{-1} = P_0^{-1} \quad (8.25)$$

O ganho de Kalman será fornecido por

$$K_t = F \Sigma_{t|t}^{-1} H^T R^{-1} \quad (8.26)$$

e as matrizes e vectores $A, B, \hat{\mathbf{a}}$ são dados pelas seguintes definições

$$A_t := (F^{-1})^T \Sigma_{t|t}^{-1} F^{-1} \quad (8.27)$$

$$B_t := A G (G^T A_t G + Q^{-1})^{-1} \quad (8.28)$$

$$\hat{\mathbf{a}}_{t|t} := \Sigma_{t|t}^{-1} \hat{\mathbf{x}}_{t|t} \quad (8.29)$$

$$\hat{\mathbf{a}}_{t|t-1} := \Sigma_{t|t-1}^{-1} \hat{\mathbf{x}}_{t|t-1} \quad (8.30)$$

Demonstração: Segue das considerações precedentes. Forma alternativa, expressa directamente em termos das variáveis de estado, encontra-se em Brown, 1983, pp.217-9.

9. Espaço de estados, problemas de identificação e estimação

Na secção anterior abordou-se a relação entre a recursividade do algoritmo de Kalman e o carácter dinâmico dos sistemas estocásticos. Que o filtro representava mais do que um simples algoritmo e pressupunha uma dada representação sistémica dos processos aleatórios foi imediatamente compreendido por Rudolf Kalman (Kalman, 1960; Kalman *et al.*, 1969). Posteriormente, em meados dos anos 70, os trabalhos de Akaike (1974 a,b) forneceram a base para a representação e estudo dos processos estocásticos como sistema dinâmico finito, através da chamada representação markoviana em espaço de estados. É a partir desses trabalhos que o modelo de Kalman $\Sigma := (F, G, H)$ encontra uma aplicação generalizada no estudo e previsão de sucessões cronológicas e, mais geralmente, na teoria das "Time Series".

No modelo de Kalman o estado — representado pelo vector aleatório x_t — tem o sentido físico de especificar completamente a informação transmitida do "passado" para o "futuro" (Priestley, 1984, p.172), de tal forma que o comportamento posterior do sistema seja completamente determinado pelo seu estado actual — x_t — e pelo *input* futuro — w_t, v_t, u_t . Se se omitir o *input* determinístico u_t e a evolução conhecida de F, G e H , cuja a consideração não é importante para o presente estudo, a simples inspecção do modelo revela esse carácter markoviano do sistema:

$$\begin{aligned}x_{t+1} &= Fx_t + Gw_t \\y_t &= Hx_t + v_t\end{aligned}\tag{9.1}$$

Como $Ew_t w_{t-s}^T = Ev_t v_{t-s}^T = Ev_t w_{t-s}^T = 0$, a informação obtida para a variável de estado x_t , através das observações, condensa toda a informação relevante para a previsão. A estimativa \hat{x}_t descreve completamente o conhecimento do estado do sistema.

O conceito de estado, associado a um vector que se pretende de dimensão mínima, parece corresponder ao conceito epistemológico de *simplicidade*. Vale a pena reler Karl Popper: simplicidade e representação do sistema por variáveis de estado aparecem identificados com parcimónia de parâmetros do modelo e com elevada exposição ao teste prático (1). É pois fundamental definir rigorosamente o conceito de estado (2).

- (1) "Falsifiability", no dizer popperiano. Esta representação aparece também como contendo uma elevada improbabilidade *a priori*, o que significa, após os testes da prática e da crítica, uma robusta aderência à realidade. V. Popper, 1980, pp.136-45.
- (2) Seguir-se-á de perto Akaike (1974,b).

Considerem-se dois processos aleatórios de segunda ordem $\{y_t\}$ e $\{z_t\}$, mutuamente estacionários em covariância e centrados: $Ey_t = 0$, $Ez_t = 0$. Denotar-se-á por $R(l)$ a matriz de covariância $Ey_{t+l}z_t^T$, e por Z_t o fecho, em média quadrática, da variedade gerada por $Z_t = \{z_t, z_{t-1}, \dots\}$. Os preditores $\hat{y}_{\tau|t}$, $\tau > t$, em notação agora óbvia, são as projecções $\hat{E}[y_{\tau}|Z_t]$. As expressões alternativas da equação de Wiener-Hopf

$$\begin{aligned} Ey_{\tau}z_k^T &= E\hat{y}_{\tau|t}z_k^T \\ E(y_{\tau} - \hat{y}_{\tau|t})z_k^T &= E\tilde{y}_{\tau|t}z_k^T = 0 \quad k = t, t-1, \dots \end{aligned}$$

mostram que toda a informação dada por Z_t sobre y_{τ} está contida na projecção $\hat{y}_{\tau|t}$ — a inovação é ortogonal a Z_t . A equação de Wiener-Hopf, releia-se (4.13'), mostra ainda que

$$E\tilde{y}_{t+l|t}z_{t-m}^T = R(l+m) \quad l, m = 0, 1, \dots \quad (9.3)$$

O espaço preditor, \hat{Y}_t , será o fecho da variedade gerada pelas projecções de y_{t+i} em Z_t . É pois um subespaço de Z_t , formado pelas combinações lineares dos preditores $\hat{y}_{\tau|t}$, $\tau > t$. Admitir-se-á sempre que \hat{Y}_t tem *dimensão finita*.

Denote-se por $\{x_{1t}, x_{2t}, \dots, x_{pt}\}$ uma base do espaço \hat{Y}_t , base não necessariamente orthonormal. Pode representar-se $\hat{y}_{t|t}$ por meio de uma transformação linear H^* de $x_t := [x_{1t} x_{2t} \dots x_{pt}]^T$, $\hat{y}_{t|t} = H^* x_t$. Pelo que, introduzindo a inovação $v_t := \tilde{y}_{t|t}$, se encontra

$$y_t = H^* x_t + v_t \quad (9.4)$$

sendo por isso $v_t \perp Z_t$, constituindo um "ruído branco".

No momento $t+1$ ter-se-á a variedade Z_{t+1} e, em geral $Z_{t+1} \neq Z_t$, o que altera os preditores e permite construir a nova base para \hat{Y}_{t+1} . O vector formado pelos elementos dessa nova base será x_{t+1} . Denote-se por $w_t := \tilde{z}_{t+1|t}$ a inovação do processo $\{z_t\}$. Ter-se-á

$$w_t = z_{t+1} - \hat{z}_{t+1|t} \quad (9.5)$$

Sendo pois $\{w_t\}$ um processo puramente aleatório (V. Teorema 4.2).

Cada elemento de \hat{Y}_{t+1} , em particular x_{t+1} , pode ser obtido como a soma de duas combinações lineares — uma do vector da antiga base e outra da inovação w_t que lhe é ortogonal:

$$x_{t+1} = F^*x_t + G^*w_t \quad (9.6)$$

Chegou-se assim a uma representação markoviana do processo $\{x_t\}$, mostrando que *todo o processo aleatório estacionário admite uma representação pelo sistema* $\Sigma := (F, G, H)$ — *na condição de o espaço de preditores \hat{Y} ter dimensão finita.* É esse carácter dimensional de \hat{Y} que Kalman (1982) refere como caracterizando a finitude dos *parâmetros intrínsecos* ou coordenadas definidoras do sistema (V. secção 6).

Tendo o sistema $\Sigma^* := (F^*, G^*, H^*)$ sido determinado por este processo, fica garantida a unicidade de F^* , uma vez que a base x_t tem dimensão mínima e que w_t lhe é ortogonal. Se a covariância $Ew_t w_t^T$ for matriz não singular G^* é também única.

Naturalmente, outras representações equivalentes seriam possíveis. Bastaria, por exemplo, que x_t não constituísse uma base mas apenas um conjunto de geradores de variedade \hat{Y}_t , para que outros sistemas fossem determinados, naturalmente com matrizes de maior dimensão. Assim, tendo em vista o processo subjacente, definido pelo *input* z_t ou, indiferentemente (Teorema 4.2 -i), pelas suas inovações w_t , e pelo *output* y_t , a representação desse processo pelo sistema (F^*, G^*, H^*) é chamada REPRESENTAÇÃO MARKOVIANA MINIMAL.

Pode agora precisar-se o sentido da expressão “estado do sistema”, seja este minimal ou não. Em cada momento t o ESTADO será o vector x_t e o ESPACO DE ESTADOS, será a variedade gerada pelos componentes de x_t .

As expressões aparecem plenas de significado.

O estado representa o conjunto de preditores $\{\hat{y}_t|t, \hat{y}_{t+1}|t, \dots\}$, concentrando assim a *totalidade da informação sobre o output futuro que está linearmente contida nos inputs passados.* Se a realização for minimal o estado tem a menor dimensão, a que corresponde a simplicidade no sentido popperiano.

A aplicação dos resultados precedentes permite lançar nova luz sobre os processos ARMA. Recorde-se que se partiu de um processo com *input* z_t e *output* y_t e que se obteve uma representação pelo sistema $\Sigma := (F, G, H)$, doravante designada por *representação em espaço de estados.* Começará por demonstrar-se que o mesmo é possível para qualquer processo ARMA multivariável (Akaike, 1974 a).

O processo ARMA $\{y_t\}$ será definido por

$$y_t + A_1 y_{t-1} + \dots + A_p y_{t-p} = w_t + B_1 w_{t-1} + \dots + B_q w_{t-q} \quad (9.7)$$

com os pressupostos $Ew_t = 0$, $Ew_t w_t^T = \delta_{st} G$, $Ey_t w_{t+i}^T = 0$ para $i > 0$, sendo A_i e B_i matrizes de coeficientes. Assumir-se-á que as equações características $|\lambda^p I + \sum \lambda^{p-i} A_i| = 0$ e $|\lambda^q I + \sum \lambda^{q-i} B_i| = 0$ têm zeros fora do círculo unitário. É então possível exprimir y_t como processo linear geral ⁽³⁾:

$$y_t = \sum_{i=0}^{\infty} W_i w_{t-i}, \quad W_0 := I \quad (9.8)$$

Os preditores $\hat{y}_{\tau|t} = \hat{E}[y_{\tau}|y_t]$ terão de satisfazer a relação

$$\begin{aligned} \hat{y}_{\tau|t} + A_1 \hat{y}_{\tau-1|t} + \dots + A_p \hat{y}_{\tau-p|t} = \\ = \hat{w}_{\tau|t} + B_1 \hat{w}_{\tau-1|t} + \dots + B_{q-1} \hat{w}_{\tau-q|t}, \end{aligned} \quad (9.9)$$

obtida pelo operador $\hat{E}[-|-]$ nos dois membros de (9.7). Note-se que $\hat{y}_{\tau|t} = y_{\tau}$ para $\tau \leq t$ e $\hat{w}_{\tau|t} = 0$ para $\tau \geq t$, pelo que o segundo membro será identicamente nulo quando $\tau - q > t$. Claramente, a inovação é $y_t - \hat{y}_{t|t-1} = \hat{y}_t = w_t$.

Pela observação de (9.9) verifica-se que $\hat{y}_{\tau|t}$, com $\tau > t$, pode ser expresso como vector da variedade gerada pelos preditores $\hat{Y}_t^M := \{\hat{y}_{t|t}, \hat{y}_{t+1|t}, \dots, \hat{y}_{t+M-1|t}\}$ em que $M = \max\{p, q+1\}$.

Definindo $A_i := 0$ para $i = p+1, p+2, \dots, M$, encontra-se

$$\hat{y}_{t+M|t} = -A_1 \hat{y}_{t+M-1|t} - \dots - A_M \hat{y}_{t|t} \quad (9.10)$$

que é apenas uma ilustração particular do facto de os preditores $\hat{y}_{\tau|t}$ pertencerem à variedade gerada por \hat{Y}_t^M — note-se o paralelo com a representação de Ansley (1979, p.61) em (5.12).

Da representação como processo linear geral pode agora encontrar-se, operando de novo o projector,

(3) V. Nota Anexa 10.



$$\hat{y}_{\tau|t+1} = \hat{y}_{\tau|t} + W_{\tau-t-1} w_{t+1}$$

expressão que pode ser reescrita na forma

$$\hat{y}_{t+1+i|t+1} = \hat{y}_{t+1+i|t} + W_i w_{t+1} \quad (9.11)$$

Defina-se o vector $x_t = [\hat{y}_{i|t}^T \hat{y}_{i+1|t}^T \cdots \hat{y}_{t+M-1|t}^T]^T$. É óbvio, por (9.10) e (9.11), que ele admite a representação

$$x_{t+1} = \begin{bmatrix} 0 & I & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \cdots & & & \\ 0 & 0 & \cdots & I \\ -A_M & -A_{M-1} & \cdots & -A_1 \end{bmatrix} x_t + \begin{bmatrix} W_0 \\ W_1 \\ \cdots \\ W_{M-2} \\ W_{M-1} \end{bmatrix} w_{t+1} \quad (9.12)$$

$$y_t = [I \ 0 \ \cdots \ 0] x_t$$

Ora isto mostra que qualquer processo ARMA pode ser representado na forma de um sistema de Kalman, com a particularidade, removível, de as observações serem feitas sem erro — $p[v_t = 0] = 1$. Assim, qualquer processo ARMA tem uma representação markoviana em espaço de estados.

Inversamente suponha-se que o processo $\{y_t\}$ tem uma representação pelo sistema $\Sigma := (F, G, H)$, igualmente sem ruído observacional:

$$\begin{aligned} x_{t+1} &= Fx_t + Gw_t \\ y_t &= Hx_t \end{aligned}$$

Exprimindo o polinómio característico de F , $|\lambda I - A| = \lambda^p + \sum_1^p a_m \lambda^{p-m}$, pelo teorema de Cayley-Hamilton $F^p + \sum_1^p a_m F^{p-m} = 0$.

Como

$$y_{t+i} = F^i y_t + F^{i-1} G w_{t+1} + \cdots + F^0 G w_{t+i},$$

basta definir $B_i := H(F^i + a_1 F^{i-1} + \cdots + a_i F^0)G$ para obter imediatamente a representação ARMA

$$y_{t+p} + a_1 y_{t+p-1} + \dots + a_p y_t = w_{t+p} + B_1 w_{t+p-1} + \dots + B_{p-1} w_{t+1} \quad (9.13)$$

Estes resultados mostram que *qualquer processo estocástico estacionário, representado em espaços de estados com matrizes F, G, H , é equivalente a um processo ARMA, tal como qualquer processo ARMA é representável em espaço de estados.*

Dadas as vantagens computacionais do Filtro de Kalman é de indagar se não existirá qualquer processo prático de gerar uma representação sistêmica de um ARMA, transformando assim um processo de relações complexas e envolvendo variáveis em p, q distintos momentos, num outro processo, com características markovianas, em que as relações entre as variáveis apenas abrangem dois períodos temporais.

A resposta a este problema foi igualmente fornecida por Akaike (V. Akaike, 1974, p.367), mas existem sempre modelizações alternativas e, na maioria dos casos, não é difícil encontrar uma formulação apropriada.

Exemplo 9.1 (AR (2))

Seja dado o processo

$$x_t + a_1 x_{t-1} + a_2 x_{t-2} = \varepsilon_t$$

Defina-se $x_t^{(2)} := x_t$, $x_t^{(1)} := -a_2 x_{t-1}$. Pode então escrever-se

$$X_t = \begin{bmatrix} x_t^{(1)} \\ x_t^{(2)} \end{bmatrix} = \begin{bmatrix} 0 & -a_2 \\ 1 & -a_1 \end{bmatrix} \begin{bmatrix} x_{t-1}^{(1)} \\ x_{t-1}^{(2)} \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \end{bmatrix} \varepsilon_t$$

o que corresponde à equação de transição $X_t = F X_{t-1} + W_t$. Para reencontrar x_t pode fazer-se $x_t = [0 \ 1] X_t$. A equação AR(2) e a de transição são absolutamente equivalentes, mas enquanto a primeira envolve uma dependência em dois passos a segunda representa um processo de Markov.

(Exemplo retirado de Priestley, 1981, p.797).

Exemplo 9.2 (Modelo de Theil e Wage)

Considere-se o modelo (8.11)

$$x_t = \mu_t + \varepsilon_t$$

$$\mu_t = \mu_{t-1} + \beta_t + \eta_t$$

$$\beta_t = \beta_{t-1} + \zeta_t$$

adaptado à descrição do fenômenos aleatórios que seguem um crescimento aproximadamente linear ("local linear trend", Gilchrist, 1976, p.64) mas em que tanto a tendência (β_t) como o nível (μ_t) evoluem

segundo "passeio aleatório" contínuo. Considerando os ruídos $\varepsilon_t, \eta_t, \zeta_t$ como processos puramente aleatórios não mutuamente correlacionados, o modelo fornece a previsão $\hat{x}_{t+1} = \hat{\mu}_t + \hat{\beta}_t$ em que $\hat{\beta}_t$ e $\hat{\mu}_t$ são parâmetros a estimar e x_t pode considerar-se como o processo observado. Aparece assim a formulação segundo o sistema estocástico.

$$\begin{bmatrix} \mu_{t+1} \\ \beta_{t+2} \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \mu_t \\ \beta_{t+1} \end{bmatrix} + \begin{bmatrix} \eta_{t+1} \\ \zeta_{t+2} \end{bmatrix}$$

$$x_t = [1 \ 0] \begin{bmatrix} \mu_t \\ \beta_{t+1} \end{bmatrix} + \varepsilon_t$$

sendo claro que o estado é dado por $x_t = [\mu_t \ \beta_{t+1}]^T$.

Mas este modelo tem também uma formulação como ARIMA (0,2,2). Expressindo as primeiras diferenças na forma

$$\begin{aligned} \Delta x_t &= \beta_{t-1} + \zeta_t + \eta_t + \varepsilon_t - \varepsilon_{t-1} \\ \Delta x_{t-1} &= \beta_{t-1} + \eta_{t-1} + \varepsilon_{t-1} - \varepsilon_{t-2} \end{aligned}$$

encontra-se

$$\Delta^2 x_t = (\varepsilon_t + \eta_t) + (-2\varepsilon_{t-1} - \eta_{t-1}) + (-\varepsilon_{t-2}) + \zeta_t$$

e pode modelizar-se o processo, por exemplo pelo sistema

$$\begin{bmatrix} \varepsilon_{t+1} \\ \eta_{t+1} \\ \varepsilon_t \\ \eta_t \\ \varepsilon_{t-1} \\ \eta_{t-1} \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} \varepsilon_t \\ \eta_t \\ \varepsilon_{t-1} \\ \eta_{t-1} \\ \varepsilon_{t-2} \\ \eta_{t-2} \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \varepsilon_{t+1} \\ \eta_{t+1} \end{bmatrix}$$

$$\Delta^2 x_t = [1 \ 1 \ -2 \ -1 \ -1 \ 0] \begin{bmatrix} \varepsilon_t \\ \eta_t \\ \varepsilon_{t-1} \\ \eta_{t-1} \\ \varepsilon_{t-2} \\ \eta_{t-2} \end{bmatrix} + \zeta_t$$

A inovação do estado é dada por $w_t = [\varepsilon_{t+1} \ \eta_{t+1}]^T$ e o "ruído" do processo observado, $\Delta^2 x_t$, é ζ_t .

Observe-se como a primeira modelização é mais clara — por fornecer um sistema em termos dos coeficientes de estimação relevante — e simultaneamente mais simples — por necessitar de um menor número de parâmetros. O contraste será reanalisado na secção seguinte, pois está na base de críticas que Harvey, Durbin e outros têm conduzido à metodologia de Box-Jenkins.

Abordar-se-ão ainda dois problemas relacionados com a modelização em espaço de estados: a identificação ou escolha do modelo e a estimação de parâmetros.

Na fase da identificação ou escolha do modelo — os econométristas preferem esta última expressão por evitar confusões com o homónimo problema em equações simultâneas — ter-se-á presente que diversas alternativas são possíveis, tal como o exemplo precedente o mostra. Este problema está relacionado com o da estimação, pois muitas vezes as matrizes F , G ou H possuem parâmetros desconhecidos e será preferível considerá-los como “parâmetros” de estado.

Pode por vezes pensar-se que o Filtro de Kalman e a teoria do espaço de estados revela neste aspecto uma deficiência não encontrada nos métodos de regressão clássica. Uma simples análise revela como essa suposição é falsa, pois qualquer equação econométrica pode ser modelizada em espaços de estados, introduzindo como variável desconhecida, a estimar, precisamente, os parâmetros dessa relação econométrica (V. exemplo 1.3).

O que acontece é que a abordagem de qualquer processo aleatório “real” pressupõe sempre algumas definições apriorísticas na sua modelização — e esse facto é característico da própria actividade científica. Como o reconhece Kalman (1960, p.40), “We shall find it convenient to start with the model [...] and regard the problem of obtaining the model itself as a separate question. To be sure, the two problems *should* be optimized jointly if possible, the author is not aware, however of any study of the *joint* optimization problem”.

Passados 27 anos, alguns progressos foram feitos no sentido de seleccionar um modelo, mas sempre dentro de uma dada classe de modelos admissíveis ⁽⁴⁾, pois sem essa restrição o problema não tem significado.

Que a selecção de modelos e a estimação de parâmetros são problemas relacionados, é o que se mostra no exemplo seguinte.

(4) Em econometria ver, por exemplo, Harvey, 1981 b, secção 5. No quadro da estimação por Filtros de Kalman ver Jazwinski, 1970, secções 8.4 e 8.9 e Anderson e Moore, 1979, capítulo 10.

O problema da relação entre a qualidade do modelo, mais do que a qualidade do algoritmo, e a precisão preditiva é bem expresso por Witten e Cleary que mostram, com exemplos bem sugestivos, que o problema fundamental está “on your cleverness and diligence in formulating and expressing the model” (1986, p.18).

Exemplo 9.3 (Identificação de coeficiente de ARMA)

No exemplo 9.2 o processo puramente aleatório (vectorial) componente de um ARMA (0,2) foi tratado como o estado do sistema e os coeficientes foram fixados e introduzidos na matriz H . Na formulação de Akaike os coeficientes autoregressivos estavam introduzidos na matriz F e os de médias móveis determinavam a matriz G . Tratando esses coeficientes como variáveis aleatórias de estado, o Filtro de Kalman fornece um método de estimação dos seus valores. Escreva-se o processo ARMA, com os coeficientes autoregressivos $a^i := a_i$ e coeficientes de médias móveis $a^{p+i} := b_i$, na forma

$$y_t + a^1 y_{t-1} + \dots + a^p y_{t-p} = \varepsilon_t + a^{p+1} \varepsilon_{t-1} + \dots + a^{p+q} \varepsilon_{t-q}$$

Considere-se que os coeficientes seguem o processo aleatório

$$a_{t+1}^j = a_t^j + w_t^j$$

Será necessário assumir valores para a variância de w_t^j , que serão ínfimos caso se considere que os coeficientes são de facto constantes; assim como é necessário assumir uma estimativa a priori para $E\varepsilon_t^2$, estimativa que possibilitará uma mais rápida convergência se corresponder à realidade do processo modelizado. As estimativas para a média de a^j também deverão revelar o conhecimento prévio do processo.

Definir-se-á

$$x_t = [a_t^1 \quad a_t^2 \quad \dots \quad a_t^{p+q}]^T$$

$$w_t = [w_t^1 \quad w_t^2 \quad \dots \quad w_t^{p+q}]^T$$

$$H_t = [-y_{t-1} \quad -y_{t-2} \quad \dots \quad -y_{t-p} \quad \varepsilon_{t-1} \quad \varepsilon_{t-2} \quad \dots \quad \varepsilon_{t-q}]$$

$$v_t = \varepsilon_t$$

e ter-se-á

$$x_{t+1} = x_t + w_t$$

$$y_t = H_t x_t + v_t$$

Em cada momento admitem-se conhecidos os componentes de H_t , sendo por exemplo ε_t estimado como resíduo da equação de medida. Uma via consiste em introduzir o coeficiente (unitário) de ε_t em x_t e considerar v_t como erro de medida ⁽⁵⁾.

(Exemplo parcialmente baseado em Anderson e Moore, 1979, pp.50-1 e Otter e Tempelaar, 1980, pp.204-5)

⁽⁵⁾ Entre outros Akaike (1974, pp. 370-2) e Schweppe (1965, pp.63-4), seguido este de numerosos trabalhos sobre a estimação dos coeficientes por máxima verosimilhança, fornecem algoritmos alternativos a este método (V. Anderson e Moore, 1979, pp.51-2).

Em situações práticas o mais comum é admitir-se incerteza apenas sobre um bloco de parâmetros e considerar o outro bloco como perfeitamente determinado. Um algoritmo apropriado para enfrentar essas situações foi desenvolvido por Schmid (V. Jazwinski, 1970, p.285), com base no princípio de que a incerteza nos parâmetros produz uma degradação das estimativas. Para enfrentar o problema muitas e diversificadas metodologias são possíveis mas, se se admitir o princípio básico de considerar como variáveis de estado os parâmetros incertos, a solução teórica aparece como óbvia ⁽⁶⁾. Trata-se afinal de definir um *sistema aumentado*

$$\begin{aligned}x_{t+1} &= F_t x_t + \Delta_t u + G_t w_t \\ y_t &= H_t x_t + \Psi_t p + v_t\end{aligned}\tag{9.14}$$

Em que os parâmetros desconhecidos da equação de transição — u — são designados como *parâmetros dinâmicos* e os da equação de medida — p — são os *parâmetros de medida*.

Assume-se que u , p , x_0 , $\{w_t\}$ e $\{v_t\}$ não estão correlacionados e, como usual, será necessário introduzir estimativas *a priori* para os momentos de u e de p .

Considerar-se-á que a transição dos parâmetros desconhecidos é feita sem ruído

$$u_{t+1} = u_t$$

$$p_{t+1} = p_t$$

Definir-se-á o *vector de estado aumentado*

$$x_t^A := \begin{bmatrix} x_t \\ u_t \\ p_t \end{bmatrix}\tag{9.15}$$

e a nova matriz de medida

$$H_t^\Psi := [H_t \quad 0 \quad \Psi_t]\tag{9.16}$$

⁽⁶⁾ Cf. Otter, 1981, que desenvolve ainda um teste para identificação dos parâmetros, e Hawkes e Moore, 1976, que colocam o problema numa perspectiva bayesiana e derivam condições de convergência das covariâncias dos erros para o seu limite inferior. O início de uma abordagem global do problema pode ver-se em Mehra (1971), que considera o caso limite do desconhecimento simultâneo dos parâmetros de F , G e H e das matrizes de covariância.

Ter-se-á o sistema aumentado

$$\mathbf{x}_{t+1}^A = \begin{bmatrix} F_t & \Lambda_t & 0 \\ 0 & I & 0 \\ 0 & 0 & I \end{bmatrix} \mathbf{x}_t^A + \begin{bmatrix} G_t \\ 0 \\ 0 \end{bmatrix} \mathbf{w}_t \quad (9.14')$$

$$\mathbf{y}_t = [H_t \quad 0 \quad \Psi_t] \mathbf{x}_t^A + \mathbf{v}_t$$

Ao qual se aplicará o filtro de Kalman produzindo as estimativas

$$\mathbf{x}_{\tau|t}^A = \begin{bmatrix} \hat{\mathbf{x}}_{\tau|t} \\ \hat{\mathbf{u}}_{\tau|t} \\ \hat{\mathbf{P}}_{\tau|t} \end{bmatrix}$$

Fica ainda de fora o problema de obter estimativas apriorísticas sobre um conjunto de elementos do sistema. Vale a pena referir o caso das matrizes de covariância dos ruídos pois, como refere Mehra (1972, p.693) "The basic source of uncertainty is due to unknown *a priori* statistics of noise".

Para enfrentar o problema da estimação de $R_t = E\mathbf{v}_t\mathbf{v}_t^T$ e de $Q_t = E\mathbf{w}_t\mathbf{w}_t^T$ conhecem-se, fundamentalmente, quatro tipos de métodos: bayesianos, de máxima verosimilhança, de correlação e de ajustamento de covariâncias (Mehra, 1972). Referir-se-á muito brevemente, e apenas como exemplo, o *método de análise de correlação das inovações*, seguindo de perto Mehra (1970, 1972), pois trata-se da abordagem mais directamente enquadrável na linha geral do que tem sido exposto.

Designem-se por $V_k = E\tilde{\mathbf{y}}_t\tilde{\mathbf{y}}_{t-k}^T$ a matriz de correlação das inovações $\tilde{\mathbf{y}}_{t|t-1} = \mathbf{y}_t - H\hat{\mathbf{x}}_{t|t-1} = H\tilde{\mathbf{x}}_{t|t-1} + \mathbf{v}_t$, supondo que elas são (localmente) estacionárias em covariância. Se o modelo estiver correctamente especificado a matriz V_k será nula para $|k| > 0$, pois o filtro de Kalman será um projector, gerando um processo puramente aleatório de inovações. É esse facto que será testado.

Para simplificar o problema omitam-se os índices sempre que possível e considere-se que F e H são constantes e que $G_t \equiv I$.

A matriz V_k será dada por

$$\begin{aligned} V_k &= E[(H\tilde{\mathbf{x}}_t + \mathbf{v}_t)(H\tilde{\mathbf{x}}_{t-k} + \mathbf{v}_{t-k})^T] = \\ &= H.E\tilde{\mathbf{x}}_t\tilde{\mathbf{x}}_{t-k}^T.H^T + H.E\tilde{\mathbf{x}}_t\mathbf{v}_{t-k}^T, \quad k > 0 \\ V_0 &= H\Sigma_0 H^T + R \end{aligned}$$

por aplicação da equação (7.3') e iterando a expressão k passos relaciona-se \tilde{x}_t com \tilde{x}_{t-k} e com v_{t-k} , como se mostra

$$\begin{aligned}\tilde{x}_t &= F(I - KH)\tilde{x}_{t-1} - FKv_{t-1} + w_{t-1} \\ &= [F(I - KH)]^k \tilde{x}_{t-k} + \sum_{j=1}^k [F(I - KH)]^{j-1} [-FKv_{t-j} + w_{t-j}]\end{aligned}\quad (9.18)$$

De onde resulta

$$E\tilde{x}_t \tilde{x}_{t-k}^T = [F(I - KH)]^k \Sigma \quad (9.19)$$

$$E\tilde{x}_t v_{t-k} = -[F(I - KH)]^{k-1} FK R \quad (9.20)$$

com $\Sigma = \Sigma_{t|t-1} = E\tilde{x}_t \tilde{x}_t^T$, que se suporá constante, admitindo que o filtro atingiu um estado limite de convergência. A expressão da matriz V_k aparece agora, para $k > 0$ e depois de algumas manipulações algébricas

$$V_k = H[F(I - KH)]^{k-1} F[\Sigma H^T - KV_0] \quad (9.21)$$

sendo relativamente fácil mostrar, por introdução da fórmula de K , que ela se anula identicamente para $k > 0$.

Se as pressupostas matrizes Q e R não corresponderem às verdadeiras matrizes de covariâncias dos ruídos, o ganho de Kalman será sub-ótimo, tal como as estimativas \hat{x}_t , e a matriz V_k não se anulará. O método proposto por Mehra para resolver o problema e fazer retornar o filtro a um estado de optimalidade pode descrever-se nos seguintes passos.

(i) Denote-se por \hat{K} e $\hat{\Sigma}$ as matrizes estimadas e que se reconhecem sub-ótimas pela análise da matriz empírica \hat{V}_k observada; calcule-se $\hat{\Sigma}H^T$ em função de V_1, V_2, \dots

(ii) Calcule-se $R = V_0 - H\hat{\Sigma}H^T$

(iii) Corrijam-se as estimativas pela forma

$$K = (\hat{\Sigma}H^T + \delta\Sigma H^T)(H\Sigma_0 H^T + R + H\delta\Sigma H^T)^{-1}$$

$$\Sigma = \hat{\Sigma} + \delta\Sigma$$

$$\text{com } \delta\Sigma = F[\delta\Sigma - K(H\hat{\Sigma} + H\delta\Sigma) + \hat{K}H\Sigma + \hat{\Sigma}H^T \hat{K}^T - \hat{K}V_0 \hat{K}^T]F^T$$

equações que se resolvem numericamente introduzindo os valores já calculados de $\hat{\Sigma}H^T$.

Esboce-se muito rapidamente a prova de (iii).

De (9.19) deriva-se

$$\hat{\Sigma} = F[\hat{\Sigma} - \hat{K}H\hat{\Sigma} - \hat{\Sigma}H^T\hat{K}^T + \hat{K}(H\hat{\Sigma}H^T + R)\hat{K}^T]F^T + Q$$

Relacionando Σ ($:= \Sigma_{t+1|t}$) com $\Sigma_{t|t}$ e relembrando que o filtro terá atingido um estado estacionário, encontra-se $\Sigma = F(\Sigma - KH\Sigma)F^T + Q$, de onde se deriva a expressão de $\delta\Sigma$ que é introduzida no ganho de Kalman, corrigindo-o.

Não vale a pena pormenorizar o estudo do problema pois reconhece-se que este algoritmo, tal como alguns outros, é sobretudo aplicável a situações de estimação próximas da estacionaridade, situações mais usualmente verificáveis em fenómenos físicos do que na previsão de realidades económicas ou sociais.

Embora sob forma diversa, agora virada para a previsão econométrica, o problema de selecção de modelos e estimação de parâmetros voltará a ser abordado nas secções seguintes, em que se procurará traçar uma panorâmica das aplicações económicas e das polémicas em curso sobre a aplicação dos filtros de Kalman.

terceira parte
problemas recentes

O desenvolvimento teórico efectuado nas secções precedentes terá conduzido a alguma familiaridade com a modelização em espaços de estados e com o algoritmo de Kalman. É agora fácil perceber as suas potencialidades na modelização econométrica e na previsão.

Redescobertas recentes, feitas em alguns trabalhos estatísticos, ficam assim enquadradas numa teoria de que estão traçados os fundamentos primários. Nesta área é preciso reconhecer que os passos maiores foram dados pelos matemáticos e engenheiros da teoria do controlo, e que o seu avanço é significativo. É precisamente a recuperação desse desfazamento que Durbin, Harvey e outros investigadores têm procurado realizar.

A secção que inicia esta parte do trabalho sumariza a introdução da teoria de Kalman no terreno econométrico e apresenta o que se considerou ser mais significativo nos desenvolvimentos recentes. Trata-se de um "survey", limitado pelo espaço e também pela preocupação de não repetir resultados já estabelecidos.

A secção seguinte — décima primeira — prossegue o "survey" numa perspectiva prática, pois apresenta estudos de aplicações, resultados comparativos e sucessos mais destacados.

10. O Filtro de Kalman como unificador de procedimentos preditivos

A redescoberta do Filtro de Kalman pela literatura econométrica é paralela à compreensão das suas vantagens na modelização e previsão.

O trabalho pioneiro parece ser o de Harrison e Stevens (1971,1976), que encararam o filtro numa perspectiva bayesiana e destacaram a flexibilidade do modelo de espaço de estados. Uma exposição parcelar desses trabalhos pode encontrar-se em Kendall e Stuart (1983, pp.528-30 e 647-9), havendo uma revisão mais actual feita por Bolstad (1986).

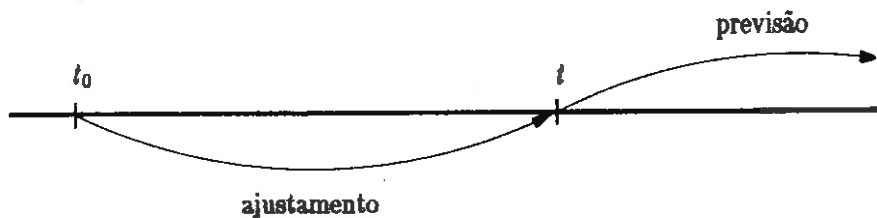
O trabalho de Harrison e Stevens, apareceu num momento de questionação dos métodos estrapolativos. Após algumas décadas de crescimento sustentado, nas quais os modelos econométricos de parâmetros fixos conheceram um grande desenvolvimento teórico e prático, surgiram diversos factores de crise — usualmente referidos aos choques petrolíferos — que levaram a mudanças insuspeitadas na evolução das variáveis económicas. O teste da constância no tempo das relações econométricas começou a ser estudado sistematicamente (Brown *et al*, 1975), e era claro que faltava uma fundamentação teórica para a construção de modelos adaptativos.

Métodos *ad hoc*, como os de alisamento exponencial, mostravam melhores resultados práticos que modelos econométricos rigorosos, e a previsão empírica subjectiva permitia ter em conta dados não abrangidos pela modelização estatística. A crítica à previsão econométrica dirigia-se a um conjunto de limitações que pareciam inultrapassáveis, ou sujeitas a correcção apenas no exterior dos modelos: a incapacidade de medir movimentos subjacentes que serão importantes no futuro, o desprezo por efeitos de substituição implícitos, e a incapacidade de conjugar o controlo dos erros com uma evolução dinâmica de parâmetros (Moyer, 1984, pp.68-70). Também a criação de pesados modelos de equações simultâneas não obtinha sucessos práticos na previsão de curto prazo. Começava a ser evidente a necessidade de numa metodologia de previsão recursiva, que incorporasse a variabilidade dos parâmetros e permitisse adaptar os modelos à evolução da conjuntura.

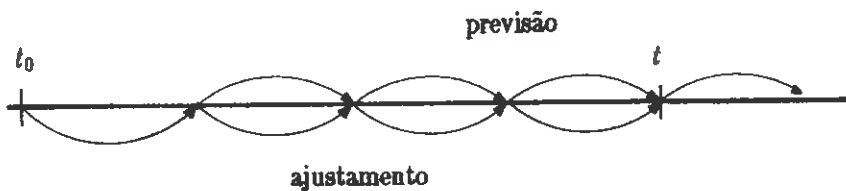
Neste contexto, Harrison e Stevens formularam uma "evidência" que durante certo tempo não foi claramente compreendida: "it is a dangerous step to extrapolate to the conclusion that because a method produces the 'best fit' on such a data series it is the best method to employ in a forecasting system" (1976, p.205).

Para verificar que tal "evidência" não era claramente aceite basta reler Kendall, que na mesma altura tinha afirmado: "The safest was to gauge the reliability of a forecasting method is to consider its performance over a period. From a suite of observed errors we can form a good estimate of the error likely to be encountered in the future." (1973, p.117).

Na realidade, a “evidência” incompreendida da posição de Harrison e Stevens só tinha sentido em face de uma modelização alternativa, que rompesse a separação rígida entre período de ajustamento e período de estimação



e introduzisse uma “dimensão extra” no problema: a recursividade



Por esse método está aberto caminho “for the estimation of time-variable parameters and states in stochastic-dynamic systems”. “Recursive analysis [...] provides a basis for sophisticated new methods” (Young, 1985, p.213).

Como a recursividade está associada ao carácter dinâmico dos sistemas (Teorema 8.1), a redescoberta do modelo de espaço de estados e do paralelo algoritmo de Kalman foi a resposta encontrada por Harrison e Stevens.

Desenvolveram o chamado MODELO LINEAR DINÂMICO, que é uma versão simplificada do sistema de Kalman. Aí as duas equações matriciais são designadas de *observação* e de *sistema*. O vector de parâmetros desconhecidos encontra-se representado por θ_t e as observações por y_t . A matriz de transição é G_t e a de medida F_t ⁽¹⁾. A distribuição é

(1) Kalman (1982, p.205), Parzen (1986, p.223) e outros, criticam justificadamente este tipo de alterações à nomenclatura usual.

suposta multinormal.

$$y_t = F_t \theta_t + v_t \quad v_t \sim N(0, V_t) \quad (10.1')$$

$$\theta_t = G_t \theta_{t-1} + w_t \quad w_t \sim N(0, W_t) \quad (10.1'')$$

Que o sistema (10.1) permite reencontrar muitos modelos de evolução com seus casos particulares, assim como existe algoritmo recursivo de estimação e previsão, são factos já amplamente ilustrados. Não vale a pena desenvolver esses temas a que foi dedicada parte substancial dos artigos de Harrison e Stevens. A justificação das potencialidades do modelo como base de um sistema preditivo terá sido a contribuição fundamental do seu trabalho.

Como primeiro ponto essencial destaca-se a possibilidade de introduzir, em cada momento, *informação adicional externa aos dados quantitativos históricos*.

Essa informação pode ser introduzida pelo menos por dois caminhos: modificando os parâmetros da matriz de transição ou considerando uma incerteza acrescida. No primeiro caso tratar-se-á de promover a alteração da matriz G_t , tal como foi mostrado no exemplo 6.4. O algoritmo nem por isso despreza a informação passada, continuando a considerá-la na estimativa do estado. No segundo caso trata-se de admitir matrizes de variâncias W_t com valores mais elevados, o que descontará o peso de uma observação "errática" ("outlier") ou de uma maior incerteza sobre período determinado de evolução. Como o ganho de Kalman será menor com maior valor das matrizes de covariâncias, isso significa que a inovação registada terá menor peso na previsão do período seguinte. Retomando a simbologia usual ter-se-á

$$\hat{y}_{t+1|t} = H_{t+1} \hat{x}_{t+1|t} = H_{t+1} F_t \hat{x}_{t|t-1} + H_{t+1} K_t \tilde{y}_t \quad (10.2)$$

o que significa que os erros de previsão em períodos de maior incerteza podem ser descontados na estimação futura.

Como segundo ponto essencial do trabalho de Harrison e Stevens destaque-se a *possibilidade de desenvolvimento do processo preditivo com reduzida ou nula informação quantitativa histórica*.

De facto, a possibilidade de introduzir informação externa aos dados quantitativos pode ter de ser levada a um limite pela inexistência destes últimos. Isto corresponde à inicialização do Filtro de Kalman com a fixação do sistema $\Sigma := (F, G, H)$ e das estimativas *a priori*

$$\hat{\mathbf{x}}_{0|-1} = \bar{\mathbf{x}}_0, \quad \Sigma_{0|-1} = P_0 \quad (10.3)$$

A primeira traduzirá o mais credível valor inicial dos parâmetros e a segunda a incerteza subjectiva que lhe é associada.

Mas julga-se que o contributo essencial do trabalho de Harrison e Stevens é o *tratamento da incerteza na modelização*. Para o efeito desenvolveram duas classes de modelos multiprocesso que se apresentarão de seguida.

Fala-se dos *modelos multiprocesso de classe I* (Harrison e Stevens, 1976, pp.220-4) quando se considera que existe um conjunto de esquemas interpretativos da realidade, sob os quais há incerteza, e quando se pretende efectuar previsões com base nesses esquemas.

Denote-se por $M^j := (F^j, G^j, M^j, Q^j, R^j)$ especificação completa relativa a dado modelo admissível. Se o decisor tiver, em cada momento, um conjunto definido $\mathcal{M} = \{M^j\}$ de modelos alternativos e se inicializar o processo com uma dada credibilidade sobre cada modelo — P_0^j — ser-lhe-á possível actualizar essa credibilidade com base nos resultados obtidos. Esquemáticamente, num dado momento t , a probabilidade de ser válido dado modelo M^j é proporcional à verosimilhança e à probabilidade que lhe foi assinalada no momento anterior

$$P_t^j = p(M^j | Y_{t-1}, \mathbf{y}_t) \propto L(\mathbf{y}_t | M^j, Y_{t-1}) \cdot p(M^j | Y_{t-1}) \quad (10.4)$$

Como o sistema foi especificado como gaussiano é fácil determinar por completo a distribuição condicional

$$p(\hat{\mathbf{y}}_{t+1|t} | Y_t, M_j), \quad (10.5)$$

assim como a credibilidade associada a cada previsão $\hat{\mathbf{y}}^j$, dada por cada modelo.

Numa perspectiva decisional Harrison e Stevens consideram a utilidade de cada previsão — ou o simétrico da sua perda — que será a função $u(\cdot)$ do erro registado. Várias abordagens são possíveis.

(i) *Fixando um modelo único* a previsão deixará de ser o valor $\hat{y}_{t+i|t}$ dado pelo filtro de Kalman, preferindo-se antes aquele valor particular que maximiza a utilidade. Como a distribuição de

$$u(\tilde{y}_{t+i|t}) \quad (10.6)$$

é determinável (pela *assumpção gaussiana* do sistema) é possível calcular, para o modelo j , a nova previsão

$$y_{t+i|t}^{*j} : \max u(\tilde{y}_{t+i|t}^j) \quad (10.7)$$

(ii) *Pretendendo-se escolher o modelo j* com base num conjunto de períodos de previsão seleccionar-se-á aquele que maximiza as utilidades esperadas

$$M^j : \max \sum_{i=1}^L E\{u(\tilde{y}_{t+i|t}^{*j})\} \quad (10.8)$$

(iii) Admitindo-se como melhor solução uma *combinação de previsões* escolher-se-á aquele $y_{t+i|t}^*$ que maximiza a ponderação das utilidades esperadas

$$y_{t+i|t}^* = \sum_j E\{u(\tilde{y}_{t+i|t}^{*j})\} \cdot P_t^j \quad (10.9)$$

Os ponderadores são as probabilidades assinaladas à validade de cada modelo.

Harrison e Stevens desenvolveram igualmente os *modelos multiprocesso de classe II*, em que consideram a sucessão como sendo explicada, em cada momento, por um modelo provavelmente diferente do válido no período anterior. Construindo uma matriz de probabilidades de transição entre modelos, $[\pi_{ij}]$, é possível calcular a probabilidade de ser válido de um modelo em dado momento. Essa credibilidade é actualizada pela regra de Bayes, ponderando a credibilidade que foi assinalada no momento anterior pelas probabilidades de transição (Harrison e Stevens, 1976, pp.224-7)

Por aqui se vê como a abordagem de Harrison e Stevens, inspirada no modelo do espaço de estados e utilizando o Filtro de Kalman para actualizar a distribuição de $\hat{y}_{t+1|t}$, se começa a afastar da metodologia clássica para adoptar uma perspectiva claramente bayesiana. O filtro é utilizado apenas para a determinação dos momentos de $\hat{y}_{t+1|t}$. Segue-se-lhe uma

sistemática incorporação de informações subjectivas e os preditores são definidos em termos da teoria bayesiana da decisão.

Os trabalhos posteriores de P.J. Harrison adoptaram essa perspectiva bayesiana de forma menos tímida ⁽²⁾. Persistindo na definição *a priori* das matrizes de covariâncias e não na sua estimação, e notando que elas condicionam o peso dado às novas observações por intermédio do ganho de Kalman, Harrison atacou directamente o problema do peso relativo dos dados, definindo modelos de *Estimação por Desconto Ponderado* ("Discounted Weighted Estimation").

Nesses modelos (V. Ameen e Harrison, 1984, 1985) a chave do problema está na definição do vector β de coeficientes de desconto (alisamento) e no uso da distribuição normal para definir relações recursivas. A arbitrariedade da escolha inicial do vector β acentua o carácter bayesiano do modelo, afirmando Harrison que essa escolha é mais directa para o decisor de que a das estimativas iniciais do Filtro de Kalman.

Outro desenvolvimento recente da mesma escola de Warwick é o tratamento, por esses métodos de desconto, dos *modelos lineares gerais*, que são aqueles em que a amostragem das observações segue uma distribuição da família exponencial (V. West *et al.*, 1985). O método é estendido a modelos lineares de sistemas não gaussianos (V. West e Harrison, 1986), com uma abordagem cem por cento bayesiana que extravasa o âmbito deste trabalho.

Pela mesma altura em que Harrison e Stevens apresentaram o seu trabalho pioneiro, Ramon Mehra desenvolveu um outro procedimento preditivo que vale a pena referir muito brevemente (V. Mehra, 1979 e Priestly, 1981, pp.804-14).

Baseando-se nos resultados de Akaike (secção 9), o método resume-se a ajustar um modelo ARMA multidimensional às observações de y_t . Por intermédio de um critério apropriado de significância determina-se a ordem p do modelo autoregressivo sobre $\{y_t\}$. Pelas correlações entre $[y_t^T y_{t-1}^T \dots y_{t-p}^T]$ e o vector $[y_t^T y_{t+1}^T \dots]$ determina-se a dimensão da matriz F_t . Trata-se depois de introduzir estimativas no modelo de espaço de estados como em (9.12).

O processo parece resumir-se a tratar um modelo ARMA na forma de espaço de estados, mas tem a vantagem de apontar a forma do modelo e estimativas iniciais, a serem depois corrigidas pela rotina do filtro de Kalman e, posteriormente, aplicadas na previsão.

Desta referência breve é importante destacar que o procedimento de Akaike e Mehra apresenta a modelização e previsão como um processo de "caixa preta", aplicando apenas uma

(2) V. nota 10 na secção 8.

análise estocástica às observações .

Mais recentemente, Harvey e outros investigadores da London School of Economics, têm vindo a propor os chamados *modelos estruturais* que permitem, simultaneamente, ultrapassar as limitações do modelo de "caixa preta" e o subjectivismo do método de Harrison.

O MODELO ESTRUTURAL procura descrever a sucessão cronológica em termos dos comportamentos que são economicamente significativos: "The models are structural in the sense that the form of the individual components can be given a direct interpretation" (Harvey, 1984). A forma básica do modelo — "basic structural model" na terminologia posteriormente adoptada (Harvey e Durbin, 1986) — corresponde à decomposição tradicional

$$y_t = \text{Tendencia} + \text{Sazonalidade} + \text{Ruido} \quad (10.10)$$

Cada um desses componentes pode ser representado em espaço de estados — o que corresponde a um ARIMA em que os componentes não são directamente observados (V. secção 9, exemplo 2)

Como foi já amplamente descrito, a *tendência* T pode corresponder a um nível "localmente fixo" ⁽³⁾, sendo nesse caso possível representá-la pelo modelo

$$\begin{aligned} T_t &= \alpha_t + v_t \\ \alpha_t &= \alpha_{t-1} + w_t \end{aligned} \quad (10.11)$$

A tendência pode também corresponder a um crescimento "localmente linear", como no modelo de Theil e Wage

$$\begin{aligned} T_t &= \alpha_t + v_t \\ \alpha_t &= \alpha_{t-1} + \beta_{t-1} + w_{1t} \\ \beta_t &= \beta_{t-1} + w_{2t} \end{aligned} \quad (10.12)$$

ou quadrático, ou polinomial de qualquer ordem.

Seja k a ordem do monómio αt^k que se admite ser componente explicativo da evolução de T_t . É fácil verificar que esse tipo de crescimento é traduzido pelo modelo

⁽³⁾ A designação de "locais" para as propriedades dos modelos parece infeliz, assim como parece polémica a interpretação de Gilchrist (1976, p.19). O sentido preciso da aplicação do termo será sempre dado pelo sistema de equações descrito.

$$\begin{aligned}
T_t &= \alpha_t^{(0)} + v_t \\
\alpha_t^{(0)} &= \alpha_{t-1}^{(0)} + \alpha_{t-1}^{(1)} + w_{1t} \\
\alpha_t^{(1)} &= \alpha_{t-1}^{(1)} + \alpha_{t-1}^{(2)} + w_{2t} \\
&\dots \\
\alpha_t^{(k)} &= \alpha_{t-1}^{(k)} + w_{kt}
\end{aligned}
\tag{10.13}$$

pelo que as diferenças $\Delta^k T_t$ serão estacionárias em média, com usuais pressupostos sobre os ruídos. O modelo em espaço de estados é pois equivalente a algum ARIMA (\cdot, k, \cdot) .

Crescimentos de outro tipo são igualmente modelizáveis. Dê-se o exemplo do crescimento exponencial (geométrico), que não necessita das transformações logarítmicas ou de Box-Cox utilizadas nos ARIMA. A tendência $T_t = \alpha_0 \cdot r^t + \varepsilon_t$, com ε_t ruído de carácter à partida indeterminado, é representado pelo sistema:

$$\begin{aligned}
T_t &= \alpha_t + v_t \\
\alpha_t &= r \cdot \alpha_{t-1} + w_t
\end{aligned}
\tag{10.14}$$

Quanto à *sazonalidade* ela pode ser estudada por vários processos. O mais intuitivo talvez seja o proposto por Harrison e Stevens (1976, p.217) que consiste em construir um vector de coeficientes sazonais aditivos cobrindo um ciclo de L períodos.

$$\mathbf{a}_t = \begin{bmatrix} a_{1t} \\ a_{2t} \\ \dots \\ a_{Lt} \end{bmatrix} \quad \sum_i^L a_{it} = 0$$

Em cada período sazonal i a matriz F corresponderá a um vector de zeros, com excepção do elemento i , ocupado pela unidade. No momento sazonal i ter-se-á

$$\begin{aligned}
S_t &= [0 \quad \dots \quad 1 \quad \dots \quad 0] \begin{bmatrix} a_{1t} \\ \dots \\ a_{it} \\ \dots \\ a_{Lt} \end{bmatrix} + v_t \\
\begin{bmatrix} a_{1t} \\ \dots \\ a_{it} \\ \dots \\ a_{Lt} \end{bmatrix} &= \begin{bmatrix} a_{1,t-1} \\ \dots \\ a_{i,t-1} \\ \dots \\ a_{L,t-1} \end{bmatrix} + \begin{bmatrix} w_{1t} \\ \dots \\ w_{it} \\ \dots \\ w_{Lt} \end{bmatrix}
\end{aligned}
\tag{10.15}$$

É evidente que são possíveis formulações diversas (V. Harrison e Stevens, 1976, p.218 e Harvey, 1984, pp.253-4). Uma forma mais compacta, que evita alterar ciclicamente a matriz F e não obriga à correcção dos coeficientes pela condição $\sum_t a_{it} = 0$, consiste em considerar o efeito da sazonalidade no período t como o complemento para zero dos efeitos registados nos $L-1$ períodos anteriores⁽⁴⁾. O coeficiente a_t passará a representar o efeito sazonal no período em causa e será pois independente, em posição, do período do ciclo.

$$S_t = [1 \ 0 \ \dots \ 0] \begin{bmatrix} a_t \\ a_{t-1} \\ \dots \\ a_{t-L+2} \end{bmatrix} + v_t \quad (10.16')$$

$$\begin{bmatrix} a_t \\ a_{t-1} \\ \dots \\ a_{t-L+2} \end{bmatrix} = \begin{bmatrix} -1 & -1 & \dots & -1 & -1 \\ 1 & 0 & \dots & 0 & 0 \\ \dots & & & & \\ 0 & 0 & \dots & 1 & 0 \end{bmatrix} \begin{bmatrix} a_{t-1} \\ a_{t-2} \\ \dots \\ a_{t-L+1} \end{bmatrix} + \begin{bmatrix} w_t \\ 0 \\ \dots \\ 0 \end{bmatrix} \quad (10.16'')$$

Outros, componentes podem ser acrescentados à equação (10.10), nomeadamente variáveis exógenas ou elementos determinísticos de efeito conhecido (V. Harvey e Durbin, 1986, p.190). No primeiro caso, essas variáveis serão tratadas como coeficientes conhecidos de parte da matriz H_t correspondente — os parâmetros multiplicativos serão componentes do estado x_t (V. exemplo 1.3). No segundo caso o efeito dos elementos determinísticos u_t é conhecido — efectua-se por intermédio da matriz Γ_t .

O modelo estrutural, na forma mais completa, poderá pois ser descrito pelo sistema seguinte

$$\begin{bmatrix} T_t \\ S_t \\ b_t \end{bmatrix} = \begin{bmatrix} F_{1t} & 0 & 0 \\ 0 & F_{2t} & 0 \\ 0 & 0 & F_{3t} \end{bmatrix} \begin{bmatrix} T_{t-1} \\ S_{t-1} \\ b_{t-1} \end{bmatrix} + G_t w_t + \Gamma_t u_t \quad (10.17')$$

$$y_t = [I \ I \ b] \begin{bmatrix} T_t \\ S_t \\ z_t \end{bmatrix} + v_t \quad (10.17'')$$

Trata-se de um esquema teórico, a adaptar conforme os modelos concretos em causa. Essa adaptação poderá ser bastante simples, não implicando, normalmente, a forma diagonal por blocos de F_t .

(4) *Mais precisamente: o complemento para zero a menos de uma perturbação aleatória. V. uma discussão do problema em Watson, 1987, pp.396-7, 404.*

É agora claro que *o modelo estrutural não é mais do que uma concretização econométrica do sistema de Kalman*. O destaque que lhe é dado por Harvey e por Durbin parece derivar de três factores de inovação metodológica e conceptual no terreno da econometria.

Em primeiro lugar este modelo permite, simultaneamente, tratar as sucessões económicas temporais *como processos estocásticos* e como a conjugação de *variáveis economicamente categorizáveis* e significantes. Parece estar assim a fechar-se um ciclo, aberto pelas posições polémicas de Yule, Kuznets e outros, em que se confrontava o estudo das oscilações e ciclos económicos por causas aleatórias — Yule — e a sua análise pela investigação de causas racionais — Kuznets (V. Murteira, 1953, p.29).

Seguindo-se a um período em que a metodologia de Box e Jenkins tinha ganho grande popularidade, Harvey e Durbin encontraram no modelo de Kalman a continuidade para as críticas que anteriormente formulavam às técnicas dos modelos ARIMA: "We have become disenchanted with the notion that the appropriate way to deal with trend and seasonal components is to eliminate them by differencing [...] the statistician should seek to identify the main observable features [...] time series usually reveal trends and seasonals as important observables of the data, and it seems desirable to model these features explicitly" (Cf. Harvey e Durbin, 1986, pp.187-8; ver também a posição de Priestley, *ibidem*, p.215, e a resposta dos autores, p.225).

Em segundo lugar, o modelo estrutural tem disponível um processo estatístico de estimação e previsão que pode ser tratável pelas *técnicas clássicas*, caminhando-se para a possibilidade de estimação de todos os seus componentes (V. Ansley e Kohn, 1985, Kohn e Ansley, 1986). O recurso à perspectiva bayesiana aparecerá como dispensável.

Por outro lado, o sistema de Kalman permite tratar com simplicidade um conjunto de problemas de estimação até ao momento complexo, nomeadamente a estimação de máxima verosimilhança (V. Jones, 1985), fornecendo o algoritmo de Kalman um procedimento recursivo automático e computacionalmente implementável.

Em terceiro lugar, Harvey e Durbin, reconhecem o sistema e o algoritmo de Kalman como *unificador geral* de um conjunto de procedimentos de estimação e previsão, que aparecem como seus casos particulares. Vale apenas desenvolver brevemente esta ideia, embora com um enfoque diverso do salientado por esses autores.

Comece-se pelo clássico *modelo de regressão linear*

$$y_t = bz_t + \varepsilon_t, \quad (10.18)$$

já exibido na forma $y_t = Hx_t + v_t$ no exemplo 1.3. O estado, aqui representado por $x_t \equiv b$, é uma variável aleatória degenerada, correspondendo a equação de transição a uma identidade. O sistema completo será

$$b_t = Ib_{t-1} + 0w_t \quad (10.19')$$

$$y_t = z_t^T b_t + \varepsilon_t \quad (10.19'')$$

É importante verificar em que condições o Filtro de Kalman reencontra o método dos mínimos quadrados.

Em vez de aplicar recursivamente o filtro ao sistema (10.19) agrupem-se as observações e os ruídos e represente-se esse agrupamento com a simbologia do sistema de Kalman

$$y = \begin{bmatrix} y_1 \\ y_2 \\ \dots \\ y_n \end{bmatrix} \quad H = \begin{bmatrix} z_1^T \\ z_2^T \\ \dots \\ z_n^T \end{bmatrix} \quad v = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \dots \\ \varepsilon_n \end{bmatrix} \quad R := E v v^T$$

Com o desconhecimento *a priori* sobre $x := b^T$ têm sentido as estimativas $\hat{b}_{0|-1}^T = 0$ e $\Sigma_{0|-1}^{-1} = 0$. Aplicando o filtro informativo (Teorema 8.3) ter-se-á sucessivamente

$$\begin{aligned} \hat{a}_{0|-1} &= \Sigma_{0|-1}^{-1} \hat{b}_{0|-1}^T = 0 \\ \Sigma_{0|0}^{-1} &= \Sigma_{0|-1}^{-1} + H^T R^{-1} H = H^T R^{-1} H \\ \hat{a}_{0|0} &= \hat{a}_{0|-1} + H^T R^{-1} y = H^T R^{-1} y \\ \hat{b}_{0|0}^T &= \Sigma_{0|0}^{-1} \hat{b}_{0|0}^T = (H^T R^{-1} H) \hat{b}_{0|0}^T \end{aligned}$$

Se a matriz $(H^T R^{-1} H)$ for não singular encontra-se imediatamente

$$\hat{b}_{0|0}^T = (H^T R^{-1} H)^{-1} (H^T R^{-1} y) \quad (10.20)$$

o que corresponde precisamente ao estimador dos mínimos quadrados generalizados (Aitken).

Este facto notável mostra como o algoritmo de Kalman, construído para um sistema estocástico, é robusto à especificação de b como vector determinístico (V. Brown, 1983, p.245).

No entanto, se não se tivesse aplicado o filtro informativo e se se tivesse utilizado o filtro original (Teorema 6.3) sobre o sistema (10.19), a solução seria diversa. Ter-se-ia uma estimativa enviesada embora de menor variância do que a dada pelos mínimos quadrados (V. Otter, 1985, pp.71-2). No limite, aumentando os valores de $\Sigma_{0|-1}$ e reduzindo os de $\hat{b}_{0|-1}$ e de R , a solução encontraria a dada pelos mínimos quadrados.

A técnica dos *mínimos quadrados recursivos*, como consequência óbvia do acima demonstrado, pode também ser vista como um caso particular do algoritmo de Kalman. Releiam-se os comentários da secção 7 à demonstração de Duncan e Horn (1972) ou veja-se uma demonstração mais directa apresentada recentemente por Diderrich (1985).

Também a *metodologia bayesiana* recursiva, que incorpora as novas observações num modelo linear dinâmico, pode ser vista como um caso particular do procedimento de Kalman. É o caso em que a informação exterior ao modelo é subjectivamente introduzida de forma sistemática. O problema já foi suficientemente desenvolvido na secção 8, pelo que apenas se chama a atenção para uma dedução apresentada recentemente por Meinhold e Singpurwalla (1983). Estes autores tratam as variáveis de estado como quantificando o "estado da natureza" — θ_t — e apresentam o modelo simplificado

$$\begin{aligned} \theta_t &= G_t \theta_{t-1} + w_t & w_t &\sim N(0, W_t) \\ Y_t &= F_t \theta_t + v_t & v_t &\sim N(0, V_t) \end{aligned} \quad (10.21)$$

Com as correspondentes alterações de nomenclatura e simplificações no modelo, a dedução de Meinhold e Singpurwalla (1983, pp.124-5) corresponde à delineada no teorema 8.2.

O problema da *regressão com coeficientes variáveis* encontra também um enquadramento geral na metodologia de Kalman. Esse problema envolve a abordagem de classes muito diversas de modelos de regressão linear

$$y_t = \beta_t x_t + \varepsilon_t \quad (10.22)$$

em que os coeficientes β_{it} evoluem de acordo com algum padrão. Nicholls e Pagan (1985) apresentam um "survey" do problema, tornando-se claro que os casos mais complexos surgem quando se ultrapassa a simples *variação aleatória*, i.e.

$$\beta_t = \bar{\beta} + \eta_t \quad E\eta_t = 0, \quad E\eta_s \eta_t^T = \delta_{st} \Sigma \quad (10.23)$$

e se aborda a chamada *evolução de coeficientes*

$$(\beta_t - \beta) + A_1(\beta_{t-1} - \beta) + \dots + A_p(\beta_{t-p} - \beta) = \eta_t, \quad (10.24)$$

com η_t nas condições de (10.23). O vector de coeficientes é pois tratado como um ARIMA (p, 1, 0), podendo mesmo evoluir como um ARIMA (p, 1, q).

Encarando-se β_t como variável de estado o problema é facilmente enquadrado no sistema de Kalman, uma vez que todo o ARIMA pode aí ser modelizado. Nicholls e Pagan mostram ainda como a estimação de máxima verosimilhança dos coeficientes $\beta_{it} \sim N(\bar{\beta}_i, \sigma_{ii})$ pode resultar facilmente da rotina do filtro (1985, pp.436-9).

Finalmente refiram-se as técnicas de *alisamento exponencial*, que ainda hoje se revelam como das mais potentes na previsão, e que também aparecem como caso particular do filtro de Kalman.

Nas condições do problema 6.3 — F_t, G_t e H_t constantes, tais como as matrizes das covariâncias Q_t e R_t — admita-se que $\Sigma_{t+1|t}$ atingiu um estado estacionário Σ . A determinação desse valor pode aliás fazer-se resolvendo (6.19) em ordem a Σ . O ganho atingirá também o estado estacionário K .

A equação preditiva, expressa em fórmula de correcção do erro

$$\hat{x}_{t+1|t} = F\hat{x}_{t|t-1} + (I - K)\tilde{y}_t$$

ou em fórmula de recorrência

$$\hat{x}_{t+1|t} = (F - KH)\hat{x}_{t|t-1} + Ky_t$$

pode ser adaptada ao processo univariável tratado pelo alisamento exponencial. Como F, K e H são nesse caso escalares, que se representarão por letras minúsculas, ter-se-á

$$\hat{y}_{t+1|t} = f\hat{y}_{t|t-1} + (1 - k)\tilde{y}_t \quad (10.25)$$

ou, alternativamente,

$$\hat{y}_{t+1|t} = (f - kh)\hat{y}_{t|t-1} + ky_t \quad (10.25')$$

resta mostrar que $(f - kh) + k = 1$.

Ora, para aplicar correctamente um alisamento exponencial (simples) pressupõe-se estacionaridade em média do processo (Gilchrist, 1976, p.56). Logo $f \equiv 1$. Além disso, o processo observado coincide com o estado, pelo que $h \equiv 1$. Daí que seja válida a condição de normalização dos ponderadores em (10.25') e que essa expressão se possa escrever:

$$\hat{y}_{t+1|t} = (1 - k)\hat{y}_{t|t-1} + ky_t, \quad (10.26)$$

o que corresponde à forma recorrente definidora do alisamento exponencial.

Não vale a pena prolongar a discussão do problema, que se encontra mais profundamente tratado em Harvey (1984, pp.248 e segs.). Aí se podem encontrar relações entre os parâmetros óptimos do alisamento e as condições iniciais do filtro, tal como a análise do alisamento exponencial duplo.

O objectivo desta breve digressão por metodologias preditivas parece suficientemente atingido. De facto, o Filtro de Kalman aparece como instrumento unificador de uma larga classe de processos de estimação e previsão, que podem assim ser encarados, como seus casos particulares.

Seguindo Durbin (1984), dir-se-á que isso acontece porque hoje se vislumbra uma lógica no desenvolvimento teórico de modelos estocásticos, lógica que corresponde a um processo de décadas. Por um lado, os modelos clássicos de regressão evoluíram dos pressupostos de homoscedasticidade e de independência das perturbações para a consideração de modelos aleatórios mais complexos, em que os próprios coeficientes passaram a ser vistos como estocásticos. Por outro lado, e por vezes em confronto com a escola econométrica, os métodos empíricos de alisamento e previsão ganharam uma dignidade teórica, ao ser demonstrada a sua correspondência com modelos aleatórios, posteriormente enquadrados na classe ARIMA. Como trave unificadora deste processo aparece hoje o sistema de Kalman o qual, englobando os processos ARIMA e permitindo o enquadramento de variáveis exógenas e de *inputs* determinísticos, parece destinar a fornecer soluções integradoras de problemas de estimação que até agora foram tratados de forma isolada e particular.

Não pode por isso deixar de se concordar com Welch (1987) que, em recente polémica com Diderrich (1985), contraria uma tentativa de indicar o filtro de Kalman como um prolongamento dos estimadores mistos de Theil e Goldberger (V. a técnica dos mínimos quadrados recursivos na secção 7). Enquanto Diderrich vê a teoria de Kalman "being essentially random parameter linear regression" (1985, p.193), Welch sublinha que o facto de o filtro poder ser deduzido a partir de um prolongamento de regressão não deve ser enfatizado

pois, na realidade, ele pode ser deduzido como extensão de muitas outras perspectivas de estimação, nomeadamente a bayesiana. "When these same viewpoints are emphasized to the exclusion of the state space and linear system concepts, however, the motivation behind and major applications of the algorithm cannot be fully appreciated" (Welch, 1987, p.91).

Talvez seja elucidativo fazer um paralelo com uma reflexão epistemológica sobre as ciências empíricas. Diz sobre o assunto Karl Popper: se "uma nova teoria com um grau de universalidade mais elevado explicar com sucesso alguma teoria antiga *corrigindo-a* então isso é um sinal seguro de que a nova teoria penetrou mais fundo do que a antiga. À exigência de que uma nova teoria contenha a antiga de forma aproximada [...] pode-se chamar (segundo nisto Bohr) o "*princípio de correspondência*" (1983, p.163 da tradução portuguesa). Parece que os parâmetros fixos são um caso particular da teoria dos parâmetros aleatórios, tal como a incorporação de informação subjectiva é um caso particular da actualização do modelo por incorporação de novos dados... A lógica da relação contrária é que não se vislumbra.

Para finalizar, e depois de visto o sistema de Kalman como integrador de uma larga classe de processos estocásticos e reconhecido o filtro como algoritmo particularizável em diversos processos de estimação, pretende-se ainda destacar duas particularidades notáveis do algoritmo.

Retorne-se, em primeiro lugar, ao problema da *recursividade*. Reconhecido que os processos reais apenas integram a estacionaridade e a aleatoriedade pura como componentes singulares, uma qualidade decisiva dos preditores será a da ponderação mais forte das observações mais recentes. Foi já amplamente mostrado como essa possibilidade é garantida pelo carácter recursivo do Filtro.

Em segundo lugar destaque-se a *característica markoviana* do algoritmo, o que permite uma grande simplicidade na implementação do sistema preditivo. De facto, o filtro de Kalman é markoviano num sentido muito preciso, que não deriva trivialmente da correspondente característica do sistema de espaço de estados.

Defina-se um *processo preditivo sequencial* (Krzysztofowicz, 1987), como composto por um conjunto de previsões sobre o vector aleatório \hat{x}_τ , elaborado com informação reunida em diversos períodos anteriores a τ .

$$\{\hat{x}_{\tau|1}, \hat{x}_{\tau|2}, \dots, \hat{x}_{\tau|\tau}; x_\tau\}$$

A informação existente até cada momento — I_t — é evidentemente markoviana

$$p(I_t | I_1, I_2, \dots, I_{t-1}) = p(I_t | I_{t-1})$$

pois, em notação óbvia $I_t \supset I_{t-1}$, $\forall t$. Mas, para o processo preditivo ser markoviano (no sentido de Krzysztofowicz) é necessário que $\hat{x}_{r|t-1} = f(I_{t-1})$ constitua uma estatística suficiente para $p(\hat{x}_{r|t} | I_{t-1})$, i.e.

$$p(\hat{x}_{r|t} | \hat{x}_{r|t-1}) = p(\hat{x}_{r|t} | f(I_{t-1})) = p(\hat{x}_{r|t} | I_{t-1}), \quad (10.27)$$

o que garante o carácter markoviano do processo $\{\hat{x}_{r|t}, t < \tau\}$ (*ibidem*, p.32).

O problema tem sido recentemente investigado ⁽⁵⁾ pois processos preditivos markovianos permitem construir de forma simples decisões sequencialmente óptimas. Se o processo preditivo for markoviano e se estiver em causa apenas a aplicação $f(I_t)$, a previsão efectuada em cada momento concentra todo o conhecimento antecipativo sobre o estado futuro.

É o que se passa com o Filtro de Kalman, pois, uma vez que as inovações são “ruído branco”, ele verifica a propriedade de Markov

$$E[\hat{x}_{r|t} | \hat{x}_{r|t-s}, s = 1, 2, \dots] = E[\hat{x}_{r|t} | \hat{x}_{r|t-1}] = \hat{x}_{r|t-1} \quad (10.28)$$

e isso implica a característica markoviana do processo preditivo (Doob, 1953, p.81). Além disso, se o sistema for gaussiano, como fica determinada a covariância da previsão a distribuição de $(\hat{x}_{r|t} | \hat{x}_{r|t-1})$ encontra-se completamente especificada.

Dois pressupostos são essenciais neste raciocínio. Um é que as matrizes F, G, H, Σ e K são determinadas *a priori*, independentemente da sequência Y_t que gera o processo preditivo. Outro é que essas matrizes se encontram correctamente especificadas, verificando-se a aleatoriedade pura da sequência de inovações $\{\tilde{y}_t\}$. Só assim se verifica (10.28).

De qualquer forma, uma vez estabelecida a rotina do filtro admite-se que esta produz estimativas centradas de x_t (e portanto de y_t), pelo que implicitamente se admite o carácter markoviano de $\{\hat{x}_{r|t}\}$. Em termos práticos isso justifica a recursividade em um passo: a última previsão e a nova observação determinam a previsão futura — o processo é de uma facilidade computacional notável.

(5) R. Krzysztofowicz e D.R. Davis (1983), “A methodology for evaluation of flood forecast-responses systems, 2. theory”, *Water Resources Research*, 19, pp.1419-55.

11. Alguns sucessos recentes

A orientação deste trabalho tem sido essencialmente teórica e não se pretende, nesta fase final, apresentar o Filtro de Kalman em competição prática com outros algoritmos preditivos. Algumas referências são necessárias, pois o valor de um método está ligado à sua eficácia; mas essa ligação não é directa e seria perigoso reduzir uma validação teórica ao pragmatismo. Vale a pena citar de novo Popper.

"A questão da sobrevivência de uma teoria é coisa que pertence ao seu destino histórico e, portanto, à história das ciências. Por outro lado, o seu uso para previsões é um assunto ligado à sua aplicação. Estas duas questões estão relacionadas, mas não intimamente, pois nós muitas vezes aplicamos teorias sem nenhuma hesitação mesmo que elas estejam mortas — isto é, falsificadas — enquanto forem aproximações suficientemente boas para o fim em vista. Assim, não há nada de paradoxal em eu estar a apostar em aplicações de uma teoria ao mesmo tempo que me recuso a apostar na sobrevivência dessa mesma teoria" (1983, p.93 da tradução portuguesa).

Substitua-se "teoria" por "sistema + algoritmo" e este ponto de vista tornar-se-á perfeitamente claro. De facto, se o sistema de espaço de estados reencontra o modelo econométrico, os modelos ARIMA e muitos outros como seus casos particulares, e se o Filtro de Kalman é algoritmo que pode conter a regressão, a máxima verosimilhança e outros procedimentos, a questão do sucesso prático do filtro fica muito reduzida, neste contexto comparativo, a um mero problema técnico ou de técnica computacional.

É por isso que a previsão competitiva ("forecast competition") deve ser encarada nos seus limites e não absolutizada. O problema é tão mais importante ser referido quanto essa competição tem sido uma das mais marcadas práticas dos últimos anos. Como o reconheceu Meade (1985, p.57) "the decade before last a number of new forecasting approaches were introduced, notably the Arima models of Box and Jenkins and the Kalman filter based approaches, such as that of Harrison and Stevens. The work of the last decade has been more concerned with consolidation and the comparative study of existing techniques [...] the most visible activity has been the forecast competition".

É neste enquadramento que devem ser lidas as linhas que seguem.

Numa primeira parte serão apresentados alguns estudos comparativos que destacam as potencialidades do Filtro de Kalman. Numa segunda parte apresentar-se-ão aplicações económicas do modelo em espaço de estados. Na fase seguinte referem-se aplicações em áreas diversificadas.

SIMULACÕES E ESTUDOS COMPARATIVOS

Que se tenha conhecimento directo a primeira aplicação sistemática do Filtro de Kalman à previsão econométrica foi realizada por Morrison e Pike, num trabalho de 1975 publicado em 1977 na revista do TIMS, *Management Science*.

O trabalho tem várias limitações teóricas ⁽¹⁾ e metodológicas mas tem o mérito de apresentar os mais conhecidos modelos de crescimento univariável na óptica do espaço de estados.

Morrison e Pike começam por apresentar os modelos de crescimento constante, linear, quadrático e exponencial, todos na formulação de espaço de estados. A equação de medida é sempre

$$y_t = \mu_t + \varepsilon_t \quad (11.1)$$

e as equações e transição (depois de simplificadas) apresentam-se na forma

$$\text{constante: } \mu_t = a_0 \quad (11.2)$$

$$\text{linear: } \mu_t = \mu_{t-1} + a_1 \quad (11.3)$$

$$\text{quadrática: } \mu_t = \mu_{t-1} + a_1 + \frac{a_2}{2} \left(\frac{t-1}{2} \right) \quad (11.4)$$

$$\text{exponencial: } \mu_t = \mu_{t-1} + \alpha \mu_{t-1} \quad (11.5)$$

A essas equações foram retiradas as componentes aleatórias e, por isso mesmo, o filtro perde a versatilidade.

Morrison e Pike estudam a sucessão cronológica que representa a procura de energia eléctrica nos Estados Unidos de 1945 a 1972. O período de estimação consiste nos primeiros catorze anos, que utilizam para inicializar o modelo, sendo os coeficientes a_0, a_1, a_2 estimados pelo método dos mínimos quadrados.

Assim, não é de espantar que a previsão *ex-post* conduza exactamente aos mesmos resultados, tanto pela regressão linear como pelo Filtro de Kalman, modelo a modelo.

Aplicam depois o modelo exponencial, mas apenas pelo filtro de Kalman, não sendo concludente o facto de ele produzir melhores resultados do que as simulações anteriores. É curioso que o método de inicialização do filtro recorra a $\hat{\alpha}$ procurado numa grelha, por

(1) Como por exemplo a de afirmar que a optimalidade do filtro apenas se verifica em processos gaussianos (p.772).

tentativas sucessivas de minimização do quadrado dos erros. Morrison e Pike parecem não se aperceber de que estão apenas a utilizar uma técnica interativa de mínimos quadrados, num modelo de coeficientes fixos.

Concentram-se em seguida sobre aspectos computacionais do filtro e aí produzem algumas observações interessantes. Reconhecem que o filtro evita a inversão matricial exigida pelos mínimos quadrados e que a sua recursividade permite manter o nível de simplicidade da computação, mesmo aumentando as observações. Reconhecem ainda a diversidade de modelos tratáveis pelo filtro.

Raciocinando sobre a equação de previsão e sobre o ganho

$$\hat{x}_{t+1|t} = F\hat{x}_{t|t-1} + K_t\tilde{y}_t$$

$$K_t = (F\Sigma_{t|t-1}H)(H\Sigma_{t|t-1}H + R)^{-1}$$

verificam que o pressuposto de erros largos na equação de medida conduz a um ganho reduzido e uma ponderação menor de observações "erráticas". Por outro lado, valores elevados na covariância de x_t levam a uma dominância de Σ e a um correlativo maior peso das observações. Aliás, é reconhecido que a subestimação dessas matrizes pode conduzir a problemas de divergência (V.Mehra, 1979, p.87).

Estudos comparativos mais fundamentados e sistemáticos foram conduzidos recentemente por Otter, Tempelaar e outros econométricos holandeses.

Num primeiro trabalho, relatado por Otter (1985, pp.114-9), foi estudado um modelo explicativo do investimento nos Estados Unidos, de 1920 a 1940. O modelo comporta os lucros e o stock de capital — P_t e K_t — como variáveis exógenas:

$$I_t = \beta_0 + \beta_1 P_t + \beta_2 P_{t-1} + \beta_3 K_{t-1} + v_t \quad (11.6)$$

A este modelo corresponde o sistema de Kalman

$$\beta_{t+1} = \beta_t$$

$$I_t = H_t \beta_t + v_t \quad (11.7)$$

com $H_t = [1 \ P_t \ P_{t-1} \ K_{t-1}]$ e β vector coluna dos coeficientes β_i , que representam o estado. Admitiu-se $E v_t = 0$, $E v_t v_s = \delta_{ts} \sigma^2$.

O modelo foi estimado recursivamente pelos mínimos quadrados e pelo Filtro de Kalman, registrando-se as previsões em um passo para $t = 5, 6, \dots, 21$. A qualidade das estimativas foi medida pelo coeficiente de erro de previsão de Theil ⁽¹⁾:

$$U = \frac{(\sum \tilde{I}_t^2)^{\frac{1}{2}}}{(\sum \hat{I}_t^2)^{\frac{1}{2}} + (\sum I_t^2)^{\frac{1}{2}}} \quad (11.8)$$

O filtro de Kalman foi inicializado com $\hat{\beta}_{0|-1} = 0$ e com diferentes valores de P_0 e $\hat{\sigma}^2$, covariância *a priori* de β e variância *a priori* de v_t . A matriz P_0 foi considerada escalar ($P_0 = \delta I$) percorrendo δ valores numa grelha entre 1 e 100; σ^2 foi testado numa grelha entre 1 e 1000.

Na melhor situação — $\delta = 100$ e $\sigma^2 = 1$ — atingiu-se um coeficiente de Theil de 0.13, indicando uma melhoria relativamente ao OLS, que obteve $U = 0.17$.

Como factos salientes deste estudo comparativo nota-se a relativa superioridade do filtro com uma *prior* difusa — $P_0 = 100I$ — e com reduzido erro na equação de medida — $\hat{\sigma}^2 = 1$. Além disso o filtro conseguiu uma convergência de $\hat{\beta}_t$ para os valores finais muito mais rápida do que o OLS. Correspondentemente, a variância empírica de $\hat{\beta}_t$ foi menos acentuada (V. quadro p.118 em Otter, 1985).

Testando depois um modelo com parâmetros estocásticos, Otter conseguiu uma ligeira melhoria das previsões ao admitir uma variância reduzida para $\beta_{0,t}$ — $\sigma^2 = 0.1$. Nesse caso atinge-se $U = 0.125$. Admitindo a variabilidade conjunta de todos os parâmetros de β a qualidade das estimativas piora, embora se mantenha geralmente superior à dada pelos mínimos quadrados.

O que parece daqui concluir-se é uma relativa superioridade do filtro, mas à custa de uma laboriosa procura interactiva da parametrização óptima. Naturalmente, a variabilidade "real" dos valores de β não era muito acentuada na sucessão cronológica em estudo.

Num outro trabalho Otter e Tempelaar (1980) comparam o filtro e os estimadores de Box-Jenkins num conjunto de sucessões simuladas. O modelo de base é um ARIMA (2, 1, 0) da forma

$$y_t = \beta_1 y_{t-1} + \beta_2 y_{t-2} + k + \varepsilon_t \quad (11.9)$$

⁽¹⁾ V. Theil, 1965, *Applied Economic Forecasting* ou, e.g., Gilchrist, 1976, pp.235-40.

em que k representa uma constante e $\{\varepsilon_t\}$ é processo estacionário puramente aleatório. A justificação para a omissão do componente de médias móveis consiste no facto de um ARMA estacionário poder ser expresso como um AR (de ordem geralmente não finita (2)) e, conseqüentemente um processo clássico de estimação consistir na truncadora desse AR a partir de dada ordem prefixada ou revelada pela análise da autocorrelação.

O sistema de Kalman foi construído por técnica semelhante à apresentada no exemplo 9.3,

$$\begin{bmatrix} \beta_1 \\ \beta_2 \\ K \end{bmatrix}_{t+1} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ K \end{bmatrix}_t + \varepsilon_t$$

$$y_t = [y_{t-1} \quad y_{t-2} \quad 1] \begin{bmatrix} \beta_1 \\ \beta_2 \\ K \end{bmatrix}_t + \varepsilon_t \quad (11.10)$$

sendo a estimação realizada pela rotina do filtro.

A estimação de Box-Jenkins foi realizada pelo método da máxima verosimilhança (Box e Jenkins, 1970, pp.274 e segs.).

Otter e Tempelaar efectuaram dois blocos de simulações, fazendo β_1 e β_2 percorrer a zona de R^2 na qual o processo (11.9) é estacionário nas primeiras diferenças. Em cada corrida geraram 50 observações de y_t .

No primeiro bloco de simulações o filtro foi inicializado com $\hat{\beta}_{0|-1} = [0 \quad -1/3 \quad 1]^T$ e $\Sigma_{0|-1} = [2/3 \quad 2/9 \quad 1/4]$. Essas estimativas *a priori* correspondem ao pressuposto de os parâmetros serem independente e uniformemente distribuídos na zona de estacionaridade. A geração das sucessões simuladas seguiu esse critério. As estimativas da constante foram as mesmas que os correspondentes parâmetros das sucessões simuladas.

No segundo bloco de simulações todos estes pressupostos se mantiveram, com excepção do valor *a priori* da constante, fazendo-se $\hat{k}_{0|-1} = 0$ com o objectivo de testar a sensibilidade do filtro de uma errónea especificação inicial.

Em quase todos os casos o filtro conseguiu uma convergência para os valores dos parâmetros muito mais rápida do que o método de Box e Jenkins. Até à décima observação as estimativas do filtro registaram um erro quadrático médio entre 1/2 e 1/3 do das correspondentes estimativas de Box-Jenkins. Entre a vigésima e a quadragésima observação as estimativas mostraram diferenças pouco significativas.

(2) V. Nota Anexa 10

Os casos em que o filtro relevou um erro de estimação inferior foram, na generalidade, aqueles em que os parâmetros se encontravam perto da fronteira de estabilidade.

Facto mais significativo é que a superioridade do filtro, quando revelada nas primeiras dez estimativas, se mantinha ao longo da sucessão de observações, embora com diferenças menos marcadas à medida que cada sucessão evoluía.

Comparando depois os métodos de acordo com o erro quadrático médio da estimação de cada parâmetro, Otter e Tempelaar definiram como critério de superioridade o registo de um erro quadrático médio menor ou igual para todos os parâmetros e menor pelo menos numa observação.

A superioridade do Filtro de Kalman é igualmente surpreendente. Em 64% dos casos o filtro revelou-se melhor técnica, enquanto o estimador de Box-Jenkins apenas mostrou melhores resultados em 4% das simulações. Os restantes 32% dos casos não foram conclusivos.

Ainda mais revelador é o facto de o filtro ter sido preterido apenas em "corridas" correspondentes ao segundo bloco de simulações: aquele em que a estimativa inicial da constante estava propositadamente falsificada. Por isso os autores concluem: "there is a strong preference for applying the Kalman Filter technique in favour of the Box-Jenkins M.L.-estimation technique even when the prior information is of poor quality" (p.211).

Dois anos mais tarde, Makridakis e outros sete investigadores promovem um estudo sistemático dos métodos de previsão utilizando 1001 sucessões cronológicas e 24 técnicas⁽³⁾. Para os fins em vista as conclusões de tal estudo não serão "definitivas", dadas algumas diferenças de critério que complicam a comparação dos métodos (V.Meade, 1985, pp.71-5). De qualquer forma é significativo que a utilização do Filtro de Kalman pelo método de Harrison e Stevens tenha conquistado o primeiro lugar no estudo das 1001 sucessões, isso quando adoptado o critério da média do erro quadrático médio de previsão em um passo (*idem* p.73)

Talvez valha a pena referir um outro estudo comparativo recente (Crato, 1986) em que a qualidade das previsões foi testada em sucessão cronológica real e num momento de alterações estruturais.

Foi estudada a evolução do Índice de Preços no Consumidor (Continente, total sem habitação) no período de Janeiro de 1983 a Outubro de 1986. Tempos antes Black e Murteira (1983) tinham analisado a mesma série num período anterior e mais extenso (1968-1982).

(3) S. Makridakis et al. (1982), "The accuracy of extrapolation (time series) methods: results at a forecasting competition", *Journal of Forecasting*, 1, pp.111-53.

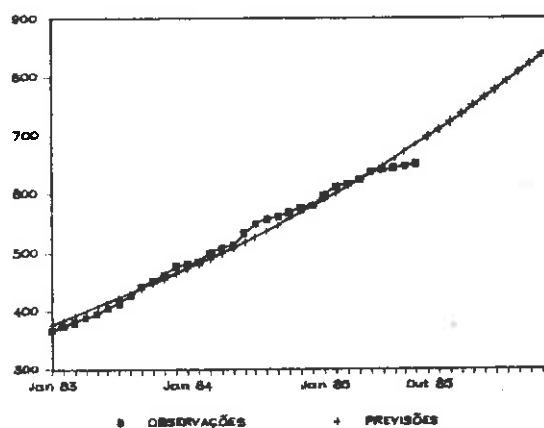
A sua conclusão era a de que, no período de 1973 a 1982 “a tendência da exponencial é dominante, por isso, todos os métodos de previsão são bons desde que partam dessa conclusão” (p.292). Comparando modelos “globais”, econométricos, com modelos “locais”, de alisamento, e com esquemas ARIMA, Black e Murteira concluem pela quase identidade das previsões fornecidas pelos dois primeiros, pondo algumas reticências à aplicabilidade, naquele contexto, do método de Box-Jenkins.

Poucos anos mais tarde a situação tinha-se modificado e, entre os últimos meses de 1985 e os primeiros meses de 1986, o IPC passou a registar um crescimento médio muito inferior. Tem então sentido considerar como período de previsão *ex-post* os meses de Novembro de 1985 ($t = 1$) a Outubro de 1986 ($t = 12$) e sobre esses momentos de mudança testar a qualidade dos preditores.

A sucessão cronológica foi previamente dessazonalizada, pelo método dos coeficientes multiplicativos médios, resultando a nova sucessão $\{I_t\}$, à qual foi ajustada o modelo econométrico de taxa de crescimento constante

$$I_t = a.e^{rt+u_t} \quad (11.11)$$

No período de previsão obteve-se um elevadíssimo erro quadrático médio (8 331.6), o que revelava um nítido desajustamento do modelo.



Uma solução graficamente óbvia parecia ser a de reter a última observação I_0 e construir um novo esquema preditivo

$$\hat{I}_t = I_0 \cdot e^{rt} \quad (11.12)$$

As estimativas são melhores mas ainda de elevado erro quadrático médio (1 703.1). O chamado modelo global parecia claramente desajustado pois o que se revelava era uma alteração significativa da taxa de crescimento r .

Ensaçou-se então um método adaptativo simples derivado directamente do alisamento exponencial. O parâmetro r passou a ser estimado por esse processo, sendo a estimação e as previsões dadas por

$$\begin{aligned} \hat{r}_{t+1} &= \hat{r}_t + \alpha(\hat{r}_t - r_t) \\ \hat{I}_{t+1} &= I_t \cdot \hat{r}_{t+1} \end{aligned} \quad (11.13)$$

com $r_t = (I_t - I_{t-1})/I_{t-1}$.

Obteve-se imediatamente uma redução drástica do erro de previsão ($EQM = 11.5$), revelando uma superioridade notória dos métodos adaptativos que pressupõem a variabilidade dos parâmetros. De facto, como foi já amplamente demonstrado, (11.13) correspondia à pressuposição de um esquema de Muth (V. eq. 8.6).

Deduziu-se em seguida o filtro de Kalman no quadro bayesiano e face ao sistema

$$\begin{aligned} r_t &= r_t^* + v_t \\ r_t^* &= g \cdot r_{t-1}^* + w_t \end{aligned} \quad (11.14)$$

em que r_t^* designa uma taxa de crescimento teórica ou implícita e os pressupostos sobre os ruídos são os usuais.

O parâmetro g foi apontado empiricamente como igual a 0.95 ⁽⁴⁾ e as previsões viram o seu erro quadrático médio ser ainda reduzido para 9.8.

O grande salto, contudo, verificou-se na passagem de modelos de coeficientes constantes para modelos que admitiam a variabilidade da taxa de crescimento r_t . Foi ainda demonstrado, na condição de constância da seguinte quantidade

$$\alpha = (g^2 \sigma_{r_t^*}^2 + \sigma_{w_t}^2) / (g^2 \sigma_{r_t^*}^2 + \sigma_{w_t}^2 + \sigma_{v_t}^2),$$

⁽⁴⁾ Ver o desenvolvimento dos pressupostos na Nota Anexa 12.

que o modelo (11.14) é equivalente ao (11.13), desempenhando α o papel de constante de alisamento.

Um reparo que hoje pode ser feito a esse estudo é o facto de as estimativas do filtro e do alisamento serem corrigidas no decorrer do período de simulação ao contrário do que admitido nos modelos econométricos.

Essa nota não parece ser qualitativamente significativa, pois a recursividade dos mínimos quadrados não iria alterar muito as correspondentes estimativas, pois existe um grande peso das numerosas observações passadas. Por outro lado, a recursividade é característica intrínseca dos modelos de alisamento e de filtragem, enquanto que a sua extensão aos modelos econométricos de coeficientes constantes é algo subsidiária.

Mas, mesmo que se não incorporassem as novas observações no Filtro de Kalman, as conclusões essenciais mantinham-se, pois nesse caso o erro quadrático médio ainda seria 1,5% do encontrado no modelo (11.11) e 4,7% do revelado pelo esquema (11.12). De facto, com a pior hipótese de $g = 1$ e sem a introdução correctiva das observações de $t = 1, 2, \dots, 12$, ter-se-ia $EQM=127.8$.

APLICAÇÕES ECONÓMICAS DO MODELO EM ESPACO DE ESTADOS

Um *case-study* repetidamente citado é o estudo feito por Szelag em 1980 referente à previsão de tráfego nas linhas telefónicas e publicado em 1982 no *Bell System Technical Journal*. O estudo teve como objectivo a implementação de medidas economicamente racionais para a mobilização de equipas de serviço.

O modelo construído por Szelag (1982, pp.73-6) é um modelo estrutural que precede a formulação de Harvey (1984). A evolução é decomposta nos componentes tendenciais, sazonais e aleatórios. A tendência é assumida como linear, reescrevendo o modelo de Theil e Wage em espaço de estados, ⁽⁵⁾

$$T_t = [1 \ 0] \begin{bmatrix} \mu_t \\ \beta_t \end{bmatrix} + v_{1t} \quad (11.15')$$

$$\begin{bmatrix} \mu_{t+1} \\ \beta_{t+1} \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \mu_t \\ \beta_t \end{bmatrix} + \begin{bmatrix} w_{1t} \\ w_{2t} \end{bmatrix} \quad (11.15'')$$

As matrizes H_T e F_T são, respectivamente, $[1 \ 0]$ e $\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$.

(5) A notação encontra-se adaptada à até aqui seguida.

A sazonalidade é descrita por um desenvolvimento trigonométrico relativamente complexo (V. Szlag, 1982, pp 73-5) que Harvey posteriormente simplifica (1984, p.254). Szlag deduz uma matriz de transição para os coeficientes sazonais $[\alpha_1 \ \alpha_2 \ \dots \ \alpha_L]$ escrevendo-se

$$\begin{aligned} S_t &= H_S \alpha_t + v_{2t} \\ \alpha_{t+1} &= F_S \alpha_t + w_{st} \end{aligned} \quad (11.16)$$

Os dois sistemas de equações são consensados num único, tal como em (10.17).

O que é particularmente interessante neste estudo são os mecanismos simplificadores da estimação. Szlag começa por recomendar que não se façam estimativas do estado sem antes reunir um conjunto elevado de observações, de $t=0$ até $t=T$.

É depois feita uma estimação de x_t e inicializado o filtro com $\hat{x}_{T+1|T} = F\hat{x}_T$.

A novidade do processo consiste em expressar \hat{x}_T função das observações $Y_T := [y_T, y_{T-1}, \dots, y_1]^T$. Para isso nota-se, no caso concreto em estudo, que F é invertível. Daí como $y_t = Hx_t + v_t = HF^{-1}(x_{t-1} - w_{t-1}) + v_t$, deduz-se:

$$\begin{aligned} y_T &= Hx_T + v_t \\ y_{T-1} &= HF^{-1}x_T - HF^{-1}w_{T-1} + v_{T-1} \\ &\dots \\ y_1 &= HF^{-(T-1)}x_T - H \left[\sum_{i=1}^{T-1} F^{T+i} w_{T-i} \right] + v_1 \end{aligned}$$

Em notação matricial, concentrando na matriz D a expressão determinística e no vector a a aleatória, pode escrever-se

$$Y_T = Dx_T + a \quad (11.17)$$

com

$$D = \begin{bmatrix} H \\ HF^{-1} \\ \dots \\ HF^{-(T-1)} \end{bmatrix} \quad a = \begin{bmatrix} a_T \\ a_{T-1} \\ \dots \\ a_1 \end{bmatrix}$$

$$a_{T-k} = - \sum_{i=1}^k HF^{-k-1+i} w_{T-i} + v_{T-k}, \quad 1 \leq k < T$$

$$a_T = v_T$$

como $Ea = 0$ o cálculo da covariância $Eaa^T = \Omega = [w_{jk}]$ produz-se elemento a elemento pela expressão

$$w_{jk} = \sum_{i=1}^{j-1} E[HF^{-j+i} w_{T-i}][HF^{-k+i} w_{T-i}]^T + E v_{T-j+1} v_{T-k+1}$$

$$= \sum_{i=1}^{j-1} HF^{-j+i} Q_{T-i} (HF^{-k+i})^T + \delta_{jk} R_k$$

tendo-se pressuposto que $Ewv = 0$.

A estimação é agora possível pelos mínimos quadrados

$$\hat{x}_T = (D^T \Omega^{-1} D)^{-1} D^T \Omega^{-1} Y_T \quad (11.18)$$

Resta o problema da determinação de \hat{Q}_t e \hat{R}_t , necessário à resolução de (11.18) e aos cálculos posteriores. Szelag, depois de simulados distintos valores, nota que as estimativas são relativamente insensíveis dentro de uma determinada gama de parâmetros.

Como, tanto para o cálculo de (11.18) como para a computação do ganho K_t há um *trade-off* entre Q_t e o escalar R_t , Szelag fixou $R_t = 1$ e procurou valores para Q_t que produzissem estimativas aceitáveis. Calculou então \hat{x}_T , o que lhe permitiu arrancar o filtro e arrancar a rotina preditiva.

Uma outra simplificação foi introduzida no cálculo de ganho K_t . Iniciada a rotina de estimação foi parado o cálculo de K_t no momento τ em que começavam a ser gerados resultados razoáveis. É então admitida a *sequência truncada* $\{K_t^*\}$:

$$\begin{aligned} K_t^* &= K_t & \text{para } t \leq \tau \\ K_t^* &= K_\tau & \text{para } t > \tau \end{aligned} \quad (11.19)$$

Além das evidentes vantagens computacionais, este processo impede uma divergência do filtro por incorrecta especificação de Q_t e R_t . De facto (V. Szelag, 1982, p.84), se o filtro iniciar um processo de divergência — revelado por uma sequência de inovações \tilde{y}_t claramente distinta de um “ruído branco”, a truncagem de K_t pode impedir a divergência. Raciocinando por analogia com o alisamento exponencial trata-se de fixar os parâmetros em valores razoáveis (normalmente $\beta \in]0; 0.3[$), o que impede a divergência provocada por parâmetros “explosivos” (e.g. $\beta > 1$ ou $\beta < 0$).

Analisando os resultados finais Szelag conclui por uma nítida superioridade do filtro face aos processos usualmente utilizados na Bell. Essa superioridade é mais notória na previsão a curto prazo — reduzindo o erro médio a metade do usual — e menos acentuada na previsão a médio-longo prazo, em que o erro médio absoluto é apenas reduzido em vinte por cento.

Ainda na área da modificação e previsão de tráfego nas linhas telefónicas é de destacar um estudo empreendido nos TLP por Tomé e Cunha (1985).

Neste último trabalho a modelização do crescimento é idêntica à de Szelag — equações (11.15) — e a sazonalidade é considerada aditiva e expressa por um *input* determinístico u_t .

O que é aqui particularmente interessante é o processo de estimação das covariâncias R_t e Q_t .

Tomé e Cunha consideram a primeira matriz com um único elemento, e a segunda como diagonal. No modelo (11.15) os parâmetros a estimar serão pois $\gamma = (r, q_{11}, q_{22})$. Para o efeito construíram um processo de simulação que gerava como inputs $x_{0|t-1}, v_t, w_{1t}, w_{2t}$ e iniciaram várias corridas de simulação com diferentes valores de γ .

Em cada corrida calcularam a função de verosimilhança para as N observações geradas

$$F = \prod_{t=1}^N p(y_t | Y_{t-1}, \gamma) \quad (11.20)$$

e escolheram γ como aquele vector de parâmetros que produziu um maior valor de $\ln F$. Intuitivamente, isso significa que as matrizes R e Q são escolhidas como aquelas a que corresponde uma sequência de observações mais concentradamente determinada pelo sistema, em termos de distribuição condicional.

Para o cálculo é necessário à partida admitir algo mais sobre as distribuições. Em rigor tal não é imprescindível se se considerar a “verosimilhança” como método de estimação e

não como a verdadeira verosimilhança (V. Hanan, 1986, p.7 e Brockwell e Davis, 1987b, p.249).

Tomé e Cunha consideraram o sistema como gaussiano, o que permite determinar

$$F = (2\pi)^{-N/2} \cdot \prod_{t=1}^N \sigma_t^{-1} \cdot \exp \left\{ -\frac{1}{2} \sum_{t=1}^N \frac{\tilde{y}_t^2}{\sigma_t^2} \right\} \quad (11.21)$$

Na expressão, σ_t^2 representa a variância de $(y_t|Y_{t-1}, \gamma)$ e, sendo $\hat{y}_{t|t-1}$ a sua média, a inovação é dada por $\tilde{y}_t := y_t - \hat{y}_{t|t-1}$. Por outro lado, sabe-se que a sucessão de inovações é um ruído branco em filtro óptimo e que os momentos de $(y_t|Y_{t-1}, \gamma)$ são dados por

$$\begin{aligned} \hat{y}_{t|t-1} &= H\hat{x}_{t|t-1} \\ \sigma_t^2 &:= E\tilde{y}_t^2 = H\Sigma_{t|t-1}H^T + R \end{aligned}$$

O próprio filtro fornece então processo para o cálculo de $\ln F$:

$$\ln F = C + \frac{1}{2} \sum_{t=1}^N [\ln |H\Sigma_{t|t-1}H^T + R| + (y_t - H\hat{x}_{t|t-1})^2 / (H\Sigma_{t|t-1}H^T + R)] \quad (11.22)$$

Com este processo de estimação Tomé e Cunha conseguiram estabelecer um sistema de previsão do tráfego telefónico com reduzido erro quadrático e, talvez potencialmente mais importante, obtiveram um filtro que rapidamente encontra um estado estacionário no ganho de Kalman e no erro de previsão (1985, p.4).

Um outro trabalho de origem nacional é a tese de doutoramento de Assis Lopes (1985). O autor estuda pormenorizadamente o mercado do cobre procurando metodologias de previsão particularmente adaptáveis. No curto prazo conclui pela utilidade dos modelos de espaço dos estados enquanto no médio-longo prazo reconhece a metodologia de Taylor⁽⁶⁾ como superior.

Como novidade ao até aqui apresentado saliente-se a utilização paralela de diferentes modelos de espaço de estados para a combinação de previsões (Lopes, 1985, pp.268-81).

(6) Ver S.J. Taylor (1979), "Price Trends in capital markets — a survey of models, autocorrelation tests, forecasts, purchasing and speculation", *Sixth Annual meeting of European Finance Association*, Bergen, e, mais recentemente, S.J. Taylor (1986), *Modelling Financial Times*, Wiley

Lopes constrói dois modelos de evolução do preço do cobre. Um de tendência "localmente" constante, correspondente ao modelo de Muth; outro de tendência "localmente" linear, tal como o modelo de Theil e Wage (ver equações 8.6 e 8.11).

A previsão combinada será

$$\hat{y}_t = w\hat{y}_t^1 + (1-w)\hat{y}_t^2 \quad (11.23)$$

em que \hat{y}_t^1 é dado pelo modelo de crescimento linear e \hat{y}_t^2 pelo modelo constante.

De forma geral a combinação de previsões (V. Granger e Newbold, 1986, pp.266-76) é dada por

$$\hat{y}_t = \sum w_i \hat{y}_t^i \quad (11.24)$$

impondo-se condições a $\sum w_i = 1$ de forma que $E\hat{y}_t = Ey_t$ e que $E\hat{y}_t \hat{y}_t^T$ seja minimizado.

Lopes utilizou um período de estimação para determinar aquele valor de w que permitia um menor erro médio da previsão combinada (7). Em seguida impôs condições à previsão final, de forma a que a resposta a variações locais acentuadas não projectasse a tendência para valores erráticos.

Para isso definiu o intervalo

$$I_t = [m_t - \sigma_t u ; m_t + \sigma_t u]$$

em que m_t e σ_t são os parâmetros da tendência (média e desvio padrão) e u é valor calculado com previsões a 1, 5, 10 e 20 passos, e calibrado de forma a obter o menor erro quadrático médio para os vários passos de previsões. Sempre que a previsão combinada não ficasse contida no intervalo I_t , seria adoptada apenas a previsão dada pelo modelo sem crescimento (Muth).

Com este processo Assis Lopes conseguiu melhorar a qualidade das previsões, revelando-se a combinação de Filtros de Kalman como "processo de modelação e estimação particularmente útil para o mais curto prazo" (Lopes, p.283).

(7) V. metodologia alternativa em Diebold e Pauly (1987, pp.25-33).

Para finalizar esta breve referência à previsão com modelos em espaço de estados refiram-se três estudos muito recentes focalizados na evolução da procura de energia eléctrica e baseados na aplicação FORECAST MASTER. Os estudos são extensos e pormenorizam a análise dos condicionantes exógenos ao consumo de energia. Vale a pena resumir as suas conclusões sobre a qualidade das estimativas do Filtro de Kalman.

Nelson (1987) comparou os métodos usualmente utilizados na Georgia Power Company — ARMA — com a modelização equivalente em espaço de estados. No que se refere à previsão do *consumo das famílias* concluiu pela superioridade do filtro, tanto em erro de previsão como no enviezamento das estimativas. Nas *vendas comerciais* encontrou erros graves nos dois processos, mais graves ainda no Filtro de Kalman, o que explicou por erradas especificações do modelo. Na previsão dos *picos mensais* concluiu por uma nítida superioridade do modelo de espaço de estados e correlativa previsão pelo filtro.

Watson, Pastuszek e Cody (1987) utilizaram diversos modelos econométricos, ARIMA, de alisamento e de espaços de estados, para prever a procura de energia num período de 12 meses. Concluíram por uma nítida superioridade dos métodos econométricos, seguidos pelos métodos de alisamento, e encontraram deficiências graves nos métodos ARIMA e no Filtro de Kalman. No que se refere a este último consideraram particularmente difícil obter uma componente sazonal que refletisse as variações periódicas. Contudo, não estudaram um modelo de coeficientes fixos.

No entanto, ao decompôr as previsões em dois períodos anuais (Outubro de 1983 a Setembro de 1984 e Outubro de 1984 a Maio de 1985), concluíram por uma qualidade de previsão semelhante nos modelos econométricos, de alisamento e de espaço de estados, embora no segundo período o filtro se deteriorasse sensivelmente. A explicação do facto não é clara mas parece residir na construção do modelo, que foi aplicado numa série diferenciada e sem tendência (pp.124 e 136).

Finalmente Gunel (1987), num estudo para o Electric Power Research Institute, do Canadá, compara modelos de Box-Jenkins, econométricos, de espaço de estados e bayesianos autoregressivos (VAR). As suas conclusões fundamentais são as de que os modelos econométricos são os mais adequados para a previsão a médio prazo (um ano) enquanto os de espaço de estados apresentam estimativas de melhor qualidade para a curto prazo (1 a 3 meses). No entanto, uma vez que o *package* utilizado (FORECAST MASTER) não permitiu a inclusão de todas as variáveis exógenas consideradas no modelo econométrico (p.150), tal conclusão tem de ser relativizada.

A terminar, note-se um facto quase sistematicamente relatado nos estudos referidos: a

qualidade do Filtro de Kalman para a previsão a curto prazo e a superioridade possível de outros modelos, nomeadamente econométricos, na previsão a longo prazo.

A verdade é que a utilização dos modelos de espaço de estados na previsão económica dá ainda os seus primeiros passos. Mas parece estar já longe o tempo em que Kendall afirmava: "The utility of these [Kalman Filter] methods outside the realm of control engineering has yet to be demonstrated" (1973, p.128).

APLICAÇÕES DO FILTRO DE KALMAN EM ÁREAS DIVERSAS

A primeira aplicação prática significativa dos resultados de Kalman terá tido lugar na aeronáutica, podendo mesmo dizer-se que o sucesso das missões Apolo está intimamente ligado à precisão conseguida pelo filtro, que foi instrumento de determinação das órbitas e de cálculo e correcção de trajectórias. Uma exposição pormenorizada dessas aplicações encontra-se em Jazwinski (1970, pp.292-301, 318-29) e algumas referências sintéticas em Brown (1983, pp.222-9, 298-300).

Como exemplos adicionais vão referir-se muito brevemente aplicações nas áreas do risco e seguros, análise de intervenção, teoria da amostragem e biologia.

Começando pela teoria do risco refira-se um "survey" empreendido por Zehnwirth (1985), que unifica diversas abordagens de estimação recursiva no quadro do Filtro de Kalman. Zehnwirth mostra como a derivação do filtro pela sequência de inovações e utilizando a teoria dos espaços de Hilbert, permite tratar vários modelos da teoria do risco, com distribuições de Poisson, Gama e normais, num mesmo enquadramento básico.

Veja-se igualmente Rantala (1986) que estuda uma política óptima de prémios formulando um modelo em espaço de estados. Pretendendo estabelecer uma tarifação de prémios (rating policy) que seja óptima num conjunto de períodos e não num momento único, como é tradicional na teoria dos seguros, Rantala é levado a construir uma função objectivo dificilmente tratável pelos métodos clássicos mas facilmente abordável pelo Filtro de Kalman.

As variáveis em causa são as indemnizações aos segurados (z_t), a margem de solvência da companhia (u_t) e os prémios (p_t).

Rantala assume que as indemnizações são estacionárias em d-ésima diferença e que podem ser modelizadas como um ARIMA.

$$\alpha(B) \Delta^d z_t = \beta(B) \epsilon_t \quad (11.25)$$

A margem de solvência é definida pela igualdade

$$u_{t+1} = r_t \cdot u_t + p_t - z_t \quad (11.26)$$

em que r_t representa um coeficiente de actualização.

Admitindo que todas as variáveis foram previamente centradas define-se como objectivo a minimização do critério quadrático de eficiência num conjunto de períodos:

$$E \left\{ \sum_{t=1}^N [u_t^2 + v \cdot (\Delta^d p_t)^2] \right\}, \quad v \geq 0 \quad (11.27)$$

As equações (11.25) e (11.26) podem ser sintetizadas na seguinte

$$(1 - rB)\alpha(B) \Delta^d u_{t+1} = \alpha(B) \Delta^d p_t - \beta(B)\varepsilon_t \quad (11.28)$$

de onde se deduzem as equações de transição expressas, por exemplo, na forma seguinte

$$x_{t+1} = Fx_t = Gw_t + \Gamma \Delta^d p_t$$

o que corresponde à modelização em espaço de estados de um ARMA (V. a construção das matrizes em Rantala, 1985, pp.22-3). Como o vector x_t pode ser aumentado de forma a integrar margens de solvência em períodos não conhecidos, e como a equação de medida pode reflectir apenas o conhecimento da última margem calculada, o Filtro de Kalman dará directamente as estimativas dos valores de u_t necessários à fixação da política óptima.

Aplicando a teoria que deduz a política óptima como função linear de \hat{x}_t (Rantala, 1986, p.23) encontra-se depois expressão directa para os valores de $\Delta^d p_t$ que minimizam (11.27).

Passando à análise de intervenção refira-se um trabalho de Harvey e Durbin (1986) destinado a estudar o efeito da legislação sobre cintos de segurança nos acidentes rodoviários. Em lugar de seguir a típica análise de intervenção de Box-Tiao com modelos ARIMA ⁽⁸⁾, Harvey e Durbin formulam um modelo estrutural em espaço de estados, incluindo uma

⁽⁸⁾ Como é feita, por exemplo, por Abraham (1987, pp.211-9) no estudo do mesmo problema nas estradas de Ontário.

variável artificial u_t na equação de medida e acrescentando ao vector de estado um componente λ_t , componente destinado a medir a influência da legislação no número de acidentes observados: y_t

Significa isso que o modelo será expresso na forma:

$$y_t = [1 \quad 1 \quad u_t] \begin{bmatrix} T_t \\ S_t \\ \lambda_t \end{bmatrix} + v_t \quad (11.29')$$

$$\begin{bmatrix} T_{t+1} \\ S_{t+1} \\ \lambda_{t+1} \end{bmatrix} = F \begin{bmatrix} T_t \\ S_t \\ \lambda_t \end{bmatrix} + w_t \quad (11.29'')$$

A variável artificial será nula antes da saída da legislação e unitária a partir desse momento t_1 .

$$u_t = \begin{cases} 0 & t < t_1 \\ 1 & t \geq t_1 \end{cases}$$

A estimação de λ_t está dependente dos valores assumidos para as variâncias de v_t e do ruído da equação de transição, w_{3t} . Harvey e Durbin consideram particularmente interessante a possibilidade de o Filtro fornecer uma análise de sensibilidade do valor de $\hat{\lambda}_t$, $t \geq t_1$, a variações do *ratio* das variâncias $q = \sigma_{w_3}^2 / \sigma_v^2$. Essa análise é tão mais importante quanto a estimativa de $\sigma_{w_3}^2$ é normalmente dificultada pelo reduzido número de observações posteriores a t_1 .

Passe-se agora à teoria da amostragem, terreno em que a modelização de Kalman parece também poder fornecer algumas soluções. Cite-se S. M. Tam (1987), que estudou o célebre problema das amostragens repetidas utilizando os modelos de espaço de estados.

Tam estuda a estimação do total de uma característica y presente numa população que é evolutiva com o tempo. Num dado momento t , o valor dessa característica em cada elemento i será y_{it} e o total será

$$T_t = \sum_{i=1}^{N_t} y_{it}$$

Formando o vector $Y_t = [y_{1t} \quad y_{2t} \quad \dots \quad y_{N_t,t}]^T$ assume-se que o valor de cada seu componente está relacionado com o de outras características conhecidas ou directamente mensuráveis, x_{ij} , agrupadas na matriz X_t .

A relação de X_t com os valores inquiridos ou inquiríveis, y_{it} , é dada por um vector de coeficientes β_t , através da equação

$$Y_t = X_t \beta_t + v_t \quad (11.30')$$

com

$$\beta_t = T \beta_{t-1} + w_t \quad (11.30'')$$

A novidade está em que não existem observações de todo o vector Y_t mas apenas a da parcela amostrada. Subscrevendo com s a população inquirida e com r o resto da população procede-se à partição seguinte

$$\begin{bmatrix} Y_{ts} \\ Y_{tr} \end{bmatrix} = \begin{bmatrix} X_{ts} \\ X_{tr} \end{bmatrix} \beta_t + \begin{bmatrix} v_{ts} \\ v_{tr} \end{bmatrix} \quad (11.31)$$

e o problema consiste em encontrar estimativas para o estado β_t e para o total T_t após amostragens sucessivas.

Como se assume conhecida toda a matriz X_t o problema pode ser resolvido em dois passos.

Com as observações $Y_{1s}, Y_{2s}, \dots, Y_{ts}$ estimam-se os vectores $\beta_1, \beta_2, \dots, \beta_t$ pela aplicação da rotina de Kalman ao primeiro bloco de (11.31). Em seguida estima-se Y_{tr} , tendo em atenção as covariâncias $E v_{ts} v_{tr}$ e $E w_t v_{tr}$ (V. Tam, 1987, pp.67-8). A estimativa \hat{T}_t é então directa.

Com este método, conclui Tam, é possível ultrapassar o principal obstáculo à aplicação de modelos das sucessões cronológicas em amostragens repetidas. Esse obstáculo reside habitualmente no pequeno número de inquéritos realizados, mas isso não impede a aplicação do Filtro de Kalman, que pode funcionar com reduzidas observações.

Passando a aplicações na área da biologia refira-se o recente trabalho de Meinhold e Singpurwalla (1987) sobre a estimação de efeitos num modelo dose-resposta.

Nesses modelos regista-se a introdução de um estímulo ou dose x_j , com um efeito $y(x_j)$ na população ou organismo biológico em estudo. Aplicações ao universo económico são óbvias. Considere-se, por exemplo, que x_j e $y(x_j)$ quantificam, respectivamente, uma medida de política e o seu efeito macro-económico.

Meinhold e Singpurwalla começam por definir uma função para a resposta esperada $Ey(x) = \exp(-\alpha x^\beta)$ e tratam como observações $y^* := \log(-\log y)$.

Ter-se-á como equação de medida

$$y_j^* = [1 \quad \log x_j] \begin{bmatrix} \log x_j \\ \beta_j \end{bmatrix} + v_j$$

Representando o estado pelo vector $\theta_j = [\log x_j \quad \beta_j]$ encontra-se o sistema

$$\begin{aligned} y_j^* &= H_j \theta_j + v_j \\ \theta_{j+1} &= \theta_j + w_j \end{aligned}$$

A equação de transição é um "passeio aleatório", com $F \equiv I$, pois se admite uma constância (em média) dos parâmetros da função resposta.

O índice j não representa agora o tempo e a totalidade das observações está simultaneamente disponível, pelo que Meinhold e Singpurwalla propõem um *alisamento em dois passos* (correspondente ao "forward-backward smoothing" na terminologia de Brown, 1983, p.276).

Por esse método, estando ordenadas as observações $j = 1, 2, \dots, N$, proceder-se-á primeiro a uma filtragem pelo algoritmo de Kalman, obtendo-se, sucessivamente,

$$\hat{\theta}_{1|1}, \hat{\theta}_{2|2}, \dots, \hat{\theta}_{N|N}$$

Efectuar-se-á depois um alisamento, ou filtragem em sentido inverso, utilizando $\hat{\theta}_{N|N}$ como estimativa inicial e obtendo

$$\hat{\theta}_{N|N}, \hat{\theta}_{N-1|N}, \dots, \hat{\theta}_{1|N}$$

É o último valor, que se designou por $\hat{\theta}_{1|N}$ para mostrar que se trata de uma estimativa de $\hat{\theta}_1$ efectuada com o conhecimento das N observações, o valor que se retém e que fornece uma proposta para os parâmetros α e β da função $y(x) = \exp(-\alpha x^\beta) + \varepsilon$.

Conclusões

O modelo de espaço de estados e o algoritmo de Kalman, uma vez descobertos pela econometria estatística, vieram a revelar-se como instrumentos teóricos e práticos de grande utilidade para a modelização, estimação e previsão estocástica.

No terreno específico que constitui o objecto deste trabalho — a previsão econométrica de curto prazo — a teoria de Kalman tem vindo a relevar-se como particularmente promissora.

Por um lado, o sistema linear dinâmico revela-se apto a modelizar fenómenos económicos e sociais em que a variabilidade dos parâmetros e a aleatoriedade da evolução são constantes. Por outro lado, o Filtro de Kalman mostra-se como processo de estimação eficiente, que enquadra estimadores econométricos, de alisamento e bayesianos como seus casos particulares, e que é capaz de competir em precisão com os outros processos práticos.

Alguns problemas, de entre os muitos que ficaram por analisar, constituem linhas de investigação de grande actualidade.

Em primeiro lugar cite-se a teoria dos sistemas na sua relação com o problema da identificação econométrica.

Partindo de perspectivas radicalmente diversas os dois desenvolvimentos teóricos entroncam-se, parecendo promissor o estudo da teoria sistémica pelas contribuições que fornece à definição e à análise da identificabilidade dos modelos.

Em segundo lugar refira-se a estimação estatística daqueles parâmetros que são aprioristicamente definidos no sistema de Kalman.

Por métodos que parecem ter raiz comum na máxima verosimilhança caminha-se hoje para a estimação das matrizes de covariâncias e dos parâmetros iniciais do modelo. É admissível que uma solução completa do problema, que actualmente parece estar a ser esboçada⁽¹⁾, venha a introduzir nova generalidade no Filtro de Kalman e a resolver muitos dos óbices práticos à sua implementação.

Em terceiro lugar cite-se o problema da selecção ou identificação de modelos.

Trata-se de problema essencial e de certa forma prévio à estimação estocástica. No caso específico da previsão o problema assume contornos diferenciados, pois a selecção do modelo é feita em função dos resultados esperados, e deverá ter em conta a evolução possível dos processos aleatórios. Critérios como os de Akaike e seus desenvolvimentos, intimamente

(1) V. Diebold e Pauly, 1987, p.38.

relacionados com o modelo de espaço de estados, constituem outra linha de investigação para o confronto e escolha de modelos dinâmicos.

Finalmente, refira-se o problema da combinação óptima de previsões.

Usualmente referida à melhoria da precisão preditiva, a metodologia da combinação de modelos tem fundamento teórico na existência de diversos esquemas explicativos parcelares. A previsão combinada é passível de uma interpretação paralela à da escolha de modelos: trata-se de seleccionar aquela combinação de sistemas explicativos que constitui modelo preditivo óptimo.

Dê entre os diversos pontos em aberto é este último que se programa como constituindo objecto de investigação futura.



Referências

A bibliografia aqui incluída compreende unicamente os trabalhos directamente consultados. Os livros e artigos conhecidos por interposta fonte encontram-se completamente referenciados no corpo do trabalho.

A lista está dividida em três partes, que correspondem às três grandes subdivisões do texto. Na segunda e terceira partes não se referem os trabalhos anteriormente citados, mesmo que eles sejam mencionados no texto respectivo. Assim, esta lista corresponde a um "processo de inovações", não se repetindo a descrição de trabalhos já referidos e permitindo o agrupamento das referências de acordo com as necessidades novas de cada parte do trabalho.

primeira parte

operadores de previsão e projectores ortogonais

ANDERSON, T.W. (1958). *An Introduction to Multivariate Statistical Methods*, Wiley.

ANDERSON, T.W. (1971). *The Statistical Analysis of Time Series*, Wiley.

ANDERSON, Brian D.O. e MOORE, John B. (1979). *Optimal Filtering*, Information and System Science Series, Prentice-Hall.

BERBERIAN, Sterling K. (1976). *Introduction to Hilbert Space*, (2nd Ed) Chelsea Publishing Company, New York.

BERGER, James O. (1985). *Statistical Decision Theory and Bayesian Analysis*, (2nd Ed.), Springer Series in Statistics, Springer-Verlag.

BOREL, Émile e DELTHEIL, Robert (1923). *Probabilités Erreurs*, Librairie Armand Colin, Paris.

DOOB, J.L. (1953). *Stochastic Processes*, Wiley.

EISENHART, Churchil (1961). "Boscovich and the combination of observations", in White, L.L., Ed. (1961), *Roger Joseph Boscovich*, Allen & Unwin, London, pp.200-212.

GRANGER, Clive W.J. (1969). "Prediction with a generalized cost of error function", *Op. Research Quart.*, 20-2, pp.199-207.

HALMOS, Paul R. (1957). *Introduction to Hilbert Space and the Theory of Spectral Multiplicity*, (2nd Ed), Chelsea Publishing Company, New York.

JAZWINSKI, Andrew H. (1970). *Stochastic Processes and Filtering Theory*, Mathematics in Science and Engineering, Academic Press.

- KALMAN, R.E. (1960). "A new approach to linear filtering and prediction problems", *Trans. ASME, Series D; Journal of Basic Engineering*, **82**, pp.35-45.
- KALMAN, R.E. (1963). "New methods in Wiener filtering Theory", in, BOGDANOFF, John L. e KOZIN, Frank, Eds., *Proceedings of the First Symposium on Engineering Applications of Random Function Theory*, Wiley.
- KARLIN, Samuel e TAYLOR, Howard M. (1975). *A First Course in Stochastic Processes*, (2nd Ed), Academic Press.
- KENDALL, M.G. (1973). *Time-Series*, Griffin, London.
- KENDALL, Maurice e STUART, Alan (1987). *The Advanced Theory of Statistics*, vol I, *Distribution Theory*, (5th Ed), Griffin, London.
- KOROLIUK, V. et al. (1983). *Aide-Mémoire de Théorie des Probabilités et de Statistique Mathématique*, tradução francesa da edição russa de 1978, Ed. Mir, Moscovo.
- LOÈVE, Michel (1963). *Probability Theory*, (3rd Ed.), Van Nostrand Reinhold Company, New York.
- MALINVAUD, E. (1980). *Statistical Methods of Econometrics*, (3rd Ed.), North-Holland.
- MURTEIRA, Bento (1979). *Probabilidades e Estatística*, (vol I), McGraw-Hill, Lisboa.
- MURTEIRA, Bento J. (1983). "Inferência e decisão estatística", ISE, Lisboa.
- PARZEN, Emanuel (1962). *Stochastic Processes*, Holden-Day, San Francisco.
- PLACKETT, R.L. (1972). "Studies in the History of probability and statistics. XXIX: The discovery of the method of least squares", *Biometrika*, **59-2**, pp.239-251.
- PRIESTLEY, M.B. (1981). *Spectral Analysis and Time Series*, Academic Press.
- ROHATGI, V.K. (1976). *An Introduction to Probability Theory and Mathematical Statistics*, Wiley.
- RUDIN, Walter (1973). *Functional Analysis*, Tata McGraw-Hill, New Delhi.
- RUDIN, Walter (1974). *Real and Complex Analysis*, (2nd Ed.), Tata McGraw-Hill, New Delhi.
- RUYMGAART, P.A. e Soong, T.T. (1985). *Mathematics of Kalman-Bucy Filtering*, Springer Series in Information Sciences, 14, Springer-Verlag.
- SORENSEN, H.W. (1970). "Least-squares estimation: from Gauss to Kalman", *IEEE Spectrum*, pp.63-68.

STRANG, Gilbert (1976). *Linear Algebra and its Applications*, Academic Press.

WIENER, Norbert (1949). *Extrapolation, Interpolation, and Smoothing of Stationary Time Series*, the MIT Press, Massachusetts, 1977 edição com o título, "Time series", the MIT Press, Massachusetts.

segunda parte

previsão com Filtro de Kalman e modelização em espaço de estados

ANSLEY, Craig F. (1979). "An algorithm for the exact likelihood of a mixed autoregressive-moving average process", *Biometrika*, **66**, 1, pp.59-65.

AKAIKE, Hirotugu (1974). "Markovian representation of stochastic processes and its application to the analysis of autoregressive moving average processes", *Ann. Inst. Statist. Math*, **26**, pp.363-387.

AKAIKE, Hirotugu (1974). "Stochastic theory of minimal realization", *IEEE, Trans. Automatic Control*, **AC-19**, pp.667-674.

BOX, George E.P. e JENKINS, Gwilym M. (1970). *Time Series Analysis, Forecasting and Control*, Holden-Day, San Francisco.

BROCKWELL, P.J. e DAVIS, R.A. (1987). "Applications of innovation representation in time series analysis", in SRIVASTAVA, J.N., Ed., *Grayhill Festschrift Volume*, North-Holland.

BROWN, Robert Goodell (1963). *Smoothing, Forecasting and Prediction of Discrete Time Series*, Prentice-Hall.

BROWN, Robert Grover (1983). *Introduction to Random Signal Analysis and Kalman Filtering*, Wiley.

COOLEY, Thomas F. e PRESCOTT, Edward C. (1973). "An adaptive regression model", *Intern. Economic Review*, **14-2**, pp.364-371.

DUNCAN, B e HORN, S.D. (1972) "Linear Dynamic Estimation from the viewpoint of regression analysis", *J. of the American Statist. Ass*, **67**, **340**, pp.815-821.

GILCHRIST, Warren (1976). *Statistical Forecasting*, Wiley.

GRANGER, C.W.J. e NEWBOLD, Paul (1986). *Forecasting Economic Time Series*, (2nd Ed), Academic Press.

- HARRISON, P.J. e STEVENS, C.F. (1971). "A Bayesian approach to short-term forecasting", *Op. Res. Quartely*, **22**, pp.341-362.
- HARRISON, P.J. e STEVENS, C.F. (1976). "Bayesian forecasting", *J.R. Statist. Soc., B*, **38**, pp.205-247.
- HARVEY, A.C. (1981a). *Time Series Models*, Philip Allan, London.
- HARVEY, A.C. (1981b). *The Econometric Analysis of Time Series*, Philip Allan, London.
- HAWKES, R.M. e MOORE, J.B. (1976). "Perfomance of Bayesian parameter estimations for linear signal models", *IEEE Trans. on Automatic Control*, pp.523-527.
- JOHNSTON, J. (1984). *Econometric Methods* (3rd Ed), McGraw-Hill.
- KAILATH, Thomas (1968). "An innovations approach to least-squares estimation — Part I: Linear Filtering in additive white noise", *IEEE Trans. Automatic Control*, **AC-13**, **6**, pp.646-654.
- KAILLATH, Thomas (1970). "The innovations approach to detection and estimation theory", *Proceedings of the IEEE*, **58**, **5**, pp.680-695.
- KALMAN, R.E. (1982). "Identifiability and problems of model selection in econometrics", *invited paper for the Fourth World Congress of the Econometrics Society*, in HILDENBRAND, Werner (Ed), *Advances in Econometrics*, Econometric Society Monographs, Cambridge University Press, Cambridge.
- KALMAN, R.E. e BUCY, R.S. (1961). "New results in linear filtering and prediction theory", *Trans. ASME, Series D, Journal of Basic Engineering*, **83**, pp.95-108.
- KALMAN, R.E., FALB, P.L. e ARBIB, M.A. (1969). *Topics in Mathematical System Theory*, McGraw-Hill.
- KNUTH, Donald (1973). *The Art of Computer Programming*, vol 1, *Fundamental Algorithms*, Addison-Wesley.
- MEHRA, Raman K. (1970). "On the identification of variances and adaptative Kalman filtering", *IEEE Trans. Automatic Control*, **AC-15**, **2**, pp.175-184.
- MEHRA, Raman K. (1971) "On line identification of linear dynamic systems with applications to Kalman filtering", *IEEE Trans. Automatic Control*, **AC-16**, **1**, pp.12-21.
- MEHRA, Raman K. (1972). "Approaches to adaptive filtering", *IEEE Trans. Automatic Control*, **AC-17**, pp.693-698.

- MEINHOLD, Richard J. e SINGPURWALLA, Nozer D. (1983). "Understanding the Kalman filter", *The American Statistician*, **37**, 2, pp.123-127.
- MONTGOMERY, Douglas C. e JOHNSON, Lynwood, A. (1976). *Forecasting and Time Series Analysis*, McGraw-Hill.
- MUTH, John F. (1960). "Optimal properties of exponentially weighted forecasts", *J. of the American Statist. Ass.*, **55**, pp.299-306.
- NEVES, João César das (1984). "Filtro de Kalman: uma apresentação sumária", *Economia*, **VIII**, 3, pp.527-542.
- OTTER, Pieter W. (1981). "Identification and estimation of discrete state-vector models with stochastic inputs", *Automatica*, **17-2**, pp.389-391.
- OTTER, P.W. e TEMPELAAR, D.T. (1980). "Kalman and Box-Jenkins estimators in a simulation study", in ANDERSON, O.D. (Ed), *Analysing Time Series*, North-Holland.
- PLACKETT, R.L. (1950). "Some theorems in least squares", *Biometrika*, **37**, pp.149-157.
- POPPER, Karl R. (1980). *The Logic of Scientific Discovery*, (10th imp., rev), Hutchinson, London.
- SAGE, Andrew P. e MELSA, James L. (1971). *An Introduction to Probability and Stochastic Processes*, Prentice Hall, New Jersey, tradução portuguesa de 1983, *Introdução às Probabilidades e Processos Estocásticos*, Editora Portuguesa de Livros Científicos e Técnicos, Lisboa.
- SHADER, Martin e SCHMID, Friedrich (1985). "Computation of M.L. estimates for the parameters of a negative binomial distribution from grouped data. A comparison of the scoring, Newton-Raphson and E-M algorithms", *Applied Stochastic Models and Data Analysis*, **1**, pp.11-23.
- THEIL, H. e GOLDBERGER, A.S. (1961). "On pure and mixed statistical estimation in economics", *International Economic Review*, **2-1**, pp.65-78.
- WITTEN, Ian H. e CLEARY, John G. (1986). "Foretelling the future by adaptive modeling", *Abacus*, **3-3**, pp.16-36.
- YOUNG, P.C. (1975). "Coments on a paper by Brown, Durbin and Evans", *J.R. Statist. Soc.*, B, **37**, pp.168-74.

terceira parte

problemas recentes

ABRAHAM, Bovas (1987). "Application of intervention analysis to a road fatality series in Ontario", *J. of Forecasting*, 6, pp.211-219.

AMEEN, J.R.M. e HARRISON, P.J. (1984). "Discount weighted estimation", *J. of Forecasting*, 3, pp.285-296.

AMEEN, J.R.M. e HARRISON, P.J. (1985). "Normal discount Bayesian models", in BERNARDO, J.M. et al. (Eds), *Bayesian Statistics 2*, pp.271-298, North-Holland.

ANSLEY, Craig F. e KOHN, Robert (1985). "Estimation, filtering and smoothing in state-space models with incompletely specified inicial conditions", *The Ann. of Statist.*, 13-4, pp.1286-1316.

BLACK, George H.J. e MURTEIRA, Bento J.F. (1983). "Inflação em 1983: ensaio de previsão não causal", *Estudos de Economia*, III, 3, pp.267-293.

BOLSTAD, William M. (1986). "Harrison-Stevens forecasting and the multiprocess dynamic linear model", *The American Statistician*, 40-2, pp.129-135.

BROCKWELL, Peter J. e DAVIS, Richard A. (1987). *Time Series: Theory and Methods*, Springer Series in Statistics, Springer-Verlag.

BROWN, R.L., DURBIN, J. e EVANS, J.M. (1975). "Techniques for testing the constancy of regression relationships over time", *J.R. Statist. Soc.*, B, 37, pp.149-163.

CRATO, Nuno (1986). "Aplicação de um método de previsão bayesiana ao IPC", CEMAPRE, Documento de trabalho n°25, ISE.

DIDERRICH, George T. (1985). "The Kalman filter from the perspective of Goldeberger-Theil estimators", *The American Statistician*, 39-3, pp.193-198.

DIEBOLD, F.X. e PAULY, P. (1987). "Structural changes and combination of forecasts", *J. of Forecasting*, 6, pp.21-40.

DURBIN, J. (1984). "Present position and potential developments: some personal views — Time series analysis", *J.R. Statist. Soc.*, A, 147-2, pp.161-173.

GUNEL, Ipek (1987). "Forecasting system energy demand", *J. of Forecasting*, 6, pp.137-156.

HANAN, E.J. (1986). "Time series and stochastic models", in Bittanti, Sergio, Ed, *Time Series and Linear Sytems*, Springer-Verlag.

- HARVEY, A.C. (1984). "A unified view of statistical forecasting procedures", *J. of Forecasting*, **3**, pp.245-275.
- HARVEY, A.C. e DURBIN, J. (1986). "The effects of seat belt legislation on British road casualties: a case study with structural time series modelling", *J.R. Statist. Soc., A*, **149**, pp.187-227.
- JONES, R.H. (1985). "Time series analysis with unequally spaced data", in HANNAN, E.J., KRISHNAIAH, P.R. e RAO, M.M., Eds, *Time Series in the Time Domain*, Handbook of statistics 5, pp.157-177, North-Holland.
- KENDALL, Maurice e STUART, Alan (1983). *The Advanced Theory of Statistics*, vol III, *Design and Analysis, and Time-Series*, Griffin, London.
- KOHN, Robert e ANSLEY, Craig, F. (1986). "Estimative, prediction and interpolation for ARIMA models with missing data", *J. of the American Statist. Ass*, **81**, **395**, pp.751-761.
- KRZYSZTOFOWICZ, Roman (1987). "Markovian forecast processes", *J. of the American Statist. Ass*, **82**, pp.31-37.
- LOPES, J.A. Assis (1985). "Metodologia integrada para previsão em mercados de preços flutuantes", (Tese de doutoramento), IST, Lisboa.
- MEADE, Nigel (1985). "Statistical forecasting" in RAND, G.K. e EGGLESE, R.W., Eds, *Further Developments in Operational Research*, Pergamon, pp.57-75.
- MEHRA, Raman K. (1979). "Kalman filters and their applications to forecasting", in MAKRIDAKIS, S. e WHEELWRIGHT, S.C., Eds, *Forecasting*, TIMS studies in the Management Sciences, **12**, pp.75-94, North-Holland.
- MEINHOLD, Richard J. e SINGPURWALLA, Nozer D. (1987). "A Kalman-Filter smothing approach for extrapolations in certain dose-response, damage-assessment, and accelerate-life-testing studies", *The American Statistician*, **41-2**, pp.101-106.
- MORRISON, G.W. e PIKE, D.M. (1977). "Kalman filtering applied to statistical forecasting", *Manag. Sc.*, **23**, pp.768-774.
- MOYER, Reed (1984). "The futility of forecasting", *Long Range Planning*, **17-1**, pp.65-72.
- MURTEIRA, Bento J.F. (1953). *Algumas Propriedades dos Processos Autoregressivos*, Lisboa.
- NELSON, Robert (1987). "State-space modelling of residential commercial and peack demands", *J. of Forecasting*, **6**, pp.97-115.

NICHOLLS, D.F. e PAGAN, A.R. (1985). "Varying coefficient regression", in HANNAN, E.J., KRISHNAIAH, P.R. e RAO, M.M., Eds, *Time Series in Time Domain*, Handbook of statistics 5, pp.413-455, North-Holland.

OTTER, Pieter W. (1985). *Dynamic Feature Space Modelling, Filtering and Self-Tuning Control of Stochastic Systems*, Lecture Notes in Economics and Mathematical Systems, 246, Springer-Verlag.

PARZEN, Emanuel (1986). "Discussion of a paper by professors Harvey and Durbin", *J.R. Statist. Soc., A*, 149, 3, pp.223-224.

POPPER, Karl R. (1983). *Realism and the Aim of Science*, Hutchinson, London [O Realismo e o Objectivo da Ciência, tradução portuguesa de 1987, publicada por D. Quixote, Lisboa].

RANTALLA, Jukka (1986). "Experience rating of ARIMA processes by the Kalman filter", *Astin Bull.*, 16-1, pp.19-31.

SZELAG, C.R. (1982). "A short-term forecasting algorithm for trunk demand service", *the Bell System Thec. J.*, 61-1, pp.67-96.

TAM, S.M. (1987). "Analysis of repeated surveys using a dynamic linear model", *Intern. Statist. Review*, 55-1, pp.63-73.

TOMÉ, F.M. e CUNHA, J.A. (1985). "Traffic forecasting with a state-space model", in *Proceedings of the Eleventh International Teletraffic Congress — September 4-11, 1985, Kyoto*, North Holland, pp.3.4b 5.1-3.4b 5.5.

WATSON, Mark W. (1987). "Uncertainty in model-based seasonal adjustment procedures and construction of minimax filters", *J. of the American Statist. Ass.*, 82, pp.395-408.

WATSON, Mark W., PASTUSZEK, Lydia M. e CODY, Eric (1987). "Forecasting commercial electricity sales", *J. of Forecasting*, 6, pp.117-136.

WELCH, Michael E. (1987). "A Kalman filtering perspective", *The American Statist.*, 41-1, pp.90-91.

WEST, Mike e HARRISON, P. Jeff (1986). "Monitoring and adaptation in Bayesian forecasting models", *J. of the American Statist. Ass.*, 81, 395, pp.741-750.

WEST, Mike e HARRISON, P. Jeff e MIGON, Helio S. (1985). "Dynamic generalized linear models and Bayesian forecasting", *J. of the American Statist. Ass.*, 80, 389, pp.73-83.

YOUNG, P. (1985). "Recursive identification, estimation and control", in HANNAN, E.J., KRISHNAIAH, P.R. e RAO, M.M., Eds, *Time Series in the Time Domain*, Handbook of statistics 5, pp.213-255, North-Holland.

ZEHNWIRTH, Ben (1985). "Linear filtering and recursive credibility estimation", *Astin Bull*, 15-1, pp. 19-35.

NOTAS ANEXAS

NOTA 1. Processos estocásticos e sucessões cronológicas

O estudo das sucessões cronológicas incide sobre dados observados de processos aleatórios que se desenvolvem no tempo. Processos aleatórios, processos estocásticos ou funções aleatórias são termos sinónimos (*), que referem famílias de variáveis aleatórias indexadas por um parâmetro que pode ou não representar o tempo.

Uma primeira definição pode ser assim estabelecida:

PROCESSO ESTOCASTICO $\{x(t)\}$ é uma família de variáveis aleatórias indexadas pelo símbolo t , em que t pertence a um conjunto de índices determinado, T , infinito, numerável ou não numerável.

Para cada dado valor t , seja t_0 , $x(t_0)$ é uma variável aleatória, razão por que o processo estocástico pode ser definido como uma correspondência $t \rightarrow x(t)$, dando um sentido mais evidente à expressão "função aleatória" (Cf. Ventsel, 1973, pp.354-60).

Um dado registo observado de $x(t)$ quando t percorre T constitui uma REALIZAÇÃO, TRAJECTORIA ou FUNÇÃO AMOSTRA do processo a qual, desse ponto de vista, pode ser encarada como função determinística de t . A colecção de todas as possíveis realizações é denominada CONJUNTO do processo ou ESPAÇO DAS REALIZAÇÕES.

(*) V. contra-argumentação terminológica em Loève 1963, pp.497-8.

Ter-se-á presente que, sendo $x(t)$ variável aleatória, existe um espaço de probabilidade subjacente. Considerar-se-á um espaço de acontecimentos Ω , com elementos w , a σ -álgebra de subconjuntos \mathcal{F} de Ω e a medida de probabilidade P . Dado o espaço de probabilidade (Ω, \mathcal{F}, P) e um conjunto arbitrário T , um processo estocástico é uma função real e finita, $x(t, w)$, definida no produto cartesiano $T \times \Omega$, função mensurável de w para cada t fixo, $t \in T$. Por simplicidade omitir-se-á sempre que possível o argumento w e escrever-se-á $x(t)$ em lugar de $x(t, w)$ assim como se denotará o processo por $\{x(t), t \in T\}$ ou mesmo por $\{x(t)\}$.

Definição formal de processo estocástico pode ser assim estabelecida

Dado um espaço de probabilidade (Ω, \mathcal{F}, P) e um conjunto $T \subset \mathbb{R}$, PROCESSO ESTOCÁSTICO é a função $x(w, t): \Omega \times T \rightarrow \mathbb{R}$, tal que, para cada t fixo, a função parcial $x(w, \cdot)$ é uma variável aleatória.

Os três elementos que fundamentalmente distinguem um processo estocástico são o espaço de estados, o conjunto de parâmetros e as relações entre as variáveis aleatórias $x(t)$.

O ESPAÇO DE ESTADOS é o conjunto de todos os possíveis valores de $x(t)$. Com $\{x(t)\}$ definido em \mathbb{R} trata-se de um processo de valores reais. Se a variável aleatória $x(t)$ for um vector de dimensão k trata-se de processo k -dimensional. Se $x(t)$ apenas puder assumir valores num conjunto discreto diz-se que o espaço de estados do processo é discreto.

O ESPAÇO DE PARAMETROS ou de INDICES é o conjunto T , conjunto

que pode também ser discreto ou contínuo.

Usualmente, um processo denominar-se-á discreto ou contínuo conforme a natureza do espaço dos parâmetros. Os casos contínuos mais frequentes surgem com $T=]-\infty, +\infty[$ ou $T=]0, +\infty[$ e os casos discretos típicos são $T=\{\dots, -2, -1, 0, +1, +2, \dots\}$ ou $T=\{0, 1, 2, \dots\}$.

No caso discreto é habitual representar a variável aleatória por x_t , convenção que será aqui adoptada.

A designação de SUCESSÃO CRONOLÓGICA será reservada para processos estocásticos em que o índice t representa o tempo e em que esse índice apenas pode tomar valores discretos.

Uma sucessão cronológica pode surgir por uma de duas vias. Haverá processos que, pela sua própria natureza, traduzem acontecimentos que tomam lugar em momentos discretos. Noutros casos os acontecimentos evoluem continuamente e é a sua medida e registo que podem ser tomados em momentos discretos, numa amostragem da realização do processo.

A caracterização de processo real constitui problema de escala de abordagem, pois é natural que processos considerados discretos possam ser estudados como contínuos quando observados com maior pormenorização, assim como é possível que processos contínuos apenas o sejam até certo grau de ampliação da realidade que lhes está subjacente. O problema é de natureza epistemológica. A questão técnica que importa sublinhar, e que em certa medida tranquiliza sobre a restrição do estudo a processos de índice discreto, é que a amostragem de um processo contínuo pode, em condições muito gerais, ser suficiente para a sua caracterização

(V. Doob, 1953, pp.50 e segs). Por outro lado, a digitalização comporta vantagens computacionais actualmente bem evidentes, conduzindo à generalização prática da abordagem discreta.

NOTA 2. Convergência em média quadrática

Considere-se o espaço de probabilidade (Ω, \mathcal{F}, P) e a sucessão (não finita) de v.a. de segunda ordem aí definidas $\{x_n\}_{n \in \mathbb{N}}$.

A sucessão dir-se-á CONVERGENTE EM MEDIA QUADRÁTICA, ou, simplesmente, CONVERGENTE, com limite $x \in S$ se a distância em média quadrática tender para zero quando $n \rightarrow \infty$, i.e.

$$(2.1) \quad x_n \rightarrow x \iff \|x_n - x\| \rightarrow 0$$

O que significa que em S se pode tomar, como definição de convergência da sucessão $\{x_n\}$,

$$(2.2) \quad \lim_{n \rightarrow \infty} E[x_n - x]^2 = 0$$

A convergência em média quadrática denotar-se-á simplificada por referência à norma, como em (2.1), ou pelo símbolo

$$\text{l.i.m. } x_n = x$$

em que l.i.m. representa, como habitualmente, "limit in mean" (V. e.g. Apostol, 1974, p. 232), com $n \rightarrow \infty$. Neste caso, e também de acordo com a terminologia consagrada, a média entende-se como média quadrática e a medida de integração como medida de probabilidade (V. e.g. Ruimgaart e Soong, 1985, p. 31).

Poder-se-ão agora referir algumas propriedades importantes de sucessões de v.a. convergentes em $L^2(\Omega)$.

UNICIDADE DO LIMITE

Admita-se a existência de dois limites, x e x' , para a sucessão $\{x_n\}$ de v.a. de $L^2(\Omega)$.

Pela desigualdade triangular da norma ter-se-á

$$\|x - x'\| \leq \|x_n - x\| + \|x_n - x'\|$$

tomando limites encontra-se $\|x - x'\| \leq 0$, o que significa, pela propriedade (iii) da distância, que $x = x'$ q.c. \square .

Notando ainda que $E\|x - x'\|^2 = \|x - x'\|^2 = 0$ e, aplicando a desigualdade de Chebishev

$$P\{\|x - x'\| > \delta\} \leq E\|x - x'\|^2 / \delta^2, \quad \delta > 0$$

conclui-se que $P\{\|x - x'\| > \delta\} = 0$, para todo o δ positivo e que o limite é único no sentido

$$P\{x = x'\} = 1$$

CONVERGENCIA EM PROBABILIDADE

Aplique-se novamente a desigualdade de Chebishev

$$P\{\|x_n - x\| > \delta\} \leq E\|x_n - x\|^2 / \delta^2, \quad \delta > 0$$

Uma vez que $\text{l.i.m. } x_n = x \Leftrightarrow E\|x_n - x\|^2 \rightarrow 0$, conclui-se que

$$(2.3) \quad \lim P\{|x_n - x| > \delta\} = 0 \quad \text{para todo } \delta > 0$$

o que significa que a convergência em média quadrática implica a convergência em probabilidade (2.3) \square .

CONVERGENCIA DA MEDIA

Aplicando (4.8, Nota 4) à v.a. $x_n - x$ encontra-se

$$\begin{aligned} |E[x_n - x]| &\leq \|x_n - x\| \\ |Ex_n - Ex| &\leq \|x_n - x\| \rightarrow 0 \end{aligned}$$

Assim, se a sucessão de v.a. $\{x_n\}$ converge para a v.a. x , também a média da sucessão converge para Ex

$$(2.4) \quad Ex_n \rightarrow Ex \quad \square.$$

CONVERGENCIA DOS SEGUNDOS MOMENTOS

Observe-se que a desigualdade triangular pode ser escrita na forma (*)

$$(2.5) \quad \left| \|x\| - \|y\| \right| \leq \|x - y\|$$

 (*) Escreva-se $\|y + (x - y)\| \leq \|x\| + \|x - y\|$

Obtem-se $\|x\| - \|y\| \leq \|x - y\|$. Multiplicando esta última desigualdade por -1 e notando que $\|x - y\| = \|y - x\|$ obtem-se (2.5).

Então, se $\{x_n\}$ converge para x ,

$$| \|x\| - \|x_n\| | \leq \|x - x_n\| \rightarrow 0$$

pelo que

$$(2.6) \quad \lim \|x_n\| = \|x\| \quad \text{ou seja,} \quad \lim E x_n^2 = E x^2 \quad \square.$$

CONVERGENCIA DAS COVARIANCIAS

Demonstrada a convergência das médias basta demonstrar a convergência dos momentos cruzados para garantir a convergência das covariâncias.

Prove-se então que $\|x_n - x\| \rightarrow 0$ e $\|y_n - y\| \rightarrow 0$ implica:

$$(2.7) \quad | E[x_n y_m] - E[xy] | \rightarrow 0$$

Note-se primeiramente que

$$\begin{aligned} | E[x_n y_m] - E[xy] | &= | E[xy] - E[x_n y] + E[x_n y] - E[x_n y_m] | \leq \\ &\leq | E[xy] - E[x_n y] | + | E[x_n y] - E[x_n y_m] | = \\ &= | E[(x - x_n) \cdot y] | + | E[x_n \cdot (y - y_m)] |. \end{aligned}$$

Aplique-se a desigualdade de Cauchy-Schwarz às duas parcelas encontradas. Conclui-se que

$$| E[x_n y_m] - E[xy] | \leq \|x - x_n\| \cdot \|y\| + \|x_n\| \cdot \|y - y_m\| \rightarrow 0 \quad \square.$$

0. ESPAÇO DOS LIMITES E SUB-ESPAÇO DE $L^2(\Omega)$

Se a sucessão $\{x_n\}$ converge para x isso implica, nomeadamente a partir de certa ordem, que $\|x_n - x\|$ represente uma distância finita.

Aplicando a desigualdade (2.5) encontra-se

$$\left| \|x_n\| - \|x\| \right| \leq \|x_n - x\| < +\infty$$

Como $\|x_n\|$ é finito isso obriga a que $\|x\|$ seja finito, ou seja, $x \in L^2(\Omega)$.

NOTA 3. Funções de perda não pares

Não foi possível o acesso directo ao trabalho de Sherman referido por Kalman (1960) e Jazwinski (1970):

S. Sherman (1958). "Non-mean-square error criteria",
Trans. IRE Prof. Group Inform. Theory, IT-4, pp.125-6.

A versão do teorema de Kalman reproduzida no texto é a apresentada por Kalman e refere-se a funções de perda pares. Jazwinski afirma, sem o demonstrar, que o teorema é aplicável a funções de perda sem esse requerimento (pp. 147-9).

Seria interessante tal generalização mas ela não parece correcta. A análise do trabalho independente de Granger (1969) confirma tal impossibilidade.

Vai dar-se um contra-exemplo.

Admita-se que $p(x|Y)$ é simétrica, tendo como valor médio m . Definindo a estimativa x' e o erro $e=x-x'$ construa-se função compatível com a definição de Jazwinski:

$$\begin{aligned} L(e) > 0 & \text{ para } e > 0 & (x > x') \\ L(e) = 0 & \text{ para } e \leq 0 & (x \leq x') \end{aligned}$$

Com qualquer estimativa x' a perda esperada será

$$\int_{-\infty}^{+\infty} L(e)p(x|Y)dx = \int_{-\infty}^{x'} 0 \cdot p(x|Y)dx + \int_{x'}^{+\infty} L(x-x')p(x|Y)$$

Compare-se esta expressão genérica com a da perda esperada com estimativa centrada. Ter-se-á que $x' > m$ implica

$$\int_{x'}^{+\infty} L(x-x')p(x|Y)dx \leq \int_m^{x'} L(x-m)p(x|Y)$$

pois $x-x' < x-m \Rightarrow L(x-x') \leq L(x-m)$ e $p(x|Y) \geq 0$.

Bastaria aliás que a função perda fosse estritamente crescente para erro positivo e que $p(x|Y)$ fosse estritamente positiva nalgum subintervalo de $[m, x']$ para que a desigualdade da perda esperada fosse estrita, o que prova que o valor médio não obteria a perda esperada mínima.

NOTA 4. Variáveis aleatórias de segunda ordem

O espaço de variáveis aleatórias de segunda ordem e o operador E_x permitem definir um espaço de Hilbert que constitui suporte fundamental ao estudo de sucessões de v.a.

Seja (Ω, \mathcal{F}, P) um espaço de probabilidade e $x: \Omega \rightarrow R$ v.a., diz-se que x é VARIÁVEL ALEATORIA DE SEGUNDA ORDEM se $E x^2 < +\infty$.

Considere-se o espaço S de todas as v.a. de segunda ordem:

$$S = \{ x: \Omega \rightarrow R, x \text{ é mensurável, } E x^2 < +\infty \}$$

Trata-se de espaço vectorial sobre R :

(i) $x \in S, a \in R \Rightarrow ax \in S$
pois ax é mensurável e $E[ax]^2 = a^2 E x^2 < +\infty$

(ii) $x, y \in S \Rightarrow x+y \in S$
pois $x+y$ é mensurável e $0 \leq x^2 + y^2 - 2xy \Rightarrow 2xy \leq x^2 + y^2 \Rightarrow$
 $\Rightarrow (x+y)^2 \leq 2x^2 + 2y^2 \Rightarrow E[x+y]^2 \leq 2E x^2 + 2E y^2 < +\infty$.

Na última implicação note-se que a desigualdade, antes de operada pela esperança matemática, é interpretada no sentido de se verificar quase certamente.

Defina-se a aplicação de S^2 em R : $(x, y) \rightarrow E xy$.

Existe E_{xy} , pois xy é função mensurável e $-\infty < E_{xy} < +\infty$ como facilmente se verifica: $|xy| \leq \frac{1}{2}x^2 + \frac{1}{2}y^2 \Rightarrow E|xy| < \infty \Rightarrow |E_{xy}| < \infty$.

A aplicação é PRODUTO ESCALAR no espaço S :

$$(4.1) \quad x \cdot y = E_{xy}$$

respeitando as propriedades características

- (i) $E_{xy} = E_{yx}$
 - (ii) $E[(x+y) \cdot z] = E_{xz} + E_{yz}$
 - (iii) $E[(\alpha x) \cdot y] = \alpha \cdot E_{xy}$
 - (iv) $E_{x^2} \geq 0$; $E_{x^2} = 0 \Rightarrow x=0$ q.c.
- para todas as v.a. $x, y, z \in S$ e $\alpha \in \mathbb{R}$.

Note-se que a última propriedade é válida com excepção de um conjunto de medida nula, ou seja, $E_{x^2} = 0 \Leftrightarrow x(w) = 0$ q.c. , devendo ser interpretada em termos de classes de equivalência das v.a.

Com o espaço S munido de um produto escalar definem-se variáveis ORTOGONAIS como aquelas v.a. em que $E_{xy} = 0$.

Sendo x_1, x_2, \dots, x_n um conjunto de v.a. ortogonais entre si obtém-se expressão da lei de Pitágoras:

$$(4.2) \quad E(\sum x_i)^2 = \sum E_{x_i^2}$$

No espaço S pode igualmente definir-se uma NORMA:

$$(4.3) \quad \|x\| = (x \cdot x)^{1/2} = E^{1/2} x^2$$

que respeita as propriedades características da aplicação $\|\cdot\|: x \rightarrow R$

- (i) $\|x+y\| \leq \|x\| + \|y\|$
 - (ii) $\|\alpha \cdot x\| = |\alpha| \cdot \|x\|$
 - (iii) $\|x\| \geq 0$; $\|x\| = 0 \Leftrightarrow x=0$ q.c.
- para todas as v.a. $x, y \in S$ e $\alpha \in R$

As duas últimas propriedades deduzem-se directamente por aplicação do operador E ou a partir do próprio produto escalar, por tradução das propriedades (iii) e (iv) deste último.

Para dedução da desigualdade triangular

$$(4.4) \quad \|x+y\| \leq \|x\| + \|y\|$$

começar-se-á por demonstrar a desigualdade de Cauchy-Schwarz:

$$(4.5) \quad |Exy| \leq \|x\| \cdot \|y\|$$

Sabe-se que $E[\alpha x + y]^2 \geq 0$, ou seja, $Ex^2 \alpha^2 + 2Exy \alpha + Ey^2 \geq 0$. O discriminante desta forma quadrática em α é não positivo: $2^2 E^2 xy - 4Ex^2 Ey^2 \leq 0$, de onde se obtém

$$(4.6) \quad E^2 xy \leq Ex^2 Ey^2$$

o que é equivalente a (4.5) \square .

Por aplicação directa de (4.6) obtém-se

$$\begin{aligned}\|x+y\|^2 &= Ex^2 + 2Exy + Ey^2 \\ &\leq \|x\|^2 + 2\|x\| \cdot \|y\| + \|y\|^2 \\ &= (\|x\| + \|y\|)^2\end{aligned}$$

Sendo a norma uma quantidade não negativa encontra-se, por radiciação, a desigualdade triangular \square .

O espaço S , normado por $\|x\| = E^{1/2}x^2$, representar-se-á doravante por $L^2(\Omega)$ ou L^2 , de acordo com a simbologia generalizada (*).

Defina-se agora a DISTANCIA entre duas v.a. de segunda ordem $d(x,y): S^2 \rightarrow R$

$$(4.7) \quad d(x,y) = \|x-y\|$$

Sendo construída a partir da norma, a distância respeitará, obrigatoriamente, as usuais propriedades características.

- (i) $d(x,y) = d(y,x)$
- (ii) $d(x,y) \leq d(x,z) + d(z,y)$
- (iii) $d(x,y) = 0 \Leftrightarrow x = y$ q.c. $[d(x,y) \geq 0]$

(* Segue-se a notação usual (V. e.g. Rudin, 1974, pp. 66 e segs) para representar os espaços de funções mensuráveis de p-norma finita:

$$L^p(\mu) = \{f: \|f\|_p = [\int_{\Omega} |f|^p d\mu]^{1/p} < \infty\}$$

Na simbologia $L^2(\Omega)$ ou L^2 subentende-se estar aplicada como medida μ a medida de probabilidade.

A demonstração pode seguir o caminho usual. $\|x-y\| = |-1| \cdot \|y-x\|$, o que prova (i). Na desigualdade triangular (4.4), substituindo x por $x-z$ e y por $-y+z$ encontra-se (ii). Aplicando a terceira propriedade da norma à v.a. $x-y$ encontra-se (iii) \square .

Chama-se de novo a atenção para o facto de a propriedade (iii) dever ser interpretada no sentido de se verificar quase certamente. A relação $x=y$ poderá ser não válida num conjunto de medida nula e, no entanto, $d(x,y)$ ser nulo. O problema pode ser ultrapassado por uma de duas vias. Ou se considera que $E^{\frac{1}{2}}[x-y]^2$ induz uma semi-métrica (V. Apostol, 1974, p.295) mantendo (iii) no sentido de q.c., ou então se considera a igualdade entre variáveis aleatórias como a relação de equivalência $x=y$ sse $x(w)=y(w)$ com excepção de w num conjunto de medida nula.

Referir-se-ão de seguida alguns resultados fundamentais em variáveis aleatórias de segunda ordem no espaço $L^2(\Omega)$ — espaço S com a métrica, a norma e o produto escalar que atrás se introduziram.

V.A. DE ORDEM 2 E TAMBEM DE ORDEM 1

Aplicando a desigualdade de Schwarz verifica-se que $|Ex| = |E[i_{\Omega} \cdot x]| \leq \|i_{\Omega}\| \cdot \|x\| = \|x\| < \infty$. Então,

$$(4.8) \quad |Ex| \leq \|x\| .$$

Quer isto dizer que $|Ex| < \infty$, o que define variável aleatória de ordem 1 \square .

O ESPAÇO $L^2(\Omega)$ É FECHADO A OPERAÇÃO DE CENTRAGEM

A partir de $x \in L^2$ define-se a v.a. centrada $(x - Ex)$.

Que $(x - Ex) \in L^2$ pode comprovar-se notando que se trata de combinação linear de variáveis de $L^2(\Omega)$: $x = \int_{\Omega} x \cdot \mathbf{1}_{\Omega} \cdot Ex$.

NOTA 5. Variedades gaussianas e processos gaussianos

Seja (Ω, \mathcal{E}, P) um espaço de probabilidade e considerem-se as variáveis aleatórias $x_i: \Omega \rightarrow \mathbb{R}$, $i=1, 2, \dots, n$.

O vector $x = [x_1 \ x_2 \ \dots \ x_n]$ é GAUSSIANO n-DIMENSIONAL, ou MULTINORMAL se a função de densidade conjunta for definida por

$$(5.1) \quad f_x(x) = (2\pi)^{-n/2} \cdot |C|^{-1/2} \cdot \exp\{-\frac{1}{2}(x-m)C^{-1}(x-m)\}$$

em que $C = [c_{ij}]_n^n$ é a matriz de covariância de x e $m^T = [m_1 \ \dots \ m_n]$ o vector das esperanças matemáticas das variáveis aleatórias x_i . A função de densidade da multinormal fica pois completamente determinada pelos parâmetros C e m .

Note-se que, na definição acima, a matriz C não pode ser singular, tendo de ser definida positiva - possuir todos os valores próprios estritamente positivos. Preferir-se-á, por isso, definir uma variável aleatória gaussiana n-dimensional pelo recurso à sua função característica:

$$(5.2) \quad \tilde{g}(u) = \exp\{iu'm - \frac{1}{2}u'Cu\}$$

que admite a possibilidade de a matriz C ser apenas semi-definida positiva - caso em que existem valores próprios nulos - havendo componentes de x linearmente dependentes.

Admita-se o caso em que as covariâncias cruzadas são nulas, i.e.

$$c_{ij} = \text{Cov}[x_i, x_j] = 0, \quad i \neq j$$

a matriz C será diagonal e ter-se-á

$$\begin{aligned} \Phi(u) &= \exp\{i\sum_j u_j m_j - \frac{1}{2}\sum_j u_j^2 c_{jj}\} \\ &= \pi_1^n \exp\{i u_j m_j - \frac{1}{2} u_j^2 c_{jj}\} \\ &= \Phi_{x_1}(u_1) \cdot \Phi_{x_2}(u_2) \dots \Phi_{x_n}(u_n) \end{aligned}$$

Assim, no caso gaussiano, a covariância nula entre variáveis implica a sua independência, propriedade que não é generalizada à maioria das distribuições (V. e.g. Rohatgi, 1976, p.233).

Com recurso à função característica prova-se também facilmente que qualquer transformação linear de um vector gaussiano é ainda um vector gaussiano:

$$(5.3) \quad x \sim N(m, C) \Rightarrow y = Ax \sim N(Am, ACA')$$

em que A é uma matriz de ordem $k \times n$

Verifica-se que

$$\begin{aligned} \Phi_y(v) &= E[e^{i v' y}] = E[e^{i (A' v)' x}] = \Phi_x(A' v) = \\ &= \exp\{i v' A m - \frac{1}{2} v' A C A' v\}, \quad v \in R^k \end{aligned}$$

então, o vector aleatório k -dimensional y segue também uma multinormal, com média $A m$ e matriz de covariância $A C A'$.

Em particular verifica-se que cada componente x_j de x é normal, com função característica

$$(5.4) \quad \Phi_{x_j}(v_j) = \exp\{i m v_j - \frac{1}{2} c_{jj} v_j^2\}$$

Note-se que A é matriz real qualquer, de ordem $k \times n$, com $k \geq 1$ (para não referir o caso trivial $k=0$). Significa isso que, dado

o conjunto de v.a. (x_1, x_2, \dots, x_n) componentes do vector multinormal x , o espaço de combinações lineares desses componentes x_j é uma variedade linear gaussiana (*), com dimensão não superior a n .

Um processo estocástico $\{x(t)\}$, de parâmetro discreto ou contínuo, é denominado GAUSSIANO quando a distribuição de probabilidade conjunta de qualquer família finita de variáveis do processo for gaussiana. Por outras palavras, $\{x(t)\}$ diz-se gaussiano quando qualquer vector aleatório, formado com um número arbitrário, finito, de variáveis do processo, tiver distribuição multinormal.

A importância dos processos gaussianos deriva não só da simplificação que permitem, no estudo de muitas das suas características, como ainda do facto de possibilitarem a caracterização de propriedades fortes, mais estritas do que as verificadas pela classe dos processos não gaussianos.

Qualquer processo estacionário está sempre em dualidade com um processo gaussiano, definido num outro espaço de probabilidade mas a com a mesma média e a mesma função de autocovariância que esse outro processo (V. Doob, 1953, pp.72 e seqs).

(*) "um conjunto (não necessariamente fechado) que contém qualquer combinação linear $\alpha x + \beta y$ desde que contiver os elementos x e y será chamado de variedade linear." (Kpimogorov, 1976, p.136).

Em termos teóricos, a correspondência estabelecida vem clarificar e organizar o estudo dos processos estocásticos. Sempre que um processo tenha uma dada propriedade P_1 , definida em termos dos primeiros e segundos momentos, o correspondente processo gaussiano terá uma propriedade P_e com características mais estritas. P_1 será chamada propriedade em sentido lato e P_e propriedade em sentido estrito (*).

A distinção entre processos estacionários em sentido lato e em sentido estrito ganha agora outro significado. Com efeito, definindo-se estacionaridade como a invariância da média e da covariância face a uma translação temporal, se a essa caracterização se somar a normalidade das variáveis aleatórias, o processo passará a ser completamente estacionário. A propriedade P_1 , que é a da estacionaridade à segunda ordem, corresponde a propriedade P_e , que é a da completa estacionaridade dos processos gaussianos que respeitam P_1 .

De igual forma, a distinção entre a independência e a não correlação de v.a. (ou a sua ortogonalidade quando $Ex_t=0$) pode ser expressa em termos de propriedades lato e estrita.

Considere-se o processo $\{x_t\}$ de variáveis centradas ($Ex_t=0$); a independência implicará a ortogonalidade

$$F(x_t x_s) = F(x_t) \cdot F(x_s) \Rightarrow E[x_t x_s] = 0$$

mas a recíproca não é, em geral, verdadeira.

 (*) A distinção deve-se a DOOB. Cf. Doob, 1953, p.77.

No entanto, se as variáveis forem gaussianas, a ortogonalidade é equivalente à independência.

P_1 representará a ortogonalidade entre variáveis centradas e a correspondente propriedade gaussiana, P_e , será a propriedade mais estrita da independência.

NOTA 6. Séries de Variáveis Aleatórias

Definir-se-á uma condição operacional para o estudo da convergência de uma sucessão de v.a., de modo a poder abordar uma classe particular de séries em $L^2(\Omega)$.

CONDIÇÃO NECESSÁRIA E SUFICIENTE DE CONVERGÊNCIA

A sucessão $\{x_n\}$ é convergente se e só se $E[x_n x_m]$ converge com $n, m \rightarrow \infty$:

$$(6.1) \quad] \text{ l.i.m. } x_n \Leftrightarrow] \lim E[x_n x_m]$$

Admita-se que $E x_n x_m \rightarrow c$ com $n, m \rightarrow \infty$. Então, $\|x_n - x_m\|^2 = E[x_n - x_m]^2 = E x_n^2 - 2E x_n x_m + E x_m^2 \rightarrow c^2 - 2c^2 + c^2 = 0$. Isso significa que $\{x_n\}$ é sucessão de Cauchy, portanto convergente em $L^2(\Omega)$.

A condição necessária deriva directamente da convergência dos momentos cruzados (2.7) \square .

Esta propriedade será utilizada para provar uma condição de existência de soma de séries de v.a. não correlacionadas.

Por aplicação directa da definição de convergência de uma sucessão de v.a., considera-se que a série de v.a. $\sum_0^\infty x_n$ (ou, mais simplificada, $\sum x_n$) tem soma S quando a sucessão de somas parciais $\{S_n\}$, com

$$S_n = \sum_0^n x_u$$

converge para S :

$$\| S - S_n \| \rightarrow 0 \quad \text{com } n \rightarrow \infty$$

CONDIÇÃO DE CONVERGÊNCIA DE SÉRIE DE V.A. NÃO CORRELACIONADAS

Considerem-se as somas parciais $S_n = \sum_0^n x_u$ e $S_m = \sum_0^m x_u$ e admita-se, sem perda de generalidade, que $n \leq m$. O momento cruzado $E[S_n \cdot S_m]$ tem a expressão:

$$E[S_n \cdot S_m] = \sum_i \sum_j E[x_i \cdot x_j]$$

com $i=1,2,\dots,n$ e $j=1,2,\dots,m$.

Se as variáveis são não correlacionadas $E[x_i \cdot x_j] = 0$, para $i \neq j$, pelo que

$$E[S_n \cdot S_m] = \sum_0^n E[x_i^2]$$

Assim, como S_n converge sse $E[S_n \cdot S_m]$ converge, a condição de existência de limite para a sucessão de somas parciais equivale à condição de existência de soma da série real:

$$(6.2) \quad \sum_0^\infty \|x_u\|^2 = \sum_0^\infty E x_u^2 \quad (*)$$

(*) A demonstração difere ligeiramente das demonstrações clássicas, baseadas nas sucessões ortonormais de vectores de L^2 (e.g. Berberian, 1976, p.49) ou na condição geral de Cauchy (e.g. Rudin, 1973, p.295).

Significa isso que a v.a. $\sum x_u$ é definida (no sentido da média quadrática) se a série real (6.2) for convergente; situação em que $E[\sum x]$ também existirá - com valor $\sum [Ex]$ como se deduz de (2.4).

Um caso particular importante surge quando $Ex_u = 0 \quad \forall u$, situação em que Ex^2 representa a variância das v.a. e em que se conclui que $\sum x$ existe quando tiver variância finita, situação na qual

$$(6.3) \quad \text{VAR} [\sum x_u] = \sum \text{VAR}[x_u] = \sum Ex_u^2$$

e, naturalmente,

$$(6.4) \quad E [\sum x_u] = \sum [Ex_u] = 0$$

NOTA 7. Pseudo-inversa

Uma matriz A^- é chamada quadrada, quando satisfaz a relação

$$(i) \quad AA^-A = A$$

Para evitar a falta de unicidade da matriz definida por (i) usual, seguindo Penrose (*), acrescentar os seguintes axiomas:

$$(ii) \quad A^-AA^- = A^-$$

$$(iii) \quad (AA^-)^T = AA^-$$

$$(iv) \quad (A^-A)^T = A^-A$$

Uma matriz assim definida, respeitando (i) a (iv), será chamada de inversa generalizada ou pseudo-inversa de Penrose e será denotada por A^+ .

Se o vector aleatório x seguir uma multinormal com matriz de covariância singular a sua definição não pode basear-se na função de distribuição mas terá de ser expressa a partir da função característica (V. Nota 5).

Em particular, se a matriz de covariância de y , seja Σ_{yy} , for singular e se x e y forem conjuntamente gaussianos, prova-se (ibidem, p.382) que:

$$E[x|y] = E_x + \Sigma_{xy} \Sigma_{yy}^+ (y - E_y)$$

$$V[x|y] = \Sigma_{xx} - \Sigma_{xy} \Sigma_{yy}^+ \Sigma_{yx}$$

(*) As definições e resultados aqui apresentados seguem de perto Kalman, 1963, pp.367-83.

NOTA 8. Processos puramente aleatórios --- "ruído branco"

As sucessões de variáveis aleatórias não correlacionadas - processos puramente aleatórios discretos - representam, naturalmente, uma abstracção de aplicabilidade directa pouco usual. O seu interesse deriva sobretudo da sua utilização como pedra base na concepção de modelos mais complexos, largamente utilizados no estudo de sucessões cronológicas.

Adoptar-se-á a definição que segue, representando-se as v.a. deste modelo (discreto) por ϵ_t .

PROCESSOS PURAMENTE ALEATORIOS $\{\epsilon_t\}$ são processos estacionários à segunda ordem com função de autocovariância $R(\tau)=0$, $\forall \tau \neq 0$.

Ter-se-á, evidentemente

$$(8.1) \quad E\epsilon_t = \mu \quad \forall t, \quad |\mu| < \infty$$

$$(8.2) \quad V[\epsilon_t] = \sigma^2 \quad \forall t, \quad \sigma^2 < \infty$$

$$(8.3) \quad R = \sigma^2 I \quad \begin{matrix} 0 & \tau \neq 0 \\ \text{ou } R(\tau) = & \\ & \sigma^2 & \tau = 0 \end{matrix}$$

$$(8.4) \quad \rho = I \quad \begin{matrix} 0 & \tau \neq 0 \\ \text{ou } \rho(\tau) = & \\ & 1 & \tau = 0 \end{matrix}$$

Outras caracterizações são igualmente utilizadas. Jenkins e Watts (1968), Jazwinski (1970) e outros, incluem na definição a independência das v.a. - exigência mais restritiva pois, a menos que a distribuição conjunta seja multinormal, a não correlação não implica independência. Chatfield (1984) inclui ainda a exigência de as v.a. serem identicamente distribuídas.

Em termos das propriedades lata e estrita vê-se que a definição adoptada corresponde a uma propriedade tipo P_1 , que se transforma numa propriedade tipo P_e em processo puramente aleatório gaussiano, processo que satisfará as condições definidas por esses autores.

Se a definição caracterizar o processo puramente aleatório discreto a partir de processo contínuo (e.g. Parzen, 1962, p.113 ou Brown, 1983, p.90) levantam-se problemas de outra complexidade. Esses problemas usualmente são ultrapassados definindo à partida o processo como gaussiano, outra razão pela qual o processo puramente aleatório discreto é muitas vezes definido com o pressuposto da normalidade.

Encontra-se ainda generalizada a designação de RUÍDO BRANCO para processos puramente aleatórios. O termo "ruído" deriva do facto de o valor actual da v.a. do processo ser não autocorrelacionado com os valores anteriores. Com processo gaussiano essa propriedade tem expressão mais estrita, em termos da distribuição condicional da v.a. :

$$(8.5) \quad f(\epsilon_t | \epsilon_{t-s}) = f(\epsilon_t) \quad s > 0$$

vendo-se que o processo é um "ruído puro", absolutamente não

previsível a partir dos valores anteriormente registados.

A adjectivação de "branco", por seu lado, exprime a "descaracterização" de $\{\epsilon_t\}$. Deriva de uma analogia espectral com a luz branca, em que todas as frequências do espectro visível estão presentes com igual relevo.

NOTA 9. O Espaço de Hilbert $L^2(\Omega)$

Vai demonstrar-se que toda a sucessão de Cauchy, i.e.

$$(9.1) \quad \langle x_n \rangle: \quad \|x_n - x_m\| \rightarrow 0 \quad \text{com } n, m \rightarrow \infty$$

é convergente em $L^2(\Omega)$ (V. e.g. Loève, 1963, p.161 ou Ash, 1972, p.85), ou seja, que

$$\text{l.i.m. } x_n \in L^2(\Omega)$$

o que significa que $L^2(\Omega)$ é um espaço completo.

A demonstração pode ser apresentada de forma muito geral, considerando qualquer espaço $L^p(\mu)$ com qualquer medida μ positiva e com $1 < p < \infty$ (V. e.g. Rudin, 1974, p. 69). É aplicável a espaços munidos de medida não finita (V. Kolmogorov, 1976, p.373). A demonstração que segue limita-se ao espaço de v.a. de segunda ordem que se defronta, e é apenas uma concretização do Teorema de Riesz-Fisher, que afirma que o espaço $L^2(\mu)$, de funções mensuráveis e de quadrado integrável, é um espaço completo (V. Berberian, 1965, p.111).

COMPLETICIDADE DE $L^2(\Omega)$

Se a sucessão $\langle x_n \rangle$ é de Cauchy então há certamente uma subsucessão $\langle x'_s \rangle$ tal que

$$\|x'_s - x'_{s-1}\| < 2^{-s} \quad s=1,2 \dots$$

Construa-se a partir de $\{x'_s\}$ a sucessão monótona crescente de somas parciais

$$y_s = |x'_1| + \sum_2^s |x'_s - x'_{s-1}|$$

Como $y_s \geq 0$ e $Ey_s \geq 0$ a desigualdade (29) permite escrever

$$\begin{aligned} Ey_s &= E|x'_1| + \sum_2^s E|x'_s - x'_{s-1}| \leq \|y_s\| = \\ &= \|x'_1 + \sum_2^s x'_s - x'_{s-1}\| \end{aligned}$$

Aplicando a desigualdade triangular ao último membro desta expressão conclui-se que

$$Ey_s \leq \|x'_1\| + \sum_2^s \|x'_s - x'_{s-1}\|$$

A sucessão $\{Ey_s\}$ é majorada, pois a série $\sum \|x'_s - x'_{s-1}\|$ é majorada por série geométrica de razão $\frac{1}{2}$. Pelo teorema de Beppo-Levi isso significa que a v.a. y_s , que é a função $y_s(w)$, converge para alguma v.a. $x(w)$, que existe e é finita. A convergência entende-se em todo o ponto $w \in \Omega$, no sentido de $\lim y_s(w) = x(w)$, com excepção de um conjunto de medida nula(*). Trata-se, em termos de teoria das probabilidades, da convergência com probabilidade 1.

 (*) O teorema da convergência monótona de Beppo-Levi permite afirmar que a sucessão de funções integráveis $\{y_s(w)\}$, sendo monótona crescente e sendo os respectivos integrais majorados, $Ey_s \leq K$, converge em quase toda a parte para uma função $x(w)$ que existe, é finita e integrável. Cf. Kolmogorov, 1976, p.293.

Mas, se a série $|x'_1| + \sum_2^{\infty} |x'_s - x'_{s-1}|$ converge, isso significa que a série $x'_1 + \sum_2^{\infty} (x'_s - x'_{s-1})$ também converge e, como esta última tem somas parciais identicamente iguais a x'_s isso significa que x'_s converge para x , para todo o ω com excepção de um conjunto de probabilidade nula.

Tendo encontrado x que é limite da subsucessão (x'_s) resta provar que esse limite por pontos é limite em média quadrática de (x_n) e pertence a $L^2(\Omega)$.

Por hipótese

$$E [x_n - x_m] \rightarrow 0$$

De (x'_s) extraia-se a subsucessão (x'_s) atrás definida. Para todo o n fixo ter-se-á

$$(x_n - x'_s) \rightarrow (x_n - x) \quad \text{q.c.}$$

Aplicando o Lema de Fatou (*) encontra-se

$$E[x_n - x]^2 \leq \liminf_s E[x_n - x'_s]^2 \rightarrow 0, \text{ com } i \rightarrow \infty.$$

Então, i.i.m. $x_n = x$ e, como $E[x_n - x]^2$ é limitado, $(x_n - x)$ pertence a L^2 pelo que também $x \in L^2$.

 (*) O Lema de Fatou permite afirmar que a sucessão de funções positivas integráveis $\{[x_n(\omega) - x'_s(\omega)]^2\}$, tendo o respectivo integral majorado, $E[x_n(\omega) - x'_s(\omega)]^2 \leq K$, admite uma função integrável tal que $E[\cdot] \leq \liminf_s E[\cdot]_s$. Cf. Berberian, 1965, p.105.

Com este resultado fundamental provou-se que $L^2(\Omega)$, pré-espço de Hilbert (*), é um espaço completo, pelo que $L^2(\Omega)$ constitui um ESPAÇO DE HILBERT (V. e.g. Berberian, 1961, p.40).

A conclusão é suficientemente importante para que se sumarizem as propriedades de $L^2(\Omega)$ que o fazem corresponder à definição de ESPAÇO DE HILBERT:

(i) $x \cdot y = Exy$ é produto escalar que define $L^2(\Omega)$ com suporte em S -
com o produto escalar define-se $\|x\| = (x \cdot x)^{1/2} = E^{1/2}x^2$,
 $d(x,y) = \|x-y\| = [(x-y) \cdot (x-y)]^{1/2} = E^{1/2}[x-y]^2$ que são,
respectivamente, uma norma e uma distância;

(ii) $L^2(\Omega)$ é espaço completo.

(*) Um pré-espço de Hilbert é um espaço vectorial munido de produto escalar. V. Berberian, 1961, pp.25-33. Em L^2 está definido o produto escalar Exy .

NOTA 10. Processos lineares gerais, AR, MA, ARMA

Os processos lineares $\{x_t\}$ definem-se a partir de combinação linear de variáveis do próprio processo e/ou de variáveis puramente aleatórias $\{\epsilon_t\}$. Nos casos não triviais essa combinação inclui variáveis desfasadas.

Dois exemplos clássicos são os processos autoregressivos e os de média móvel.

PROCESSO AUTOREGRESSIVO de ordem p , finita, representado por $AR(p)$ é o processo $\{x_t\}$ definido pela equação

$$(10.1) \quad x_t + a_1 x_{t-1} + a_2 x_{t-2} + \dots + a_p x_{t-p} = \epsilon_t,$$

em que $\{\epsilon_t\}$ é processo puramente aleatório e a_1, a_2, \dots, a_p são constantes.

A equação (10.1) pode ser expressa de forma equivalente escrevendo $x_t = \sum_1^p a_u x_{t-u}$ ou ainda $\sum_0^p a_u x_{t-u} = \epsilon_t$, o que apenas obriga a redefinir as constantes de ponderação a_u . Sem perda de generalidade pode considerar-se $a_0 = 1$.

Se o conjunto de índices tiver um mínimo será ainda necessário redefinir $x_0 = \epsilon_0, x_1 + a_1 x_0 = \epsilon_1, \dots, x_{p-1} + a_1 x_{p-2} + \dots + a_{p-1} x_{t-p+1} = \epsilon_{p-1}$.

PROCESSO DE MÉDIAS MOVEIS de ordem q , finita, representado por $MA(q)$ é o processo $\{x_t\}$ que respeita a equação

$$(10.2) \quad x_t = \epsilon_t + b_1 \epsilon_{t-1} + \dots + b_q \epsilon_{t-q}$$

em que b_1, b_2, \dots, b_q são constantes e $\{\epsilon_t\}$ é processo

puramente aleatório.

As observações feitas à forma de descrição do processo AR(p) são aqui igualmente válidas. A designação generalizada de MA(q) provém da terminologia inglesa "Moving Average".

As equações (10.1) e (10.2) são exemplo de equações de descrição de processos lineares que podem ser reescritas simplificadaamente com recurso a operadores de desfasamento.

Introduza-se o OPERADOR DE DESFASAMENTO REGRESSIVO ("backward shift") representado por B (*) e definido pela expressão

$$(10.3) \quad B y_t = y_{t-1}$$

O desfasamento de mais de um período pode ser igualmente expresso com esse operador, na forma

$$(10.4) \quad B^k y_t = y_{t-k}$$

definindo-se ainda

$$B^0 y_t = y_t$$

Poderiam ainda introduzir-se o OPERADOR DE DESFASAMENTO PROGRESSIVO ("forward shift"), representado por F e definido pela

(*) É natural encontrarem-se designações alternativas para este operador. A letra B é introduzida por Box e Jenkins (1970, p.8) enquanto Harvey (1981, p.26) e outros autores da escola econométrica inglesa preferem a letra L de "lag".

expressão

$$(10.5) \quad F^k y_t = y_{t+k}$$

o OPERADOR DE DIFERENCIAÇÃO D

$$(10.6) \quad Dy_t = y_t - y_{t-1}$$

Ter-se-á, como é evidente,

$$(10.7) \quad F = B^{-1} \quad F^{-1} = B$$

$$(10.8) \quad D = (1-B)$$

Note-se ainda que $D^{-1} = (1-B)^{-1}$ corresponde a um operador de soma

$$\begin{aligned} D^{-1}y_t &= (1-B)^{-1}y_t \\ &= (1 + B + B^2 + \dots)y_t \\ &= y_t + y_{t-1} + y_{t-2} + \dots \\ &= \sum_0^{\infty} y_{t-u} \end{aligned}$$

D que se pode confirmar directamente a partir das identidades

$$\begin{aligned} y_t &= D^{-1}Dy_t \\ &= (y_t - y_{t-1}) + (y_{t-1} - y_{t-2}) + \dots \\ y_t &= DD^{-1}y_t \\ &= (y_t + y_{t-1} + \dots) - (y_{t-1} + y_{t-2} + \dots) \end{aligned}$$

Poderá agora reescrever-se o processo autoregressivo de forma mais compacta

$$\sum_0^p a_u B^u x_t = \epsilon_t$$

ou ainda

$$(10.9) \quad \alpha(B)x_t = \epsilon_t$$

onde o polinómio $\alpha(\cdot)$ se define por

$$(10.10) \quad \alpha(z) = 1 + a_1 z + a_2 z^2 + \dots + a_p z^p$$

O processo de médias móveis, por seu turno, terá a expressão

$$x_t = \sum_0^q b_u B^u \epsilon_t$$

ou ainda

$$(10.11) \quad x_t = \beta(B)\epsilon_t$$

com o polinómio $\beta(\cdot)$ definido por

$$(10.12) \quad \beta(z) = 1 + b_1 z + \dots + b_q z^q$$

Uma generalização dos processos lineares é o chamado Processo Linear Geral, conceito que é devido a Bartlett (1955) que assim designou uma classe de processos resultante da combinação linear (infinita) de variáveis puramente aleatórias. A definição aqui adoptada é de tipo P_1 , diferindo da proposta por esse autor

 (*) Enquanto Bartlett construía o processo linear como combinação de v.a. independentes aqui apenas se supõe a não correlação. O conceito de Bartlett pode ser visto como uma concretização

(*). Seguir-se-á de perto Priestley (1981).

PROCESSO LINEAR GERAL é o processo $\{x_t\}$ que pode ser expresso na forma

$$x_t = \sum_0^{\infty} g_u \epsilon_{t-u}$$

em que $\{\epsilon_t\}$ é processo puramente aleatório e $\{g_u\}$ sucessão de constantes tais que $\sum_0^{\infty} g_u^2 < \infty$.

A série $\sum_0^{\infty} g_u \epsilon_{t-u}$ representa o limite em média quadrática da sucessão de somas parciais $S_n = \sum_0^n g_u \epsilon_{t-u}$. Como foi demonstrado na Nota 6 tal limite existe se a série real $\sum \|g_u \epsilon_{t-u}\|^2$ tiver soma. Ora

$$\sum \|g_u \epsilon_{t-u}\|^2 = \sigma_{\epsilon}^2 \cdot \sum g_u^2$$

Assim, a condição $\sum_0^{\infty} g_u^2 < \infty$ assegura que $x_t \in L^2(\Omega)$.

Assuma-se que $E\epsilon_t = 0$, pressuposto que será de ora em diante sempre admitido. Nenhuma restrição séria se coloca ao estudo subsequente. Por uma simples transformação de variáveis, que corresponde a uma deslocação para a origem, é sempre possível transformar processo estacionário em processo estacionário centrado - com valor esperado nulo. Por outro lado, a não correlação entre variáveis ($E[\epsilon_t \epsilon_s] = E\epsilon_t \cdot E\epsilon_s$), transforma-se agora na sua ortogonalidade ($E[\epsilon_t \epsilon_s] = 0$), o que simplifica o estudo.

Passará a ter-se, como em (6.4),

em sentido estrito do PLG, e sê-lo-á quando o processo for gaussiano.

$$(10.13) \quad E x_t = 0$$

e, como em (6.3)

$$(10.14) \quad V[x_t] = \sigma_x^2 = \sigma_\varepsilon^2 \cdot \sum_0^{\infty} g_u^2$$

Para determinar as funções de autocovariância e autocorrelação reescrever-se-á o processo de forma que permite encontrar expressões mais simples para essas funções:

$$(10.15) \quad x_t = \sum_{-\infty}^{+\infty} g_u \varepsilon_{t-u} \quad \text{com } g_u = 0 \text{ para } u < 0$$

Em teoria dos sistemas esta equação define um sistema linear invariante que, com a condição $g_u = 0$ para $u < 0$, representa um sistema fisicamente realizável ou causal, terminologia evidente por si própria (Cf. Chatfield, 1975, pp.187-8).

Com a expressão (10.15) encontrar-se-á imediatamente a função de autocovariância

$$(10.16) \quad R(\tau) = E[x_t, x_{t-\tau}] = \sigma_\varepsilon^2 \cdot \sum_{-\infty}^{+\infty} g_u g_{u-\tau}$$

sendo nulos os termos da série quando $u=0$ ou $\tau > u$.

 (*) A convolução de duas funções f e g , representada por $f * g$, é o integral funcional $h(x) = \int_{\mathbb{R}} f(t)g(x-t)dt = \int_{\mathbb{R}} g(t)f(x-t)dt$. V. e.g. Apostol, 1974, p.327.

Em particular, a convolução de duas famílias $\{x_u\}$ $\{y_u\}$ $u \in \mathbb{Z}$, de números reais ou complexos, somáveis absolutamente, é a família de termo geral $z_\tau = (x * y)_\tau = \sum_{-\infty}^{+\infty} x_u y_{u-\tau}$.

A função de autocovariância é assim a autoconvolução (*) de $\{g_u\}$, o que se transforma numa regra prática para o seu cálculo, nomeadamente quando o número de parâmetros $g_u \neq 0$ é finito.

A função de autocorrelação será dada por

$$(10.17) \quad \rho(\tau) = \frac{\sum_{-\infty}^{+\infty} g_u g_{t-\tau}}{\sum_0^{+\infty} g_u^2}$$

A existência das séries (10.16) e (10.17) é garantida no pressuposto de $E\varepsilon_t = 0$, conforme se verá de seguida.

CONDIÇÕES DE ESTACIONARIDADE

Com os pressupostos $E\varepsilon_t = 0$ e $\sum_u^2 < +\infty$ o processo linear geral é centrado e estacionário à segunda ordem. Com efeito

$$E\varepsilon_t = 0 \quad V[x_t] = \sigma_x^2 \quad \forall t$$

Por outro lado, como $|R(\tau)| \leq R(0)$, a série $\sum_u g_u g_{u-\tau}$ necessariamente converge, confirmando a generalidade da expressão encontrada para $R(\tau)$. Ora nessa expressão o valor de $R(\tau)$ é função unicamente da amplitude do desfasamento τ , pelo que assim se cumpre a última condição de estacionaridade do processo.

Admitindo $E\varepsilon_t = 0$, o que torna imediatamente o processo estacionário à primeira ordem, a condição de estacionaridade resume-se assim à condição $\sum_u^2 < \infty$, incluída na definição anteriormente estabelecida, pelo que, nessas condições, todo o

V. Tykhomirov, 1982, p.504.

processo linear geral é estacionário à segunda ordem.

Pode ainda demonstrar-se (V. Bartlett, 1955, p.147) que o pressuposto adicional de independência das variáveis ϵ_t implica a completa estacionaridade do processo. Admitindo que este é gaussiano fica automaticamente introduzida essa propriedade de tipo P_e face à propriedade de tipo P_1 de estacionaridade à segunda ordem.

Introduza-se agora a reescrita do processo em termos de operadores de desfasamento, de forma a encontrar outra forma para a condição de estacionaridade $\sum g_u^2 < \infty$.

A expressão $x_t = \sum g_u B^u \epsilon_t$ aparece na forma mais compacta

$$(10.18) \quad x_t = G(B) \epsilon_t$$

tendo sido introduzida a função

$$(10.19) \quad G(z) = \sum_0^{\infty} g_u z^u$$

Trata-se pois de função representada em série de potências, com centro na origem. Ver-se-á que a condição de estacionaridade pode ser expressa em termos do carácter analítico (*) dessa função $G(z)$.

(*) Uma função é analítica num conjunto aberto S se a sua primeira derivada existe em todo o ponto de S .

As propriedades, definições e teoremas base da demonstração que segue podem encontrar-se em Apostol, 1974, pp.193, 234-7, 434 e 449-50.

Em primeiro lugar verifique-se que

$$(i) \sum g_u^2 < \infty \Rightarrow |\sum g_u z^u| < \infty \text{ para } |z| < 1$$

o que significa que $G(z)$ será analítica no interior do círculo unitário.

Pelo critério da raiz a convergência de $\sum g_u^2$ obriga à desigualdade $L = \limsup (g_u^2)^{1/u} \leq 1$. Ora, como $\limsup (g_u^2)^{1/u} = \limsup |g_u|^{1/u}$, L é precisamente o inverso do raio de convergência da série de potências, pelo que esse raio r verificará a desigualdade $r \geq 1$, o que implica que a série seja convergente para $|z| < 1$, portanto, que $G(z)$ seja função analítica no interior do círculo unitário ■.

Demonstre-se agora que

$$(ii) \text{ se } G(z) \text{ for função analítica para } |z|=1 \text{ necessariamente a série } \sum g_u^2 \text{ converge.}$$

Para $G(z)$ ser analítica para $|z|=1$ terá de ser representável, em cada ponto da fronteira do círculo unitário, e numa vizinhança de cada desses pontos, por uma série de potências. Isso significa que o raio de convergência da série terá de ser superior à unidade. Então, se $r > 1$, $L = 1/r < 1$.

Mas $L = \limsup |g_u|^{1/u} = \limsup (g_u^2)^{1/u} < 1$, pelo que a série $\sum g_u^2$ converge ■.

Conclui-se assim que a condição de estacionaridade do processo linear geral pode ser expressa em termos da função $G(z)$, ficando essa condição assegurada se a função $G(z)$ for analítica no interior e na fronteira do círculo unitário : $|z| \leq 1$.

PROCESSOS DE MEDIAS MOVEIS E AUTOREGRESSIVOS

É imediatamente verificável que o processo linear geral encontra os processos de médias móveis como um seu caso particular. De facto, definindo $g_u=0$ para $u>q$ encontra-se (10.1)

$$x_t = \sum_0^q b_u \epsilon_{t-u},$$

ou ainda, identificando a função $G(z)$ com o polinómio finito $\beta(z)$,

$$x_t = \beta(B)\epsilon_t$$

É facto ainda, sob determinadas "condições de invertibilidade", que um processo autoregressivo pode ser expresso como processo linear geral. O processo AR(1) por exemplo

$$x_t + ax_{t-1} = \epsilon_t$$

permite encontrar recursivamente

$$x_t = \epsilon_t - ax_{t-1}$$

$$x_t = \epsilon_t - a(\epsilon_{t-1} - a\epsilon_{t-2})$$

...

$$x_t = \epsilon_t - a\epsilon_{t-1} + a^2\epsilon_{t-2} - \dots + (-a)^u \epsilon_{t-u} + \dots$$

conduzindo ao processo

$$x_t = \sum_0^{\infty} (-a)^u \epsilon_{t-u} \quad \text{com } (-a)^u = g_u$$

É também admissível que o processo $x_t = G(B)\epsilon_t$ possa ser

reescrito na forma

$$(10.20) \quad G^{-1}(B)x_t = \epsilon_t$$

Se for possível desenvolver $G^{-1}(B)$ em série de potências

$$G^{-1}(z) = \sum_0^{\infty} h_u z^u,$$

encontrar-se-á processo autoregressivo.

A condição, naturalmente, é que a série $\sum h_u B^u x_t$ seja convergente em média quadrática. Raciocínio semelhante ao já atrás efectuado leva à conclusão de que tal requisito é satisfeito se $G^{-1}(z)$ for função analítica no interior e fronteira do círculo unitário:

$$(10.21) \quad G^{-1}(z) < \infty \quad |z| \leq 1.$$

Efectue-se o raciocínio inverso: se $AR(p)$ modelo autoregressivo de ordem p , estacionário

$$\alpha(B)x_t = \epsilon_t$$

com $\alpha(z)$ polinómio finito, admitir a inversão

$$x_t = \alpha^{-1}(B)\epsilon_t$$

isso significa que ele pode ser escrito, equivalentemente, na forma de processo de médias móveis de ordem (em geral) infinita; ou seja, na forma de processo linear geral. A condição para que $\alpha^{-1}(z)$ seja invertível em $\alpha(z)$, gerando processo estacionário

equivalente é a condição

$$|\alpha^{-1}(z)| < \infty \quad |z| < 1 .$$

Isso implica que todas as raízes do polinómio finito $\alpha(z)$ estejam no exterior do círculo unitário, o que é, afinal, forma equivalente de expressar a condição de invertibilidade em processo estacionário, mas agora directamente a partir do polinómio $\alpha(z)$.

Considere-se o problema contrário: expressar o processo estacionário $MA(q)$, processo de médias móveis finito

$$x_t = \beta(B)\epsilon_t$$

na forma de processo autoregressivo:

$$\beta^{-1}(B)x_t = \epsilon_t$$

O polinómio $\beta(B)$ é finito, de ordem q , e o polinómio $\beta^{-1}(B)$ será, em geral, de ordem infinita.

A condição de invertibilidade em processo estacionário equivalente é, tal como em (10.21), expressa sobre a função $\beta^{-1}(z)$:

$$|\beta^{-1}(z)| < \infty \quad |z| \leq 1 ;$$

requisito que é sempre satisfeito se o polinómio (finito) $\beta(z)$ tiver todas as raízes exteriores ao círculo unitário.

Estes resultados podem ser sintetizados nas seguintes condições

de invertibilidade ou de equivalência entre processos de Médias Móveis e Autoregressivos estacionários (*):

(i) MA(q) \rightarrow AR(∞)

$$x_t = \beta(B)\epsilon_t \Leftrightarrow \beta^{-1}(B)x_t = \epsilon_t$$

transformação possível se todas as raízes do polinómio finito $\beta(B)$ forem exteriores ao círculo unitário

(ii) AR(p) \rightarrow MA(∞)

$$\alpha(B)x_t = \epsilon_t \Leftrightarrow x_t = \alpha^{-1}(B)\epsilon_t$$

transformação possível se todas as raízes do polinómio finito $\alpha(B)$ forem exteriores ao círculo unitário

As condições de estacionaridade do processo linear geral, que se deduzem directamente da expressão (10.21) concretizam-se na seguinte:

(iii) MA(q)

$$x_t = \beta(B)\epsilon_t$$

processo sempre estacionário pois o polinómio de ordem q $\beta(z)$ tem sempre valor finito para $|z| \leq 1$

(iv) AR(p)

$$\alpha(B)x_t = \epsilon_t$$

processo estacionário desde que a função $\alpha^{-1}(z)$ seja analítica para $|z| \leq 1$, ou seja, desde que as raízes do polinómio $\alpha(z)$ sejam todas exteriores ao círculo unitário.

(*) V. Priestley, 1981, pp.132-46.

PROCESSOS ARMA

O processo linear geral permite ainda encontrar, como caso particular de especial interesse, os conhecidos modelos mistos, autoregressivos e de médias móveis, que se representam por ARMA(p,q) e que são da forma:

$$(10.22) \quad x_t + a_1 x_{t-1} + \dots + a_p x_{t-p} = \epsilon_t + b_1 \epsilon_{t-1} + \dots + b_q \epsilon_{t-q}$$

De modo mais compacto escrever-se-á

$$(10.23) \quad \alpha(B)x_t = \beta(B)\epsilon_t$$

E o modelo reduzir-se-á a um processo linear geral quando puder ser escrito na forma

$$(10.24) \quad x_t = \alpha^{-1}(B)\beta(B)\epsilon_t$$

o que é sempre possível se todas as raízes do polinómio $\alpha(z)$ forem exteriores ao círculo unitário. Essa é, aliás, tanto a condição de equivalência de um ARMA a um processo linear geral (estacionário), como a condição da sua estacionaridade, que reside apenas na sua parcela autoregressiva.

PROCESSOS ARIMA

Considere-se o processo (x_t) e encontrem-se as suas d-ésimas diferenças

$$x \rightarrow D^d x_t = (1-B)^d x_t$$

Se o novo processo segue um ARMA(p,q) dir-se-á que $\{x_t\}$ segue um ARIMA(p,d,q).

De forma compacta pode escrever-se

$$(10.25) \quad \alpha(B)(1-B)^d x_t = \beta(B)\epsilon_t.$$

NOTA 11. Processos estacionários

O estudo de processos cujas propriedades estatísticas não mudam com o tempo - processos estacionários - é particularmente importante na análise de sucessões cronológicas. Os modelos que se encontram mais completamente estudados são aqueles em que a sucessão é directamente estacionária ou pode tornar-se estacionária por meio de uma transformação simples.

PROCESSO COMPLETAMENTE ESTACIONARIO é aquele em que são idênticas as distribuições de probabilidade conjuntas de

$(x_{t_0}, x_{t_1}, x_{t_2}, \dots, x_{t_n})$ e $(x_{t_0+k}, x_{t_1+k}, x_{t_2+k}, \dots, x_{t_n+k})$ para toda a escolha admissível de índices, t_1, t_2, \dots, t_n , e de k .

A propriedade de completa estacionaridade é muito restritiva. No estudo prático das sucessões cronológicas é usual reduzir a exigência de invariância face a uma translação temporal à da invariância dos momentos da distribuição até certa ordem m .

PROCESSO ESTACIONARIO ATÉ A ORDEM m é aquele em que existem e são idênticos, para qualquer deslocação temporal, os momentos conjuntos até à ordem m

$$E[(x_{t_1})^{m_1} \dots (x_{t_n})^{m_n}] = E[(x_{t_1+k})^{m_1} \dots (x_{t_n+k})^{m_n}]$$

para toda a escolha admissível de k e do conjunto de índices, e para todos os expoentes inteiros positivos tais que $m_1 + \dots + m_n \leq m$.

Particularmente importantes são os processos estacionários até à segunda ordem. Trata-se de classe de processos estocásticos

profundamente estudada, pois a descrição até à segunda ordem é, sob muitos aspectos, perfeitamente suficiente. Para a análise espectral, por exemplo, sabe-se que tanto a frequência como a função potência dependem unicamente dos dois primeiros momentos. Para a importante classe dos processos gaussianos, por outro lado, a descrição até à segunda ordem é perfeitamente completa.

Tais considerações, seguindo Rozanov (1973, p. 213) e Priestley (1981, pp.112-3), entre outros, conduzem a que se adopte a seguinte definição:

PROCESSO ESTACIONARIO $\{x_t\}$ é aquele que é estacionário até à ordem 2, definindo-se assim pelas seguintes equações, válidas para todos os índices $t, s, t+k, s+k$ admissíveis:

- (i) $E x_t = E x_{t+k} = \mu$, μ constante
- (ii) $E[x_t x_s] = E[x_{t+k} x_{s+k}] = \phi$, função unicamente de $t-s$.

As seguintes propriedades são de verificação imediata

$$(11.1) \quad [P1] \quad E x_t = \mu < \infty, \quad t \in T$$

$$(11.2) \quad [P2] \quad V[x_t] = \sigma^2 < +\infty, \quad t \in T$$

$$(11.3) \quad [P3] \quad \text{Cov}[x_t, x_s] = R(s-t), \quad s, t \in T$$

[P1] estabelece, conforme à definição, que o valor esperado da variável é constante. Como, por (ii), x_t é v.a. de L^2 , o seu valor esperado é necessariamente finito.



[P3] deduz-se imediatamente da definição $Cov[x_t, x_s] = E[x_t \cdot x_s] - E x_t E x_s = E[x_t \cdot x_s] - \mu^2 = E[x_{t+k} \cdot x_{s+k}] - \mu^2$ e, portanto, apenas função de $s-t$, acima denotada por $R(\cdot)$.

[P2] aparece como caso particular de [P3], pois quando $s-t=0$ vem $E[x_t \cdot x_s] = \text{constante}$ e, portanto, $V[x_t] = E x_t^2 - \mu^2 = \text{constante}$.

Note-se que a presente definição de processo estacionário corresponde, em alguns autores, à definição de PROCESSO ESTACIONARIO EM COVARIANCIA (e.g. Karlin e Taylor, 1975, p.445) ou EM SENTIDO LATO (e.g. Murteira, 1981, p.34 ou Papoulis, 1984, p.220). Note-se ainda que a designação de ESTACIONARIO EM COVARIANCIA, ou EM SENTIDO LATO, por vezes corresponde apenas à verificação das propriedades [P2] e [P3], não sendo tais processos obrigatoriamente estacionários em média (V. e.g. Parzen, 1962, p.71) - o que quer dizer que são independentes as características (i) e (ii) contidas na definição acima adoptada.

Pode no entanto sustentar-se que a condição de estacionaridade em média - componente (i) da definição - "is unnatural mathematically and has nothing to do with the essential properties of interest in these [wide sense stationary] processes" (Doob, 1953, p.95).

Leia-se ainda Ventsel (1973, p.406): a condição de estacionaridade da função de autocovariância "c'est l'unique condition essentielle imposé à une fonction aléatoire stationnaire".

Apesar da coerência dessas objecções, a definição por que se optou mantém-se como a mais corrente e a tradicional no estudo

das sucessões cronológicas. O facto é que a identificação da estacionaridade com a estacionaridade dos primeiros e segundos momentos facilita a organização do estudo prático desses processos particulares.

Com a definição adoptada é óbvio que qualquer processo estacionário até à ordem m é também estacionário até às ordens $m-1$, $m-2$, ..., pelo que qualquer processo estacionário à ordem $m > 2$, é também estacionário, no sentido da definição utilizada. No entanto, um processo completamente estacionário pode não ser estacionário a nenhuma ordem, por não possuir momentos finitos.

Em processos estacionários o facto de $\text{Cov}[x_t, x_s]$ ser função unicamente da diferença $t-s$, leva a definir algumas funções de considerável interesse prático. Admitir-se-á, como até aqui esteve implícito, que o conjunto T é linear.

A FUNÇÃO DE AUTO-COVARIÂNCIA, denotada $R(\tau)$ e definida para todos os valores de τ admissíveis, tem como argumento o desfaseamento temporal entre as variáveis $x_t, x_{t+\tau}$ do processo estacionário $\{x_t\}$, representa a covariância entre essas variáveis e é definida pela equação

$$(11.4) \quad R(\tau) = E[(x_t - \mu)(x_{t+\tau} - \mu)].$$

E elementar a dedução das seguintes propriedades

$$(11.5) \quad [P1] \quad R(0) = \sigma^2$$

$$(11.6) \quad [P2] \quad |R(\tau)| \leq R(0)$$

$$(11.7) \quad [P3] \quad R(\tau) = R(-\tau)$$

A primeira propriedade deriva directamente das definições de variância e de covariância.

[P2] pode demonstrar-se verificando que $0 \leq V[y+z] = V[y]+V[z]+2.Cov[y,z]$. Substituindo y e z por duas variáveis do processo desfasadas de τ , encontra-se $\sigma^2+\sigma^2+2.R(\tau) \geq 0$, de onde deriva que $\sigma^2 \geq -R(\tau)$. Cálculo similar, considerando a variância da diferença entre as duas variáveis conduz a $\sigma^2 \geq R(\tau)$. As duas inequações implicam [P2].

A terceira propriedade deduz-se imediatamente da equação (11.4), trocando a ordem dos factores do produto operado no segundo membro. Note-se que este raciocínio é válido porque a v.a. é definida como real; se fosse complexa a igualdade apenas se verificaria em módulo (V. Priestley, 1981, p.111) \square .

Função normalizada pode ser obtida introduzindo não a covariância mas sim o coeficiente de correlação entre as variáveis. É esta última que tem maior utilização prática. Em conjunto com o conhecimento de σ^2 , fornece a mesma informação sobre o processo que a contida em $R(\tau)$.

A FUNÇÃO DE AUTOCORRELAÇÃO, denotada por $\rho(\tau)$ e definida para todos os valores de τ admissíveis, tem como argumento o desfasamento temporal entre as variáveis do processo estacionário $\{x_t\}$, representa o coeficiente de correlação entre as variáveis e é definida pela equação

$$(11.8) \quad \rho(\tau) = R(\tau)/R(0)$$

A partir de (11.4) verifica-se que esta equação é equivalente à seguinte

$$(11.9) \quad \rho(\tau) = \text{Cov}[x_t, x_{t+\tau}] / (\text{V}[x_t] \cdot \text{V}[x_{t+\tau}])^{1/2}$$

reencontrando-se a expressão habitual para o coeficiente de correlação.

E também elementar o estabelecimento das seguintes propriedades

$$(11.10) \quad [P1] \quad \rho(0) = 1$$

$$(11.11) \quad [P2] \quad |\rho(\tau)| \leq 1, \quad \tau \in T$$

$$(11.12) \quad [P3] \quad \rho(\tau) = \rho(-\tau), \quad \tau \in T$$

A primeira propriedade deduz-se directamente da definição, a segunda de (11.6) e a terceira de (11.7) ■.

Dado qualquer conjunto sequencial de variáveis do processo, de índices um a um distintos, ordenados naturalmente, pode definir-se MATRIZ DE VARIANCIA-COVARIANCIA DO PROCESSO, R , com elemento genérico $r_{ij} = R(i-j)$.

Trata-se de matriz simétrica e semi-definida positiva

$$(11.13) \quad u' \cdot R \cdot u \geq 0, \quad \forall u \in \mathbb{R}^n$$

com a particularidade ainda, em processo estacionário, de os

elementos diagonais serem todos idênticos e iguais a σ^2 . A característica semi-definida tem de ser respeitada (tal como as três propriedades atrás apresentadas) ao se postular determinada função como representando a autocovariância de um processo.

Quando as variáveis do processo estacionário são não autocorrelacionadas a matriz será escalar, podendo-se indicar, de forma compacta, que o processo é estacionário e de variáveis não autocorrelacionadas escrevendo as seguintes equações:

$$(11.14) \quad \begin{aligned} E x_t &= \mu \\ R &= \sigma^2 I \quad \text{ou,} \quad V[x_t] = \sigma^2 I \end{aligned}$$

Pode também definir-se, de forma análoga, MATRIZ DE CORRELAÇÃO do processo, matriz que respeitará as mesmas características, com a particularidade de ter os elementos diagonais identicamente iguais à unidade.

Outra função de especial interesse na análise de sucessões cronológicas é a FUNÇÃO DE AUTOCORRELAÇÃO PARCIAL, construída com os coeficientes parciais

$$r_{t, t+\tau} = \frac{x_t x_{t+\tau} + x_{t+1} x_{t+2} + \dots + x_{t+\tau-1} x_{t+\tau}}{\tau}$$

que representam a correlação linear de desfasamento τ que não é induzida pela correlação entre as variáveis de desfasamento intermédio 1, 2, ..., $\tau-1$.

Tais coeficientes podem ser calculados a partir da matriz de autocorrelação (V. Cramer, 1958, pp.305-7), quando finito o número de variáveis.

NOTA 12. Estudo de um caso: previsão inflacionária

O objectivo fundamental do trabalho "Aplicação de um método de previsão bayesiana ao IPC" foi o de mostrar como a previsão quantitativa de curto prazo difere radicalmente do mero ajustamento de modelos, tal como é usual em técnicas de análise com objectivos descritivos.

Nos finais de 1985 e princípios de 1986 a situação foi particularmente complexa do ponto de vista da previsão do IPC. O ritmo de inflação sofreu um abrandamento, por um conjunto de motivos já na altura conhecidos: adesão às comunidades, deliberadas medidas de política económica, alterações nas expectativas dos agentes económicos.

A série estudada inicia-se em Janeiro de 1983 e prolonga-se até Outubro de 1986. Como momento zero, a partir do qual se estudam previsões, considera-se Outubro de 1985. O período de previsão estende-se de Novembro desse ano ($t=1$) até Outubro de 1986 ($t=12$). A série original foi dessazonalizada, com coeficientes multiplicativos, sendo todos os raciocínios efectuados a partir da nova série, designada por (I_t) . Procurou-se assim uma maior simplicidade e clareza na análise dos diversos modelos sem se introduzirem problemas adicionais, irrelevantes para os propósitos do trabalho.

MODELOS DE PREVISÃO ESTUDADOS

A previsão econométrica típica resulta da linearização e pos-

terior ajustamento do modelo:

$$\text{MODELO 1} \quad I_t = a \cdot \exp(rt + u_t)$$

em que r é a taxa de crescimento teórica, instantânea, desse índice, t é variável discreta com incremento unitário mensal e u_t é variável aleatória residual. Admita-se, que u_t é v.a. de média nula, variância constante e que não é autocorrelacionada.

A simples análise gráfica revela serem naturalmente exageradas as previsões resultantes do modelo (1), pelo que o agente decisor é provavelmente levado a "deslocar" a curva ajustada obrigando-a, por exemplo, a conter o valor registado para I_0 , seja por um processo de estimação com restrições, seja retendo do modelo (1) apenas o parâmetro estimado r .

$$\text{MODELO 2} \quad I_t = I_0 \cdot \exp(rt)$$

Este novo modelo, confrontado com o rigor teórico do procedimento anterior, afigura-se eclético e de fundamentação estatística duvidosa - mas produz melhores estimativas, como se verifica pela observação do Erro Quadrático Médio.

Admita-se a mudança de variáveis que transforma cada índice na sua taxa de crescimento discreta empírica

$$(I_t - I_{t-1}) / I_{t-1} = y_t$$

conduzindo ao modelo teórico extremamente simples

$$y_t = r + v_t$$

em que se inclui, aditivamente, a variável aleatória v_t que se considera de média nula.

Trata-se de modelo funcionalmente equivalente aos modelos anteriores se as componentes aleatórias forem identicamente nulas. Como modelo estocástico derivado do modelo (1) possui variância superior.

A solução econométrica é trivial. O processo dos mínimos quadrados, admitindo que v_t é não autocorrelacionada e tem variância constante, conduz à equação preditiva

$$\text{MODELO 3} \quad r = y_i/n$$

obtendo-se

$$I_t = (1+r)^t \cdot I_0$$

Trata-se de modelo homólogo ao modelo (2), em que se fixou I_0 .

Foi igualmente testado o método do alisamento exponencial.

Esse modelo preditivo poderá ser assim expresso:

$$\text{MODELO 4} \quad r_{t+1} = r_t + a \cdot (r_t - r_t) \quad \text{com } 0 < a < 1$$

$$y_t = r_t$$

Foram calculadas recursivamente as estimativas do modelo (4) e feita a previsão para o momento seguinte, desde

t=1 (Nov.85) até t=12 (Out.86), com $\alpha=.05$.

Um outro modelo estudado foi o de espaço de estados:

MODELO 5
$$r_t = r_t + v_t$$

$$r_t = g \cdot r_{t-1} + w_t \quad (5)$$

a que foi aplicado o Filtro de Kalman.

Foi admitido que as variâncias são constantes e possuem os valores $V[v]=V[w]=1$. Introduzindo $\sigma_t^2=1$ (valor elevado que traduz uma "prior" relativamente difusa) e $g=.95$, obtiveram-se as estimativas com menor Erro Quadrático Médio.

O quadro síntese de valores e resultados é o que em seguida se reproduz.

| Meses | IPC | It | Previsões dos Modelos | | | | | |
|-----------------------|-------|--------|-----------------------|---------|---------|--------|--------|--|
| | | | (1) | (2) | (3) | (4) | (5) | |
| | 355.0 | | | | | | | |
| JAN 83 | 369.2 | 367.15 | | | | | | |
| FEB | 377.2 | 373.59 | | | | | | |
| MAR | 386.3 | 379.72 | | | | | | |
| ABR | 395.1 | 387.94 | | | | | | |
| MAY | 394.3 | 394.37 | | | | | | |
| JUN | 401.4 | 403.81 | | | | | | |
| JUL | 410.1 | 412.64 | | | | | | |
| AGO | 423.4 | 425.10 | | | | | | |
| SEP | 438.1 | 441.73 | | | | | | |
| OCT | 446.4 | 451.92 | | | | | | |
| NOV | 456.9 | 461.60 | | | | | | |
| DEC | 475.3 | 477.26 | | | | | | |
| JAN 84 | 482.4 | 479.72 | | | | | | |
| FEB | 490.3 | 485.60 | | | | | | |
| MAR | 509.0 | 500.33 | | | | | | |
| ABR | 518.9 | 509.50 | | | | | | |
| MAY | 514.1 | 514.19 | | | | | | |
| JUN | 530.4 | 533.59 | | | | | | |
| JUL | 544.9 | 548.27 | | | | | | |
| AGO | 555.0 | 557.22 | | | | | | |
| SEP | 556.7 | 561.31 | | | | | | |
| OCT | 561.7 | 568.65 | | | | | | |
| NOV | 570.8 | 576.67 | | | | | | |
| DEC | 576.3 | 578.68 | | | | | | |
| JAN 85 | 600.9 | 597.56 | | | | | | |
| FEB | 617.4 | 611.49 | | | | | | |
| MAR | 627.4 | 616.72 | | | | | | |
| ABR | 634.4 | 622.91 | | | | | | |
| MAY | 636.2 | 636.31 | | | | | | |
| JUN | 637.0 | 640.83 | | | | | | |
| JUL | 639.2 | 643.16 | | | | | | |
| AGO | 644.3 | 646.88 | | | | | | |
| SEP | 645.0 | 650.34 | | | | | | |
| OCT | 651.7 | 659.77 | | | | | | |
| NOV | 662.2 | 669.00 | 710.16 | 672.18 | 671.57 | 668.21 | 666.46 | |
| DEC | 673.4 | 676.18 | 723.52 | 684.83 | 683.59 | 677.60 | 676.87 | |
| JAN 86 | 686.3 | 682.49 | 737.14 | 697.72 | 695.83 | 684.79 | 683.28 | |
| FEB | 695.4 | 688.74 | 751.01 | 710.85 | 708.28 | 691.06 | 688.81 | |
| MAR | 703.4 | 691.42 | 765.14 | 724.22 | 720.96 | 697.27 | 694.72 | |
| ABR | 712.6 | 699.69 | 779.54 | 737.85 | 733.86 | 699.70 | 695.21 | |
| MAY | 709.1 | 709.23 | 794.21 | 751.74 | 746.99 | 708.06 | 705.90 | |
| JUN | 711.3 | 715.58 | 809.15 | 765.88 | 760.36 | 717.76 | 717.07 | |
| JUL | 711.1 | 715.50 | 824.38 | 780.29 | 773.97 | 724.07 | 722.19 | |
| AGO | 718.1 | 720.98 | 839.89 | 794.98 | 787.82 | 723.57 | 717.93 | |
| SEP | 720.9 | 726.87 | 855.70 | 809.94 | 801.92 | 728.97 | 725.05 | |
| OCT | 725.7 | 734.68 | 871.80 | 825.18 | 816.27 | 734.82 | 731.80 | |
| Erro Quadratico Medio | | | 8331.61 | 2703.17 | 2189.07 | 11.52 | 9.80 | |

REFERÊNCIAS ADICIONAIS

- APOSTOL, Tom A. (1974) *Mathematical Analysis (Sec Ed)*, Addison-Wesley
- ASH, R. B. (1972) *Real Analysis and Probability*, Academic Press
- BARTLETT, M.S. (1955) *An Introduction to Stochastic Processes with Special Reference to Methods and Applications*, Cambridge University Press, London.
- BERBERIAN, Sterling K. (1965,1970). *Measure and Integration*, Chelsea Publishing Company, New York.
- CHATFIELD, C. (1975, 1984) *The Analysis of Time Series : An Introduction (3rd Ed)*, Chapman and Hall, London.
- CRAMER, Harald (1946) *Mathematical Methods of Statistics*, Princeton University Press, Princeton.
- JENKINS, G.M. e WATTS, D.G. (1968) *Spectral Analysis and its Applications*, Holden-Day, San Francisco.
- KOLMOGOROV, A. N. e FOMIN, S. V. (1976, 1982) *Elementos de Teoria das Funções e de Análise Funcional*, Mir, Moscovo.
- MURTEIRA, Bento J. F. (1982) *Elementos sobre a Teoria dos Processos Estocásticos*, Instituto Superior de Economia, Lisboa.
- PAPDULIS, Athanasios (1965, 1984) *Probability, Random Variables, and Stochastic Processes*, McGraw-Hill.
- ROZANOV, Yuri (1973) *Procesos Aleatorios*, MIR, Moscovo.
- TYKHOMIROV, V.M. (1952) "Algebras de Banach", in KOLMOGOROV, A. N. e FOMIN, S. V. (1976, 1982) *Elementos de Teoria das Funções e de Análise Funcional*, Mir, Moscovo.
- VENTSEL, Hélène (1973) *Théorie des Probabilités*, Ed. MIR, Moscovo.

NOTAS ANEXAS