

Processos Estocásticos

Paulo Lopes dos Santos
T-P Perdicoulis

4 de dezembro de 2024

Conteúdo

1	Estimação de Espectro Processos Estocásticos	1
1.1	Introdução	1
1.2	Estimação Paramétrica	2
1.2.1	Processos AR	2
1.2.2	Determinação da ordem de modelos AR - Autocorrelação parcial	9
1.2.3	Modelos MA	13
1.2.4	Determinação da ordem de modelos MA - Sequência de Autocorrelação	25
1.2.5	Modelos ARMA	25

Resumo

Capítulo 1

Estimação de Espectro Processos Estocásticos

1.1 Introdução

A estimação do espectro de processos estocásticos é uma ferramenta essencial no processamento estatístico de sinal, desempenhando um papel crucial em uma ampla gama de aplicações científicas e tecnológicas. Análise da densidade espectral dum sinal e revela como é que a sua energia distribui pelas várias frequências, permitindo uma compreensão profunda e o processamento eficaz de sinais em campos tão diversos como telecomunicações, processamento de áudio, geofísica, entre outros.

A capacidade de decompor um sinal em componentes de frequência facilita não só a análise detalhada do mesmo, mas também a implementação de técnicas de filtragem eficazes, que são fundamentais para a extração ou supressão de componentes específicas de frequência, aprimorando aplicações de remoção de ruído e compressão de dados. Além disso, a análise espectral é utilizada para modelar e prever séries temporais em economia e finanças, ajudando na tomada de decisões estratégicas através da identificação de ciclos económicos ou tendências de mercado.

As técnicas para a estimação do espectro podem ser classificadas em métodos paramétricos e não-paramétricos. Os métodos paramétricos assumem que o processo estocástico pode ser descrito por modelos auto-regressivos (AR), média móvel (MA) e auto-regressivo-média móvel (ARMA). Estes modelos são particularmente úteis quando os dados disponíveis são limitados ou quando existe um conhecimento a priori que pode ser incorporado na modelagem.

Por outro lado, os métodos não-paramétricos, incluindo o periodograma e a estimativa de Blackman-Tukey, não dependem de suposições sobre a natureza do processo estocástico, oferecendo uma abordagem mais flexível e amplamente aplicável. Estes métodos, contudo, podem sofrer de problemas como alta variância e baixa resolução espectral, especialmente em conjuntos de dados de pequena dimensão.

1.2 Estimação Paramétrica

1.2.1 Processos AR

Densidade espectral dum processo AR

Os processos AR são descritos pela equação

$$A(q^{-1})y(t) = e(t), \quad (1.1)$$

onde

$$A(q^{-1}) = 1 + a_1 q^{-1} + \dots + a_{n_a} q^{-n} \quad (1.2)$$

e $e(t)$ representa um ruído branco com média nula e variância σ^2 . A partir de (1.1), temos que

$$y(t) = \frac{1}{A(q^{-1})} e(t). \quad (1.3)$$

Como os processos AR são, por definição, estacionários, a densidade espectral de potência é dada por:

$$\Phi_{yy}(\omega) = \frac{1}{|A(e^{-j\omega})|^2} \sigma^2 = \frac{1}{|1 + a_1 e^{-j\omega} + \dots + a_{n_a} e^{-j\omega n_a}|^2} \sigma^2. \quad (1.4)$$

Desta forma, concluímos que a estimação do espectro está intimamente relacionada com a identificação do modelo AR.

Estimador de Yule Walker

Os parâmetros dum modelo AR podem ser determinados a partir de estimativas da sequência de autocovariância do processo. Vimos, aquando da apresentação dos modelos AR, que os primeiros $n+1$ elementos da sequência de autocovariância dum processo AR(n) pode ser determinada através sistema de equações:

$$\begin{cases} \lambda_{yy}(0) + a_1 \lambda_{yy}(1) + \dots + a_n \lambda_{yy}(n) = \sigma^2 \\ \lambda_{yy}(1) + a_1 \lambda_{yy}(0) + \dots + a_n \lambda_{yy}(n-1) = 0 \\ \lambda_{yy}(2) + a_1 \lambda_{yy}(1) + \dots + a_n \lambda_{yy}(n-2) = 0 \\ \vdots \\ \lambda_{yy}(n) + a_1 \lambda_{yy}(n-1) + \dots + a_n \lambda_{yy}(0) = 0 \end{cases} \quad (1.5)$$

As equações deste sistema são designadas por equações de Yule Walker [13, 12, 2, 6] e também podem ser utilizadas para estimar os parâmetros do modelo dada a sequência de autocovariância $\lambda_{yy}(\tau)$, $\tau = 1, \dots, n$. Temos então:

$$\begin{cases} \lambda_{yy}(0)a_1 + \lambda_{yy}(1)a_2 + \dots + \lambda_{yy}(n-1)a_n = -\lambda_{yy}(1) \\ \lambda_{yy}(1)a_1 + \lambda_{yy}(0)a_2 + \dots + \lambda_{yy}(n-2)a_n = -\lambda_{yy}(2) \\ \vdots \\ \lambda_{yy}(n-1)a_1 + \lambda_{yy}(n-2)a_2 + \dots + \lambda_{yy}(0)a_n = -\lambda_{yy}(n), \end{cases} \quad (1.6)$$

Definindo

$$\boldsymbol{\theta}^{(n)} = [a_1 \ a_2 \ \cdots \ a_n]^T \quad (1.7)$$

$$\mathbf{\Lambda}_n = \begin{bmatrix} \lambda_{yy}(0) & \lambda_{yy}(1) & \cdots & \lambda_{yy}(n-1) \\ \lambda_{yy}(1) & \lambda_{yy}(0) & \cdots & \lambda_{yy}(n-2) \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_{yy}(n-1) & \lambda_{yy}(n-2) & \cdots & \lambda_{yy}(0) \end{bmatrix} \quad (1.8)$$

$$\boldsymbol{\lambda}_{1:n} = [\lambda_{yy}(1) \ \lambda_{yy}(2) \ \cdots \ \lambda_{yy}(n)]^T \quad (1.9)$$

este sistema de equações pode ser rescrito na forma matricial:

$$\mathbf{\Lambda}_n \boldsymbol{\theta}^{(n)} = -\boldsymbol{\lambda}_{1:n}, \quad (1.10)$$

sendo os parâmetros calculados através de

$$\boldsymbol{\theta}^{(n)} = -\mathbf{\Lambda}_n^{-1} \boldsymbol{\lambda}_{1:n}. \quad (1.11)$$

Uma vez determinados os parâmetros $a_1, a_2, \dots, a_n$, a variância do ruído branco, σ^2 , pode ser calculada a partir da primeira equação do sistemas de equações (1.6):

$$\sigma^2 = \lambda_{yy}(0) + \lambda_{yy}(1)a_1 + \cdots + \lambda_{yy}(n)a_n. \quad (1.12)$$

Se, como acontece em problemas do mundo real, tivermos estimativas da sequência autocovariância, podemos estimar os parâmetros através da equação (1.11) substituindo os elementos da sequência de autocovariância pelas estimativas.

Algoritmo de Levinson-Durbin

A matriz $\mathbf{\Lambda}_n$ é uma matriz de Toeplitz, caracterizada por ter elementos constantes ao longo de cada diagonal [8]. Além disso, $\mathbf{\Lambda}_n$ é simétrica. O algoritmo de Levinson-Durbin [6] explora esta estrutura para obter um tempo de resolução de ordem n^2 , o que é significativamente mais eficiente do que os métodos de eliminação de Gauss-Jordan, que têm uma complexidades de ordem n^3 . Este algoritmo calcula recursivamente os parâmetros de modelos auto-regressivos de ordens de 1 a n , ou seja, os coeficientes dos polinômios:

$$A^{(k)}(q^{-1}) = 1 + a_1^{(k)}q^{-1} + \cdots + a_k^{(k)}q^{-k}, \quad k = 1, \dots, n. \quad (1.13)$$

A ideia principal da recursividade é calcular $\boldsymbol{\theta}^{(k+1)}$ do modelo de ordem $k+1$ a partir de $\boldsymbol{\theta}^{(k)}$ do modelo de ordem k .

Para $k = 1$, o sistema de equações (1.6) reduz-se a uma única equação cuja solução é

$$\boldsymbol{\theta}^{(1)} = a_1^{(1)} = -\frac{\lambda_{yy}(1)}{\lambda_{yy}(0)}. \quad (1.14)$$

A recursividade é desenvolvida resolvendo a equação (1.10) para a ordem $k+1$:

$$\mathbf{\Lambda}_{k+1} \boldsymbol{\theta}^{(k+1)} = -\boldsymbol{\lambda}_{1:k+1}. \quad (1.15)$$

Especificando os valores de $\mathbf{\Lambda}_{k+1}$, $\boldsymbol{\theta}^{(k+1)}$ e $\boldsymbol{\lambda}_{1:k+1}$, obtemos:

$$\begin{bmatrix} \lambda_{yy}(0) & \lambda_{yy}(1) & \cdots & \lambda_{yy}(k-1) & \lambda_{yy}(k) \\ \lambda_{yy}(1) & \lambda_{yy}(0) & \cdots & \lambda_{yy}(k-2) & \lambda_{yy}(k-1) \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \lambda_{yy}(k-1) & \lambda_{yy}(k-2) & \cdots & \lambda_{yy}(2) & \lambda_{yy}(1) \\ \lambda_{yy}(k) & \lambda_{yy}(k-1) & \cdots & \lambda_{yy}(1) & \lambda_{yy}(0) \end{bmatrix} \begin{bmatrix} a_1^{(k+1)} \\ a_2^{(k+1)} \\ \vdots \\ a_k^{(k+1)} \\ a_{k+1}^{(k+1)} \end{bmatrix} = - \begin{bmatrix} \lambda_{yy}(1) \\ \lambda_{yy}(2) \\ \vdots \\ \lambda_{yy}(k) \\ \lambda_{yy}(k+1) \end{bmatrix} \quad (1.16)$$

As relações entre $\mathbf{\Lambda}_{k+1}$, e $\mathbf{\Lambda}_k$ e entre $\boldsymbol{\lambda}_{1:k+1}$ e $\boldsymbol{\lambda}_{1:k}$, são as seguintes:

$$\mathbf{\Lambda}_{k+1} = \left[\begin{array}{c|c} & \begin{matrix} \lambda_{yy}(k) \\ \lambda_{yy}(k-1) \\ \vdots \\ \lambda_{yy}(1) \end{matrix} \\ \hline \begin{matrix} \lambda_{yy}(k) & \lambda_{yy}(k-1) & \cdots & \lambda_{yy}(1) \end{matrix} & \lambda_{yy}(0) \end{array} \right] \quad (1.17)$$

$$\boldsymbol{\lambda}_{1:k+1} = \left[\begin{array}{c} \boldsymbol{\lambda}_{1:k} \\ \lambda_{yy}(k+1) \end{array} \right]. \quad (1.18)$$

Definindo $\boldsymbol{\lambda}_{k:1}$ como $\boldsymbol{\lambda}_{1:k}$ revertido:

$$\boldsymbol{\lambda}_{k:1} = [\lambda_{yy}(k) \quad \lambda_{yy}(k-1) \quad \cdots \quad \lambda_{yy}(1)]^T, \quad (1.19)$$

podemos compactar a relação entre $\mathbf{\Lambda}_{k+1}$ e $\mathbf{\Lambda}_k$ como:

$$\mathbf{\Lambda}_{k+1} = \left[\begin{array}{c|c} \mathbf{\Lambda}_k & \boldsymbol{\lambda}_{k:1} \\ \hline \boldsymbol{\lambda}_{k:1}^T & \lambda_{yy}(0) \end{array} \right]. \quad (1.20)$$

A representação de $\boldsymbol{\theta}^{(k+1)}$ em função de $\boldsymbol{\theta}^{(k)}$ tem a seguinte forma:

$$\boldsymbol{\theta}^{(k+1)} = \begin{bmatrix} \boldsymbol{\theta}^{(k)} \\ 0 \end{bmatrix} + \begin{bmatrix} \boldsymbol{\epsilon}_k \\ a_{k+1}^{(k+1)} \end{bmatrix}. \quad (1.21)$$

em que $\boldsymbol{\epsilon}_k \in \mathbb{R}^k$ é o termo de correcção e $a_{k+1}^{(k+1)} \in \mathbb{R}$ é o $k+1$ ésimo parâmetro do modelo de ordem $k+1$. Se na equação (1.15) substituirmos $\boldsymbol{\theta}^{(k+1)}$, $\mathbf{\Lambda}_{k+1}$ e $\boldsymbol{\lambda}_{1:k+1}$ pelos valores de (1.21), (1.20) e (1.18), temos:

$$\left[\begin{array}{c|c} \mathbf{\Lambda}_k & \boldsymbol{\lambda}_{k:1} \\ \hline \boldsymbol{\lambda}_{k:1}^T & \lambda_{yy}(0) \end{array} \right] \left(\begin{bmatrix} \boldsymbol{\theta}^{(k)} \\ 0 \end{bmatrix} + \begin{bmatrix} \boldsymbol{\epsilon}_k \\ a_{k+1}^{(k+1)} \end{bmatrix} \right) = - \begin{bmatrix} \boldsymbol{\lambda}_{1:k} \\ \lambda_{yy}(k+1) \end{bmatrix}. \quad (1.22)$$

Isolando as primeiras k equações:

$$\mathbf{\Lambda}_k \boldsymbol{\theta}^{(k)} + \mathbf{\Lambda}_k \boldsymbol{\epsilon}_k + \boldsymbol{\lambda}_{k:1} a_{k+1}^{(k+1)} = -\boldsymbol{\lambda}_{1:k} \quad (1.23)$$

e a última equação:

$$\boldsymbol{\lambda}_{k:1}^T \boldsymbol{\theta}^{(k)} + \boldsymbol{\lambda}_{k:1}^T \boldsymbol{\epsilon}_k + \lambda_{yy}(0) a_{k+1}^{(k+1)} = -\lambda_{yy}(k+1) \quad (1.24)$$

Substituindo $-\lambda_{1:k}$ pelo valor da equação (1.10) (com n substituído por k) na equação (1.23) temos:

$$\mathbf{\Lambda}_k \boldsymbol{\theta}^{(k)} + \mathbf{\Lambda}_k \boldsymbol{\epsilon}_k + \boldsymbol{\lambda}_{k:1} a_{k+1}^{(k+1)} = \mathbf{\Lambda}_k \boldsymbol{\theta}^{(k)} \quad (1.25)$$

de onde

$$\mathbf{\Lambda}_k \boldsymbol{\epsilon}_k = -\lambda_{k:1} a_{k+1}^{(k+1)} \quad (1.26)$$

Dado que $\hat{\mathbf{\Lambda}}_k$ é uma matriz de Toeplitz simétrica, ela não se altera se revertermos as suas linhas e colunas [6]. Consequentemente, as equações Yule Walker do modelo de ordem k invertidas e com as parcelas revertidas têm seguinte forma matricial:

$$\begin{bmatrix} \lambda_{yy}(0) & \lambda_{yy}(1) & \cdots & \lambda_{yy}(k-1) \\ \lambda_{yy}(1) & \lambda_{yy}(0) & \cdots & \lambda_{yy}(k-2) \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_{yy}(k-1) & \lambda_{yy}(k-2) & \cdots & \lambda_{yy}(0) \end{bmatrix} \begin{bmatrix} a_k^{(k)} \\ a_{k-1}^{(k)} \\ \vdots \\ a_1^{(k)} \end{bmatrix} = - \begin{bmatrix} \lambda_{yy}(k) \\ \lambda_{yy}(k-1) \\ \vdots \\ \lambda_{yy}(1) \end{bmatrix}. \quad (1.27)$$

Isto pode ser compactado na seguinte equação:

$$\mathbf{\Lambda}_k \boldsymbol{\theta}_R^{(k)} = -\lambda_{k:1} \quad (1.28)$$

onde

$$\boldsymbol{\theta}_R^{(k)} = \begin{bmatrix} a_k^{(k)} & a_{k-1}^{(k)} & \cdots & a_1^{(k)} \end{bmatrix}^T. \quad (1.29)$$

Substituindo este valor em (1.26) obtemos:

$$-\mathbf{\Lambda}_k \boldsymbol{\epsilon}_k = -\mathbf{\Lambda}_k \boldsymbol{\theta}_R^{(k)} a_{k+1}^{(k+1)} \quad (1.30)$$

o que resulta em:

$$\boldsymbol{\epsilon}_k = \boldsymbol{\theta}_R^{(k)} a_{k+1}^{(k+1)}. \quad (1.31)$$

Para completar a recursão é necessário determinar uma expressão para $a_{k+1}^{(k+1)}$. Isto pode ser feito multiplicando a equação (1.31) à esquerda por $\lambda_{k:1}^T$:

$$\lambda_{k:1}^T \boldsymbol{\epsilon}_k = \lambda_{k:1}^T \boldsymbol{\theta}_R^{(k)} a_{k+1}^{(k+1)} \quad (1.32)$$

Como $\boldsymbol{\theta}_R^{(k)}$ e $\lambda_{k:1}$ são os vectores revertidos de $\boldsymbol{\theta}^{(k)}$ e $\lambda_{1:k}$, respectivamente, temos

$$\lambda_{k:1}^T \boldsymbol{\theta}_R^{(k)} = \lambda_{1:k}^T \boldsymbol{\theta}^{(k)}. \quad (1.33)$$

Isto permite rescrever a equação (1.32) na forma:

$$\lambda_{k:1}^T \boldsymbol{\epsilon}_k = \lambda_{1:k}^T \boldsymbol{\theta}^{(k)} a_{k+1}^{(k+1)}. \quad (1.34)$$

Podemos então obter $a_{k+1}^{(k+1)}$ substituindo e $\lambda_{1:k}^T \boldsymbol{\epsilon}_k$ por este valor na equação (1.24):

$$\lambda_{k:1}^T \boldsymbol{\theta}^{(k)} + \lambda_{1:k}^T \boldsymbol{\theta}^{(k)} a_{k+1}^{(k+1)} + \lambda_{yy}(0) a_{k+1}^{(k+1)} = -\lambda_{yy}(k+1) \quad (1.35)$$

o que resulta na seguinte expressão para $\hat{a}_{k+1}^{(k+1)}$:

$$a_{k+1}^{(k+1)} = -\frac{\lambda_{yy}(k+1) + \lambda_{k:1}^T \boldsymbol{\theta}^{(k)}}{\lambda_{yy}(0) + \lambda_{1:k}^T \boldsymbol{\theta}^{(k)}}. \quad (1.36)$$

As equações (1.22), (1.31) e (1.36) definem a recursão necessária à determinação dos parâmetros do modelo de ordem $k+1$ a partir dos parâmetros do modelo de ordem k . Este processo pode ser sumarizado no seguinte algoritmo:

Algoritmo de Levinson-Durbin

1. Inicialização:

$$\theta_1 = a_1^{(1)} = -\frac{\lambda_{yy}(1)}{\lambda_{yy}(0)}.$$

2. Recursão Para $k = 1$ até $n - 1$

2.1. Calcular:

$$a_{k+1}^{(k+1)} = -\frac{\lambda_{yy}(k+1) + \lambda_{k:1}^T \theta^{(k)}}{\lambda_{yy}(0) + \lambda_{1:k}^T \theta^{(k)}}.$$

2.2. Determinar a correcção:

$$\epsilon_k = a_{k+1}^{(k+1)} \theta_R^{(k)}.$$

2.3. Actualizar o vector de parâmetros:

$$\theta^{(k+1)} = \begin{bmatrix} \theta^{(k)} \\ 0 \end{bmatrix} + \begin{bmatrix} \epsilon_k \\ a_{k+1}^{(k+1)} \end{bmatrix} = \begin{bmatrix} a_1^{(k)} \\ a_2^{(k)} \\ \vdots \\ a_k^{(k)} \\ 0 \end{bmatrix} + a_{k+1}^{(k+1)} \begin{bmatrix} a_k^{(k)} \\ a_{k-1}^{(k)} \\ \vdots \\ a_1^{(k)} \\ 1 \end{bmatrix}.$$

Estimador do erro de previsão

Os algoritmos de erro de previsão são ferramentas fundamentais na estimação de modelos paramétricos de processos estocástico. Estes algoritmos minimizam a diferença entre os valores observados e os gerados pelo previsor de um passo do modelo, ajustando os parâmetros de forma a melhorar a correspondência entre o modelo e os dados observados.

Uma das principais propriedades destes algoritmos é a consistência das estimativas. A consistência refere-se à capacidade do algoritmo em fornecer estimativas de parâmetros que convergem para os valores verdadeiros à medida que o número de observações aumenta. Isto significa que, com um conjunto de dados suficientemente grande, as estimativas obtidas pelos algoritmos de erro de previsão aproximam-se, com probabilidade 1, dos parâmetros reais do processo subjacente. Esta característica é crucial em aplicações onde a precisão da estimativa influencia a qualidade das previsões futuras e a compreensão do comportamento do processo.

Outra propriedade relevante é a optimalidade em processos gaussianos. Em particular, quando o processo estocástico é gaussiano, os algoritmos de erro de previsão proporcionam estimativas que minimizam o erro médio quadrático dos parâmetros forma ótima. Em contextos onde o ruído ou as incertezas no sistema seguem uma distribuição normal, os algoritmos de erro de previsão tornam-se uma escolha preferencial, garantindo estimativas precisas e robustas [9]. Vimos, anteriormente, que o previsor a um passo de um processo AR é dado por:

$$\hat{y}(t, a_1, \dots, a_n) = [1 - A(q^{-1})] y(t) = -a_1 y(t-1) - \dots - a_n y(t-n), \quad (1.37)$$

sendo uma combinação linear dos parâmetros do modelo, $a_1, \dots, a_n$. Assim, o erro de previsão é dado por:

$$\hat{e}(t, a_1, \dots, a_n) = y(t) - \hat{y}(t, a_1, \dots, a_n) = y(t) + a_1 y(t-1) - \dots + y(t-n) \quad (1.38)$$

o que implica que também é uma função linear dos parâmetros. Isso significa que o critério quadrático

$$\begin{aligned} V(a_1, \dots, a_n) &= \frac{1}{2} \sum_{t=1}^N \hat{e}^2(t, a_1, \dots, a_n) = \\ &= \frac{1}{2} \sum_{t=1}^N [y(t) + a_1 y(t-1) + \dots + a_n y(t-n)]^2 \end{aligned} \quad (1.39)$$

é uma função quadrática de $a_1, \dots, a_n$ e que, conseqüentemente, tem um único mínimo. Este critério pode ser minimizado utilizando o bem conhecido algoritmo de mínimos quadrados que consiste na resolução do sistema de equações normais:

$$\Phi^T \Phi \theta = \Phi^T Y \quad (1.40)$$

onde:

$$\theta = [a_1 \ \dots \ a_n]^T \in \mathbb{R}^n \quad (1.41)$$

$$\Phi = \begin{bmatrix} -y(0) & -y(-1) & \dots & -y(1-n) \\ -y(1) & -y(0) & \dots & -y(2-n) \\ \vdots & \vdots & \ddots & \vdots \\ -y(t-1) & -y(t-2) & \dots & -y(t-n) \\ \vdots & \vdots & \ddots & \vdots \\ -y(N-1) & -y(N-2) & \dots & -y(N-n) \end{bmatrix} \in \mathbb{R}^{N \times n} \quad (1.42)$$

$$Y = [y(1) \ y(2) \ \dots \ y(N)] \in \mathbb{R}^N. \quad (1.43)$$

Se a matriz $\Phi^T \Phi$ for não singular, a solução das equações normais é dada por:

$$\hat{\theta} = (\Phi^T \Phi)^{-1} \Phi^T Y = \Phi^\dagger Y, \quad (1.44)$$

onde $\Phi^\dagger$ é o pseudo-inverso de Φ . Caso os valores de $y(0), \dots, y(n-1)$ não estejam disponíveis, podemos atribuir-lhes o valor zero, embora isso possa enviesar as estimativas do modelo, principalmente quando o número de observações N é pequeno. Se não quisermos que isso aconteça, podemos remover estes valores de Φ e Y , redefinindo estas variáveis da seguinte forma:

$$\Phi = \begin{bmatrix} -y(n) & -y(n-1) & \dots & -y(1) \\ -y(n+1) & -y(n) & \dots & -y(2) \\ \vdots & \vdots & \ddots & \vdots \\ -y(t-1) & -y(t-2) & \dots & -y(t-n) \\ \vdots & \vdots & \ddots & \vdots \\ -y(N-1) & -y(N-2) & \dots & -y(N-n) \end{bmatrix} \in \mathbb{R}^{(N-n) \times n} \quad (1.45)$$

$$Y = [y(n+1) \ y(n+2) \ \dots \ y(N)] \in \mathbb{R}^{N-n}. \quad (1.46)$$

Esta reformulação permite lidar com a ausência de observações iniciais, evitando enviesar a estimação dos parâmetros, especialmente em situações em que o número de amostras disponíveis é limitado.

Relação entre os estimadores de Yule-Walker e de erro de previsão

A equação (1.44) pode ser reescrita na seguinte forma

$$\hat{\theta} = \left(\frac{1}{N-n} \Phi^T \Phi \right)^{-1} \frac{1}{N-n} \Phi^T \mathbf{Y}. \quad (1.47)$$

Como

$$\begin{aligned} & \frac{1}{N-n} \Phi^T \Phi = \\ &= \frac{1}{N-n} \left(\begin{bmatrix} -y(n) & -y(n+1) & \cdots & -y(N-1) \\ -y(n-1) & -y(n) & \cdots & -y(N-2) \\ \vdots & \vdots & \ddots & \vdots \\ -y(1) & -y(2) & \vdots & -y(N-n) \end{bmatrix} \begin{bmatrix} -y(n) & -y(n-1) & \cdots & -y(1) \\ -y(n+1) & -y(n) & \cdots & -y(2) \\ \vdots & \vdots & \ddots & \vdots \\ -y(N-1) & -y(N-2) & \cdots & -y(N-n) \end{bmatrix} \right) = \\ &= \begin{bmatrix} \frac{1}{N-n} \sum_{t=n}^{N-1} y^2(t) & \frac{1}{N-n} \sum_{t=n}^{N-1} y(t)y(t-1) & \cdots & \frac{1}{N-n} \sum_{t=n}^{N-1} y(t)y(t-n+1) \\ \frac{1}{N-n} \sum_{t=n-1}^{N-2} y(t)y(t+1) & \frac{1}{N-n} \sum_{t=n-1}^{N-2} y^2(t) & \cdots & \frac{1}{N-n} \sum_{t=n-1}^{N-2} y(t)y(t-n+2) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{1}{N-n} \sum_{t=1}^{N-n} y(t)y(t+n-1) & \frac{1}{N-n} \sum_{t=1}^{N-n} y(t)y(t+n-2) & \cdots & \frac{1}{N-n} \sum_{t=1}^{N-n} y^2(t) \end{bmatrix} \end{aligned} \quad (1.48)$$

e

$$\begin{aligned} \frac{1}{N-n} \Phi^T \mathbf{Y} &= \frac{1}{N-n} \begin{bmatrix} -y(n) & -y(n+1) & \cdots & -y(N-1) \\ -y(n-1) & -y(n) & \cdots & -y(N-2) \\ \vdots & \vdots & \ddots & \vdots \\ -y(1) & -y(2) & \vdots & -y(N-n) \end{bmatrix} \begin{bmatrix} y(n+1) \\ y(n+2) \\ \vdots \\ y(N) \end{bmatrix} \\ &= \begin{bmatrix} -\frac{1}{N-n} \sum_{t=n}^{N-1} y(t)y(t+1) \\ -\frac{1}{N-n} \sum_{t=n-1}^{N-2} y(t)y(t+2) \\ \vdots \\ -\frac{1}{N-n} \sum_{t=1}^{N-n} y(t)y(t+n) \end{bmatrix}, \end{aligned} \quad (1.49)$$

Quando $N \rightarrow \infty$

$$\lim_{N \rightarrow \infty} \hat{\theta} = - \begin{bmatrix} \lambda_{yy}(0) & \lambda_{yy}(1) & \cdots & \lambda_{yy}(n-1) \\ \lambda_{yy}(1) & \lambda_{yy}(0) & \cdots & \lambda_{yy}(n-2) \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_{yy}(n-1) & \lambda_{yy}(n-2) & \cdots & \lambda_{yy}(0) \end{bmatrix}^{-1} \begin{bmatrix} \lambda_{yy}(1) \\ \lambda_{yy}(2) \\ \vdots \\ \lambda_{yy}(n) \end{bmatrix}. \quad (1.50)$$

Podemos constatar que esta equação é igual à equação (1.11), de onde concluímos que, tal como as estimativas que resultam da resolução do sistema de equações de Yule-Walker, as estimativas de mínimos quadrados convergem para o valor verdadeiro dos parâmetros com probabilidade 1, sendo isto representado por:

$$\hat{\theta} \xrightarrow{prob; 1} \theta = [a_1 \quad \cdots \quad a_n]. \quad (1.51)$$

Diz-se, por isso que o estimador de erro de previsão é fortemente consistente. Dada a equivalência assintótica com o estimador de Yule-Walker podemos concluir que o estimador de Yule-Walker também é consistente.

1.2.2 Determinação da ordem de modelos AR - Autocorrelação parcial

Vimos que um processo AR $y(t)$ é modelado como uma combinação linear de cópias de $y(t)$ atrasadas no tempo. Para determinar a ordem do modelo temos que avaliar quais as cópias é que influenciam $y(t)$. Podemos fazer isso através da função de autocorrelação parcial. A autocorrelação parcial entre $y(t)$ e $y(t-k)$ é o coeficiente de autocorrelação das componentes destes sinais que não estão correlacionadas com as cópias com atrasos intermédios, isto é, com $y(t-1)$, $y(t-2)$, $\dots$, $y(t-k+1)$.

Como não há nenhuma versão de $y(t)$ com atraso entre 0 e 1, e, por outro lado, $E\{y(t)\} = E\{y(t-1)\} = 0$, então a autocorrelação parcial entre $y(t)$ e $y(t-1)$ é o coeficiente de autocorrelação:

$$\phi_{yy}(1) = \rho_{yy}(1) = \frac{E\{y(t)y(t-1)\}}{\sqrt{E\{y^2(t)\}E\{y^2(t-1)\}}} = \frac{\Lambda_{yy}(1)}{\lambda_{yy}(0)}. \quad (1.52)$$

Entre $y(t)$ e $y(t-2)$ existe $y(t-1)$. Assim, para calcular a correlação parcial entre $y(t)$ e $y(t-2)$, temos que, em primeiro lugar, determinar as componentes de $y(t)$ e de $y(t-2)$ que não estão correlacionadas com $y(t-1)$, isto é, temos que decompor $y(t)$ e $y(t-2)$ da seguinte forma:

$$y(t) = y(t-1)\hat{\alpha}_1^{(1)} + e_0^{(1)}(t) \quad (1.53)$$

$$y(t-2) = y(t-1)\hat{\beta}_1^{(1)} + e_2^{(1)}(t), \quad (1.54)$$

em que $e_0^{(1)}(t)$ e $e_2^{(1)}(t)$ não são correlacionados com $y(t-1)$. A autocorrelação parcial entre $y(t)$ e $y(t-2)$ é o coeficiente de correlação entre $e_0^{(1)}(t)$ e $e_2^{(1)}(t)$:

$$\phi_{yy}(2) = \frac{E\{e_0^{(1)}(t)e_2^{(1)}(t)\}}{\sqrt{E\{e_0^{(1)2}(t)\}E\{e_2^{(1)2}(t)\}}}. \quad (1.55)$$

A correlação parcial de $y(t)$ com $y(t - k)$ tem uma definição idêntica à de $y(t)$ com $y(t - 2)$ só que aqui existem $k - 1$ cópias de $y(t)$ com atraso intermédio: $y(t - 1), \dots, y(t - k + 1)$. Temos então:

$$y(t) = [y(t - 1) \ \dots \ y(t - k + 1)] \hat{\alpha}^{(k-1)} + e_0^{(k-1)}(t) \quad (1.56)$$

$$y(t - k) = [y(t - 1) \ \dots \ y(t - k + 1)] \hat{\beta}^{(k-1)} + e_k^{(k-1)}(t) \quad (1.57)$$

com

$$\hat{\alpha}^{(k-1)} = [\hat{\alpha}_1^{(k-1)} \ \dots \ \hat{\alpha}_{k-1}^{(k-1)}]^T \quad (1.58)$$

$$\hat{\beta}^{(k-1)} = [\hat{\beta}_1^{(k-1)} \ \dots \ \hat{\beta}_{k-1}^{(k-1)}]^T \quad (1.59)$$

onde $e_0^{(k-1)}(t)$ e $e_{k-1}^{(k)}(t)$ não são correlacionados com $[y(t - 1) \ \dots \ y(t - k - 1)]$. O coeficiente de correlação entre $e_0^{(k-1)}(t)$ e $e_k^{(k-2)}(t)$ corresponde à correlação parcial entre $y(t)$ e $y(t - k)$:

$$\phi_{yy}(k) = \frac{\mathbf{E} \left\{ e_0^{(k-1)}(t) e_k^{(k-1)}(t) \right\}}{\sqrt{\mathbf{E} \left\{ e_0^{(k-1)}(t)^2 \right\} \mathbf{E} \left\{ e_k^{(k-1)}(t)^2 \right\}}}. \quad (1.60)$$

Os valores de $e_0^{(k-1)}$ e $e_k^{(k-1)}$, são os resíduos dos estimadores de mínimos quadrados dos parâmetros $\alpha_i^{(k-1)}$ e $\beta_i^{(k-1)}$, $i = 1, \dots, n$. Consequentemente, para estimar uma sequência de n coeficientes de correlação parcial são necessárias $2n$ regressões lineares, o que pode ser computacionalmente pesado. Estas regressões são da forma:

$$\hat{\alpha}^{(k-1)} = \left(\frac{1}{N - k - 1} X_f^{(k-1)T} X_f^{(k-1)} \right)^{-1} \frac{1}{N - k - 1} X_f^{(k-1)T} Y_f^{(k-1)} \quad (1.61)$$

e

$$\hat{\beta}^{(k-1)} = \left(\frac{1}{N - k - 1} X_b^{(k-1)T} X_b^{(k-1)} \right)^{-1} \frac{1}{N - k - 1} X_b^{(k-1)T} Y_b^{(k-1)} \quad (1.62)$$

onde

$$X_f^{(k-1)} = \begin{bmatrix} y(k-1) & y(k-2) & \dots & y(1) \\ y(k) & y(k-1) & \dots & y(2) \\ \vdots & \vdots & \ddots & \vdots \\ y(N) & y(N-1) & \dots & y(N-k+1) \end{bmatrix} \quad (1.63)$$

$$Y_f^{(k-1)} = [y(k) \ y(k+1) \ \dots \ y(N)] \quad (1.64)$$

$$X_b^{(k-1)} = \begin{bmatrix} y(N) & y(N-1) & \dots & y(N-k+2) \\ y(N-1) & y(N-2) & \dots & y(N-k+1) \\ \vdots & \vdots & \ddots & \vdots \\ y(k) & y(k-1) & \dots & y(2) \end{bmatrix} \quad (1.65)$$

$$Y_b^{(k-1)} = [y(N-k+1) \ y(N-k) \ \dots \ y(1)]^T. \quad (1.66)$$

Quando $N \rightarrow \infty$,

$$\begin{aligned}
\lim_{N \rightarrow \infty} \hat{\alpha}^{(k-1)} &= \begin{bmatrix} \lambda_{yy}(0) & \lambda_{yy}(1) & \cdots & \lambda_{yy}(k-2) \\ \lambda_{yy}(1) & \lambda_{yy}(0) & \cdots & \lambda_{yy}(k-3) \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_{yy}(k-2) & \lambda_{yy}(k-3) & \cdots & \lambda_{yy}(0) \end{bmatrix}^{-1} \begin{bmatrix} \lambda_{yy}(1) \\ \lambda_{yy}(2) \\ \vdots \\ \lambda_{yy}(k-1) \end{bmatrix} = \\
&= \mathbf{\Lambda}_{k-1}^{-1} \boldsymbol{\lambda}_{1:k-1} = \boldsymbol{\alpha}^{(k-1)}, \\
\lim_{N \rightarrow \infty} \hat{\beta}^{(k-1)} &= \begin{bmatrix} \lambda_{yy}(0) & \lambda_{yy}(1) & \cdots & \lambda_{yy}(k-2) \\ \lambda_{yy}(1) & \lambda_{yy}(0) & \cdots & \lambda_{yy}(k-3) \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_{yy}(k-2) & \lambda_{yy}(k-3) & \cdots & \lambda_{yy}(0) \end{bmatrix}^{-1} \begin{bmatrix} \lambda_{yy}(k-1) \\ \lambda_{yy}(k-2) \\ \vdots \\ \lambda_{yy}(1) \end{bmatrix} = \\
&= \mathbf{\Lambda}_{k-1}^{-1} \boldsymbol{\lambda}_{k-1:1} = \boldsymbol{\beta}^{(k-1)}.
\end{aligned}$$

Podemos ver, nestas equações, que

$$\boldsymbol{\beta}^{(k-1)} = \boldsymbol{\alpha}_R^{(k-1)} = \begin{bmatrix} \alpha_k^{(k-1)} \\ \alpha^{k-1}(k) \\ \vdots \\ \alpha^{(k-1)} \end{bmatrix} \quad (1.67)$$

Também podemos ver que as regressões são assintoticamente equivalentes à resolução dum sistema de equações de Yule-Walker. Consequentemente, os parâmetros modelos de regressão de ordens $k = 1, \dots, \infty$ também são os da recursão de Levinson-Durbin. A seguir demonstra-se que $\alpha_k^{(k)}$ é o coeficiente de correlação parcial $\Phi_{yy}(k)$

Lema 1 (Coeficiente de Autocorrelação Parcial). *Seja $y(t)$ um processo estocástico estacionário de média nula, $\Phi_{yy}(k)$ o Coeficiente de Autocorrelação Parcial entre $y(t)$ e $y(t-k)$ definido na equação (1.60) e $\boldsymbol{\alpha}^{(k)} = [\alpha_1^{(k)} \cdots \alpha_k^{(k)}]^T$ o vector de parâmetros do modelo $AR(k)$ em (1.56). Então*

$$\Phi_{yy}(k) = \alpha_k^{(k)}. \quad (1.68)$$

Demonstração. Utilizando o operador de atraso, a equação (1.56) que descreve o modelo de ordem $k-1$ pode ser reescrita na forma:

$$y(t) = \alpha_1^{(k-1)} q^{-1} y(t) + \cdots + \alpha_{k-1}^{(k-1)} y(t) + e_0^{(k-1)}(t) \Leftrightarrow \quad (1.69)$$

$$\Leftrightarrow \left(1 - \alpha_1^{(k-1)} q^{-1} - \cdots - \alpha_{k-1}^{(k-1)} q^{-k+1}\right) y(t) = e_0^{(k-1)}(t) \quad (1.70)$$

Dado que este é um modelo de regressão, então

$$1 - \alpha_1^{(k-1)} q^{-1} - \cdots - \alpha_{k-1}^{(k-1)} q^{-k+1} = 1 + a_1^{(k-1)} q^{-1} + \cdots + a_{k-1}^{(k-1)} q^{-k+1} = A^{(k-1)}(q^{-1}) \quad (1.71)$$

em que $A^{(k-1)}(q^{-1})$ é o polinómio que descreve o modelo de regressão de ordem $k-1$ na recursão de Levinson (1.13). Temos, então, $\alpha_i^{(k)} = -a_i^{(k)}$, $i = 1, \dots, k$.

Consequentemente, $\theta^{(k)} = -\alpha^{(k)}$, para $k = 1, \dots, \infty$. Substituindo $a_k^{(k)}$ por $-\alpha_k^{(k)}$ e $\theta^{(k)}$ por $-\alpha^{(k)}$ na equação (1.36) para o modelo de ordem k obtemos:

$$\alpha_k^{(k)} = \frac{\lambda_{yy}(k) - \lambda_{k-1:1}^T \alpha^{(k-1)}}{\lambda_{yy}(0) - \lambda_{1:k-1}^T \alpha^{(k-1)}}. \quad (1.72)$$

Das equações (1.56), (1.57) e (1.67)

$$e_0^{(k-1)}(t) = y(t) - [y(t-1) \ \cdots \ y(t-k+1)] \alpha^{(k-1)} \quad (1.73)$$

$$e_k^{(k-1)}(t) = y(t-k) - [y(t-1) \ \cdots \ y(t-k+1)] \alpha_R^{(k-1)} \quad (1.74)$$

Como $e_0^{(k-1)}(t)$ e $e_k^{(k-1)}(t)$ não são correlacionados com $[y(t-1) \ \cdots \ y(t-k+1)]$, então

$$\begin{aligned} & \mathbf{E} \left\{ e_0^{(k-1)}(t) e_k^{(k-1)}(t) \right\} = \\ &= \mathbf{E} \left\{ e_0^{(k-1)}(t) \left(y(t-k) - [y(t-1) \ \cdots \ y(t-k+1)] \alpha_R^{(k-1)} \right) \right\} = \\ &= \mathbf{E} \left\{ e_0^{(k-1)}(t) y(t-k) \right\} = \\ &= \mathbf{E} \left\{ \left(y(t) - [y(t-1) \ \cdots \ y(t-k+1)] \alpha^{(k-1)} \right) y(t-k) \right\} = \\ &= \lambda_{yy}(k) - [\lambda_{yy}(k-1) \ \cdots \ \lambda_{yy}(1)] \alpha^{(k-1)} = \\ &= \lambda_{yy}(k) - \lambda_{k-1:1}^T \alpha^{(k-1)}, \end{aligned} \quad (1.75)$$

$$\begin{aligned} & \mathbf{E} \left\{ e_0^{(k-1)}(t)^2 \right\} = \\ & \mathbf{E} \left\{ e_0^{(k-1)}(t) \left(y(t) - [y(t-1) \ \cdots \ y(t-k+1)] \alpha^{(k-1)} \right) \right\} = \\ &= \mathbf{E} \left\{ e_0^{(k-1)}(t) y(t) \right\} = \\ &= \mathbf{E} \left\{ \left(y(t) - [y(t-1) \ \cdots \ y(t-k+1)] \alpha^{(k-1)} \right) y(t) \right\} = \\ &= \lambda_{yy}(0) - [\lambda_{yy}(1) \ \cdots \ \lambda_{yy}(k-1)] \alpha^{(k-1)} = \\ &= \lambda_{yy}(0) - \lambda_{1:k-1}^T \alpha^{(k-1)} \end{aligned} \quad (1.76)$$

e

$$\begin{aligned} & \mathbf{E} \left\{ e_k^{(k-1)}(t)^2 \right\} = \\ & \mathbf{E} \left\{ e_k^{(k-1)}(t) \left(y(t-k) - [y(t-1) \ \cdots \ y(t-k+1)] \alpha_R^{(k-1)} \right) \right\} = \\ &= \mathbf{E} \left\{ e_k^{(k-1)}(t) y(t-k) \right\} = \\ &= \mathbf{E} \left\{ \left(y(t-k) - [y(t-1) \ \cdots \ y(t-k+1)] \alpha_R^{(k-1)} \right) y(t-k) \right\} = \\ &= \lambda_{yy}(0) - [\lambda_{yy}(k-1) \ \cdots \ \lambda_{yy}(1)] \alpha_R^{(k-1)} = \\ &= \lambda_{yy}(0) - \lambda_{k-1:1} \alpha_R^{(k-1)} = \\ &= \lambda_{yy}(0) - \lambda_{1:k}^T \alpha^{(k-1)}, \end{aligned} \quad (1.77)$$

pois $\lambda_{k-1:1}^T \alpha_R^{(k-1)} = \lambda_{1:k}^T \alpha^{(k-1)}$ Temos, então,

$$\phi_{yy}(k) = \frac{\mathbf{E} \left\{ e_0^{(k-1)}(t) e_k^{(k-1)}(t) \right\}}{\sqrt{\mathbf{E} \left\{ e_0^{(k-1)}(t)^2 \right\} \mathbf{E} \left\{ e_k^{(k-1)}(t)^2 \right\}}} = \frac{\lambda_{yy}(k) - \lambda_{k-1:1}^T \alpha^{(k-1)}}{\lambda_{yy}(0) - \lambda_{1:k}^T \alpha^{(k-1)}} = \alpha_k^{(k)}. \quad (1.78)$$

□

Como se referiu anteriormente, a ordem dum processo AR pode ser determinada através dos coeficientes de autocorrelação parcial. O procedimento baseia-se na observação do comportamento desses coeficientes à medida que se aumenta esfasamento, ou seja, a distância temporal entre as observações. Num processo AR(n), os primeiros n coeficientes de autocorrelação parcial não são zero, enquanto os coeficientes para esfasamentos superiores a n são nulos. A ordem do modelo é sinalizada pelo último esfasamento em que o coeficiente de correlação parcial estimado, $\hat{\phi}_{yy}(n)$, é significativamente diferente de zero. A partir de $n + 1$, os coeficientes devem ser aproximadamente zero, indicando que a dependência temporal se esgota na ordem n . Para determinar se os coeficientes de autocorrelação parcial são significativamente diferentes de zero, é utilizado um intervalo de confiança com base numa distribuição normal. Sob a hipótese nula de que o coeficiente de autocorrelação parcial é zero, a variância do coeficiente estimado é aproximadamente $\frac{1}{N}$, onde N é o tamanho da amostra. Assim, os limites dum intervalo com 95% de confiança são $\pm \frac{1.96}{N}$.

Para determinar a ordem dum modelo AR podemos utilizar um gráfico com os coeficientes de autocorrelação parcial. Para um modelo AR(n) espera-se que os primeiros n elementos sejam significativos (estejam fora do intervalo de confiança) e que os seguintes caiam dentro dos limites, sendo aproximadamente zero (ver Figura 1.1).

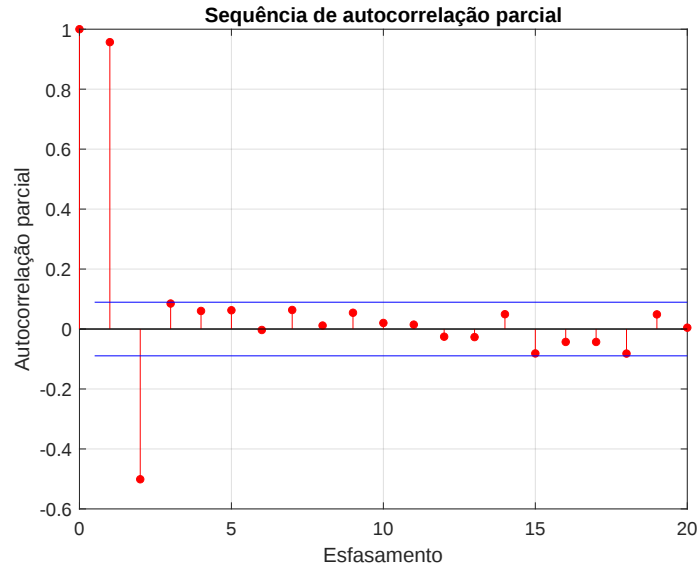


Figura 1.1: Coeficientes de autocorrelação parcial estimados a partir duma realização dum processo AR(2)

1.2.3 Modelos MA

Os processos MA (Média Móvel) podem ser representados pela equação:

$$y(t) = C(q^{-1})e(t) \quad (1.79)$$

onde

$$C(q^{-1}) = 1 + c_1 q^{-1} + \dots + c_n q^{-n}, \quad (1.80)$$

e $e(t)$ é ruído branco de média nula e variância σ^2 . A densidade espectral de potência é dada por:

$$\Phi_{yy}(\omega) = |C(e^{-j\omega})|^2 \sigma^2 = |1 + c_1 e^{-j\omega} + \dots + c_n e^{-j\omega n}|^2 \sigma^2. \quad (1.81)$$

Assim, a densidade espectral pode ser obtida através da estimação dos parâmetros de $C(q^{-1})$.

Método da função geradora da autocovariância

Consideremos o seguinte processo MA de ordem 1:

$$y(t) = e(t) + c_1 e(t-1). \quad (1.82)$$

Os parâmetros deste processo são c_1 , o coeficiente do polinômio $C(q^{-1})$ e σ^2 , a variância de $e(t)$. A sequência de de covariância deste processo é obtida pelas equações de Yule-Walker[1] que, neste caso, são:

$$\lambda_{yy}(0) = (1 + c_1^2) \sigma^2 \quad (1.83)$$

$$\lambda_{yy}(1) = c_1 \sigma^2 \quad (1.84)$$

$$\lambda_{yy}(-1) = \lambda_{yy}(1) \quad (1.85)$$

$$\lambda_{yy}(\tau) = 0, \quad |\tau| > 1. \quad (1.86)$$

Para resolver este sistema de equações, observemos que $y(t)$ pode ser interpretado como a saída dum sistema linear, estável e invariante no tempo, cuja a função de transferência é:

$$H(z) = 1 + c_1 z^{-1}, \quad (1.87)$$

e que é excitado por ruído branco com variância σ^2 . Dado que a densidade espectral de $e(t)$ é $\Phi_{ee}(e^{j\omega}) = \sigma^2$, a densidade espectral de $y(t)$, que é a transformada de Fourier da função da função de auto-covariância de $y(t)$, será [9, 7]

$$\Phi_{yy}(e^{j\omega}) = H(e^{j\omega}) H(e^{-j\omega}) \sigma^2 = \mathcal{F}\{\lambda_{yy}(\tau)\} = \sum_{\tau=-\infty}^{\infty} \lambda_{yy}(\tau) e^{-j\omega\tau}, \quad (1.88)$$

o que resulta em:

$$\Phi_{yy}(e^{j\omega}) = \lambda_{yy}(-1) e^{j\omega} + \lambda_{yy}(0) + \lambda_{yy}(1) e^{-j\omega}. \quad (1.89)$$

Substituindo $z = e^{j\omega}$, obtemos a função geradora de autocovariância $\Phi_{yy}(z)$, que é a transformada z de $\lambda_{yy}(\tau)$ [3, 10, 5]:

$$\Phi_{yy}(z) = H(z) H(z^{-1}) \sigma^2 = \lambda_{yy}(-1) z + \lambda_{yy}(0) + \lambda_{yy}(1) z^{-1}. \quad (1.90)$$

Como a sequência de autocovariância é simétrica, temos $\lambda_{yy}(-1) = \lambda_{yy}(1)$, o que permite rescrever $\Phi_{yy}(z)$ na forma

$$\Phi_{zz}(z) = \lambda_{yy}(1) z + \lambda_{yy}(0) + \lambda_{yy}(1) z^{-1}. \quad (1.91)$$

A partir de (1.90), podemos vermos que os zeros de $\Phi_{yy}(z)$ são o zero de $H(z)$ e o seu inverso. Assim:

$$\Phi_{yy}(z) = (z - z_1)(z - z_1^{-1})z^{-1}\sigma^2 \quad (1.92)$$

onde z_1 é o zero de $H(z)$, i.e., $z_1 = -c_1$. Como $H(z)$ é um polinómio mónico inversamente estável, a função de transferência pode ser expressa como:

$$H(z) = (z - z_1)z^{-1} = 1 - z_1z^{-1} \quad (1.93)$$

em que z_1 é o zero de $\Phi_{yy}(z)$ no interior do círculo unitário. A partir da sequência de autocorrelação de $y(t)$, podemos determinar $H(z)$ com o seguinte procedimento [1, 3, 10]:

1. Determinar o zero de $\Phi_{yy}(z)$ no interior do círculo unitário:

$$\lambda_{yy}(1)z^2 + \lambda_{yy}(0)z + \lambda_{yy}(1) = 0 \quad (1.94)$$

Resolvendo para z_1 :

$$z_1 = \begin{cases} -\frac{\lambda_{yy}(0)}{2\lambda_{yy}(1)} + \frac{\sqrt{\left(\frac{\lambda_{yy}(0)}{\lambda_{yy}(1)}\right)^2 - 4}}{2}, & \lambda_{yy}(1) > 0 \\ -\frac{\lambda_{yy}(0)}{2\lambda_{yy}(1)} - \frac{\sqrt{\left(\frac{\lambda_{yy}(0)}{2\lambda_{yy}(1)}\right)^2 - 4}}{2}, & \lambda_{yy}(1) < 0 \end{cases} \quad (1.95)$$

Expressando a equação de z_1 em termos do coeficiente de correlação $\rho_{yy}(1) = \frac{\lambda_{yy}(1)}{\lambda_{yy}(0)}$:

$$z_1 = -\frac{1 - \sqrt{1 - 4\rho_{yy}^2(1)}}{2\rho_{yy}(1)} \quad (1.96)$$

Dado que $z_1 = -c_1$, obtemos:

$$c_1 = \frac{1 - \sqrt{1 - 4\rho_{yy}^2(1)}}{2\rho_{yy}(1)}. \quad (1.97)$$

2. Uma vez determinado c_1 , podemos resolver (1.83) para encontrar σ^2 :

$$\sigma^2 = \frac{\lambda_{yy}(0)}{1 + c_1^2}. \quad (1.98)$$

Observemos que, sendo c_1 um coeficiente real, então, de (1.97),

$$1 - 4\rho_{yy}^2(1) \leq 0 \Rightarrow |\rho_{yy}(1)| \leq \frac{1}{2} \quad (1.99)$$

Dado que:

$$\rho_{yy}(1) = \frac{\lambda_{yy}(1)}{\lambda_{yy}(0)} = \frac{c_1}{1 + c_1^2}, \quad (1.100)$$

temos que $\max \{|\rho_{yy}(1)|\}$ é $\frac{1}{2}$, atingido quando $|c_1| = 1$, ou seja, quando o zero de $H(z)$ está na circunferência unitária, tornando a função de transferência não seja invertível. Contudo, quando a sequência de correlação é estimada a partir de uma realização de $y(t)$, e $|c_1|$ está próximo de 1, é natural que o ruído faça com que a estimativa de $\rho_{yy}(1)$ tome valores com módulo superior a 0.5. Nestas condições, a transformada de z da função de autocovariância, calculada com as estimativas, resulta num polinómio com um par de zeros complexos:

$$\hat{z}_{1,2} = \frac{1}{2\hat{r}_{yy}(1)} \pm j \frac{\sqrt{4\hat{r}_{yy}^2(1) - 1}}{2\hat{r}_{yy}(1)} \quad (1.101)$$

Uma vez que:

$$|\hat{z}_{1,2}| = \frac{1 + 4\hat{r}_{yy}^2(1) - 1}{4\hat{r}_{yy}^2(1)} = 1, \quad (1.102)$$

estes zeros estão na circunferência unitária. Assim, sempre que a transformada z da estimativa da função de autocovariância apresenta um par de zeros complexos, esse par de zeros complexos está na circunferência unitária, correspondendo a um processo MA(1) com o zero próximo dessa circunferência. Portanto, a melhor estimativa do processo, será um modelo MA(1) com $c_1 = 1$ se $\rho_{yy}(1) > 0$ e $c_1 = -1$ se $\rho_{yy}(1) < 0$ [3, 10].

A estimação dos parâmetros dum processo MA(n), descrito por:

$$y(t) = e(t) + c_1 e(t-1) + \dots + c_n e(t-n) \quad (1.103)$$

pode ser realizada de forma semelhante à de um processo MA(1). Primeiramente, determinam-se os zeros da função geradora de autocovariância, resolvendo a seguinte equação:

$$\Lambda_{yy}(n)z^{2n} + \dots + \Lambda_{yy}(1)z^{n+1} + \Lambda_{yy}(0)z^n + \Lambda_{yy}(1)z^{n-1} + \dots + \Lambda_{yy}(n) = 0. \quad (1.104)$$

Em seguida, seleccionam-se os zeros $z_1, \dots, z_n$ que se encontram no interior do círculo unitário para determinar a função de transferência do processo:

$$H(z) = (1 - z_1 z^{-1}) \dots (1 - z_n z^{-1}) - \quad (1.105)$$

Os parâmetros do processo são os coeficientes de $H(z)$ [1, 9]:

$$c_1 = -z_1 - \dots - z_n \quad (1.106)$$

$$c_2 = z_1 z_2 + \dots + z_1 z_n + z_2 z_3 + \dots + z_{n-1} z_n \quad (1.107)$$

$$\vdots \quad (1.108)$$

$$c_n = (-1)^n z_1 z_2 \dots z_n. \quad (1.109)$$

A variância do erro de previsão pode ser determinada pela primeira equação de Yule-Walker:

$$(1 + c_1^2 + \dots + c_n^2) \sigma^2 = \lambda_{yy}(0), \quad (1.110)$$

resultando em:

$$\sigma^2 = \frac{\lambda_{yy}(0)}{1 + c_1^2 + \dots + c_n^2}. \quad (1.111)$$

Tal como nos processos MA(1), a função geradora de autocovariância pode apresentar zeros na circunferência unitária. Quando isto ocorre, a distribuição destes zeros para $H(z)$ e $H(z^{-1})$ resulta numa função de transferência com coeficientes complexos, o que corresponde a uma sequência de autocovariância não admissível.

Se o modelo do processo tiver um ou mais zeros próximos da circunferência unitária, a sequência de autocovariância estimada a partir duma realização pode ser inadmissível, dando origem a uma função geradora com um par de zeros ou mais na circunferência unitária. Se for só um par de zeros, este deve ser projectado para 1 ou -1 conforme a parte real seja positiva ou negativa. No cenário improvável, serem mais do que um par, devemos verificar quais os pares que têm multiplicidade par, sendo metade destes atribuídos a $H(s)$. Os restantes zeros na circunferência devem ser projectados para 1 ou -1 , de acordo com a sua parte real [9, 7].

Método da resposta impulsional truncada

Recordemos que a função de transferência $H(z)$ dum processo MA(n) é:

$$H(z) = 1 + c_1 z^{-1} + \dots + c_n z^{-n} = \frac{1}{\left(\frac{1}{H(z)}\right)} \quad (1.112)$$

A ideia é expandir $\frac{1}{H(z)}$ numa soma infinita de coeficientes:

$$\frac{1}{H(z)} = 1 + a_1 z^{-1} + \dots + a_p z^{-p} + \dots = A^\infty(z^{-1}) \quad (1.113)$$

onde os coeficientes a_k , $k = 1, \dots, \infty$ são os valores da resposta impulsional de $\frac{1}{H(z)}$ para $k = 1, \dots, \infty$. Estes valores também podem ser interpretados como os coeficientes do polinómio auto-regressivo dum processo AR de ordem infinita [1]:

$$H(z) = \frac{1}{\left(\frac{1}{H(z)}\right)} = \frac{1}{A^\infty(z^{-1})} \Rightarrow A^{(\infty)}(q^{-1})y(t) = e(t). \quad (1.114)$$

Dado que $H(z)$ é inversamente estável, então a resposta impulsional de $\frac{1}{H(z)}$ é absolutamente somável, ou seja:

$$1 + \sum_{k=1}^{\infty} |a_k| = C < \infty. \quad (1.115)$$

Podemos concluir daqui que $\lim_{p \rightarrow \infty} \sum_{k=p+1}^{\infty} a_k = 0$ e que, consequentemente, podemos aproximar o processo MA(n) por um processo AR(p) com p suficientemente elevado [3]. A determinação dos coeficientes a_k pode ser feita através da divisão longa de 1 por $H(z)$ ou através da transformada z inversa de $\frac{1}{H(z)}$

Exemplo 1 - Aproximação dum processo MA(1) por um modelo AR(p)

Seja $H(z) = 1 + c_1 z^{-1}$ a função de transferência de um processo MA(1). Podemos obter $\frac{1}{H(z)}$ dividindo 1 por $H(z)$

$$\begin{array}{r|l}
 1 & 1 + c_1 z^{-1} \\
 1 & 1 - c_1 z^{-1} + c_1^2 z^{-2} - c_1^3 z^{-3} + \dots \\
 \hline
 0 & -c_1 z^{-1} \\
 & -c_1 z^{-1} \quad -c_1^2 z^{-2} \\
 \hline
 & 0 \quad c_1^2 z^{-2} \\
 & c_1^2 z^{-2} \quad c_1^3 z^{-3} \\
 \hline
 & 0 \quad -c_1^3 z^{-3} \\
 & \vdots
 \end{array}$$

que resulta em

$$\frac{1}{H(z)} = 1 - c_1 z^{-1} + c_1^2 z^{-2} - c_1^3 z^{-3} + \dots + (-1)^p c_1^p z^{-p} + \dots \quad (1.116)$$

Se truncarmos esta soma para $k \geq p$ obtemos o polinómio da aproximação AR(p):

$$A(z^{-1}) = 1 - c_1 z^{-1} + c_1^2 z^{-2} - c_1^3 z^{-3} + \dots + (-c_1)^p z^{-p} \quad (1.117)$$

Uma vez obtida esta aproximação o modelo AR é:

$$A(q^{-1})y(t) = e(t), \quad (1.118)$$

ou seja

$$y(t) = e(t) + c_1 y(t-1) - c_1^2 y(t-2) + c_1^3 y(t-3) - \dots - (-c_1)^p y(t-p) \quad (1.119)$$

Observemos que os coeficientes do polinómio $A(z)$ são os valores da transformada inversa de $\frac{1}{H(z)}$:

$$\mathcal{Z}^{-1} \left\{ \frac{1}{H(z)} \right\} = \mathcal{Z}^{-1} \left\{ \frac{1}{1 + c_1 z^{-1}} \right\} = \begin{cases} 0, & n < 0 \\ (-c_1)^p, & p \geq 0 \end{cases} \quad (1.120)$$

□

Dada a sequência de autocovariância dum processo MA(n), podemos determinar a ordem da aproximação AR através dos coeficientes de correlação parcial e estimar os parâmetros do modelo AR através da resolução das equações de Yule-Walker [7]. O Modelo MA vai ser a resposta impulsional do modelo AR truncada em $t = n$, ou seja, os coeficientes de modelo MA(n) são os valores da resposta impulsional para $t = 1$ a n desta aproximação, ou equivalentemente, os valores da transformada inversa de $H(z) = \frac{1}{A(z^{-1})}$ nestes instantes.

Exemplo 2 - Aproximação dum modelo MA(1) por um modelo AR utilizando a sequência de autocorrelação parcial

A figura (1.2) mostra uma estimativa de sequência de autocorrelação parcial estimada a partir de uma realização dum processo MA(1). Como só a partir do esfasamento 5 (inclusivé) os coeficientes estão dentro do intervalo de confiança, esta

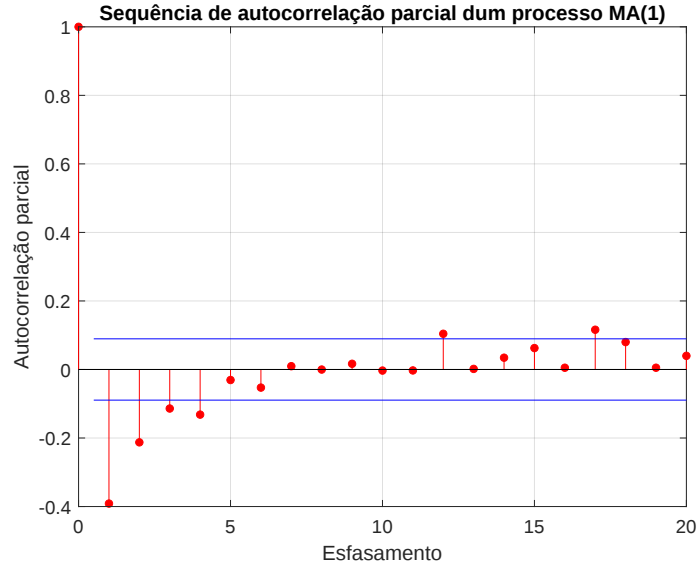


Figura 1.2: Sequência de autocorrelação parcial duma realização dum processo MA(1) e intervalo de confiança de 95%.

sequência sugere um modelo AR(4). O modelo foi estimado através da resolução das equações de Yule-Walker [3, 5], e tem a seguinte função de transferência:

$$\hat{H}(z) = \frac{1}{1 + 0.5124z^{-1} + 0.2998z^{-2} + 0.1777z^{-3} + 0.1271z^{-4}}. \quad (1.121)$$

A Figura 1.3 mostra a resposta impulsional de $\hat{H}(z)$. Como sabemos que o processo é MA(1), o modelo MA estimado é

$$\hat{\hat{H}}(z) = 1 + \hat{h}(1)z^{-1} = 1 - 0.5124z^{-1} \quad (1.122)$$

a que corresponde a equação

$$y(t) = e(t) - 0.5124e(t-1). \quad (1.123)$$

□

Método dos Mínimos Quadrados para Modelos MA

O Teorema da densidade espectral [10] permite-nos interpretar um processos estocástico estacionário como sendo a saída de um sistema linear e invariante no tempo, estável e excitado por ruído branco. Esta representação está ilustrada na figura 1.4. Quando $y(t)$ é um processo MA e assumimos que $H(q) = \frac{1}{A(q)}$ é uma boa aproximação AR, podemos calcular erro de previsão como:

$$\hat{e}(t) = A(q)y(t) = y(t) + a_1y(t-1) + \dots + a_nay(t-na) \quad (1.124)$$

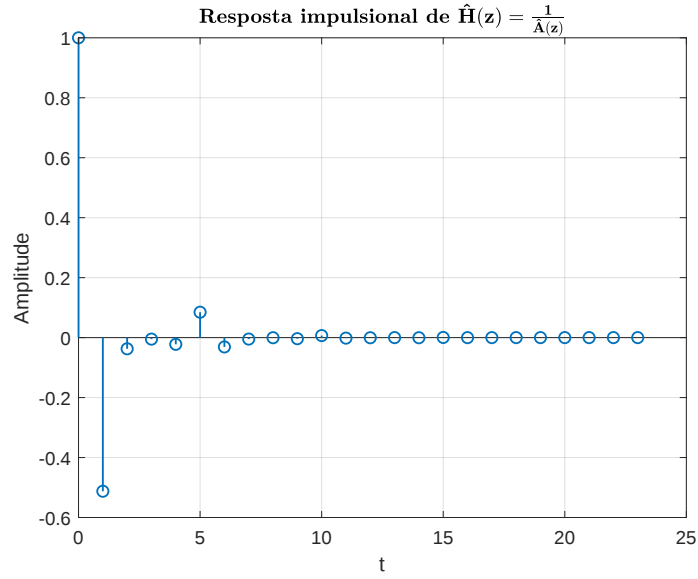


Figura 1.3: Resposta impulsional de $\hat{H}(z)$

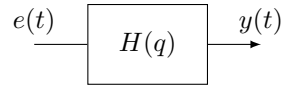


Figura 1.4: O processo estocástico estacionário $y(t)$ como saída de um sistema excitado por ruído branco $e(t)$.

onde $\hat{e}(t)$ representa uma estimativa do termo de ruído $e(t)$.

Com base nesta estimativa $\hat{e}(t)$, podemos prosseguir para estimar o modelo AR através de uma regressão linear, assumindo que $y(t)$ pode ser expresso em termos de $\hat{e}(t)$ da seguinte forma:

$$y(t) = C(q)\hat{e}(t) = \hat{e}(t) + c_1\hat{e}(t-1) + \cdots c_n\hat{e}(t-n). \quad (1.125)$$

Os coeficientes de $C(q)$ do modelo MA são os que minimizam a soma dos quadrados dos resíduos:

$$S(\boldsymbol{\theta}) = \sum_{t=n+1}^N (y(t) - \hat{e}(t) - c_1\hat{e}(t-1) - \cdots - c_n\hat{e}(t-n))^2 \quad (1.126)$$

onde $\boldsymbol{\theta}$ é o vector de parâmetros definido como:

$$\boldsymbol{\theta} = [c_1 \quad \cdots \quad c_n]^T. \quad (1.127)$$

A solução que minimiza $S(\boldsymbol{\theta})$ é obtida através do estimador de mínimos quadrados, conforme apresentado em [9]:

$$\hat{\boldsymbol{\theta}} = \left(\Phi^T \Phi \right)^{-1} \Phi^T \mathbf{Y} \quad (1.128)$$

onde Φ é a matriz de regressão formada pelos valores passados de $\hat{e}(t)$:

$$\Phi = \begin{bmatrix} \hat{e}(n) & \cdots & \hat{e}(1) \\ \hat{e}(n+1) & \cdots & \hat{e}(2) \\ \vdots & \vdots & \vdots \\ \hat{e}(N-1) & \cdots & \hat{e}(N-n) \end{bmatrix} \quad (1.129)$$

e Y é o vetor das amostras de $y(t)$:

$$Y = [y(n+1) \quad \cdots \quad y(N)] \quad (1.130)$$

Rescrevendo a expressão para $\hat{\theta}$, temos:

$$\left(\frac{1}{N-n} \Phi^T \Phi \right) \frac{1}{N-n} \Phi^T Y \quad (1.131)$$

Quando o número de amostras, N , tende para o infinito, obtemos uma estimativa que converge para:

$$\lim_{N \rightarrow \infty} \hat{\theta} = \begin{bmatrix} \lambda_{\hat{e}\hat{e}}(0) & \lambda_{\hat{e}\hat{e}}(1) & \cdots & \lambda_{\hat{e}\hat{e}}(n-1) \\ \lambda_{\hat{e}\hat{e}}(1) & \lambda_{\hat{e}\hat{e}}(0) & \cdots & \lambda_{\hat{e}\hat{e}}(n-2) \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_{\hat{e}\hat{e}}(n-1) & \lambda_{\hat{e}\hat{e}}(n-1) & \cdots & \lambda_{\hat{e}\hat{e}}(0) \end{bmatrix}^{-1} \begin{bmatrix} \lambda_{y\hat{e}}(1) \\ \lambda_{y\hat{e}}(2) \\ \vdots \\ \lambda_{y\hat{e}}(n) \end{bmatrix} \quad (1.132)$$

Assumindo que $\hat{e}(t)$ é um processo de ruído branco, temos $\lambda_{\hat{e}\hat{e}}(\tau) = 0$ para $\tau \neq 0$, e portanto, os coeficientes c_k podem ser estimados como:

$$c_k = \hat{\theta}_k = \frac{\lambda_{y\hat{e}}(k)}{\lambda_{\hat{e}\hat{e}}(0)}, \quad k = 1, \dots, n. \quad (1.133)$$

Este método iterativo fornece uma forma prática e eficiente de estimar os parâmetros de um modelo MA através da construção inicial de um modelo AR e, em seguida, aplicando mínimos quadrados sobre os resíduos estimados [1].

Exemplo 3 - Estimação dum MA(1) por um modelo AR utilizando o método de mínimos quadrados iterativo Fez-se um teste de brancura de ruído aos resíduos do modelo AR(4) estimado no exemplo 2. A figura 1.5 mostra qyue há evidência de que estes resíduos são ruído branco. Utilizámos estes resíduos para estimar o parâmetro do MA(1) através dos mínimos quadrados e obtivemos $\hat{c}_1 = -0.5154$ que corresponde ao modelo MA(1):

$$y(t) = e(t) - 0.5154e(t-1), \quad (1.134)$$

que á muiro semelhante ao estimado no exemplo 2. Com o método da correlação obtivemos $\hat{c}_1 = -0.5167$ que também é um valor muito semelhante. $\square$

Método de Minimização do Erro de Previsão

O método de minimização do erro de previsão permite uma estimação consistente dos parâmetros dum modelo MA, minimizando a função custo dada pela soma do quadrado dos erros de previsão a um passo:

$$J(\theta) = \frac{1}{2} \sum_{t=1}^N \hat{e}^2(t, \theta), \quad (1.135)$$

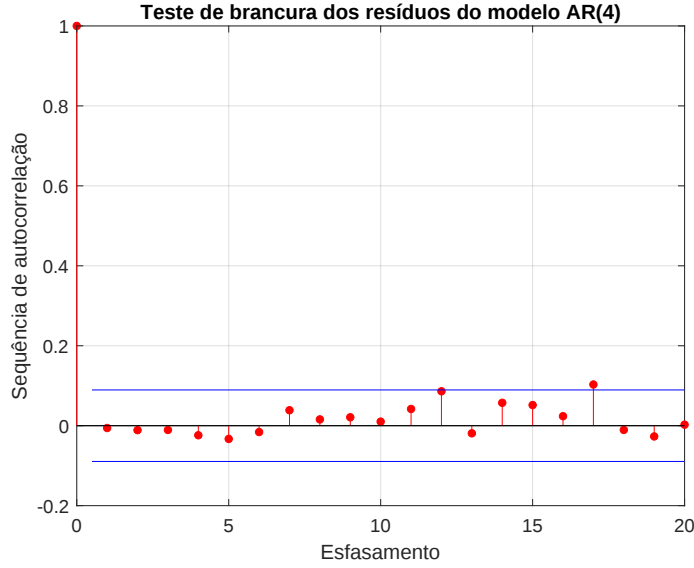


Figura 1.5: Sequência de autocorrelação dos resíduos de modelo AR(4) e intervalo de confiança de 95%.

onde $\hat{e}(t, \theta)$ é o erro de previsão e θ é o vector dos parâmetros do polinómio $C(q^{-1})$ definido em (1.127) [1]. O mínimo ocorre quando o gradiente da função custo é zero [9]:

$$\left(\frac{dJ(\theta)}{d\theta} \right)^T = \left[\frac{\partial J(\theta)}{\partial c_1} \quad \dots \quad \frac{\partial J(\theta)}{\partial c_n} \right]^T = \mathbf{0}. \quad (1.136)$$

Dado que esta equação não é linear nos parâmetros, utiliza-se um método iterativo, como o de descida de gradiente, actualização do parâmetros em cada iteração [7]:

$$\theta^{(i+1)} = \theta^{(i)} - \eta \left(\frac{dJ(\theta)}{d\theta} \right)_{\theta=\theta^{(i)}}^T \quad (1.137)$$

onde $\eta > 0$ é a taxa de aprendizagem.

A derivada da função custo relativamente a c_k é dada por

$$\frac{\partial J(\theta)}{\partial c_k} = \sum_{t=1}^N \frac{\partial \hat{e}(t, \theta)}{\partial c_k} \hat{e}(t, \theta), \quad k = 1, \dots, n. \quad (1.138)$$

onde $\hat{e}(t, \theta) = y(t) - \hat{y}(t, \theta)$ e $\hat{y}(t, \theta)$ é a previsão de $y(t)$.

A previsão pode ser expressa como:

$$\hat{y}(t, \theta) = \frac{C(q^{-1}) - 1}{C(q^{-1})} y(t). \quad (1.139)$$

Ao expandir esta expressão, rearranjar os termos e substituir $y(t-k) - \hat{y}(t-k, \theta)$ por $\hat{e}(t-k, \theta)$, obtemos a forma final predictor [1, 9]:

$$\hat{y}(t, \theta) = c_1 \hat{e}(t-1, \theta) + \dots + c_n \hat{e}(t-n, \theta), \quad (1.140)$$

com o erro de previsão dado por:

$$\hat{e}(t, \boldsymbol{\theta}) = y(t) - \hat{y}(t, \boldsymbol{\theta}) = y(t) - c_1 \hat{e}(t-1, \boldsymbol{\theta}) - \dots - c_n \hat{e}(t-n, \boldsymbol{\theta}) = \frac{1}{C(q^{-1})} y(t). \quad (1.141)$$

Assim, $\hat{e}(t, \boldsymbol{\theta})$ pode ser interpretado como a saída dum sistema linear e invariante no tempo com função de transferência $\frac{1}{C(z^{-1})}$, excitado por $y(t)$ [7].

Para implementar o algoritmo, calcula-se a derivada

$$\frac{\partial \hat{e}(t, \boldsymbol{\theta})}{\partial c_k} = -\hat{e}(t-k, \boldsymbol{\theta}) - c_1 \frac{\partial \hat{e}(t-1, \boldsymbol{\theta})}{\partial c_k} - \dots - c_n \frac{\partial \hat{e}(t-n, \boldsymbol{\theta})}{\partial c_k}, \quad (1.142)$$

resultando em

$$\frac{\partial \hat{e}(t, \boldsymbol{\theta})}{\partial c_k} = -\frac{1}{C(q^{-1})} \hat{e}(t-k, \boldsymbol{\theta}), \quad k = 1, \dots, n. \quad (1.143)$$

Definindo $\hat{e}_F(t, \boldsymbol{\theta})$ como a saída dum sistema linear e invariante no tempo com função de transferência $-\frac{1}{C(z^{-1})}$ excitado por $\hat{e}(t, \boldsymbol{\theta})$ [4]:

$$\hat{e}_F(t, \boldsymbol{\theta}) = -\frac{1}{C(q^{-1})} \hat{e}(t, \boldsymbol{\theta}), \quad (1.144)$$

então, usando as propriedades dos sistemas lineares e invariantes no tempo, temos que:

$$\frac{\partial \hat{e}(t, \boldsymbol{\theta})}{\partial c_k} = \hat{e}_F(t-k, \boldsymbol{\theta}), \quad k = 1, \dots, n. \quad (1.145)$$

A figura 1.6 mostra o diagrama de blocos que resume o cálculo do erro de previsão e

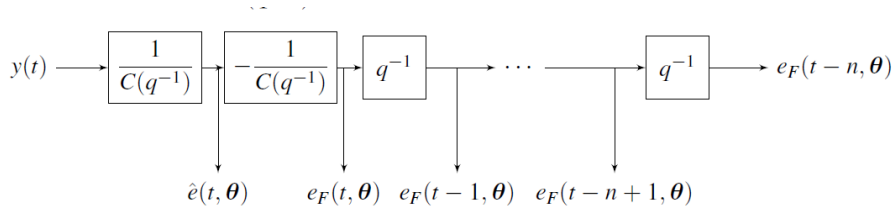


Figura 1.6: Derminação de $\hat{e}(t, \boldsymbol{\theta})$ e das suas derivadas a partir de $y(t)$.

das suas derivadas.

Este algoritmo pode ser resumido no seguintes passos, estando o seu organigrama representado na figura 1.7:

Descrição do Algoritmo

Inicialização: – Obter uma estimativa inicial $\boldsymbol{\theta}^{(0)}$.

– Definir o índice de iteração $i = 0$.

Iteração até convergência: – Calcular o erro estimado:

$$\hat{e}(t, \boldsymbol{\theta}^{(i)}) = \frac{1}{C^{(i)}(q^{-1})} y(t), \quad t = 1, \dots, N$$

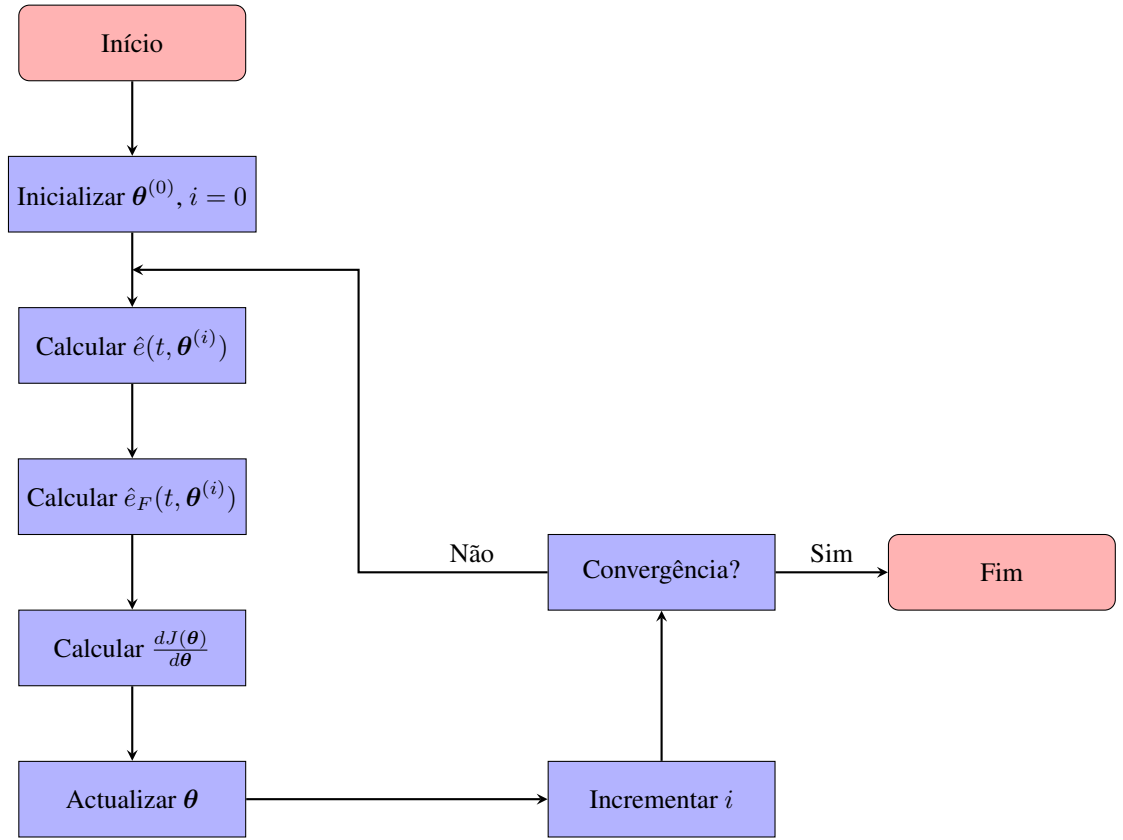


Figura 1.7: Organigrama do Algoritmo de Minimização do Erro de Previsão de Modelos MA

- Calcular o erro filtrado:

$$\hat{e}_F(t, \theta^{(i)}) = -\frac{1}{C^{(i)}(q^{-1})} \hat{e}(t, \theta^{(i)}), \quad t = 1, \dots, N$$

- Calcular o gradiente do custo:

$$\left(\frac{dJ(\theta)}{d\theta} \right)_{\theta=\theta^{(i)}}^T = \sum_{t=1}^N \begin{bmatrix} \hat{e}_F(t-1, \theta^{(i)}) \\ \vdots \\ \hat{e}_F(t-n, \theta^{(i)}) \end{bmatrix} \hat{e}(t, \theta^{(i)}).$$

- Actualizar os parâmetros:

$$\theta^{(i+1)} = \theta^{(i)} - \eta \left(\frac{dJ(\theta)}{d\theta} \right)_{\theta=\theta^{(i)}}^T$$

- Incrementar i .

Como $\hat{e}(t)$ não é uma função linear dos parâmetros, $J(\theta)$ não é uma função convexa, apresentando vários mínimos locais. A convergência do algoritmo depende da escolha

adequada dos valores iniciais e da taxa de aprendizagem η . Valores baixos de η resultam em convergência mais lenta e segura; valores altos aceleram a convergência, mas podem causar saltos para fora da região de convergência [9, 11]. Depois de obtidos os coeficientes de $C(q^{-1})$ a variância do ruído branco pode ser estimada através de:

$$\hat{\sigma}^2 = \frac{1}{N} \sum_{t=1}^N \hat{e}^2(t, \theta^{(i)}). \quad (1.146)$$

1.2.4 Determinação da ordem de modelos MA - Sequência de Autocorrelação

Num modelo MA de ordem n , a função de autocorrelação, denotada por $\rho_{yy}(\tau)$, é diferente de zero apenas até ao esfasamento n . Para valores de esfasamento superiores, $\rho_{yy}(k)$ é nulo. Formalmente, temos:

$$\rho_{yy}(\tau) = 0, \quad \tau > n \quad (1.147)$$

Isto implica que, ao observar a sequência de autocorrelação de uma realização do processo, é possível identificar o ponto que indica a ordem do modelo n , em que os valores de autocorrelação se aproximam de zero. Assim para estimar a ordem do modelo temos que, em primeiro lugar, calcular a sequência de autocorrelação a partir dos dados observados. Após calcular a sequência de autocorrelação, deve-se determinar o ponto em que $\rho_{yy}(\tau)$ se torna praticamente zero. Este ponto marca o esfasamento máximo n até onde a autocorrelação é significativa. Para valores de $\tau > n$, espera-se que $\rho_{yy}(\tau)$ seja aproximadamente zero. Para determinar se os coeficientes de autocorrelação são significativamente diferentes de zero, é utilizado um intervalo de confiança com base numa distribuição normal. Sob a hipótese nula de que o coeficiente de autocorrelação é zero, a variância do coeficiente estimado é aproximadamente $\frac{1}{N}$, onde N é o tamanho da amostra. Assim, os limites dum intervalo com 95% de confiança são $\pm \frac{1.96}{N}$. A ordem do modelo será o último valor de τ para o qual todos os coeficientes de autocorrelação estão fora do intervalo. A figura (1.8) mostra a sequência de autocorrelação dum processo AR. Esta figura evidencia ue o processo é de ordem 2 pois, só para $\tau > 2$ é que os coeficientes de autocorrelação estão dentro do intervalo de confiança.

1.2.5 Modelos ARMA

Os modelos ARMA são amplamente utilizados para representar séries temporais estacionárias, combinando componentes AR. Em termos matemáticos, um modelo ARMA(n, m) de ordem n para a parte AR e ordem m para a parte MA é descrito pela seguinte equação:

$$A(q^{-1})y(t) = C(q^{-1})e(t) \quad (1.148)$$

onde

$$A(q^{-1}) = 1 + a_1q^{-1} + \dots + a_nq^{-n} \quad (1.149)$$

e

$$C(q^{-1}) = 1 + c_1q^{-1} + \dots + c_mq^{-m} \quad (1.150)$$

A estimação destes modelos é complexa devido à presença simultânea de duas componentes interdependentes: a componente AR, que depende diretamente dos valores

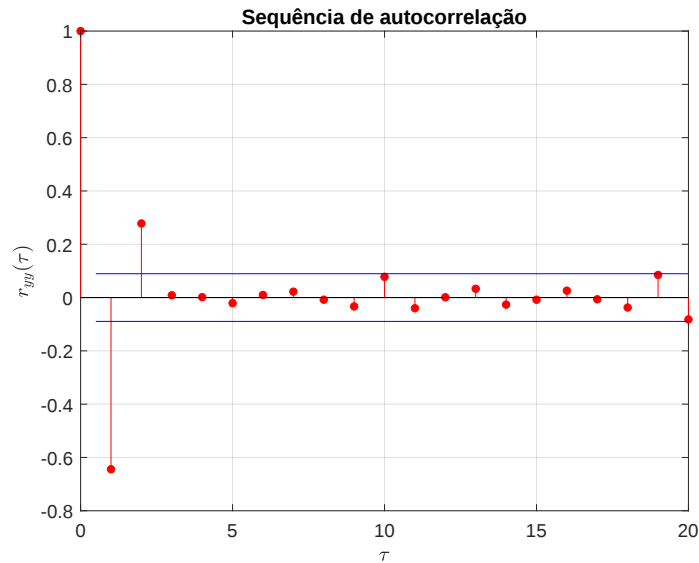


Figura 1.8: Sequência de autocorrelação duma realização dum processo MA(2) e intervalo de confiança de 95%.

passados da série, e a componente MA, que depende dos erros passados. Essa interdependência faz com que a estimação das duas partes necessite de ser cuidadosamente realizada, de forma a minimizar a interferência mútua. Diversos métodos foram desenvolvidos para abordar esse problema, permitindo a estimação eficiente dos parâmetros AR e MA, tanto isoladamente quanto de forma conjunta.

Nesta secção, abordaremos três métodos para a estimação de modelos ARMA:

1. **Método dos Esfasamentos Superiores:** Este método isola a componente AR ao aplicar as equações de Yule-Walker para esfasamentos superiores à ordem da parte MA, onde a influência desta se torna desprezível. Isso permite estimar a parte AR de forma inicial, minimizando a interferência da componente MA.
2. **Método de Ajuste da Resposta Impulsional:** Neste método, aproximamos a resposta impulsional do modelo ARMA pela dum modelo AR. Primeiramente, modela-se a série como um modelo AR de ordem elevada p (designado por $AR(p)$), para capturar a dinâmica essencial do processo. A ordem p é escolhida de modo a que a resposta impulsional do modelo AR reproduza bem a do modelo ARMA. A partir do modelo AR obtido, calcula-se a sua resposta impulsional, que será ajustada para coincidir com os primeiros valores da resposta impulsional do modelo $ARMA(n, m)$.
3. **Método de Mínimos Quadrados para Modelos ARMA:** Este método é uma extensão do Método de Mínimos Quadrados para Modelos MA, agora aplicado a modelos ARMA. Assim como no caso de modelos MA, começa por estimar uma aproximação AR do processo e, em seguida, utiliza mínimos quadrados aplicados aos resíduos para determinar as componentes AR e MA.

4. **Método de Minimização do Erro de Previsão:** Este método otimiza os parâmetros ao minimizar o erro de previsão a um passo da série, utilizando técnicas de otimização para ajustar conjuntamente os parâmetros AR e MA, de modo a reduzir o erro de previsão global.

Cada um desses métodos será detalhado, com explicações sobre o processo e as fórmulas necessárias. A escolha do método mais adequado dependerá das características específicas da série e do nível de precisão desejado na estimação dos parâmetros.

Método dos Esfasamentos Superiores para Estimativa da Parte Auto-Regressiva de um Modelo ARMA

Método dos Esfasamentos Superiores é uma abordagem eficaz para estimar os parâmetros da componente AR de um modelo ARMA(n, m) sem a influência inicial da componente MA. Este método baseia-se na aplicação das equações de Yule-Walker em esfasamentos (superiores à ordem m da componente MA, onde a influência desta parte é nula. Se $y(t)$ for um processo ARMA(n, m), a sua sequência de autocovariância $\lambda_{yy}(\tau)$, para $\tau > m$, satisfaz as seguintes equações:

$$\sum_{k=1}^n \lambda_{yy}(\tau - k) a_k = -\lambda_{yy}(\tau) \quad (1.151)$$

Assim, para $m < \tau \leq n + m$ obtemos o seguinte sistema:

$$\begin{aligned} \lambda_{yy}(m) a_1 + \lambda_{yy}(m-1) a_2 + \cdots + \lambda_{yy}(m+1-n) a_n &= -\lambda_{yy}(m+1) \\ \lambda_{yy}(m+1) a_1 + \lambda_{yy}(m) a_2 + \cdots + \lambda_{yy}(m+2-n) a_n &= -\lambda_{yy}(m+2) \\ &\vdots \\ \lambda_{yy}(m+n) a_1 + \lambda_{yy}(m+n-1) a_2 + \cdots + \lambda_{yy}(m) a_n &= -\lambda_{yy}(m+n+1) \end{aligned}$$

Este é um sistema de equações de Yule-Walker muito semelhante ao utilizado para estimar modelos AR, onde a componente MA não interfere. Assim, podemos estimar os coeficientes $a_1, a_2, \dots, a_n$, utilizando, por exemplo, o algoritmo de Levinson-Durbin.

Após obter os coeficientes AR, calculamos os resíduos da série usando esses coeficientes. Esses resíduos aproximam-se de um processo MA, permitindo-nos modelar a componente MA independentemente. Podemos ajustar a parte MA aos resíduos obtidos utilizando os métodos que estudamos para estimar modelos MA.

Método de Ajuste da Resposta Impulsional

Este método é muito semelhante ao método da resposta impulsional truncada para estimar modelos MA, pois também se baseia no ajuste da resposta impulsional. A ideia é fazer coincidir os primeiros valores da resposta impulsional de um modelo ARMA com os de uma aproximação AR. No entanto, aqui o ajuste envolve resolver um sistema de equações lineares, o que torna o processo um pouco mais complexo.

Para começar, modelamos a série temporal como um modelo AR de ordem p , designado por AR(p). Este modelo AR serve como uma aproximação inicial, capturando a dinâmica essencial do processo, incluindo as componentes AR e MA do modelo ARMA original. Para que a resposta impulsional do modelo AR seja representativa, a ordem p deve ser suficientemente alta, permitindo uma boa aproximação da resposta

impulsional do modelo ARMA.

A determinação de p pode ser feita a partir da sequência de autocovariância do processo, com base nos coeficientes de correlação parcial. Para obter os parâmetros do modelo AR, resolvem-se as equações de Yule-Walker.

Com o modelo AR estimado, calculamos a sua resposta impulsional. A resposta impulsional representa a reação do sistema a um impulso de Dirac e reflete a dinâmica essencial do processo. O objetivo é que esta resposta se aproxime da resposta impulsional do modelo ARMA que estamos a tentar modelar.

Agora, introduzimos um modelo ARMA(n, m) de baixa ordem (com valores de n e m definidos previamente) e ajustamos a sua resposta impulsional para coincidir com os primeiros valores da resposta impulsional do modelo AR.

O objetivo é encontrar os coeficientes do modelo ARMA que reproduzam os valores iniciais da resposta impulsional da aproximação AR. Para isso, podemos:

- Resolver um sistema de equações lineares se quisermos igualar precisamente os primeiros $n + m + 1$ pontos da resposta impulsional, ou
- utilizar um estimador de mínimos quadrados, caso desejemos aproximar um número de pontos maior que $n + m + 1$.

As equações para os parâmetros do modelo ARMA(n, m) baseiam-se na equação às diferenças que descreve o processo:

$$y(t) = e(t) + c_1 e(t-1) + \dots + c_m e(t-m) - a_1 y(t-1) - \dots - a_n y(t-n). \quad (1.152)$$

onde $y(t)$ é a saída e $e(t)$ é a entrada. A resposta impulsional $h(t)$ corresponde é o valor de $y(t)$ quando a entrada é o impulso de Dirac, definido por:

$$e(t) = \delta(t) = \begin{cases} 1, & t = 0 \\ 0, & t \neq 0 \end{cases} \quad (1.153)$$

e o sistema está em repouso (ou seja, $y(t-1) = y(t-2) = \dots = 0$).
ara calcular os primeiros valores de $h(t)$, obtemos

$$\begin{aligned} h(0) &= e(0) = 1 \\ h(1) &= c_1 - a_1 h(0) \\ h(2) &= c_2 - a_1 h(1) - a_2 h(0) \\ &\vdots \\ h(m) &= -c_m - a_1 h(m-1) - a_2 h(m-2) - \dots - a_m h(0) \\ &\vdots \\ h(n+m) &= -a_1 h(n+m-1) - \dots - a_n h(m). \end{aligned}$$

Podemos reescrever essas equações na forma matricial como:

$$\left[\begin{array}{c|c} \mathbf{I}_m & -\mathcal{H}_{n+m} \end{array} \right] \boldsymbol{\theta} = \mathbf{h}_{1:n+m} \quad (1.154)$$

onde

- $\mathcal{H}_{n+m}$ é a matriz:

$$\mathcal{H}_{n+m} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ h(1) & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ h(n) & h(n-1) & \cdots & 1 \\ \vdots & \vdots & \vdots & \vdots \\ h(n+m) & h(n+m-1) & \cdots & h(m) \end{bmatrix} \in \mathbb{R}^{n+m \times n}, \quad (1.155)$$

- θ é o vector de parâmetros:

$$\theta = [c_1 \quad \cdots \quad c_m \quad a_1 \quad \cdots \quad a_n]^T \in \mathbb{R}^{n+m} \quad (1.156)$$

- $h_{1:n+m}$ o vector das amostras da resposta impulsional:

$$h_{1:n+m} = [h(1) \quad \cdots \quad h(n+m)]^T \in \mathbb{R}^{n+m} \quad (1.157)$$

Finalmente, os parâmetros do modelo são obtidos resolvendo:

$$\theta = \left[\frac{\mathbf{I}_m}{\mathbf{0}_{n \times m}} \mid -\mathcal{H}_{n+m} \right]^{-1} h_{1:n+m} \quad (1.158)$$

Desta forma, o ajuste da resposta impulsional permite encontrar os parâmetros do modelo ARMA, reproduzindo os primeiros valores da dinâmica do processo representado pelo modelo AR.

Método de Mínimos Quadrados para Modelos ARMA

O Método de Mínimos Quadrados para Modelos ARMA expande a abordagem utilizada no Método de Mínimos Quadrados para Modelos MA, que foi anteriormente discutido. Este método estima simultaneamente a componentes AR e MA do modelo.

Tal como no Método de Ajuste da Resposta Impulsional, começamos por aproximar a série temporal com um modelo AR de ordem elevada, utilizando esta aproximação inicial para capturar a dinâmica do processo antes de ajustar conjuntamente as componentes AR e MA. A ideia é que uma ordem p suficientemente elevada permitirá que a resposta impulsional do modelo AR seja representativa da resposta do modelo ARMA. Para seleccionar p , podemos, uma vez mais, utilizar os coeficientes de correlação parcial, e os parâmetros do modelo AR podem ser obtidos através das equações de Yule-Walker.

Assim, tal como no Método de Mínimos Quadrados para Modelos MA, esta aproximação AR fornece uma base inicial sobre a qual as componentes AR e MA do modelo ARMA serão ajustadas.

Com o modelo AR inicializado, avançamos para a estimação dos parâmetros do modelo ARMA. Ao contrário do caso MA, onde só precisávamos ajustar a componente MA, aqui procuramos minimizar o erro quadrático total entre os valores observados e a combinação das componentes AR e MA, ou seja, o ajuste considera o impacto de

ambas as partes do modelo.

A abordagem de Mínimos Quadrados visa encontrar os valores dos coeficientes c_i , $i = 1, \dots, m$ e a_j , $j = 1, \dots, n$ do modelo (1.152) que minimizem a soma dos quadrados dos resíduos dados por:

$$S(\boldsymbol{\theta}) = \sum_{t=1}^N (y(t) - \hat{y}(t, \boldsymbol{\theta}))^2, \quad (1.159)$$

onde $\hat{y}(t, \boldsymbol{\theta})$ é a estimativa do modelo e $\boldsymbol{\theta}$ o vector de parâmetros definido em (1.156). A solução que minimiza $S(\boldsymbol{\theta})$ é novamente o estimador de mínimos quadrados (1.128) mas agora com a matriz de regressão Φ dada por:

$$\Phi = \begin{bmatrix} \hat{e}(r) & \cdots & \hat{e}(r-m) & -y(r) & \cdots & -y(r-n) \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \hat{e}(N-1) & \cdots & \hat{e}(N-1-m) & -y(N-1) & \cdots & -y(N-1-n) \end{bmatrix} \quad (1.160)$$

e o vector $\mathbf{Y}$ das amostras dado por:

$$\mathbf{Y} = [y(r+1) \quad \cdots \quad y(N)]^T \quad (1.161)$$

onde $r = \max(m, n)$.

Depois de ajustar o modelo ARMA, é importante verificar se este reflete bem a dinâmica do processo. Isso pode ser feito através dum teste de brancura de ruído aos resíduos do modelo. Caso necessário, é possível ajustar iterativamente a ordem p do modelo AR inicial ou e recalibrar os parâmetros do modelo ARMA para melhorar a aproximação aos dados da série.

Método de Minimização do Erro de Previsão

O Método de Minimização do Erro de Previsão permite a estimação consistente dos parâmetros de um modelo ARMA, minimizando a função de custo dada pela soma dos quadrados dos erros de previsão a um passo. Esse método expande o conceito aplicado previamente para modelos MA, onde o foco era apenas a componente MA. Para modelos ARMA(n, m), o método considera agora a interdependência das componentes AR e MA, ajustando ambos os conjuntos de parâmetros simultaneamente para minimizar o erro de previsão.

A função de custo é idêntica à do método para modelos MA e está definida em (1.135). Para a minimizar, buscamos os valores dos parâmetros que anulam o seu gradiente, isto é:

$$\left(\frac{dJ(\boldsymbol{\theta})}{d\boldsymbol{\theta}} \right)^T = \left[\frac{\partial J(\boldsymbol{\theta})}{\partial c_1} \quad \cdots \quad \frac{\partial J(\boldsymbol{\theta})}{\partial c_m} \quad \frac{\partial J(\boldsymbol{\theta})}{\partial a_1} \quad \cdots \quad \frac{\partial J(\boldsymbol{\theta})}{\partial a_n} \right]^T = \mathbf{0}. \quad (1.162)$$

Tal como no caso dos modelos MA, esta função não é linear nos parâmetros e, por isso, a minimização requer um método iterativo, como o do gradiente descendente onde, em cada iteração, os parâmetros são actualizados através de (1.137).

As derivadas da função custo relativamente a c_k e a_k são dada por

$$\frac{\partial J(\boldsymbol{\theta})}{\partial c_k} = \sum_{t=1}^N \frac{\partial \hat{e}(t, \boldsymbol{\theta})}{\partial c_k} \hat{e}(t, \boldsymbol{\theta}), \quad k = 1, \dots, m \quad (1.163)$$

$$\frac{\partial J(\boldsymbol{\theta})}{\partial a_k} = \sum_{t=1}^N \frac{\partial \hat{e}(t, \boldsymbol{\theta})}{\partial a_k} \hat{e}(t, \boldsymbol{\theta}), \quad k = 1, \dots, n \quad (1.164)$$

onde $\hat{e}(t, \boldsymbol{\theta}) = y(t) - \hat{y}(t, \boldsymbol{\theta})$ é o erro de previsão sendo a previsão $\hat{y}(t, \boldsymbol{\theta})$ dada por:

$$\hat{y}(t, \boldsymbol{\theta}) = \frac{C(q^{-1}) - A(q^{-1})}{C(q^{-1})} y(t) \quad (1.165)$$

a que corresponde a seguinte equação às diferenças:

$$\hat{y}(t, \boldsymbol{\theta}) = c_1 \hat{e}(t-1, \boldsymbol{\theta}) + \dots + c_m \hat{e}(t-m, \boldsymbol{\theta}) - a_1 y(t-1) - \dots - a_n y(t-n), \quad (1.166)$$

com o erro de previsão dado por

$$\begin{aligned} \hat{e}(t, \boldsymbol{\theta}) &= -c_1 \hat{e}(t-1, \boldsymbol{\theta}) - \dots - c_m \hat{e}(t-m, \boldsymbol{\theta}) + y(t) + a_1 y(t-1) + \dots + a_n y(t-n) = \\ &= \frac{A(q^{-1})}{C(q^{-1})} y(t). \end{aligned} \quad (1.167)$$

Tal como anteriormente, $\hat{e}(t, \boldsymbol{\theta})$ pode ser interpretado como a saída dum sistema linear e invariante no tempo excitado $y(t)$, mas agora com função de transferência $\frac{A(z^{-1})}{C(z^{-1})}$.

As derivadas do erro de previsão relativamente aos parâmetros do modelo são

$$\frac{\partial \hat{e}(t, \boldsymbol{\theta})}{\partial c_k} = -\hat{e}(t-k, \boldsymbol{\theta}) - c_1 \frac{\partial \hat{e}(t-1, \boldsymbol{\theta})}{\partial c_k} - \dots - c_m \frac{\partial \hat{e}(t-m, \boldsymbol{\theta})}{\partial c_k}, \quad (1.168)$$

e

$$\frac{\partial \hat{e}(t, \boldsymbol{\theta})}{\partial a_k} = y(t-k) - c_1 \frac{\partial \hat{e}(t-1, \boldsymbol{\theta})}{\partial a_k} - \dots - c_m \frac{\partial \hat{e}(t-m, \boldsymbol{\theta})}{\partial a_k} \quad (1.169)$$

resultando em

$$\frac{\partial \hat{e}(t, \boldsymbol{\theta})}{\partial c_k} = -\frac{1}{C(q^{-1})} \hat{e}(t-k, \boldsymbol{\theta}), \quad k = 1, \dots, m. \quad (1.170)$$

e

$$\frac{\partial \hat{e}(t, \boldsymbol{\theta})}{\partial a_k} = \frac{1}{C(q^{-1})} y(t-k), \quad k = 1, \dots, n. \quad (1.171)$$

Definindo

$$\hat{e}_F(t, \boldsymbol{\theta}) = -\frac{1}{C(q^{-1})} \hat{e}(t, \boldsymbol{\theta}) \quad (1.172)$$

e

$$y_F(t, \boldsymbol{\theta}) = \frac{1}{C(q^{-1})} y(t) \quad (1.173)$$

temos que, utilizando as propriedades dos sistemas lineares e invariantes no tempo

$$\frac{\partial \hat{e}(t, \boldsymbol{\theta})}{\partial c_k} = \hat{e}_F(t-k, \boldsymbol{\theta}) \quad (1.174)$$

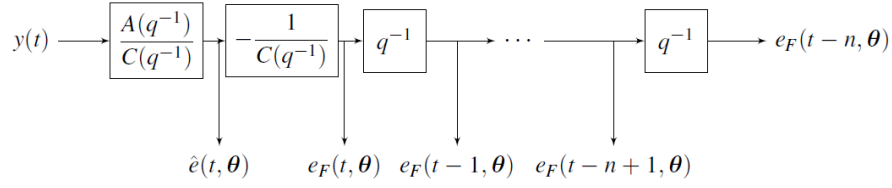


Figura 1.9: Derminação de $\hat{e}(t, \theta)$ e das suas derivadas em ordem a c_k a partir de $y(t)$.

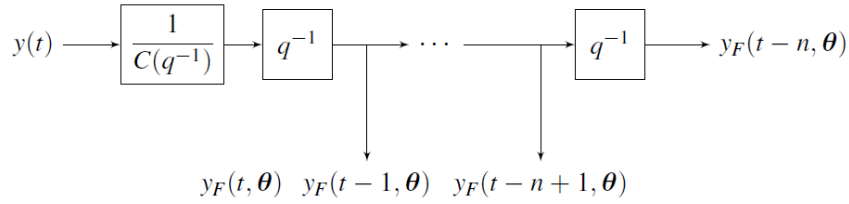


Figura 1.10: Derminação das derivadas $\hat{e}(t, \theta)$ e das suas derivadas em ordem a a_k a partir de $y(t)$.

e

$$\frac{\partial \hat{e}(t, \theta)}{\partial a_k} = y_F(t - k, \theta) \quad (1.175)$$

A figuras 1.9 e 1.9 mostram os diagrama de blocos que resumem o cálculo do erro de previsão e das suas derivadas. O algoritmo iterativo para o minimizar o erro de previsão de modelos ARMA pode ser resumido nos seguintes passos:

Inicialização:

Obter $\theta^{(0)}$ e definir $i = 0$.

Repetir até convergir:

- Calcular $\hat{e}(t, \theta^{(i)}) = \frac{A^{(i)}(q^{-1})}{C^{(i)}(q^{-1})} y(t)$, $t = 1, \dots, n$
- Calcular $\hat{e}_F(t, \theta^{(i)}) = -\frac{1}{C^{(i)}(q^{-1})} \hat{e}(t)$, $t = 1, \dots, n$
- Calcular $y_F(t, \theta^{(i)}) = \frac{1}{C^{(i)}(q^{-1})} y(t)$, $t = 1, \dots, n$
- Para $k = 1$ até m , calcular $\frac{\partial \hat{e}(t, \theta^{(i)})}{\partial c_k^{(i)}} = \hat{e}_F(t - k, \theta^{(i)})$, $t = 1, \dots, n$
- Para $k = 1$ até n , calcular $\frac{\partial \hat{e}(t, \theta^{(i)})}{\partial a_k^{(i)}} = y_F(t - k, \theta^{(i)})$, $t = 1, \dots, n$

- Calcular $\left(\frac{dJ(\boldsymbol{\theta})}{d\boldsymbol{\theta}}\right)_{\boldsymbol{\theta}=\boldsymbol{\theta}^{(i)}}^T = \sum_{t=1}^N \begin{bmatrix} \hat{e}_F(t-1, \boldsymbol{\theta}^{(i)}) \\ \vdots \\ \hat{e}_F(t-m, \boldsymbol{\theta}^{(i)}) \\ y_F(t-1, \boldsymbol{\theta}^{(i)}) \\ \vdots \\ y_F(t-n, \boldsymbol{\theta}^{(i)}) \end{bmatrix} \hat{e}(t, \boldsymbol{\theta}^{(i)}).$
- Actualizar $\boldsymbol{\theta} : \boldsymbol{\theta}^{(i+1)} = \boldsymbol{\theta}^{(i)} - \eta \left(\frac{dJ(\boldsymbol{\theta})}{d\boldsymbol{\theta}}\right)_{\boldsymbol{\theta}=\boldsymbol{\theta}^{(i)}}^T$
- incrementar i

$$\text{Calcular } \hat{\sigma}^2 = \frac{1}{N} \sum_{t=1}^N \hat{e}^2(t, \boldsymbol{\theta}^{(i)}).$$

Como no caso dos modelos MA, o erro de previsão, $\hat{e}(t, \boldsymbol{\theta})$ não é uma função linear dos parâmetros, tornando $J(\boldsymbol{\theta})$ uma função não convexa que pode apresentar múltiplos mínimos locais. Assim, a convergência do algoritmo depende de uma boa escolha de valores iniciais e de uma taxa de aprendizagem η adequada. Valores baixos de η garantem uma convergência mais estável, mas lenta; valores altos aceleram o processo, porém podem comprometer a estabilidade.

Bibliografia

- [1] G. E. P. Box and G. M. Jenkins. *Time Series Analysis: Forecasting and Control*. Holden-Day, 1976.
- [2] George EP Box, Gwilym M Jenkins, Gregory C Reinsel, and Greta M Ljung. *Time Series Analysis: Forecasting and Control*. John Wiley & Sons, 5 edition, 2015.
- [3] Peter J. Brockwell and Richard A. Davis. *Introduction to Time Series and Forecasting*. Springer, New York, 2nd edition, 2002.
- [4] G. C. Goodwin and K. S. Sin. *Adaptive Filtering Prediction and Control*. Prentice Hall, 1984.
- [5] James D. Hamilton. *Time Series Analysis*. Princeton University Press, Princeton, NJ, 1994.
- [6] Monson H Hayes. *Statistical Digital Signal Processing and Modeling*. John Wiley & Sons, 1996.
- [7] S. Haykin. *Adaptive Filter Theory*. Prentice Hall, 1996.
- [8] Roger A Horn and Charles R Johnson. *Matrix Analysis*. Cambridge University Press, Cambridge, UK, 2nd edition, 2013.
- [9] Lennart Ljung. *System Identification: Theory for the User*. Prentice Hall, Upper Saddle River, NJ, 2nd edition, 1999.
- [10] Maurice B. Priestley. *Spectral Analysis and Time Series*. Academic Press, London, 1981.
- [11] T. Soderstrom and P. Stoica. *System Identification*. Prentice Hall, 1989.
- [12] Gilbert Walker. On periodicity in series of related terms. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*, 131(818):518–532, 1931.
- [13] George Udny Yule. On a method of investigating periodicities in disturbed series, with special reference to wolfer’s sunspot numbers. *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 226(636-646):267–298, 1927.