(2)

ESTIMADORES ESPECTRAIS
AUTO-REGRESSIVOS

AMILTON BRAIO ARA

DISSERTAÇÃO APRESENTADA

AO

INSTITUTO DE MATEMÁTICA E ESTATÍSTICA

DA

UNIVERSIDADE DE SÃO PAULO

PARA OBTENÇÃO DO GRAU DE MESTRE

EM

ESTATÍSTICA

ORIENTADOR: Prof. Dr. PEDRO ALBERTO MORETTIN

- São Paulo, maio de 1982 -

À meus Pais

AGRADECIMENTOS

Ao Professor Doutor Pedro Alberto Morettin, pela dedicação com que realizou a árdua tarefa de orientação, propiciando sempre um ambiente agradável e amigo.

Ao Professor Doutor Clóvis de Araujo Peres, pelo incentivo recebido desde o início do curso de Mestrado e pela sua orientação durante o período em que fui bolsista.

Ao colega e amigo Francisco Alberto Pino, pela ajuda na parte de computação.

As colegas Clélia Maria de Castro Toloi e Lúcia Pereira da Silva, pelo apoio constante.

Ao Professor Doutor Afranio Rubens de Mesquita, pelas valiosas sugestões.

Ao Sr. João Baptista Esteves de Oliveira, pelo excelente trabalho de datilografia

Í N D I C E

CAP. 1 - INTRODUÇÃO	1
1.1 - Preliminares	1
1.2 - Objetivos do Trabalho.	2
1.3 - Definições	3
1.4 - Estimadores Espectrais	8
CAP. 2 - PROCESSO AUTO-REGRESSIVO	14
2.1 - Definição.	14
2.2 - Condições de Estacionaridade para um Processo Auto- Regressivo	16
2.3 - Função de Auto-correlação e Espectro de um Processo Auto-regressivo.	21
2.3.1 - Função de auto-correlação.	21
2.3.2 - Equações de Yule-Walker.	29
2.3.3 - Variância.	31
2.3.4 - Espectro	32
CAP. 3 - ESTIMADORES DOS PARÂMETROS DO PROCESSO AR.	34
3.1 - Introdução	34
3.2 - Estimativas de Máxima Verossimilhança dos Parâmetros no Caso de Distribuição Normal - Estimativas de Yule- Walker.	36
3.3 - Propriedades dos Estimadores dos Parâmetros de um Processo AR(M)	48
3.4 - Estimação da Função Densidade Espectral e Propriedades Assintóticas do Estimador Espectral Auto-Regressivo	58
CAP. 4 - DETERMINAÇÃO DA ORDEM DO PROCESSO AR A SER AJUSTADO À SÉ- RIE DE TEMPO	66
4.1 - Introdução	66
4.2 - O Método de Anderson	67
4.3 - O Método de Akaike	71
4.3.1 - Definição de FPE de um processo.	71
4.3.2 - FPE das estimativas de mínimos quadrados de um modelo auto-regressivo.	76
4.3.3 - Uma estimativa de $(FPE)_M$ e o procedimento de mínimo FPE	86
4.3.4 - Propriedades estatísticas de $(FPE)(M)$	92

4.4 - O Método de Parzen	100
4.4.1 - Definição do CAT de um previsor.	100
4.4.2 - Base teórica do critério	101
CAP. 5 - DETERMINAÇÃO DA ORDEM DE UM MODELO ARMA A SER AJUSTADO À SÉRIE DE TEMPO	107
5.1 - Introdução	107
5.2 - Processo de Médias Móveis (MA)	108
5.3 - Processos Mistos Auto-regressivos - Médias Móveis (ARMA)	109
5.4 - O Método de Akaike	111
5.4.1 - Definição do Critério de Informação de Akaike (AIC)	111
5.4.2 - A média do log-verossimilhança como uma medida do ajuste	112
5.4.3 - Procedimento MAICE	112
5.5 - O Método de Gray, Kelley e McIntire.	120
5.5.1 - Definições e Teoremas.	120
5.5.2 - Determinação de ordem através das quantidades R e S.	124
5.5.3 - A estatística D(n,m)	133
CAP. 6 - APLICAÇÕES	136
6.1 - Introdução	136
6.2 - Resultados dos Estudos de Simulação.	136
6.3 - Aplicação do MME à série de Manchas Solares	142
6.4 - Comparação dos Métodos de Determinação da Ordem dos Processos AR e ARMA.	149
6.5 - Comentários Finais	158
APÊNDICE 1 - ESTIMADORES AR E O CONCEITO DE ENTROPIA.	160
A1.1 - Método de Máxima Entropia	160
A1.2 - Informação e Entropia	160
A1.3 - A Correspondência entre um Processo AR e MME.	165
APÊNDICE 2 - ESTIMATIVAS DE BURG DOS COEFICIENTES AR.	170
BIBLIOGRAFIA.	173

CAPÍTULO 1

INTRODUÇÃO

1.1 - PRELIMINARES

Os processos auto-regressivos (AR) foram usados pela primeira vez por Yule (1927) no estudo da série das manchas solares de Wolfer. Na década de 30 os processos AR apareceram em Economia, quando, com os trabalhos de Frisch (1933) e Tinberger (1937), introduziu-se o uso de elementos estocásticos em modelos dinâmicos de sistemas econômicos.

A aplicação dos processos AR à análise espectral, a qual ocupa-se da estimação do espectro de um processo estocástico estacionário, foi sugerida por Parzen (1969). O método tradicional de estimação do espectro baseado na análise harmônica ou de Fourier é não paramétrico, isto é, não é assumido a priori nenhum modelo. Parzen (1969) sugeriu que, usando o processo AR, a estimação espectral fosse feita através de um modelo paramétrico. Tal estimador é conhecido como estimador espectral auto-regressivo. Esse estimador é idêntico ao estimador baseado no princípio da máxima entropia introduzido por Burg (1967, 1968).

O método de estimação espectral AR geralmente produz um espectro mais suavizado do que aquele obtido pela análise es-

pectral tradicional, e os exemplos dados por Jones (1974) e Makhoul (1975) sugerem que as estimativas espectrais AR podem ser mais facilmente interpretadas. Embora produza um espectro suave, o método também tem "alta resolução", isto é, ele pode detectar picos estreitos no espectro (Griffiths & Pietro Diaz, 1977). O método parece útil para dados mostrando uma variação cíclica com mudanças aleatórias na amplitude, na fase e na frequência (veja Tong, 1977).

O problema crucial passou a ser o da determinação da ordem do processo AR e ser usado, isto é, a determinação do número de parâmetros do modelo. Anderson (1963) desenvolveu um procedimento de testes de hipóteses; Akaike (1969) definiu um critério de estimação da ordem do processo AR denominado FPE (erro final de previsão), o qual minimiza a variância esperada do erro de previsão; Parzen (1974) sugeriu o critério CAT (critério de função de transferência auto-regressiva), que minimiza o erro quadrático médio quando se supõe que o verdadeiro modelo é auto-regressivo infinito.

1.2 - OBJETIVOS DO TRABALHO

Os principais objetivos deste trabalho são:

- a) apresentação do método de estimação espectral auto-regressiva ou método da máxima entropia e comparação com o método tradicional de suavização da função de auto-covariância.

b) Apresentação e comparação dos vários métodos de determinação da ordem dos processos AR e ARMA (autoregressivos-médias móveis).

c) Aplicar o método da máxima entropia a séries reais.

Para a comparação dos métodos de estimação espectral e dos métodos de determinação da ordem, foram simuladas várias séries. Para cada série simulada determinamos as estimativas espectrais AR e de suavização do espectro e aplicamos os vários métodos de determinação da ordem.

Foram simuladas séries geradas por processos AR de ordens 1, 2 e 5 e por um processo ARMA(3,2). Para cada um desses processos foram simuladas 10 séries "longas" com 300 pontos e 10 séries "curtas" com 50 pontos. Em cada série simulada foram ignorados os primeiros 200 pontos para eliminar os efeitos dos valores iniciais iguais a zero.

1.3 - DEFINIÇÕES

DEFINIÇÃO 1.1 - PROCESSO ESTOCÁSTICO

Seja (Ω, A, P) um espaço de probabilidades. Uma família indexada de variáveis aleatórias (a valores reais ou complexos), $\{X(t, \omega), t \in T, \omega \in \Omega\}$ definida nesse espaço, é chamada um processo estocástico (real ou complexo). Se T for o conjunto dos inteiros, o processo é discreto. Se T for o conjunto dos reais, o processo é contínuo.

Então, para t fixo $X(t, \omega)$ é uma variável aleatória e para ω fixo $X(t, \omega)$ é uma particular realização do processo es-

tocástico.

DEFINIÇÃO 1.2 - SÉRIE TEMPORAL

Uma série temporal é um conjunto de observações $\{X(t), t \in T\}$ geradas seqüencialmente no tempo. Se os valores futuros de uma série temporal são perfeitamente determinados por uma função matemática do tempo, dizemos que ela é determinística. Se os valores futuros puderem ser determinados apenas por uma distribuição de probabilidade, dizemos que ela é não determinística ou estocástica.

Logo, as séries temporais não-determinísticas, as quais serão o nosso objeto de estudo, podem ser consideradas como uma particular realização de um processo estocástico.

EXEMPLO 1.1 - São exemplos de séries temporais:

- a) Série das temperaturas médias mensais em São Paulo.
- b) Série das alturas de marés em Ubatuba.
- c) Série das manchas solares de Wolfer.
- d) Série dos preços de café pagos aos produtores de São Paulo.

DEFINIÇÃO 1.3 - PROCESSO ESTACIONÁRIO

Dizemos que um processo é *estritamente estacionário* se o seu "mecanismo gerador" não varia no decorrer do tempo, isto é, se as funções de distribuição permanecem as mesmas sob translações no tempo.

Então, um processo estocástico é *estritamente estacionário* se

$$P(X(t_1+k) \in S_1, \dots, X(t_N+k) \in S_N) = P(X(t_1) \in S_1, \dots, X(t_N) \in S_N), (1.1)$$

para todo $t_1 < \dots < t_N$, eventos reais $S_1, \dots, S_N$ e $-\infty < k < \infty$.

Num processo estritamente estacionário, temos

$$E[X(t)] = E[X(0)] = m, \quad -\infty < t < \infty \quad (1.2)$$

e

$$E[X(t+k)X(t)] = E[X(k)X(0)] = \gamma_k, \quad -\infty < t < \infty \quad (1.3)$$

Isto é, os valores médios são constantes no tempo e podemos supor, sem perda de generalidade, que $m = 0$ e as covariâncias dependem apenas do deslocamento de tempo k , mas não do instante t .

A função γ_k chamamos de *função de auto-covariância* do processo e observando que $\gamma_0 = E\{[X(t)]^2\} = \text{Var}[X(t)]$, chamamos $\rho_k = \frac{\gamma_k}{\gamma_0}$ de *função de auto-correlação* do processo.

Se nós substituirmos a condição (1.1) pela suposição de que as variáveis aleatórias do processo tenham a propriedade

$$\text{Var}[X(t)] = \gamma_0 < \infty \quad (1.4)$$

e ainda satisfaçam as condições (1.2) e (1.3), então dizemos que o processo é *fracamente estacionário* ou *estacionário de segunda ordem*. No restante do texto, chamaremos simplesmente de processos estacionários, aos processos fracamente estacionários.

DEFINIÇÃO 1.4 - RUÍDO BRANCO

Chamamos de ruído branco a uma sequência $\epsilon(t)$, $t = 0, \pm 1, \dots$

de variáveis aleatórias independentes e identicamente distribuídas, com média zero e variância σ^2 . Isto é,

$$E[\varepsilon(t)] = 0$$

e

$$E[\varepsilon(s)\varepsilon(t)] = \begin{cases} \sigma^2, & \text{se } s = t \\ 0, & \text{se } s \neq t \end{cases}$$

DEFINIÇÃO 1.5 - ESPECTRO

Seja $X(t)$ um processo estocástico estacionário, com função de auto-covariância γ_k . Então, o *espectro* de $X(t)$ é definido como a transformada de Fourier de γ_k . Consideremos os dois casos seguintes:

a) CASO DISCRETO

Seja $\{X(t), t \in \mathbb{Z}\}$ um processo estocástico estacionário, com função de auto-covariância γ_k . Se estiver satisfeita a condição

$$\sum_{k=-\infty}^{\infty} |\gamma_k| < \infty, \quad (1.5)$$

isto é, se $\gamma_k \rightarrow 0$, quando $|k| \rightarrow \infty$, então o espectro de $X(t)$ é dado por

$$f(\lambda) = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} \gamma_k e^{-i\lambda k}, \quad -\pi \leq \lambda \leq \pi \quad (1.6)$$

b) CASO CONTÍNUO

Seja o processo $\{X(t), t \in \mathbb{R}\}$ de parâmetro contínuo. Se a con-

dição (1.5) for substituída por

$$\int_{-\infty}^{\infty} \gamma(u) du < \infty, \quad (1.7)$$

o espectro de $X(t)$ é dado por

$$f(\lambda) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \gamma(u) e^{-i\lambda u} du, \quad -\infty < \lambda < \infty \quad (1.8)$$

Pode ser demonstrada (veja Morettin, 1979) a seguinte propriedade, para o caso discreto.

PROPOSIÇÃO - O espectro $f(\lambda)$ é limitado, não negativo e uniformemente contínuo. Além disso, $f(\lambda)$ é par e periódico de período 2π .

Como $f(\lambda)$ é par e periódico de período 2π , podemos nos restringir a representá-lo apenas no intervalo $[0, \pi]$.

DEFINIÇÃO 1.6 - REPRESENTAÇÃO ESPECTRAL DE UM PROCESSO ESTACIONÁRIO

Seja $X(t)$, $-\infty < t < \infty$, um processo estacionário real com $E[X(t)] = 0$ e função de auto-covariância $\gamma_k = E[X(t+k)X(t)]$, tal que γ_k é contínua para $k = 0$. Sob essas suposições, existe uma função de conjunto aleatória a valores complexos, $Z(\Lambda)$, chamada *medida espectral aleatória* do processo e uma interpretação de integral tal que o processo tem a *representação espectral*

$$X(t) = \int_{-\infty}^{\infty} e^{i\lambda t} Z(d\lambda).$$

A medida espectral aleatória $Z(A)$ satisfaz a seguinte propriedade:

$$E[Z(d\lambda)\overline{Z(d\mu)}] = \begin{cases} F(d\lambda), & \mu = \lambda \\ 0, & \mu \neq \lambda, \end{cases}$$

onde $F(A)$ é chamada a *distribuição espectral* de $X(t)$ e

$$F(d\lambda) = f(\lambda)d\lambda.$$

1.4 - ESTIMADORES ESPECTRAIS

Do ponto de vista prático, a análise espectral é baseada na conversão dos dados indexados ao tempo, em estimativa do espectro.

Os métodos tradicionais de estimação espectral introduzidos por Bartlett (1948) e por Tukey & Hamming (1949), descritos a seguir, são conhecidos, no domínio do tempo, como estimadores suavizados de covariâncias e, no domínio de frequências, como estimadores suavizados de periodograma.

Dados os valores $X(1), X(2), \dots, X(N)$ do processo estacio-

nário real $X(t)$, $t = 0, \pm 1, \pm 2, \dots$ definimos a transformada de Fourier finita desses valores como

$$d^{(N)}(\lambda) = \frac{1}{\sqrt{2\pi N}} \sum_{t=1}^N X(t) e^{-i\lambda t}, \quad -\infty < \lambda < \infty$$

Como $d^{(N)}(\lambda)$ é periódica de período 2π , podemos nos restringir às frequências de forma $\lambda_v = \frac{2\pi v}{N}$, com $-\left[\frac{N-1}{2}\right] \leq v \leq \left[\frac{N}{2}\right]$; então, escrevemos

$$d_v^{(N)} = \frac{1}{\sqrt{2\pi N}} \sum_{t=1}^N X(t) e^{-i\lambda_v t}$$

O periodograma, introduzido por Schwter (1894) é o estimador de $f(\lambda_v)$ definido por

$$I_v^{(N)} = |d_v^{(N)}|^2 = \frac{1}{2\pi N} \left| \sum_{t=1}^N X(t) e^{-i\lambda_v t} \right|^2, \quad (1.9)$$

para frequência $\lambda_v = \frac{2\pi v}{N}$, $-\left[\frac{N-1}{2}\right] \leq v \leq \left[\frac{N}{2}\right]$.

A sua média e variância assintóticas são

$$E[I_v^{(N)}] = f(\lambda_v)$$

e

$$\text{Var}[I_v^{(N)}] = f^2(\lambda_v), \quad v \neq 0, \quad v \neq \frac{N}{2}$$

$$\text{Var}[I_v^{(N)}] = \begin{cases} 2f^2(0), & v = 0 \\ 2f^2(\pi), & v = \frac{N}{2} \end{cases}$$

Logo, embora o periodograma seja assintoticamente não vi-

ciado ele é muito instável. Para se obter estimadores que além de serem assintoticamente não viciados, sejam mais estáveis, foram introduzidos os estimadores suavizados do espectro.

Um estimador natural do espectro

$$f(\lambda) = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} \gamma_k e^{-i\lambda k}, \quad -\pi \leq \lambda \leq \pi,$$

é obtido substituindo-se as auto-covariâncias γ_k pelas suas estimativas $C(k)$ dadas pela relação (3.27) do Capítulo 3. No entanto, pode ser mostrado que

$$\frac{1}{2\pi} \sum_{k=-\infty}^{\infty} e^{-i\lambda k} C(k) = I^{(N)}(\lambda).$$

Logo esse estimador coincide com o periodograma anterior.

Tukey (1949) introduziu, então, o *estimador suavizado de covariâncias*

$$\hat{f}(\lambda) = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} e^{-i\lambda k} w_M(k) C(k), \quad -\pi \leq \lambda \leq \pi \quad (1.10)$$

onde, para cada inteiro $M < N$, $w_M(k)$, $k = 0, \pm 1, \pm 2, \dots$ é uma sequência de pesos satisfazendo:

- i) $0 \leq w_M(k) \leq w_M(0) = 1$
- ii) $w_M(-k) = w_M(k)$, para todo k
- iii) $w_M(k) = 0$, $|k| > M$.

A janela espectral correspondente à função peso (núcleo) $w_M(k)$

é definida por

$$W_M(k) = \frac{1}{2\pi} \sum_k e^{-i\lambda k} w_M(k)$$

e satisfaz

- i) $W_M(-\lambda) = W_M(\lambda)$ para todo λ
- ii) $\int_{-\pi}^{\pi} W_M(\lambda) d\lambda = w_M(0) = 1.$

Da relação (1.10) obtemos que $\hat{f}(\lambda)$ é a convolução das transformadas de Fourier de $w_M(k)$ e $C(k)$, isto é,

$$\hat{f}(\lambda) = \int_{-\pi}^{\pi} W_M(\lambda - \alpha) I^{(N)}(\alpha) d\alpha. \quad (1.11)$$

Substituindo a integral em (1.11) por suas somas de Riemann, obtemos o *estimador suavizado de periodograma*

$$\hat{f}(\lambda) = \sum_{v=-\left[\frac{N-1}{2}\right]}^{\left[\frac{N}{2}\right]} W(\lambda - \lambda_v) I_v^{(N)}, \quad (1.12)$$

onde $W(\lambda)$ é uma função peso real, par, periódica e tal que

$$\sum_{v=-\left[\frac{N-1}{2}\right]}^{\left[\frac{N}{2}\right]} W(\lambda_v) = 1.$$

Os estimadores suavizados de covariância e os estimadores suavizados de periodograma são assintoticamente equivalentes.

Se a função peso $w_M(k)$ for uma função contínua tal que

$$w_M(k) = \begin{cases} w\left(\frac{k}{M}\right), & |k| \leq M \\ 0, & |k| > M \end{cases}$$

e $0 \leq w\left(\frac{k}{M}\right) \leq w(0) = 1$, onde M é o ponto de truncamento, então os estimadores suavizados do espectro são assintoticamente não viesados, com variância

$$\text{Var}[\hat{f}(\lambda_k)] \approx f^2(\lambda_k) c_w \cdot \frac{M}{N}$$

onde

$$c_w = \int_{-1}^1 w^2(v) dv, \quad v = \frac{k}{M}.$$

Então se M é controlado de modo que $\frac{M}{N} \rightarrow 0$, os estimadores são consistentes.

Finalmente, $\hat{f}(\lambda_k)$ tem uma distribuição aproximada

$$\frac{f(\lambda_k)}{r} \chi^2(r),$$

onde r é o número equivalente de graus de liberdade e é dado por

$$r = \frac{N}{\pi \int_{-\pi}^{\pi} W_M^2(\lambda) d\lambda},$$

se $\hat{f}(\lambda_k)$ for o estimador suavizado de covariâncias e

$$r = \frac{2}{\sum_{v=-\left[\frac{N-1}{2}\right]}^{\left[\frac{N}{2}\right]} W^2(\lambda_v)},$$

se $\hat{f}(\lambda_k)$ for o estimador suavizado de periodograma.

CAPÍTULO 2

PROCESSO AUTO-REGRESSIVO

2.1 - DEFINIÇÃO

Seja $\epsilon(t)$, $t = 0, \pm 1, \pm 2, \dots$ um ruído branco com média zero e variância σ^2 e sejam $a_1, a_2, \dots, a_p$ números reais constantes tais que $a_p \neq 0$. Então dizemos que $X(t)$ é um *processo auto-regressivo de ordem p*, $AR(p)$, se $X(t)$ for um processo fracamente estacionário satisfazendo a equação de diferenças

$$X(t) = a_1 X(t-1) + a_2 X(t-2) + \dots + a_p X(t-p) + \epsilon(t), \quad t = 0, \pm 1, \pm 2, \dots \quad (2.1)$$

e para o qual

$$E[X(s)\epsilon(t)] = 0, \text{ para todo } s \leq t-1, \quad t = 0, \pm 1, \pm 2, \dots \quad (2.2)$$

Então, o valor da auto-regressão em qualquer instante t consiste de uma função linear dos valores do processo para as p observações anteriores, mais um ruído aleatório ou *ínovação* $\epsilon(t)$, o qual é não correlacionado com os valores passados do processo (condição (2.2)). Se $X(t-1), \dots, X(t-p)$ fossem funções conhecidas fixadas, então (2.1) seria um modelo de regressão usual. Contudo, a regressão do processo é feita sobre os seus próprios valores, justificando-se então o nome auto-regressão.

Analogamente, um processo é dito auto-regressivo de or-

dem infinita se satisfaz a condição (2.2) e a equações da forma

$$X(t) = \sum_{k=1}^{\infty} a_k X(t-k) + \varepsilon(t), \quad t = 0, \pm 1, \pm 2, \dots \quad (2.3)$$

onde

$$\sum_{k=1}^{\infty} a_k^2 < \infty.$$

São muito úteis, na prática, os modelos auto-regressivos de 1^a ordem ($p=1$) e de 2^a ordem ($p=2$), respectivamente,

$$X(t) = a_1 X(t-1) + \varepsilon(t)$$

e

$$X(t) = a_1 X(t-1) + a_2 X(t-2) + \varepsilon(t).$$

Usando o operador translação para o passado, B , definido por

$$BX(t) = X(t-1) \quad \text{e} \quad B^m X(t) = X(t-m), \quad \text{para } 1 \leq m \leq p,$$

a expressão (2.1) pode ser escrita na forma

$$(1 - a_1 B - a_2 B^2 - \dots - a_p B^p) X(t) = \varepsilon(t)$$

ou, sendo

$$A(B) = 1 - a_1 B - a_2 B^2 - \dots - a_p B^p,$$

$$A(B)X(t) = \varepsilon(t), \quad (2.4)$$

a qual é um filtro linear que transforma $X(t)$ em $\varepsilon(t)$.

Definimos a transformada - Z desse filtro linear, como sendo a função a valores complexos

$$A(Z) = 1 - \sum_{k=1}^p a_k Z^k$$

a qual está bem definida em todo campo complexo.

Então a função de transferência (função resposta da frequência) do filtro linear (2.4) é dada por

$$A(\lambda) = A(e^{-i\lambda}) = 1 - a_1 e^{-i\lambda} - a_2 e^{-2i\lambda} - \dots - a_p e^{-pi\lambda}.$$

Se $A(\lambda) \neq 0$, para todo λ , tal que $-\pi \leq \lambda \leq \pi$, a condição de invertibilidade está satisfeita e podemos escrever:

$$X(t) = A^{-1}(B)\varepsilon(t) \quad (2.5)$$

e o processo auto-regressivo, de ordem p , pode ser pensado como a saída $X(t)$ de um filtro linear, com função de transferência

$$A^{-1}(\lambda) = A^{-1}(e^{-i\lambda}) = \frac{1}{1 - a_1 e^{-i\lambda} - a_2 e^{-2i\lambda} - \dots - a_p e^{-pi\lambda}}$$

e entrada um ruído branco $\varepsilon(t)$.

2.2 - CONDIÇÕES DE ESTACIONARIDADE PARA UM PROCESSO AUTO-REGRESSIVO

Vejamos as condições que o conjunto de parâmetros $a_1, a_2, \dots, a_p$, de um processo $AR(p)$,

$$X(t) = a_1 X(t-1) + a_2 X(t-2) + \dots + a_p X(t-p) + \epsilon(t), \quad t=0, \pm 1, \pm 2, \dots$$

deve satisfazer para que o processo seja estacionário.

Já vimos que o processo AR(p) pode ser escrito nas seguintes formas equivalentes:

$$X(t) - \sum_{k=1}^p a_k X(t-k) = \epsilon(t),$$

$$\left(1 - \sum_{k=1}^p a_k B^k\right) X(t) = \epsilon(t),$$

$$A(B)X(t) = \epsilon(t),$$

onde

$$A(B) = 1 - \sum_{k=1}^p a_k B^k,$$

e, se a condição de invertibilidade estiver satisfeita, teremos:

$$X(t) = \frac{1}{A(B)} \epsilon(t).$$

Como a transformada - Z do filtro linear (2.4),

$$A(Z) = 1 - \sum_{k=1}^p a_k Z^k$$

é um polinômio, se a condição de invertibilidade estiver satisfeita, isto é, se a função de transferência $A(\lambda) = A(e^{-i\lambda})$ não se anular para $-\pi \leq \lambda \leq \pi$ (ou de modo equivalente, se $A(Z)$

não tiver zeros no círculo unitário), então $A(\lambda)$ é realmente limitada fora de uma vizinhança do zero.

Assim, escrevendo $|A(\lambda)| \geq M > 0$, vemos que o filtro linear com função de transferência $A(\lambda)$ pode ser invertido, pois:

$$\int_{-\pi}^{\pi} \left| \frac{1}{A(\lambda)} \right|^2 f_{\epsilon}(\lambda) d\lambda = \left(\frac{\sigma^2}{2\pi} \right) \int_{-\pi}^{\pi} \left| \frac{1}{A(\lambda)} \right|^2 d\lambda \leq \frac{\sigma^2}{M^2}.$$

Logo,

$$X(t) = \int_{-\pi}^{\pi} e^{i\lambda t} \left(\frac{1}{A(\lambda)} \right) Z_{\epsilon}(d\lambda)$$

é uma solução fracamente estacionária para a equação de diferenças (2.1). Além disso, como $\frac{1}{A(\lambda)}$ é uma função de quadrado integrável ela tem uma única expansão em série de Fourier dada por

$$\frac{1}{A(\lambda)} = \sum_{j=-\infty}^{\infty} b_j e^{-i\lambda j}, \text{ com } \sum_{j=-\infty}^{\infty} b_j^2 < \infty.$$

Então, substituindo na expressão de $X(t)$ acima, temos

$$\begin{aligned} X(t) &= \int_{-\pi}^{\pi} e^{i\lambda t} \sum_{j=-\infty}^{\infty} b_j e^{-i\lambda j} Z_{\epsilon}(d\lambda) = \\ &= \sum_{j=-\infty}^{\infty} b_j \int_{-\pi}^{\pi} e^{i\lambda(t-j)} Z_{\epsilon}(d\lambda) = \\ &= \sum_{j=-\infty}^{\infty} b_j \epsilon(t-j). \end{aligned}$$

Para que a condição (2.2) esteja satisfeita é necessário e

suficiente que:

$$E[X(s)\varepsilon(t)] = \sum_{j=-\infty}^{\infty} b_j E[\varepsilon(s-j)\varepsilon(t)] = 0, \text{ para } s \leq t-1.$$

Mas,

$$E[\varepsilon(s-j)\varepsilon(t)] = \begin{cases} \sigma^2, & \text{se } s-j = t \\ 0, & \text{se } s-j \neq t, \end{cases}$$

logo,

$$E[X(s)\varepsilon(t)] = b_{s-t}\sigma^2 = 0,$$

para $s \leq t-1$, implica que $b_{s-t} = 0$ para $s = t-1, t-2, \dots$ ou, de modo equivalente, $b_{-1} = b_{-2} = \dots = 0$.

Além disso,

$$E[X(t)\varepsilon(t)] = \sum_{j=-\infty}^{\infty} E b_j \varepsilon(t-j)\varepsilon(t) = b_0 \sigma^2.$$

Mas, das expressões (2.1) e (2.2) temos:

$$\begin{aligned} E[X(t)\varepsilon(t)] &= E\left\{\left[\sum_{j=1}^p a_j X(t-j) + \varepsilon(t)\right]\varepsilon(t)\right\} = \\ &= \sum_{j=1}^p a_j E[X(t-j)\varepsilon(t)] + E[\varepsilon^2(t)] = E[\varepsilon^2(t)] = \sigma^2. \end{aligned}$$

Logo, $b_0 = 1$ e portanto para que a condição (2.2) esteja satisfeita devemos ter

$$X(t) = \varepsilon(t) + \sum_{j=1}^{\infty} b_j \varepsilon(t-j).$$

A condição de estacionaridade pode ser enunciada univocamente em termos da transformada $-Z$ na seguinte forma:

Para que a equação de diferenças (2.1) tenha uma solução fracamente estacionária satisfazendo a condição (2.2), isto é, para que uma auto-regressão finita exista, é suficiente que $A(Z)$ tenha todos os seus zeros fora do círculo unitário. A solução será um processo de médias móveis com coeficientes iguais aos coeficientes da expansão em série de

$$\frac{1}{A(Z)} = B(Z) = 1 + \sum_{j=1}^{\infty} b_j Z^j.$$

No caso particular de um processo auto-regressivo de 1ª ordem

$$(1 - a_1 B)X(t) = \varepsilon(t), \quad (2.6)$$

cuja função de transferência $A(\lambda) = 1 - a_1 e^{-i\lambda}$, para $a_1 \neq \pm 1$, é diferente de zero para todo λ , $-\pi \leq \lambda \leq \pi$, a condição de invertibilidade está satisfeita e a expressão (2.6) pode ser escrita na forma:

$$X(t) = \left[\frac{1}{1 - a_1 B} \right] \varepsilon(t) \quad (2.7)$$

e como o 2º membro da expressão (2.7) acima é a soma da série geométrica $(1 + a_1 B + a_1^2 B^2 + \dots) \varepsilon(t)$, temos:

$$X(t) = (1 + a_1 B + a_1^2 B^2 + \dots) \varepsilon(t)$$

ou

$$X(t) = \sum_{j=0}^{\infty} a_1^j \epsilon(t-j)$$

e para que o processo AR(1) seja estacionário os zeros do polinômio $A(Z) = 1 - a_1 Z$ devem estar fora do círculo unitário. Logo, como

$$A(Z) = 0 \implies 1 - a_1 Z = 0 \implies Z = a_1^{-1},$$

devemos ter:

$$|a_1^{-1}| > 1, \text{ ou seja, } |a_1| < 1.$$

2.3 - FUNÇÃO DE AUTO-CORRELAÇÃO E ESPECTRO DE UM PROCESSO AUTO-REGRESSIVO

2.3.1 - Função de Auto-correlação

Consideremos um processo auto-regressivo de ordem p

$$X(t) = a_1 X(t-1) + a_2 X(t-2) + \dots + a_p X(t-p) + \epsilon(t), \quad t = 0, \pm 1, \pm 2, \dots$$

Multiplicando ambos os membros por $X(t-k)$, temos:

$$\begin{aligned} X(t-k)X(t) &= a_1 X(t-k)X(t-1) + a_2 X(t-k)X(t-2) + \\ &+ \dots + a_p X(t-k)X(t-p) + X(t-k)\epsilon(t). \end{aligned} \quad (2.8)$$

Tomando a esperança em ambos os membros e observando que $E[X(t-k)\epsilon(t)] = 0$ para $k > 0$, pois $X(t-k)$ envolve apenas os ruídos $\epsilon(j)$ relativos ao instante $t-k$ e instantes anteriores, os quais são não correlacionados com $\epsilon(t)$, obtemos a equação de

diferenças:

$$\gamma_k = a_1 \gamma_{k-1} + a_2 \gamma_{k-2} + \dots + a_p \gamma_{k-p}, \quad k > 0 \quad (2.9)$$

a qual é uma importante relação de recorrência para a função de auto-covariância de um processo AR(p) estacionário.

Dividindo ambos os membros da relação (2.9) por γ_0 , obtemos uma equação de diferenças da mesma forma para a função de auto-correlação:

$$\rho_k = a_1 \rho_{k-1} + a_2 \rho_{k-2} + \dots + a_p \rho_{k-p}, \quad k > 0 \quad (2.10)$$

ou ainda,

$$\rho_k - a_1 \rho_{k-1} - a_2 \rho_{k-2} - \dots - a_p \rho_{k-p} = 0, \quad k > 0 \quad (2.11)$$

a qual é uma equação de diferenças homogênea de ordem p, com coeficientes constantes. Portanto, se G^{-1} é uma raiz da equação característica correspondente

$$A(Z) = 1 - a_1 Z - a_2 Z^2 - \dots - a_p Z^p = 0,$$

então $\rho_k = G^k$ será uma solução da equação (2.11).

Logo, como a transformada - Z,

$$A(Z) = 1 - \sum_{j=1}^p a_j Z^j,$$

pode ser fatorada na forma

$$A(Z) = \prod_{j=1}^p (1 - G_j Z),$$

onde $G_1^{-1}, G_2^{-1}, \dots, G_p^{-1}$ são as p raízes da equação característica $A(Z) = 0$, temos que

$$\rho_k^{(1)} = G_1^k, \rho_k^{(2)} = G_2^k, \dots, \rho_k^{(p)} = G_p^k$$

são p soluções de equação (2.11) e portanto a solução geral dessa equação é da forma:

$$\rho_k = A_1 G_1^k + A_2 G_2^k + \dots + A_p G_p^k \quad (2.12)$$

onde $A_1, A_2, \dots, A_p$ são constantes reais ou complexas.

Supondo as raízes $G_1^{-1}, G_2^{-1}, \dots, G_p^{-1}$ distintas, dois casos podem ocorrer:

1º) Uma raiz G_j^{-1} é real.

2º) Um par de raízes G_i^{-1}, G_j^{-1} são conjugadas complexas.

Cada raiz real G_j^{-1} contribui para a solução (2.12) com um termo $A_j G_j^k$, onde A_j é uma constante real, e cada par de raízes complexas G_i^{-1}, G_j^{-1} , contribui para a solução (2.12) com um termo da forma $A_i G_i^k + A_j G_j^k$, onde A_i e A_j são constantes complexas.

Para que a função de auto-correlação seja real, cada termo correspondente a um par de raízes complexas deve ser real e, para que isso ocorra, devemos escolher os seus coeficientes A_i e A_j como sendo dois números complexos conjugados da forma

$$A_i = a(\cos\phi' + i \sin\phi') \quad \text{e} \quad A_j = a(\cos\phi' - i \sin\phi').$$

De fato, como as raízes G_i^{-1} e G_j^{-1} são conjugadas complexas, G_i e G_j também o serão e podemos escrever

$$G_i = d(\cos \lambda + i \sin \lambda) \quad \text{e} \quad G_j = d(\cos \lambda - i \sin \lambda)$$

então,

$$G_i^k = d^k(\cos k\lambda + i \sin k\lambda) \quad \text{e} \quad G_j^k = d^k(\cos k\lambda - i \sin k\lambda)$$

e portanto

$$\begin{aligned} A_i G_i^k + A_j G_j^k &= \text{ad}^k[\cos(k\lambda + \phi') + i \sin(k\lambda + \phi')] + \text{ad}^k[\cos(k\lambda + \phi') - \\ &\quad - i \sin(k\lambda + \phi')] = 2\text{ad}^k \cos(k\lambda + \phi'), \end{aligned}$$

o qual é um termo real.

Fazendo as constantes $A = 2a$, $\phi = \phi' + \frac{\pi}{2}$, e observando que $\cos(k\lambda + \phi') = \cos(k\lambda + \phi - \frac{\pi}{2}) = \sin(k\lambda + \phi)$, verificamos que cada par de raízes complexas contribui para a solução (2.12) com um termo da forma $\text{Ad}^k \sin(k\lambda + \phi)$.

Como a condição de estacionaridade é que as raízes de $A(Z) = 0$ estejam fora do círculo unitário, isto é, $|G_j^{-1}| > 1$ e portanto $|G_j| < 1$, para $j=1,2,\dots,p$, a função de auto-correlação de um processo AR(p), tem as seguintes características:

1º) Os termos da forma $A_i G_i^k$, correspondentes a raízes reais, decaem exponencialmente para zero quando k cresce. Este fato é conhecido como amortecimento exponencial da função de auto-correlação.

2º) Os termos da forma $A_i G_i^k + A_j G_j^k = \text{Ad}^k \sin(k\lambda + \phi)$, cor-

respondentes a pares de raízes complexas seguem ondas senóides amortecidas, pois

$$d^k = |G_i|^k = |G_j|^k < 1.$$

Portanto, em geral, a função de auto-correlação de um processo auto-regressivo consistirá de uma mistura de exponenciais amortecidas e de ondas senóides amortecidas.

Consideremos, como exemplos, os casos particulares dos processos auto-regressivos de 1^a e 2^a ordem.

EXEMPLO 2.1 - Processo auto-regressivo de 1^a ordem - Consideremos o processo

$$X(t) = a_1 X(t-1) + \epsilon(t), \quad t=0, \pm 1, \pm 2, \dots$$

onde $|a_1| < 1$, para satisfazer a condição de estacionaridade.

Função de auto-correlação

Usando a relação (2.11) no caso particular de $p=1$, verificamos que a função de auto-correlação satisfaz a equação de diferenças de 1^a ordem.

$$\rho_k = a_1 \rho_{k-1}, \quad k > 0$$

a qual, com $\rho_0 = 1$, tem a solução

$$\rho_k = a_1^k, \quad k \geq 0.$$

Então, conforme está ilustrado pela Figura 1, a função de auto-correlação de um processo auto-regressivo de 1^a ordem de-

cai exponencialmente para zero quando a_1 é positivo e decai exponencialmente para zero oscilando em sinal quando a_1 é negativo.

Figura 1

EXEMPLO 2.2 - Processo auto-regressivo de 2^a ordem - Consideremos o processo

$$X(t) = a_1 X(t-1) + a_2 X(t-2) + \epsilon(t), \quad t=0, \pm 1, \pm 2, \dots$$

Para que a condição de estacionaridade esteja satisfeita, os zeros do polinômio $A(Z) = 1 - a_1 Z - a_2 Z^2$ devem estar fora do círculo unitário.

É fácil ver que os parâmetros a_1 e a_2 devem estar na região triangular, da Figura 2, determinada pelas relações:

$$\begin{cases} a_2 + a_1 < 1 \\ a_2 - a_1 < 1 \\ -1 < a_2 < 1 \end{cases}$$

Figura 2

Da relação (2.11), temos

$$\rho_k = a_1 \rho_{k-1} + a_2 \rho_{k-2}, \quad k > 0$$

com valores iniciais $\rho_0 = 1$ e $\rho_1 = \frac{a_1}{1-a_2}$.

Da relação (2.12), a solução geral da equação de diferenças acima é

$$\rho_k = A_1 G_1^k + A_2 G_2^k$$

ou, determinando A_1 e A_2 em função de G_1 e G_2 ,

$$\rho_k = \frac{G_1(1-G_2^2)G_1^k - G_2(1-G_1^2)G_2^k}{(G_1-G_2)(1+G_1G_2)}$$

onde G_1^{-1} e G_2^{-1} são as raízes da equação característica

$$A(Z) = 1 - a_1Z - a_2Z^2 = 0$$

Quando as raízes são reais, e portanto $a_1^2 + 4a_2 \geq 0$, a função de auto-correlação consiste de uma mistura de exponenciais amortecidas. Isto ocorre para os valores dos parâmetros correspondentes às regiões 1 e 2 localizadas acima da parábola na Figura 2 e as funções de auto-correlação correspondentes a essas regiões tem respectivamente o aspecto daquelas mostradas na Figura 1-a e na Figura 1-b. O caso em que a raiz dominante é positiva corresponde à região 1 e o caso em que a raiz dominante é negativa corresponde à região 2.

Quando as raízes são complexas e portanto $a_1^2 + 4a_2 < 0$, a função de auto-correlação terá a forma de uma senóide amortecida dada por

$$\rho_k = \frac{d^k \text{sen}(k\lambda + \phi)}{\text{sen}\phi}$$

onde d é o fator de amortecimento, λ é a frequência e ϕ é a fase, e são determinados em função dos parâmetros do processo por

$$d = \sqrt{-a_2}$$

$$\cos \lambda = \frac{|a_1|}{2\sqrt{-a_2}}$$

$$\operatorname{tg} \phi = \frac{1+d^2}{1-d^2} \operatorname{tg} \lambda$$

Esse caso ocorre para os valores dos parâmetros correspondentes às regiões 3 e 4 localizadas abaixo da parábola na região triangular da Figura 2 e as funções de auto-correlação correspondentes a essas regiões estão representadas respectivamente nas Figuras 3-a e 3-b.

Figura 3

2.3.2 - Equações de Yule-Walker

Se nós substituirmos $k = 1, 2, \dots, p$ na relação (2.10), ob-

teremos o seguinte conjunto de equações lineares para $a_1, a_2, \dots, a_p$ em termos de $\rho_1, \rho_2, \dots, \rho_p$, denominadas *equações de Yule-Walker*.

$$\begin{cases} \rho_1 = a_1 + a_2 \rho_1 + \dots + a_p \rho_{p-1} \\ \rho_2 = a_1 \rho_1 + a_2 + \dots + a_p \rho_{p-2} \\ \vdots \\ \rho_p = a_1 \rho_{p-1} + a_2 \rho_{p-2} + \dots + a_p \end{cases} \quad (2.13)$$

Na forma matricial, fazendo-se

$$\underline{\tilde{A}} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{bmatrix}, \quad \underline{\tilde{\rho}}_p = \begin{bmatrix} \rho_1 \\ \rho_2 \\ \vdots \\ \rho_p \end{bmatrix} \quad \text{e} \quad \underline{P}_p = \begin{bmatrix} 1 & \rho_1 & \rho_2 & \dots & \rho_{p-1} \\ \rho_1 & 1 & \rho_1 & \dots & \rho_{p-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho_{p-1} & \rho_{p-2} & \rho_{p-3} & \dots & 1 \end{bmatrix}$$

temos

$$\underline{P}_p \underline{\tilde{A}} = \underline{\tilde{\rho}}_p, \quad (2.14)$$

logo a solução do sistema de equações (2.13) é dado por

$$\underline{\tilde{A}} = \underline{P}_p^{-1} \underline{\tilde{\rho}}_p. \quad (2.15)$$

As equações (2.13) são úteis na estimação dos parâmetros $a_1, a_2, \dots, a_p$ do modelo, substituindo-se as auto-correlações teóricas ρ_j por suas estimativas.

2.3.3 - Variância

Tomando a esperança em ambos os membros da relação (2.8) para $k = 0$, temos:

$$E[X(t)]^2 = a_1 E[X(t)X(t-1)] + a_2 E[X(t)X(t-2)] + \dots + a_p E[X(t)X(t-p)] + E[X(t)\epsilon(t)].$$

Como a única parte de $X(t)$ que será correlacionada com $\epsilon(t)$ é o último ruído $\epsilon(t)$, temos que

$$E[X(t)\epsilon(t)] = E[\epsilon(t)]^2 = \sigma^2.$$

Logo:

$$\gamma_0 = a_1 \gamma_{-1} + a_2 \gamma_{-2} + \dots + a_p \gamma_{-p} + \sigma^2$$

e como $\gamma_{-k} = \gamma_k$, temos:

$$\gamma_0 = a_1 \gamma_1 + a_2 \gamma_2 + \dots + a_p \gamma_p + \sigma^2 \quad (2.16)$$

e, ainda, dividindo ambos os membros por $\gamma_0 = \sigma_x^2$, temos:

$$\frac{\sigma^2}{\sigma_x^2} = 1 - a_1 \rho_1 - a_2 \rho_2 - \dots - a_p \rho_p,$$

logo, a variância do processo AR(p), $\sigma_x^2 = \text{var}[X(t)]$, é dada por:

$$\sigma_x^2 = \frac{\sigma^2}{1 - a_1 \rho_1 - a_2 \rho_2 - \dots - a_p \rho_p} \quad (2.17)$$

2.3.4 - Espectro

Conforme já vimos, um processo AR(p) pode ser pensado como a saída $X(t)$ de um filtro linear, com função de transferência $A^{-1}(\lambda)$ e entrada um ruído branco $\varepsilon(t)$. Logo o espectro do processo $X(t)$ será dado por:

$$f_{XX}(\lambda) = \frac{1}{|A(\lambda)|^2} f_{\varepsilon\varepsilon}(\lambda). \quad (2.18)$$

Por outro lado, o espectro do ruído branco $\varepsilon(t)$ é dado por:

$$f_{\varepsilon\varepsilon}(\lambda) = \frac{1}{2\pi} \sum_{\tau=-\infty}^{\infty} \gamma_{\varepsilon\varepsilon}(\tau) e^{-i\lambda\tau}$$

e como

$$\gamma_{\varepsilon\varepsilon}(\lambda) = \begin{cases} \sigma^2, & \text{se } \tau = 0 \\ 0, & \text{se } \tau \neq 0 \end{cases}$$

temos,

$$f_{\varepsilon\varepsilon}(\lambda) = \frac{\sigma^2}{2\pi}. \quad (2.19)$$

Logo, substituindo na expressão (2.18), temos:

$$f_{XX}(\lambda) = \frac{1}{|A(\lambda)|^2} \cdot \frac{\sigma^2}{2\pi} \quad (2.20)$$

e como

$$A(\lambda) = 1 - a_1 e^{-i\lambda} - a_2 e^{-2i\lambda} - \dots - a_p e^{-pi\lambda},$$

temos, finalmente,

$$f_{XX}(\lambda) = \frac{1}{2\pi} \cdot \frac{\sigma^2}{\left| 1 - \sum_{j=1}^p a_j e^{-i\lambda j} \right|^2}, \quad -\pi \leq \lambda \leq \pi. \quad (2.21)$$

CAPÍTULO 3

ESTIMADORES DOS PARÂMETROS DO PROCESSO AR

3.1 - INTRODUÇÃO

Nós vamos supor que a série de tempo sob observação seja a realização de um processo puramente não-determinístico, fracamente estacionário $X(t)$, com $E[X(t)] = 0$ e tal que a memória da história passada infinita de $X(t)$ seja nula, no sentido de média quadrática.

Nessas condições $X(t)$ tem a representação de um processo de médias móveis do tipo:

$$X(t) = \sum_{j=0}^{\infty} b_j \varepsilon(t-j) \quad (3.1)$$

onde $b_0 = 1$ e $\varepsilon(t)$ é um ruído branco, isto é, os $\varepsilon(t)$'s são variáveis aleatórias não correlacionadas com média 0 e variância σ^2 . Então $X(t)$ pode ser aproximado por:

$$X(t) \approx \sum_{m=1}^{\ell} a_{\ell m} X(t-m) + \varepsilon(t)$$

e a aproximação torna-se mais precisa com o aumento do valor de ℓ .

Já vimos também que as raízes de

$$A(Z) = 1 - \sum_{m=1}^{\ell} a_{\ell m} Z^m = 0 \quad (3.2)$$

estão fora do círculo unitário.

Logo todo processo não-determinístico, fracamente estacionário, pode ser aproximado, com a precisão desejada, por um processo auto-regressivo de ordem suficientemente alta. Esse fato explica a razão da grande aplicação do modelo auto-regressivo para estimação do espectro de potências em situações práticas.

Tendo em vista as observações anteriores, vamos supor que a série de tempo sob observação seja a realização de um processo auto-regressivo de ordem M , definido por:

$$X(t) = \sum_{m=1}^M a_m X(t-m) + \epsilon(t) \quad (3.3)$$

onde $\{\epsilon(t)\}$ é um ruído branco, e que as raízes da equação característica

$$A(Z) = 1 - \sum_{m=1}^M a_m Z^m = 0$$

estejam fora do círculo unitário.

Nesse capítulo vamos supor que por algum método foi determinado o valor conveniente da ordem M , do modelo auto-regressivo a ser ajustado à série, e nos preocuparemos com a esti-

mação dos parâmetros e com o cálculo da variabilidade assintótica desse tipo de estimativa do espectro de potências. No próximo capítulo discutiremos a questão da determinação de ordem M do processo a ser ajustado à série de tempo.

3.2 - ESTIMATIVAS DE MÁXIMA VEROSSIMILHANÇA DOS PARÂMETROS NO CASO DA DISTRIBUIÇÃO NORMAL -- ESTIMATIVAS DE YULE-WALKER

Nesta secção vamos supor o caso mais geral em que $E[X(t)] \neq 0$ e faremos a hipótese adicional que os $\epsilon(t)$'s são normalmente distribuídos com $E[\epsilon(t)] = 0$ e $\text{Var}[\epsilon(t)] = E[\epsilon^2(t)] = \sigma^2$. Nós trataremos então do problema da estimação dos coeficientes para o modelo

$$X(t) = \sum_{m=1}^M a_m X(t-m) + a_0 + \epsilon(t), \quad t=1, \dots, N. \quad (3.4)$$

Por conveniência, para obter as estimativas de máxima verossimilhança, nós suporemos que as observações de $X(t)$ começam para $t = -(M-1)$ e que $X[-(M-1)]$, $X[-(M-2)]$, ..., $X(0)$ são números conhecidos. Então a distribuição conjunta de $\epsilon(1), \dots, \epsilon(N)$ determina a distribuição conjunta de $X(1), \dots, X(N)$.

A equação de diferenças (3.4) pode ser escrita na forma matricial, do seguinte modo:

$$X(t) = \underline{A}' \tilde{X}_{t-1}^* + \epsilon(t), \quad t=1, \dots, N \quad (3.5)$$

onde

$$\underline{\tilde{A}} = \begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ \vdots \\ \vdots \\ a_M \end{bmatrix} = \begin{bmatrix} a_0 \\ \underline{a} \end{bmatrix}, \quad \tilde{X}_{t-1}^* = \begin{bmatrix} 1 \\ X(t-1) \\ X(t-2) \\ \vdots \\ \vdots \\ X(t-M) \end{bmatrix} \quad (3.6)$$

logo,

$$\varepsilon(t) = X(t) - \underline{\tilde{A}}' \tilde{X}_{t-1}^* = X(t) - \tilde{X}_{t-1}^{*'} \underline{\tilde{A}}, \quad t=1, \dots, N. \quad (3.7)$$

Como $\varepsilon(1), \dots, \varepsilon(t)$ são independentes de $X(-M+1), \dots, X(0)$, a densidade de probabilidade condicional de $\varepsilon(1), \dots, \varepsilon(N)$ dado $X(-M+1), \dots, X(0)$ é a mesma que a densidade não condicional e é dada por:

$$\frac{1}{(2\pi\sigma^2)^{\frac{1}{2}N}} \exp \left[-\frac{1}{2\sigma^2} \sum_{t=1}^N \varepsilon^2(t) \right]. \quad (3.8)$$

Quando $X(-M+1), \dots, X(0)$ são fixados (3.4) define uma transformação 1 a 1 de $X(1), \dots, X(N)$ em $\varepsilon(1), \dots, \varepsilon(N)$ e o Jacobiano dessa transformação é:

$$J_{(N \times N)} = \left| \frac{\partial \varepsilon(t)}{\partial X(t)} \right| = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 & \dots & 0 \\ -a_1 & 1 & 0 & \dots & 0 & \dots & 0 \\ -a_2 & -a_1 & 1 & \dots & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & & \vdots & & \vdots \\ \vdots & \vdots & \vdots & & \vdots & & \vdots \\ 0 & 0 & 0 & \dots & -a_M & -a_{M-1} & \dots & -a_2 & -a_1 & 1 \end{bmatrix} = 1$$

Logo, a densidade de probabilidade conjunta de $X(1), \dots, X(N)$ dado $X(-M+1), \dots, X(0)$ é

$$\frac{1}{(2\pi\sigma^2)^{\frac{1}{2}N}} \exp \left[-\frac{1}{2\sigma^2} \sum_{t=1}^N \left(X(t) - \tilde{A}' \tilde{X}_{t-1}^* \right)^2 \right]. \quad (3.9)$$

Para maximizar esta expressão em relação a $\tilde{A}$, e obter o estimador de máxima verossimilhança, nós precisamos minimizar a expressão

$$\sum_{t=1}^N [X(t) - \tilde{A}' \tilde{X}_{t-1}^*]^2 = \sum_{t=1}^N [X(t) - \tilde{X}_{t-1}^{*'} \tilde{A}]^2 \quad (3.10)$$

Este é da forma do problema de mínimos quadrados para um modelo de regressão linear com intercepto, do tipo

$$\tilde{Y} = X^* \tilde{A} + \varepsilon \quad (3.11)$$

onde

$$\tilde{Y} = \begin{bmatrix} X(1) \\ X(2) \\ \vdots \\ X(N) \end{bmatrix}, \quad X^* = \begin{bmatrix} 1 & X(0) & X(-1) & \dots & X(-M+1) \\ 1 & X(1) & X(0) & \dots & X(-M+2) \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & X(N-1) & X(N-2) & \dots & X(-M+N) \end{bmatrix}$$

$$\tilde{A} = \begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_M \end{bmatrix} \quad e \quad \varepsilon = \begin{bmatrix} \varepsilon(1) \\ \varepsilon(2) \\ \vdots \\ \varepsilon(N) \end{bmatrix}$$

cuja equações normais são

$$(X^{*'} X^*) \hat{\tilde{A}} = X^{*'} \tilde{Y} \quad (3.12)$$

Ocorre porém que, nesse caso, a matriz de planejamento X^* não é uma matriz de constantes, mas sim uma matriz constituída

da por variáveis aleatórias. No entanto, conforme foi mostrado por Durbin (1959), assintoticamente as propriedades das estimativas de mínimos quadrados são as mesmas que para o caso usual em que a matriz X^* é formada por constantes. Isto é, se a matriz $[E(X^* ' X^*)]^{-1} (X^* ' X^*)$ convergir, estocasticamente, para a matriz identidade I , teremos assintoticamente

$$E(\hat{A}) = \underline{A} + o(n^{-r})$$

e

$$\text{Var}(\hat{A}) = \sigma^2 [E(X^* ' X^*)]^{-1} [1 + o(n^{-s})], \quad (3.13)$$

onde r e s em geral são iguais a 1. Logo, as estimativas de mínimos quadrados serão consistentes.

Então, para N grande, fazendo

$$\underline{1}_N = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ \vdots \\ \vdots \\ 1 \end{bmatrix}, \quad \tilde{X}_t = \begin{bmatrix} X(t) \\ X(t-1) \\ \vdots \\ \vdots \\ X(t-M+1) \end{bmatrix},$$

$$X = \begin{bmatrix} X(0) & X(-1) & \dots & X(-M+1) \\ X(1) & X(0) & \dots & X(-M+2) \\ \vdots & \vdots & & \vdots \\ \vdots & \vdots & & \vdots \\ X(N-1) & X(N-2) & \dots & X(-M+N) \end{bmatrix} = \begin{bmatrix} X'_0 \\ X'_1 \\ \vdots \\ \vdots \\ X'_{N-1} \end{bmatrix} \quad (3.14)$$

e o vetor de médias

$$\bar{X}' = [\bar{X}_{(1)} \quad \bar{X}_{(2)} \quad \dots \quad \bar{X}_{(M)}]$$

onde

$$\bar{X}_{(\ell)} = \frac{1}{N} \sum_{t=1}^N X(t-\ell), \quad \ell=0,1,\dots,M \quad (3.15)$$

com $\bar{X}_{(0)} = \bar{X}$, temos

$$X^* = \begin{bmatrix} \mathbf{1}_N & X \end{bmatrix}$$

e portanto

$$(X^{*'} X^*) \hat{\tilde{A}} = X^{*'} \tilde{Y}$$

pode ser colocado na forma

$$\begin{bmatrix} \mathbf{1}_N' \\ X' \end{bmatrix} \begin{bmatrix} \mathbf{1}_N & X \end{bmatrix} \hat{\tilde{A}} = \begin{bmatrix} \mathbf{1}_N' \\ X' \end{bmatrix} \tilde{Y}$$

ou

$$\begin{bmatrix} N & N\bar{X}' \\ N\bar{X} & X'X \end{bmatrix} \begin{bmatrix} \hat{\tilde{a}}_0 \\ \hat{\tilde{a}} \end{bmatrix} = \begin{bmatrix} N\bar{X} \\ X'\tilde{Y} \end{bmatrix}, \quad (3.16)$$

logo

$$\begin{cases} N\hat{\tilde{a}}_0 + N\bar{X}'\hat{\tilde{a}} = N\bar{X} \end{cases} \quad (3.17)$$

$$\begin{cases} N\bar{X}\hat{\tilde{a}}_0 + X'X\hat{\tilde{a}} = X'\tilde{Y} \end{cases} \quad (3.18)$$

Mas,

$$X'X = [X_0 X_1 \dots X_{N-1}] \begin{bmatrix} X'_0 \\ X'_1 \\ \vdots \\ X'_{N-1} \end{bmatrix} = \sum_{t=1}^N X_{t-1} X'_{t-1}$$

a qual é uma matriz quadrada de ordem M, cujo elemento da linha ℓ e da coluna m é:

$$b_{\ell m} = \sum_{t=1}^N X(t-\ell)X(t-m), \quad \ell, m=1, 2, \dots, M \quad (3.19)$$

e

$$X'Y = [X_0 X_1 \dots X_{N-1}] \begin{bmatrix} X(1) \\ X(2) \\ \vdots \\ X(N) \end{bmatrix} = \sum_{t=1}^N X_{t-1} X(t)$$

o qual é um vetor $M \times 1$, cujo ℓ -ésimo elemento é:

$$b_{\ell 0} = \sum_{t=1}^N X(t-\ell)X(t), \quad \ell=1, 2, \dots, M. \quad (3.20)$$

Logo as equações (3.18) e (3.17) são dadas respectivamente por:

$$N\bar{X}_{(\ell)} \hat{a}_0 + \sum_{m=1}^M \sum_{t=1}^N X(t-\ell)X(t-m) \hat{a}_m = \sum_{t=1}^N X(t-\ell)X(t), \quad \ell=1, 2, \dots, M$$

$$N\hat{a}_0 + N \sum_{m=1}^M \bar{X}_{(m)} \hat{a}_m = N\bar{X}$$

ou ainda

$$\sum_{m=1}^M b_{\ell m} \hat{a}_m + N \bar{X}_{(\ell)} \hat{a}_0 = b_{\ell 0}, \quad \ell = 1, 2, \dots, M \quad (3.21)$$

$$N \sum_{m=1}^M \bar{X}_{(m)} \hat{a}_m + N \hat{a}_0 = N \bar{X} \quad (3.22)$$

onde

$$b_{\ell m} = \sum_{t=1}^N X(t-\ell) X(t-m), \quad \ell, m=1, \dots, M,$$

$$b_{\ell 0} = \sum_{t=1}^N X(t-\ell) X(t), \quad \ell = 1, \dots, M,$$

$$\bar{X}_{(\ell)} = \frac{1}{N} \sum_{t=1}^N X(t-\ell), \quad \ell = 0, 1, \dots, M$$

e

$$\bar{X}_0 = \bar{X}.$$

Se nós subtrairmos da relação (3.21) a relação (3.22) multiplicada por $\bar{X}_{(\ell)}$, obteremos

$$\sum_{m=1}^M [b_{\ell m} - N \bar{X}_{(\ell)} \bar{X}_{(m)}] \hat{a}_m = b_{\ell 0} - N \bar{X}_{(\ell)} \bar{X}, \quad \ell=1, \dots, M$$

ou

$$\sum_{m=1}^M \left[\sum_{t=1}^N X(t-\ell) X(t-m) - N \bar{X}_{(\ell)} \bar{X}_{(m)} \right] \hat{a}_m = \sum_{t=1}^N X(t-\ell) X(t) - N \bar{X}_{(\ell)} \bar{X}, \quad \ell=1, \dots, M \quad (3.23)$$

ou, de modo equivalente

$$\begin{aligned} \sum_{m=1}^M \sum_{t=1}^N [X(t-\ell) - \bar{X}_{(\ell)}] [X(t-m) - \bar{X}_{(m)}] \hat{a}_m &= \\ &= \sum_{t=1}^N [X(t-\ell) - \bar{X}_{(\ell)}] [X(t) - \bar{X}], \quad \ell=1, \dots, M \end{aligned} \quad (3.24)$$

Como, para N relativamente grande, temos $\bar{X}_{(\ell)} \approx \bar{X}$ para $\ell = 1, 2, \dots, M$, as equações (3.24) são aproximadamente iguais a

$$\sum_{m=1}^M C_{XX}(\ell-m) \hat{a}_m = C_{XX}(\ell), \quad \ell=1, \dots, M \quad (3.25)$$

ou

$$\sum_{m=1}^M r_{\ell-m} \hat{a}_m = r_{\ell}, \quad \ell=1, \dots, M \quad (3.26)$$

onde as auto-covariâncias estimadas $C_{XX}(k)$'s e as auto-correlações estimadas r_k 's são definidas, em termos das N observações $X(1), X(2), \dots, X(N)$, por

$$C_{XX}(k) = \begin{cases} \frac{1}{N} \sum_{t=1}^{N-|k|} [X(t+|k|) - \bar{X}] [X(t) - \bar{X}], & |k| \leq N-1 \\ 0 & , |k| > N-1 \end{cases} \quad (3.27)$$

e

$$r_k = \frac{C_{XX}(k)}{C_{XX}(0)}, \quad k=0, 1, \dots, N-1 \quad (3.28)$$

onde

$$\bar{X} = \frac{1}{N} \sum_{t=1}^N X(t).$$

Note que a expressão (3.26) é obtida da expressão (3.25) dividindo cada $C_{XX}(k)$ por $C_{XX}(0)$. As equações (3.26) são conhecidas como equações de Yule-Walker amostrais, as quais podem ser obtidas das equações (2.13) substituindo as auto-correlações teóricas ρ_k pelas auto-correlações estimadas r_k .

Se nós escrevermos

$$R = \begin{bmatrix} 1 & r_1 & \cdots & r_{M-1} \\ r_1 & 1 & \cdots & r_{M-2} \\ \vdots & \vdots & & \vdots \\ r_{M-1} & r_{M-2} & \cdots & 1 \end{bmatrix}, \quad r = \begin{bmatrix} r_1 \\ r_2 \\ \vdots \\ r_M \end{bmatrix}, \quad (3.29)$$

as equações de Yule-Walker amostrais (3.26) podem ser escritas na forma matricial

$$R\hat{a} = r \quad (3.30)$$

e a solução de (3.30) é dada por

$$\hat{a} = R^{-1}r \quad (3.31)$$

As equações (3.30) têm a vantagem que são necessárias apenas M estatísticas e R é positiva definida. (A matriz R pode ser considerada como proporcional à matriz cujo elemento da ℓ -ésima linha e m -ésima coluna é

$$\sum_{t=-M}^M [X(t+\ell) - \bar{X}][X(t+m) - \bar{X}],$$

onde $X(t+l) - \bar{X}$ é igual a zero se $t+l < 0$ ou $t+l > N$. R é simétrica não apenas com relação à diagonal principal, mas também com relação à diagonal secundária.

As formas especiais de R e $\underline{r}$ conduzem a um método recursivo simples de calcular os coeficientes para equação de ordem $(M+1)$, a partir da solução para uma equação de ordem M . As equações para o vetor de coeficientes de uma equação de ordem $(M+1)$, digamos $\hat{a}(M+1)$, são:

$$R_{M+1} \hat{a}(M+1) = \underline{r}_{M+1} \quad (3.32)$$

onde

$$R_{M+1} = \begin{bmatrix} R_M & \tilde{r}_M \\ \tilde{r}_M' & 1 \end{bmatrix}, \quad \hat{a}_{M+1} = \begin{bmatrix} \hat{a}^{(1)}(M+1) \\ \hat{a}_{M+1}(M+1) \end{bmatrix}, \quad \underline{r}_{M+1} = \begin{bmatrix} \underline{r}_M \\ r_{M+1} \end{bmatrix} \quad (3.33)$$

e R_M e $\underline{r}_M$ são dadas por (3.29), e $\tilde{r}_M = (r_M, r_{M-1}, \dots, r_1)'$ o qual é $\underline{r}_M$ com suas coordenadas na ordem inversa.

As equações particionadas de (3.32) são:

$$R_M \hat{a}^{(1)}(M+1) + \tilde{r}_M \hat{a}_{M+1}(M+1) = \underline{r}_M \quad (3.34)$$

$$\tilde{r}_M' \hat{a}^{(1)}(M+1) + \hat{a}_{M+1}(M+1) = r_{M+1} \quad (3.35)$$

A eliminação de $\hat{a}^{(1)}(M+1)$ de (3.34) e (3.35) produz:

$$\hat{a}_{M+1}(M+1) = \frac{r_{M+1} - \tilde{r}_M' R_M^{-1} \underline{r}_M}{1 - \tilde{r}_M' R_M^{-1} \tilde{r}_M} = \frac{r_{M+1} - \tilde{r}_M' \hat{a}^{(M)}}{1 - \tilde{r}_M' \hat{a}^{(M)}}, \quad (3.36)$$

onde $\hat{\underline{a}}(M) = (\hat{a}_M(M), \hat{a}_{M-1}(M), \dots, \hat{a}_1(M))'$, o qual é $\hat{\underline{a}}(M)$ com suas coordenadas na ordem inversa.

De fato, pré-multiplicando a equação (3.34) por $(-\tilde{r}_M' R_M^{-1})$ e adicionando à equação (3.35), temos,

$$[1 - \tilde{r}_M' R_M^{-1} \tilde{r}_M] \hat{a}_{M+1}(M+1) = r_{M+1} - \tilde{r}_M' R_M^{-1} r_M$$

ou

$$\hat{a}_{M+1}(M+1) = \frac{r_{M+1} - \tilde{r}_M' R_M^{-1} r_M}{1 - \tilde{r}_M' R_M^{-1} \tilde{r}_M}$$

Da relação (3.34), temos

$$\begin{aligned} \hat{\underline{a}}^{(1)}(M+1) &= R_M^{-1} r_M - R_M^{-1} \tilde{r}_M \hat{a}_{M+1}(M+1) = \\ &= \hat{\underline{a}}(M) - \hat{\underline{a}}(M) \hat{a}_{M+1}(M+1), \end{aligned} \quad (3.37)$$

onde $\hat{a}_{M+1}(M+1)$ é dado pela relação (3.36).

Em termos dos componentes (3.36) e (3.37) são respectivamente:

$$\begin{aligned} \hat{a}_{M+1}(M+1) &= \frac{r_{M+1} - \sum_{m=1}^M \hat{a}_m(M) r_{M+1-m}}{1 - \sum_{m=1}^M \hat{a}_m(M) r_m} = \\ &= \frac{C(M+1) - \sum_{m=1}^M \hat{a}_m(M) C(M+1-m)}{C(0) - \sum_{m=1}^M \hat{a}_m(M) C(m)} \end{aligned} \quad (3.38)$$

e

$$\hat{a}_m(M+1) = \hat{a}_m(M) - \hat{a}_{M+1-m}(M) \cdot \hat{a}_{M+1}(M+1), \quad m=1, \dots, M, \quad (3.39)$$

Essas equações fornecem um método recursivo simples de calcular as estimativas dos coeficientes para um processo de dada ordem. Os passos intermediários de cálculo fornecem os coeficientes para todos os processos de ordem inferior. Tais estimativas são conhecidas como estimativas de Yule-Walker dos coeficientes.

A estimativa de máxima verossimilhança de σ^2 é dada por:

$$\hat{\sigma}^2 = \frac{1}{N} \sum_{t=1}^N [X(t) - \underline{A}' \tilde{X}_{t-1}^*]^2. \quad (3.40)$$

De fato, a função de verossimilhança é

$$\begin{aligned} L(\sigma^2) &= \frac{1}{(2\pi\sigma^2)^{N/2}} \exp \left[-\frac{1}{2\sigma^2} \sum_{t=1}^N (X(t) - \underline{A}' \tilde{X}_{t-1}^*)^2 \right] = \\ &= \frac{1}{(2\pi\sigma^2)^{N/2}} \exp \left[-\frac{1}{2\sigma^2} \sum_{t=1}^N \epsilon^2(t) \right], \end{aligned}$$

logo,

$$\begin{aligned} \frac{\partial L(\sigma^2)}{\partial \sigma^2} &= -\frac{1}{(2\pi)^{N/2}} \cdot \frac{N}{2} (\sigma^2)^{-\frac{N}{2}-1} \exp \left[-\frac{1}{2\sigma^2} \sum \epsilon^2(t) \right] + \\ &+ \frac{1}{(2\pi\sigma^2)^{N/2}} \exp \left[-\frac{1}{2\sigma^2} \sum \epsilon^2(t) \right] \frac{1}{2(\sigma^2)^2} \sum \epsilon^2(t) = \\ &= -\frac{1}{(2\pi)^{N/2}} \exp \left[-\frac{1}{2\sigma^2} \sum \epsilon^2(t) \right] \left\{ \frac{N}{2} (\sigma^2)^{-\frac{N}{2}-1} - \right. \\ &\left. - \frac{1}{(\sigma^2)^{N/2}} \frac{1}{2(\sigma^2)^2} \sum \epsilon^2(t) \right\} \end{aligned}$$

$$\begin{aligned}\frac{\partial L(\sigma^2)}{\partial \sigma^2} = 0 &\implies \frac{N}{2}(\hat{\sigma}^2)^{-\frac{N}{2}-1} - \frac{1}{(\hat{\sigma}^2)^{N/2}} \frac{1}{2(\hat{\sigma}^2)^2} \sum \epsilon^2(t) = 0 \implies \\ &(\hat{\sigma}^2)^{-\frac{N}{2}-1} (\hat{\sigma}^2)^{\frac{N}{2}} (\hat{\sigma}^2)^2 = \frac{1}{N} \sum \epsilon^2(t) \implies \\ &(\hat{\sigma}^2)^{-\frac{N}{2}-1 + \frac{N}{2} + 2} = \frac{1}{N} \sum \epsilon^2(t) \implies \hat{\sigma}^2 = \frac{1}{N} \sum \epsilon^2(t).\end{aligned}$$

Então a estimativa de máxima verossimilhança de σ^2 é dada por:

$$\hat{\sigma}^2 = \frac{1}{N} \sum_{t=1}^N \epsilon^2(t)$$

ou

$$\hat{\sigma}^2 = \frac{1}{N} \sum_{t=1}^N [X(t) - \hat{A}' \tilde{X}_{t-1}^*]^2.$$

3.3 - PROPRIEDADES DOS ESTIMADORES DOS PARÂMETROS DE UM PROCESSO AR(M)

Nas condições descritas na introdução desse capítulo, isto é, supondo $E[X(t)] = 0$ e sendo os ruídos $\epsilon(t)$'s variáveis aleatórias não correlacionadas com média zero e variância σ^2 , a função densidade espectral $f_{XX}(\lambda)$ do processo $\{X(t)\}$ é dada por:

$$f_{XX}(\lambda) = \frac{1}{2\pi} \frac{\sigma^2}{\left| 1 - \sum_{m=1}^M a_m \exp(-i\lambda m) \right|^2}, \quad -\pi \leq \lambda \leq \pi \quad (3.41)$$

onde $\gamma_{XX}(\ell) = E[X(t+\ell)X(t)]$ é dada por:

$$\begin{aligned}
 \gamma_{XX}(\ell) &= E \left(\int_{-\pi}^{\pi} e^{i\lambda(t+\ell)} Z_X(d\lambda) \right) \overline{ \left(\int_{-\pi}^{\pi} e^{i\mu t} Z_X(d\mu) \right) } = \\
 &= E \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} e^{i\lambda(t+\ell)} e^{-i\mu t} Z_X(d\lambda) \overline{Z_X(d\mu)} = \\
 &= \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} e^{i(\lambda-\mu)t} e^{i\lambda\ell} E[Z_X(d\lambda) \overline{Z_X(d\mu)}] = \\
 &= \int_{-\pi}^{\pi} e^{i\lambda\ell} F_{XX}(d\lambda) = \int_{-\pi}^{\pi} e^{i\lambda\ell} f_{XX}(\lambda) d\lambda = \\
 &= \left(\frac{\sigma^2}{2\pi} \right) \int_{-\pi}^{\pi} e^{i\lambda\ell} \frac{1}{|A(\lambda)|^2} d\lambda \quad (3.42)
 \end{aligned}$$

e

$$A(\lambda) = 1 - \sum_{m=1}^M a_m \exp(-i\lambda m).$$

Consideremos, então a situação em que possuímos um conjunto de dados $\{X(t); t=1, 2, \dots, N\}$, correspondentes à série de tempo amostrada. Nessas condições, as estimativas $\{\hat{a}_m; m=1, 2, \dots, M\}$ dos coeficientes auto-regressivos podem ser obtidas resolvendo-se as equações de Yule-Walker amostrais dadas pela expressão (3.26), onde

$$r_\ell = \frac{C_{XX}(\ell)}{C_{XX}(0)} \quad (3.43)$$

e

$$C_{XX}(\ell) = \frac{1}{N} \sum_{t=1}^{N-|\ell|} X(t+|\ell|)X(t) \quad (3.44)$$

ou de modo equivalente resolvendo-se as equações (3.25), com $C_{XX}(\ell)$ definido pela expressão (3.44).

Observemos que nós estamos supondo $E[X(t)] = 0$. No entanto, em situações práticas, $C_{XX}(\ell)$ seria definido pela expressão (3.27).

Tendo em vista a relação (2.16), definimos uma estimativa $S^2(M)$ de σ^2 , obtida pelo método dos momentos, por:

$$S^2(M) = C_{XX}(0) - \sum_{m=1}^M \hat{a}_m C_{XX}(m). \quad (3.45)$$

Dessa forma estamos ajustando um modelo com as primeiras $M+1$ covariâncias iguais a $\{C_{XX}(\ell), \ell=0,1,2,\dots,M\}$. Se o processo $X(t)$ for estritamente estacionário e ergódico, $\hat{a}_m$ e $S^2(M)$ convergem, respectivamente, para a_m e σ^2 com probabilidade 1, quando N tende ao infinito. Ver Grenander, Roseblatt (1957) ou Doob (1953).

Usando o resultado de Anderson & Walker (1964) sobre a distribuição limite dos coeficientes das auto-correlações amostrais de um processo linear nós podemos obter,

TEOREMA 1 - Sob a suposição de estacionaridade estrita e mútua independência dos $\{\epsilon(t)\}$, a distribuição de

$$\sqrt{N}\Delta a_m = \sqrt{N}(\hat{a}_m - a_m), \quad (m=1,2,\dots,M)$$

converge, quando N tende ao infinito, para uma distribuição Gaussiana M -dimensional (com vetor de média zero e matriz de variância $\sigma^2 \Gamma_{MM}^{-1}$, onde Γ_{MM} é uma matriz de ordem $M \times M$ com o ele-

mento da linha ℓ e coluna m igual a

$$\gamma_{XX}(m-\ell) = E[X(t-\ell)X(t-m)].$$

O teorema pode ser provado usando o fato que a distribuição limite de $\{\sqrt{N}\Delta a_m; m=1,2,\dots,M\}$ é idêntica à distribuição limite de $\{\eta_m; m=1,2,\dots,M\}$, onde

$$\begin{bmatrix} \eta_1 \\ \eta_2 \\ \vdots \\ \eta_M \end{bmatrix} = \sqrt{N}\Gamma_{MM}^{-1} \begin{bmatrix} C_{X\epsilon}(1) \\ C_{X\epsilon}(2) \\ \vdots \\ C_{X\epsilon}(M) \end{bmatrix} \quad (3.46)$$

e

$$C_{X\epsilon}(\ell) = \frac{1}{N} \sum_{t=1}^N X(t-\ell)\epsilon(t), \quad \ell=1,2,\dots,M$$

Do resultado de Anderson & Walker (1964) pode ser mostrado que a distribuição de $\{\sqrt{N} C_{X\epsilon}(\ell); \ell=1,2,\dots,M\}$ tende assintoticamente a ser Gaussiana com vetor de média zero e matriz de variância $\sigma^2\Gamma_{MM}$. A afirmação do teorema é consequência direta dessas observações.

Como a matriz Γ_{MM}^{-1} é real e simétrica ela pode ser escrita como

$$\Gamma_{MM}^{-1} = B_M' D_M D_M B_M \quad (3.47)$$

onde

$$B_M = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ -a_{11} & 1 & 0 & \dots & 0 \\ -a_{22} & -a_{21} & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -a_{M-1,M-1} & -a_{M-1,M-2} & \dots & \dots & 1 \end{bmatrix}, \quad (3.48)$$

$$D_M = \begin{bmatrix} 1/\sigma(0) & 0 & 0 & \dots & 0 \\ 0 & 1/\sigma(1) & 0 & \dots & 0 \\ 0 & 0 & 1/\sigma(2) & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 1/\sigma(M-1) \end{bmatrix}, \quad (3.49)$$

$\sigma(\ell)$ e $\{a_{\ell,m}; m=1,2,\dots,\ell\}$ sendo dados pela relação

$$\begin{aligned} \sigma^2(\ell) &= E \left| X(t-\ell) - \sum_{m=1}^{\ell} a_{\ell,m} X(t-\ell+m) \right|^2 = \\ &= \min_{\{\beta_m\}} E \left| X(t-\ell) - \sum_{m=1}^{\ell} \beta_m X(t-\ell+m) \right|^2, \end{aligned} \quad (3.50)$$

e B'_M indica a matriz transposta de B_M .

Dessa relação temos que as variáveis aleatórias

$$Y(t-\ell) = X(t-\ell) - \sum_{m=1}^{\ell} a_{\ell,m} X(t-\ell+m), \quad \ell=0,1,\dots,M$$

são mutuamente não correlacionadas e com variâncias iguais a $\sigma^2(\ell)$, respectivamente.

Nós estamos assumindo:

$$a_{M,k} = a_k, \quad k = 1, 2, \dots, M$$

$$\sigma^2(M) = \sigma^2 = E[\varepsilon^2(t)]$$

$$Y(t) = X(t)$$

$$\sigma^2(0) = E[X^2(t)]$$

Provemos que as variáveis aleatórias $Y(t-\ell), \ell=0,1,\dots, M$ são não correlacionadas.

De fato, temos:

$$\begin{aligned} E[Y(t-\ell)Y(t-s)] &= E\left\{\left[X(t-\ell) - \sum_{m=1}^{\ell} a_{\ell,m}X(t-\ell+m)\right]\left[X(t-s) - \right. \right. \\ &\quad \left. \left. - \sum_{j=1}^s a_{s,j}X(t-s+j)\right]\right\} = E[X(t-\ell)X(t-s)] - \\ &\quad - \sum_{m=1}^{\ell} a_{\ell,m}E[X(t-\ell+m)X(t-s)] - \sum_{j=1}^s a_{s,j}E[X(t-\ell)X(t-s+j)] + \\ &\quad + \sum_{m=1}^{\ell} \sum_{j=1}^s a_{\ell,m}a_{s,j}E[X(t-\ell+m)X(t-s+j)] = \gamma_{XX}(\ell-s) - \\ &\quad - \sum_{m=1}^{\ell} a_{\ell,m}\gamma_{XX}(\ell-s-m) - \sum_{j=1}^s a_{s,j}\gamma_{XX}(\ell-s+j) + \\ &\quad + \sum_{j=1}^s \sum_{m=1}^{\ell} a_{s,j}a_{\ell,m}\gamma_{XX}(\ell-s+j-m) = \gamma_{XX}(\ell-s) - \\ &\quad - \sum_{m=1}^{\ell} a_m\gamma_{XX}(\ell-s-m) - \sum_{j=1}^s a_j\left[\gamma_{XX}(\ell-s+j) - \sum_{m=1}^{\ell} a_m\gamma_{XX}(\ell-s+j-m)\right] \end{aligned}$$

fazendo $\ell - s = k$, temos, pela relação (2.9) do capítulo anterior, para todo $\ell > s$, $(\ell, s = 0, 1, 2, \dots, M)$,

$$\gamma_{XX}(\ell-s) - \sum_{m=1}^{\ell} a_m \gamma_{XX}(\ell-s-m) = 0$$

e

$$\gamma_{XX}(\ell-s+j) - \sum_{m=1}^{\ell} a_m \gamma_{XX}(\ell-s+j-m) = 0,$$

para todo $j = 1, 2, \dots, s$, logo $E[Y(t-\ell)Y(t-s)] = 0$, e portanto, $Y(t-\ell)$, $\ell = 0, 1, 2, \dots, M$ são mutuamente não correlacionadas.

Temos ainda, $\sigma^2(\ell) = E[Y^2(t-\ell)]$, $\ell = 0, 1, 2, \dots, M$, e como $X(t)$ tem média zero, os $Y(t-\ell)$, ($\ell = 0, 1, 2, \dots, M$) também tem média zero e portanto $\sigma^2(\ell)$ são as variâncias de $Y(t-\ell)$, $\ell = 0, 1, 2, \dots, M$.

Aplicando a transformação $D_M B_M$ ao vetor

$$X_M = (X(t), X(t-1), \dots, X(t-M+1))',$$

nós obtemos

$$D_M B_M X_M = (Z_0(t), Z_1(t), \dots, Z_{M-1}(t))'$$

com

$$E[Z_{\ell}(t)Z_m(t)] = \begin{cases} 0, & \ell \neq m \\ 1, & \ell = m. \end{cases}$$

De fato, temos:

$$D_{MM}^B X_M = \begin{bmatrix} \frac{1}{\sigma(0)} & 0 & 0 & \dots & 0 \\ -\frac{a_{11}}{\sigma(1)} & \frac{1}{\sigma(1)} & 0 & \dots & 0 \\ -\frac{a_{22}}{\sigma(2)} & -\frac{a_{21}}{\sigma(2)} & \frac{1}{\sigma(2)} & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -\frac{a_{M-1,M-1}}{\sigma(M-1)} & -\frac{a_{M-1,M-2}}{\sigma(M-1)} & \dots & \dots & \frac{1}{\sigma(M-1)} \end{bmatrix} \begin{bmatrix} X(t) \\ X(t-1) \\ X(t-2) \\ \vdots \\ X(t-M+1) \end{bmatrix} =$$

$$\begin{bmatrix} \frac{X(t)}{\sigma(0)} \\ \frac{X(t-1)}{\sigma(1)} - \frac{a_{11}X(t)}{\sigma(1)} \\ \frac{X(t-2)}{\sigma(2)} - \frac{a_{21}X(t-1)}{\sigma(2)} - \frac{a_{22}X(t)}{\sigma(2)} \\ \vdots \\ \frac{X(t-(M-1))}{\sigma(M-1)} - \sum_{j=1}^{M-1} \frac{a_{M-1,j}X(t-(M-1)+j)}{\sigma(M-1)} \end{bmatrix} = \begin{bmatrix} Z_0(t) \\ Z_1(t) \\ Z_2(t) \\ \vdots \\ Z_{M-1}(t) \end{bmatrix}$$

logo, para $\ell \neq m$

$$\begin{aligned} E[Z_\ell(t)Z_m(t)] &= E\left\{\left[\frac{X(t-\ell)}{\sigma(\ell)} - \sum_{j=1}^{\ell} \frac{a_{\ell j}X(t-\ell+j)}{\sigma(\ell)}\right]\left[\frac{X(t-m)}{\sigma(m)} - \sum_{k=1}^m \frac{a_{mk}X(t-m+k)}{\sigma(m)}\right]\right\} = \frac{1}{\sigma(\ell)\sigma(m)}E\left\{\left[X(t-\ell) - \sum_{j=1}^{\ell} a_{\ell j}X(t-\ell+j)\right]\left[X(t-m) - \sum_{k=1}^m a_{mk}X(t-m+k)\right]\right\} = \\ &= \frac{1}{\sigma(\ell)\sigma(m)}E[Y(t-\ell)Y(t-m)] = 0, \end{aligned}$$

pois as variáveis aleatórias $Y(t-\ell)$, $\ell=0,1,\dots,M$ são não correlacionadas; para $\ell=m$, temos

$$\begin{aligned} E[Z_\ell(t)Z_m(t)] &= E[Z_\ell^2(t)] = E\left\{\left[\frac{X(t-\ell)}{\sigma(\ell)} - \sum_{j=1}^{\ell} \frac{a_{\ell j} X(t-\ell+j)}{\sigma(\ell)}\right]^2\right\} = \\ &= \frac{1}{\sigma^2(\ell)} E\left\{\left[X(t-\ell) - \sum_{j=1}^{\ell} a_{\ell j} X(t-\ell+j)\right]^2\right\} = \\ &= \frac{1}{\sigma^2(\ell)} E\{Y^2(t-\ell)\} = \frac{1}{\sigma^2(\ell)} \sigma^2(\ell) = 1. \end{aligned}$$

Como a matriz de covariâncias da distribuição limite

$$\{\sqrt{N}C_{x\epsilon}(\ell); \ell=1,2,\dots,M\}$$

é igual a $\sigma^2\Gamma_{MM}$, onde $\Gamma_{MM} = E[X_M X_M']$, fazendo

$$C_{x\epsilon} = [C_{x\epsilon}^{(1)}, C_{x\epsilon}^{(2)}, \dots, C_{x\epsilon}^{(M)}]',$$

temos que a matriz de covariâncias da distribuição limite de $U_M = \sigma^{-1}\sqrt{N}D_M B_M C_{x\epsilon}$ é uma matriz I_M (identidade de ordem M).

De fato, temos:

$$\begin{aligned} \text{Var}(U_M) &= \text{Var}(\sigma^{-1}D_M B_M \sqrt{N}C_{x\epsilon}) = \\ &= \sigma^{-2}D_M B_M \text{Var}(\sqrt{N}C_{x\epsilon})B_M' D_M = \\ &= \sigma^{-2}D_M B_M \sigma^2 \Gamma_{MM} B_M' D_M = D_M B_M \Gamma_{MM} B_M' D_M \end{aligned}$$

e como $\Gamma_{MM}^{-1} = B_M' D_M^{-1} D_M^{-1} B_M$ e portanto $\Gamma_{MM} = B_M^{-1} D_M^{-1} D_M^{-1} (B_M')^{-1}$, temos:

$$\begin{aligned} \text{Var}(U_M) &= D_M B_M B_M^{-1} D_M^{-1} D_M^{-1} (B_M')^{-1} B_M' D_M = \\ &= (D_M B_M) (D_M B_M)^{-1} (B_M' D_M)^{-1} (B_M' D_M) = I_M \end{aligned}$$

Então, como a distribuição limite de $\{\sqrt{N}\Delta a_m; m=1, 2, \dots, M\}'$ é idêntica à de $\sqrt{N}\Gamma_{MM}^{-1}C_{x\varepsilon}$ e como

$$\sqrt{N}\Gamma_{MM}^{-1}C_{x\varepsilon} = \sqrt{N}B_M'D_M D_M B_M C_{x\varepsilon} = \sigma B_M'D_M U_M,$$

nós obtivemos uma representação assintótica de

$$\sqrt{N}\Delta a = \{\sqrt{N}\Delta a_m; m=1, \dots, M\}'$$

por variáveis que em sua distribuição limite não mutuamente ortogonais e normalizadas, isto é,

$$\sqrt{N}\Delta a \sim \sigma B_M'D_M U_M, \quad (3.51)$$

onde $\sim$ significa que ambos os membros têm a mesma distribuição limite. Usando esta relação nós podemos facilmente obter a distribuição limite de qualquer transformação linear de $\sqrt{N}\Delta a$.

Supondo, além da estrita estacionaridade e independência dos $\{\varepsilon(t)\}$, que $\{\varepsilon(t)\}$ possui momento de 4ª ordem finito, discutiremos a variabilidade amostral da estimativa do espectro. Nessas condições e sendo

$$C_{\varepsilon\varepsilon}(\ell) = \frac{1}{N} \sum_{t=1}^N \varepsilon(t+\ell)\varepsilon(t),$$

temos:

TEOREMA 2 - Se $\{\varepsilon(t)\}$ é estritamente estacionário e $\varepsilon(t)$'s são mutuamente independentes com momento de 4ª ordem $m_4 = E[\varepsilon^4(t)]$, a distribuição simultânea de

$$\sqrt{N} \left(\frac{C_{\varepsilon\varepsilon}(0) - \sigma^2}{\sqrt{m_4 - \sigma^2}} \right) \text{ e } \sqrt{N} \frac{C_{\varepsilon\varepsilon}(\ell)}{\sigma^2}, \quad \ell=1, 2, \dots, L,$$

(L; qualquer inteiro positivo) converge para uma distribuição Gaussiana (L+1)-dimensional com vetor de média zero e matriz de variância I_{L+1} , onde I_{L+1} é uma matriz identidade (L+1)×(L+1).

A prova desse teorema é obtida usando o Teorema do Limite Central para com processo estacionário m-dependente devido a Diananda (1953).

3.4 - ESTIMAÇÃO DA FUNÇÃO DENSIDADE ESPECTRAL E PROPRIEDADES

ASSINTÓTICAS DO ESTIMADOR ESPECTRAL AUTO-REGRESSIVO

O estimador espectral auto-regressivo proposto por Parzen (1969) é dada por

$$\hat{f}_M(\lambda) = \frac{1}{2\pi} \frac{S^2(M)}{\left| 1 - \sum_{m=1}^M \hat{a}_m \exp(-i\lambda m) \right|^2}, \quad -\pi \leq \lambda \leq \pi \quad (3.52)$$

onde $\{\hat{a}_m\}$ e $S^2(M)$ são dados pelas relações (3.25) e (3.45).

Com suposições sobre a regularidade do processo estacionário, e com suposições sobre a razão relativa assintótica de M e N, Berk (1974) mostrou que os estimadores espectrais auto-regressivos são consistentes e, com suposições adicionais, mostrou também que esses estimadores são assintoticamente normais, não correlacionados para diferentes frequências. A distribuição limite também foi investigada por Kromer (1969).

Os resultados demonstrados por Berk (1974) são enunciados nos teoremas seguintes:

TEOREMA 3 - Seja $X(t) = \epsilon(t) + b_1\epsilon(t-1) + b_2\epsilon(t-2) + \dots$ onde $\{\epsilon(t)\}$ é uma seqüência de variáveis aleatórias independentes identicamente distribuídas com média 0 e variância σ^2 . Suponha que

$$B(Z) = 1 + b_1Z + b_2Z^2 + \dots = \frac{1}{1 - a_1Z - a_2Z^2 - \dots} = \frac{1}{A(Z)}$$

seja não nula para $|Z| < 1$. Suponha

- i) $A(e^{i\lambda})$ é não nula, $-\pi < \lambda \leq \pi$.
- ii) $E[\epsilon(t)]^4 < \infty$, $\forall t$.
- iii) A escolha de M em termos de N é tal que $\frac{M^3}{N} \rightarrow 0$.
- iv) A escolha de M em termos de N é tal que

$$M^{\frac{1}{2}}(|a_{M+1}| + |a_{M+2}| + \dots) \rightarrow 0.$$

Então, quando N tende ao infinito, $\hat{f}_M(\lambda)$ converge para $f(\lambda)$ em probabilidade.

TEOREMA 4 - Suponha que $\{X(t)\}$ é um processo linear,

$$X(t) = \epsilon(t) + b_1\epsilon(t-1) + b_2\epsilon(t-2) + \dots,$$

onde $\{\epsilon(t)\}$ é uma seqüência de variáveis aleatórias independentes identicamente distribuídas, $E[\epsilon(t)]$, tal que

$$B(Z) = 1 + b_1Z + b_2Z^2 + \dots$$

é não nula dentro do círculo unitário e

- i) a densidade espectral $f(\lambda) = \frac{\sigma^2}{2\pi} |B(e^{i\lambda})|^2$ é positiva.

- ii) $E[\epsilon(t)]^4 < \infty$, $\forall t$.
- iii) M é escolhido de modo que $M \rightarrow \infty$ e $\frac{M^3}{N} \rightarrow 0$, $N \rightarrow \infty$.
- iv) M é escolhido de modo que

$$N^{\frac{1}{2}}(|a_{M+1}| + |a_{M+2}| + \dots) \rightarrow 0,$$

$N \rightarrow \infty$, onde

$$A(Z) = 1 - a_1 Z - a_2 Z^2 - \dots = \frac{1}{B(Z)}.$$

A estimativa de densidade espectral é

$$\hat{f}_M(\lambda) = \frac{1}{2\pi} \frac{S^2(M)}{\left| 1 - \sum_{m=1}^M \hat{a}_m \exp(-i\lambda m) \right|^2} = \frac{1}{2\pi} \frac{S^2(M)}{[\hat{C}_M(\lambda)^2 + \hat{S}_M(\lambda)^2]}$$

onde

$$\hat{C}_M(\lambda) = 1 - \sum_{m=1}^M \hat{a}_m \cos m\lambda$$

$$\hat{S}_M(\lambda) = \sum_{m=1}^M \hat{a}_m \sin m\lambda.$$

Então,

$$\left(\frac{N}{M}\right)^{\frac{1}{2}} [\hat{f}_M(0) - f(0)], \left(\frac{N}{M}\right)^{\frac{1}{2}} [\hat{f}_M(\lambda_1) - f(\lambda_1)], \dots$$

$$\left(\frac{N}{M}\right)^{\frac{1}{2}} [\hat{f}_M(\lambda_p) - f(\lambda_p)], \left(\frac{N}{M}\right)^{\frac{1}{2}} [\hat{f}_M(\pi) - f(\pi)], \quad 0 < \lambda_1 < \dots < \lambda_p < \pi,$$

são variáveis aleatórias assintoticamente independentes e normais, com médias 0 e variâncias

$$4f^2(0), 2f^2(\lambda_1), \dots, 2f^2(\lambda_p), 4f^2(\pi),$$

respectivamente.

Então, assumindo que a ordem M é assintoticamente suficiente para eliminar o vício, o Teorema 3 mostra que a auto-regressão produz um estimador consistente de densidade espectral de $\{X(t)\}$. Além disso, assumindo que M tende ao infinito de modo que o vício de se usar uma auto-regressão finita se anule com rapidez suficiente, o Teorema 4 mostra que as estimativas espectrais auto-regressivas são assintoticamente normais, não correlacionadas para diferentes frequências fixadas, e a variância assintótica é a mesma que a das estimativas espectrais baseadas no periodograma truncado.

Sob a suposição que o processo observado $\{X(t)\}$ é normal com média zero e tem uma densidade espectral estritamente positiva, limitada e "suficientemente suave", Kromer (1969) mostrou que os estimadores espectrais auto-regressivos são consistentes e assintoticamente normais quando primeiro $N \rightarrow \infty$ e então $M \rightarrow \infty$. Além disso, forneceu uma medida da razão de decréscimo do vício.

Apresentamos em seguida esses resultados.

Seja:

γ = função de covariância de $\{X(t)\}$.

f = função densidade espectral de $\{X(t)\}$.

$\hat{f}_M(\lambda)$ = estimador espectral dado pela relação (3.52).

Suponha que f é estritamente positiva e limitada. Então existe uma sequência $\{b_j\}_{j=1}^{\infty}$ de números reais e um número real

positivo σ tal que

$$\sum_{j=1}^{\infty} b_j^2 < \infty$$

e

$$\frac{\sigma^2}{2\pi} \left| 1 + \sum_{j=1}^{\infty} b_j \exp(-i\omega j) \right|^2 = f(\omega),$$

em quase toda parte, $\omega \in [-\pi, \pi]$.

Sem suposições adicionais, para cada $M \geq 1$, as soluções σ_M^2 e $\{a_m\}_{m=1}^M$ das equações

$$\sum_{m=1}^M a_m \gamma(\ell-m) = \gamma(\ell), \quad \ell=1, 2, \dots, M$$

$$\sigma_M^2 = \gamma(0) - \sum_{m=1}^M a_m \gamma(m)$$

são tais que

$$f_M(\lambda) = \frac{1}{2\pi} \frac{\sigma_M^2}{\left| 1 - \sum_{m=1}^M a_m \exp(-i\lambda m) \right|^2}$$

é uma função contínua em $[-\pi, \pi]$.

Se, além disso, assumirmos que $\{X(t)\}$ é normal e que pelo menos uma das condições seguintes é verdadeira

$$\sum_{j=1}^{\infty} |b_j| < \infty$$

$$\sum_{j=1}^{\infty} j b_j^2 < \infty,$$

então, para cada $M \geq 1$,

$$\lim_{N \rightarrow \infty} P \left[\sqrt{N} \left(\frac{\hat{f}_M(\lambda) - f_M(\lambda)}{\sqrt{\Sigma_M(\lambda)}} \right) \leq x \right] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{1}{2}t^2} dt, \quad -\pi \leq \lambda \leq \pi, \quad (3.53)$$

onde a convergência é uniforme em $x \in (-\infty, \infty)$,

$$\Sigma_M(\lambda) = 2\pi \int_{-\pi}^{\pi} \tilde{\Sigma}_M(\lambda, \omega) f^2(\omega) d\omega,$$

e $\tilde{\Sigma}_M(\lambda, \omega)$ é dado em Kromer (1969), pela equação (1.12), na página 15.

O limite iterado

$$\lim_{M \rightarrow \infty} \lim_{N \rightarrow \infty} a_{NM} = a$$

significa

$$\lim_{M \rightarrow \infty} (\lim_{N \rightarrow \infty} |a_{NM} - a|) = 0$$

isto é, para cada $\epsilon > 0$ existem inteiros $M_{\epsilon}^*, N_{1\epsilon}^*, N_{2\epsilon}^*, \dots$ tais que se o par (N, M) satisfaz $M \geq M_{\epsilon}^*, N \geq N_{M_{\epsilon}^*}$, então

$$|a_{NM} - a| < \epsilon$$

Um resultado de normalidade assintótica é

$$\lim_{M \rightarrow \infty} \lim_{N \rightarrow \infty} P \left[\sqrt{N} \left(\frac{\hat{f}_M(\lambda) - f_M(\lambda)}{\sqrt{\Sigma_M(\lambda)}} \right) \leq x \right] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{1}{2}t^2} dt, \quad -\pi \leq \lambda \leq \pi, \quad (3.54)$$

onde a convergência é uniforme em $x \in (-\infty, \infty)$.

A "suavidade" da densidade espectral $\tilde{f}$ é indicada por um inteiro não negativo r e um número real ν de acordo com uma das seguintes condições:

- i) Para $r=0$, $\tilde{f}$ é suposta coincidir em quase toda parte com uma função contínua g satisfazendo

$$\sup_{t \leq h} \left(\int_{-\pi}^{\pi} |g(\lambda+t) + g(\lambda-t) - 2g(\lambda)|^2 d\lambda \right)^{\frac{1}{2}} = o(h^{\nu})$$

para algum $\nu \in (\frac{1}{2}, 1]$.

- ii) Para $r \geq 1$, $\tilde{f}$ é suposta coincidir em quase toda parte com uma função contínua g , a qual tem uma derivada absolutamente contínua $g^{(r-1)}$ e uma r -ésima derivada $g^{(r)}$ satisfazendo

$$\sup_{t \leq h} \left(\int_{-\pi}^{\pi} |g^{(r)}(\lambda+t) + g^{(r)}(\lambda-t) - 2g^{(r)}(\lambda)|^2 d\lambda \right)^{\frac{1}{2}} = o(h^{\nu})$$

para algum $\nu \in (0, 1]$.

Se é suposto que $\tilde{f}$ satisfaz ii) para $r \geq 2$, então

$$\lim_{M \rightarrow \infty} \frac{1}{M} \sum_M(\lambda) = \sum(\lambda) = \begin{cases} 2f^2(\lambda), & \lambda \neq 0 \text{ ou } \pm\pi \\ 4f^2(\lambda), & \lambda = 0 \text{ ou } \pm\pi \end{cases}$$

em quase toda parte, e

$$\lim_{M \rightarrow \infty} \lim_{N \rightarrow \infty} P \left[\sqrt{\frac{N}{M}} \left(\frac{\hat{f}_M(\lambda) - f_M(\lambda)}{\sqrt{\Sigma(\lambda)}} \right) \leq x \right] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{1}{2}t^2} dt,$$

em quase toda parte, $-\pi \leq \lambda \leq \pi$ (3.55)

onde a convergência é uniforme em $x \in (-\infty, \infty)$.

Se é suposto que f satisfaz i) e ii) para $r \geq 0$, então para qualquer $\epsilon > 0$

$$\lim_{M \rightarrow \infty} \lim_{N \rightarrow \infty} P[|\hat{f}_M(\lambda) - f(\lambda)| > \epsilon] = 0, \quad (3.56)$$

em quase toda parte, $-\pi \leq \lambda \leq \pi$.

Finalmente, se é suposto $r \geq 0$ e $\xi \in [0, r + \nu - \frac{1}{2})$, então

$$|\hat{f}_M(\lambda) - f(\lambda)| = o(M^{-\xi}), \quad (3.57)$$

em quase toda parte, $-\pi \leq \lambda \leq \pi$.

As relações (3.54) a (3.56) são afirmações de normalidade assintótica e consistência no limite iterado. A relação (3.57) fornece uma medida da razão de decréscimo do vício.

CAPÍTULO 4

DETERMINAÇÃO DA ORDEM DO PROCESSO AR A SER AJUSTADO À SÉRIE DE TEMPO

4.1 - INTRODUÇÃO

A ordem de uma equação de diferenças estocástica é determinada pelo mais alto "lag" o qual tem um coeficiente não nulo. Isto é, uma equação de diferenças estocástica é de ordem M se

$$X(t) = a_1 X(t-1) + a_2 X(t-2) + \dots + a_M X(t-M) + a_0 + \epsilon(t) \quad (4.1)$$

com $a_M \neq 0$.

Numa equação de diferenças de ordem M , $X(t)$ é afetado, no máximo, pelas M variáveis $X(t-1), X(t-2), \dots, X(t-M)$. Logo, em previsão é necessário usar apenas essas variáveis. Assim sendo, a determinação da ordem M do modelo auto-regressivo a ser ajustado à séries de tempo é de crucial importância.

Nós apresentaremos três métodos para a determinação da ordem, os quais foram propostos por Anderson (1963), Akaike (1970) e Parzen (1977).

Anderson (1963) desenvolveu um procedimento de múltipla decisão que consiste em testar seqüencialmente as hipóteses

$a_j = 0$, $j = p, p-1, \dots, m$, onde é estabelecido "a priori" que a "verdadeira" ordem M^* é tal que $m \leq M^* \leq p$.

Cada teste pode ser feito através da estatística

$$t_M = \frac{\hat{a}_M(M)}{\sqrt{\hat{\text{var}}(\hat{a}_M(M))}} \quad (4.2)$$

que tem aproximadamente uma distribuição normal.

Akaike(1970) desenvolveu um procedimento simples e prático que consiste em calcular a quantidade amostral

$$\text{FPE}(M) = [1+N^{-1}(M+1)][1-N^{-1}(M+1)]^{-1}S^2(M) \quad (4.3)$$

e escolher a ordem M^* que minimiza $\text{FPE}(M)$.

A estatística FPE é denominada erro final de previsão e é uma estimativa amostral da variância da previsão, do previsor um passo à frente para o modelo (4.1).

Parzen (1976) sugeriu usar a quantidade amostral

$$\text{CAT}(M) = \frac{1}{N} \sum_{j=0}^M \frac{N-j}{NS^2(j)} - \frac{N-M}{NS^2(M)} \quad (4.4)$$

e escolher M^* que minimiza essa quantidade. CAT representa "criterion autoregressive transfer function".

4.2 - O MÉTODO DE ANDERSON

Consideremos inicialmente o problema de determinar se a ordem é $M-1$ ou M . Isto é, decidir se $a_M = 0$. Esse problema po-

de ser colocado na forma de testar a hipótese nula $H_0: a_M = 0$ contra a alternativa $a_M \neq 0$.

O teste da razão de verossimilhança é análogo ao teste t bicaudal do problema da regressão, isto é, nós rejeitamos a hipótese nula se

$$\frac{|\hat{a}_M|}{s\sqrt{b^{*MM}}} > t(\epsilon), \quad (4.5)$$

onde

$$b^{*ij} = (b^*_{ij})^{-1}, \quad b^*_{ij} = \sum_{t=1}^N X(t-i)X(t-j) - N\bar{X}_{(i)}\bar{X}_{(j)},$$

$s^2 = NS^2(M)/(N-M-1)$ e $t(\epsilon)$ é o valor crítico da distribuição normal padronizada a um nível de significância ϵ . Para N grande, este teste tem, aproximadamente, nível de significância ϵ .

Suponha que a hipótese nula seja falsa. Em geral

$$p \lim_{N \rightarrow \infty} \left(\frac{1}{N} b^*_{ij} \right) = \Gamma \quad (4.6)$$

onde Γ é a matriz quadrada de ordem M , cujo elemento da linha i e coluna j é dado por

$$\begin{aligned} \gamma_{ij} &= E \left\{ [X(t-i) - E(X(t-i))] [X(t-j) - E(X(t-j))] \right\} = \\ &= E \left\{ [X(s) - E(X(s))] [X(s+i-j) - E(X(s+i-j))] \right\} = \\ &= \gamma(i-j), \end{aligned} \quad (4.7)$$

$$p \lim_{N \rightarrow \infty} s = \sigma \quad e \quad p \lim_{N \rightarrow \infty} \hat{a}_M = a_M$$

Então o teste estatístico comporta-se grosseiramente como $\sqrt{N}|a_M|/\sigma\sqrt{\gamma^{MM}}$, onde $\gamma^{ij} = (\gamma_{ij})^{-1}$. Este argumento pode ser tomado rigoroso para provar que o teste é consistente. Isto é, para uma dada alternativa $a_M \neq 0$, a probabilidade de rejeitar a hipótese nula aproxima-se de 1 quando N cresce.

Além disso, pelo Teorema 1 do Capítulo 3, sabemos que a matriz de covariâncias da distribuição normal limite de

$$\sqrt{N}[\hat{a}_1(M) - a_{1M}], \dots, \sqrt{N}[\hat{a}_M(M) - a_{MM}]$$

é $\sigma_M^2 \Gamma^{-1}$ onde $\Gamma = (\gamma_{ij})$ é a matriz dos elementos γ_{ij} definidos por (4.7). Então, como

$$\text{Var}\left\{\sqrt{N}[\hat{\underline{a}}(M) - \underline{a}_M]\right\} = N \text{Var}[\hat{\underline{a}}(M)] = \sigma_M^2 \Gamma^{-1},$$

temos

$$\text{Var}[\hat{\underline{a}}(M)] = \frac{1}{N} \sigma_M^2 \Gamma^{-1}$$

e como $\hat{\Gamma} = C(0)R_M$, onde R_M é a matriz de correlação amostral dada por (3.29), uma estimativa de $\text{Var}[\hat{\underline{a}}(M)]$ é dada por

$$\widehat{\text{Var}}[\hat{\underline{a}}(M)] = \frac{S^2(M)}{NC(0)} R_M^{-1} \quad (4.8)$$

onde $C(0)$ é dado por (3.27). Substituindo $S^2(M)$ pelo seu valor dado pela relação (3.45), temos

$$\widehat{\text{Var}}[\hat{\underline{a}}(M)] = \frac{C(0) - \sum_{j=1}^M \hat{a}_j(M)C(j)}{NC(0)} R_M^{-1} = \frac{1}{N} \left[1 - \sum_{j=1}^M \hat{a}_j(M)r_j \right] R_M^{-1} \quad (4.9)$$

Então,

$$\widehat{\text{Var}}[\hat{a}_{(M+1)}] = \frac{1}{N} \left[1 - \sum_{j=1}^{M+1} \hat{a}_j (M+1) r_j \right] R_{M+1}^{-1}$$

e calculando a inversa da matriz simétrica particionada R_{M+1} (Searle, 1971, seção 1.7), temos

$$\begin{aligned} \text{Var}[\hat{a}_{M+1}(M+1)] &= \frac{1}{N} \frac{\left[1 - \sum_{j=1}^{M+1} \hat{a}_j (M+1) r_j \right]}{1 - \tilde{r}_M' R_M^{-1} \tilde{r}_M} = \\ &= \frac{1}{N} \frac{\left[1 - \sum_{j=1}^{M+1} \hat{a}_j (M+1) r_j \right]}{1 - \sum_{j=1}^M \hat{a}_j (M) r_j} = \\ &= \frac{1}{N} \frac{\left[C(0) - \sum_{j=1}^{M+1} \hat{a}_j (M+1) C(j) \right]}{C(0) - \sum_{j=1}^M \hat{a}_j (M) C(j)} \\ &= \frac{1}{N} \frac{S^2(M+1)}{S^2(M)} \end{aligned} \quad (4.10)$$

Observemos que esta estimativa é obtida substituindo σ_M^2 por $S^2(M)$. Poderia ser usada uma estimativa não viciada de σ_M^2 , dada por $NS^2(M)/(N-M-1)$.

Dessa forma, usando a estatística definida por (4.5), podemos introduzir um procedimento de múltipla decisão. Tal procedimento consiste em especificar uma máxima ordem p e uma mínima ordem m ($0 \leq m < p$) e testar sucessivamente as hipóteses $a_p = 0$, $a_{p-1} = 0$, etc., até que a hipótese $a_{M^*} = 0$ seja rejeitada. Então

M^* será a ordem do modelo auto-regressivo a ser ajustado à série. (Esse procedimento é descrito detalhadamente em Anderson, 1971).

Logo, para a determinação da ordem de dependência entre as observações, com base nas observações $X(1), X(2), \dots, X(N)$, nós supomos que se possa especificar uma ordem de dependência mínima m e uma ordem de dependência máxima p , e se queira decidir se a ordem de dependência é $m, m+1, \dots, p-1$ ou p . Se houver independência, m será zero, mas quase sempre há razão para se esperar que a ordem de dependência seja pelo menos algum inteiro positivo m .

Para N grande a teoria normal pode ser usada. A máxima ordem de dependência p pode ser fixada como sendo muito grande, como por exemplo $N/10$ ou $N/4$, se ϵ_i , $i = p, p-1, \dots, m+1$, forem tomados muito pequenos, tal como da ordem de 10^{-3} .

Para testar a hipótese $a_M = 0$, com base numa amostra grande, precisamos apenas de $\hat{a}_M(M)$. Logo, para realizar esse procedimento de múltipla decisão precisamos apenas de $\hat{a}_j(j)$ para $j = p$ e para valores de j menores que p , até que a hipótese $a_j = 0$ seja rejeitada. O cálculo dessas estimativas dos coeficientes, com base em $r_1, \dots, r_p$, é feito através das relações (3.36) e (3.37) ou resolvendo a equação vetorial (3.30).

4.3 - O MÉTODO DE AKAIKE

4.3.1 - Definição de FPE de um Previsor

Daremos as bases teórica e empírica de um procedimento

simples e prático para a determinação da ordem do processo auto-regressivo a ser ajustado à série de tempo, desenvolvido por Akaike (1969, 1970a, 1970b).

O procedimento é baseado na precisão do previsor, medida através do chamado erro final de previsão (FPE), o qual é definido como o erro quadrático médio de previsão do previsor, isto é, para um previsor $\hat{X}(t)$ de $X(t)$,

$$\text{FPE de } \hat{X}(t) = E[X(t) - \hat{X}(t)]^2. \quad (4.11)$$

A determinação da ordem é realizada ajustando modelos auto-regressivos de sucessivas ordens dentro de uma faixa pré-estabelecida, calculando estimativas do FPE para os modelos, e adotando aquele com a estimativa mínima.

Em situações práticas $\hat{X}(t)$ é dada como uma função de valores recentes de $X(t)$ e a estrutura ou os parâmetros da função são determinados, ou identificados, usando toda a história passada de $X(t)$. Supondo que a dependência dessa estrutura identificada aos valores recentes de $X(t)$, os quais são usados para se obter $\hat{X}(t)$, seja decrescente quando se aumenta o comprimento da história passada usada para a identificação, nós consideraremos a situação idealizada onde a dependência se anula completamente.

Isto é equivalente à situação onde a estrutura do previsor é identificada usando uma observação do processo $X(t)$ e, usando essa estrutura, a previsão é feita com outro processo $\hat{Y}(t)$ o qual é independente de $X(t)$, mas possui as mesmas pro-

priedades estatísticas de $X(t)$.

Na notação que usaremos, vamos denotar por $[u(\ell)]$ o vetor coluna de elementos $u(\ell)$, $(\ell = 1, 2, \dots, M)$ e por v ou $[v(\ell, m)]$ a matriz com o elemento da linha ℓ e coluna m , igual a $v(\ell, m)$, $(\ell, m = 1, 2, \dots, M)$. Quando a dimensão M foi de interesse especial será adicionado o índice M e então u_M e v_M serão usados para u e v acima. O símbolo $'$ denota a transposta de uma matriz ou vetor.

Quando o processo $X(t)$ é estacionário e o previsor $\hat{Y}(t)$ de $Y(t)$ é linear e dado por

$$\hat{Y}(t) = \sum_{m=1}^M \hat{a}_M(m) Y(t-m) + \hat{a}_M(0), \quad (4.12)$$

onde $\hat{a}_M(m)$ é uma função de $\{X(t)\}$, nós temos

$$\text{FPE de } \hat{Y}(t) = \sigma^2(M) + \sum_{\ell=0}^M \sum_{m=0}^M E[\Delta a_M(\ell) \Delta a_M(m)] V_{M+1}(\ell, m) \quad (4.13)$$

onde

$$\begin{aligned} \sigma^2(M) &= E \left[Y(t) - \sum_{m=1}^M a_M(m) Y(t-m) - a_M(0) \right]^2 = \\ &= \min_{[a(m)]} E \left[Y(t) - \sum_{m=1}^M a(m) Y(t-m) - a(0) \right]^2, \end{aligned}$$

$$V_{M+1}(\ell, m) = E[Y(t-\ell) Y(t-m)], \quad \ell, m = 1, 2, \dots, M$$

$$V_{M+1}(0, m) = V_{M+1}(m, 0) = EY(t), \quad m = 1, 2, \dots, M \quad (4.15)$$

$$V_{M+1}(0, 0) = 1,$$

e

$$\Delta a_M(m) = \hat{a}_M(m) - a_M(m), \quad m=0, 1, \dots, M, \quad (4.16)$$

onde $a_M(m)$ está definido em (4.14) e fornece o melhor (no sentido de média quadrática) preditor linear.

De fato, temos:

$$\begin{aligned} \text{FPE de } \hat{Y}(t) &= E[Y(t) - \hat{Y}(t)]^2 \\ &= E\left[Y(t) - \sum_{m=1}^M \hat{a}_M(m)Y(t-m) - \hat{a}_M(0)\right]^2. \end{aligned}$$

Mas, fazendo

$$\Delta a_M(m) = \hat{a}_M(m) - a_M(m), \quad m=0, 1, \dots, M,$$

temos

$$\begin{aligned} \text{FPE de } \hat{Y}(t) &= E\left\{Y(t) - \sum_{m=1}^M [\Delta a_M(m) + a_M(m)]Y(t-m) - \hat{a}_M(0)\right\}^2 = \\ &= E\left\{\left[Y(t) - \sum_{m=1}^M a_M(m)Y(t-m) - a_M(0)\right] - \left[\Delta a_M(0) + \sum_{m=1}^M \Delta a_M(m)Y(t-m)\right]\right\}^2 = \\ &= E\left[Y(t) - \sum_{m=1}^M a_M(m)Y(t-m) - a_M(0)\right]^2 + E\left[\Delta a_M(0) + \sum_{m=1}^M \Delta a_M(m)Y(t-m)\right]^2 - \\ &- 2E\left\{\left[Y(t) - \sum_{m=1}^M a_M(m)Y(t-m) - a_M(0)\right] \left[\Delta a_M(0) + \sum_{m=1}^M \Delta a_M(m)Y(t-m)\right]\right\}, \end{aligned}$$

fazendo

$$\sigma^2(M) = E\left[Y(t) - \sum_{m=1}^M a_M(m)Y(t-m) - a_M(0)\right]^2 =$$

$$= \min_{[a(m)]} E \left[Y(t) - \sum_{m=1}^M a(m) Y(t-m) - a(0) \right]^2$$

e observando que a esperança do duplo produto se anula, pois $\hat{a}_M(m)$ é função de $X(t)$, que por sua vez é independente de $Y(t)$ e $E[\Delta a_M(m)] = E[\hat{a}_M(m) - a_M(m)] = 0$ (pelo Teorema 1 do Capítulo 3, $\hat{a}_M(m)$ é um estimador assintoticamente não viciado de $a_M(m)$), temos

$$\begin{aligned} \text{FPE de } \hat{Y}(t) &= \sigma^2(M) + E[\Delta a_M(0)]^2 + \\ &+ E \left[\sum_{\ell=1}^M \sum_{m=1}^M \Delta a_M(\ell) \Delta a_M(m) Y(t-\ell) Y(t-m) \right] + \\ &+ 2E \left[\Delta a_M(0) \sum_{m=1}^M \Delta a_M(m) Y(t-m) \right] = \\ &\stackrel{\text{indep}}{=} \sigma^2(M) + E[\Delta a_M(0)]^2 + \\ &+ \sum_{\ell=1}^M \sum_{m=1}^M E[\Delta a_M(\ell) \Delta a_M(m)] \cdot E[Y(t-\ell) Y(t-m)] + \\ &+ 2 \sum_{m=1}^M E \Delta a_M(0) \Delta a_M(m) E[Y(t-m)]. \end{aligned}$$

Como $X(t)$ é estacionário e $Y(t)$ tem as mesmas propriedades estatísticas de $X(t)$, $Y(t)$ também é estacionário, e então,

$$E[Y(t-m)] = E[Y(t)], \quad m=1, 2, \dots, M$$

logo,

$$\text{FPE de } \hat{Y}(t) = \sigma^2(M) + \sum_{\ell=0}^M \sum_{m=0}^M E[\Delta a_M(\ell) \Delta a_M(m)] V_{M+1}(\ell, m)$$

onde

$$V_{M+1}(\ell, m) = E[Y(t-\ell)Y(t-m)], \quad \ell, m=1, 2, \dots, M$$

$$V_{M+1}(0, m) = V_{M+1}(m, 0) = E[Y(t-m)] = E[Y(t)], \quad m=1, 2, \dots, M$$

e

$$V_{M+1}(0, 0) = 1. \quad \text{c.q.d.}$$

A relação (4.13) mostra que o FPE é composto de duas componentes: a primeira delas correspondente ao FPE do melhor previsor linear para um dado M e a segunda delas devida ao desvio estatístico de $\hat{a}_M(m)$ em relação a $a_M(m)$. Geralmente, conforme o valor de M cresce, o primeiro termo $\sigma^2(M)$ decresce, mas o segundo termo cresce para um número finito de observações de $X(t)$.

Dado um conjunto de previsores, a definição de FPE sugere naturalmente que adotemos como previsor ótimo aquele com o FPE mínimo. O nosso objetivo será a realização de uma boa aproximação da escolha ótima de M sob a hipótese de estacionaridade e no caso linear descrito acima, usando a informação obtida de um número finito de observações de $X(t)$.

4.3.2 - FPE das Estimativas de Mínimos Quadrados de um Modelo Auto-regressivo

Daqui em diante nós assumiremos que $X(t)$ é um processo au-

to-regressivo estacionário gerado pela relação

$$X(t) = \sum_{m=1}^M a(m)X(t-m) + a(0) + \epsilon(t), \quad (4.17)$$

onde os $\epsilon(t)$'s são ruídos brancos com $E[\epsilon(t)] = 0$ e $E[\epsilon^2(t)] = \sigma^2$.

Dado um conjunto de dados $\{X(t); t = -M+1, -M+2, \dots, N\}$, o vetor de parâmetros $[\hat{a}_M(m)]$ do previsor é definido como a estimativa de mínimos quadrados de $[a(m)]$, isto é, $\hat{a}_M(m)$, $m = 1, 2, \dots, M$ são as soluções das equações (3.23), ou de modo equivalente, de

$$\sum_{m=1}^M C_{XX}(m, \ell) \hat{a}_M(m) = C_{XX}(0, \ell), \quad \ell = 1, 2, \dots, M, \quad (4.18)$$

onde

$$C_{XX}(m, \ell) = \frac{1}{N} \sum_{t=1}^N [X(t-m) - \bar{X}_m][X(t-\ell) - \bar{X}_\ell],$$

$$\bar{X}_m = \frac{1}{N} \sum_{t=1}^N X(t-m), \quad m = 0, 1, 2, \dots, M$$

e

$$\hat{a}_M(0) = \bar{X}_0 - \sum_{m=1}^M \hat{a}_M(m) \bar{X}_m \quad (4.19)$$

Seguindo a definição de $Y(t)$ dada na secção anterior, o previsor $\hat{Y}(t)$ de $Y(t)$ é dado nesse caso por

$$\begin{aligned} \hat{Y}(t) &= \sum_{m=1}^M \hat{a}_M(m) Y(t-m) + \hat{a}_M(0) = \\ &= \sum_{m=1}^M \hat{a}_M(m) Y(t-m) + \bar{X}_0 - \sum_{m=1}^M \hat{a}_M(m) \bar{X}_m \end{aligned}$$

logo,

$$\hat{Y}(t) = \sum_{m=1}^M \hat{a}_M(m) [Y(t-m) - \bar{X}_m] + \bar{X}_0. \quad (4.20)$$

Nós estamos supondo que $Y(t)$ é gerado pela relação

$$Y(t) = \sum_{m=1}^M a_m Y(t-m) + a_0 + \delta(t),$$

onde $\delta(t)$ tem as mesmas propriedades estatísticas de $\varepsilon(t)$. Nós temos

$$Y(t) - \hat{Y}(t) = \delta(t) - \sum_{m=1}^M \Delta a_M(m) y(t-m) - \left[\Delta \bar{X}_0 - \sum_{m=1}^M \hat{a}_M(m) \Delta \bar{X}_m \right],$$

onde

$$y(t) = Y(t) - E[Y(t)] \quad \text{e} \quad \Delta \bar{X}_\ell = \bar{X}_\ell - E[X(t)].$$

De fato, temos

$$\begin{aligned} Y(t) - \hat{Y}(t) &= \sum_{m=1}^M a_m Y(t-m) + a_0 + \delta(t) - \sum_{m=1}^M \hat{a}_M(m) [Y(t-m) - \bar{X}_m] - \bar{X}_0 = \\ &= \sum_{m=1}^M [a_m - \hat{a}_M(m)] Y(t-m) + \sum_{m=1}^M \hat{a}_M(m) \bar{X}_m + \delta(t) - \bar{X}_0 + a_0. \end{aligned}$$

fazendo

$$y(t) = Y(t) - E[Y(t)] \quad \text{e} \quad \Delta \bar{X}_\ell = \bar{X}_\ell - E[X(t)], \text{ temos}$$

$$\begin{aligned} Y(t) - \hat{Y}(t) &= \sum_{m=1}^M [a_m - \hat{a}_M(m)] [y(t-m) + E(Y(t-m))] + \sum_{m=1}^M \hat{a}_M(m) [\Delta \bar{X}_m + E(X(t))] + \delta(t) - \\ &= [\Delta \bar{X}_0 + E(X(t))] + a_0 = \sum a_m y(t-m) + \sum a_m E(Y(t-m)) - \sum \hat{a}_M(m) y(t-m) - \\ &= \sum \hat{a}_M(m) E(Y(t-m)) + \sum \hat{a}_M(m) \Delta \bar{X}_m + \sum \hat{a}_M(m) E(X(t)) + \delta(t) - \Delta \bar{X}_0 - E(X(t)) + a_0. \end{aligned}$$

Por outro lado, temos $Y(t) = \sum_m a_m Y(t-m) + a_0 + \delta(t)$ e como $\delta(t)$ tem as mesmas propriedades estatísticas de $\varepsilon(t)$, temos

$$E[\delta(t)] = 0 \quad \text{e} \quad E[Y(t)] = \sum_m a_m E[Y(t-m)] + a_0,$$

então

$$a_0 = E[Y(t)] - \sum_m a_m E[Y(t-m)],$$

o qual substituído na expressão acima, produz

$$\begin{aligned} Y(t) - \hat{Y}(t) &= \sum_m a_m Y(t-m) + \sum_m a_m E[Y(t-m)] - \sum_m \hat{a}_M(m) Y(t-m) - \sum_m \hat{a}_M(m) E[Y(t-m)] + \\ &+ \sum_m \hat{a}_M(m) \Delta \bar{X}_m + \sum_m \hat{a}_M(m) E[X(t)] + \delta(t) - \Delta \bar{X}_0 - E[X(t)] + E[Y(t)] - \sum_m a_m E[Y(t-m)]. \end{aligned}$$

Finalmente, como $Y(t)$ tem as mesmas propriedades estatísticas $X(t)$, ambos são estacionários e tem médias iguais, então

$$E[X(t)] = E[X(t-m)] = E[Y(t-m)] = E[Y(t)],$$

e portanto

$$Y(t) - \hat{Y}(t) = \delta(t) - \sum_m [\hat{a}_M(m) - a_m] Y(t-m) - \left[\Delta \bar{X}_0 - \sum_m \hat{a}_M(m) \Delta \bar{X}_m \right]$$

ou

$$Y(t) - \hat{Y}(t) = \delta(t) - \sum_{m=1}^M \Delta a_M(m) Y(t-m) - \left[\Delta \bar{X}_0 - \sum_{m=1}^M \hat{a}_M(m) \Delta \bar{X}_m \right]$$

onde

$$\Delta a_M(m) = \hat{a}_M(m) - a_m, \quad y(t) = Y(t) - E[Y(t)]$$

e

$$\Delta \bar{X}_\ell = \bar{X}_\ell - E[X(t)].$$

c.q.d.

Levando em conta a independência de $y(t)$ de Δa_m e $\Delta \bar{X}_\ell$, nós obtemos:

$$\begin{aligned}
 \text{FPE de } \hat{Y}(t) &= E[Y(t) - \hat{Y}(t)]^2 = \\
 &= \sigma^2 + \sum_{m=1}^M \sum_{\ell=1}^M E[\Delta a_M(m) \Delta a_M(\ell)] \gamma_{XX}(\ell-m) + \\
 &+ E\left[\Delta \bar{X}_0 - \sum_{m=1}^M \hat{a}_M(m) \Delta \bar{X}_m\right]^2 \quad (4.21)
 \end{aligned}$$

onde

$$\gamma_{XX}(\ell-m) = E[X(t-\ell)X(t-m)] - [E(X(t))]^2.$$

De fato, temos

$$\begin{aligned}
 \text{FPE de } \hat{Y}(t) &= E[Y(t) - \hat{Y}(t)]^2 = \\
 &= E\left[\delta(t) - \sum_{m=1}^M \Delta a_M(m) y(t-m) - \left(\Delta \bar{X}_0 - \sum_{m=1}^M \hat{a}_M(m) \Delta \bar{X}_m\right)\right]^2 = \\
 &\stackrel{\text{indep}}{=} E[\delta^2(t)] + E \sum_{m=1}^M \sum_{\ell=1}^M \Delta a_M(m) \Delta a_M(\ell) y(t-m) y(t-\ell) + \\
 &+ E\left[\Delta \bar{X}_0 - \sum_{m=1}^M \hat{a}_M(m) \Delta \bar{X}_m\right]^2 = \\
 &= \sigma^2 + \sum_{m=1}^M \sum_{\ell=1}^M E \Delta a_M(m) \Delta a_M(\ell) E y(t-m) y(t-\ell) + \\
 &= E\left[\Delta \bar{X}_0 - \sum_{m=1}^M \hat{a}_M(m) \Delta \bar{X}_m\right]^2.
 \end{aligned}$$

Mas,

$$\begin{aligned}
 E[y(t-m)y(t-\ell)] &= E[Y(t-m) - EY(t-m)][Y(t-\ell) - EY(t-\ell)] = \\
 &= E[Y(t-m)Y(t-\ell)] - E[Y(t-m)EY(t-\ell)] - \\
 &- E[EY(t-m)Y(t-\ell)] + E[EY(t-m)EY(t-\ell)]
 \end{aligned}$$

e como $E[Y(t)] = \text{constante}$, temos

$$\begin{aligned} E[y(t-m)y(t-l)] &= E[Y(t-m)Y(t-l)] - EY(t-l)EY(t-m) - \\ &\quad - EY(t-m)EY(t-l) + E[EY(t-m)EY(t-l)] = \\ &= EY(t-m)Y(t-l) - [EY(t)]^2 - [EY(t)]^2 + [EY(t)]^2 = \\ &= E[Y(t-m)Y(t-l)] - [EY(t)]^2 \end{aligned}$$

e como $Y(t)$ tem as mesmas propriedades estatísticas que $X(t)$, temos

$$\begin{aligned} E[y(t-m)y(t-l)] &= E[X(t-m)X(t-l)] - [EX(t)]^2 = \\ &= \gamma_{XX}(\ell-m). \end{aligned}$$

Logo

$$\begin{aligned} \text{FPE de } \hat{Y}(t) &= \sigma^2 + \sum_{m=1}^M \sum_{\ell=1}^M E[\Delta a_M(m) \Delta a_M(\ell)] \gamma_{XX}(\ell-m) + \\ &\quad + E \left[\Delta \bar{X}_0 - \sum_{m=1}^M \hat{a}_M(m) \Delta \bar{X}_m \right]^2. \quad \text{c.q.d.} \end{aligned}$$

Para o cálculo assintótico do FPE de $\hat{Y}(t)$, nós faremos uso do seguinte teorema, o qual é derivado do Teorema 1 do capítulo anterior.

TEOREMA 1 - Sob a presente suposição sobre $X(t)$, a distribuição limite de

$$\sqrt{N} \Delta \bar{X}_0 = \sqrt{N} [\bar{X}_0 - E(X(t))] \quad \text{e} \quad \sqrt{N} \Delta a_M(m) = \sqrt{N} [\hat{a}_M(m) - a(m)]$$

$(m=1,2,\dots,M)$, quando N tende ao infinito, é uma distribuição Gaussiana de dimensão $(M+1)$ com média zero e matriz de covariâncias

$$\sigma^2 \begin{bmatrix} \delta^{-2} & 0_M' \\ 0_M & \Gamma_{MM}^{-1} \end{bmatrix} \quad (4.22)$$

onde

$$\delta = 1 - \sum_{m=1}^M a(m),$$

Γ_{MM} é uma matriz de ordem $M \times M$ de elementos $\gamma(\ell, m) = \gamma_{XX}(\ell - m)$ e 0 indica um vetor de zeros.

Da ergodicidade do processo $X(t)$ nós sabemos que $C_{XX}(\ell, m)$ converge para $\gamma_{XX}(\ell - m)$, quando N tende ao infinito, com probabilidade um. Então $\hat{a}_M$ é uma estimativa consistente de a_M , nesse caso com convergência com probabilidade um.

Da relação (4.18) nós temos, para $\ell = 1, 2, \dots, M$,

$$[\hat{a}_M(\ell)] = [C_{XX}(m, \ell)]^{-1} [C_{XX}(0, \ell)]. \quad (4.23)$$

Então nós obtemos (Akaike, 1969),

$$[\Delta a_M(\ell)] = [C_{XX}(m, \ell)]^{-1} [C_{\epsilon X}(\ell)]$$

onde

$$C_{\epsilon X}(\ell) = N^{-1} \sum_{t=1}^N \epsilon(t) [X(t-\ell) - \bar{X}_\ell]. \quad (4.24)$$

Da consistência de $C_{xx}(m, \ell)$ nós sabemos que a distribuição limite de $\sqrt{N}\Delta\bar{X}_0$ e $\sqrt{N}\Delta a_M$ é idêntica à distribuição limite de $\sqrt{N}\Delta\bar{X}_0$ e $N\Gamma_{MM}^{-1}C_{ex}$ (ver Teorema 1 do Capítulo anterior). Aplicando o Teorema do Limite Central de Diananda para um processo estacionário m-dependente, conforme foi feito por Anderson & Walker (1964), nós podemos obter

LEMA 1 - A distribuição limite de $\sqrt{N}\Delta\bar{X}_0$ e $\sqrt{N}C_{ex}$, quando N tende ao infinito, é Gaussiana (M+1)-dimensional com média zero e matriz de covariâncias

$$\sigma^2 \begin{bmatrix} \delta^{-2} & 0'_M \\ 0_M & \Gamma_{MM} \end{bmatrix}.$$

Observemos que assim como a densidade espectral de $X(t) - E[X(t)]$ na frequência zero é $\sigma^2\delta^{-2}$, onde

$$\delta = 1 - \sum_{m=1}^M a_M(m),$$

a variância de distribuição limite de $\sqrt{N}\Delta\bar{X}_0$ também é igual a $\sigma^2\delta^{-2}$.

A afirmação do Teorema 1 é uma consequência direta desse Lema e da observação seguinte à relação (4.24).

Retornemos agora ao cálculo do FPE de $\hat{Y}(t)$. Ao invés de tomar a esperança de $[Y(t) - \hat{Y}(t)]^2$ diretamente, tomemos a esperança condicional de $[Y(t) - \hat{Y}(t)]^2$ para um dado $X(t)$. Nós denotaremos essa esperança condicional por $E_X[Y(t) - \hat{Y}(t)]^2$. Da in-

dependência de $Y(t)$ e $X(t)$ nós temos

$$\begin{aligned} E_X[Y(t) - \hat{Y}(t)]^2 &= \sigma^2 + \sum_{m=1}^M \sum_{\ell=1}^M \Delta a_M(m) \Delta a_M(\ell) \gamma_{XX}(\ell-m) + \\ &+ \left[\Delta \bar{X}_0 - \sum_{m=1}^M \hat{a}_M(m) \Delta \bar{X}_m \right]^2. \end{aligned} \quad (4.25)$$

Levando em conta o fato que no limite as diferenças entre $\sqrt{N}\Delta\bar{X}_0$ e $\sqrt{N}\Delta\bar{X}_m$ ($m=1,2,\dots,M$) são estocasticamente nulas, nós podemos concluir do Teorema 1 que $NE_X\{[Y(t) - \hat{Y}(t)]^2 - \sigma^2\}$ tem uma distribuição limite com esperança igual a $(M+1)\sigma^2$.

De fato, temos

$$\begin{aligned} N\{E_X[Y(t) - \hat{Y}(t)]^2 - \sigma^2\} &= N \sum_m \sum_{\ell} \Delta a_M(m) \Delta a_M(\ell) \gamma_{XX}(\ell-m) + \\ &+ N[\Delta \bar{X}_0 - \sum_m \hat{a}_M(m) \Delta \bar{X}_m]^2 = \\ &= \sum_m \sum_{\ell} \sqrt{N} \Delta a_M(m) \sqrt{N} \Delta a_M(\ell) \gamma_{XX}(\ell-m) + \\ &+ [\sqrt{N} \Delta \bar{X}_0 - \sum_m \sqrt{N} \hat{a}_M(m) \Delta \bar{X}_m]^2, \end{aligned}$$

mas

$$\begin{aligned} E_{\infty} \left[\sum_m \sum_{\ell} \sqrt{N} \Delta a_M(m) \sqrt{N} \Delta a_M(\ell) \gamma_{XX}(\ell-m) \right] &= \\ &= \sum_{m=1}^M \sum_{\ell=1}^M \gamma_{XX}(\ell-m) E_{\infty} [\sqrt{N} \Delta a_M(m) \sqrt{N} \Delta a_M(\ell)]. \end{aligned}$$

Denotando por $R_{XX}(\ell-m)$ ao elemento da linha ℓ e coluna

m da matriz $R_{MM} = \Gamma_{MM}^{-1}$ temos, pelo Teorema 1, que

$$E_{\infty}[\sqrt{N}\Delta a_M(m)\sqrt{N}\Delta a_M(\ell)] = \sigma^2 R_{XX}(\ell-m), \text{ para } m, \ell=1, 2, \dots, M,$$

logo

$$\begin{aligned} E_{\infty} \left[\sum_{m=1}^M \sum_{\ell=1}^M \sqrt{N}\Delta a_M(m) \sqrt{N}\Delta a_M(\ell) \gamma_{XX}(\ell-m) \right] &= \sum_{m=1}^M \sum_{\ell=1}^M \gamma_{XX}(\ell-m) \sigma^2 R_{XX}(\ell-m) = \\ &= \sigma^2 \sum_{m=1}^M \sum_{\ell=1}^M \gamma_{XX}(\ell-m) R_{XX}(\ell-m) \end{aligned}$$

levando em conta o fato que as matrizes Γ_{MM} e R_{MM} são simétricas e sendo $(u_{m,m})$ o m-ésimo elemento diagonal da matriz identidade $I = \Gamma_{MM} R_{MM}$, temos

$$\begin{aligned} E_{\infty} \left[\sum_{m=1}^M \sum_{\ell=1}^M \sqrt{N}\Delta a_M(m) \sqrt{N}\Delta a_M(\ell) \gamma_{XX}(\ell-m) \right] &= \\ &= \sigma^2 \sum_{m=1}^M \left[\sum_{\ell=1}^M \gamma_{XX}(m, \ell) R_{XX}(\ell, m) \right] = \\ &= \sigma^2 \sum_{m=1}^M (u_{m,m}) = \\ &= M\sigma^2. \end{aligned}$$

Como $\hat{a}_M(m)$ é uma estimativa consistente de $a(m)$ e como $\sqrt{N}\Delta \bar{X}_m$ tende para $\sqrt{N}\Delta \bar{X}_0$ ($m=1, 2, \dots, M$) quando $N \rightarrow \infty$, temos ainda

$$\begin{aligned} E_{\infty} \left[\sqrt{N}\Delta \bar{X}_0 - \sum_{m=1}^M \sqrt{N}\hat{a}_M(m) \Delta \bar{X}_m \right]^2 &= E_{\infty} \left\{ \sqrt{N}\Delta \bar{X}_0 \left[1 - \sum_{m=1}^M a(m) \right] \right\}^2 = \\ &= E_{\infty} \left[\sqrt{N}\Delta \bar{X}_0 \right]^2 \delta^2 = \end{aligned}$$

$$\underset{\text{Teor.1}}{=} \sigma^2 \delta^{-2} \delta^2 = \sigma^2$$

logo,

$$E_{\infty} \left\{ N [E_X (Y(t) - \hat{Y}(t))^2 - \sigma^2] \right\} = (M+1) \sigma^2$$

c.q.d.

Esse resultado sugere a seguinte definição de $(FPE)_M$, como uma avaliação assintótica do FPE de $\hat{Y}(t)$:

$$(FPE)_M \text{ de } \hat{Y}(t) = \left(1 + \frac{M+1}{N} \right) \sigma^2. \quad (4.26)$$

O procedimento de identificação da ordem do predictor será baseado em alguma estimativa de $(FPE)_M$.

4.3.3 - UMA ESTIMATIVA DE $(FPE)_M$ E O PROCEDIMENTO DE MÍNIMO FPE

Conforme as observações seguintes à relação (3.45) do Capítulo anterior, sabemos que

$$S^2(M) = C_{XX}(0,0) - \sum_{\ell=1}^M \hat{a}_M(\ell) C_{XX}(0,\ell) \quad (4.27)$$

$$\sum_{m=1}^M a_M(m) C_{XX}(m, \ell) = C_{XX}(0, \ell) - \sum_{m=1}^M C_{XX}(m, \ell) a(m),$$

também obtida de (4.18), nós obtemos

$$\begin{aligned} S^2(M) &= C_{XX}(0, 0) - 2 \sum_{m=1}^M a(m) C_{XX}(0, m) + \sum_{\ell=1}^M \sum_{m=1}^M a(\ell) a(m) C_{XX}(m, \ell) - \\ &= \sum_{\ell=1}^M \sum_{m=1}^M \Delta a_M(\ell) \Delta a_M(m) C_{XX}(m, \ell). \end{aligned} \quad (4.28)$$

De fato, temos

$$S^2(M) = C_{XX}(0, 0) - \sum_{\ell=1}^M \sum_{m=1}^M \hat{a}_M(\ell) C_{XX}(\ell, m) \hat{a}_M(m);$$

lembrando que

$$\Delta a_M(m) = \hat{a}_M(m) - a(m)$$

e usando a relação

$$\sum_{m=1}^M \Delta a_M(m) C_{XX}(m, \ell) = C_{XX}(0, \ell) - \sum_{m=1}^M C_{XX}(m, \ell) a(m),$$

temos

$$\begin{aligned} S^2(M) &= C_{XX}(0, 0) - \sum_{\ell=1}^M \sum_{m=1}^M \hat{a}_M(\ell) C_{XX}(\ell, m) \hat{a}_M(m) = \\ &= C_{XX}(0, 0) - \sum_{\ell} \sum_m [\Delta a_M(\ell) + a(\ell)] C_{XX}(\ell, m) [\Delta a_M(m) + a(m)] = \\ &= C_{XX}(0, 0) - \sum_{\ell} \sum_m [\Delta a_M(\ell) C_{XX}(\ell, m) + a(\ell) C_{XX}(\ell, m)] [\Delta a_M(m) + a(m)] = \end{aligned}$$

$$\begin{aligned}
 &= C_{XX}(0,0) - \sum_{\ell} \sum_m \Delta a_M(\ell) \Delta a_M(m) C_{XX}(m,\ell) - \sum_{\ell} \sum_m a(\ell) C_{XX}(m,\ell) \Delta a_M(m) - \\
 &- \sum_{\ell} \sum_m a(m) \Delta a_M(\ell) C_{XX}(m,\ell) - \sum_{\ell} \sum_m a(\ell) a(m) C_{XX}(m,\ell) = \\
 &= C_{XX}(0,0) - \sum_{\ell=1}^M \sum_{m=1}^M \Delta a_M(\ell) \Delta a_M(m) C_{XX}(m,\ell) - \\
 &- \sum_{\ell=1}^M a(\ell) \left[\sum_{m=1}^M \Delta a_M(m) C_{XX}(m,\ell) \right] - \sum_{m=1}^M a(m) \left[\sum_{\ell=1}^M \Delta a_M(\ell) C_{XX}(m,\ell) \right] - \\
 &- \sum_{\ell=1}^M \sum_{m=1}^M a(\ell) a(m) C_{XX}(m,\ell) = C_{XX}(0,0) - \\
 &- \sum_{\ell=1}^M \sum_{m=1}^M \Delta a_M(\ell) \Delta a_M(m) C_{XX}(m,\ell) - \sum_{\ell=1}^M a(\ell) \left[C_{XX}(0,\ell) - \right. \\
 &- \left. \sum_{m=1}^M C_{XX}(m,\ell) a(m) \right] - \sum_{m=1}^M a(m) \left[C_{XX}(0,m) - \sum_{\ell=1}^M C_{XX}(m,\ell) a(\ell) \right] - \\
 &- \sum_{\ell=1}^M \sum_{m=1}^M a(\ell) a(m) C_{XX}(m,\ell) = C_{XX}(0,0) - \sum_{\ell} \sum_m \Delta a_M(\ell) \Delta a_M(m) C_{XX}(m,\ell) - \\
 &- \sum_{\ell=1}^M a(\ell) C_{XX}(0,\ell) + \sum_{\ell} \sum_m a(\ell) a(m) C_{XX}(m,\ell) - \sum_{m=1}^M a(m) C_{XX}(0,m) + \\
 &+ \sum_{\ell} \sum_m a(\ell) a(m) C_{XX}(m,\ell) - \sum_{\ell} \sum_m a(\ell) a(m) C_{XX}(m,\ell) = C_{XX}(0,0) - \\
 &- 2 \sum_{m=1}^M a(m) C_{XX}(0,m) + \sum_{\ell=1}^M \sum_{m=1}^M a(\ell) a(m) C_{XX}(m,\ell) - \\
 &- \sum_{\ell=1}^M \sum_{m=1}^M \Delta a_M(\ell) \Delta a_M(m) C_{XX}(m,\ell).
 \end{aligned}$$

c.q.d.

Da definição de $C_{xx}(m, \ell)$, nós temos

$$\begin{aligned} H(M) &= C_{xx}(0,0) - 2 \sum_{m=1}^M a(m) C_{xx}(0,m) + \sum_{\ell=1}^M \sum_{m=1}^M a(m) a(\ell) C_{xx}(m,\ell) = \\ &= \frac{1}{N} \sum_{t=1}^N [\varepsilon(t) - \bar{\varepsilon}]^2. \end{aligned} \quad (4.29)$$

De fato, temos

$$\begin{aligned} H(M) &= \frac{1}{N} \sum_{t=1}^N [X(t) - \bar{X}_0]^2 - 2 \sum_{m=1}^M a(m) \frac{1}{N} \sum_{t=1}^N [X(t) - \bar{X}_0] [X(t-m) - \bar{X}_m] + \\ &+ \sum_{\ell=1}^M \sum_{m=1}^M a(m) a(\ell) \frac{1}{N} \sum_{t=1}^N [X(t-m) - \bar{X}_m] [X(t-\ell) - \bar{X}_\ell] = \\ &= \frac{1}{N} \sum_{t=1}^N \left\{ [X(t) - \bar{X}_0]^2 - 2 \sum_{m=1}^M a(m) [X(t) - \bar{X}_0] [X(t-m) - \bar{X}_m] + \right. \\ &+ \left. \sum_{\ell=1}^M \sum_{m=1}^M a(m) a(\ell) [X(t-m) - \bar{X}_m] [X(t-\ell) - \bar{X}_\ell] \right\} = \\ &= \frac{1}{N} \sum_{t=1}^N \left\{ [X(t) - \bar{X}_0] - \sum_{m=1}^M a(m) [X(t-m) - \bar{X}_m] \right\}^2 = \\ &= \frac{1}{N} \sum_{t=1}^N \left\{ X(t) - \sum a(m) X(t-m) + \sum a(m) \bar{X}_m - \bar{X}_0 \right\}^2 = \end{aligned}$$

$$\begin{aligned}
 \bar{X}_0 - \sum a(m) \bar{X}_m - a(0) &= \frac{1}{N} \sum_{t=1}^N X(t) - \sum_{m=1}^M a(m) \frac{1}{N} \sum_{t=1}^N X(t-m) - a(0) = \\
 &= \frac{1}{N} \sum_{t=1}^N X(t) - \frac{1}{N} \sum_{t=1}^N \sum_{m=1}^M a(m) X(t-m) - a(0) = \\
 &= \frac{1}{N} \sum_{t=1}^N \left[X(t) - \sum_{m=1}^M a(m) X(t-m) - a(0) \right] = \\
 &= \frac{1}{N} \sum_{t=1}^N \varepsilon(t) = \bar{\varepsilon},
 \end{aligned}$$

logo,

$$H(M) = \frac{1}{N} \sum_{t=1}^N [\varepsilon(t) - \bar{\varepsilon}]^2.$$

c.q.d.

Também obtemos

$$E[H(M)] = (1 - N^{-1})\sigma^2. \quad (4.30)$$

De fato, temos

$$\begin{aligned}
 H(M) &= \frac{1}{N} \sum_{t=1}^N [\varepsilon(t) - \bar{\varepsilon}]^2 = \\
 &= \frac{1}{N} \left[\sum_{t=1}^N \varepsilon^2(t) - N(\bar{\varepsilon})^2 \right]
 \end{aligned}$$

logo,

$$E[H(M)] = E \left\{ \frac{1}{N} \left[\sum_{t=1}^N \varepsilon^2(t) - N(\bar{\varepsilon})^2 \right] \right\} =$$

$$= \frac{1}{N} \sum_{t=1}^N E \epsilon^2(t) - E[(\bar{\epsilon})^2] = \frac{1}{N} N \sigma^2 - E[(\bar{\epsilon})^2] = \sigma^2 - E[(\bar{\epsilon})^2]$$

mas,

$$E[(\bar{\epsilon})^2] = E \left[\frac{1}{N} \sum_{t=1}^N \epsilon(t) \right]^2 \underset{\text{indep}}{=} \frac{1}{N^2} \sum_{t=1}^N E[\epsilon^2(t)] = \frac{1}{N^2} N \sigma^2 = \frac{\sigma^2}{N}$$

então

$$E[H(M)] = \sigma^2 - \frac{\sigma^2}{N} = \left(1 - \frac{1}{N}\right) \sigma^2 = \left(1 - N^{-1}\right) \sigma^2.$$

c.q.d.

Do Teorema 1, nós sabemos que quando assumimos o modelo (4.17), a distribuição limite de

$$Q(M) = N \sum_{\ell=1}^M \sum_{m=1}^M \Delta a_M(m) \Delta a_M(\ell) C_{XX}(m, \ell) \quad (4.31)$$

tem esperança $M\sigma^2$, isto é,

$$E_{\infty}[Q(M)] = M\sigma^2, \quad (4.32)$$

onde E_{∞} denota a esperança da distribuição limite da quantidade entre colchetes, quando N tende ao infinito.

De fato, da ergodicidade do processo $X(t)$ sabemos que $C_{XX}(\ell, m)$ converge para $\gamma_{XX}(\ell, m)$ quando N tende ao infinito e como já demonstramos que $E_{\infty} \left[N \sum_{\ell} \sum_m \Delta a_M(m) \Delta a_M(\ell) \gamma_{XX}(m, \ell) \right] = M\sigma^2$, temos que

$$E_{\infty}[Q(M)] = M\sigma^2.$$

Logo, das relações (4.28), (4.30), (4.31) e (4.32), temos

$$E_{\infty}[S^2(M)] = \left(1 - \frac{1}{N} - \frac{M}{N}\right)\sigma^2 = \left(1 - \frac{M+1}{N}\right)\sigma^2 = [1 - N^{-1}(M+1)]\sigma^2.$$

Essas observações sugerem que seria razoável adotar

$$[1 - N^{-1}(M+1)]^{-1}S^2(M)$$

como um estimador de σ^2 , e tendo em vista a relação (4.26), definir uma estimativa (FPE)(M) de $(FPE)_M$ por

$$(FPE)(M) = [1 + N^{-1}(M+1)][1 - N^{-1}(M+1)]^{-1}S^2(M). \quad (4.33)$$

Daquilo que foi exposto até aqui concluimos naturalmente que quando há muitos previsores obtidos aplicando o método dos mínimos quadrados, seria razoável escolhermos aquele com o mínimo valor de (FPE)(M). Seguindo esse critério, para a identificação do previsor (isto é, para a determinação da ordem do processo) através de um único registro de $X(t)$, nós procedemos como segue; calculamos (FPE)(M) sucessivamente para $M=0,1,2,\dots,L$ (L suposto inteiro positivo) e adotamos o valor $M = M_0$ como a ordem do processo, onde $(FPE)(M_0) =$ ao mínimo de $(FPE)(M)$ ($M=0,1,2,\dots,L$).

Esse processo será chamado procedimento do mínimo FPE.

4.3.4 - Propriedades Estatísticas de (FPE)(M)

Para percebermos a utilidade prática do procedimento do mínimo FPE, nós temos que analisar primeiramente as características estatísticas de (FPE)(M) ($M=0,1,2,\dots,L$) para um mo-

delo fixado para $X(t)$.

Nós vamos supor que a ordem de $X(t)$ é K ($K \geq 0$), isto é, que $a_K \neq 0$ e $a_m = 0$ para $m > K$ na relação (4.17). Suporemos também que o conjunto de dados é dado na forma $\{X(t), t = -L+1, -L+2, \dots, 1, 2, \dots, N\}$. Da relação (4.29) nós podemos verificar que $H(M)$ permanece constante para $M \geq K$ (pois $a_m = 0$, para $m > K$) e portanto o comportamento de $S^2(M)$ depende apenas de $Q(M)$ dado pela relação (4.31). Da discussão da secção 2.2 nós sabemos que a distribuição limite de $Q(M)$ ($M=K, K+1, \dots, L$) é idêntica àquela de $NC'_{\epsilon XM} \Gamma_{MM}^{-1} C_{\epsilon XM}$, onde $C_{\epsilon XM} = [C_{\epsilon X}(\ell)]$, ($\ell=1, 2, \dots, M$).

De fato, da ergodicidade do processo $X(t)$ sabemos que $C_{XX}(\ell-m)$ converge para $\gamma_{XX}(\ell-m)$ quando $N \rightarrow \infty$ com probabilidade 1 e da consistência de $C_{XX}(m-\ell)$ sabemos que a distribuição limite de $\sqrt{N}\Delta a_M$ é idêntica àquela de $\sqrt{N}\Gamma_{MM}^{-1}C_{\epsilon X}$, logo

$$\begin{aligned} Q(M) &= N \sum_{\ell=1}^M \sum_{m=1}^M \Delta a_M(m) \Delta a_M(\ell) C_{XX}(m, \ell) = \\ &= \sum_{\ell=1}^M \sqrt{N} \Delta a_M(\ell) \sum_{m=1}^M \sqrt{N} \Delta a_M(m) C_{XX}(m, \ell) = \\ &= [\sqrt{N}(\Delta a_m)'] [C_{XX}(m, \ell)] [\sqrt{N} \Delta a_M] = \end{aligned}$$

$$= [\sqrt{N} \Delta a_M(1) \dots \sqrt{N} \Delta a_M(M)] \begin{bmatrix} C(1,1) & C(2,1) & \dots & C(M,1) \\ C(1,2) & C(2,2) & \dots & C(M,2) \\ \vdots & \vdots & & \vdots \\ C(1,M) & C(2,M) & \dots & C(M,M) \end{bmatrix} \begin{bmatrix} \sqrt{N} \Delta a_M(1) \\ \vdots \\ \sqrt{N} \Delta a_M(M) \end{bmatrix}$$

tem distribuição limite

$$Q(M) \sim \sqrt{N} C'_{\epsilon XM} \Gamma_{MM}^{-1} \Gamma_{MM} \sqrt{N} \Gamma_{MM}^{-1} C_{\epsilon XM} = N C'_{\epsilon XM} \Gamma_{MM}^{-1} C_{\epsilon XM},$$

onde $C_{\epsilon XM} = C_{\epsilon X} = [C_{\epsilon X}(\ell)]$, $(\ell=1,2,\dots,M)$.

Como foi visto no Lema 1 da secção 4.3.2 a matriz de covariância da distribuição limite de $\sqrt{N} C_{\epsilon XM}$ é idêntica àquela de $\{X(t-m) - EX(t-m); m=1,2,\dots,M\}$ multiplicada por σ^2 , isto é, $\sigma^2 \Gamma_{MM}$. Então o procedimento de sucessivas ortogonalizações de $X(t-m) - EX(t-m)$ ($m=1,2,\dots,M$), descrito na secção 3 do capítulo anterior, pode ser aplicado a $\sqrt{N} C_{\epsilon XM}$ para se obter uma variável aleatória vetorial

$$U_M = \sigma^{-1} \sqrt{N} D_M B_M C_{\epsilon XM}$$

cuja distribuição limite é uma distribuição normal M -dimensional unitária. Definindo

$$T_M = \sigma^{-1} D_M B_M$$

temos que

$$U_M = T_M \sqrt{N} C_{\epsilon XM}$$

com

$$\sigma^2 T_M \Gamma_{MM} T_M' = I_M$$

onde I denota a matriz identidade e a matriz T_M da transformação é uma matriz com zeros acima da diagonal.

De fato, temos

$$\sigma^2 T_M \Gamma_{MM} T_M' = \sigma^2 \sigma^{-1} D_M B_M \Gamma_{MM} \sigma^{-1} B_M' D_M = D_M B_M \Gamma_{MM} B_M' D_M = I_M$$

conforme foi demonstrado na pg.56 do capítulo anterior.

Da estrutura de T_M podemos ver facilmente que

$$T_M(\ell, m) = T_L(\ell, m) \quad (\ell, m=1, 2, \dots, M) \text{ para } M < L$$

isto é, a submatriz formada pelos primeiros $M \times M$ são elementos de T_L é idêntica a T_M .

Então nós temos que U_M é o vetor dos primeiros M elementos de U_L . Dessa observação nós podemos concluir que a distribuição limite de $Q(M)$ é idêntica àquela de

$$\sigma^2 \sum_{\ell=1}^M U^2(\ell), \quad (M=K, K+1, \dots, L).$$

De fato, temos

$$\begin{aligned} \sigma^2 \sum_{\ell=1}^M U^2(\ell) &= \sigma^2 U_M' U_M = \\ &= \sigma^2 \sigma^{-1} \sqrt{N} C'_{\epsilon x M} B_M' D_M \sigma^{-1} \sqrt{N} D_M B_M C_{\epsilon x M} = \\ &= N C'_{\epsilon x M} B_M' D_M D_M B_M C_{\epsilon x M} \end{aligned}$$

mas da relação (3.47) temos $B_M' D_M D_M B_M = \Gamma_{MM}^{-1}$, logo,

$$\sigma^2 \sum_{\ell=1}^M U^2(\ell) = N C'_{\epsilon x M} \Gamma_{MM}^{-1} C_{\epsilon x M} \sim Q_M$$

conforme foi visto na pg.94 .

A distribuição limite de $U^2(\ell)$, $(\ell=1,2,\dots,L)$ é então a distribuição de variáveis qui-quadrado mutuamente independentes cada uma com 1 grau de liberdade. Então, nós obtemos:

TEOREMA 2 - Para $M \geq K$, $\sigma^{-2}Q(M)$ é assintoticamente distribuído como a soma parcial dos M primeiros termos da seqüência de variáveis com distribuição qui-quadrado, mutuamente independentes, cada uma com 1 grau de liberdade.

Agora nós faremos a análise do comportamento estatístico de $(FPE)(M)$ $(M=0,1,\dots,L)$, conforme pode ser visto em Akaike (1970-a). Para qualquer inteiro positivo M , nós definimos $a_M(m)$ independente da ordem K , como a solução de (4.18) quando $C_{XX}(\ell-m)$ é substituído por $\gamma_{XX}(\ell-m)$ e, levando em conta a relação (4.27), definimos $\sigma^2(M)$ por

$$\sigma^2(M) = \gamma_{XX}(0) - \sum_{m=1}^M a_M(m) \gamma_{XX}(m). \quad (4.34)$$

Nós denotamos $\Delta a_M(m) = \hat{a}_M(m) - a_M(m)$, onde $\hat{a}_M(m)$ é a solução de (4.18). M não está restrito a ser maior ou igual do que a ordem K . Da relação (4.28) temos que

$$\begin{aligned} S^2(M) &= C_{XX}(0,0) - \sum_{\ell=1}^M \hat{a}_M(\ell) C_{XX}(0,\ell) = \\ &= C_{XX}(0,0) - 2 \sum_{m=1}^M a_M(m) C_{XX}(0,m) + \sum_{\ell=1}^M \sum_{m=1}^M a_M(\ell) a_M(m) C_{XX}(m,\ell) - \\ &= \sum_{\ell=1}^M \sum_{m=1}^M \Delta a_M(\ell) \Delta a_M(m) C_{XX}(m,\ell). \end{aligned} \quad (4.35)$$

Ignorando os termos de ordem N^{-2} , nós temos aproximadamente

$$\begin{aligned} (\text{FPE})(M_1) - (\text{FPE})(M_2) &= [1 - N^{-1}(M_1+1)]^{-1} [1 - N^{-1}(M_2+1)]^{-1} \cdot \\ &\cdot [S^2(M_1) - S^2(M_2) - N^{-1}(M_2 - M_1)(S^2(M_1) + S^2(M_2))] \end{aligned} \quad (4.36)$$

onde $0 \leq M_1, M_2 \leq L$.

Se nós assumirmos que a igualdade (4.36) vale exatamente, teremos para $M_1 < M_2$

$$\begin{aligned} P\left\{(\text{FPE})(M_1) - (\text{FPE})(M_2) > 0\right\} &\geq P\left\{S^2(M_1) - S^2(M_2) - 2N^{-1}(M_2 - M_1)S^2(M_1) > 0\right\} \geq \\ &\geq P\left\{[S^2(M_1) - S^2(M_2)] [(M_2 - M_1)C_{xx}(0)]^{-1} > 2N^{-1}\right\} \end{aligned} \quad (4.37)$$

Da relação (4.35) nós temos

$$\begin{aligned} S^2(M_1) - S^2(M_2) &= \sigma^2(M_1) - \sigma^2(M_2) + \Delta[\sigma^2(M_1) - \sigma^2(M_2)] + \\ &+ N^{-1}[Q(M_2) - Q(M_1)] \end{aligned} \quad (4.38)$$

onde $\Delta[\sigma^2(M_1) - \sigma^2(M_2)]$ é obtido substituindo $\gamma_{xx}(m, \ell)$ por

$$\gamma_{XX}^{-1}(0)[\sigma^2(M_1) - \sigma^2(M_2)]$$

será da ordem de $N^{-\frac{1}{2}}$, tendo em vista a relação (4.34) e o fato que $\gamma_{XX}(m) \leq \gamma_{XX}(0)$. Enquanto

$$C_{XX}^{-1}(0,0)\Delta[\sigma^2(M_1) - \sigma^2(M_2)] \quad \text{e} \quad C_{XX}^{-1}(0,0)[Q(M_2) - Q(M_1)]N^{-1}$$

são estocasticamente da ordem de N^{-1} . Então a probabilidade da relação (4.37) será, nesse caso, muito próxima de 1.

Em geral nós temos para $M < K$,

$$\lim_{N \rightarrow \infty} P\left\{ (FPE)(M) - (FPE)(K) > 0 \right\} = 1 \quad (4.39)$$

Pela relação (4.36) e o fato que $S^2(M)$ é uma estimativa consistente de σ^2 , nós podemos verificar que, para $M \geq K$, a distribuição limite de $N[(FPE)(K) - (FPE)(M)]$ ($M=K+1, K+2, \dots, L$) é idêntica aquela de $N[S^2(K) - S^2(M)] - 2\sigma^2(M-K)$.

Como temos, para $M \geq K$, que $S^2(K) - S^2(M) = N^{-1}[Q(M) - Q(K)]$, nós podemos verificar do Teorema 2 que a distribuição limite de $N\sigma^{-2}[(FPE)(K) - (FPE)(M)] + 2(M-K)$ é idêntica à distribuição da soma de $(M-K)$ variáveis aleatórias qui-quadrado $\chi_1^2(i)$ ($i=1, 2, \dots, M-K$), mutuamente independentes, cada uma com 1 grau de liberdade.

Então para $M \geq K$, nós temos

$$\lim_{N \rightarrow \infty} P\left\{ (FPE)(K) > (FPE)(M) \right\} = P\left\{ \sum_{i=1}^{M-K} \chi_1^2(i) > 2(M-K) \right\}. \quad (4.40)$$

Dá relação (4.39), nós verificamos que usando $(FPE)(M)$ pa-

ra o procedimento de mínimo FPE, a probabilidade de adotar M menor do que K , como por exemplo $M = M_0$, poderá ser feita arbitrariamente pequena quando N é aumentado indefinidamente, enquanto a relação (4.40) mostra que para $M > K$ a probabilidade de observar $(FPE)(M)$ menor do que $(FPE)(K)$ tende a uma constante não nula. Essa última observação mostra que o valor M_0 adotado para M pelo procedimento de mínimo FPE como a ordem do previsor, isto é como a ordem do processo auto-regressivo a ser ajustado à série, não é uma estimativa consistente de K .

Isso não significa necessariamente uma séria desvantagem do procedimento para aplicações práticas. A própria probabilidade, como sugerida em (4.40), de adotar M_0 maior do que K não é intolerável para aplicações práticas. Além disso será mais comum nós encontrarmos com a situação onde o valor teórico de K pode ser considerado infinito. Para esse caso o resultado da discussão precedente feita em seguida da relação (4.38) sugere que a probabilidade de M_0 ser igual a um valor M para o qual $|a_M(m) - a(m)|$ seja maior do que $N^{-\frac{1}{2}}$ e o correspondente $\sigma^2(M)$ difira de σ^2 por uma quantidade maior que $N^{-\frac{1}{2}}\gamma_{xx}(0)$ será muito pequena. A relação (4.38) também sugere que neste caso a relação (4.40) valerá aproximadamente quando a diferença de $\sigma^2(K)$ a σ^2 é feita significativamente menor do que N^{-1} .

O procedimento o qual é obtido substituindo a definição de $(FPE)(M)$ no procedimento do mínimo FPE por

$$(FPE)^\alpha(M) = [1 + N^{-\alpha}(M+1)][1 - N^{-1}(M+1)]^{-1} S^2(M),$$

onde $0 < \alpha < 1$, será chamado procedimento do mínimo $(FPE)^\alpha$. Através desta modificação nós obteremos a consistência de M_0 como uma estimativa da ordem K do processo auto-regressivo finito. Mas a modificação pode aumentar muito a tendência de M_0 , tomando valores muito pequenos para a minimização de FPE.

4.4 - O MÉTODO PARZEN

4.4.1 - Definição do CAT de um Previsor

No método de Akaike supõem-se que exista uma verdadeira ordem M^* tal que a série de tempo é $AR(M^*)$. O problema consiste então em estimar essa "verdadeira" ordem do modelo auto-regressivo que se ajusta aos dados.

O método de Parzen para modelar uma série de tempo estacionária, descrito em Parzen (1974), é baseado no conceito que o problema é estimar a função de transferência auto-regressiva infinita $A_\infty(e^{i\lambda})$, $-\pi \leq \lambda \leq \pi$. A estimação é feita através de um critério chamado CAT (critério de função de transferência auto-regressiva). CAT é suficiente para a escolha da ordem do modelo aproximador auto-regressivo, o qual é "ótimo" no sentido que sua função de transferência é um estimador de erro quadrático médio mínimo (chamado ARTFACT) da função de transferência auto-regressiva infinita (ARTF) do filtro, a qual transforma a série de tempo em um ruído branco.

Dada uma amostra $X(t)$, $t=1,2,\dots,N$ de uma série de tempo

estacionária normal com média zero, a ordem de um ARTFACT (Aproximador da função de transferência auto-regressiva convergindo para a verdadeira), isto é, a ordem M^* do modelo aproximador auto-regressivo "ótimo", é definida como sendo o valor de M que minimiza

$$CAT(M) = \frac{1}{N} \sum_{j=0}^M \frac{1}{\hat{S}^2(j)} - \frac{1}{\hat{S}^2(M)}, \quad (4.41)$$

onde

$$\hat{S}^2(j) = \frac{N}{N-j} S^2(j) \quad (4.42)$$

é um estimador não viciado de $\sigma^2(j)$ o qual é o erro quadrático médio de previsão um passo à frente com memória de comprimento j . Ou, de modo equivalente,

$$CAT(M) = \frac{1}{N} \sum_{j=0}^M \frac{N-j}{NS^2(j)} - \frac{N-M}{NS^2(M)} \quad (4.43)$$

onde $S^2(M)$ é dado pela relação (3.45).

4.4.2 - Base Teórica do Critério

Conforme vimos, o método de Parzen para determinar a ordem do modelo auto-regressivo a ser ajustado a uma série de tempo, consiste na estimação do filtro (com função de transferência A_∞), o qual transforma a série de Tempo $X(\cdot)$ num ruído branco $\varepsilon(\cdot)$.

Então, pela relação (2.4), temos

$$A(B)X(t) = \varepsilon(t).$$

Consideremos a expansão em série de potências de $A_\infty(Z)$,

$$A_\infty(Z) = 1 - a_\infty(1)Z - a_\infty(2)Z^2 - \dots - a_\infty(M)Z^M - \dots \quad (4.44)$$

A série de tempo $X(\cdot)$ será um processo auto-regressivo de ordem M , se o erro de previsão de memória infinita $\varepsilon(t)$ for de fato uma combinação linear finita de $X(t-1), \dots, X(t-M)$ e o coeficiente de $X(t-M)$ for não nulo; então $a_\infty(j) = a_M(j)$. Nesse caso a função de transferência auto-regressiva de memória M , abreviadamente ARTF(M), será

$$A_M(Z) = 1 - a_M(1)Z - a_M(2)Z^2 - \dots - a_M(M)Z^M, \quad (4.45)$$

e nós podemos representar o erro de previsão de memória M $\varepsilon^{(M)}(t)$, como a saída de um filtro com entrada $X(t)$,

$$\varepsilon^{(M)}(t) = A_M(B)X(t)$$

então

$$\begin{aligned} \varepsilon^{(M)}(t) &= X(t) - a_M(1)X(t-1) - a_M(2)X(t-2) - \dots - \\ &\quad - a_M(M)X(t-M) = A_M(B)X(t). \end{aligned} \quad (4.46)$$

TEOREMA 3 - Suponha $a_M(M)$ não nulo. Então $X(\cdot)$ é AR(M) se, e somente se, $\varepsilon^{(M)}(\cdot)$ for um ruído branco.

A parte "somente se" do teorema segue da definição de AR(M). A parte "se" pode ser provada por indução. (Veja Parzen (1977).

Para justificar e motivar o uso de modelos auto-regressivos com memória finita como modelos aproximativos para séries de tempo estacionários, Parzen usa a estrutura matemática de aproximação no domínio de frequência. De modo mais geral, essa estrutura produz aproximações paramétricas finitas para qualquer função que tenha as propriedades matemáticas das funções densidades espectrais.

Seja $f(\lambda)$ uma função densidade espectral no intervalo $[-\pi, \pi]$; nós definimos um espaço de Hilbert $L_2(f)$ das funções mensuráveis a valores complexos em $[-\pi, \pi]$, com o produto interno

$$((g, h))_f = \int_{-\pi}^{\pi} g(\lambda) \overline{h(\lambda)} f(\lambda) d\lambda, \quad (4.47)$$

e para as quais,

$$\|g\|_f^2 = \int_{-\pi}^{\pi} |g(\lambda)|^2 f(\lambda) d\lambda < \infty \quad (4.48)$$

tal que

$$\|g\|_f^2 = ((g, g))_f,$$

onde $\|g\|_f$ é a norma de g . (Conforme é mostrado em Hannan (1970), pg. 497).

O problema de previsão um passo à frente com memória fi-

nita de comprimento M pode ser colocado como um problema de otimização em $L_2(f)$ como segue. O erro de previsão $\eta(t)$ de um previsor usual de memória finita de comprimento M é da forma $\eta(t) = p(B)X(t)$, onde $p(Z)$ é um polinômio de grau M da forma $p(Z) = 1 - c_1 Z - c_2 Z^2 - \dots - c_M Z^M$.

A variância do erro de previsão pode ser expressa como a norma de p .

$$\begin{aligned} S^2(\eta) &= E[\eta^2(t)] = ((p(e^{-i\lambda}), p(e^{-i\lambda})))_f = \\ &= \int_{-\pi}^{\pi} |p(e^{-i\lambda})|^2 f(\lambda) d\lambda \end{aligned} \quad (4.49)$$

Obter um erro de previsão mínimo é equivalente a encontrar o polinômio $A_{M^*}(\cdot)$ de grau M^* como norma mínima, no sentido que a $ARTF(M^*)A_{M^*}(\cdot)$ satisfaz

$$S^2(M^*) = ((A_{M^*}(e^{-i\lambda}), A_{M^*}(e^{-i\lambda})))_f < ((p(e^{-i\lambda}), p(e^{-i\lambda})))_f, \quad (4.50)$$

para todo polinômio $p(e^{-i\lambda})$.

Para obter um teorema sobre as propriedades da $ARTF(M)\hat{A}_M(e^{-i\lambda})$ amostral, nós precisamos modificar nossas definições dessas funções de transferência como segue:

$$\begin{aligned} g_{\infty}(e^{-i\lambda}) &= \frac{A_{\infty}(e^{-i\lambda})}{\sigma^2(\infty)} \\ g_M(e^{-i\lambda}) &= \frac{A_M(e^{-i\lambda})}{\sigma^2(M)} \end{aligned}$$

$$\hat{g}_M(e^{-i\lambda}) = \frac{\hat{A}_M(e^{-i\lambda})}{S^2(M)} \quad (4.51)$$

onde

$$\sigma^2(p) = \gamma_0 - \sum_{j=1}^p a_j \gamma_j, \quad p=1,2,\dots \quad (4.52)$$

Como uma medida do erro quadrático médio total de estimação, nós temos

$$\begin{aligned} J_M &= E \|\hat{g}_M(e^{-i\lambda}) - g_\infty(e^{-i\lambda})\|_f^2 = \\ &= E \int_{-\pi}^{\pi} |\hat{g}_M(e^{-i\lambda}) - g_\infty(e^{-i\lambda})|^2 f(\lambda) d\lambda. \end{aligned} \quad (4.53)$$

Usando relações recursivas da teoria dos filtros (veja Kailath, 1974), o seguinte teorema básico pode ser provado.

TEOREMA 4 - Para N grande, temos aproximadamente

$$J_M = \frac{1}{N} \sum_{j=1}^M \frac{1}{\sigma^2(j)} + \frac{1}{\sigma^2(\infty)} - \frac{1}{\sigma^2(M)} \quad (4.54)$$

Esta expressão para o erro quadrático médio é a soma de dois termos, um representando o vício:

$$((g_M - g_\infty, g_M - g_\infty))_f = \frac{1}{\sigma^2(\infty)} - \frac{1}{\sigma^2(M)} \quad (4.55)$$

e outro representando a variabilidade do estimador $\hat{g}_M$:

$$E \|\hat{g}_M - g_M\|_f^2 = \frac{1}{N} \sum_{j=1}^M \frac{1}{\sigma^2(j)}. \quad (4.56)$$

Como $\frac{1}{\sigma^2(\infty)}$ não depende de M , para encontrar a ordem M^* é suficiente minimizar

$$J'_M = J_M - \frac{1}{\sigma^2(\infty)}.$$

A motivação e a justificação para o CAT(M) é que ele é um estimador, a partir de uma amostra de tamanho N , de J'_M . O procedimento de otimalidade desta estimação continua sendo investigado.

CAPÍTULO 5

DETERMINAÇÃO DA ORDEM DE UM MODELO ARMA A SER AJUSTADO À SÉRIE DE TEMPO

5.1 - INTRODUÇÃO

Com o ajuste de um modelo auto-regressivo muitas vezes não se consegue obter uma representação paramétrica "parcimoniosa" (com poucos parâmetros) de um processo estocástico estacionário, corresponde a uma série de tempo observada. Nesses casos tal representação paramétrica parcimoniosa pode ser conseguida através do ajuste dos modelos mistos auto-regressivo-médias-móveis (ARMA), que definiremos na secção 3 e que são mais gerais que os modelos (AR).

Observemos ainda que quando se dispõe de uma pequena quantidade de dados, caso em que a obtenção de um modelo com poucos parâmetros se faz necessária, o uso dos modelos mistos ARMA adquire grande importância.

No entanto, o problema crucial continua sendo a determinação da ordem do modelo ARMA a ser ajustado à série de tempo observada. Para resolver esse problema apresentaremos dois métodos, o método de Akaike (1974) denominado AIC (critério de informação de Akaike) e o método de Gray et alii (1978).

5.2 - PROCESSO DE MÉDIAS MÓVEIS (MA)

Consideremos um ruído branco, isto é, uma sequência de variáveis aleatórias não correlacionadas $\epsilon(t)$, $t=0, \pm 1, \pm 2, \dots$ todas com média $E[\epsilon(t)] = 0$ e variância $E[\epsilon^2(t)] = \sigma^2$. Definimos então, um processo de médias móveis de ordem q pela expressão

$$X(t) = \epsilon(t) - b_1\epsilon(t-1) - b_2\epsilon(t-2) - \dots - b_q\epsilon(t-q), \quad (5.1)$$

$t = 0, \pm 1, \pm 2, \dots$, onde os b_j 's são constantes reais.

Analogamente, um processo é dito de médias móveis de ordem infinita se

$$X(t) = \epsilon(t) - \sum_{j=1}^{\infty} b_j \epsilon(t-j) \quad (5.2)$$

onde $\sum_{j=1}^{\infty} b_j^2 < \infty$.

Da teoria de filtros lineares, um processo de médias móveis de ordem q pode ser pensado como a saída $X(t)$ de um filtro linear, com função de transferência

$$B(\lambda) = 1 - \sum_{j=1}^q b_j e^{-i\lambda j} \quad (5.3)$$

e entrada um ruído branco $\epsilon(t)$.

Como o espectro de um ruído branco é contínuo com função densidade espectral $f_{\epsilon}(\lambda) = \sigma^2/2\pi$, a relação (5.1) define um processo fracamente estacionário com espectro contínuo e função densidade espectral

$$f_x(\lambda) = |B(\lambda)|^2 f_\varepsilon(\lambda) = \frac{\sigma^2}{2\pi} \left| 1 - \sum_{j=1}^q b_j e^{-i\lambda j} \right|^2 \quad (5.4)$$

O processo (5.1) tem média zero e variância $\sigma^2 \sum_{j=0}^q b_j^2$, onde $b_0 = 1$.

Um processo de médias móveis tem a seguinte interpretação intuitiva. No instante t , a "natureza" produz um "choque" aleatório ou inovação $\varepsilon(t)$, o qual é não correlacionado com os choques em outros instantes. A quantidade $X(t)$, observada no instante t , depende então das inovações do passado $\varepsilon(t-j)$ para $j=1,2,\dots,q$ e do presente $\varepsilon(t)$, sendo uma média dessas inovações.

5.3 - PROCESSOS MISTOS AUTO-REGRESSIVOS-MÉDIAS-MÓVEIS (ARMA)

Uma generalização natural dos modelos auto-regressivos é considerar como entrada do filtro linear correspondente à equação de diferenças (2.1), não apenas um único ruído branco, mas sim um processo de médias móveis.

Obteremos então, o processo misto auto-regressivo-médias-móveis de ordem (p,q)

$$X(t) = a_1 X(t-1) + a_2 X(t-2) + \dots + a_p X(t-p) + \varepsilon(t) - b_1 \varepsilon(t-1) - \dots - b_q \varepsilon(t-q)$$

ou

$$\phi(B)X(t) = \theta(B)\varepsilon(t) \quad (5.5)$$

onde

$$\phi(B) = 1 - a_1 B - a_2 B^2 - \dots - a_p B^p$$

$$\theta(B) = 1 - b_1 B - b_2 B^2 - \dots - b_q B^q$$

o qual indicamos abreviadamente por ARMA(p,q).

Se $\theta(B)$ não tiver zeros dentro do círculo unitário e $\phi(B)$ não tiver zeros dentro ou sobre o círculo unitário, a relação (5.5) pode ser escrita na forma

$$X(t) = \frac{\theta(B)}{\phi(B)} \varepsilon(t) = \left(\frac{1 - \sum_{j=1}^q b_j B^j}{1 - \sum_{k=1}^p a_k B^k} \right) \varepsilon(t) \quad (5.6)$$

e o processo ARMA(p,q) pode ser pensado como a saída $X(t)$ de um filtro linear, com função de transferência

$$C(\lambda) = \frac{1 - \sum_{j=1}^q b_j e^{-i\lambda j}}{1 - \sum_{k=1}^p a_k e^{-i\lambda k}} \quad (5.7)$$

e entrada um ruído branco.

Logo a relação (5.5) define um processo fracamente estacionário com espectro contínuo e função densidade espectral

$$f_X(\lambda) = |C(\lambda)|^2 f_\varepsilon(\lambda) = \frac{\sigma^2}{2\pi} \left| \frac{1 - \sum_{j=1}^q b_j e^{-i\lambda j}}{1 - \sum_{k=1}^p a_k e^{-i\lambda k}} \right|^2. \quad (5.8)$$

5.4 - O MÉTODO DE AKAIKE

5.4.1 - Definição do Critério de Informação de Akaike (AIC)

O procedimento desenvolvido por Akaike (1972) é baseado no Critério de Informação definido por

$$\begin{aligned} \text{AIC} = & (-2) \log(\text{máxima-verossimilhança}) \\ & + 2 (\text{número de parâmetros independentes do mo-} \\ & \text{delo ajustado}), \end{aligned} \quad (5.9)$$

onde $\log$ indica o logaritmo natural.

Esse critério é uma estimativa de menos duas vezes o logaritmo da verossimilhança esperada. O oposto do logaritmo da verossimilhança esperada é, exceto por uma constante aditiva, idêntico à entropia (generalizada) da verdadeira estrutura estocástica em relação à estrutura assumida pelo modelo e provém uma medida natural do desvio da estrutura assumida em relação à verdadeira estrutura.

AIC mede o ajuste e a inexatidão do modelo. O ajuste é medido pela 1ª parte da definição, a qual decresce quando o número de parâmetros do modelo é aumentado, e a inexatidão é medida pelo último termo, o qual cresce com o número de parâmetros. O valor total de AIC decresce rapidamente conforme o número de parâmetros do modelo ajustado vai sendo aumentado e então, cresce quase linearmente quando um número excessivo de parâmetros é incluído no modelo. O modelo correspondente ao con-

junto de parâmetros que produz o mínimo de AIC, é escolhido como aquele que melhor representa a verdadeira estrutura do processo estocástico gerador da série observada.

5.4.2 - A Média do log-verossimilhança como uma Medida do Ajuste

O fato bem conhecido que o estimador de máxima verossimilhança (EMV) é, sob condições de regularidade, assintoticamente eficiente (Cramer, 1946), mostra que a função de verossimilhança tende a ser um critério muito sensível ao desvio dos parâmetros do modelo dos verdadeiros valores.

Consideremos a situação onde $x_1, x_2, \dots, x_N$ são obtidos como o resultado de N observações independentes de uma variável aleatória com função densidade de probabilidade $g(x)$. Se uma família paramétrica de funções densidade é dada por $f(x|\theta)$ com um vetor de parâmetros θ , o log-verossimilhança média, ou log-verossimilhança dividido por N , é dado por

$$\frac{1}{N} \sum_{i=1}^N \log f(x_i|\theta) \quad (5.9a)$$

Quando N tende ao infinito, esta média tende, com probabilidade 1, a

$$S(g; f(\cdot|\theta)) = \int g(x) \log f(x|\theta) dx \quad (5.10)$$

onde é suposta a existência da integral.

Da eficiência do EMV pode ser verificado que o log-verossimilhança esperado $S(g; f(\cdot|\theta))$ é um critério muito sensível a

pequenos desvios de $f(x|\theta)$ de $g(x)$.

A diferença

$$I(g;f(\cdot|\theta)) = S(g;g) - S(g;f(\cdot|\theta)) \quad (5.11)$$

é conhecida como a *informação média de Kullback-Leibler* para a discriminação entre $g(x)$ e $f(x)$ e assume valores positivos, a menos que $f(x|\theta) = g(x)$ em quase toda parte.

Essas observações mostram que $S(g;f(\cdot|\theta))$ será um critério razoável para definir o modelo que melhor se ajusta aos dados pela sua maximização ou, da analogia ao conceito de entropia, pela minimização de $-S(g;f(\cdot|\theta))$.

Uma das características mais importantes de $S(g;f(\cdot|\theta))$ é que sua estimativa natural, o log-verossimilhança média (5.9a), pode ser obtida sem o conhecimento de $g(x)$. Observamos ainda que, quando somente uma família $f(x|\theta)$ é considerada, maximizando a estimativa (5.9a) de $S(g;f(\cdot|\theta))$ com relação a θ , obtemos o EMV $\hat{\theta}$.

5.4.3 - Procedimento MAICE

A estatística AIC é introduzida como uma estimativa da quantidade de informação.

$$\begin{aligned} I(g;f(\cdot|\theta)) &= S(g;g) - S(g;f(\cdot|\theta)) = \\ &= \int g(x) \log[g(x)/f(x|\theta)] dx \end{aligned} \quad (5.12)$$

e tendo em vista o que foi exposto na secção anterior, devemos

adotar o modelo $f(x|\theta)$ que minimiza esta quantidade, e portanto para o qual se obtém a maximização do log-verossimilhança esperada

$$\int g(x) \log f(x|\theta) dx.$$

Apresentaremos a seguir um esboço da derivação do AIC. Suponha que $g(x) = f(x|\theta)$, onde $\theta = (\theta_1, \theta_2, \dots)$ e seja $x_1, \dots, x_N$ uma amostra de N observações independentes de $f(x|\theta)$. Então uma estimativa consistente de

$$(-2) \int f(x|\theta) \log f(x|\theta) dx \quad (5.13)$$

é dada por

$$-2 N^{-1} \sum_{i=1}^N \log f(x_i | {}_k\hat{\theta}), \quad (5.14)$$

onde ${}_k\hat{\theta}$ é a estimativa de máxima verossimilhança de ${}_k\theta = (\theta_1, \theta_2, \dots, \theta_k, 0, \dots, 0)$. Como essa estimativa é viciada, podemos ajustá-la em relação ao vício, adotando a estimativa

$$N^{-1} \left[-2 \sum_{i=1}^N \log f(x_i | {}_k\hat{\theta}) + k \right].$$

Além disso, se nós usamos a estimativa ${}_k\hat{\theta}$ para medir o desvio da distribuição definida por $f(x|{}_k\hat{\theta})$ da verdadeira distribuição $f(x|\theta)$, então $2 \log f(\cdot | {}_k\hat{\theta})$ fornece, em média, valores menores do que $2 \log f(\cdot | \theta)$, sendo a diferença esperada igual a

k/N quando $k\hat{\theta}$ é suficientemente próxima de θ . Então, finalmente

$$N^{-1} \left[-2 \sum_{i=1}^N \log f(x_i | k\hat{\theta}) + 2k \right] \quad (5.15)$$

é introduzida como uma estimativa de (5.13).

Quando o número N de observações é fixado a quantidade

$$(-2) \sum_{i=1}^N \log f(x_i | k\hat{\theta}) + 2k \quad (5.16)$$

ou, de modo equivalente,

$$(-2) \log (\text{máxima-verossimilhança}) + 2k$$

é suficiente para comparar desvios dos modelos ajustados em relação à verdadeira estrutura g . Portanto o procedimento MAICE consiste em adotar o modelo que produz o mínimo valor de AIC, entre um dado conjunto de possíveis modelos, como o modelo que melhor se ajusta aos dados.

Observemos ainda que para o número N de observações fixado, a relação (5.16) mostra que quando a máxima verossimilhança é idêntica para dois modelos, o MAICE é aquele definido com o menor número de parâmetros. Logo, pela sua definição, o MAICE fornece uma formulação matemática do princípio da parcimônia.

Para o ajuste de modelos ARMA, a definição do AIC é justificada pelos resultados de Hannan (1973) sobre a distribui-

ção assintótica do estimador de máxima verossimilhança de modelos ARMA e pelos resultados de Whittle (1962) sobre a distribuição assintótica das estatísticas da razão de verossimilhança de modelos ARMA.

Dado um conjunto de observações $\{X(1), X(2), \dots, X(N)\}$, o log-verossimilhança do modelo (5.5) é dado por

$$\begin{aligned} \log \ell(\phi, \theta, \sigma^2 | X(1), \dots, X(N)) = & -\frac{1}{2} N \log \sigma^2 - \{S(\phi, \theta) | 2\sigma^2\} + \\ & + h(\phi, \theta) - \frac{1}{2} N \log 2\pi \end{aligned} \quad (5.17)$$

onde $S(\phi, \theta) = \sum_{t=-\infty}^N \varepsilon^2(t)$ e $h(\phi, \theta)$ é uma função de ϕ e θ e pode ser negligenciada quando N é grande. Então o procedimento de estimação (aproximada) da função de máxima-verossimilhança é reduzido à procura dos parâmetros que minimizam

$$\hat{\sigma}^2 = \{S(\phi, \theta) | N\}. \quad (5.18)$$

Adotando o método de Box & Jenkins para minimizar $\hat{\sigma}^2$, a definição do AIC de um modelo ARMA é dada por

$$AIC(p, q) = N \log \hat{\sigma}^2 + 2(p+q+2) + N \log 2\pi + N. \quad (5.19)$$

Para o ajuste de um modelo ARIMA (p, d, q) do tipo

$$\phi(B) \nabla^d X(t) = \theta(B) \varepsilon(t),$$

onde

$$\nabla^m X(t) = (1-B)^m X(t), \quad (5.20)$$

no qual tomamos d diferenças, o número de dados diminui de N

para $N - d$, tal que se nós queremos comparar o ajuste de modelos ARIMA com diferenças, nós precisamos normalizar o AIC para compensar o efeito dessa mudança no número de dados. Logo, o AIC normalizado para um modelo ARIMA (p, d, q) é dado por

$$AIC(p, d, q) = N \log \hat{\sigma}^2 + \left\{ N / (N - d) \right\} \cdot 2(p + q + 1 + \delta_{d_0}) + N \log 2\pi + N,$$

onde

$$\delta_{d_0} = \begin{cases} 1, & \text{se } d = 0 \\ 0, & \text{se } d \neq 0 \end{cases} \quad (5.21)$$

Para a comparação de vários modelos, quando N é fixado, a definição de AIC pode ser substituída por

$$N \log \hat{\sigma}^2 + \left\{ N / (N - d) \right\} \cdot 2(p + q + 1 + \delta_{d_0}). \quad (5.22)$$

Observemos ainda, que o procedimento MAICE pode ser usado para o ajuste de um modelo auto-regressivo. Em particular, quando o processo é AR de ordem M com erros Gaussianos, a função de verossimilhança é dada por

$$L(\sigma^2) = \frac{1}{(2\pi\sigma^2)^{N/2}} \exp \left[-\frac{1}{2\sigma^2} \sum_{t=1}^N \varepsilon^2(t) \right]$$

e o estimador de máxima verossimilhança de σ^2 é

$$S^2(M) = \frac{1}{N} \sum_{t=1}^N \varepsilon^2(t)$$

então

$$AIC = -2 \log (\text{máxima verossimilhança}) + 2(M+1) =$$

$$= -2 \left[\log \frac{1}{(2\pi S^2(M))^{N/2}} - \frac{N}{2} \right] + 2(M+1) =$$

$$= N \log 2\pi S^2(M) + N + 2(M+1) =$$

$$= N \log S^2(M) + 2(M+1) + N \log 2\pi + N$$

e, quando N é fixado, podemos para efeito de comparação usar apenas

$$AIC(M) = N \log S^2(M) + 2(M+1). \quad (5.23)$$

Por outro lado, a estatística FPE para o ajuste de um modelo AR é dada pela relação (4.33) por

$$(FPE)(M) = \frac{1 + (M+1)/N}{1 - (M+1)/N} S^2(M)$$

e como, para $N \gg M$,

$$\log \left| \frac{1 + (M+1)/N}{1 - (M+1)/N} \right| \approx 2 \left(\frac{M+1}{N} \right)$$

temos

$$\log (FPE)(M) = \log S^2(M) \left| \frac{1 + (M+1)/N}{1 - (M+1)/N} \right| \approx \log S^2(M) + 2 \left(\frac{M+1}{N} \right).$$

Ou seja, pela relação (5.23), temos

$N \log (\text{FPE})(M) \longrightarrow \text{AIC}(M)$, quando $N \longrightarrow \infty$.

Portanto, verificamos que, quando usado para o ajuste de um modelo AR, o procedimento MAICE é assintoticamente equivalente ao procedimento de mínimo FPE.

O procedimento MAICE não tem propriedades ótimas claramente definidas. Especialmente quando é usado para a determinação da ordem de um modelo AR, o procedimento MAICE não produz uma estimativa consistente. No entanto, apesar de sua inconsistência, a estimativa correspondente do espectro é consistente (veja Shibata, 1976).

Uma modificação do AIC para o ajuste de modelos AR ou ARMA é dada por

$$\text{BIC} = (N-K) \log \left[\frac{S(K)}{N-K} \right] + K \log \left[\frac{S(0)-S(K)}{K} \right], \quad (5.24)$$

onde N é o número de observações, K é o número de coeficiente do modelo e $S(K)$ indica N vezes a estimativa do log (máxima-verossimilhança), sob a suposição Gaussiana, da variância do ruído $\epsilon(t)$, a qual é dada pela relação (3.40). Minimizando o BIC, nós obtemos estimativas consistentes das ordens.

Quando o procedimento BIC é aplicado a uma situação onde de fato seria necessário um modelo de ordem infinita, o procedimento tende a produzir uma estimativa muito suavizada do espectro, quando comparado com o procedimento de mínimo AIC. Isso decorre do fato que o procedimento de mínimo AIC produz

um limite superior da ordem do modelo a ser ajustado aos dados, enquanto o procedimento de mínimo BIC produz resultados mais consistentes em aplicações repetidas.

5.5 - O MÉTODO DE GRAY, KELLEY E McINTIRE

Gray, Kelley e McIntire (1978) sugeriram o uso das quantidades R e S , definidas abaixo, para determinar p e q em um processo ARMA(p, q).

5.5.1 - Definições e Teoremas

DEFINIÇÃO 1 - Seja m um inteiro, $h > 0$ e f uma função a valores reais. Seja também $f_m = f(mh)$,

$$H_n[f_m] = \begin{vmatrix} f_m & f_{m+1} & \dots & f_{m+n-1} \\ f_{m+1} & f_{m+2} & \dots & f_{m+n} \\ \vdots & \vdots & & \vdots \\ f_{m+n-1} & f_{m+n} & \dots & f_{m+2n-2} \end{vmatrix},$$

$$H_0[f_m] \equiv 1$$

e

$$H_{n+1}[1; f_m] = \begin{vmatrix} 1 & 1 & \dots & 1 \\ f_m & f_{m+1} & \dots & f_{m+n} \\ f_{m+1} & f_{m+2} & \dots & f_{m+n+1} \\ \vdots & \vdots & & \vdots \\ f_{m+n-1} & f_{m+n} & \dots & f_{m+2n-1} \end{vmatrix} \quad (5.25)$$

Então, nós definimos

$$R_n(f_m) = \frac{H_n[f_m]}{H_n[1; f_m]} \quad (5.26)$$

e

$$S_n(f_m) = \frac{H_{n+1}[1; f_m]}{H_n[f_m]} \quad (5.27)$$

Pye & Atchison (1973) mostraram que $R_n(f_m)$ e $S_n(f_m)$ podem ser calculadas recursivamente pelas seguintes relações: Definimos

$$S_0(f_m) = 1, \quad m = 0, \pm 1, \pm 2, \dots$$

$$R_1(f_m) = f_m, \quad m = 0, \pm 1, \pm 2, \dots$$

então

$$R_{n+1}(f_m) = R_n(f_{m+1}) \left[\frac{S_n(f_{m+1})}{S_n(f_m)} - 1 \right] \quad (5.28)$$

e

$$S_n(f_m) = S_{n-1}(f_{m+1}) \left[\frac{R_n(f_{m+1})}{R_n(f_m)} - 1 \right] \quad (5.29)$$

para $n = 1, 2, \dots$ e $m = 0, \pm 1, \pm 2, \dots$.

Nós estamos supondo em (5.28) e (5.29) $S_n(f_m) \neq 0$ e $R_n(f_m) \neq 0$. Se $R_n(f_{m+1}) = R_n(f_m) = 0$, nós não definimos $S_n(f_m)$.

Contudo, se $R_n(f_m) = 0$ e $R_n(f_{m+1}) \neq 0$, nós definimos $S_n(f_m) = +\infty$. Nós usamos uma definição similar para $R_{n+1}(f_m)$.

DEFINIÇÃO 2 - Dizemos que uma função f é um elemento de $L(n, \Delta)$ sobre um conjunto de inteiros $I = \{m_0, m_0+1, \dots\}$ se existe um menor inteiro $n > 0$ e um conjunto de a_i 's tal que f é uma solução da equação de diferenças

$$Y_m - a_1 Y_{m-1} - \dots - a_n Y_{m-n} = 0, \text{ para } m \in I.$$

As provas dos seguintes teoremas serão omitidas e podem ser encontradas em Gray, Kelley & McIntire (1978).

TEOREMA 1 - Seja $n > 0$, e suponha que $S_n(f_m)$ e $R_n(f_m)$ em (5.26) e (5.27) estejam definidos, isto é, os denominadores em (5.26) e (5.27) não sejam nulos e $S_n(f_m) \neq 0$. Então $f_m \in L(n, \Delta)$ para $m \geq m_0 + n$ se e somente se $S_n(f_m)$ é constante, como uma função de m , para $m \geq m_0$.

TEOREMA 2 - Seja $X(t)$ um processo estacionário ARMA(p,q), dado pela relação (5.5) e com auto-correlações ρ_m . Suponha que $S_n(\rho_m)$ e $R_n(\rho_m)$ estejam definidos, $p > 0$ e $S_n(\rho_m) \neq 0$. Então $S_n(\rho_m)$ é dado por

$$S_n(\rho_m) = \begin{cases} C_1 & , m \geq q-p+1 \\ C_2 & , m \leq -q-p \\ \text{não constante,} & -q-p < m < q-p+1 \end{cases} \quad (5.30)$$

se e somente se $n = p$, onde as constantes C_1 e C_2 são dadas por

$$C_1 = (-1)^p \left[1 - \sum_{k=1}^p a_k \right] \quad (5.31)$$

$$C_2 = -C_1/a_p \quad (5.32)$$

TEOREMA 3 - Se as condições do Teorema 2 são satisfeitas com ρ_m substituído por $(-1)^m \rho_m$, então $S_n((-1)^m \rho_m)$ é dado por

$$S_n((-1)^m \rho_m) = \begin{cases} C_3 & , m \geq q-p+1 \\ C_4 & , m \leq -q-p \\ \text{não constante, } -q-p < m < q-p+1. \end{cases} \quad (5.33)$$

se e somente se $n=p$, onde as constantes C_3 e C_4 são dadas por

$$C_3 = (-1)^p \left[1 - \sum_{k=1}^p (-1)^k a_k \right] \quad (5.34)$$

$$C_4 = (-1)^{p+1} C_3/a_p. \quad (5.35)$$

COROLÁRIO 1 - Sob as suposições do Teorema 2, $R_n(\rho_m)$ e $R_n((-1)^m \rho_m)$, são dadas por

$$R_n(\rho_m) = R_n((-1)^m \rho_m) \begin{cases} =0, & m \geq q-p+1 \text{ e } m < -q-p \\ \neq 0, & -q-p \leq m < q-p+1. \end{cases} \quad (5.36)$$

se e somente se $n=p+1$.

Os Teoremas 2 e 3 mostram que $S_p(\rho_m)$ e $S_p((-1)^m \rho_m)$ são constantes para $m \geq q-p+1$ e para $m \leq -q-p$. O corolário 1 mostra que $R_{p+1}(\rho_m)$ e $R_{p+1}((-1)^m \rho_m)$ são nulos fora do intervalo

$[-q-p; q-p+1)$.

5.5.2 - Determinação de Ordem através das Quantidades R e S

Com base nos resultados da secção anterior, a ordem (p, q) de um modelo ARMA a ser ajustado à série de tempo observada pode ser determinada através do seguinte procedimento.

Calculam-se inicialmente as quantidades R e S, através das relações (5.28) e (5.29), para vários valores de m e n. Essas quantidades são dispostas como nas Tabelas 1 e 2, onde simplificamos a notação fazendo $R_n(f_m) = R_n(m)$ e $S_n(f_m) = S_n(m)$. Chamaremos essas disposições de números das Tabelas 1 e 2 respectivamente de "arranjos S e R".

Tabela 1

		$S_n(f_m)$		
m \ n	n	1	...	k
-l		$S_1(-l)$	...	$S_k(-l)$
-l+1		$S_1(-l+1)$	...	$S_k(-l+1)$
⋮		⋮		⋮
-1		$S_1(-1)$	...	$S_k(-1)$
0		$S_1(0)$	...	$S_k(0)$
1		$S_1(1)$	...	$S_k(1)$
⋮		⋮		⋮
j		$S_1(j)$	...	$S_k(j)$

Tabela 2

$\begin{array}{c c} & n \\ \hline m & \end{array}$		$R_n(f_m)$			
		1	2	...	k+1
-ℓ		$R_1(-ℓ)$	$R_2(-ℓ)$	...	$R_{k+1}(-ℓ)$
-ℓ+1		$R_1(-ℓ+1)$	$R_2(-ℓ+1)$	...	$R_{k+1}(-ℓ+1)$
⋮		⋮	⋮	...	⋮
-1		$R_1(-1)$	$R_2(-1)$	...	$R_{k+1}(-1)$
0		$R_1(0)$	$R_2(0)$	...	$R_{k+1}(0)$
1		$R_1(1)$	$R_2(1)$	...	$R_{k+1}(1)$
⋮		⋮	⋮	...	⋮
j		$R_1(j)$	$R_2(j)$	...	$R_{k+1}(j)$

Tabela 3

$\begin{array}{c c} & n \\ \hline m & \end{array}$		$(S(\rho_m))$			
		1	...	p	p+1
-ℓ		$S_1(-ℓ)$	...	c_2	$c_2[\frac{0}{0}-1]$
-ℓ+1		$S_1(-ℓ+1)$	...	c_2	$c_2[\frac{0}{0}-1]$
⋮		⋮	...	⋮	⋮
-q-p-2		$S_1(-q-p-2)$	...	c_2	$c_2[\frac{0}{0}-1]$
-q-p-1		$S_1(-q-p-1)$	...	c_2	$\pm\infty$
-q-p		$S_1(-q-p)$	...	c_2	$\left. \begin{array}{l} 2q \text{ não} \\ \text{constantes} \end{array} \right\}$
⋮		⋮	...	⋮	
q-p		$S_1(q-p)$	...	c_1	
q-p+1		$S_1(q-p+1)$	...	c_1	$c_1[\frac{0}{0}-1]$
⋮		⋮	...	⋮	⋮
j		$S_1(j)$	...	c_1	$c_1[\frac{0}{0}-1]$

Então, em termos de S e R conforme estão dispostos nas Tabelas 1 e 2 com $f_m = \rho_m$, pelo Teorema 2 e o Corolário 1, temos que um processo é estacionário ARMA(p,q) se e somente se os arranjos S e R associados forem da forma das Tabelas 3 e 4.

Tabela 4

		$R(\rho_m)$			
m	n	1	2	...	$p+1$
$-l$		$\rho(-l)$	$R_2(-l)$	...	0
$-l+1$		$\rho(-l+1)$	$R_2(-l+1)$	...	0
$\vdots$		$\vdots$	$\vdots$		$\vdots$
$-q-p-1$		$\rho(-q-p+1)$	$R_2(-q-p-1)$	...	0
$-q-p$		$\rho(-q-p)$	$R_2(-q-p)$	...	não nulo
$\vdots$		$\vdots$	$\vdots$		$\vdots$
$q-p$		$\rho(q-p)$	$R_2(q-p)$	...	não nulo
$q-p+1$		$\rho(q-p+1)$	$R_2(q-p+1)$	...	0
$\vdots$		$\vdots$	$\vdots$		$\vdots$
j		$\rho(j)$	$R_2(j)$	...	0

Nas Tabelas 3 e 4 nós temos $f_m = \rho_m$. Os arranjos não diferem em seu padrão se $f_m = (-1)^m \rho_m$. Obviamente, nesse caso a coluna 1 do arranjo $-R$ será formada pelos valores de $(-1)^m \rho_m$. Na coluna $p+1$ da Tabela 3 temos vários elementos indefinidos, indicados na forma $\begin{bmatrix} 0 \\ 0 \end{bmatrix} - 1$. Logo, numa série amostrada espera-se que a coluna com as características da coluna p seja seguida por uma coluna altamente variável, pois na prática a

quantidade $\frac{0}{0}$ será substituída por um quociente de dois números pequenos. Analogamente os zeros da coluna $p+1$ no arranjo R , na prática são substituídos por números pequenos.

No exemplo seguinte nós indicaremos com asterisco os elementos $S_m(-m)$, $S_m(-m+1)$ e $R_{m+1}(-m)$, pois os elementos constantes no arranjo $-S$ devem ocorrer simetricamente sobre um desses pares de pontos, e os elementos nulos no arranjo $-R$ devem ocorrer simetricamente sobre um dos $R_{m+1}(-m)$ elementos.

EXEMPLOS 1

Para ilustrar os teoremas anteriores, considere os arranjos R e S para o processo

$$X(t) - 1,32X(t-1) + 0,68X(t-2) = \varepsilon(t) - 0,8\varepsilon(t-1) \quad (5.37)$$

As Tabelas 5 e 6 mostram esses arranjos quando ρ_m é conhecido, usando, respectivamente, $f_m = (-1)\rho_m$ e $f_m = \rho_m$.

As Tabelas 7 e 8 mostram os arranjos R e S , quando ρ_m é substituído pelas auto-correlações amostrais r_m dadas pela relação (3.28), onde $X(t)$ são dados simulados de (5.37) com $\varepsilon(t) \sim N(0,1)$.

Em cada caso os arranjos R e S são mostrados com pelo menos uma coluna e várias linhas a mais do que o necessário. Na prática, o número de linhas e colunas deve ser pré-fixado. Esses números devem ser escolhidos muito maiores do que se espera que seja necessário para não criar problemas. Para as Ta-

TABELA 5 - Valores teóricos de ρ_m

$$f_m = (-1)^m \rho_m$$

Arranjo R			m	Arranjo S		
R_1 (m)	R_2 (m)	R_3 (m)		S_1 (m)	S_2 (m)	S_3 (m)
0,13829	0,0378	0,000000	-10	-2,2806	4,4117	.
-0,17710	-0,0553	0,000000	-9	-1,7928	4,4117	.
0,14042	0,1693	0,000000	-8	-1,0863	4,4117	.
-0,01212	0,2224	0,000000	-7	14,0844	4,4117	indefinido
-0,18295	-0,1005	0,000000	-6	-3,0386	4,4117	indefinido
0,37198	0,0992	0,000000	-5	-2,2198	4,4117	indefinido
-0,45498	-0,1530	0,000000	-4	-1,7356	4,4117	∞
0,33468	0,5636	0,143284	-3	-0,9420	4,4117	6,2513
0,01940	0,4758	0,299323*	-2	-28,3076	5,7403*	-1,5639*
-0,52985	-0,4701*	0,074883	-1	-2,8873*	2,0857*	-3,0000
1,00000*	0,1708	0,000000	0	-1,5298*	3,0000	indefinido
-0,52985	-0,3235	0,000000	1	-1,0366	3,0000	indefinido
0,01940	-0,3832	0,000000	2	16,2492	3,0000	indefinido
0,33468	0,1040	0,000000	3	-2,3592	3,0000	.
-0,45498	-0,0674	0,000000	4	-1,8197	3,0000	.
0,37298	0,0683	0,000000	5	-1,4905	3,0000	.
-0,18295	-0,1512	0,000000	6	-0,9337	3,0000	.
-0,01212	-0,1151	0,000000	7	-12,5774	3,0000	.
0,14042	0,0376	0,000000	8	-2,2612	3,0000	.
-0,17710	-0,0257	0,000000	9	-1,7808	3,0000	.

TABELA 6 - $f_m = \rho_m$

Arranjo R			m	Arranjo S		
$R_1(m)$	$R_2(m)$	$R_3(m)$		$S_1(m)$	$S_2(m)$	$S_3(m)$
0,1382	-0,3078	0,000000	-10	0,2806	0,5294	
0,1771	0,4789	0,000000	-9	-0,2071	0,5294	
0,1404	0,2014	0,000000	-8	-0,9136	0,5294	
0,0121	0,1947	0,000000	-7	-16,0844	0,5294	
-0,1829	0,2940	0,000000	-6	1,0386	0,5294	indefinido
-0,3729	1,0021	0,000000	-5	0,2198	0,5294	indefinido
-0,4549	-1,0045	0,000000	-4	-0,2643	0,5294	∞
-0,3346	-0,5018	-1,194030	-3	-1,0579	0,5294	-3,203096*
0,0194	-0,5119	0,973987*	-2	26,3076	1,7640*	-0,801331*
0,5298	-1,5298*	-0,243666	-1	0,8873*	0,6409*	-0,360000
1,0000*	0,5558	0,000000	0	-0,4701*	0,3600	indefinido
0,5298	0,3481	0,000000	1	-0,9633	0,3600	indefinido
0,0194	0,3412	0,000000	2	-18,2492	0,3600	indefinido
-0,3346	0,6831	0,000000	3	0,3594	0,3600	indefinido
-0,4549	0,6814	0,000000	4	-0,1802	0,3600	.
-0,3729	-0,1999	0,000000	5	-0,5094	0,3600	.
-0,1829	-0,1324	0,000000	6	-1,0662	0,3600	.
0,0121	-0,1369	0,000000	7	10,5774	0,3600	
0,1404	-0,3256	0,000000	8	0,2612	0,3600	
0,1771	0,2093	0,000000	9	-0,2191	0,3600	

TABELA 7 - $f_m = (-1)^m r_m$

Arranjo R				m	Arranjo S		
$R_1(m)$	$R_2(m)$	$R_3(m)$	$R_4(m)$		$S_1(m)$	$S_2(m)$	$S_3(m)$
0,2404	0,0679	-0,0138	0,0229	-10	-2,2615	3,9279	4,4171
0,3032	0,0841	-0,0271	0,1038	-9	-1,7551	4,5711	0,6872
0,2290	0,2619	-0,0316	0,0156	-8	-1,1107	4,0987	-1,5689
-0,0253	0,3654	-0,0184	0,1423	-7	10,3686	3,7441	-0,2366
-0,2881	-0,1261	-0,0174	0,0569	-6	-2,7831	4,2895	1,7045
0,5139	0,1313	-0,0253	-0,2514	-5	-2,1009	3,7228	-2,1323
-0,5657	-0,1718	-0,0127	-0,1365	-4	-1,6131	4,2711	-44,4542
0,3469	0,7293	0,1215	-0,3159*	-3	-0,8142	4,1969	5,4827*
0,0644	0,4241	0,2449*	0,0894	-2	-10,0291	5,3992*	-1,5889*
-0,5818	-0,4182*	0,0710	0,0190	-1	-2,7187*	2,2373*	-3,5896
1,0000*	0,1733	-0,0098	0,0257	0	-1,5818*	3,1537	3,3574
-0,5818	-0,3187	-0,0199	0,0173	1	-1,1108	3,2507	-0,9606
-0,0694	-0,5551	-0,0143	-0,1352	2	4,3825	3,3675	0,2076
0,3469	0,1554	-0,0152	0,0292	3	-2,6310	3,0586	2,0505
-0,5657	-0,0937	-0,0240	-0,0530	4	-1,9084	3,5558	-0,4409
0,5139	0,1197	-0,0203	-0,0222	5	-1,5606	2,8430	-1,5926
-0,2881	-0,2535	-0,0098	-0,0123	6	-0,9120	3,0704	-5,2171
-0,0253	-0,1759	0,0060	-0,0253	7	-10,0349	3,2407	5,5352
-0,2290	0,0694	0,0153	0,0063	8	-2,3244	3,5189	-3,6087
-0,3032	-0,0668	-0,0051	-0,0051	9	-1,7927	2,7119	0,8652
0,2404	0,0731	0,0068		10	-1,2946	2,5241	
-0,0708	4,4244			11	0,0424	2,5202	
-0,0738	-0,0352			12	-2,5003		

TABELA 8 - $f_m = r_m$

Arranjo R				m	Arranjo S		
R_1 (m)	R_2 (m)	R_3 (m)	R_4 (m)		S_1 (m)	S_2 (m)	S_3 (m)
0,2404	-0,5872	-0,1090	-0,0655	-10	0,2615	0,4962	-1,5419
0,3032	0,6024	0,3045	0,0589	-9	-0,2449	0,4064	-1,2102
0,2290	0,3271	-0,1651	0,0314	-8	-0,8893	0,7847	-0,7783
-0,0253	0,3064	0,1900	0,0371	-7	-12,3686	0,3618	-0,9070
-0,2881	0,4471	-0,1444	-0,0851	-6	0,7838	0,5154	-1,1402
-0,5139	2,7360	0,1929	0,8414	-5	0,1009	0,4882	-0,6371
-0,5657	-0,7162	-0,1517	2,0942	-4	-0,3869	0,3567	2,8973
-0,3469	-0,5008	-1,0973	-0,6579*	-3	-1,1858	0,4647	-2,6322*
0,0644	-0,5298	0,9264*	0,1861	-2	8,0291	1,4273*	-0,7628*
0,5818	-1,5818*	-0,2685	-0,2913	-1	0,7187*	0,5914*	-0,2340*
1,0000*	0,6555	-0,0886	-0,0858	0	-0,4182*	0,3492	-1,0032
-0,5818	0,3981	0,2388	-0,0259	1	-0,8892	0,2715	-0,6426
0,0694	0,3811	-0,1087	0,0353	2	-6,3825	0,4416	-0,7958
-0,3469	0,6479	0,1267	0,0588	3	0,6310	0,3675	-1,0172
-0,5657	-1,9509	-0,2484	-0,0301	4	-0,0916	0,3436	-0,7764
-0,5139	-0,4252	0,1059	0,0636	5	-0,4384	0,5443	-0,5559
-0,2881	-0,2125	-0,1098	-0,2739	6	-1,0880	0,2730	-0,2337
0,0253	-0,2197	-0,0471	-0,0880	7	8,0349	0,4094	-1,5920
0,2290	-0,4970	0,1202	0,0530	8	-0,3244	0,4482	-0,4273
0,3032	0,5775	0,0254		9	-0,2073	0,5415	-1,3207
0,2404	0,1342	-0,0267		10	-0,7054	0,6439	
0,0708	0,0918			11	-2,0424	0,4571	
-0,0738	0,1758			12	0,5003		
-0,1107				13	-0,2941		
-0,0782				14			

belas 7 e 8 foi usada uma amostra de 200 observações para se obter r_m . No cálculo dessas estimativas, as quais foram usadas no cálculo dos elementos dos arranjos R e S, o número de "lags" utilizados aumenta conforme os elementos se afastam para a direita da tabela e conforme os elementos verticalmente de $S_i(i+1)$, $S_i(-i)$ e $R_{i+1}(-i)$. Por essa razão os valores mais afastados desses pontos serão mais variáveis. No entanto, para determinar p e q são necessários apenas os valores próximos desses pontos.

A partir da observação dos arranjos conforme estão apresentadas nas Tabelas 7 e 8, podemos identificar o modelo de vários modos. A melhor maneira parece ser primeiro inspecionar o arranjo -S para $f_m = r_m$ e $f_m = (-1)^m r_m$. Escolher o caso que parece mais claro e identificar p como igual ao número de primeira coluna que tem comportamento constante, seguida por uma coluna altamente variável. Esse comportamento é claro na Tabela 7 onde obviamente $p = 2$. A mesma tabela mostra que $q-p+1=0$, logo $q = 1$. Note que $p = 2$ e $q = 1$ podem ser identificados num primeiro exame da tabela uma vez que os dois elementos com asterisco em $S_2(m)$ são claramente distinguíveis.

Note, também, que esta escolha de p e q é confirmada pelo fato que $|S_3(-4)| = 44,4542$, o maior valor da tabela, e $S_3(-1) = -3,5896$. Isto é, certamente, significativo pois se $p = 2$ e $q = 1$ o verdadeiro valor de $|S_3(-4)| = \infty$ enquanto $S_3(-1) = -C_1 = -S_2(0)$. Então, a partir da análise de todas as implicações dos Teoremas 2 e 3, nenhuma outra escolha de p e q parece pos-

sível.

Neste exemplo p e q são ambos facilmente obtidos do arranjo $-R$ ou do arranjo $-S$ para $f_m = (-1)^m r_m$. Isso não ocorre sempre. Em geral, o arranjo $-S$ é melhor para determinar p , enquanto o arranjo $-R$ é melhor para determinar q . Além disso, embora não seja sempre o caso, usualmente $f_m = r_m$ produz uma melhor identificação para séries com alta frequência, enquanto $f_m = (-1)^m r_m$ é usualmente melhor para séries de baixa frequência.

5.5.3 - A estatística $D(n,m)$

Com base nas propriedades de R e S , pode ser definida uma estatística, a qual será denominada estatística $-D$, como uma medida da concordância dos dados observados com o padrão característico de um processo $ARMA(p,q)$ estacionário.

Com essa estatística perde-se alguma informação. No entanto, ela é suficientemente sensível para diferentes valores de p e q , os quais em muitos problemas são adequados para identificar o modelo.

DEFINIÇÃO 3 - Para cada par de inteiros (n,m) tal que $n=0,1,2,\dots,N$ e $m=0,1,2,\dots,M$, definimos

$$D(n,m) = \frac{\left| \frac{S_{n+1}(f_{-m-n-1})}{S_n(f_{-m-n})} \right|}{V}$$

$$V = \left[\frac{S_n(f_{m-n+1}) + S_{n+1}(f_{m-n})}{S_{n-1}(f_{m-n+2}) + S_n(f_{m-n+1})} \right]^2 + \sum_{i=1}^3 \left[\left(\frac{R_{n+1}(f_{-m-n-i})}{R_n(f_{-m-n-i+1})} \right)^2 + \left(\frac{R_{n+1}(f_{m-n+i})}{R_n(f_{m-n+i+1})} \right)^2 \right] \quad (5.38)$$

e

$$S_{-1}(f_m) \equiv 0, \quad R_0(f_m) \equiv 1.$$

A motivação para a definição de $D(n,m)$ é fornecida pela inspeção das Tabelas 3 e 4 e das relações recursivas (5.28) e (5.29). Embora não seja claro das Tabelas 3 e 4, as relações (5.28) e (5.29) mostram que o comportamento-zero no arranjo-R tende a continuar nas colunas $p+2, p+3, \dots$ devido a multiplicação por R_n no cálculo de R_{n+1} . Dessas relações, é claro também que "o efeito infinito e $-C_1$ " tenderá a continuar nas colunas $p+1, p+2, \dots$. No entanto, esse efeito é anulado em $D(n,m)$ pelas divisões no numerador e no denominador, e desse modo um valor grande de $D(n,m)$ é indicativo de um ajuste com um processo ARMA(n,m).

Então a estatística-D para p e q é definida pela seguinte regra: Selecione o modelo ARMA(p,q) onde

$$D(p,q) = \max_{\substack{n=1,2,\dots,N \\ m=0,1,\dots,M}} D(n,m).$$

Portanto, a identificação deverá ser feita através da inspeção dos arranjos R e S e dos valores da estatística-D.

Quando o processo for estacionário normalmente a estatística -D é suficiente para se obter o modelo. No entanto, se o processo for não estacionário ou sazonal, caso em que as colunas no arranjo -R tem séries de valores "próximas de zero" intercalados entre valores não próximos de zero, a estatística -D normalmente não será de grande valor.

Alguns valores da estatística -D, para o Exemplo 1, são dados abaixo

(n,m)	D(n,m)	(n,m)	D(n,m)
(1,0)	0,013	(2,2)	0,540
(1,1)	0,001	(2,3)	0,042
(1,2)	1,041	(3,0)	0,003
(1,3)	1,045	(3,1)	0,001
(2,0)	0,811	(3,2)	0,035
(2,1)	110,755	(3,3)	0,004

Para o caso do processo ser não estacionário, com a não estacionaridade produzida pelo fato das raízes da equação característica estarem sobre o círculo unitário, existe um método para se identificar o modelo ARIMA apropriado para o ajuste. Mas, nesse caso, na prática é mais conveniente fazer-se diferenças e cair no caso estacionário.

CAPÍTULO 6

APLICAÇÕES

6.1 - INTRODUÇÃO

Neste capítulo final apresentaremos os resultados dos estudos de simulação, da aplicação do método da máxima entropia à série de Manchas Solares e da comparação dos vários métodos de determinação de ordem dos processos AR e ARMA.

6.2 - RESULTADOS DOS ESTUDOS DE SIMULAÇÃO

Calculamos as estimativas espectrais auto-regressiva e suavizadas de covariâncias para séries simuladas com 50 e 300 pontos, geradas pelos processos AR(2) e AR(5). Na estimação espectral auto-regressiva a ordem do processo foi determinada usando-se o FPE de AKAIKE e os coeficientes do modelo foram determinados pelos métodos de Yule-Walker e de Burg. Na estimação espectral suavizada de covariâncias foi usada a janela de Parzen, variando-se o ponto de truncamento (PT) até obter-se o valor mais adequado. Obtivemos os seguintes resultados:

(i) Processo AR(2): $X(t) = 0,5X(t-1) - 0,25X(t-2) + \epsilon(t)$.

À série simulada de 50 pontos foi ajustado um modelo AR de ordem 1 e o estimador espectral auto-regressivo não produ-

ziu bons resultados. O estimador espectral suavizado de covariâncias também produziu resultados insatisfatórios.

No entanto, para uma série simulada com 300 pontos foi identificada a ordem correta, ambos os métodos de estimação dos coeficientes produziram valores muito próximos dos verdadeiros e a função densidade espectral (fde) estimada pelo MME praticamente coincide com a do modelo simulado, conforme verificamos na Figura 7.1

Por outro lado, usando-se o método de suavização de covariâncias para um PT de 20, os resultados não foram tão bons, conforme verificamos na Figura 7.2.

ii) Processo AR(5)

$$X(t) = -1,5X(t-1) - 2,1X(t-2) - 1,4X(t-3) - 0,65X(t-4) - 0,14X(t-5) + \epsilon(t).$$

Com uma série simulada de 50 pontos, usando-se o método de Yule-Walker para a estimação dos coeficientes do modelo foi ajustado um modelo de ordem 2, sub-estimando-se a verdadeira ordem do processo e produzindo resultados insatisfatórios. No entanto, usando-se o método de Burg para a estimação dos coeficientes do modelo foi detectada a verdadeira ordem 5 e a estimação espectral auto-regressiva produziu excelentes resultados, obtendo-se ótima resolução na discriminação do pico do espectro.

O método de suavização de covariâncias não apresentou a mesma resolução do estimador espectral auto-regressivo, conforme é mostrado nas Figuras 7.3 e 7.4.

FIGURA 7.1 - Processo AR(2): f.d.e. e estimativa AR de uma série de 300 pontos.

FIGURA 7.2 - Processo AR(2): f.d.e. e estimativa SC de uma série de 300 pontos.

FIGURA 7.3 - Processo $AR(5)$: f.d.e. e estimativa AR de uma série de 50 pontos.

FIGURA 7.4 - Processo AR(5): f.d.e. e estimativa SC de uma série de 50 pontos.

Com 300 pontos, estimando-se os coeficientes pelo método de Burg, o estimador espectral auto-regressivo produz uma estimativa espectral de extrema precisão. Na Figura 7.5 verificamos que essa estimativa é praticamente coincidente com a f.d.e. do modelo simulado.

Usando-se o método de Yule-Walker para estimar os coeficientes, a ordem contínua sendo sub-estimado com o valor 3. A Figura 7.6 mostra que com o estimador suavizado de covariância não se obtém com precisão a f.d.e. do modelo simulado.

6.3 - APLICAÇÃO DO MME À SÉRIE DE MANCHAS SOLARES

Aplicando-se o MME a 88 pontos da série de Manchas Solares, com um intervalo de amostragem $\Delta t = 2$ anos e estimando-se os coeficientes pelo método de Yule-Walker, foi ajustado um modelo AR(5) e detectado um período de 10,353 anos, enquanto usando-se o método de Burg foi ajustado um modelo AR(6) e detectado o período conhecido de 11 anos com grande discriminação.

Pelo método de suavização de covariâncias, com PT de 20, obtivemos um pico pouco definido. Com o aumento do PT para 40 e 60, obtém-se uma maior definição do pico correspondente a um período de 11,429 anos, mas começam a aparecer picos laterais indesejáveis.

A Figura 7.7 compara os estimadores espectrais auto-regressivo e suavizado de covariâncias e a Figura 7.8 compara os estimadores espectrais auto-regressivos, quando se estima os coeficientes pelos métodos de Yule-Walker e de Burg.

FIGURA 7.5 - Processo $AR(5)$: f.d.e. e estimativa
AR de uma série de 300 pontos.

FIGURA 7.6 - Processo AR(5) f.d.e. e estimativa SC de uma série de 300 pontos.

FIGURA 7.7 - Estimativas espectrais AR e SC de 88 pontos da série de manchas solares.

FIGURA 7.8 - Estimativas espectrais de Burg e de Y-W, de 88 pontos de série de manchas solares.

FIGURA 7.9 - Estimativas espectrais AR e SC de 176 pontos da série de manchas solares.

FIGURA 7.10 - Estimativas espectrais de Burg e de Y-W de 176 pontos da série de manchas solares.

A aplicação do MME a 176 pontos da série de Manchas Solares ($\Delta t = 1$ ano), com a estimação dos coeficientes pelo método de Yule-Walker, identificou um modelo AR(8) e indicou um período de 10,353 anos; e com a estimação dos coeficientes pelo método de Burg identificou um modelo AR(11) e detectou o período de 11 anos.

O método de suavização de covariâncias, com PT de 40, continua indicando um período de 11,429 anos, com pouca discriminação conforme é mostrado na Figura 7.9.

A Figura 7.10 compara os estimadores espectrais auto-regressivos, quando se estima os coeficientes pelos métodos de Yule-Walker e de Burg. Com a série de 176 pontos os dois métodos têm praticamente a mesma resolução na determinação do pico do espectro.

6.4 - COMPARAÇÃO DOS MÉTODOS DE DETERMINAÇÃO DA ORDEM DOS PROCESSOS AR E ARMA

Aplicamos inicialmente o critério FPE a 10 séries longas com 300 pontos e a 10 séries curtas com 50 pontos, geradas por cada um dos processos AR(1), AR(2) e AR(5). A estimação dos coeficientes do modelo foi feita usando os métodos de Yule-Walker e de Burg. Os resultados estão resumidos nas Tabelas 7.1, 7.2 e 7.3, onde indicamos as frequências das estimativas de cada ordem.

Na Tabela 7.1 verificamos que para o processo AR(1) foi estimada a ordem correta com grande frequência, não havendo

diferença entre os métodos de estimação dos coeficientes. Para o processo AR(2), a Tabela 7.2 mostra que foi detectada a ordem correta com maior frequência apenas para as séries de 300 pontos, enquanto que para as séries curtas de 50 pontos obteve-se com maior frequência ordem 1. Finalmente, para o processo AR(5) verificamos que quando usamos o método de Burg para estimação dos coeficientes obtemos melhores resultados do que quando usamos o método de Yule-Walker, nesse último caso o FPE sub-estima a ordem do processo.

TABELA 7.1 - Processo AR(1)

N	Método	Ordem									
		1	2	3	4	5	6	7	8	9	10
300	Y-W	50	20	10	-	10	-	-	10	-	-
	Burg	50	20	10	-	10	-	-	10	-	-
50	Y-W	90	10	-	-	-	-	-	-	-	-
	Burg	90	10	-	-	-	-	-	-	-	-

TABELA 7.2 - Processo AR(2)

N	Método	Ordem									
		1	2	3	4	5	6	7	8	9	10
300	Y-W	-	40	20	10	20	-	-	10	-	-
	Burg	-	30	20	10	30	-	-	10	-	-
50	Y-W	70	20	10	-	-	-	-	-	-	-
	Burg	60	30	10	-	-	-	-	-	-	-

Devido ao custo computacional muito elevado, as comparações seguintes foram feitas aplicando os vários métodos de determinação da ordem do processo a apenas uma série longa de

TABELA 7.3 - Processo AR(5)

N	Método	Ordem									
		1	2	3	4	5	6	7	8	9	10
300	Y-W	-	10	40	30	20	-	-	-	-	-
	Burg	-	-	-	20	30	40	-	10	-	-
50	Y-W	-	80	10	10	-	-	-	-	-	-
	Burg	-	-	-	60	30	-	-	-	10	-

300 pontos e a uma série curta de 50 pontos, geradas por cada processo.

Os resultados da comparação do critério FPE com os critérios AIC e CAT estão indicados nas Tabelas 7.4 e 7.5, cujos valores indicam as ordens estimadas por cada um dos métodos.

TABELA 7.4 - AR(2)

N	FPE		AIC	CAT
	YW	Burg		
300	4	4	2	1
50	1	1	2	2

TABELA 7.5 - AR(5)

N	FPE		AIC	CAT
	YW	Burg		
300	3	6	10	6
50	2	4	4	4

Para o processo AR(1) os três métodos indicaram a ordem correta. Os resultados para o processo AR(2) estão indicados na Tabela 7.4, onde verificamos que apenas o AIC indicou em ambos os casos a ordem correta. No entanto, para o processo AR(5) o AIC não forneceu um bom resultado, conforme verificamos na Tabela 7.5.

Para a determinação da ordem do processo ARMA a ser ajustado a uma série temporal usamos os critérios AIC de Akaike,

$$(F(M) = FAC(M))$$
[illegible]

MATRIZ S

[illegible]

TABELA 7.8 - (N = 300)

ESTATISTICA D				
ORDEM DE AR	ORDEM DE MA			
	0	1	2	3
0	.1139E+00	.1289E+02	.2703E+03	.1122E+00
1	.9275E-02	.4004E-01	.4528E-02	.6535E-01
2	.2328E+01	.2039E+01	.2960E+01	.1393E+00
3	.3823E-02	.2534E+02	.4688E-07	.1691E-03
4	.2249E-06	.1832E-09	.1715E+00	.5934E-01

ORDEM DE AUTO-REGRESSAO SELECIONADA - P = 0
 ORDEM DE MEDIAS MOVEIS SELECIONADA - Q = 2

.2025E+02

.1029E-01

.2403E+01

.3945E-04

.9420E-02

$(N=50)$
$$C(M) \cap A = \{M\}$$

МАТРИЦА

[illegible]

MA7012 S

[illegible]

TABELA 7.11 - (N=50)

ESTATISTICA D				
ORDEM DE AP	ORDEM DE MA			
	0	1	2	3
0	.3361E-01	.3991E+01	.1038E+04	.1973E-01
1	.2695E-05	.6769E-05	.2488E-05	.2112E-01
2	.6160E+01	.7950E+00	.7426E-04	.1858E-02
3	.1918E-02	.1407E-03	.2650E-01	.4029E-03
4	.3636E-01	.8221E-01	.1712E-01	.8678E-01

ORDEM DE AUTO-REGRESSAO SFLECCIONADA - P = 0
 ORDEM DE MEDIAS MOVEIS SELECCIONADA - Q = 2

e os arranjos R e S e a estatística D, do método de Gray, Kelley e McIntire.

Obtivemos melhores resultados com o AIC de Akaike o qual, para uma série de 300 pontos indicou o modelo correto ARMA(3,2), e para uma série de 50 pontos indicou o modelo ARMA(1,2).

Usando o método de Gray, Kelley e McIntire para as mesmas séries, obtivemos os seguintes resultados: Para a série de 300 pontos os arranjos R e S indicam o modelo ARMA(3,1) e a estatística D seleciona o modelo ARMA(0,2). Enquanto que para a série de 50 pontos, tanto os arranjos R e S quanto a estatística D indicam modelo ARMA(0,2). Esses resultados estão indicados nas Tabelas 7.6 a 7.11.

6.5 - COMENTÁRIOS FINAIS

Comparando as estimativas espectrais auto-regressiva e suavizada de covariância com o verdadeiro espectro do modelo simulado, podemos constatar que, exacto para o processo AR(2) com 50 observações, em todos os outros casos considerados o estimador espectral auto-regressivo produziu estimativas muito próximas do verdadeiro espectro, obtendo-se ótima resolução na discriminação dos picos do espectro. Por outro lado, o estimador suavizado de covariâncias não produziu bons resultados.

Na estimação espectral auto-regressiva, usando-se o FPE de Akaike para a determinação da ordem do processo, verificamos que se obtém melhores resultados quando se estima os coeficientes pelo método de Burg, do que quando se usa o método

de Yule-Walker, o qual em geral sub-estima a ordem do processo.

Destaquemos, finalmente, que uma vantagem do estimador espectral auto-regressivo é a possibilidade de se usar métodos objetivos de determinação da ordem do modelo AR a ser ajustado à série, enquanto que usando-se o estimador suavizado de covariâncias o ponto de truncamento deve ser fixado subjetivamente. No entanto, as estimativas da ordem nem sempre coincidem com a verdadeira ordem do processo.

APÊNDICE 1

ESTIMADORES AR E O CONCEITO DE ENTROPIA

A1.1 - MÉTODO DA MÁXIMA ENTROPIA

O método da máxima entropia (MME) de análise espectral foi proposto originalmente por Burg (1967,1968) que apresentou não apenas a formulação resultante da aplicação da entropia para a determinação espectral, mas também um método de cálculo dos coeficientes necessários para se obter o erro de previsão. O MME, usando o algoritmo de Burg, foi aplicado por Ulrych (1972), que mostrou as excelentes propriedades de resolução desse método.

A obtenção do espectro de um processo estocástico estacionário amostrado uniformemente pelo método da máxima entropia (MME), é feita maximizando-se a entropia desse processo. Para um melhor entendimento da razão da maximização da entropia, vamos apresentar a relação entre entropia e probabilidade a priori.

A1.2 - INFORMAÇÃO E ENTROPIA

Começemos com uma definição estatística de informação.

Consideremos a situação na qual M diferentes saídas m_i de um sistema possam ocorrer, sendo p_i as probabilidades de ocorrência. Se soubermos, a priori, que todos os p_i são iguais, então não dispomos de nenhuma informação especial sobre o sistema. Se, por outro lado, nós formos capazes de especificar o valor de um particular p_i , nós ganhamos uma certa quantidade de informação sobre o sistema. Está claro então, que a probabilidade de ocorrência de um evento está relacionada com a informação. Além disso, se o evento para o qual nós temos uma especificada probabilidade é pouco provável, a especificação transmite muita informação sobre o sistema. Portanto, nós escrevemos a relação entre informação e probabilidade como

$$I = \log_2 \frac{1}{p_i}$$

O uso do logaritmo na equação acima é justificado, pelo fato que a informação deve ser uma quantidade aditiva.

Suponhamos agora que nós observamos o sistema acima por um longo intervalo de tempo T . Então, se T for muito grande, nós esperamos que ocorram nesse intervalo de tempo T , $p_1 T$ vezes m_1 , $p_2 T$ vezes m_2 , etc. A informação total sobre o sistema será então

$$I_{\text{total}} = p_1 T \log_2 \frac{1}{p_1} + p_2 T \log_2 \frac{1}{p_2} + \dots$$

Define-se, então, (conforme Shannon, 1948), a *entropia* H

como sendo a informação média por intervalo de tempo. Isto é,

$$H = \frac{I_{\text{total}}}{T} = - \sum_{i=1}^M p_i \log_2 p_i.$$

Está claro, da discussão acima da relação entre informação e probabilidade, que entropia é uma medida da incerteza descrita por um conjunto de probabilidades.

A entropia é zero para um sistema onde todos os p_i são iguais a zero, exceto um, o qual é unitário. Neste caso o sistema está perfeitamente determinado, e não existe nenhuma incerteza. Em todos os outros casos a entropia é positiva. Entropia é então, uma medida da desordem de um sistema. De modo alternativo, entropia é uma medida de nossa ignorância sobre a estrutura real do sistema (Brillouin, 1956).

Tendo definido entropia, vejamos como o conceito de máxima entropia se aplica à determinação da função densidade espectral.

O objetivo é obter um método de determinação de função densidade espectral que seja consistente com o conhecimento a priori e que estime a probabilidade a priori fixa que descreve a informação a priori, sem supor nada além disso.

Tendo em vista a relação entre entropia e incerteza, Jaynes (1963, 1968) formulou o seguinte princípio:

A probabilidade a priori fixa que descreve a informação

disponível mas é maximalmente livre com respeito à informação não disponível é aquela com a máxima entropia.

O uso desse princípio elimina as restrições não razoáveis que são exigidas pela análise espectral convencional.

Em geral, o método de determinação da distribuição de probabilidade de máxima entropia é o seguinte (Jaynes, 1968):

O processo x_t que está sendo considerado pode assumir os valores $x_1, x_2, \dots, x_n$. Nós supomos que a informação disponível sobre x_t está na forma dos valores médios $\langle f_1(x_t) \rangle, \langle f_2(x_t) \rangle, \dots, \langle f_m(x_t) \rangle$ de várias funções $f_1(x_t), f_2(x_t), \dots, f_m(x_t)$, $m < n$. A distribuição de probabilidade $p_i = p(x_i)$ que é consistente com essa informação, mas é maximalmente livre de outras restrições é aquela que maximiza a entropia

$$H = -\sum_i p_i \log_2 p_i.$$

A maximização de H está sujeita à restrição

$$\sum_i p_i = 1$$

e

$$\sum_i p_i f_k(x_i) = \langle f_k(x_i) \rangle, \quad k=1, 2, \dots, m.$$

A solução desse problema variacional é

$$p_i = \frac{1}{Z(\lambda_1, \dots, \lambda_m)} \exp [\lambda_1 f_1(x_i) + \dots + \lambda_m f_m(x_i)]$$

com a função partição

$$Z(\lambda_1, \dots, \lambda_m) = \sum_i \exp[\lambda_1 f_1(x_i) + \dots + \lambda_m f_m(x_i)].$$

Os multiplicadores de Lagrange λ_k são determinados das restrições, as quais se reduzem às relações

$$\langle f_k(x_i) \rangle = \frac{\partial}{\partial \lambda_k} \log Z(\lambda_1, \dots, \lambda_m).$$

Jaynes (1968) mostrou também que "a distribuição de probabilidade que maximiza a entropia é numericamente igual à distribuição de freqüências a qual é obtida o maior número de vezes". Essa distribuição de freqüências é a distribuição de $u_i = \frac{m_i}{M}$, onde m_i é o número de vezes que ocorre x_i em M realizações do processo x_t .

Aplicando o conceito de máxima entropia à análise espectral, nós apresentamos inicialmente a relação entre entropia (estritamente falando, a razão de entropia para um processo infinito) e a densidade espectral $f(\lambda)$ de um processo Gaussiano estacionário,

$$H = \frac{1}{4\lambda_N} \int_{-\frac{\lambda_N}{2\pi}}^{\frac{\lambda_N}{2\pi}} \log f(\lambda) d\lambda \quad (A1.1)$$

onde λ_N é a freqüência angular de Nyquist. A derivação de (A1.1) é mostrada em detalhe em Smylie et alii (1973). Reescrevendo (A1.1) em termos das auto-correlações ρ_k do processo, obtemos

$$H = \frac{1}{4\lambda_N} \int_{-\frac{\lambda_N}{2\pi}}^{\frac{\lambda_N}{2\pi}} \log \left[\sum_{-\infty}^{\infty} \rho(k) \exp(-i\lambda k \Delta t) \right] d\lambda \quad (A1.2)$$

onde Δt é a razão uniforme de amostragem.

Maximizando (A1.2) em relação aos valores desconhecidos $\rho(k)$ com a restrição que $f(\lambda)$ deve ser consistente com as auto-correlações conhecidas $\rho(0), \rho(1), \dots, \rho(M-1)$ resulta na estimativa espectral pelo MME. Essa estimativa expressa máxima incerteza com relação à informação desconhecida mas é consistente com a informação conhecida.

O procedimento variacional conduz à expressão para a densidade espectral pelo MME (Smylie et alii, 1973; Edward & Fildelson, 1973), a qual para um processo linear x_t é

$$P_E(\lambda) = \frac{P_M}{\frac{\lambda_N}{2\pi} \left| 1 + \sum_{j=1}^{M-1} \beta_j \exp(-i\lambda j \Delta t) \right|^2} \quad (A1.3)$$

Em (A1.3) P_M é uma constante e os β_j são os coeficientes do erro de previsão que são determinados pelos dados.

A1.3 - A CORRESPONDÊNCIA ENTRE UM PROCESSO AR E O MME

A relação entre o MME de análise espectral e a representação AR de um processo estocástico foi estabelecida por Van den Bos (1971).

Smylie et alii, (1973) mostraram que a entropia de um pro-

cesso discreto estacionário Gaussiano $X(t)$, $t = 1, 2, \dots, M+1$ a qual é equivalente a (A1.1), é dada por

$$H = \frac{1}{2} \log \{ \det[C(M)] \} \quad (A1.4)$$

onde $C(M)$ é a matriz Toeplitz semi-positiva definida das auto-covariâncias do processo $X(t)$.

$$C(M) = \begin{bmatrix} \gamma(0) & \gamma(1) & \dots & \gamma(M) \\ \gamma(1) & \gamma(0) & \dots & \gamma(M-1) \\ \vdots & \vdots & & \vdots \\ \gamma(M) & \gamma(M-1) & \dots & \gamma(0) \end{bmatrix} \quad (A1.5)$$

Vamos assumir que os valores $\gamma(0), \gamma(1), \dots, \gamma(M)$ da função de auto-covariância para os primeiros $(M+1)$ "lags" sejam conhecidos. A idéia na qual está baseado o MME é escolher as auto-covariâncias desconhecidas $\gamma(M+1), \gamma(M+2)$, etc., de tal modo que a entropia do processo seja maximizada em cada passo.

Então da relação (A1.4), segue que $\gamma(M+1)$ é determinada maximizando $\det[C(M+1)]$ em relação a $\gamma(M+1)$. Uma vez que $\gamma(M+1)$ tenha sido obtida, $\gamma(M+2)$ é determinada substituindo o valor de $\gamma(M+1)$ em $C(M+2)$ e maximizando em relação a $\gamma(M+2)$. Esse procedimento é repetido para as outras auto-covariâncias.

Pode ser mostrado que a solução para a maximização de $\det[C(M+1)]$ em relação a $\gamma(M+1)$ é dada por

$$\det \begin{bmatrix} \gamma(1) & \gamma(0) & \dots & \gamma(M-1) \\ \gamma(2) & \gamma(1) & \dots & \gamma(M-2) \\ \vdots & \vdots & & \vdots \\ \gamma(M+1) & \gamma(M) & \dots & \gamma(1) \end{bmatrix} = 0. \quad (A1.6)$$

Por outro lado, considerando um processo AR de ordem M e substituindo $k = 1, 2, \dots, M+1$ na equação (2.9), temos as equações de Yule-Walker na forma

$$\begin{cases} \gamma(1) - \alpha_1 \gamma(0) - \dots - \alpha_M \gamma(M-1) = 0 \\ \gamma(2) - \alpha_1 \gamma(1) - \dots - \alpha_M \gamma(M-2) = 0 \\ \vdots \\ \gamma(M+1) - \alpha_1 \gamma(M) - \dots - \alpha_M \gamma(1) = 0 \end{cases} \quad (A1.7)$$

Suponha agora que os valores $\gamma(0), \gamma(1), \dots, \gamma(M)$ da função de auto-covariância para os primeiros $(M+1)$ "lags" sejam conhecidos. A substituição desses valores nas M primeiras equações de (A1.7) produz os coeficientes $\alpha_1, \alpha_2, \dots, \alpha_M$. Tendo em vista a equação (2.9) para $M = 1, 2, \dots, M+2$, temos

$$\begin{cases} \gamma(1) - \alpha_1 \gamma(0) - \dots - \alpha_M \gamma(M-1) - \alpha_{M+1} \gamma(M) = 0 \\ \gamma(2) - \alpha_1 \gamma(1) - \dots - \alpha_M \gamma(M-2) - \alpha_{M+1} \gamma(M-1) = 0 \\ \vdots \\ \gamma(M+1) - \alpha_1 \gamma(M) - \dots - \alpha_M \gamma(1) - \alpha_{M+1} \gamma(0) = 0 \end{cases} \quad (A1.8)$$

e para que o sistema de equações (A1.8) seja compatível e re-

caia no sistema formado pelas M primeiras equações de (A1.7), a auto-covariância desconhecida $\gamma(M+1)$ deve satisfazer a condição

$$\det \begin{bmatrix} \gamma(1) & \gamma(0) & \dots & \gamma(M-1) \\ \gamma(2) & \gamma(1) & \dots & \gamma(M-2) \\ \vdots & \vdots & & \vdots \\ \gamma(M+1) & \gamma(M) & \dots & \gamma(1) \end{bmatrix} = 0 \quad (\text{A1.9})$$

A comparação de (A1.9) e (A1.6), mostra que exatamente o mesmo valor $\gamma(M+1)$ é determinado maximizando a entropia do processo. Essa equivalência estende-se a todas as outras auto-covariâncias. Portanto, concluímos que o MME de análise espectral é equivalente a ajustar um modelo AR a um processo estocástico.

O espectro de um processo AR, dado por (2.21), compara-se com (A1.3).

Logo, a representação de um processo estocástico por modelos AR é a representação que exhibe a máxima entropia.

Um processo AR

$$X(t) = a_1 X(t-1) + a_2 X(t-2) + \dots + a_M X(t-M) + \varepsilon(t) \quad (\text{A1.10})$$

pode ser interpretado como o processo que identifica $X(t)$ com o valor que é previsto dos M valores anteriores do processo, com um erro de previsão igual a $\varepsilon(t)$. Então $a_1, a_2, \dots, a_M$ são os coeficientes de um filtro de previsão. Reescrevendo (A1.10)

como

$$\varepsilon(t) = X(t) - a_1X(t-1) - a_2X(t-2) - \dots - a_MX(t-M)$$

identificamos $1, -a_1, -a_2, \dots, -a_M$ com os $M+1$ coeficientes do filtro do erro de previsão. Se os coeficientes $-a_1, -a_2, \dots, -a_n$ forem indicados por $\beta_1, \beta_2, \dots, \beta_M$ os coeficientes do filtro do erro de previsão tornam-se $1, \beta_1, \beta_2, \dots, \beta_M$.

APÊNDICE 2

ESTIMATIVAS DE BURG DOS COEFICIENTES AR

Burg (1967,1968) sugeriu um método de estimar os coeficientes AR (ou de modo equivalente os coeficientes do filtro do erro de previsão) que não requer a estimativa a priori da função de auto-covariância. O método de Burg usa uma recursão muito similar à desenvolvida para as estimativas de Yule-Walker.

Nós ilustraremos a recursão calculando as estimativas $\hat{a}_3(j)$, $j = 1, 2, 3$ das estimativas $\hat{a}_2(j)$, $j = 1, 2$. Usando as equações de Yule-Walker (3.25) com $M = 3$ para $\ell = 1, 2$, as estimativas $\hat{a}_3(1)$ e $\hat{a}_3(2)$ podem ser expressas em termos de $\hat{a}_3(3)$ na forma

$$\begin{bmatrix} C(0) & C(1) \\ C(1) & C(0) \end{bmatrix} \begin{bmatrix} \hat{a}_3(1) \\ \hat{a}_3(2) \end{bmatrix} = \begin{bmatrix} C(1) \\ C(2) \end{bmatrix} - \hat{a}_3(3) \begin{bmatrix} C(2) \\ C(1) \end{bmatrix},$$

fazendo

$$V(1) = \begin{bmatrix} C(0) & C(1) \\ C(1) & C(0) \end{bmatrix}$$

as equações acima podem ser reescritas como

$$\begin{bmatrix} \hat{a}_3(1) \\ \hat{a}_3(2) \end{bmatrix} = V(1)^{-1} \begin{bmatrix} C(1) \\ C(2) \end{bmatrix} - V(1)^{-1} \hat{a}_3(3) \begin{bmatrix} C(2) \\ C(1) \end{bmatrix}.$$

Além disso, como substituindo $M = 2$ em (3.25) temos

$$\begin{bmatrix} \hat{a}_2(1) \\ \hat{a}_2(2) \end{bmatrix} = V(1)^{-1} \begin{bmatrix} C(1) \\ C(2) \end{bmatrix},$$

segue que

$$\begin{bmatrix} \hat{a}_3(1) \\ \hat{a}_3(2) \end{bmatrix} = \begin{bmatrix} \hat{a}_2(1) \\ \hat{a}_2(2) \end{bmatrix} - \hat{a}_3(3) \begin{bmatrix} \hat{a}_2(2) \\ \hat{a}_2(1) \end{bmatrix}. \quad (A2.1)$$

Como a relação (A2.1) não depende de $C(k)$, nós podemos usá-la para relacionar $\hat{a}_3(1)$ e $\hat{a}_3(2)$ com $\hat{a}_2(1)$ e $\hat{a}_2(2)$, bastando para isso determinar $\hat{a}_3(3)$. Pela técnica de Burg, seguindo o princípio de mínimos quadrados, nós determinamos $\hat{a}_3(3)$ minimizando $\sum \epsilon^2(t)$ em relação a $\hat{a}_3(3)$. Isto é equivalente a minimizar a variância do erro de previsão em relação a $\hat{a}_3(3)$. Então $\hat{a}_3(3)$ pode ser obtido da expressão

$$\frac{\partial S(a_3(3))}{\partial a_3(3)} = 0 \quad (A2.2)$$

onde $S(a_3(3))$ é a soma de quadrados residual para o processo AR de 3ª ordem e é dada por

$$S(\hat{a}_3(3)) = \sum_{t=4}^N [X(t) - \hat{a}_3(1)X(t-1) - \hat{a}_3(2)X(t-2) - \hat{a}_3(3)X(t-3)]^2 \quad (A2.3)$$

A solução de (A2.2) para $\hat{a}_3(3)$ usando as estimativas $\hat{a}_3(1)$ e $\hat{a}_3(2)$ obtidas da recursão dada por (A2.1) corresponde, aproximadamente, à estimação de máxima verossimilhança desses coeficientes (veja Box & Jenkins, 1976).

Uma importante extensão para (A2.3) foi proposta por Burg (1968). Burg sugeriu que a variância do erro de previsão seja calculada passando o filtro do erro de previsão sobre os dados nos dois sentidos, para trás e para frente. A expressão para a variância do erro para o processo AR de 3.^a ordem é

$$P_4 = \frac{1}{2(N-3)} \sum_{t=4}^N \left\{ (X(t) - \hat{a}_3(1)X(t-1) - \hat{a}_3(2)X(t-2) - \hat{a}_3(3)X(t-3))^2 + \right. \\ \left. + (X(t-3) - \hat{a}_3(1)X(t-2) - \hat{a}_3(2)X(t-1) - \hat{a}_3(3)X(t))^2 \right\} \quad (A2.4)$$

e $\hat{a}_3(3)$ é determinado de $\frac{\partial P_4}{\partial \hat{a}_3(3)} = 0$.

O ponto importante a ressaltar sobre (A2.4) é que P_4 é determinada passando o filtro sobre os dados, mas não desprezando nenhuma informação obtida dos dados. Isto é, não é feita nenhuma suposição sobre a extensão dos dados fora do espaço paramétrico, e portanto, a estimação dos coeficientes AR é consistente com o princípio da máxima entropia.

BIBLIOGRAFIA

- [1] - AKAIKE, H. (1969a), Fitting autoregressive models for prediction, *Annals of the Institute of Statistical Mathematics*, 21: 243-247.
- [2] - AKAIKE, H. (1969b), Power spectrum estimation through autoregressive model fitting, *Annals of the Institute of Statistical Mathematics*, 21: 407-419.
- [3] - AKAIKE, H. (1970a), Statistical predictor identification, *Annals of the Institute Statistical Mathematics*, 22: 203-217.
- [4] - AKAIKE, H. (1970b), On a semi-automatic power spectrum estimation procedure, In: INTERNATIONAL CONFERENCE ON SYSTEM SCIENCES, 3^o, Hawaii, *Proceedings*.
- [5] - AKAIKE, H. (1974), A new look at the statistical model identification, *IEEE Transactions on Automatic Control*, 19 (6): 716-723.
- [6] - AKAIKE, H. (1977), On entropy maximization principle. In: KRISHNAIAH, P. R. ed. *Applications of statistics: proceedings of the Symposium Held at Wright State University, Dayton, 1976*, Amsterdam, North-Holland, pp. 27-41.
- [7] - AKAIKE, H. (1978), Time series analysis and control through parametric models, In: FINDLEY, D.F. ed., *Applied time series analysis*, New York, Academic Press, pp. 1-24.
- [8] - ANDERSON, T.W. (1963), Determination of the order of dependence in normally distributed time series, In: ROSENBLATT, M. ed. *Proceedings of the Symposium on Time Series Analysis Providence, 1962*, New York, John Wiley, pp. 425-466. (The SIAM Series Applied Mathematics).
- [9] - ANDERSON, T.W. & WALKER, A.M. (1964), On the asymptotic distribution of the autocorrelations, of a sample from a linear stochastic process, *The Annals of Mathematical Statistics*, 35 (3): 1296-1303.

- [10] - ANDERSON, T.W. (1971), *The statistical analysis of time series*, New York, John Wiley, 704p. (Wiley Series in Probability and Mathematical Statistics).
- [11] - BARTLETT, M.S. (1948), Smoothing periodograms from time-series with continuous spectra, *Nature*, 161: 686-687.
- [12] - BEAMISH N. & PRIESTLEY, M.B. (1981), A study of autoregressive and window spectral estimation, *Applied Statistics*, 30 (1): 41-58.
- [13] - BERK, K.N. (1974), Consistent autoregressive spectral estimates, *Annals of Statistics*, 2 (3): 489-502.
- [14] - BOX, G.E. & JENKINS, G.M. (1976), *Time series analysis: forecasting and control*, Rev. ed. San Francisco, Holden-Day, 575p. (Holden-Day Series in Time Series Analysis).
- [15] - BRILLINGER, D.R. (1975), *Time series, data analysis and theory*, New York, Holt, 500p. (International Series in Decision Processes).
- [16] - BRILLOUIN, L. (1956), *Science and information theory*, New York, Academic Press, 320p.
- [17] - BURG, J.P. (1967), Maximum entropy spectral analysis, paper presented at the 37th Annual International Meeting Soc. of Explor. Geophys, Oklahoma.
- [18] - BURG, J.P. (1968), A new analysis technique for time series data, NATO ADVANCED STUDY INSTITUTE ON SIGNAL PROCESSING, Netherlands.
- [19] - CRAMÉR, H. (1946), *Mathematical methods of statistics*, Princeton, University Press, 575p.
- [20] - CRAMÉR, H. & LEADBETTER, M.R. (1967), *Stationary and related stochastic processes: sample function properties and their applications*, New York, John Wiley, 348p.
- [21] - DOOB, J.L. (1953), *Stochastic processes*, New York, John Wiley, 654p. (Wiley Publications in Statistics).
- [22] - DURBIN, J. (1959), Estimation of parameters in time-series regression models, *Journal of the Royal Statistical Society, Series B*, 22: 139-153.

- [23] - EDWARD, J.A. & FITELSON, M.M. (1973), Notes on maximum entropy spectral analysis, *IEEE Transactions on Information Theory*, 19:232.
- [24] - FRISCH, R. (1933), Propagation problems and impulse problems in dynamical economics, *Economic Essays in Honor of Gustav Cassal*, pp 171-205, London, Allen & Unwin.
- [25] - GRAY, H.L., KELLEY, G.D. & MCINTIRE, D.D. (1978), A new approach to arma modeling, *Communications in Statistics*, B7 (1): 1-77.
- [26] - GRENANDER, V. & ROSENBLATT, M. (1957), *Statistical analysis of stationary time series*, New York, John Wiley, 300p. (Wiley Series in Probability and Mathematical Statistics).
- [27] - GRIFFITHS, L.J. & PRIETRO-DIAZ, R. (1977), Spectral analysis of natural seismic events using autoregressive techniques, *IEEE Transactions on Geoscience Electronics*, 15: 13-25.
- [28] - HANNAN, E.J. (1970), *Multiple time series*, N.York, John Wiley, 536p, (Wiley series in Probability and Mathematical Statistics).
- [29] - HANNAN, E.J. (1973), The asymptotic theory of linear time-series models, *Journal of Applied Probability*, 10: 130-145.
- [30] - JAYNES, E.T. (1963), New engineering applications of information theory, In: BOGDANOFF, J.L. & KOZIN, F. eds. *Proceedings of the First Symposium on Engineering Applications of Random Function Theory and Probability*, 1962, New York, John Wiley, pp. 163-203.
- [31] - JAYNES, E.T. (1968), Prior probabilities, *IEEE Transactions on Systems Science and Cybernetics*, 4: 227-241.
- [32] - JENKINS, G.M. & WATTS, D.G. (1968), *Spectral analysis and its applications*, San Francisco, Holden-Day, 525p.
- [33] - JONES, R.H. (1974), Identification and autoregressive spectrum estimation, *IEEE Transactions on Automatic Control*, 19 (6): 894-898.
- [34] - KOOPMANS, L.H. (1974), *The spectral analysis of time series*, N.York, Academic Press, 366p. (Probability and Mathematical Statistics, 22).
- [35] - KROMER, R.E. (1969), *Asymptotic properties of the autoregressive spectral estimator*, Ph.D. Thesis, Stanford University.

- [36] - MAKHOUL, J. (1975), Linear prediction a tutorial review, *Proceedings of the IEEE*, 63 (4): 561-580.
- [37] - MILLER, K.S. (1968), *Linear difference equations*, New York, W. A. Benjamin, 105p. (Mathematics Monograph Series).
- [38] - MORETTIN, P.A. (1979), *Análise hârmônica de processos estocásticos*, Rio de Janeiro, IMPA, 176p. Trabalho apresentado ao 12º Colóquio Brasileiro de Matemática, Poços de Caldas, 1979.
- [39] - NELSON, C.R. (1973), *Applied time series analysis for managerial forecasting*, San Francisco, Holden-Day, 231p.
- [40] - OZAKI, T. (1977), On the order determination of ARIMA models, *Applied Statistics*, 26 (3): 290-301.
- [41] - PARZEN, E. (1969), Multiple time series modeling, In: KRISHNAIAH, P.R. ed. *Multivariate analysis: proceedings of the Second International Symposium on Multivariate Analysis*, Dayton, 1968, New York, Academic Press, pp. 389-409.
- [42] - PARZEN, E. (1974), Some recent advances in time series modeling, *IEEE Transactions on Automatic Control*, 19 (6): 723-730.
- [43] - PARZEN, E. (1977), Multiple time series: determining the order of approximating autoregressive schemes, In: KRISHNAIAH, P.R. ed., *Multivariate analysis: proceedings of the Fourth International Symposium on Multivariate Analysis*, Dayton, 1975, Amsterdam, North-Holland, pp. 283-295.
- [44] - SEARLE, S.R. (1971), *Linear Models*, New York, John Wiley, 532p. (Wiley Series in Probability and Mathematical Statistics).
- [45] - SHANNON, C.E. (1948), A mathematical theory of communication, *Bell. Syst. Tech. J.*, 27: 379-423.
- [46] - SHIBATA, R. (1976), Selection of the order of an autoregressive model by Akaike's information criterium, *Biometrika*, 63(1): 117-126.
- [47] - SMYLIE, O.E.; CLARKE, G.K.C. & ULRICH, T.J. (1973), Analysis of irregularities on the Earth's rotation, *Methods in Computational Physics*, 13: 391-430.

- [48] - TINBERGEN, J. (1937), *An economic approach to business cycle problems*, Paris, Hermann.
- [49] - TONG, H. (1977), Some comments on the canadian lynx data, *Journal of the Royal Statistical Society, Serie A*, 140(4): 432-436.
- [50] - ULRYCH, T.J. (1972), Maximum entropy power spectrum of truncated sinewaves, *Journal of Geophysical Research*, 77: 1396-1400.
- [51] - ULRYCH, T.J. & BISHOP, T.N. (1975), Maximum entropy spectral analysis and autoregressive decomposition, *Reviews of Geophysics and Space Physics*, 13 (1): 183-200.
- [52] - VAN DE BOS, A. (1971), Alternative interpretation of maximum entropy spectral analysis, *IEEE Transaction on Information*, 17: 493-494.
- [53] - YAGLOM, Q.M. (1962), *An introduction to the theory of stationary random functions*, Englewood Cliffs, Prentice-Hall, 235p.
- [54] - YULE, G.U. (1927), On a method of investigating periodicities in disturbed series with special inference to Wolfer's sunspot numbers, *Philosophical transactions of the Royal Society of London, Serie A*, 226: 267-298.