



UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL  
INSTITUTO DE MATEMÁTICA  
PROGRAMA DE PÓS-GRADUAÇÃO EM MATEMÁTICA

**Análise Estatística Bayesiana em Processos  
com Longa Dependência**

Dissertação de Mestrado

**Avelino Viana Dias Junior**

Porto Alegre, 20 de Outubro de 2010.

Dissertação submetida por Avelino Viana Dias Junior <sup>1</sup> como requisito parcial para a obtenção do grau de Mestre em Matemática pelo Programa de Pós-Graduação em Matemática do Instituto de Matemática da Universidade Federal do Rio Grande do Sul.

Professora Orientadora:

Dr.<sup>a</sup>. Sílvia Regina Costa Lopes

Banca Examinadora:

Dr. Artur Oscar Lopes

Dr. Cleber Bisognin

Dr.<sup>a</sup>. Simone Maffini Cerezer

Data da Defesa: 20 de Outubro de 2010.

---

<sup>1</sup>Conselho Nacional de Desenvolvimento Científico e Tecnológico

## AGRADECIMENTOS

Aos meus pais pelo apoio incondicional. Por me ensinarem a enfrentar os problemas com tranquilidade e nunca desistir de buscar meus objetivos.

À minha orientadora Dr<sup>a</sup>. Sílvia Regina Costa Lopes pela oportunidade, confiança e dedicação.

Aos colegas do Programa de Pós-Graduação em Matemática da UFRGS.

Aos professores do Programa de Pós-Graduação em Matemática da UFRGS pelo comprometimento com a qualidade de trabalho. Especialmente ao professor Cleber Bisognin pelas sugestões e comentários.

Aos bolsistas do LCPM pelo auxílio e amizade.

Ao Marcus Alexandre Nunes pelo auxílio na parte computacional. Agradeço também ao Guilherme Pumi, à Taiane Schaedler Prass e à Raquel Romes Linhares. Suas sugestões e comentários foram de grande importância para o desenvolvimento desta dissertação.

À banca examinadora pela atenção e sugestões.

À Rosane, secretária do Programa de Pós-Graduação, pela atenção e pela simpatia.

À CAPES (Coordenação de Aperfeiçoamento de Pessoal de Ensino Superior) pelo apoio financeiro.

À Universidade Federal do Rio Grande do Sul pela oportunidade.

Ao pessoal da CEFVAV, casa do estudante que me aceitou como morador e onde eu tive a felicidade de fazer muitas amizades.

E a todos que torceram ou que, de alguma forma, contribuíram para a realização deste trabalho. Muito obrigado.

## RESUMO

A abordagem Bayesiana na inferência estatística tem sido muito utilizada como uma alternativa aos métodos clássicos. Neste trabalho, apresentamos uma abordagem Bayesiana para a estimação dos parâmetros dos modelos *autoregressivos de médias móveis* de ordens  $p$  e  $q$ , denotados por  $\text{ARMA}(p, q)$  e do modelo *autoregressivo fracionariamente integrado de médias móveis*, denotado por  $\text{ARFIMA}(p, d, q)$ . Para o último modelo, a abordagem Bayesiana é realizada assumindo  $p = q = 0$ . Considerando o modelo  $\text{AR}(p)$ , que é um caso particular do modelo  $\text{ARMA}(p, q)$  onde  $q = 0$ , um estimador é proposto através da abordagem Bayesiana. A eficiência do estimador é verificada através de simulações de Monte Carlo e os resultados são comparados com o método clássico da máxima verossimilhança. No caso do modelo  $\text{ARFIMA}(0, d, 0)$ , um estudo teórico é realizado através de uma abordagem Bayesiana. Para estimar os parâmetros desse modelo, é utilizada a sua representação autoregressiva.

Alguns algoritmos computacionais Bayesianos são apresentados nesse trabalho já que desempenham um papel importante na inferência Bayesiana. Alguns desses algoritmos, como o amostrador de Gibbs e o Metropolis-Hastings, foram utilizados na construção dos estimadores para os parâmetros dos modelos ARMA e ARFIMA.

## ABSTRACT

The Bayesian approach in statistical inference has been widely used as an alternative to traditional methods. In this work, we present a Bayesian approach for estimating the parameters of the *autoregressive moving average processes of order  $p$  and  $q$* , denoted by  $\text{ARMA}(p, q)$  and of the *autoregressive fractionally integrated moving average process*, denoted by  $\text{ARFIMA}(p, d, q)$ . For the later model, the Bayesian approach is performed assuming  $p = q = 0$ . Whereas  $\text{AR}(p)$ , which is a particular case of the  $\text{ARMA}(p, q)$  model when  $q = 0$ , an estimator is proposed via the Bayesian approach. The efficiency of the estimator is verified by Monte Carlo simulations and the results are compared with the classical maximum likelihood estimator. In the case of  $\text{ARFIMA}(0, d, 0)$  process, a theoretical study is performed by the Bayesian approach. For estimating the parameters of that process we consider its infinite autoregressive representation.

Some Bayesian computational algorithms are presented in this work since they play an important role in Bayesian inferences. Some of these algorithms, such as Gibbs sampler and Metropolis-Hastings algorithm, were used in building the estimators for the parameters of ARMA and ARFIMA models.

# ÍNDICE

<b>1</b>	<b>Introdução</b>	<b>1</b>
<b>2</b>	<b>Noções Preliminares</b>	<b>4</b>
2.1	Conceitos e Resultados Básicos . . . . .	4
2.2	Processos ARMA( $p, q$ ) . . . . .	9
2.3	Processos com Longa Dependência . . . . .	11
<b>3</b>	<b>Estimação Bayesiana em Processos de Longa Dependência</b>	<b>16</b>
3.1	Estimação Clássica . . . . .	17
3.1.1	Método da Máxima Verossimilhança . . . . .	17
3.1.2	Método da Regressão Utilizando a Função Periodograma	19
3.2	Análise e Decisão Bayesiana . . . . .	21
3.2.1	Análise Bayesiana de um Modelo ARMA( $p, q$ ) . . . . .	24
3.3	Análise Bayesiana de um Processo ARFIMA( $0, d, 0$ ) . . . . .	34
3.4	Estimação em Processos ARFIMA( $0, d, 0$ ) . . . . .	41
<b>4</b>	<b>Alguns Algoritmos Computacionais Bayesianos</b>	<b>47</b>
4.1	Método de Monte Carlo Ordinário . . . . .	47
4.2	Método de Monte Carlo por Importância . . . . .	48
4.3	Método de Monte Carlo em Cadeias de Markov . . . . .	49
4.4	Amostrador de Gibbs . . . . .	51
4.5	Algoritmo de Metropolis-Hastings . . . . .	52
4.6	Amostrador de Fatia . . . . .	54
<b>5</b>	<b>Conclusão</b>	<b>58</b>
	<b>Referências</b>	<b>60</b>
	<b>Anexo A</b>	<b>63</b>
	<b>Anexo B</b>	<b>65</b>

# Capítulo 1

## Introdução

No estudo de séries temporais, aquelas que apresentam a característica de longa dependência ou longa memória, têm sido muito mencionadas na literatura. A longa dependência tem sido observada em diversas áreas, como na astronomia, economia e hidrologia. Acreditava-se que as observações, desde que separadas por um longo período de tempo, poderiam ser consideradas como sendo independentes. Verificou-se, no entanto, que mesmo para observações distantes, a correlação entre elas não poderia ser considerada desprezível.

A longa dependência de um processo  $\{X_t\}_{t \in \mathbb{Z}}$  pode ser caracterizada no domínio do tempo através da sua função de autocorrelação, que não é absolutamente convergente. No domínio da frequência, a longa dependência é caracterizada através da função densidade espectral, que assume valores ilimitados para frequências próximas de zero. Estudos iniciais envolvendo séries temporais com esta propriedade foram apresentados pelo hidrologista Harold E. Hurst, em 1951, ao estudar a série temporal das vazões do rio Nilo. Um modelo muito utilizado para descrever o comportamento de uma série temporal que possui longa dependência foi proposto por Granger e Joyeux (1980) e Hosking (1981 e 1984), chamado modelo *autoregressivo fracionalmente integrado de médias móveis*, denotado por  $\text{ARFIMA}(p, d, q)$ , onde  $d$  é o parâmetro fracionário e  $p$  e  $q$  são, respectivamente, os graus do polinômio autoregressivo e o de médias móveis (ver Definição 2.16). A longa dependência do processo  $\{X_t\}_{t \in \mathbb{Z}}$  é caracterizada pelo parâmetro  $d$  quando  $d \in (0, 0.5)$ . Alguns autores definem o processo como sendo de memória intermediária quando  $d < 0$  e de curta memória quando  $d = 0$ . Quando  $d = 0$ , o processo  $\text{ARFIMA}(0, d, 0)$  se reduz ao *autoregressivo de médias móveis* de ordens  $p$  e  $q$ , denotado por  $\text{ARMA}(p, q)$ . Em Beran (1994) encontramos um estudo completo dos modelos  $\text{ARFIMA}(p, d, q)$ . Lopes (2008) apresenta um estudo completo sobre estimação e previsão de processos lineares e não lineares com a característica da longa dependência.

Na estimação clássica, diversos métodos são utilizados para estimar os parâmetros do modelo  $\text{ARFIMA}$ . Entre os mais utilizados aparece o método

que pertence a classe paramétrica, proposto por Fox e Taqqu (1986) e Sowel (1992). Esse método está baseado na função de verossimilhança e a estimação de todos os parâmetros do modelo é feita simultaneamente. Na classe semi-paramétrica destaca-se o método de regressão proposto por Geweke Porter-Hudak (1983). Nesse método, o parâmetro  $d$  é calculado em um primeiro passo e os demais parâmetros no segundo passo.

O método clássico para a estimação dos parâmetros de um modelo qualquer considera esses parâmetros como sendo constantes a serem estimadas. Alternativamente ao método clássico, o método Bayesiano permite tratar os parâmetros de um modelo como sendo variáveis aleatórias e devem, portanto, seguir alguma distribuição de probabilidade. Os métodos Bayesianos, na inferência estatística, tem sido muito utilizados como alternativa aos métodos clássicos. Uma característica importante da inferência Bayesiana é que ela permite incorporar uma informação inicial aos parâmetros de um modelo, dada pela função de distribuição *a priori*, a qual traz informações dos parâmetros antes que os dados da série temporal tenham sido analisados. A fundamentação da teoria Bayesiana está no teorema de Bayes, que relaciona a função de distribuição *a priori* com a função de verossimilhança, obtendo assim a função de distribuição *a posteriori*. A partir dessa função de distribuição, toda a inferência sobre os parâmetros pode ser realizada.

O objetivo desse trabalho é apresentar uma abordagem Bayesiana para a estimação dos parâmetros dos modelos  $ARMA(p, q)$  e  $ARFIMA(0, d, 0)$ , assim como apresentar alguns métodos computacionais que são utilizados para obter inferências da função de distribuição *a posteriori*.

No Capítulo 2 desse trabalho, apresentamos alguns conceitos básicos que envolvem a análise de séries temporais. Definimos os processos  $ARMA(p, q)$  e  $ARFIMA(p, d, q)$  e o processo  $ARFIMA(0, d, 0)$  como sendo um caso particular do último. Também são apresentadas as principais propriedades desses processos. No Capítulo 3, apresentamos os estimadores clássicos para o parâmetro de diferenciação  $d$  do modelo  $ARFIMA(p, d, q)$  onde é feita uma breve descrição dos métodos da máxima verossimilhança e daqueles baseados em uma regressão linear simples. A seguir, é feita a descrição da metodologia Bayesiana, onde caracterizamos as funções de distribuição *a priori* e *a posteriori*. A partir daí, apresentamos uma abordagem Bayesiana para o modelo  $ARMA(p, q)$ . Resultados análogos são considerados para o modelo autoregressivo de ordem  $p$  ( $AR(p)$ ), já que este é um caso particular do modelo  $ARMA(p, q)$  quando assumimos  $q = 0$ . Para este último modelo, um estimador é proposto utilizando resultados da metodologia Bayesiana. Na construção do estimador, utilizamos o algoritmo Metropolis-Hasting, que é apresentado no Capítulo 4. Para verificar o desempenho desse estimador foram feitas simulações e os resultados foram comparados com o método clássico da máxima verossimilhança. Além disso, o estimador proposto é utilizado para estimar o parâmetro  $d$  de um processo  $ARFIMA(0, d, 0)$ , con-

siderando a sua representação autoregressiva infinita. Essa representação permite que o processo ARFIMA(0,  $d$ , 0) possa ser aproximado por um modelo autoregressivo de ordem  $m$ . O valor que  $m$  deve assumir, para que se tenha uma boa aproximação, depende do valor do parâmetro fracionário  $d$ . No Anexo B apresentamos um estudo que determina os valores de  $d$  para uma maior ou menor ordem de aproximação. Ainda no Capítulo 3 apresentamos uma análise Bayesiana do processo ARFIMA(0,  $d$ , 0). A partir do vetor de parâmetros para esse modelo, é definida a função de distribuição *a priori*, a função de verossimilhança e, conseqüentemente, a função de distribuição *a posteriori*, juntamente com o desenvolvimento dos cálculos que definem tais funções. O Capítulo 4 apresenta alguns algoritmos computacionais Bayesianos clássicos. Eles desempenham um avanço na inferência Bayesiana já que permitem obter amostras da função de distribuição *a posteriori* de interesse.

No Capítulo 5 apresentamos as conclusões finais, assim como propostas para futuros trabalhos.

# Capítulo 2

## Noções Preliminares

Apresentamos nesse capítulo alguns conceitos fundamentais necessários para o estudo de séries temporais. Uma série temporal é uma sequência de observações de uma determinada variável aleatória, ordenadas ao longo do tempo. Como exemplos de séries temporais podemos citar as medidas da temperatura de uma cidade, registros de nascimentos ou óbitos em uma determinada região ou a taxa de desemprego no país. O número de manchas solares e o volume anual mínimo de rio Nilo, estudados no Capítulo 3, também são exemplos de séries temporais.

Ao estudar séries temporais, queremos construir modelos que permitam explicar o comportamento da série temporal no período observado. Dessa forma, podemos fazer previsões futuras para a série temporal tendo como base os valores passados. Alguns conceitos são importantes para a compreensão desses modelos e são apresentados no início desse capítulo. Iniciamos definindo conceitos básicos, como o de processos estocásticos, estacionaridade, função de autocovariância, de autocorrelação e densidade espectral. Definimos alguns processos estocásticos importantes, como o processo Gaussiano e o processo de ruído branco. Definimos também os processos *autoregressivos de médias móveis* (ARMA) e *autoregressivo fracionalmente integrado de médias móveis* (ARFIMA), assim como suas principais propriedades. Um estudo mais completo pode ser encontrado em Brockwell e Davis (1991) e Wei (1990).

### 2.1 Conceitos e Resultados Básicos

Nesta seção, daremos os conceitos e resultados básicos para este trabalho.

**Definição 2.1** (Processo Estocástico). Um *processo estocástico* é uma família de variáveis aleatórias  $\{X_t\}_{t \in T}$  definidas em um mesmo espaço de probabilidade  $(\Omega, \mathbb{A}, \mathbb{P})$  sendo  $T$  um conjunto de índices (que pode assumir valores em um conjunto discreto ou contínuo),  $\Omega$  o espaço amostral,  $\mathbb{A}$  a  $\sigma$ -álgebra

da classe de eventos aleatórios e  $\mathbb{P} : \mathbb{A} \rightarrow [0; 1]$  a função que associa probabilidade a um evento aleatório qualquer.

**Observação 2.1.** Neste trabalho consideramos os casos em que  $T = \mathbb{Z}$  e  $T = \{1, 2, \dots, n\}$ . Quando  $T = \{1, 2, \dots, n\}$ , temos uma série temporal  $\{X_t\}_{t=1}^n$  que é uma amostra aleatória advinda do processo  $\{X_t\}_{t \in \mathbb{Z}}$ .

Quando desejamos ter uma ideia do grau de dependência entre um número finito de variáveis aleatórias, podemos calcular a função de autocovariância.

**Definição 2.2** (Função de Autocovariância). Seja  $\{X_t\}_{t \in \mathbb{Z}}$  um processo estocástico tal que  $Var(X_t) < \infty$ , para todo  $t \in \mathbb{Z}$ . A *função de autocovariância* do processo, denotada por  $\gamma_X(\cdot, \cdot)$ , é dada por

$$\gamma_X(r, s) = Cov(X_r, X_s) = \mathbb{E}[(X_r - \mathbb{E}(X_r))(X_s - \mathbb{E}(X_s))], \text{ para todo } r, s \in \mathbb{Z}.$$

**Definição 2.3** (Função de Autocorrelação). Seja  $\{X_t\}_{t \in \mathbb{Z}}$  um processo estocástico tal que  $Var(X_t) < \infty$ , para todo  $t \in \mathbb{Z}$ . A *função de autocorrelação* do processo, denotada por  $\rho_X(\cdot, \cdot)$ , é dada por

$$\rho_X(r, s) = \frac{\gamma_X(r, s)}{\sqrt{Var(X_r)}\sqrt{Var(X_s)}}, \text{ para todo } r, s \in \mathbb{Z}.$$

**Definição 2.4** (Processo Estacionário). Um processo  $\{X_t\}_{t \in \mathbb{Z}}$  é dito *estacionário* se e somente se

- (a)  $\mathbb{E}(|X_t|^2) < \infty$ , para todo  $t \in \mathbb{Z}$
- (b)  $\mathbb{E}(X_t) = \mu$ , para todo  $t \in \mathbb{Z}$
- (c)  $\gamma_X(r, s) = \gamma_X(r + t, s + t)$ , para todo  $r, s, t \in \mathbb{Z}$ .

**Observação 2.2.**

1- A estacionariedade definida acima é também chamada *estacionariedade fraca*, *estacionariedade de segunda ordem* ou *estacionariedade de covariância*.

2- Se  $\{X_t\}_{t \in \mathbb{Z}}$  é estacionário então  $\gamma_X(r, s) = \gamma_X(r - s, 0)$ , para todo  $r, s \in \mathbb{Z}$  (basta tomar  $t = -s$  no item (c) da Definição 2.4). Portanto, podemos redefinir a *função de autocovariância* por

$$\gamma_X(k) \equiv \gamma_X(k, 0) = Cov(X_{t+k}, X_t), \text{ para todo } t, k \in \mathbb{Z}. \quad (2.1)$$

3- Se  $\{X_t\}_{t \in \mathbb{Z}}$  é estacionário, então a função de autocorrelação é definida por

$$\rho_X(k) \equiv \frac{\gamma_X(k)}{\gamma_X(0)} = Corr(X_{t+k}, X_t), \text{ para todo } t, k \in \mathbb{Z},$$

onde  $\gamma_X(0) \equiv \gamma_X(0, 0) = Var(X_t)$ .

**Definição 2.5** (Processo Gaussiano). Um processo estocástico  $\{X_t\}_{t \in \mathbb{Z}}$  é dito *Gaussiano* se, para qualquer conjunto  $t_1, t_2, \dots, t_n \in \mathbb{Z}$ , as variáveis aleatórias  $X_{t_1}, X_{t_2}, \dots, X_{t_n}$  têm uma distribuição normal  $n$ -dimensional.

**Definição 2.6** (Ruído Branco). O processo  $\{\varepsilon_t\}_{t \in \mathbb{Z}}$  é denominado um *ruído branco* com média zero e variância  $\sigma_\varepsilon^2$ , denotado por  $\varepsilon_t \sim WN(0, \sigma_\varepsilon^2)$ , se

$$\mathbb{E}(\varepsilon_t) = 0, \quad \text{Var}(\varepsilon_t) = \mathbb{E}(\varepsilon_t^2) = \sigma_\varepsilon^2 \quad \text{e} \quad \gamma_\varepsilon(k) = \begin{cases} \sigma_\varepsilon^2, & k = 0, \\ 0, & k \neq 0. \end{cases} \quad (2.2)$$

**Observação 2.3.** Observe que um processo ruído branco é formado por variáveis aleatórias com média zero, variância finita e constante e não correlacionadas.

Muitas vezes desejamos estimar a função de autocovariância a partir das observações de uma amostra finita  $\{X_t\}_{t=1}^n$  de variáveis aleatórias. Essa estimativa nos dá informações sobre o grau de dependência do processo  $\{X_t\}_{t \in \mathbb{Z}}$ . Para estimar a função de autocovariância  $\gamma_X(\cdot)$ , usamos o estimador definido como segue.

**Definição 2.7** (Função de Autocovariância Amostral). A *função de autocovariância amostral* do processo  $\{X_t\}_{t \in \mathbb{Z}}$  é definida a partir de uma série temporal  $\{X_t\}_{t=1}^n$  e é dada por

$$\hat{\gamma}_X(k) = \frac{1}{n} \sum_{j=1}^{n-k} (X_{j+k} - \bar{X})(X_j - \bar{X}), \quad k = 0, 1, 2, \dots, n-1, \quad (2.3)$$

e  $\hat{\gamma}_X(-k) = \hat{\gamma}_X(k)$ ,  $-n < k \leq 0$ , onde  $\bar{X} = \frac{1}{n} \sum_{j=1}^n X_j$  é a média amostral.

**Definição 2.8** (Função de Autocorrelação Amostral). A *função de autocorrelação amostral* do processo  $\{X_t\}_{t \in \mathbb{Z}}$  é definida a partir de uma série temporal  $\{X_t\}_{t=1}^n$  e é dada por

$$\hat{\rho}_X(k) \equiv \frac{\hat{\gamma}_X(k)}{\hat{\gamma}_X(0)}, \quad k = 0, 1, 2, \dots, n-1,$$

onde  $\hat{\gamma}_X(\cdot)$  é dada pela expressão (2.3).

O teorema a seguir caracteriza a função de autocovariância que pode ser escrita na forma

$$\gamma_X(k) = \int_{(-\pi, \pi]} e^{i\lambda k} dF_X(\lambda),$$

para alguma função  $F_X(\cdot)$  com massa concentrada em  $(-\pi, \pi]$ .

**Teorema 2.1** (Herglotz). *Seja  $\{X_t\}_{t \in \mathbb{Z}}$  um processo estocástico estacionário e seja  $\gamma_X(\cdot)$  sua função de autocovariância, conforme expressão (2.2). Então, a função  $\gamma_X(\cdot)$  é função definida não negativa se e somente se*

$$\gamma_X(k) = \int_{(-\pi, \pi]} e^{i\lambda k} dF_X(\lambda), \quad \text{para todo } k \in \mathbb{Z}, \quad (2.4)$$

onde  $F_X(\cdot)$  é uma função contínua à direita, não decrescente e limitada em  $[-\pi, \pi]$  e  $F_X(-\pi) = 0$ . A função  $F_X(\cdot)$  é denominada função de distribuição espectral de  $\gamma_X(\cdot)$  (ou de  $\{X_t\}_{t \in \mathbb{Z}}$ ). Se  $F_X(\lambda) = \int_{-\pi}^{\lambda} f_X(\nu) d\nu$ ,  $-\pi \leq \lambda \leq \pi$ , então  $f_X(\cdot)$  é denominada função densidade espectral de  $\gamma_X(\cdot)$  (ou de  $\{X_t\}_{t \in \mathbb{Z}}$ ).

**Demonstração:** Se  $\gamma_X(\cdot)$  tem representação dada por (2.4) então  $\gamma_X(\cdot)$  é Hermitiana, ou seja, satisfaz  $\gamma_X(-k) = \overline{\gamma_X(k)}$ . Além disso, se  $a_r \in \mathbb{C}$ , onde  $r = 1, 2, \dots, n$ , então

$$\begin{aligned} \sum_{r,s=1}^n a_r \gamma_X(r-s) \bar{a}_s &= \int_{-\pi}^{\pi} \sum_{r,s=1}^n a_r \bar{a}_s \exp(i\lambda(r-s)) dF_X(\lambda) \\ &= \int_{-\pi}^{\pi} \left| \sum_{r=1}^n a_r \exp(i\lambda r) \right|^2 dF_X(\lambda) \geq 0, \end{aligned}$$

de modo que  $\gamma_X(\cdot)$  também é função definida não negativa.

Reciprocamente, supor que  $\gamma_X(\cdot)$  é uma função definida não negativa. Então, definindo

$$\begin{aligned} f_N(\lambda) &= \frac{1}{2\pi N} \sum_{r,s=1}^N e^{-ir\lambda} \gamma_X(r-s) e^{is\lambda} \\ &= \frac{1}{2\pi N} \sum_{|m| < N} (N - |m|) e^{-im\lambda}, \end{aligned}$$

e utilizando que  $\gamma_X(\cdot)$  é definida não negativa, temos que

$$f_N(\lambda) \geq 0, \quad \text{para todo } \lambda \in (-\pi, \pi].$$

Seja  $F_N(\cdot)$  definida como sendo a função de distribuição correspondente a função densidade  $f_N(\cdot)$ . Assim,

$$F_N(\lambda) = \begin{cases} 0, & \lambda \leq -\pi, \\ \int_{-\pi}^{\lambda} f_N(\lambda) d\lambda, & -\pi \leq \lambda \leq \pi, \\ F_N(\pi), & \lambda \geq \pi. \end{cases} \quad (2.5)$$

Então, para todo inteiro  $k$ ,

$$\int_{(-\pi, \pi]} e^{ik\lambda} dF_N(\lambda) = \frac{1}{2\pi} \sum_{|m| \leq N} \left(1 - \frac{|m|}{N}\right) \gamma_X(m) \int_{-\pi}^{\pi} e^{i(k-m)\lambda} d\lambda, \quad (2.6)$$

ou seja,

$$\int_{(-\pi, \pi]} e^{ik\lambda} dF_N(\lambda) = \begin{cases} \left(1 - \frac{|k|}{N}\right) \gamma_X(k), & |k| < N, \\ 0, & |k| \geq N. \end{cases} \quad (2.7)$$

Podemos aplicar o teorema de Helly (ver Ash, 1972) para deduzir que existe uma função de distribuição  $F_X$  e uma subsequência  $\{F_{N_j}\}$  de uma sequência  $\{F_N\}$  tal que para alguma função contínua limitada  $g(\cdot)$  com  $g(\pi) = g(-\pi)$ , tem-se

$$\int_{(-\pi, \pi]} g(\lambda) dF_{N_j}(\lambda) \rightarrow \int_{(-\pi, \pi]} g(\lambda) dF_X(\lambda), \quad \text{quando } j \rightarrow \infty. \quad (2.8)$$

Substituindo  $N$  por  $N_j$  em (2.7) e considerando o limite  $j \rightarrow \infty$ , obtemos

$$\gamma_X(k) = \int_{(-\pi, \pi]} e^{ik\lambda} dF_X(\lambda), \quad (2.9)$$

a qual é a representação espectral de  $\gamma_X(\cdot)$ .

□

A seguir definimos dois estimadores da função densidade espectral.

**Definição 2.9** (Função Periodograma). Sejam  $\{X_t\}_{t \in \mathbb{Z}}$  um processo estocástico estacionário e  $\{X_t\}_{t=1}^n$  uma série temporal obtida a partir deste processo. A *função periodograma* calculada a partir da série temporal  $\{X_t\}_{t=1}^n$ , é definida por

$$I_n(\lambda) = \frac{1}{2\pi n} \left| \sum_{t=1}^n e^{-i\lambda t} X_t \right|^2, \quad \lambda \in (-\pi, \pi]. \quad (2.10)$$

**Observação 2.4.** Podemos reescrever a expressão (2.10) da forma

$$I_n(\lambda) = \frac{1}{2\pi} \left( \widehat{\gamma}_X(0) + 2 \sum_{k=1}^{n-1} \widehat{\gamma}_X(k) \cos(\lambda k) \right), \quad \lambda \in (-\pi, \pi], \quad (2.11)$$

onde  $\widehat{\gamma}_X(\cdot)$  é a função de autocovariância amostral do processo  $\{X_t\}_{t \in \mathbb{Z}}$  dada na Definição 2.3.

A *função periodograma* é um estimador não-viciado mas, inconsistente, para a função densidade espectral  $f_X(\cdot)$  (ver Brockwell e Davis, 1991; pág. 343). Na próxima definição apresentamos um estimador consistente para a função densidade espectral.

**Definição 2.10** (Função Periodograma Suavizado). Sejam  $\{X_t\}_{t \in \mathbb{Z}}$  um processo estocástico estacionário e  $\{X_t\}_{t=1}^n$  uma série temporal obtida a partir deste processo. A *função periodograma suavizado* calculada a partir da série temporal  $\{X_t\}_{t=1}^n$ , denotada por  $f_s(\cdot)$ , é definida por

$$f_s(\lambda) = \frac{1}{2\pi} \sum_{k=-(n-1)}^{n-1} \kappa\left(\frac{k}{n-1}\right) \widehat{\gamma}_X(k) \cos(\lambda k), \quad \lambda \in (-\pi, \pi],$$

onde  $\widehat{\gamma}_X(\cdot)$  é a função de autocovariância amostral do processo definida em (2.3) e  $\kappa(\cdot)$  é uma função de ponderação, sendo função par e contínua, satisfazendo  $\kappa(0) = 1$ ,  $|\kappa(x)| \leq 1$ , para todo  $x$ , e  $\kappa(x) = 0$ , para todo  $|x| > 1$ .

Um estudo mais completo pode ser encontrado em Brockwell e Davis (1991) e Wei (1990).

## 2.2 Processos ARMA( $p, q$ )

Nessa seção, definimos os processos *autoregressivos de média móvel* de ordens  $p$  e  $q$  para um processo  $\{X_t\}_{t \in \mathbb{Z}}$ , denotados por ARMA( $p, q$ ), que são definidos em termos de equações de diferenças lineares com coeficientes constantes. Um processo ARMA( $p, q$ ) é frequentemente referido como um processo de curta memória pois a função de autocovariância (ou dependência) entre  $X_t$  e  $X_{t+k}$  decresce rapidamente quando  $k \rightarrow \infty$ . De fato, a sua função de autocorrelação é geometricamente limitada, ou seja,

$$|\rho_X(k)| \leq Cr^k, \quad k \in \mathbb{N} - \{0\},$$

onde  $C > 0$  e  $0 < r < 1$ . Dessa forma, a função de autocorrelação de um processo ARMA( $p, q$ ) tem a característica de ser absolutamente somável (ver Brockwell e Davis, 1991).

Apresentamos também algumas propriedades desse processos, como causalidade, invertibilidade e a sua função de densidade espectral.

**Definição 2.11** (Processo ARMA( $p, q$ )). O processo estocástico  $\{X_t\}_{t \in \mathbb{Z}}$  é denominado um *processo autoregressivo de média móvel* com média  $\mu$ , denotado por ARMA( $p, q$ ), se é um processo estacionário satisfazendo as equações

$$\Phi(\mathcal{B})(X_t - \mu) = \Psi(\mathcal{B})\varepsilon_t, \quad \text{para todo } t \in \mathbb{Z}, \quad (2.12)$$

onde  $\mathcal{B}$  é o *operador defasagem*, isto é,  $\mathcal{B}^j(X_t) = X_{t-j}$ , para  $j \in \mathbb{N} \cup \{0\}$ .  $\{\varepsilon_t\}_{t \in \mathbb{Z}}$  é o processo de ruído branco com média zero e variância  $\sigma_\varepsilon^2 > 0$ ,  $\Phi(\cdot)$  e  $\Psi(\cdot)$  são os polinômios de ordens  $p$  e  $q$  dados, respectivamente, por

$$\begin{aligned} \Phi(z) &= 1 - \phi_1 z - \phi_2 z^2 - \dots - \phi_p z^p \\ \Psi(z) &= 1 + \psi_1 z + \psi_2 z^2 + \dots + \psi_q z^q \end{aligned} \quad (2.13)$$

onde  $\phi_i$ ,  $1 \leq i \leq p$ , e  $\psi_j$ ,  $1 \leq j \leq q$  são constantes reais.

Podemos supor, sem perda de generalidade, que  $\mu = 0$ .

**Definição 2.12** (ARMA( $p, q$ ) Causal). Um processo ARMA( $p, q$ ), dado pela expressão (2.12) é denominado *causal* se existe uma sequência de constantes  $\{\pi_j\}_{j \geq 0}$ , tal que

$$\sum_{j=0}^{\infty} |\pi_j| < \infty \quad \text{e} \quad X_t = \sum_{j=0}^{\infty} \pi_j \varepsilon_{t-j}, \quad \text{para todo } t \in \mathbb{Z}. \quad (2.14)$$

**Definição 2.13** (ARMA( $p, q$ ) Inversível). Um processo ARMA( $p, q$ ), definido pela expressão (2.12), é denominado *inversível* se existe uma sequência de constantes  $\{\varphi_j\}_{j \geq 0}$ , tal que

$$\sum_{j=0}^{\infty} |\varphi_j| < \infty \quad \text{e} \quad \varepsilon_t = \sum_{j=0}^{\infty} \varphi_j X_{t-j}, \quad \text{para todo } t \in \mathbb{Z}. \quad (2.15)$$

**Observação 2.5.**

1- A causalidade e a invertibilidade não são apenas propriedades do processo  $\{X_t\}_{t \in \mathbb{Z}}$  sozinho, mas sim da relação entre os processos  $\{X_t\}_{t \in \mathbb{Z}}$  e  $\{\varepsilon_t\}_{t \in \mathbb{Z}}$  que aparecem na Definição 2.11.

2- Se  $p = 0$  na equação (2.12), isto é, se  $\Phi(z) \equiv 1$ , então obtemos  $X_t = \Psi(\mathcal{B})\varepsilon_t$  e  $\{X_t\}_{t \in \mathbb{Z}}$  é dito ser um *processo média móvel de ordem  $q$*  e denotado por MA( $q$ ).

Se  $q = 0$  na equação (2.12), isto é, se  $\Psi(z) \equiv 1$ , então obtemos  $\Phi(\mathcal{B})X_t = \varepsilon_t$  e  $\{X_t\}_{t \in \mathbb{Z}}$  é dito ser um *processo autoregressivo de ordem  $p$*  e denotado por AR( $p$ ).

No teorema, a seguir, apresentamos a função densidade espectral de um processo ARMA( $p, q$ ).

**Teorema 2.2.** *Seja  $\{X_t\}_{t \in \mathbb{Z}}$  um processo estacionário ARMA( $p, q$ ), definido em (2.12), onde  $\Phi(\cdot)$  e  $\Psi(\cdot)$  não possuem raízes em comum e  $\Phi(\cdot)$  não tem raízes no círculo unitário. Então, o processo  $\{X_t\}_{t \in \mathbb{Z}}$  tem função densidade espectral dada por*

$$f_X(\lambda) = \frac{\sigma_\varepsilon^2}{2\pi} \left| \frac{\Psi(e^{-i\lambda})}{\Phi(e^{-i\lambda})} \right|^2, \quad \text{para todo } \lambda \in (-\pi, \pi]. \quad (2.16)$$

**Demonstração:** A solução estacionária do processo ARMA( $p, q$ ), definido em (2.12), pode ser escrita como

$$X_t = \sum_{j \in \mathbb{Z}} \pi_j Z_{t-j}, \quad \text{onde} \quad \sum_{j \in \mathbb{Z}} |\pi_j| < \infty. \quad (2.17)$$

Os coeficientes  $\pi_j$  podem ser calculados a partir da expressão

$$\frac{\Psi(z)}{\Phi(z)} = \sum_{j \in \mathbb{Z}} \pi_j z^j = \Pi(z).$$

Por outro lado, a função densidade espectral do processo  $\{\varepsilon_t\}_{t \in \mathbb{Z}}$  é dada por  $f_\varepsilon(\lambda) = \frac{\sigma_\varepsilon^2}{2\pi}$ . Assim o processo  $\{X_t\}_{t \in \mathbb{Z}}$  possui função densidade espectral.

Fazendo  $U_t = \Phi(\mathcal{B})X_t = \Psi(\mathcal{B})\varepsilon_t$ , obtemos que

$$f_U(\lambda) = |\Phi(e^{-i\lambda})|^2 f_X(\lambda) = |\Psi(e^{-i\lambda})|^2 f_\varepsilon(\lambda). \quad (2.18)$$

Como  $\Phi(\cdot)$  não tem raízes no círculo unitário, temos que  $\Phi(e^{-i\lambda}) \neq 0$ , para todo  $\lambda \in (-\pi, \pi]$ . Logo,

$$f_X(\lambda) = \frac{|\Psi(e^{-i\lambda})|^2}{|\Phi(e^{-i\lambda})|^2} f_\varepsilon(\lambda) = \frac{\sigma_\varepsilon^2}{2\pi} \frac{|\Psi(e^{-i\lambda})|^2}{|\Phi(e^{-i\lambda})|^2},$$

para todo  $\lambda \in (-\pi, \pi]$ .

□

## 2.3 Processos com Longa Dependência

A longa dependência pode ser definida a partir da característica da função de autocorrelação do processo  $\{X_t\}_{t \in \mathbb{Z}}$ . Se a função de autocorrelação  $\gamma_X(\cdot)$  não é absolutamente somável, ou seja, se

$$\sum_{k \in \mathbb{Z}} |\rho_X(k)| = \infty$$

então há dependência significativa entre observações separadas por um grande período de tempo.

Outra forma de se definir um processo com longa dependência é a partir da função densidade espectral que pode assumir valores arbitrariamente grandes para frequências próximas de zero. A seguir introduzimos uma definição mais formal para a propriedade de longa dependência.

**Definição 2.14** (Processos com Longa Dependência). Um processo estacionário  $\{X_t\}_{t \in \mathbb{Z}}$  tem a característica de *longa dependência* (ou de *longa memória*) se

$$\rho_X(k) \sim Ck^{2d-1}, \quad \text{quando } k \rightarrow \infty,$$

onde  $C \in \mathbb{R} - \{0\}$  e  $0 < d < 1/2$ . Equivalentemente, se existe um número  $b \in (0, 1)$  tal que  $f_X(\lambda) \sim C_1|\lambda|^b$ ,  $C_1 > 0$ , onde  $f_X(\cdot)$  é a função densidade espectral do processo, então  $\{X_t\}_{t \in \mathbb{Z}}$  é um *processo estacionário com longa dependência*.

**Observação 2.6.** Alguns autores diferenciam entre processos de memória intermediária quando  $d < 0$  e, portanto,

$$\sum_{k \in \mathbb{Z}} |\rho_X(k)| < \infty,$$

e processos de longa dependência, quando  $0 < d < \frac{1}{2}$  e, neste caso,

$$\sum_{k \in \mathbb{Z}} |\rho_X(k)| = \infty.$$

**Definição 2.15** (Operador Diferença). Seja  $\mathcal{B}$  o *operador defasagem*, isto é,  $\mathcal{B}^j(X_t) = X_{t-j}$ , para  $j \in \mathbb{N} \cup \{0\}$ . Para qualquer número real  $d > -1$  define-se o *operador diferença*  $\nabla^d = (1 - \mathcal{B})^d$  através da expansão binomial

$$\nabla^d = (1 - \mathcal{B})^d = \sum_{j=0}^{\infty} \varphi_j(d) \mathcal{B}^j, \quad (2.19)$$

onde

$$\varphi_j(d) = \frac{\Gamma(j-d)}{\Gamma(j+1)\Gamma(-d)} = \prod_{0 < k \leq j} \frac{k-1-d}{k}, \quad \text{para todo } j \in \mathbb{N}, \quad (2.20)$$

com  $\Gamma(\cdot)$  a função gama definida por

- (i)  $\Gamma(x) = \int_0^{\infty} t^{x-1} e^{-t} dt$ , para todo  $x > 0$
- (ii)  $\Gamma(0) = \infty$
- (iii)  $\Gamma(x+1) = x\Gamma(x)$ , para todo  $x \in \mathbb{R}$ .

A seguir apresentamos uma classe de modelos que caracteriza a longa dependência. Eles foram introduzidos, independentemente, por Granger e Joyeux (1980) e Hosking (1981).

**Definição 2.16** (Processo ARFIMA( $p, d, q$ )). Um processo  $\{X_t\}_{t \in \mathbb{Z}}$  é dito ser um processo ARMA *fracionariamente integrado*, denotado por ARFIMA( $p, d, q$ ), com  $d \in (-0.5, 0.5)$  se, e somente se,  $\{X_t\}_{t \in \mathbb{Z}}$  é estacionário e satisfaz a equação de diferenças

$$\Phi(\mathcal{B})\nabla^d(X_t) = \Psi(\mathcal{B})\varepsilon_t, \quad \text{para todo } t \in \mathbb{Z}. \quad (2.21)$$

O parâmetro  $d$  é o *grau de diferenciação*,  $\nabla^d$  é o operador diferença definido em (2.19),  $\{\varepsilon_t\}_{t \in \mathbb{Z}}$  é o processo ruído branco com média zero e variância  $\sigma_\varepsilon^2 > 0$ ,  $\mathcal{B}$  é o operador defasagem e  $\Phi(\cdot)$  e  $\Psi(\cdot)$  são os polinômios de ordem  $p$  e  $q$ , também chamados de *autoregressivo e média móvel*, respectivamente, dados pela expressão (2.13).

Um caso particular do processo ARFIMA( $p, d, q$ ) é aquele onde  $p = 0 = q$ , ou seja, o ARFIMA( $0, d, 0$ ), também chamado de *processo ARFIMA puro*, o qual satisfaz a equação

$$\nabla^d(X_t) = (1 - \mathcal{B})^d X_t = \varepsilon_t, \quad \text{para todo } t \in \mathbb{Z}. \quad (2.22)$$

O próximo teorema apresenta algumas propriedades do processo estocástico ARFIMA( $0, d, 0$ ), tais como a sua representação nas formas autoregressiva de ordem infinita e média móvel de ordem infinita, denotadas, respectivamente, por AR( $\infty$ ) e MA( $\infty$ ). Além disso, apresenta a função de autocovariância, a função de autocorrelação, a função de autocorrelação parcial e a função densidade espectral do processo  $\{X_t\}_{t \in \mathbb{Z}}$ , quando assumimos que  $d \in (-0.5, 0.5)$ .

**Teorema 2.3.** *Seja  $\{X_t\}_{t \in \mathbb{Z}}$  um processo ARFIMA( $0, d, 0$ ) definido em (2.22). Então,*

- (i) *Quando  $d > -0.5$ ,  $\{X_t\}_{t \in \mathbb{Z}}$  é um processo inversível e apresenta-se como um processo autoregressivo de ordem infinita (AR( $\infty$ )), ou seja,*

$$(1 - \mathcal{B})^d X_t = \sum_{j=0}^{\infty} \varphi_j(d) X_{t-j} = \varepsilon_t, \quad (2.23)$$

onde

$$\varphi_j(d) = \frac{\Gamma(j-d)}{\Gamma(j+1)\Gamma(-d)} = \prod_{0 < k \leq j} \frac{k-1-d}{k}, \quad \text{para todo } j \in \mathbb{N}. \quad (2.24)$$

- (ii) *Quando  $d < 0.5$ ,  $\{X_t\}_{t \in \mathbb{Z}}$  é um processo estacionário e apresenta-se como um processo média móvel de ordem infinita (MA( $\infty$ )), ou seja,*

$$X_t = (1 - \mathcal{B})^{-d} \varepsilon_t = \sum_{j=0}^{\infty} \psi_j(d) \varepsilon_{t-j}, \quad (2.25)$$

onde

$$\psi_j(d) = \frac{\Gamma(j+d)}{\Gamma(j+1)\Gamma(d)} = \prod_{0 < k \leq j} \frac{k-1+d}{k}, \quad \text{para todo } j \in \mathbb{N}. \quad (2.26)$$

Nos itens a seguir assumimos que  $d \in (-0.5, 0.5)$ .

- (iii) *O processo  $\{X_t\}_{t \in \mathbb{Z}}$  possui função de autocovariância dada por*

$$\begin{aligned} \gamma_X(k) &= \sigma_\varepsilon^2 \frac{(-1)^k \Gamma(1-2d)}{\Gamma(k-d+1)\Gamma(1-k-d)}, \quad \text{para todo } k \in \mathbb{N} - \{0\}, \\ \gamma_X(0) &= \sigma_\varepsilon^2 \frac{\Gamma(1-2d)}{\Gamma^2(1-d)}. \end{aligned} \quad (2.27)$$

(iv) O processo  $\{X_t\}_{t \in \mathbb{Z}}$  possui função de autocorrelação dada por

$$\rho_X(k) = \frac{\Gamma(k+d)\Gamma(1-d)}{\Gamma(k-d+1)\Gamma(d)} = \prod_{0 < j \leq k} \frac{j-1+d}{j-d}, \quad \text{para todo } k \in \mathbb{N}. \quad (2.28)$$

(v) O processo  $\{X_t\}_{t \in \mathbb{Z}}$  possui função de autocorrelação parcial dada por

$$\alpha_X(k) = \frac{d}{1-d}, \quad k \in \mathbb{N} - \{0\}. \quad (2.29)$$

(vi) O processo  $\{X_t\}_{t \in \mathbb{Z}}$  possui função densidade espectral dada por

$$f_X(\lambda) = \frac{\sigma_\varepsilon^2}{2\pi} |1 - e^{-i\lambda}|^{-2d} = \frac{\sigma_\varepsilon^2}{2\pi} \left| 2 \operatorname{sen}\left(\frac{\lambda}{2}\right) \right|^{-2d}, \quad \text{para todo } \lambda \in (-\pi, \pi]. \quad (2.30)$$

**Demonstração:** Ver Brockwell e Davis (1991). □

**Observação 2.7.** Pela fórmula de Stirling  $\Gamma(x) \sim \sqrt{2\pi} e^{-x+1} (x-1)^{x-\frac{1}{2}}$ , quando  $x \rightarrow \infty$ , as expressões (2.24), (2.26) e (2.28) podem ser aproximadas por

$$\begin{aligned} \varphi_j(d) &\sim \frac{j^{-d-1}}{\Gamma(-d)}, \quad \text{quando } j \rightarrow \infty \\ \psi_j(d) &\sim \frac{j^{d-1}}{\Gamma(d)}, \quad \text{quando } j \rightarrow \infty \\ \rho_X(k) &\sim k^{2d-1} \frac{\Gamma(1-d)}{\Gamma(d)}, \quad \text{quando } k \rightarrow \infty. \end{aligned}$$

onde “ $\sim$ ” indica que a razão entre os dois termos tende a 1 quando  $j \rightarrow \infty$  ou  $k \rightarrow \infty$ .

**Observação 2.8.** Como  $\operatorname{sen}(\lambda) \sim \lambda$ , quando  $\lambda \rightarrow 0$ , vemos que  $f_X(\lambda) \sim \frac{\sigma_\varepsilon^2}{2\pi} \lambda^{-2d}$ , quando  $\lambda \rightarrow 0$ , ou seja,  $f_X(0)$  é finita se, e somente se,  $d < 0$ .

O próximo teorema estabelece as condições para a existência e unicidade da solução de um processo ARFIMA( $p, d, q$ ), assim como as condições para a sua causalidade e invertibilidade. Além disso, apresenta as suas funções de autocorrelação e densidade espectral.

**Teorema 2.4** (Existência e Unicidade da Solução Estacionária). *Seja  $\{X_t\}_{t \in \mathbb{Z}}$  um processo ARFIMA( $p, d, q$ ) dado em (2.21). Suponha que  $d \in (-0.5, 0.5)$  e que  $\Phi(\cdot)$  e  $\Psi(\cdot)$  não tem zeros em comum.*

(i) Se  $\Phi \neq 0$ , para  $|z| = 1$ , então existe uma única solução estacionária  $\{X_t\}_{t \in \mathbb{Z}}$  para (2.21), dada por

$$X_t = \sum_{j=0}^{\infty} \varphi_j(d) \varepsilon_{t-j} \quad (2.31)$$

onde  $\varphi(z) = \sum_{j=0}^{\infty} \varphi_j(d) z^j = \frac{\Psi(z)}{\Phi(z)} (1 - z)^{-d}$ .

(ii) A solução  $\{X_t\}_{t \in \mathbb{Z}}$  é causal se e somente se  $\Phi(z) \neq 0$ , para todo  $|z| \leq 1$ .

(iii) A solução  $\{X_t\}_{t \in \mathbb{Z}}$  é inversível se e somente se  $\Psi(z) \neq 0$ , para todo  $|z| \leq 1$ .

(iv) Se a solução  $\{X_t\}_{t \in \mathbb{Z}}$  é causal e inversível então a sua função de autocorrelação  $\rho_X(\cdot)$  e a função de densidade espectral  $f_X(\cdot)$  satisfazem, para  $d \neq 0$ ,

$$\rho_X(k) \sim C k^{2d-1}, \quad \text{quando } k \rightarrow \infty$$

onde  $C \neq 0$ , e

$$f_X(\lambda) = \frac{\sigma_\varepsilon^2 |\Psi(e^{-i\lambda})|^2}{2\pi |\Phi(e^{-i\lambda})|^2} |1 - e^{-i\lambda}|^{-2d} \sim \frac{\sigma_\varepsilon^2}{2\pi} \left( \frac{\Psi(1)}{\Phi(1)} \right)^2 \lambda^{-2d},$$

quando  $\lambda \rightarrow 0$ .

**Demonstração:** Ver Brockwell e Davis (1991).

□

## Capítulo 3

# Estimação Bayesiana em Processos de Longa Dependência

Iniciamos este capítulo apresentando estimadores clássicos para o parâmetro de diferenciação  $d$  do processo ARFIMA( $p, d, q$ ). São descritos os métodos de máxima verossimilhança e o método obtido de uma regressão linear simples, utilizando a função periodograma. No que segue, toda a análise estatística desenvolvida considera a metodologia Bayesiana para a estimação dos parâmetros nos modelos AR( $p$ ), ARMA( $p, q$ ) e ARFIMA( $0, d, 0$ ). São caracterizadas as funções de distribuição *a priori*, de verossimilhança e de distribuição *a posteriori*. Para ilustrar a metodologia Bayesiana, apresentamos um exemplo cujo parâmetro é unidimensional e estabelecemos uma estimativa pontual para esse parâmetro.

Apresentamos uma abordagem Bayesiana para o modelo ARMA( $p, q$ ) (ver Sáfyadi e Andrade, 2007). Para o modelo AR( $p$ ), um estimador é proposto utilizando resultados da metodologia Bayesiana. Para avaliar a eficiência desse estimador, são feitas simulações e os resultados são comparados com o método clássico da máxima verossimilhança. Uma aplicação do modelo AR( $p$ ) é feita utilizando uma série temporal com dados reais. O estimador proposto também é utilizado para estimar os parâmetro  $d$  e  $\sigma_\varepsilon^2$  de um processo ARFIMA( $0, d, 0$ ). Nesse caso, o modelo ARFIMA( $0, d, 0$ ) é aproximado por um modelo autoregressivo de ordem  $m$ . Como aplicação para o modelo ARFIMA( $0, d, 0$ ), utilizamos a série temporal que representa o volume anual do rio Nilo, já que esta série temporal apresenta característica de longa dependência (ver Palma, 2007). O final desse capítulo apresenta uma análise Bayesiana do processo ARFIMA( $0, d, 0$ ).

## 3.1 Estimação Clássica

Nesta seção apresentamos alguns métodos de estimação do parâmetro  $d$  do processo ARFIMA( $p, d, q$ ). O primeiro é o método da máxima verossimilhança proposto por Fox e Taqqu (1986), utilizando a aproximação sugerida por Whittle (1953) que é um estimador paramétrico, onde todos os parâmetros são estimados simultaneamente.

Na classe de estimadores semiparamétricos, apresentamos o método da regressão, utilizando a função periodograma, proposto por Geweke e Porter-Hudak (1983). Nesse método de estimação o parâmetro de diferenciação  $d$  é estimado primeiro enquanto que os demais parâmetros são estimados em um segundo passo.

**Definição 3.1** (Estimador Não Viciado, U.M.V.U e Consistente). Sejam  $\{X_t\}_{t \in \mathbb{Z}}$  um processo estocástico estacionário e  $\{X_t\}_{t=1}^n$  uma amostra desse processo. Denotamos por  $\tau(\theta)$  um parâmetro qualquer do processo  $\{X_t\}_{t \in \mathbb{Z}}$  e  $T(X_1, X_2, \dots, X_n)$  um estimador qualquer de  $\tau(\theta)$ . Então,

(a)  $T(X_1, X_2, \dots, X_n)$  é um *estimador não viciado* para  $\tau(\theta)$  se, e somente se,  $\mathbb{E}(T(X_1, X_2, \dots, X_n)) = \tau(\theta)$ .

(b) Se  $Var(T^*(X_1, X_2, \dots, X_n)) \leq Var(T(X_1, X_2, \dots, X_n))$ , para todo estimador não viciado  $\tau(\theta)$ , então  $T^*(X_1, X_2, \dots, X_n)$  é dito ser um estimador *U.M.V.U (Uniformly Minimum Variance Unbiased)* para o parâmetro  $\tau(\theta)$ .

(c) Seja  $\{T_n(X_1, X_2, \dots, X_n)\}_{n \geq 1}$  uma sequência de estimadores para  $\tau(\theta)$ . Então,  $T_n$  é um *estimador consistente em probabilidade* para  $\tau(\theta)$  se, e somente se,

$$\lim_{n \rightarrow \infty} \mathbb{P}(|T_n(X_1, X_2, \dots, X_n) - \tau(\theta)| \leq \epsilon) = 0, \quad (3.1)$$

para todo  $\epsilon > 0$ .

### 3.1.1 Método da Máxima Verossimilhança

O estimador de máxima verossimilhança sugerido por Fox e Taqqu (1986) está baseado na função de verossimilhança aproximada, sugerida por Whittle (1953), e por isso é denotado por  $W$ . O estimador  $W$ , sob certas condições de regularidade é consistente e possui distribuição assintótica normal. Além disso, é não viciado e de mínima variância. Vamos assumir que  $\{X_t\}_{t \in \mathbb{Z}}$  é um processo Gaussiano estacionário com média  $\mu$  e variância  $\sigma^2$ . Além disso,  $f_X(\cdot) = f_X(\cdot, \boldsymbol{\eta})$  é a função densidade espectral do processo  $\{X_t\}_{t \in \mathbb{Z}}$ , onde  $\boldsymbol{\eta} = (\phi_1, \dots, \phi_p, \psi_1, \dots, \psi_q, d, \sigma_\epsilon^2)' \in \mathbb{R}^{p+q+1} \times \mathbb{R}_+$  é o vetor de parâmetros desconhecidos de tamanho  $m = p + q + 2$ .

Através da amostra  $\{X_t\}_{t=1}^n$  do processo  $\{X_t\}_{t \in \mathbb{Z}}$ , podemos estimar os parâmetros do vetor  $\boldsymbol{\eta}$  utilizando a função de verossimilhança, dada por

$$L(\mathbf{x}; \boldsymbol{\eta}) = (2\pi)^{-\frac{n}{2}} |\Sigma_n(\boldsymbol{\eta})|^{-\frac{1}{2}} \exp\left[-\frac{1}{2} \mathbf{x}' (\Sigma_n(\boldsymbol{\eta}))^{-1} \mathbf{x}\right], \quad (3.2)$$

onde  $\mathbf{x} = (x_1, x_2, \dots, x_n)' \in \mathbb{R}^n$ ,  $\mathbf{x}'$  é o vetor transposto do vetor  $\mathbf{x}$  e  $\Sigma_n(\boldsymbol{\eta})$  é uma matriz quadrada  $n \times n$ , da forma

$$\Sigma_n(\boldsymbol{\eta}) = \left[ \gamma_X^{\boldsymbol{\eta}}(i, j) \right]_{i,j=1}^n, \quad (3.3)$$

onde

$$\gamma_X^{\boldsymbol{\eta}}(i, j) = \gamma_X^{\boldsymbol{\eta}}(i - j), \quad \text{para todo } i, j = 1, \dots, n, \quad (3.4)$$

e  $\gamma_X^{\boldsymbol{\eta}}(\cdot) = \gamma_X(\cdot)$  é a função de autocovariância do processo  $\{X_t\}_{t \in \mathbb{Z}}$ , dada pela expressão (2.1).

O logaritmo da função de verossimilhança em (3.2) é dado por

$$\mathcal{L}_n(\mathbf{x}; \boldsymbol{\eta}) = \ln(L(\mathbf{x}; \boldsymbol{\eta})) = -\frac{n}{2} \ln(2\pi) - \frac{1}{2} \ln(|\Sigma_n(\boldsymbol{\eta})|) - \frac{1}{2} \mathbf{x}' [\Sigma_n(\boldsymbol{\eta})]^{-1} \mathbf{x}. \quad (3.5)$$

Assim, obtemos o vetor  $m$ -dimensional

$$\begin{aligned} \mathcal{L}'_n(\mathbf{x}; \boldsymbol{\eta}) &= \left( \frac{\partial}{\partial \boldsymbol{\eta}_j} \mathcal{L}_n(\mathbf{x}; \boldsymbol{\eta}) \right)_{j=1}^m \\ &= \left( -\frac{1}{2} \frac{\partial}{\partial \boldsymbol{\eta}_j} \ln |\Sigma_n(\boldsymbol{\eta})| - \frac{1}{2} \mathbf{x}' \left[ \frac{\partial}{\partial \boldsymbol{\eta}_j} [\Sigma_n(\boldsymbol{\eta})]^{-1} \right] \mathbf{x} \right)_{j=1}^m. \end{aligned} \quad (3.6)$$

O estimador de máxima verossimilhança pode ser obtido maximizando (3.5) com relação ao vetor  $\boldsymbol{\eta}$ . Sob algumas condições de regularidade, isso pode ser feito resolvendo o sistema de  $m$  equações

$$\mathcal{L}_n(\mathbf{x}; \hat{\boldsymbol{\eta}}) = 0,$$

dado pela expressão (3.6), onde  $\hat{\boldsymbol{\eta}}$  é o estimador de máxima verossimilhança de  $\boldsymbol{\eta}$ .

Caso a série temporal  $\{X_t\}_{t=1}^n$  tenha um tamanho amostral relativamente grande (por exemplo, 5000 dados), fica difícil obter a inversa da matriz de autocovariâncias  $[\Sigma(\boldsymbol{\eta})]^{-1}$  pois, nesse caso, será necessário inverter uma matriz quadrada de ordem  $n$ . No entanto, é possível usar uma forma aproximada para a matriz  $[\Sigma(\boldsymbol{\eta})]^{-1}$  estabelecida por Whittle (1953). Utilizamos o método da máxima verossimilhança aproximada para estimar  $\boldsymbol{\eta}$ , maximizando a função

$$L(\mathbf{x}; \boldsymbol{\eta}) = \left( \frac{1}{\sqrt{2\pi\sigma_X}} \right)^n \exp\left\{ -\frac{\mathbf{Z}' \mathbf{A}_n(\boldsymbol{\eta}) \mathbf{Z}}{2n\sigma_X} \right\}, \quad (3.7)$$

ou seja, minimizando

$$\frac{\mathbf{Z}' \mathbf{A}_n(\boldsymbol{\eta}) \mathbf{Z}}{n}, \quad (3.8)$$

onde  $\mathbf{Z} = (X_1 - \bar{X}, \dots, X_n - \bar{X})'$ ,  $\bar{X}$  é a média amostral e

$$\mathbf{A}_n(\boldsymbol{\eta}) = [\alpha(k)]_{k=1}^n \quad (3.9)$$

é a matriz  $n \times n$  proposta por Whittle (1953) para aproximar  $\Sigma_n(\boldsymbol{\eta})$  com

$$\alpha(k) = \frac{1}{(2\pi)^2} \int_{-\pi}^{\pi} \frac{1}{f_X(\lambda; \boldsymbol{\eta})} e^{i\lambda k} d\lambda, \quad (3.10)$$

onde  $f_X(\cdot, \boldsymbol{\eta})$  é a função densidade espectral caracterizada pelo vetor de parâmetros  $\boldsymbol{\eta}$ .

Computacionalmente, o estimador  $W$  é obtido minimizando a forma discreta

$$\mathcal{L}_n(\boldsymbol{\eta}) = \frac{1}{2\pi} \sum_{j=1}^{\lfloor \frac{n}{2} \rfloor} \left( \ln(f_X(\lambda_j, \boldsymbol{\eta})) + \frac{I_n(\lambda_j)}{f_X(\lambda_j, \boldsymbol{\eta})} \right), \quad (3.11)$$

onde  $\boldsymbol{\eta}$  é o vetor de parâmetros desconhecidos,  $[x]$  é a parte inteira de  $x$  e  $\lambda_j = \frac{2\pi j}{n}$ ,  $j \in \{1, \dots, \lfloor \frac{n}{2} \rfloor\}$  são as frequências de Fourier.

A variância assintótica do estimador  $W$  (ver Fox e Taquq, 1986) é dada por

$$Var(W) \sim \frac{6}{\pi^2}. \quad (3.12)$$

### 3.1.2 Método da Regressão Utilizando a Função Periodograma

Nesta sub-seção, apresentamos um método de estimação para o parâmetro de diferenciação  $d$ , proposto por Geweke e Porter-Hudak (1983). Esse estimador está baseado no método de regressão utilizando a função periodograma dada na Definição 2.9.

Seja  $\{X_t\}_{t \in \mathbb{Z}}$  um processo ARFIMA( $p, d, q$ ) dado pela expressão (2.21), com  $d \in (-0.5, 0.5)$ . A função densidade espectral desse processo é dada por

$$f_X(\lambda) = |1 - e^{-i\lambda}|^{-2d} f_U(\lambda), \quad (3.13)$$

onde  $f_U(\cdot)$  é a função densidade espectral do processo ARMA( $p, q$ ), ou seja,

$$\Phi(\mathcal{B})U_t = \Psi(\mathcal{B}) \quad \text{onde} \quad U_t = (1 - \mathcal{B})^d X_t. \quad (3.14)$$

Tomando o logaritmo na expressão (3.13), temos que

$$\ln(f_X(\lambda)) = \ln\left(|1 - e^{-i\lambda}|^{-2d}\right) + \ln(f_U(\lambda)) \quad (3.15)$$

ou, escrevendo de outra maneira,

$$\ln(f_X(\lambda)) = \ln(f_U(0)) - d \ln\left(|1 - e^{-i\lambda}|^2\right) + \ln\left(\frac{f_U(\lambda)}{f_U(0)}\right). \quad (3.16)$$

Substituindo  $\lambda$  pela frequência de Fourier  $\lambda_j = \frac{2\pi j}{n}, j \in \{0, 1, \dots, [\frac{n}{2}]\}$  e adicionando  $\ln(I_n(\lambda_j))$  em ambos os lados da igualdade na expressão (3.16), onde  $I_n(\cdot)$  é a função periodograma (ver Definição 2.9), obtemos

$$\ln(I_n(\lambda_j)) = \ln(f_U(0)) - d \ln\left(|1 - e^{-i\lambda_j}|^2\right) + \ln\left(\frac{I_n(\lambda_j)}{f_X(\lambda_j)}\right) + \ln\left(\frac{f_U(\lambda_j)}{f_U(0)}\right). \quad (3.17)$$

Considerando o limite máximo de  $j$  igual a  $g(n)$ , o qual é escolhido de modo que  $\frac{g(n)}{n} \rightarrow 0$ , quando  $n \rightarrow \infty$ , e  $\lambda_j \leq \lambda_{g(n)}$ , onde  $g(n)$  é pequeno, o termo  $\ln\left(\frac{f_U(\lambda_j)}{f_U(0)}\right)$  é desprezível se comparado com os outros termos do lado direito da equação (3.17). Dessa forma, a equação (3.17) pode ser aproximada por

$$\ln(I_n(\lambda_j)) \simeq \ln(f_U(0)) - d \ln\left(|1 - e^{-i\lambda_j}|^2\right) + \ln\left(\frac{I_n(\lambda_j)}{f_X(\lambda_j)}\right), \quad (3.18)$$

e pode ser reescrita de forma aproximada por uma equação de regressão linear simples dada por

$$y_j = a + bx_j + \varepsilon_j, \quad \text{para todo } j = 1, 2, \dots, g(n) \quad (3.19)$$

onde

$$\begin{aligned} y_j &= \ln(I_n(\lambda_j)), & x_j &= \ln\left(|1 - e^{-i\lambda_j}|^2\right), & \varepsilon_j &= \ln\left(\frac{I_n(\lambda_j)}{f_X(\lambda_j)}\right), \\ a &= \ln(f_U(0)) & \text{e } b &= -d. \end{aligned} \quad (3.20)$$

Podemos utilizar a regressão de mínimos quadrados para encontrar o estimador para o parâmetro  $d$  pelo método proposto por Geweke e Porter-Hudak (1983), denotado por GPH, utilizando a função periodograma. Dessa forma, da expressão (3.19) obtemos o estimador

$$GPH = \frac{\sum_{i=1}^{g(n)} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{g(n)} (x_i - \bar{x})^2}, \quad \text{onde } \bar{x} = \frac{1}{g(n)} \sum_{i=1}^{g(n)} x_i, \quad (3.21)$$

com

$$\mathbb{E}(GPH) \cong d \quad \text{e} \quad \text{Var}(GPH) \cong \frac{\pi^2}{6 \sum_{i=1}^{g(n)} (x_i - \bar{x})^2}. \quad (3.22)$$

## 3.2 Análise e Decisão Bayesiana

Suponhamos que temos um conjunto de dados  $\mathbf{y} = (y_1, y_2, \dots, y_n)'$  e admitimos que eles tenham sido gerados a partir de uma distribuição de probabilidade condicional a qual depende de uma quantidade  $\boldsymbol{\theta}$ , ou seja, temos uma distribuição condicional  $p(\mathbf{y}|\boldsymbol{\theta})$ , onde  $\boldsymbol{\theta} \in \Theta$  pode ser um escalar ou um vetor de parâmetros.

O objetivo é estimar  $\boldsymbol{\theta}$  a partir dos valores conhecidos de  $\mathbf{y}$ , ou saber os valores dessa quantidade que são os mais prováveis de gerarem  $\mathbf{y}$ . Outro interesse ao estudar uma série temporal é encontrar um modelo matemático que nos permita fazer previsões sobre os seus futuros valores.

Antes de observar os valores de  $\mathbf{y}$ , podemos supor uma distribuição de probabilidade para  $\boldsymbol{\theta}$ , denotada por  $p(\boldsymbol{\theta})$  (*a função de distribuição a priori*). Em seguida, combinamos essa informação com os dados através da *função de verossimilhança*, denotada por  $f(\mathbf{y}|\boldsymbol{\theta})$ . Como resultado obtemos a função de *distribuição a posteriori*, denotada por  $p(\boldsymbol{\theta}|\mathbf{y})$ , onde a inferência sobre  $\boldsymbol{\theta}$  será feita em função desta distribuição.

O teorema de Bayes será a base desse procedimento. No enfoque Bayesiano o parâmetro  $\boldsymbol{\theta}$  é considerado uma quantidade aleatória e, caso assumamos valores contínuos num certo intervalo, podemos escrever

$$p(\boldsymbol{\theta}|\mathbf{y}) = \frac{f(\mathbf{y}|\boldsymbol{\theta})p(\boldsymbol{\theta})}{\int_{\Theta} f(\mathbf{y}|\boldsymbol{\theta})p(\boldsymbol{\theta})d\boldsymbol{\theta}}. \quad (3.23)$$

Note que o denominador na expressão (3.23) é uma constante com relação a  $\boldsymbol{\theta}$ , o que nos permite reescrever o teorema de Bayes como

$$p(\boldsymbol{\theta}|\mathbf{y}) \propto f(\mathbf{y}|\boldsymbol{\theta})p(\boldsymbol{\theta}),$$

o que significa que  $p(\boldsymbol{\theta}|\mathbf{y})$  é proporcional ao produto da função de verossimilhança pela função de distribuição a priori de  $\boldsymbol{\theta}$ , ou seja,

$$p(\boldsymbol{\theta}|\mathbf{y}) = C f(\mathbf{y}|\boldsymbol{\theta})p(\boldsymbol{\theta}),$$

onde

$$C = \left( \int_{\Theta} f(\mathbf{y}|\boldsymbol{\theta})p(\boldsymbol{\theta})d\boldsymbol{\theta} \right)^{-1}.$$

O tratamento estatístico clássico pode ser difícil quando envolve questões como estimação de parâmetros e identificação de modelos lineares e não lineares com variáveis *correlacionadas*. Nesse contexto a abordagem Bayesiana pode ser um caminho vantajoso. Do ponto de vista Bayesiano, existe a possibilidade de se obter a função densidade para os parâmetros e, portanto, regiões de maior probabilidade, não limitando-se apenas a estimativas pontuais. O método de estimação clássica baseado na função de verossimilhança considera os parâmetros de um modelo como valores constantes a serem calculados. Já a abordagem Bayesiana considera os parâmetros do modelo como

variáveis aleatórias, e como tais, devem ter uma distribuição de probabilidade.

O método de inferência Bayesiana de um modelo ARMA( $p, q$ ), por exemplo, requer a aplicação de uma função densidade de probabilidade *a priori*  $p(\Phi, \Psi, \tau)$ , onde  $\Phi = (\phi_1, \dots, \phi_p)'$ ,  $\Psi = (\psi_1, \dots, \psi_q)'$  e  $\tau = \frac{1}{\sigma^2}$ . A função densidade *a priori* expressa as incertezas sobre  $(\Phi, \Psi, \tau)$  antes das observações dos dados  $\mathbf{X} = (X_{p+1}, X_{p+2}, \dots, X_n)'$ . Na abordagem Bayesiana a função de verossimilhança  $L(\Phi, \Psi, \tau|\mathbf{X})$  e a função densidade *a priori*  $p(\Phi, \Psi, \tau)$  são combinadas utilizando o Teorema de Bayes, de modo a se obter a função densidade de probabilidade *a posteriori*  $f(\Phi, \Psi, \tau|\mathbf{X})$ , como na equação

$$f(\Phi, \Psi, \tau|\mathbf{X}) = \frac{p(\Phi, \Psi, \tau)p(\mathbf{X}|\Phi, \Psi, \tau)}{p(\mathbf{X})} \quad (3.24)$$

onde  $p(\mathbf{X}|\Phi, \Psi, \tau) = L(\Phi, \Psi, \tau|\mathbf{X})$ ,  $p(\mathbf{X})$  é a constante normalizadora para assegurar que  $\int_{\Xi} f(\Phi, \Psi, \tau|\mathbf{X})d\Phi d\Psi d\tau = 1$ , e  $\Xi$  é um subconjunto do  $\mathbb{R}^{p+q+1}$  que representa o domínio de definição dos parâmetros  $(\Phi, \Psi, \tau)$ . A função densidade *a posteriori* mostra como as variáveis aleatórias  $\Phi$ ,  $\Psi$  e  $\tau$  são distribuídas após a análise de  $\mathbf{X}$ .

A seguir podemos então calcular  $f(\Phi|\mathbf{X})$ ,  $f(\Psi|\mathbf{X})$  e  $f(\tau|\mathbf{X})$ . Essas são as funções densidades *a priori* marginais de cada parâmetro  $\Phi$ ,  $\Psi$  e  $\tau$ , respectivamente, e a partir daí, extrair informações sobre os parâmetros. Podemos, por exemplo, fazer uma estimação pontual para  $\hat{\Psi} = \mathbb{E}[f(\Psi|\mathbf{X})]$  assim como  $Var(\hat{\Psi})$  e o intervalo de confiança  $\mathbb{P}(\psi_{\alpha/2} \leq \psi_j \leq \psi_{1-\alpha/2}) = 1 - \alpha$ , para  $j = 1, \dots, q$ . Dessa forma a abordagem Bayesiana é mais flexível do que o método de máxima verossimilhança pois ela permite que se teste várias hipóteses para a função densidade *a priori*  $p(\Phi, \Psi, \tau)$  e se escolha a que resulta em uma função densidade *a posteriori* aceitável para o problema considerado.

Para ilustrar o método de estimação Bayesiana, consideramos um caso simples onde as variáveis aleatórias são independentes e identicamente distribuídas.

**Exemplo 3.1.** Seja  $Y$  uma variável aleatória com distribuição de Poisson, denotada por  $\mathcal{P}(\theta)$ , com parâmetro  $\theta$ . Assumimos que  $\theta$  é uma variável aleatória com distribuição  $\Gamma(a, b)$  com hiperparâmetros  $a$  e  $b$  conhecidos. Esta é a função de distribuição *a priori* para o parâmetro  $\theta$ . Assim, a função densidade condicional de  $Y$  dado  $\Theta = \theta$  é dada por

$$f(y|\theta) = \frac{\exp(-\theta)\theta^y}{y!}, \quad \text{para todo } y \in \mathbb{N} \cup \{0\} \quad (3.25)$$

e, a função de distribuição *a priori* da variável  $\theta$  é dada por

$$p(\theta) = \frac{b^a}{\Gamma(a)}\theta^{a-1}\exp(-b\theta), \quad \text{para todo } \theta > 0. \quad (3.26)$$

## Estimação Pontual

Considerando  $\mathbf{y} = (y_1, y_2, \dots, y_n)'$  uma amostra aleatória de tamanho  $n$  da distribuição de Poisson com parâmetro  $\theta$ , a função de verossimilhança para  $\theta$  é dada por

$$\begin{aligned} L(\theta; \mathbf{y}) &= \prod_{i=1}^n f(y_i | \theta) = \\ &= \prod_{i=1}^n \frac{\exp(-\theta)\theta^{y_i}}{y_i!} = \frac{\exp(-n\theta)\theta^{n\bar{y}}}{\prod_{i=1}^n y_i!}, \end{aligned} \quad (3.27)$$

onde  $n\bar{y} = \sum_{i=1}^n y_i$ . De (3.23), temos que a função de distribuição *a posteriori* de  $\theta$  é dada por

$$\begin{aligned} p(\theta | \mathbf{y}) &= \frac{p(\theta)L(\theta; \mathbf{y})}{\int_0^\infty p(\theta)L(\theta; \mathbf{y})d\theta} = \\ &= \frac{\theta^{a+n\bar{y}-1} \exp[-(n+b)\theta]}{\int_0^\infty \theta^{a+n\bar{y}-1} \exp[-(n+b)\theta]d\theta}. \end{aligned} \quad (3.28)$$

Sabemos que

$$\int_0^\infty x^{a-1} \exp(-bx)dx = \frac{\Gamma(a)}{b^a}, \quad \text{para todo } a > 0 \text{ e } b > 0. \quad (3.29)$$

A expressão (3.28) pode ser reescrita como

$$p(\theta | \mathbf{y}) = \frac{(n+b)^{a+n\bar{y}} \theta^{a+n\bar{y}-1} \exp[-(n+b)\theta]}{\Gamma(a+n\bar{y})}. \quad (3.30)$$

Isso mostra que a função de distribuição *a posteriori* para  $\theta$  também é uma distribuição gama, mas com parâmetros  $a+n\bar{y}$  e  $n+b$ .

Um estimador pontual para  $\theta$  é dado por  $\hat{\theta} = \mathbb{E}[p(\theta | \mathbf{y})] = \frac{a+n\bar{y}}{n+b}$ .

## Estimação por Intervalo

Seja  $\theta$  um parâmetro unidimensional e assumamos que a distribuição *a posteriori* para  $\theta$  seja unimodal. Um estimador por intervalo Bayesiano, com probabilidade  $1 - \alpha$ , é dado por  $(\theta_*, \theta^*)$ , onde

$$\int_{-\infty}^{\theta_*} p(\theta | \mathbf{y})d\theta = \frac{\alpha}{2} \quad \text{e} \quad \int_{\theta^*}^{\infty} p(\theta | \mathbf{y})d\theta = \frac{\alpha}{2}.$$

O intervalo  $(\theta_*, \theta^*)$  é chamado um *intervalo de credibilidade para  $\theta$  com probabilidade  $1 - \alpha$* .

No procedimento clássico, se  $(y_1, y_2, \dots, y_n)'$  é uma amostra aleatória de uma população com densidade  $f(y_i; \theta)$  e  $S = g(y_1, y_2, \dots, y_n)$  é uma estatística com uma distribuição amostral  $f_S(s; \theta)$ , utilizamos  $f_S(s; \theta)$  para encontrar um intervalo de confiança para  $\theta$ .

Com uma confiança de  $100(1 - \alpha)\%$ , temos  $\mathbb{P}(s_*(\theta) \leq S \leq s^*(\theta)) = 1 - \alpha$  e assim encontramos  $h^*(s)$  e  $h_*(s)$  tal que  $\mathbb{P}(h_*(s) \leq \theta \leq h^*(s)) = 1 - \alpha$ . O intervalo de confiança  $[h_*(s); h^*(s)]$  é interpretado da seguinte forma:  $100(1 - \alpha)\%$  dos intervalos construídos contém o verdadeiro valor de  $\theta$ . Considerar um intervalo de confiança de  $95\%$  para o parâmetro  $\theta$ , por exemplo, significa que se retirássemos da população um grande número de amostras de tamanho  $n$ ,  $95\%$  dessas amostras iriam gerar intervalos de confiança que contém o verdadeiro valor de  $\theta$  enquanto que  $5\%$  delas gerariam intervalos de confiança sem conter o verdadeiro valor de  $\theta$ .

Observamos que os intervalos de credibilidade Bayesianos obtidos da função de distribuição *a posteriori* tem uma interpretação mais simples do que os intervalos de confiança clássicos.

### 3.2.1 Análise Bayesiana de um Modelo ARMA( $p, q$ )

Esta sub-seção apresenta uma abordagem Bayesiana para a estimação dos parâmetros de um processo ARMA( $p, q$ ), apresentado na Definição 2.11. A partir daí será considerado o processo autoregressivo de ordem  $p$ , denotado por AR( $p$ ), que aparece na Observação 2.5 e é um caso particular do processo ARMA( $p, q$ ). Neste modelo,  $\boldsymbol{\eta} = (\phi_1, \phi_2, \dots, \phi_p, \sigma_\varepsilon^2) \in \mathbb{R}^p \times \mathbb{R}_+$  é o vetor que contém os parâmetros a serem estimados. Será útil denotar  $\sigma_\varepsilon^2$  por  $\tau^{-1}$ .

Utilizando os resultados da abordagem Bayesiana para o modelo AR( $p$ ), o objetivo é construir, através do aplicativo R-project, os estimadores para os parâmetros desse modelo. Os resultados obtidos foram comparados, via simulação, com o método tradicional da máxima verossimilhança.

A primeira parte dessa sub-seção apresenta a abordagem Bayesiana para o modelo ARMA( $p, q$ ). Serão definidas a função de distribuição *a priori*, a função de máxima verossimilhança e, conseqüentemente, a função de distribuição *a posteriori*, que contém as informações a respeito dos parâmetros a serem estimados. Depois disso, estudamos o modelo AR( $p$ ) como um caso particular. Um estudo de simulação utilizando a abordagem Bayesiana é feito através dos Métodos de Monte Carlo em Cadeias de Markov a serem estudados (ver Seção 4.3), e os resultados são comparados com o método tradicional da máxima verossimilhança. No final dessa sub-seção, uma série de dados reais é analisada.

### Análise Bayesiana de um Modelo ARMA( $p, q$ )

Seja  $\{X_t\}_{t \in \mathbb{Z}}$  um processo ARMA( $p, q$ ), dado pela expressão (2.12). O vetor de parâmetros do modelo é dado por

$$\boldsymbol{\eta} = (\phi_1, \dots, \phi_p, \psi_1, \dots, \psi_q, \sigma_\varepsilon^2)' \in \mathbb{R}^{p+q} \times \mathbb{R}_+. \quad (3.31)$$

A função de verossimilhança desse processo, a partir de uma série temporal de tamanho  $n$  e condicionada às primeiras  $p$  observações, é dada por

$$\begin{aligned} L(\boldsymbol{\Phi}, \boldsymbol{\Psi}, \tau | X_n) &= \prod_{t=p+1}^n \left( \frac{1}{2\pi} \right)^{\frac{1}{2}} \tau^{\frac{1}{2}} \exp \left\{ -\frac{\tau}{2} \sum_{t=p+1}^n \varepsilon_t^2 \right\} \\ &= (2\pi)^{-\frac{T}{2}} \tau^{\frac{T}{2}} \exp \left\{ -\frac{\tau}{2} \sum_{t=p+1}^n \varepsilon_t^2 \right\} \\ &= C \tau^{\frac{T}{2}} \exp \left\{ -\frac{\tau}{2} \sum_{t=p+1}^n \varepsilon_t^2 \right\}, \end{aligned} \quad (3.32)$$

onde  $C = (2\pi)^{-\frac{T}{2}}$ ,  $T = n - p$  e  $\tau^{-1} = \sigma_\varepsilon^2$ .

O processo estocástico (2.12) pode ser reescrito na forma matricial por

$$\mathbf{X} = \mathbf{Z}\boldsymbol{\Phi} + \mathbf{G}\boldsymbol{\varepsilon}, \quad (3.33)$$

onde  $\mathbf{X} = (X_{p+1}, X_{p+2}, \dots, X_n)'$ ,  $\boldsymbol{\Phi} = (\phi_1, \dots, \phi_p)'$  e  $\boldsymbol{\varepsilon} = (\varepsilon_{p+1}, \varepsilon_{p+2}, \dots, \varepsilon_n)'$ . As matrizes  $\mathbf{Z}$  e  $\mathbf{G}$  são dadas por

$$\mathbf{Z} = \begin{pmatrix} X_p & \cdots & X_1 \\ X_{p+1} & \cdots & X_2 \\ \vdots & \vdots & \vdots \\ X_{n-1} & \cdots & X_{n-p} \end{pmatrix}_{T \times p} \quad (3.34)$$

e

$$\mathbf{G} = \begin{pmatrix} 1 & 0 & 0 & 0 & \cdots & 0 \\ \psi_1 & 1 & 0 & 0 & \cdots & 0 \\ \psi_2 & \psi_1 & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 1 \end{pmatrix}_{T \times T}. \quad (3.35)$$

A função de verossimilhança pode ser reescrita como

$$\begin{aligned} L(\boldsymbol{\Phi}, \boldsymbol{\Psi}, \tau | X_n) &= \frac{C}{|\mathbf{V}|^{\frac{1}{2}}} \tau^{\frac{T}{2}} \exp \{ (\mathbf{X} - \mathbf{Z}\boldsymbol{\Phi})' \mathbf{V} (\mathbf{X} - \mathbf{Z}\boldsymbol{\Phi}) \} \\ &= C_1 \tau^{\frac{T}{2}} |\mathbf{V}|^{\frac{1}{2}} \exp \{ (\mathbf{X} - \mathbf{Z}\boldsymbol{\Phi})' \mathbf{V} (\mathbf{X} - \mathbf{Z}\boldsymbol{\Phi}) \}, \end{aligned} \quad (3.36)$$

onde  $C_1 = \frac{C}{|\mathbf{V}|^{\frac{1}{2}}}$ ,  $C = (2\pi)^{-\frac{T}{2}}$  e  $\mathbf{V} = (\mathbf{G}\mathbf{G}')^{-1}$ . Pode-se notar que, na função de verossimilhança em (3.36), os parâmetros de  $\Psi$  entram apenas na inversa da matriz de covariâncias  $\mathbf{V}$ .

Toda inferência sobre os parâmetros é feita através do conhecimento da função de distribuição *a posteriori*. Sabemos que a função de distribuição *a posteriori* é proporcional ao produto da função de distribuição *a priori* pela função de verossimilhança, de modo que uma função de distribuição *a priori* deve ser utilizada. Vamos assumir que a função de distribuição *a priori* é da forma  $p(\Phi, \Psi, \tau) = C_2\tau^{-1}$ , onde  $\tau^{-1} = \sigma_\varepsilon^2$ , pois temos pouco conhecimento a respeito da distribuição dos parâmetros. A integral da função de distribuição *a priori*  $p(\Phi, \Psi, \tau) = C_2\tau^{-1}$  não é finita. No entanto, podemos usar essa distribuição, já que a função de distribuição *a posteriori* resultante  $p(\Phi, \Psi, \tau|\mathbf{X})$  tem integral igual a 1. Para determinar o valor para a constante  $C_2$ , podemos considerar um subconjunto  $D$  no domínio dos parâmetros  $(\Phi, \Psi, \tau)$  tal que

$$\int_D p(\Phi, \Psi, \tau) d\Phi d\Psi d\tau = 1. \quad (3.37)$$

Então,

$$\int_D p(\Phi, \Psi, \tau) d\Phi d\Psi d\tau = \int_D C_2\tau^{-1} d\Phi d\Psi d\tau = C_2 \int_D \tau^{-1} d\Phi d\Psi d\tau = 1. \quad (3.38)$$

Portanto,

$$C_2 = \left( \int_D \tau^{-1} d\Phi d\Psi d\tau \right)^{-1}. \quad (3.39)$$

Dessa forma, a função de distribuição *a posteriori* é dada por

$$\begin{aligned} p(\Phi, \Psi, \tau|\mathbf{X}) &= C_3 L(\Phi, \Psi, \tau|X_n) p(\Phi, \Psi, \tau) \\ &= C_3 C_1 \tau^{\frac{T}{2}} |\mathbf{V}|^{\frac{1}{2}} \exp \left\{ -\frac{\tau}{2} (\Phi - \hat{\Phi})' \Sigma_G^{-1} (\Phi - \hat{\Phi}) + \nu_G S_G^2 \right\} \\ &\quad C_2 \tau^{-1}, \\ &= C_3 C_2 C_1 \tau^{\frac{T}{2}-1} |\mathbf{V}|^{\frac{1}{2}} \exp \left\{ -\frac{\tau}{2} (\Phi - \hat{\Phi})' \Sigma_G^{-1} (\Phi - \hat{\Phi}) + \nu_G S_G^2 \right\} \end{aligned}$$

onde

$$\begin{aligned} C_3 &= \left( \int L(\Phi, \Psi, \tau|X_n) p(\Phi, \Psi, \tau) d\Phi d\Psi d\tau \right)^{-1}, \\ \hat{\Phi} &= (\mathbf{Z}'\mathbf{V}\mathbf{Z})^{-1} (\mathbf{Z}'\mathbf{V}\mathbf{X}), \quad \Sigma_G^{-1} = \mathbf{Z}'\mathbf{V}\mathbf{Z} \end{aligned}$$

e

$$\nu_G S_G^2 = -\frac{\tau}{2} (\mathbf{X} - \mathbf{Z}\hat{\Phi})' \mathbf{V} (\mathbf{X} - \mathbf{Z}\hat{\Phi}).$$

Considerando a função de distribuição *a posteriori* dada em (3.40), é possível utilizar o algoritmo Amostrador de Gibbs (ver Seção 4.4) para obter amostras

dos parâmetros. Para tanto, será necessário o conhecimento das distribuições condicionais completas desses parâmetros. Para  $\Phi$ , temos que

$$p(\Phi|\Psi, \tau, \mathbf{X}) = C_4 \exp\left\{-\frac{\tau}{2}(\Phi - \hat{\Phi})'\Sigma_G^{-1}(\Phi - \hat{\Phi})\right\},$$

onde

$$C_4 = \left(\int \exp\left\{-\frac{\tau}{2}(\Phi - \hat{\Phi})'\Sigma_G^{-1}(\Phi - \hat{\Phi})\right\}d\Phi\right)^{-1}, \quad (3.40)$$

o que implica que  $\Phi|\Psi, \tau, \mathbf{X} \sim \mathcal{N}(\hat{\Phi}, \Sigma_G)$ . Para o parâmetro  $\tau$ , temos que

$$p(\tau|\Phi, \Psi, \mathbf{X}) = C_5 \tau^{\frac{T}{2}-1} \exp\left\{-\frac{\tau}{2}(\Phi - \hat{\Phi})'\mathbf{Z}'\mathbf{V}\mathbf{Z}(\Phi - \hat{\Phi}) + \nu_G S_G^2\right\},$$

onde

$$C_5 = \left(\int \tau^{\frac{T}{2}-1} \exp\left\{-\frac{\tau}{2}(\Phi - \hat{\Phi})'\mathbf{Z}'\mathbf{V}\mathbf{Z}(\Phi - \hat{\Phi}) + \nu_G S_G^2\right\}d\tau\right)^{-1}. \quad (3.41)$$

Portanto,  $\tau|\Phi, \Psi, \mathbf{X} \sim \Gamma\left(\frac{T}{2}, b\right)$ , onde

$$b = (\Phi - \hat{\Phi})'\mathbf{Z}'\mathbf{V}\mathbf{Z}(\Phi - \hat{\Phi}) + (\mathbf{X} - \mathbf{Z}\hat{\Phi})'\mathbf{V}(\mathbf{X} - \mathbf{Z}\hat{\Phi}),$$

e  $T = n - p$ .

Para gerar as amostras de  $\Psi$ , vamos considerar a matriz  $\mathbf{V} = (\mathbf{G}\mathbf{G}')^{-1}$ . Utilizando a decomposição de Cholesky para  $\mathbf{V}^{-1}$  encontramos  $\mathbf{G}$  e, imediatamente, os valores de  $\Psi$ . Para gerar amostras da matriz  $\mathbf{V}$ , temos

$$p(\mathbf{V}|\Phi, \tau, \mathbf{X}) = C_6 |\mathbf{V}|^{\frac{1}{2}} \exp\{(\Phi - \hat{\Phi})'\mathbf{Z}'\mathbf{V}\mathbf{Z}(\Phi - \hat{\Phi})\},$$

e, portanto,

$$\mathbf{V}|\Phi, \tau, \mathbf{X} \sim \text{Wishart}\left(T + 2, \frac{\tau}{2}(\Phi - \hat{\Phi})'\mathbf{Z}'\mathbf{Z}(\Phi - \hat{\Phi})\right),$$

onde  $\Phi = (\phi_1, \dots, \phi_p)'$ ,  $\mathbf{Z}$  é a matriz  $T \times p$  dada em (3.34),  $T = n - p$  e  $\tau^{-1} = \sigma_\varepsilon^2$ .

### Análise Bayesiana de um Modelo AR( $p$ )

A análise feita para o modelo AR( $p$ ) é similar à análise para o modelo ARMA( $p, q$ ), já que um modelo autoregressivo é um caso particular do modelo ARMA( $p, q$ ), onde  $q = 0$ . Nesse caso, temos que  $\Psi = \mathbf{0}$ , ou seja,  $\psi_j = 0$ , para todo  $j \in \{1, 2, \dots, q\}$ . Então, a matriz  $\mathbf{G}$  é agora dada por

$$\mathbf{G} = \begin{pmatrix} 1 & 0 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & 0 & \dots & 0 \\ 0 & 0 & 1 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \dots & 1 \end{pmatrix}_{T \times T},$$

ou seja,  $\mathbf{G}$  é a matriz identidade  $T \times T$ , o que implica que  $\mathbf{V} = (\mathbf{G}\mathbf{G}')^{-1} = \mathbb{I}_{T \times T}$ . Assim, a matriz  $\mathbf{V}$  poderá ser ignorada em alguns cálculos envolvendo o modelo  $\text{AR}(p)$ .

A função de distribuição *a posteriori* dada em (3.40) pode ser reescrita por

$$p(\Phi, \tau | \mathbf{X}) = C_3 \tau^{\frac{T}{2}} \exp \left\{ -\frac{\tau}{2} (\Phi - \hat{\Phi})' \Sigma_G^{-1} (\Phi - \hat{\Phi}) + \nu_G S_G^2 \right\}, \quad (3.42)$$

onde

$$\begin{aligned} \hat{\Phi} &= (\mathbf{Z}'\mathbf{Z})^{-1} (\mathbf{Z}'\mathbf{X}), \\ \Sigma_G^{-1} &= \mathbf{Z}'\mathbf{Z}, \end{aligned}$$

e

$$\nu_G S_G^2 = -\frac{\tau}{2} (\mathbf{X} - \mathbf{Z}\hat{\Phi})' (\mathbf{X} - \mathbf{Z}\hat{\Phi}).$$

As distribuições condicionais completas para o modelo  $\text{AR}(p)$  são dadas a seguir

$$p(\Phi | \tau, \mathbf{X}) = C_4 \tau^{\frac{T}{2}-1} \exp \left\{ -\frac{\tau}{2} (\Phi - \hat{\Phi})' \Sigma_G^{-1} (\Phi - \hat{\Phi}) \right\},$$

o que implica que  $\Phi | \tau, \mathbf{X} \sim \mathcal{N}(\hat{\Phi}, \Sigma_G)$ . Para o parâmetro  $\tau$ , temos que  $\tau | \Phi, \mathbf{X} \sim \Gamma\left(\frac{T}{2}, b\right)$ , onde

$$b = (\Phi - \hat{\Phi})' \mathbf{Z}'\mathbf{Z} (\Phi - \hat{\Phi}) + (\mathbf{X} - \mathbf{Z}\hat{\Phi})' (\mathbf{X} - \mathbf{Z}\hat{\Phi}).$$

Assim, utilizando a função de distribuição *a posteriori*  $p(\Phi, \tau | \mathbf{X})$ , foram encontradas as condicionais completas  $\Phi | \tau, \mathbf{X} \sim \mathcal{N}(\hat{\Phi}, \Sigma_G)$  e  $\tau | \Phi, \mathbf{X} \sim \Gamma\left(\frac{T}{2}, b\right)$ . A partir daí utilizamos o Amostrador de Gibbs (ver Seção 4.4) para obter amostras desses parâmetros.

## Seleção de Modelos

Os parâmetros do modelo  $\text{AR}(p)$  foram ajustados para diferentes valores de  $p$ . Então, é preciso escolher o valor de  $p$  mais adequado. O modelo deve ter o menor número de parâmetros possível além de ajustar corretamente o conjunto de dados.

Dois critérios muito utilizados para se escolher qual o valor de  $p$  a ser utilizado são: o critério de informação de Akaike (AIC) e o critério Bayesiano de informação (BIC). Esses critérios são dados por:

$$\text{AIC} = \ln(\tau^{-1}) + \frac{2k}{n} \quad (3.43)$$

e

$$\text{BIC} = \ln(\tau^{-1}) + \frac{k \ln(n)}{n}, \quad (3.44)$$

onde  $k = p + 1$  e  $n$  é o tamanho da série temporal. A ordem  $p$  do modelo é escolhida encontrando o valor de  $k$  que minimize (3.43) ou (3.44), conforme o critério a ser utilizado.

### Simulações

Para avaliar a eficiência do estimador vamos realizar simulações e descrever os resultados encontrados. Utilizando a função `arma.sim` presente no R-project, foi considerado  $re = 1000$  replicações de um processo  $AR(p)$ , para  $p \in \{1, 2\}$ . Para o modelo  $AR(1)$  foi utilizado  $\phi_1 = 0.3$  e para o modelo  $AR(2)$ ,  $\phi_1 = 0.3$  e  $\phi_2 = 0.6$ .

Para avaliar a eficiência do estimador proposto através da abordagem Bayesiana e comparar com o estimador presente no R-project, utilizamos suas média, erro quadrático médio (eqm), variância (Var) e vício. Elas são dadas, respectivamente, por

$$\begin{aligned} \text{média} &= \frac{1}{re} \sum_{j=1}^{re} \hat{\theta}_j, & \text{eqm} &= \frac{1}{re} \sum_{j=1}^{re} (\hat{\theta}_j - \theta_0)^2, \\ \text{var} &= \frac{1}{re} \sum_{j=1}^{re} (\hat{\theta}_j - \text{média})^2 & \text{e} & \text{vício} = \text{média} - \theta_0 \end{aligned} \quad (3.45)$$

onde  $\hat{\theta}_j$ , para cada  $j \in \{1, 2, \dots, re\}$ , é o valor do estimador na  $j$ -ésima replicação;  $\theta_0$  é o valor nominal do parâmetro. Os resultados dessas simulações são apresentados nas Tabelas 3.1 e 3.2.

**Tabela 3.1:** Resultado da Estimação Clássica e Bayesiana de um Modelo  $AR(1)$ , com  $\phi_1 = 0.3$ .

Clássica				Bayesiana			
$\hat{\phi}_1$	vício( $\hat{\phi}_1$ )	Var( $\hat{\phi}_1$ )	eqm( $\hat{\phi}_1$ )	$\hat{\phi}_1$	vício( $\hat{\phi}_1$ )	Var( $\hat{\phi}_1$ )	eqm( $\hat{\phi}_1$ )
0.29725	0.00275	0.00095	0.00096	0.29802	0.00198	0.00094	0.00095

**Tabela 3.2:** Resultado da Estimação Clássica e Bayesiana de um Modelo  $AR(2)$ , com  $\phi_1 = 0.3$  e  $\phi_2 = 0.6$ .

Clássica							
$\hat{\phi}_1$	vício( $\hat{\phi}_1$ )	Var( $\hat{\phi}_1$ )	eqm( $\hat{\phi}_1$ )	$\hat{\phi}_2$	vício( $\hat{\phi}_2$ )	Var( $\hat{\phi}_2$ )	eqm( $\hat{\phi}_2$ )
0.29763	0.00237	0.00067	0.00070	0.59484	0.00516	0.00071	0.00073
Bayesiana							
$\hat{\phi}_1$	vício( $\hat{\phi}_1$ )	Var( $\hat{\phi}_1$ )	eqm( $\hat{\phi}_1$ )	$\hat{\phi}_2$	vício( $\hat{\phi}_2$ )	Var( $\hat{\phi}_2$ )	eqm( $\hat{\phi}_2$ )
0.29872	0.00128	0.00066	0.00069	0.59561	0.00439	0.00065	0.00067

De acordo com os valores encontrados nas Tabelas 3.1 e 3.2, é possível verificar que, em ambos os modelos  $AR(1)$  e  $AR(2)$ , os resultados foram um

pouco melhores quando o estimador proposto foi utilizado, já que a média está mais próxima do valor nominal do parâmetro e os valores para o erro quadrático médio e para a variância são um pouco menores.

A seguir, apresentamos uma aplicação para o modelo  $AR(p)$ . O valor de  $p$  é escolhido utilizando os critérios de seleção de modelos AIC e BIC, como está descrito no que segue.

### Aplicação: Manchas Solares

Para aplicar o modelo  $AR(p)$ , proposto nesta seção através de uma abordagem Bayesiana, apresentamos a análise de uma série temporal de dados reais. A série temporal escolhida possui  $n = 100$  e expressa o número de manchas solares no período entre os anos 1770 a 1869 (ver Figura 3.1). A série temporal das manchas solares descreve a atividade presente na superfície do sol. Essa atividade influencia no campo magnético e também no clima da Terra.

O modelo utilizado na análise da série temporal é o  $AR(p)$ , onde  $p \in \{1, 2, 3, 4\}$ .

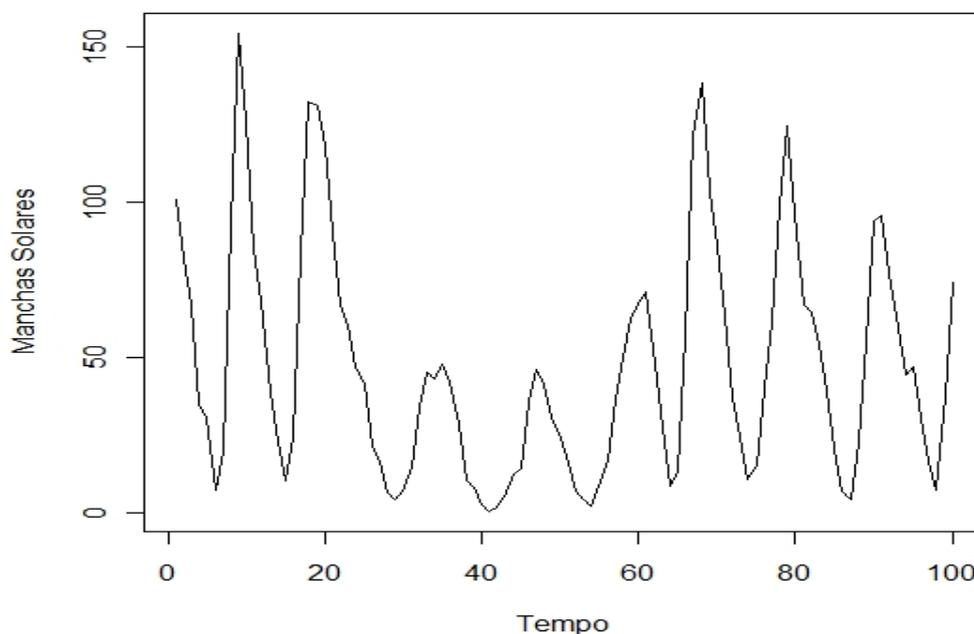
Para encontrar o valor de  $p$  mais apropriado são utilizadas as fórmulas para AIC e BIC dadas em (3.43) e (3.44), respectivamente. A forma utilizada para os cálculos dos AIC e BIC considera os métodos clássico e Bayesiano. Os resultados são apresentados nas Tabelas 3.3 e 3.4 e mostram que o valor  $p = 3$  é o que minimiza ambos os critérios, considerando o método clássico ou Bayesiano. Com base nesses resultados, o melhor modelo que descreve a série temporal das manchas solares, para o período considerado, é um  $AR(3)$ .

**Tabela 3.3:** Resultado da Seleção do Modelo  $AR(p)$  pelo Método Clássico, onde  $p \in \{1, 2, 3, 4\}$ .

$p$	$\hat{\tau}$	$\hat{\sigma}_\varepsilon^2$	AIC	BIC
1	0.00217	460.83	6.16992	6.22202
2	0.00402	248.76	5.57463	5.65279
3	0.00457	218.70	<b>5.46770</b>	<b>5.57190</b>
4	0.00454	220.26	5.49362	5.62388

**Tabela 3.4:** Resultado da Seleção do Modelo AR( $p$ ) pelo Método Bayesiano, onde  $p \in \{1, 2, 3, 4\}$ .

$p$	$\hat{\tau}$	$\hat{\sigma}_\varepsilon^2$	AIC	BIC
1	0.00204	490.20	6.23093	6.283033
2	0.00323	309.60	5.79398	5.87214
3	0.00385	259.70	<b>5.63952</b>	<b>5.74373</b>
4	0.00389	257.07	5.64868	5.77894



**Figura 3.1:** Série Temporal das Manchas Solares, no período de 1770 a 1869.

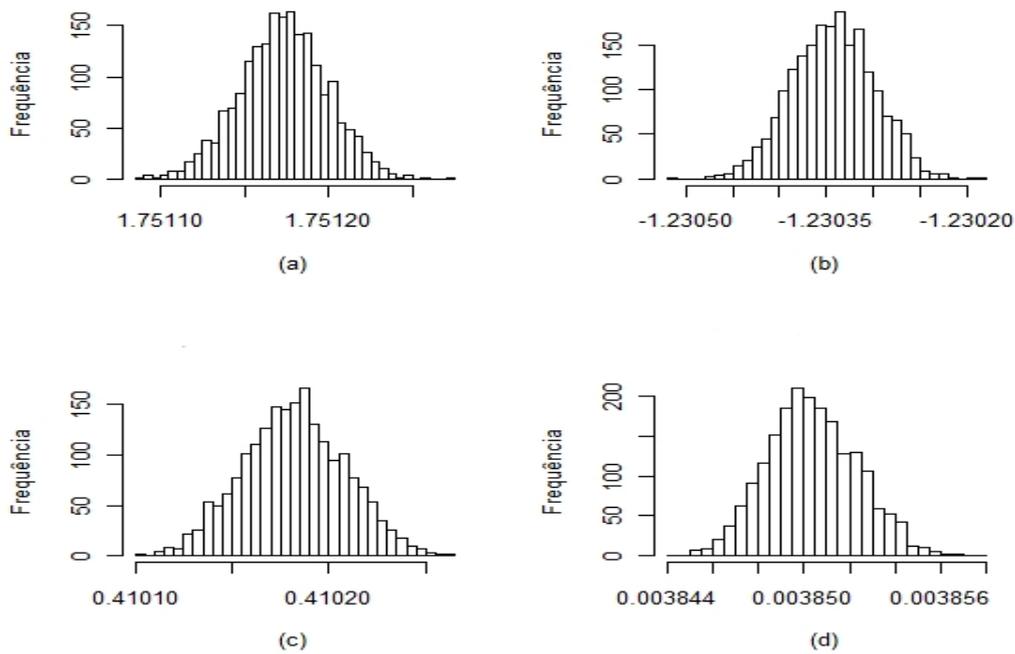
Os parâmetros do modelo AR(3) para a série temporal das manchas solares, foram estimados através do método clássico e Bayesiano e os valores encontrados estão na Tabela 3.5.

A Figura 3.2 mostra os histogramas das amostras estimadas para cada um dos parâmetros do modelo AR(3). Para obter esses histogramas, o estimador Bayesiano foi utilizado e cada parâmetro foi estimado 2000 vezes.

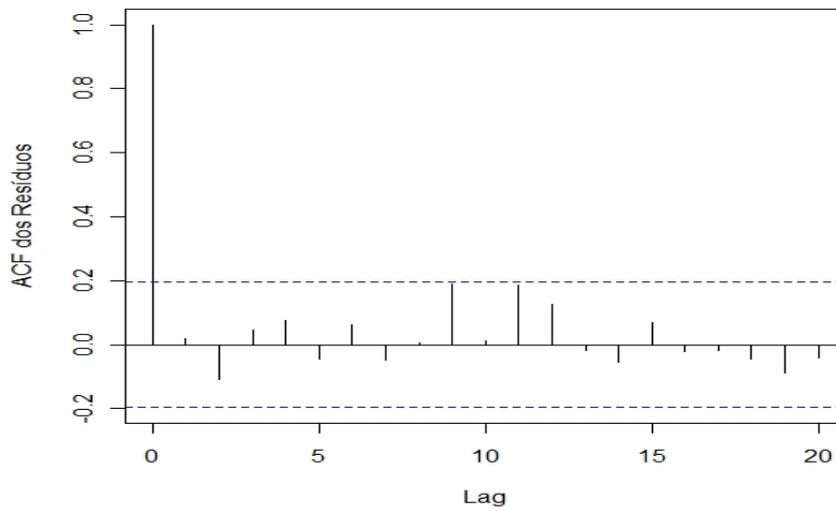
Os gráficos das funções de autocorrelação dos resíduos, apresentados nas nas Figuras 3.3 e 3.4, foram realizados considerando os metodos clássico e Bayesiano, respectivamente. Analisando os gráficos, é possível verificar que

**Tabela 3.5:** Resultado da Estimação Clássica e Bayesiana através de um Modelo AR(3) para a Série Temporal das Manchas Solares.

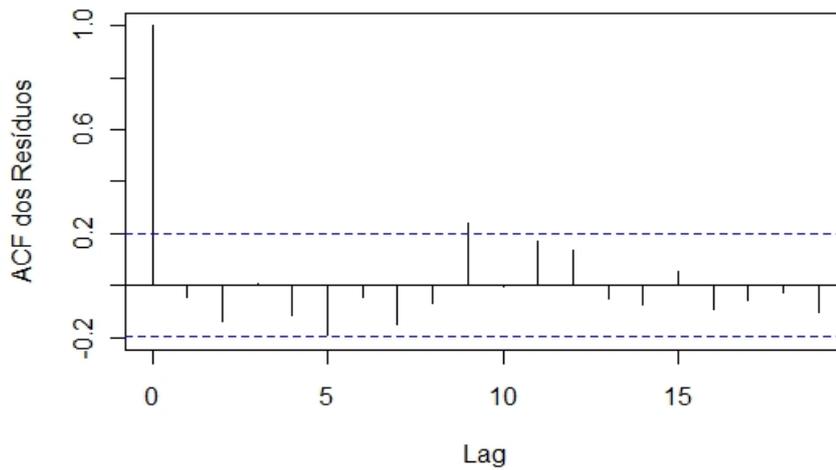
Clássica				
$\hat{\phi}_1$	$\hat{\phi}_2$	$\hat{\phi}_3$	$\hat{\tau}$	$\hat{\sigma}_\varepsilon^2$
1.5512	-0.9984	0.2043	0.00457	218.7
Bayesiana				
$\hat{\phi}_1$	$\hat{\phi}_2$	$\hat{\phi}_3$	$\hat{\tau}$	$\hat{\sigma}_\varepsilon^2$
1.7511	-1.2303	0.4102	0.00385	259.7



**Figura 3.2:** Histogramas das Amostras para cada um dos Estimadores: (a)  $\hat{\phi}_1$ ; (b)  $\hat{\phi}_2$ ; (c)  $\hat{\phi}_3$  e (d)  $\hat{\tau}$ .



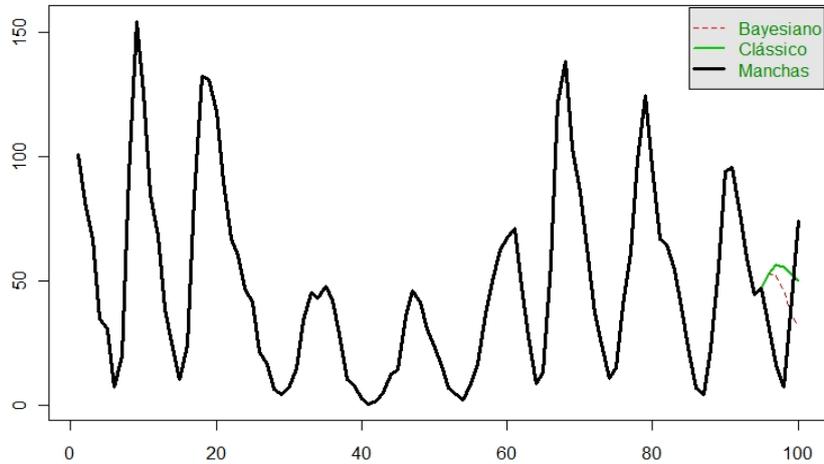
**Figura 3.3:** Função de Autocorrelação dos Resíduos Utilizando a Estimação Clássica.



**Figura 3.4:** Função de Autocorrelação dos Resíduos Utilizando a Estimação Bayesiana.

os resíduos estimados são praticamente não correlacionados. Isso indica que o modelo AR(3) está adequado.

A Figura 3.5 mostra a a série temporal original das manchas solares e a previsão de 5 passos à frente, com base no modelo AR(3). Nesse caso, a série temporal original com  $n = 100$  foi considerada e os seus cinco últimos valores foram retirados desta série. A partir da nova série temporal de tamanho  $n = 95$ , foi feita uma previsão para os próximos cinco valores da série, considerando os métodos clássico e Bayesiano.



**Figura 3.5:** Série Temporal das Manchas Solares e a Previsão para os Cinco Últimos Valores nos Casos Clássico e Bayesiano.

A Seção 3.3, a seguir, apresenta a teoria de estimação sob o enfoque Bayesiano para os processos ARFIMA(0,  $d$ , 0).

### 3.3 Análise Bayesiana de um Processo ARFIMA(0, $d$ , 0)

A análise desenvolvida para o modelo ARFIMA(0,  $d$ , 0) está baseada na função de verossimilhança aproximada resultante do truncamento do processo ARFIMA(0,  $d$ , 0) por sua representação média móvel infinita dada em (2.25), com coeficientes  $\psi_j(d)$ , para  $j = 0, 1, \dots, m$ , dados em (2.26).

Para a análise Bayesiana desse processo desejamos calcular a função de distribuição conjunta *a posteriori* de todos os parâmetros envolvidos para, a partir daí, obter as informações sobre esses parâmetros.

Será necessário definir a função de verossimilhança e a função de distribuição *a priori* do processo e, conseqüentemente, a função de distribuição *a posteriori*.

### Função de Verossimilhança de um ARFIMA(0, d, 0)

Antes de apresentar a função de verossimilhança  $p(\mathbf{X}|\boldsymbol{\eta})$  de um processo ARFIMA(0, d, 0), vamos considerar uma função de distribuição conjunta qualquer de um vetor de variáveis aleatórias  $\mathbf{X} = (X_1, X_2, \dots, X_n)'$ . Se as variáveis aleatórias são independentes e possuem função densidade  $f_{X_i}(\cdot)$ , para  $i = 1, 2, \dots, n$ , então

$$f_{\mathbf{X}}(x_1, \dots, x_n) = \prod_{i=1}^n f_{X_i}(x_i), \quad \text{onde } (x_1, \dots, x_n)' \in \mathbb{R}^n, \quad (3.46)$$

é a função de distribuição conjunta das variáveis aleatórias  $X_1, X_2, \dots, X_n$ .

Quando as variáveis aleatórias não são independentes, é possível calcular a sua função de distribuição conjunta através das funções de distribuição condicionais. Nesse caso, a função de distribuição conjunta é dada por

$$\begin{aligned} f_{\mathbf{X}}(x_1, \dots, x_n) &= f_{X_1}(x_1) f_{X_2|X_1}(x_2|x_1) f_{X_3|X_1, X_2}(x_3|x_1, x_2) \cdots \\ &\quad f_{X_n|X_1, X_2, \dots, X_{n-1}}(x_n|x_1, x_2, \dots, x_{n-1}) \\ &= f_{X_1}(x_1) \prod_{i=2}^n f_{X_i|X_1, \dots, X_{i-2}, X_{i-1}}(x_i|x_1, \dots, x_{i-1}). \end{aligned} \quad (3.47)$$

Se o processo  $\{X_t\}_{t \in \mathbb{Z}}$  é Gaussiano, então a função de distribuição conjunta  $f_{\mathbf{X}}(x_1, \dots, x_n)$  tem distribuição normal multivariada. Nesse caso, as funções de distribuições condicionais dadas em (3.47) também podem ser calculadas. Um caso particular onde  $n = 2$  está dado no Anexo A.

Neste texto, a função de verossimilhança do processo ARFIMA(0, d, 0) é definida de acordo com as aproximações de ordem  $m$ , ou seja, a partir da aproximação média móvel de ordem  $m$ , dada na forma

$$X_t = \sum_{j=0}^m \psi_j(d) \varepsilon_{t-j}, \quad (3.48)$$

onde  $\varepsilon_t \sim \mathcal{N}(0, \sigma_\varepsilon^2)$ , para todo  $t = 1, 2, \dots, n$ . Os coeficientes  $\psi_j(d)$ ,  $j = 1, 2, \dots, m$ , são dados em (2.26).

Consideramos  $\boldsymbol{\varepsilon}_0 = (\varepsilon_0, \varepsilon_{-1}, \dots, \varepsilon_{1-m})'$  o vetor com  $m$  variáveis auxiliares e  $\mathbf{X} = (X_1, X_2, \dots, X_n)'$  uma série temporal de tamanho  $n$ . A expressão (3.48) pode ser reescrita na forma

$$X_t = \varepsilon_t + \sum_{j=1}^m \psi_j(d) \varepsilon_{t-j} = \varepsilon_t + \mu_t(d), \quad \text{para todo } t = 1, 2, \dots, n, \quad (3.49)$$

onde

$$\mu_t(d) = \sum_{j=1}^m \psi_j(d) \varepsilon_{t-j}. \quad (3.50)$$

Como existe dependência entre as variáveis do processo, a função de verossimilhança deve ser calculada através do produto das funções de distribuição condicionais, dada por

$$p(\mathbf{X}|\boldsymbol{\eta}) = p(X_1|\boldsymbol{\eta})p(X_2|X_1, \boldsymbol{\eta}) \cdots p(X_n|X_1, X_2, \dots, X_{n-1}, \boldsymbol{\eta}), \quad (3.51)$$

onde  $\boldsymbol{\eta} = (d, \sigma_\varepsilon^2, \boldsymbol{\varepsilon}_0)$  é o vetor de parâmetros do modelo. Para calcular  $p(X_1|\boldsymbol{\eta})$ , que aparece na expressão (3.51), basta ver que assumindo  $\boldsymbol{\eta}$  conhecido,  $\mu_1(d) = \sum_{j=1}^m \psi_j(d) \varepsilon_{1-j}$  também é conhecido. Por outro lado,  $\varepsilon_1$  é desconhecido e deve ser tratado como uma variável aleatória que tem distribuição normal com média zero e variância  $\sigma_\varepsilon^2$ . Assim, utilizando a expressão (3.49), temos que

$$X_1 = \varepsilon_1 + \sum_{j=1}^m \psi_j(d) \varepsilon_{1-j} = \varepsilon_1 + \mu_1(d). \quad (3.52)$$

Além disso,

$$\begin{aligned} \mathbb{E}(X_1) &= \mathbb{E}(\varepsilon_1 + \mu_1(d)) = \mathbb{E}(\varepsilon_1) + \mathbb{E}(\mu_1(d)) = \mu_1(d) \\ \text{e } \text{Var}(X_1) &= \text{Var}(\varepsilon_1 + \mu_1(d)) = \text{Var}(\varepsilon_1) + \text{Var}(\mu_1(d)) = \sigma_\varepsilon^2. \end{aligned}$$

Portanto,  $X_1|\boldsymbol{\eta} \sim \mathcal{N}(\mu_1(d), \sigma_\varepsilon^2)$ .

Para calcular  $p(X_2|X_1, \boldsymbol{\eta})$ , assumimos  $\boldsymbol{\eta}$  e  $X_1$  dados. Então, também é conhecido  $\varepsilon_1 = X_1 - \mu_1(d)$ . Decorre que  $\mu_2(d) = \sum_{j=1}^m \psi_j(d) \varepsilon_{2-j}$  é conhecido. Por outro lado,  $\varepsilon_2 \sim \mathcal{N}(0, \sigma_\varepsilon^2)$ . Utilizando expressão (3.49), temos que

$$X_2 = \varepsilon_2 + \sum_{j=1}^m \psi_j(d) \varepsilon_{2-j}. \quad (3.53)$$

Assim,

$$\begin{aligned} \mathbb{E}(X_2) &= \mathbb{E}(\varepsilon_2 + \mu_2(d)) = \mathbb{E}(\varepsilon_2) + \mathbb{E}(\mu_2(d)) = \mu_2(d) \\ \text{e } \text{Var}(X_2) &= \text{Var}(\varepsilon_2 + \mu_2(d)) = \text{Var}(\varepsilon_2) + \text{Var}(\mu_2(d)) = \sigma_\varepsilon^2. \end{aligned}$$

Portanto,  $X_2|\boldsymbol{\eta} \sim \mathcal{N}(\mu_2, \sigma_\varepsilon^2)$ .

Seguindo esse procedimento, temos que a sequência  $\varepsilon_1, \dots, \varepsilon_n$  pode ser calculada a partir dos valores  $X_1, \dots, X_n$ , através da equação

$$X_t = \varepsilon_t + \sum_{j=1}^m \psi_j(d) \varepsilon_{t-j}, \quad \text{para } t = 1, 2, \dots, n. \quad (3.54)$$

Se  $1 < t \leq n$ , podemos calcular  $p(X_t|X_1, X_2, \dots, X_{t-1}, \boldsymbol{\eta})$  observando que o valor  $\mu_t(d)$  dado na expressão (3.50) é conhecido. Então,

$$\begin{aligned}\mathbb{E}(X_t) &= \mathbb{E}(\varepsilon_t + \mu_t(d)) = \mathbb{E}(\varepsilon_t) + \mathbb{E}(\mu_t(d)) = \mu_t(d) \\ \text{e } \text{Var}(X_t) &= \text{Var}(\varepsilon_t + \mu_t(d)) = \text{Var}(\varepsilon_t) + \text{Var}(\mu_t(d)) = \sigma_\varepsilon^2.\end{aligned}$$

Portanto,  $X_t|\boldsymbol{\eta} \sim \mathcal{N}(\mu_t(d), \sigma_\varepsilon^2)$ . Logo, a função de verossimilhança do processo é dada por

$$\begin{aligned}p(\mathbf{X}|\boldsymbol{\eta}) &= p(X_1|\boldsymbol{\eta})p(X_2|X_1, \boldsymbol{\eta}) \cdots p(X_n|X_1, X_2, \dots, X_{n-1}, \boldsymbol{\eta}) = \\ &= \prod_{t=1}^n (2\pi\sigma_\varepsilon^2)^{-\frac{1}{2}} \exp\left[-\frac{1}{2\sigma_\varepsilon^2}(X_t - \mu_t)^2\right] \\ &= (2\pi\sigma_\varepsilon^2)^{-\frac{n}{2}} \exp\left[-\frac{1}{2\sigma_\varepsilon^2} \sum_{t=1}^n (X_t - \mu_t)^2\right].\end{aligned}$$

### Função de Distribuição *a priori* de um ARFIMA(0, $d$ , 0)

Como não temos informações a respeito das funções de distribuição *a priori* de cada parâmetro do modelo ARFIMA(0,  $d$ , 0), elas são escolhidas de tal forma que tenha o mínimo de influência sobre a função de distribuição *a posteriori* resultante. Dessa forma, grande parte da informação está baseada na função de verossimilhança. As funções de distribuição *a priori* devem respeitar as regiões onde estão definidos os parâmetros do modelo. Desta forma, utilizamos uma função de distribuição *a priori* para o parâmetro  $d$  uniforme no intervalo  $(-0.5, 0.5)$ . Além disso, vamos supor que os parâmetros são independentes. As funções de distribuição *a priori* são dadas como segue

$p(d) = \mathbf{1}_{(-0.5, 0.5)}(d)$ , onde  $\mathbf{1}_{(-0.5, 0.5)}(d)$  é igual a 1 se  $d \in (-0.5, 0.5)$  e 0 caso contrário;

$\boldsymbol{\varepsilon}_0|\sigma_\varepsilon^2 \sim \mathcal{N}(\mathbf{0}, \sigma_\varepsilon^2\mathbb{I}_m)$ , onde  $\mathbb{I}_m$  denota a matriz identidade  $m \times m$ .

Então,  $\boldsymbol{\varepsilon}_0|\sigma_\varepsilon^2$  tem distribuição normal multivariada cuja função densidade de probabilidade conjunta é da forma

$$\begin{aligned}p(\boldsymbol{\varepsilon}_0|\sigma_\varepsilon^2) &= \frac{1}{(2\pi)^{\frac{m}{2}} \sqrt{\det(\sigma_\varepsilon^2\mathbb{I}_m)}} \exp\left[-\frac{1}{2} \left(\boldsymbol{\varepsilon}_0' (\sigma_\varepsilon^2\mathbb{I}_m)^{-1} \boldsymbol{\varepsilon}_0\right)\right] \\ &= \frac{1}{(2\pi)^{\frac{m}{2}} (\sigma_\varepsilon^2)^{\frac{m}{2}}} \exp\left[-\frac{1}{2} (\varepsilon_0, \dots, \varepsilon_{1-m})' \frac{1}{\sigma_\varepsilon^2} \mathbb{I}_m (\varepsilon_0, \dots, \varepsilon_{1-m})\right]\end{aligned}$$

$$= (2\pi\sigma_\varepsilon^2)^{-\frac{m}{2}} \exp\left[-\frac{1}{2\sigma_\varepsilon^2} \sum_{j=1}^m \varepsilon_{1-j}^2\right];$$

$\sigma_\varepsilon^2 \sim GI(a, b)$ , ou seja,  $\sigma_\varepsilon^2$  é uma variável aleatória cuja função densidade de probabilidade é da forma

$$p(\sigma_\varepsilon^2) = \frac{(\sigma_\varepsilon^2)^{-a-1} b^a}{\Gamma(a)} \exp\left(-\frac{b}{\sigma_\varepsilon^2}\right).$$

Vamos admitir que  $\sigma_\varepsilon^2$ ,  $\varepsilon_0|\sigma_\varepsilon^2$  e  $d$  são independentes e portanto a função de distribuição conjunta *a priori* é da forma

$$\begin{aligned} p((\varepsilon_0|\sigma_\varepsilon^2), \sigma_\varepsilon^2, d) &= p(\varepsilon_0|\sigma_\varepsilon^2)p(\sigma_\varepsilon^2)p(d) = \\ &= \frac{1}{(2\pi)^{\frac{m}{2}} \sqrt{\det(\sigma_\varepsilon^2 \mathbb{I}_m)}} \exp\left[-\frac{1}{2}(\varepsilon_0'(\sigma_\varepsilon^2 \mathbb{I}_m)^{-1}\varepsilon_0)\right] \\ &\frac{(\sigma_\varepsilon^2)^{-a-1} b^a}{\Gamma(a)} \exp\left(-\frac{b}{\sigma_\varepsilon^2}\right) \mathbf{1}_{(-0.5, 0.5)}(d) \\ &= (2\pi\sigma_\varepsilon^2)^{-\frac{m}{2}} \exp\left[-\frac{1}{2\sigma_\varepsilon^2} \sum_{j=1}^m \varepsilon_{1-j}^2\right] \frac{(\sigma_\varepsilon^2)^{-a-1} b^a}{\Gamma(a)} \exp\left(-\frac{b}{\sigma_\varepsilon^2}\right) \mathbf{1}_{(-0.5, 0.5)}(d). \\ &= (2\pi\sigma_\varepsilon^2)^{-\frac{m}{2}} \frac{(\sigma_\varepsilon^2)^{-a-1} b^a}{\Gamma(a)} \exp\left[-\frac{1}{2\sigma_\varepsilon^2} \sum_{j=1}^m \varepsilon_{1-j}^2 - \frac{b}{\sigma_\varepsilon^2}\right] \mathbf{1}_{(-0.5, 0.5)}(d). \end{aligned}$$

### Função de Distribuição *a posteriori* de um ARFIMA(0, $d$ , 0)

A função de distribuição *a posteriori* é proporcional ao produto da função de verossimilhança pela função de distribuição *a priori*, ou seja,

$$p(\boldsymbol{\eta}|\mathbf{X}) = Cp(\mathbf{X}|\boldsymbol{\eta})p(\boldsymbol{\eta}), \quad (3.55)$$

onde  $C = \left(\int p(\mathbf{X}|\boldsymbol{\eta})p(\boldsymbol{\eta})d\boldsymbol{\eta}\right)^{-1}$ .

Assim, temos que

$$\begin{aligned} p(\boldsymbol{\eta}|\mathbf{X}) &= Cp(\mathbf{X}|\boldsymbol{\eta})p(\boldsymbol{\eta}) = C(2\pi\sigma_\varepsilon^2)^{-\frac{n}{2}} \exp\left[-\frac{1}{2\sigma_\varepsilon^2} \sum_{t=1}^n (X_t - \mu_t)^2\right] \\ &(2\pi\sigma_\varepsilon^2)^{-\frac{m}{2}} \exp\left[-\frac{1}{2\sigma_\varepsilon^2} \sum_{j=1}^m \varepsilon_{1-j}^2\right] \frac{(\sigma_\varepsilon^2)^{-a-1} b^a}{\Gamma(a)} \exp\left[-\frac{b}{\sigma_\varepsilon^2}\right] \\ &= C_1(\sigma_\varepsilon^2)^{-\frac{n+m}{2}-a-1} \exp\left[-\frac{1}{2\sigma_\varepsilon^2} \left(\sum_{t=1}^n (X_t - \mu_t)^2 + \sum_{j=1}^m \varepsilon_{1-j}^2\right) - \frac{b}{\sigma_\varepsilon^2}\right], \end{aligned} \quad (3.56)$$

onde

$$C_1 = C(2\pi)^{-\frac{m+n}{2}} \frac{b^a}{\Gamma(a)}.$$

### Distribuições Condicionais Completas de um ARFIMA(0, d, 0)

A partir da função de distribuição conjunta *a posteriori*  $p(\boldsymbol{\eta}|\mathbf{X})$  do vetor de parâmetros  $\boldsymbol{\eta} = (d, \sigma_\varepsilon^2, \boldsymbol{\varepsilon}_0)$ , podemos calcular  $p(\xi|\boldsymbol{\eta}_{(\xi)}, \mathbf{X})$ , que é a distribuição condicional do parâmetro  $\xi$  dado todos os parâmetros restantes do modelo e a série temporal  $\mathbf{X} = (X_1, X_2, \dots, X_n)'$ . Essa distribuição é chamada de *condicional completa* do parâmetro  $\xi$ . As funções de distribuições condicionais podem ser calculadas através da função de distribuição *a posteriori*, notando que

$$p(\xi|\boldsymbol{\eta}_{(\xi)}, \mathbf{X}) = \frac{p(\boldsymbol{\eta}|\mathbf{X})}{\int_{\Xi} p(\boldsymbol{\eta}|\mathbf{X}) d\xi}, \quad (3.57)$$

onde  $\Xi$  é o conjunto formado pelos valores nos quais  $\xi$  está definido e  $\boldsymbol{\eta}_{(\xi)}$  é o vetor que difere de  $\boldsymbol{\eta}$  apenas por não ter o parâmetro  $\xi$ . Além disso, a integral dada em (3.57) é constante como função de  $\xi$ .

Podemos, dessa forma, calcular as funções de distribuição condicionais completas dos parâmetros  $d$  e  $\sigma_\varepsilon^2$  assim como de todas as variáveis auxiliares. Então, considerando  $d = x$  e  $\sigma_\varepsilon^2 = y$ , obtemos

$$\begin{aligned} p(x|\boldsymbol{\eta}_{(x)}, \mathbf{X}) &= \frac{p(\boldsymbol{\eta}|\mathbf{X})}{\int_{-0.5}^{0.5} p(\boldsymbol{\eta}|\mathbf{X}) dx} = \\ &= \frac{C_1(\sigma_\varepsilon^2)^{-\frac{n+m}{2}-a-1} \exp\left[-\frac{1}{2\sigma_\varepsilon^2} \left(\sum_{t=1}^n (X_t - \mu_t)^2 + \sum_{j=1}^m \varepsilon_{1-j}^2\right) - \frac{b}{\sigma_\varepsilon^2}\right]}{\int_{-0.5}^{0.5} C_1(\sigma_\varepsilon^2)^{-\frac{n+m}{2}-a-1} \exp\left[-\frac{1}{2\sigma_\varepsilon^2} \left(\sum_{t=1}^n (X_t - \mu_t)^2 + \sum_{j=1}^m \varepsilon_{1-j}^2\right) - \frac{b}{\sigma_\varepsilon^2}\right] dx} \\ &= \frac{C_1(\sigma_\varepsilon^2)^{-\frac{n+m}{2}-a-1} \exp\left[-\frac{1}{2\sigma_\varepsilon^2} \sum_{j=1}^m \varepsilon_{1-j}^2 - \frac{b}{\sigma_\varepsilon^2}\right] \exp\left[-\frac{1}{2\sigma_\varepsilon^2} \sum_{t=1}^n (X_t - \mu_t)^2\right]}{C_1(\sigma_\varepsilon^2)^{-\frac{n+m}{2}-a-1} \exp\left[-\frac{1}{2\sigma_\varepsilon^2} \sum_{j=1}^m \varepsilon_{1-j}^2 - \frac{b}{\sigma_\varepsilon^2}\right] \int_{-0.5}^{0.5} \exp\left[-\frac{1}{2\sigma_\varepsilon^2} \sum_{t=1}^n (X_t - \mu_t)^2\right] dx} \\ &= C_2 \exp\left[-\frac{1}{2\sigma_\varepsilon^2} \sum_{t=1}^n (X_t - \mu_t)^2\right] \mathbf{1}_{(-0.5, 0.5)}(x), \end{aligned}$$

onde

$$C_2 = \frac{1}{\int_{-0.5}^{0.5} \exp\left[-\frac{1}{2\sigma_\varepsilon^2} \sum_{t=1}^n (X_t - \mu_t)^2\right] dx}.$$

Logo,

$$p(d|\boldsymbol{\eta}_{(d)}, \mathbf{X}) = C_2 \exp\left[-\frac{1}{2\sigma_\varepsilon^2} \sum_{t=1}^n (X_t - \mu_t)^2\right] \mathbf{1}_{(-0.5, 0.5)}(d). \quad (3.58)$$

De forma análoga, podemos calcular a função de distribuição completa para o parâmetro  $\sigma_\varepsilon^2$ . Então,

$$\begin{aligned} p(y|\boldsymbol{\eta}_{(y)}, \mathbf{X}) &= \frac{p(\boldsymbol{\eta}|\mathbf{X})}{\int_0^\infty p(\boldsymbol{\eta}|\mathbf{X}) dy} \\ &= \frac{C_1 y^{-\frac{n+m}{2}-a-1} \exp\left[-\frac{1}{2y} \left(\sum_{t=1}^n (X_t - \mu_t)^2 + \sum_{j=1}^m \varepsilon_{1-j}^2\right) - \frac{b}{y}\right]}{\int_0^\infty C_1 y^{-\frac{n+m}{2}-a-1} \exp\left[-\frac{1}{2y} \left(\sum_{t=1}^n (X_t - \mu_t)^2 + \sum_{j=1}^m \varepsilon_{1-j}^2\right) - \frac{b}{y}\right] dy} \\ &= \frac{y^{-(\frac{T+m}{2}+a)-1} \exp\left\{-\frac{\frac{1}{2} \left(\sum_{t=1}^n (X_t - \mu_t)^2 + \sum_{j=1}^m \varepsilon_{1-j}^2\right) - b}{y}\right\}}{\int_0^\infty y^{-\frac{n+m}{2}-a-1} \exp\left[-\frac{1}{2y} \left(\sum_{t=1}^n (X_t - \mu_t)^2 + \sum_{j=1}^m \varepsilon_{1-j}^2\right) - \frac{b}{y}\right] dy} \\ &= C_3 y^{-(\frac{T+m}{2}+a)-1} \exp\left\{-\frac{\frac{1}{2} \left(\sum_{t=1}^n (X_t - \mu_t)^2 + \sum_{j=1}^m \varepsilon_{1-j}^2\right) - b}{y}\right\}, \end{aligned}$$

onde

$$C_3 = \frac{1}{\int_0^\infty y^{-\frac{n+m}{2}-a-1} \exp\left[-\frac{1}{2y} \left(\sum_{t=1}^n (X_t - \mu_t)^2 + \sum_{j=1}^m \varepsilon_{1-j}^2\right) - \frac{b}{y}\right] dy}.$$

Logo,  $(\sigma_\varepsilon^2|\boldsymbol{\eta}_{(\sigma_\varepsilon^2)}, \mathbf{X}) \sim GI(a^*, b^*)$ , onde

$$a^* = a + \frac{n+m}{2} \quad \text{e} \quad b^* = b + \frac{1}{2} \left(\sum_{t=1}^n (X_t - \mu_t)^2 + \sum_{j=1}^m \varepsilon_{1-j}^2\right), \quad (3.59)$$

ou seja,  $(\sigma_\varepsilon^2|\boldsymbol{\eta}_{(\sigma_\varepsilon^2)}, \mathbf{X})$  tem distribuição gama inversa, agora com parâmetros  $a^*$  e  $b^*$ .

As variáveis auxiliares obtidas como distribuições univariadas são da forma

$$(\varepsilon_j|\boldsymbol{\eta}_{(\varepsilon_j)}, \mathbf{X}) \sim \mathcal{N}(\mu_j, \sigma_\varepsilon^2(j)), \quad \text{para } j = 1-m, 2-m, \dots, 0 \quad (3.60)$$

$$\text{onde } \mu_j = \frac{\sigma_\varepsilon^2(j)}{\sigma_\varepsilon^2} \sum_{j=-1}^{1-m} \psi_{1-j}(d) X_{1-j} - \sum_{k=1; k \neq 1-j}^m \psi_k(d) \varepsilon_{1-j-k}$$

$$e \quad \sigma_\varepsilon^2(j) = \frac{\sigma_\varepsilon^2}{1 + \sum_{j=-1}^{1-m} [\psi_{1-j}(d)]^2}.$$

A Seção 3.4, a seguir, apresenta as simulações da estimação Bayesiana para os processos ARFIMA(0,  $d$ , 0).

### 3.4 Estimação em Processos ARFIMA(0, $d$ , 0)

Nesta seção, o estimador proposto através da abordagem Bayesiana é utilizado para estimar os parâmetros  $d$  e  $\sigma_\varepsilon^2$  de um processo ARFIMA(0,  $d$ , 0), definido em (2.22), embora esse estimador tenha sido construído inicialmente para um modelo AR( $p$ ).

A primeira parte desta seção descreve o método utilizado para estimar os parâmetros do processo ARFIMA(0,  $d$ , 0), que é analisado através da sua representação autoregressiva infinita. A seguir são feitas simulações para verificar a eficiência do estimador. O final da seção apresenta a análise de uma série temporal de dados reais. A série temporal escolhida é muito estudada na Estatística e representa o nível mínimo anual do Rio Nilo no período entre 622 e 1281 d.C.

#### Análise Autoregressiva do Processo ARFIMA(0, $d$ , 0)

A estimação dos parâmetros  $d$  e  $\sigma_\varepsilon^2$  é proposta considerando o processo  $\{X_t\}_{t \in \mathbb{Z}}$  na sua representação autoregressiva de ordem infinita (AR( $\infty$ )), como vista na expressão (2.23). Esta representação infinita é truncada em um certo ponto  $m$ , de modo que o processo ARFIMA(0,  $d$ , 0) é agora aproximado por um processo autoregressivo de ordem  $m$  (AR( $m$ )) da forma

$$X_t - \varphi_1(d)X_{t-1} - \cdots - \varphi_m(d)X_{t-m} = \sum_{j=0}^m \varphi_j(d)X_{t-j} = \varepsilon_t, \quad (3.61)$$

com coeficientes  $\varphi_j(d)$ , para todo  $j \in \{1, 2, \dots, m\}$ , dados em (2.24). O estimador proposto através da abordagem Bayesiana para o modelo AR( $p$ ), que foi estudado na Seção 3.2.1, pode agora ser utilizado para estimar os parâmetros de um processo ARFIMA(0,  $d$ , 0). Basta observar que os coeficientes  $\varphi_j(d)$  são, para todo  $j \in \{1, 2, \dots, m\}$ , dados como função de  $d$ . Temos,

$$\begin{aligned} \varphi_1(d) &= \frac{\Gamma(1-d)}{\Gamma(2)\Gamma(-d)} = \frac{(-d)\Gamma(-d)}{1\Gamma(1)\Gamma(-d)} = -d, \\ \varphi_2(d) &= \frac{\Gamma(2-d)}{\Gamma(3)\Gamma(-d)} = \frac{(1-d)\Gamma(1-d)}{2\Gamma(2)\Gamma(-d)} = \frac{1-d}{2}\varphi_1(d), \end{aligned}$$

$$\varphi_3(d) = \frac{\Gamma(3-d)}{\Gamma(4)\Gamma(-d)} = \frac{(2-d)\Gamma(2-d)}{3\Gamma(3)\Gamma(-d)} = \frac{2-d}{3}\varphi_2(d).$$

Em geral, os coeficientes dados pela expressão (2.24), podem ser calculados recursivamente por

$$\varphi_1(d) = -d \quad \text{e} \quad \varphi_{j+1}(d) = \frac{j-d}{j+1}\varphi_j(d), \quad \text{para todo } 1 \leq j \leq m-1. \quad (3.62)$$

### Simulações

Para verificar a eficiência desse método, foram feitas simulações. As ordens de aproximação são  $m = \{5, 10, 20, 30, 40, 50\}$ . Não é desejável que se tenha valores muito altos para  $m$ , pois isso torna o tempo computacional inviável. No entanto, é esperado que valores bem maiores para  $m$  realizem uma melhor aproximação na estimação dos parâmetros  $d$  e  $\sigma_\varepsilon^2$ . O valor de  $m$  depende de  $d$ , ou seja,  $m = m(d)$  e temos uma melhor aproximação autorregressiva para valores de  $d$  que estão próximos de 0.5 (ver Anexo B). A estimação dos parâmetros  $d$  e  $\sigma_\varepsilon^2$  de um processo ARFIMA(0,  $d$ , 0) foi realizada para diferentes valores para  $d$ , onde  $0 < d < 0.5$  e também para  $m \in \{5, 10, 20, 30, 40, 50\}$ . Através do comando `fracdiff.sim` foram geradas, para cada valor de  $d$ , 100 replicações de um processo ARFIMA(0,  $d$ , 0), onde cada série temporal tem tamanho  $n = 1000$ . Além disso, todas as séries temporais têm a mesma variância  $\sigma_\varepsilon^2 = 1$ . O valor estimado para ambos os parâmetros, denotado por  $\hat{d}$  e  $\hat{\sigma}_\varepsilon^2$ , foi calculado como sendo a média estimada relativa às 100 amostras; o vício, a variância (var) e o erro quadrático médio (eqm), cujas fórmulas são mostradas em (3.45), também foram calculados.

Os resultados encontrados a partir das simulações estão apresentados nas Tabelas 3.6 a 3.10.

**Tabela 3.6:** Resultados da Estimação dos Parâmetros de um Processo ARFIMA(0,  $d$ , 0), com  $d = 0.1$  e  $\sigma_\varepsilon^2 = 1$ .

$m$	$\hat{d}$	$\hat{\sigma}_\varepsilon^2$	vício( $\hat{d}$ )	vício( $\hat{\sigma}_\varepsilon^2$ )	var( $\hat{d}$ )	var( $\hat{\sigma}_\varepsilon^2$ )	eqm( $\hat{d}$ )	eqm( $\hat{\sigma}_\varepsilon^2$ )
5	0.10303	1.00031	0.00303	0.00031	0.00522	0.04043	0.00535	0.04055
10	0.10110	0.99823	0.00110	-0.00177	0.00367	0.04757	0.00376	0.04778
20	0.10114	1.00566	0.00114	0.00566	0.00537	0.03243	0.00544	0.03359
30	0.10598	1.00143	0.00598	0.00143	0.00512	0.04132	0.00531	0.04278
40	0.10437	0.99241	0.00437	-0.00759	0.00598	0.04335	0.00611	0.04411
50	0.10228	0.98021	0.00228	-0.01979	0.00616	0.03889	0.00640	0.03908

**Tabela 3.7:** Resultados da Estimação dos Parâmetros de um Processo ARFIMA(0,  $d$ , 0), com  $d = 0.25$  e  $\sigma_\varepsilon^2 = 1$ .

$m$	$\hat{d}$	$\hat{\sigma}_\varepsilon^2$	vício( $\hat{d}$ )	vício( $\hat{\sigma}_\varepsilon^2$ )	var( $\hat{d}$ )	var( $\hat{\sigma}_\varepsilon^2$ )	eqm( $\hat{d}$ )	eqm( $\hat{\sigma}_\varepsilon^2$ )
5	0.25712	1.00655	0.00712	0.00655	0.00490	0.04731	0.00508	0.04761
10	0.25440	1.00657	0.00440	0.00657	0.00655	0.03798	0.00673	0.03881
20	0.25395	1.00732	0.00732	0.00732	0.00571	0.04381	0.00588	0.04455
30	0.25299	1.00328	0.00328	0.00328	0.00187	0.04411	0.00213	0.04443
40	0.25019	0.98357	0.00019	-0.01643	0.00098	0.05343	0.00108	0.05574
50	0.24890	0.97601	-0.00110	-0.02399	0.00158	0.05561	0.00180	0.05987

**Tabela 3.8:** Resultados da Estimação dos Parâmetros de um Processo ARFIMA(0,  $d$ , 0), com  $d = 0.35$  e  $\sigma_\varepsilon^2 = 1$ .

$m$	$\hat{d}$	$\hat{\sigma}_\varepsilon^2$	vício( $\hat{d}$ )	vício( $\hat{\sigma}_\varepsilon^2$ )	var( $\hat{d}$ )	var( $\hat{\sigma}_\varepsilon^2$ )	eqm( $\hat{d}$ )	eqm( $\hat{\sigma}_\varepsilon^2$ )
5	0.36873	1.01985	0.01873	0.01985	0.00468	0.03891	0.00509	0.03941
10	0.36439	1.01779	0.01439	0.01779	0.00551	0.04431	0.00564	0.04458
20	0.36245	1.01335	0.01245	0.01335	0.00311	0.04078	0.00343	0.04117
30	0.35721	1.00315	0.00721	0.00315	0.00479	0.03858	0.00484	0.03880
40	0.35299	0.99298	0.00299	-0.0702	0.00522	0.03649	0.00557	0.03663
50	0.35074	0.98088	0.00074	-0.01912	0.00581	0.03536	0.00593	0.03666

**Tabela 3.9:** Resultados da Estimação dos Parâmetros de um Processo ARFIMA(0,  $d$ , 0), com  $d = 0.45$  e  $\sigma_\varepsilon^2 = 1$ .

$m$	$\hat{d}$	$\hat{\sigma}_\varepsilon^2$	vício( $\hat{d}$ )	vício( $\hat{\sigma}_\varepsilon^2$ )	var( $\hat{d}$ )	var( $\hat{\sigma}_\varepsilon^2$ )	eqm( $\hat{d}$ )	eqm( $\hat{\sigma}_\varepsilon^2$ )
5	0.48715	1.03695	0.03715	0.03695	0.00816	0.07880	0.00834	0.07891
10	0.48288	1.04487	0.03288	0.04487	0.00897	0.05133	0.00981	0.05344
20	0.47881	1.02887	0.02881	0.02887	0.00659	0.06255	0.00663	0.06454
30	0.47761	1.03061	0.02761	0.03061	0.00755	0.05568	0.00774	0.05596
40	0.46791	0.99265	0.01791	-0.00735	0.00869	0.04439	0.00881	0.04461
50	0.45527	0.98023	0.00527	-0.01977	0.00943	0.03930	0.00955	0.03965

**Tabela 3.10:** Resultados da Estimação dos Parâmetros de um Processo ARFIMA(0,  $d$ , 0), com  $d = 0.49$  e  $\sigma_\varepsilon^2 = 1$ .

$m$	$\hat{d}$	$\hat{\sigma}_\varepsilon^2$	vício( $\hat{d}$ )	vício( $\hat{\sigma}_\varepsilon^2$ )	var( $\hat{d}$ )	var( $\hat{\sigma}_\varepsilon^2$ )	eqm( $\hat{d}$ )	eqm( $\hat{\sigma}_\varepsilon^2$ )
5	0.53004	1.03212	0.04004	0.03212	0.00829	0.05628	0.00878	0.05692
10	0.52934	1.02234	0.03934	0.02234	0.00827	0.06650	0.00904	0.06669
20	0.52276	1.04554	0.03276	0.04554	0.00549	0.05289	0.00568	0.05329
30	0.50866	1.03334	0.01866	0.03334	0.00210	0.00428	0.00229	0.00481
40	0.49771	0.98861	-0.00771	-0.1139	0.00109	0.06436	0.00175	0.06479
50	0.49347	0.97975	-0.00347	-0.02025	0.00085	0.04021	0.00118	0.04113

Os resultados apresentados nas Tabelas 3.6 a 3.10 mostram que o valor de  $\hat{d}$  está mais próximo do valor real à medida que  $m$  assume valores mais altos. Os resultados para  $\hat{\sigma}_\varepsilon^2$  mostram que valores mais altos para  $m$  aproximam o valor real  $\sigma_\varepsilon^2 = 1$  por valores menores que 1. Além disso, a variância e o erro quadrático médio foi sempre menor para  $\hat{d}$  quando comparado a  $\hat{\sigma}_\varepsilon^2$ .

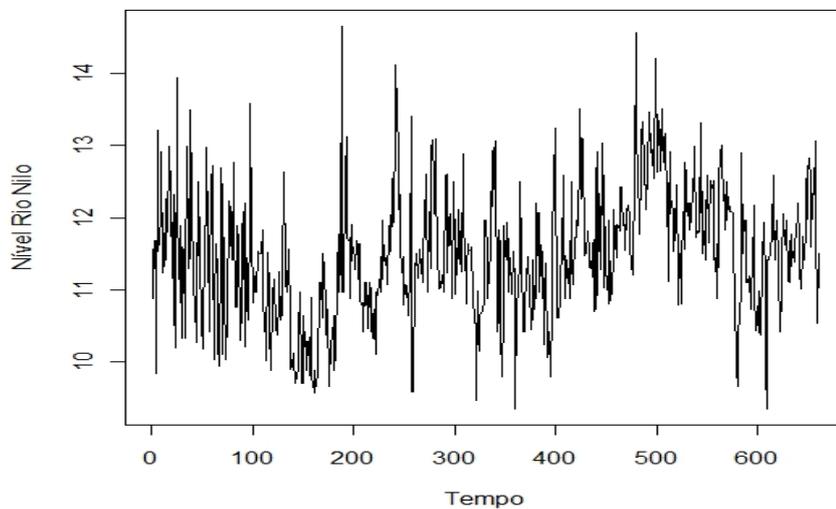
A seguir, apresentamos uma aplicação para o modelo ARFIMA(0,  $d$ , 0). Utilizamos a série temporal contendo o volume anual mínimo do Rio Nilo para o período entre os anos 622 e 1281.

### Aplicação: Nível Mínimo Anual do Rio Nilo

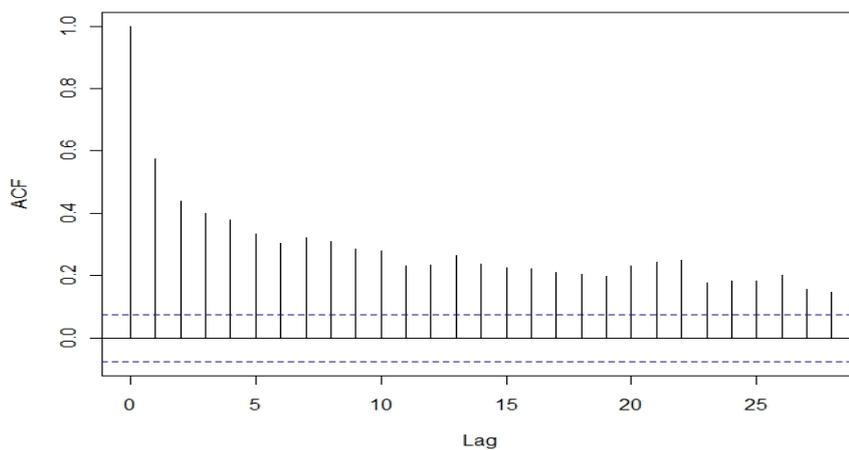
A série temporal do volume anual mínimo do Rio Nilo, representada através da Figura 3.6, fornece dados reais para o período entre os anos 622 e 1281, com um total de  $n = 660$  observações. Esta série temporal aparece em diversos trabalhos na Estatística pois ela apresenta a característica da longa dependência. De acordo com esses trabalhos o parâmetro  $d$  encontra-se no intervalo (0.35; 0.42). A Figura 3.7 apresenta o gráfico da função de autocorrelação amostral, e mostra que a série temporal parece exibir longa dependência. A Figura 3.8 apresenta a função periodograma da série temporal do rio Nilo. O pico que aparece próximo da frequência zero, na Figura 3.8, também indica que a série temporal possui longa dependência. Em Palma (2007) encontra-se um estudo mais completo sobre os dados do Rio Nilo. Em Palma (2007), o estudo é realizado através de uma abordagem Bayesiana, onde o modelo ARFIMA(0,  $d$ , 0) foi utilizado para descrever a longa dependência e os valores estimados para os parâmetros foram  $\hat{d} = 0.4195$  e  $\hat{\sigma}_\varepsilon^2 = 0.4884$ .

O modelo ARFIMA( $p, d, q$ ) permite analisar séries temporais que apresentam a característica da longa dependência, mas não considera a sazonalidade. Um estudo de uma série temporal, cujos dados representam o fluxo mensal do Rio Nilo, pode ser encontrada em Bisognin e Lopes (2007). Para a análise desta série temporal foi utilizado o modelo *autoregressivo fracionalmente integrado média móvel com sazonalidade  $s$* , denotado por SARFIMA( $p, d, q$ )  $\times$  ( $P, D, Q$ ) $_s$ . Esse modelo permite que sejam consideradas tanto a longa dependência como a sazonalidade presentes no processo.

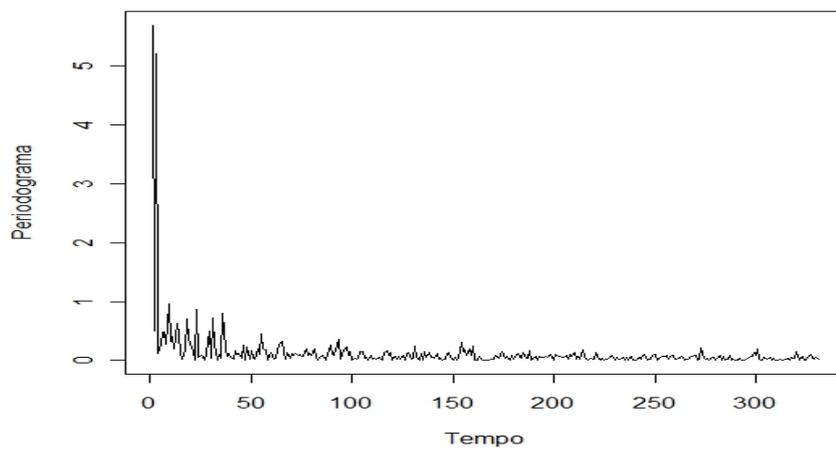
Neste trabalho, realizamos um estudo de simulação para o modelo ARFIMA(0,  $d$ , 0). O método utilizado foi baseado na aproximação autoregressiva, dada pela expressão (3.61), considerando o estimador proposto através da abordagem Bayesiana. O valor escolhido para  $m$  na expressão (3.61) foi  $m = 50$ , já que a aproximação do processo ARFIMA(0,  $d$ , 0) por um processo AR(50) apresenta resultados próximos do esperado na estimação dos parâmetros  $d$  e  $\sigma_\varepsilon^2$ . Os resultados encontrados foram  $\hat{d} = 0.4722$  e  $\hat{\sigma}_\varepsilon^2 = 0.4908$ .



**Figura 3.6:** Nível Mínimo Anual do Rio Nilo no período de 622 a 1281.



**Figura 3.7:** Função de Autocorrelação Amostral da Série Temporal do Rio Nilo.



**Figura 3.8:** Função Periodograma da Série Temporal do Rio Nilo.

# Capítulo 4

## Alguns Algoritmos Computacionais Bayesianos

Quando desejamos obter informações sobre a função de distribuição *a posteriori* de interesse, muitas vezes é necessário resolver integrais que não apresentam solução analítica. Uma alternativa para contornar esse problema é utilizar a aproximação de Laplace para resolver essas integrais. No entanto, a aproximação de Laplace pode não ser o método mais indicado quando as integrais são muito complexas. Nesse caso, o tempo computacional gasto para resolver essas integrais pode ser muito grande, o que inviabiliza esse método.

A inferência Bayesiana vem obtendo grandes avanços devido a algumas técnicas de simulação que permitem, de forma relativamente simples, obter amostras da função de distribuição *a posteriori* de interesse. Em particular, os métodos de Monte Carlo em Cadeias de Markov (MCMC), têm recebido grande atenção. Esse método de amostragem consiste basicamente em produzir uma cadeia de Markov cuja distribuição de equilíbrio seja a função de distribuição *a posteriori*  $p(\boldsymbol{\theta}|\mathbf{y})$  de interesse. Alguns métodos de amostragem mais conhecidos são apresentados nesse capítulo. Um estudo envolvendo métodos de MCMC pode ser encontrado em Gelman et al. (2004). Em Robert e Casella (1994) encontramos várias técnicas de amostragem utilizando métodos de Monte Carlo em Cadeias de Markov. Em Palma (2007) encontramos a descrição de técnicas que permitem verificar a convergência da cadeia de Markov.

### 4.1 Método de Monte Carlo Ordinário

Vamos supor que desejamos aproximar a integral da forma

$$\mathbb{E}(g(\boldsymbol{\theta})|\mathbf{y}) = \int g(\boldsymbol{\theta})p(\boldsymbol{\theta}|\mathbf{y})d\boldsymbol{\theta}, \quad (4.1)$$

onde  $g(\cdot)$  é uma função contínua,  $\boldsymbol{\theta} \in \Theta \subseteq \mathbb{R}^k$  e  $\mathbf{y} \in \mathbb{R}^m$ , onde os valores de  $k$  e  $m$  dependem, respectivamente, da dimensão do vetor de parâmetros e do tamanho da série temporal. Além disso, vamos supor que seja possível simular uma amostra  $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \dots, \boldsymbol{\theta}_n$  a partir da função de distribuição *a posteriori*  $p(\boldsymbol{\theta}|\mathbf{y})$ . O método de Monte Carlo ordinário consiste em, a partir da amostra  $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \dots, \boldsymbol{\theta}_n$  de  $p(\boldsymbol{\theta}|\mathbf{y})$ , aproximar (4.1) por

$$\hat{\mathbb{E}}(g(\boldsymbol{\theta})|\mathbf{y}) = \frac{1}{n} \sum_{i=1}^n g(\boldsymbol{\theta}_i). \quad (4.2)$$

Pela lei forte dos grandes números, temos que

$$\hat{\mathbb{E}}(g(\boldsymbol{\theta})|\mathbf{y}) \xrightarrow{q.c.} \mathbb{E}(g(\boldsymbol{\theta})|\mathbf{y}), \quad (4.3)$$

onde “ $\xrightarrow{q.c.}$ ” indica a *convergência quase certa*.

A precisão dessa aproximação pode ser medida pelo *erro padrão de Monte Carlo*, dado por

$$\frac{1}{\sqrt{n(n-1)}} \left\{ \sum_{i=1}^n \left( g(\boldsymbol{\theta}_i) - \frac{1}{n} \sum_{i=1}^n g(\boldsymbol{\theta}_i) \right)^2 \right\}^{\frac{1}{2}}. \quad (4.4)$$

O método de Monte Carlo ordinário permite criar intervalos de confiança. Para isso, consideramos uma amostra ordenada da função de distribuição *a posteriori*  $p(\boldsymbol{\theta}|\mathbf{y})$ , dada por  $\boldsymbol{\theta}_{(1)} < \boldsymbol{\theta}_{(2)} < \dots < \boldsymbol{\theta}_{(n)}$ . Então,  $[\boldsymbol{\theta}_{\frac{\alpha}{2}}, \boldsymbol{\theta}_{1-\frac{\alpha}{2}}]$  é um intervalo a  $100(1 - \alpha)\%$  para  $\boldsymbol{\theta}$ , onde os extremos do intervalo são tais que

$$\mathbb{P}[\boldsymbol{\theta} \leq \boldsymbol{\theta}_{\frac{\alpha}{2}}|\mathbf{y}] = \frac{\alpha}{2} \quad \text{e} \quad \mathbb{P}[\boldsymbol{\theta} \leq \boldsymbol{\theta}_{1-\frac{\alpha}{2}}|\mathbf{y}] = 1 - \frac{\alpha}{2}. \quad (4.5)$$

Esse método de simulação pode ser encontrado em Gelman et al. (2004) e Gamerman (1997). Em Albert (2008), são apresentados exemplos do método de Monte Carlo utilizando o R-project.

## 4.2 Método de Monte Carlo por Importância

Para obter inferência a partir da função de distribuição *a posteriori*, muitas vezes é necessário resolver integrais que não apresentam solução analítica. Na literatura encontramos muitos métodos numéricos para aproximar essas integrais (ver Gelman et al., 2004 e Gamerman, 1997).

Uma forma de simulação que permite amostrar valores é o método de Monte Carlo por importância. Notar que o método de Monte Carlo ordinário considera que simulamos uma amostra diretamente da função de distribuição *a posteriori*  $p(\boldsymbol{\theta}|\mathbf{y})$ . Quando isso não é possível, uma alternativa é simular uma amostra de uma função de distribuição que seja semelhante a  $p(\boldsymbol{\theta}|\mathbf{y})$ ,

dita amostragem via função de importância. Vamos considerar que  $l(\boldsymbol{\theta})$  seja uma função densidade na qual seja simples simular amostras e que aproxime a função de distribuição *a posteriori*  $p(\boldsymbol{\theta}|\mathbf{y})$ . Assim, observando que

$$p(\boldsymbol{\theta}|\mathbf{y}) = \frac{p(\boldsymbol{\theta})f(\mathbf{y}|\boldsymbol{\theta})}{\int p(\boldsymbol{\theta})f(\boldsymbol{\theta}|\mathbf{y})d\boldsymbol{\theta}},$$

onde  $f(\mathbf{y}|\boldsymbol{\theta})$  e  $p(\boldsymbol{\theta})$  são as funções de verossimilhança e de distribuição *a priori* para  $\boldsymbol{\theta}$ , podemos escrever (4.1) da forma

$$\begin{aligned} \int g(\boldsymbol{\theta})p(\boldsymbol{\theta}|\mathbf{y})d\boldsymbol{\theta} &= \frac{\int g(\boldsymbol{\theta})f(\mathbf{y}|\boldsymbol{\theta})p(\boldsymbol{\theta})d\boldsymbol{\theta}}{\int f(\mathbf{y}|\boldsymbol{\theta})p(\boldsymbol{\theta})d\boldsymbol{\theta}} \\ &= \frac{\int g(\boldsymbol{\theta})\frac{f(\mathbf{y}|\boldsymbol{\theta})p(\boldsymbol{\theta})}{l(\boldsymbol{\theta})}l(\boldsymbol{\theta})d\boldsymbol{\theta}}{\int \frac{f(\mathbf{y}|\boldsymbol{\theta})p(\boldsymbol{\theta})}{l(\boldsymbol{\theta})}l(\boldsymbol{\theta})d\boldsymbol{\theta}} \\ &= \frac{\int g(\boldsymbol{\theta})w(\boldsymbol{\theta})l(\boldsymbol{\theta})d\boldsymbol{\theta}}{\int w(\boldsymbol{\theta})l(\boldsymbol{\theta})d\boldsymbol{\theta}}, \end{aligned} \tag{4.6}$$

onde  $w(\boldsymbol{\theta}) = \frac{f(\mathbf{y}|\boldsymbol{\theta})p(\boldsymbol{\theta})}{l(\boldsymbol{\theta})}$ . A partir da amostra  $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \dots, \boldsymbol{\theta}_n$  de  $l(\boldsymbol{\theta})$ , onde  $l(\boldsymbol{\theta})$  é a função que aproxima  $p(\boldsymbol{\theta}|\mathbf{y})$ , encontramos a aproximação de Monte Carlo para  $\mathbb{E}(g(\boldsymbol{\theta})|\mathbf{y})$  dada por

$$\hat{\mathbb{E}}(g(\boldsymbol{\theta})|\mathbf{y}) = \frac{1}{\sum_{i=1}^n w_i} \sum_{i=1}^n w_i g(\boldsymbol{\theta}_i), \tag{4.7}$$

onde  $w_i = \frac{f(\mathbf{y}|\boldsymbol{\theta}_i)p(\boldsymbol{\theta}_i)}{l(\boldsymbol{\theta}_i)}$ , para  $i = 1, 2, \dots, n$ .

Esse método de amostragem pode ser encontrado em Gelman et al. (2004). Em Albert (2008) o R-project é utilizado para ilustrar esse método de amostragem.

### 4.3 Método de Monte Carlo em Cadeias de Markov

O seguinte procedimento é muito utilizado para gerar amostras a partir da função de distribuição *a posteriori*  $p(\boldsymbol{\theta}|\mathbf{y})$ , principalmente para o caso onde  $\boldsymbol{\theta}$  tem dimensão muito elevada. Um estudo a respeito desse método pode ser encontrado em Gelman et al. (2004), Robert e Casella (1994) e Gamerman (1997).

Vamos supor que desejamos gerar uma amostra a partir de uma função de distribuição *a posteriori*  $p(\boldsymbol{\theta}|\mathbf{y})$ ,  $\boldsymbol{\theta} \in \Theta \subseteq \mathbb{R}^k$ , mas não podemos fazer

isso diretamente. Podemos construir uma cadeia de Markov com espaço de estados no espaço de parâmetros  $\Theta$  (conjunto formado por todos os valores possíveis para  $\theta$ ) que é simples de simular e cuja distribuição de equilíbrio seja dada por  $p(\boldsymbol{\theta}|\mathbf{y})$ . Se temos muitas simulações dessa cadeia, os seus valores simulados podem ser usados como base para encontrar as características da função de distribuição *a posteriori*  $p(\boldsymbol{\theta}|\mathbf{y})$ .

Se a função de distribuição conjunta *a posteriori*  $p(\boldsymbol{\theta}|\mathbf{y})$  for positiva em  $\Theta_1, \Theta_2, \dots, \Theta_k$ , com  $\Theta_i$  sendo suporte para a distribuição de  $\theta_i$ ,  $i = 1, 2, \dots, k$ , então a função de distribuição *a posteriori*  $p(\boldsymbol{\theta}|\mathbf{y})$  é unicamente determinada pelas distribuições condicionais completas  $p(\theta_i|\mathbf{y}, \boldsymbol{\theta}_{(i)})$ , para  $i = 1, 2, \dots, k$ , onde

$$\boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_k)' \quad \text{e} \quad \boldsymbol{\theta}_{(i)} = (\theta_1, \dots, \theta_{i-1}, \theta_{i+1}, \dots, \theta_k)'. \quad (4.8)$$

Se  $\boldsymbol{\theta}^{(n)} = (\theta_1^{(n)}, \theta_2^{(n)}, \dots, \theta_k^{(n)})'$  é uma realização de uma cadeia, temos que,

$$\boldsymbol{\theta}^{(n)} \xrightarrow{d} \boldsymbol{\theta} \sim p(\boldsymbol{\theta}|\mathbf{y}), \quad (4.9)$$

onde “ $\xrightarrow{d}$ ” indica *convergência em distribuição*, quando  $n \rightarrow \infty$ .

Suponha que  $g(\boldsymbol{\theta})$  seja uma função contínua do parâmetro  $\boldsymbol{\theta} \in \Theta$ . Para estimar o valor esperado de  $g(\boldsymbol{\theta})$  relativo à função de distribuição *a posteriori*  $p(\boldsymbol{\theta}|\mathbf{y})$ , isto é,

$$\mathbb{E}(g(\boldsymbol{\theta}|\mathbf{y})) = \int g(\boldsymbol{\theta})p(\boldsymbol{\theta}|\mathbf{y})d\boldsymbol{\theta}, \quad (4.10)$$

podemos utilizar  $\frac{1}{n} \sum_{i=1}^n g(\boldsymbol{\theta}^{(i)})$  pois sabemos que

$$\frac{1}{n} \sum_{i=1}^n g(\boldsymbol{\theta}^{(i)}) \xrightarrow{q.c.} \mathbb{E}[g(\boldsymbol{\theta}|\mathbf{y})], \quad (4.11)$$

onde “ $\xrightarrow{q.c.}$ ” indica a *convergência quase certa*.

Na prática,  $\boldsymbol{\theta}^{(i)}$  pode estar correlacionado, mas podemos considerar espaços adequados entre os  $\boldsymbol{\theta}^{(i)}$  gerados para garantir uma amostra aleatória de  $p(\boldsymbol{\theta}|\mathbf{y})$ . Na geração das amostras devemos considerar as  $l$  primeiras iterações como um período de aquecimento (“burn-in-samples”) que devem ser descartadas para eliminar o efeito dos valores iniciais. Após esse período podemos selecionar 10 mil amostras, por exemplo, e então de cada dez observações geradas, selecionar uma observação para compor a amostra desejada. Temos como resultado  $B = 1000$  amostras cujo efeito da correlação é agora bem menor.

Em Robert e Casella (1994) e Gelman et al. (2004) encontramos uma descrição sobre técnicas computacionais envolvendo métodos de Monte Carlo em cadeias de Markov.

## 4.4 Amostrador de Gibbs

O amostrador de Gibbs é um algoritmo usado para a geração de amostras em distribuições *a posteriori* que não tenham uma forma fechada conhecida ou que tenham um difícil tratamento analítico. O amostrador de Gibbs foi proposto por Geman e Geman (1984) e difundido por Gelfand e Smith (1990).

É um método de amostragem iterativo de uma cadeia de Markov, onde a transição de um estado para o outro é feita a partir das distribuições condicionais completas, isto é, a partir de um vetor de parâmetros  $\theta$  *a posteriori* define-se a distribuição condicional completa de um subvetor de parâmetros com a distribuição destes condicionados a todos os demais.

Suponhamos que desejamos obter inferências da distribuição *a posteriori* conjunta  $p(\boldsymbol{\theta}|\mathbf{y})$ , para  $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_k)'$ . Para isso, simulamos quantidades aleatórias de distribuições condicionais completas que produzem uma cadeia de Markov, dadas por  $p(\theta_i|\mathbf{y}, \boldsymbol{\theta}_{(i)})$ , onde  $i = 1, \dots, k$  e  $\boldsymbol{\theta}_{(i)}$  é o vetor dado em (4.8).

As condicionais  $p(\theta_i|\mathbf{y}, \boldsymbol{\theta}_{(i)})$  são identificadas como função de  $\theta_i$  por inspeção da forma de  $p(\boldsymbol{\theta}|\mathbf{y})$  que é a distribuição *a posteriori* de  $\boldsymbol{\theta}$  dado  $y$  (ver Gamerman, 1997).

Suponha que atribuímos um conjunto arbitrário de valores iniciais, dado por  $\boldsymbol{\theta}^{(0)} = (\theta_1^{(0)}, \theta_2^{(0)}, \dots, \theta_k^{(0)})'$ , para o vetor de parâmetros  $\boldsymbol{\theta}$ . Então escrevemos o algoritmo:

- (i) Gerar  $\theta_1^{(1)}$  de  $p(\theta_1|\mathbf{y}, \theta_2^{(0)}, \dots, \theta_k^{(0)})$
- (ii) Gerar  $\theta_2^{(1)}$  de  $p(\theta_2|\mathbf{y}, \theta_1^{(1)}, \theta_3^{(0)}, \dots, \theta_k^{(0)})$
- (iii) Gerar  $\theta_3^{(1)}$  de  $p(\theta_3|\mathbf{y}, \theta_1^{(1)}, \theta_2^{(1)}, \theta_4^{(0)}, \dots, \theta_k^{(0)})$
- ⋮
- (k) Gerar  $\theta_k^{(1)}$  de  $p(\theta_k|\mathbf{y}, \theta_1^{(1)}, \theta_2^{(1)}, \dots, \theta_{k-1}^{(1)})$ .

Então, substituir os valores iniciais  $\boldsymbol{\theta}^{(0)}$  pela nova realização que gera  $\boldsymbol{\theta}^{(1)} = (\theta_1^{(1)}, \theta_2^{(1)}, \dots, \theta_k^{(1)})'$  de  $\boldsymbol{\theta}$  e repetir o processo acima.

Para  $n$  suficientemente grande, observar que o valor  $(\theta_1^{(n)}, \theta_2^{(n)}, \dots, \theta_k^{(n)})'$  converge para um valor da quantidade aleatória com distribuição  $p(\boldsymbol{\theta}|\mathbf{y})$  (ver Geman e Geman, 1984). Além disso,  $\theta_j^{(n)}$  pode ser considerado como uma observação simulada da distribuição *a posteriori* marginal  $p(\theta_j|\mathbf{y})$ , para todo  $j = 1, 2, \dots, k$ . Replicando o processo acima  $B$  vezes obtemos  $B$  vetores  $(\theta_{1b}^{(n)}, \theta_{2b}^{(n)}, \dots, \theta_{kb}^{(n)})'$ ; para  $g = 1, 2, \dots, B$ .

Da convergência do amostrador de Gibbs, qualquer característica da função densidade *a posteriori* marginal  $p(\theta_j|\mathbf{y})$  pode ser obtida. Em particular, se  $p(\theta_j|\mathbf{y}, \boldsymbol{\theta}_{(j)})$  é dada em forma fechada, então

$$\widehat{p(\theta_j|\mathbf{y})} = \frac{1}{B} \sum_{b=1}^B p(\theta_j|\mathbf{y}, \boldsymbol{\theta}_{(j)}^b), \quad \text{para } j = 1, 2, \dots, k.$$

Existem várias técnicas para verificar a convergência do algoritmo. Gelman e Smith (1990) sugerem o uso de técnicas gráficas; dessa forma, considera-se várias cadeias paralelas a partir de valores iniciais diferentes. Após um grande número de iterações em cada cadeia, compara-se os histogramas para a  $j$ -ésima componente  $\theta_j$  de  $\boldsymbol{\theta}$ . Histogramas similares indicam a convergência da cadeia.

Geweke (1992) sugere métodos gráficos baseados em séries temporais das amostras selecionadas. Uma técnica para monitorar a convergência do algoritmo é proposta por Gelman et al. (2004), baseado na análise da variância. Esse teste de convergência segue o seguinte princípio: diversas cadeias são criadas em paralelo, cada uma tendo uma certa distribuição inicial e essas distribuições iniciais devem assumir valores dispersos. Para que o algoritmo convirja para uma distribuição estacionária, é necessário que todas as cadeias sigam um mesmo padrão de comportamento. Assim, para cada cadeia gerada, sua variância amostral é calculada e então é possível estabelecer uma média das variâncias dentro dessas cadeias. A partir daí, a variância entre as cadeias também é calculada.

Esse método de monitoramento da convergência propõe que se a variância entre as cadeias for bem menor do que a variância dentro das mesmas, então a cadeia converge para uma distribuição de equilíbrio.

Uma descrição mais detalhada a respeito desse método pode ser encontrada em Gelman et al. (2004) e em Palma (2007).

## 4.5 Algoritmo de Metropolis-Hastings

O amostrador de Gibbs permite gerar amostras da função de distribuição *a posteriori* conjunta desde que as distribuições condicionais completas possuam uma forma fechada. Por outro lado, o algoritmo Metropolis-Hastings permite gerar amostras da distribuição *a posteriori* conjunta que tenham distribuições condicionais completas possuindo ou não uma forma fechada.

Quando utilizamos uma função de distribuição *a priori* conjugada, temos em geral funções de distribuição *a posteriori* com uma forma fechada conhecida como, por exemplo, as distribuições Normal, Gama, Poisson, Beta, entre outras. As simulações de amostras dessas distribuições estão disponíveis em qualquer *software*.

Quando as distribuições condicionais completas não são facilmente identificadas podemos utilizar o algoritmo de Metropolis-Hastings ou métodos de amostragem por importância. Para isso, devemos supor que desejamos simular amostras da função densidade não-regular  $p(\theta_i|\boldsymbol{\theta}_{(i)}, \mathbf{y})$ , ou simplesmente  $p(\theta_i|\boldsymbol{\theta}_{(i)})$ , em que  $\boldsymbol{\theta}_{(i)} = (\theta_1, \theta_2, \dots, \theta_{i-1}, \theta_{i+1}, \dots, \theta_k)'$ .

Precisamos definir o núcleo de transição  $q(\theta, \beta)$  da distribuição  $\pi(\theta)$  que representa  $p(\theta_i|\boldsymbol{\theta}_{(i)})$  e que transforma  $\theta$  em  $\beta$ . Se  $\theta$  é uma variável real com amplitude em toda a reta  $\mathbb{R}$ , podemos construir  $q(\cdot, \cdot)$  tal que  $\beta \leftarrow \theta + \sigma z$ , onde  $Z \sim \mathcal{N}(0, \sigma^2)$ , em que  $\sigma^2$  reflete a variância condicional de  $\theta$  em  $\pi(\theta)$ .

Se  $\theta$  é limitado com amplitude  $(a, b)$  devemos utilizar uma transformação que leva  $(a, b)$  em  $\mathbb{R}$  e então utilizar o núcleo de transição  $q(\cdot, \cdot)$  e aplicar o algoritmo de Metropolis-Hasting para a densidade da variável transformada. O algoritmo é dado pelos seguintes passos:

**Passo 1:** Iniciar com um valor inicial  $\boldsymbol{\theta}^{(0)}$  e o indicador de estágio,  $j = 0$ ;

**Passo 2:** Gerar um ponto  $\beta$  do núcleo de transição  $q(\boldsymbol{\theta}^{(j)}, \beta)$ ;

**Passo 3:** Atualizar  $\boldsymbol{\theta}^{(j+1)} = \beta$  com probabilidade,

$$\alpha = \min \left\{ 1, \frac{\pi(\beta)q(\boldsymbol{\theta}^{(j)}, \beta)}{\pi(\boldsymbol{\theta}^{(j)})q(\beta, \boldsymbol{\theta}^{(j)})} \right\} \quad (4.12)$$

Escolher  $\boldsymbol{\theta}^{(j)}$  com probabilidade  $1 - \alpha$ ;

**Passo 4:** Repetir os passos 2 e 3 e até conseguir uma distribuição estacionária.

#### Observação 4.1.

1- O algoritmo Metropolis-Hastings é especificado pela função densidade candidata para a geração da  $q(\cdot, \cdot)$ ;

2- Se um valor candidato é rejeitado, o valor atual é considerado na próxima etapa;

3- O cálculo de  $p$  em (4.12) não depende da constante normalizadora;

4- Se a função densidade candidata para a geração das amostras é simétrica, isto é, se  $q(x, y) = q(y, x)$ , a probabilidade de movimento se reduz à  $\frac{\pi(\beta)}{\pi(\boldsymbol{\theta}^{(j)})}$ .

Assim, se  $\pi(\beta) > \pi(\boldsymbol{\theta}^{(j)})$ , a cadeia se move para  $\beta$ ; caso contrário, ela se move para  $\beta$  com probabilidade  $\frac{\pi(\beta)}{\pi(\boldsymbol{\theta}^{(j)})}$ , ou seja, um salto para a região “ascendente” é sempre aceito e um salto para a região “descendente” é aceito com uma dada probabilidade.

5- Se a escolha de  $q(\cdot, \cdot)$  não for adequada, podemos não conseguir uma distribuição estacionária no **Passo 4**.

Um estudo mais completo sobre o Metropolis-Hastings pode ser encontrado em Gelman et al. (2004), assim como técnicas para verificar a convergência do algoritmo.

## 4.6 Amostrador de Fatia

É um método utilizado para gerar amostras de um grande número de distribuições, em particular aquelas cujas funções densidade de probabilidade estejam em um conjunto fechado e limitado.

Vamos supor que desejamos gerar valores de uma distribuição que tenha função densidade  $g(x)$ ,  $x \in D \subseteq \mathbb{R}$ . Considerar uma variável auxiliar  $Z$ , onde  $Z|x \sim U(0, g(x))$ . Então, a função de distribuição conjunta de  $(Z, X)$  é uniforme na região  $\{(z, x) : 0 \leq z \leq g(x)\}$ , com função densidade dada por

$$f_{(Z,X)}(z, x) = \begin{cases} \frac{1}{k}, & 0 \leq z \leq g(x) \\ 0, & \text{caso contrário.} \end{cases}$$

A função de distribuição condicional de  $X|Z = z$  é tal que

$$f_{X|Z}(x|z) \propto f(z, x) = \begin{cases} \frac{1}{k}, & 0 \leq z \leq g(x) \\ 0, & \text{caso contrário.} \end{cases}$$

Logo,

$$X|Z = z \sim U(A(z)), \quad \text{onde } A(z) = \{x : g(x) \geq z\}.$$

Utilizando o algoritmo Amostrador de Gibbs é possível simular valores a partir das funções de distribuições condicionais  $X|Z = z$  e  $Z|X = x$  que, como foi visto, possuem função de distribuição uniforme. Um procedimento iterativo pode ser dado por:

**Passo 1:** Gerar  $Z^{(i)} \sim U(0, g(x^{(i-1)}))$ , para  $i = 1, 2, \dots, n$ . Nesse caso, deve ser atribuído um valor inicial  $x^{(0)}$  para gerar  $Z^{(1)}$ .

**Passo 2:** Gerar  $X^{(i)} \sim U(A(z^{(i)}))$ , onde  $A(z) = \{x : g(x) \geq z^{(i)}\}$ , para  $i = 1, 2, \dots, n$ .

Esse método de amostragem apresenta a vantagem de utilizar apenas a distribuição uniforme para gerar valores. Além disso, não é necessário assumir uma função de distribuição proposta como no caso do algoritmo Metropolis-Hastings, que em muitos casos pode ser difícil de encontrar. Por outro lado, a determinação de  $A(z)$  pode ser uma tarefa difícil. Um estudo mais completo sobre esse método de amostragem pode ser encontrado em Neal (2003), onde são apresentados métodos para encontrar  $A(z)$  assim como uma generalização do algoritmo envolvendo variáveis multidimensionais. Em

Albert (2008) encontramos vários exemplos de distribuições onde o método pode ser aplicado.

Outra característica desse algoritmo é que os valores gerados costumam ser fortemente correlacionados. Então podemos considerar espaços adequados entre os valores gerados, garantindo que a amostra seja aleatória.

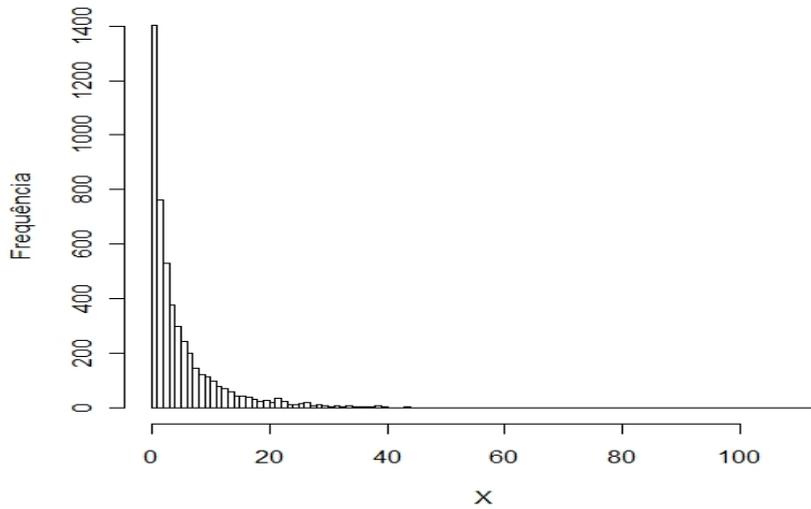
A seguir são apresentados dois exemplos para a utilização do algoritmo Amostrador de Fatia, para gerar valores das variáveis aleatórias a partir da sua função densidade de probabilidade.

**Exemplo 4.1.** Considere  $X$  uma variável aleatória cuja função densidade é dada por  $g_X(x) = \frac{1}{2}e^{-\sqrt{x}}$ , para todo  $x > 0$ . Nesse caso temos  $Z|X = x \sim U\left(0, \frac{1}{2}e^{-\sqrt{x}}\right)$  e

$$A(z) = \{x : g_X(x) \geq z\} = \left\{x : \frac{1}{2}e^{-\sqrt{x}} \geq z\right\} = \{x : x < (\ln(2z))^2\}.$$

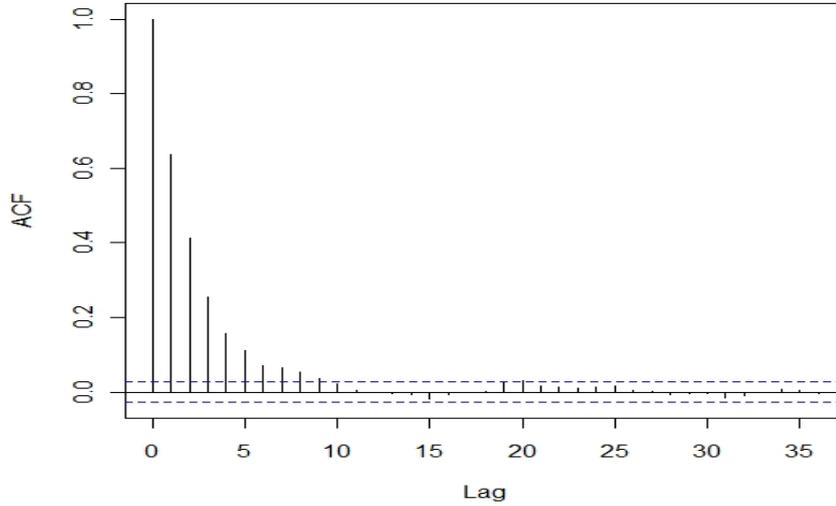
Logo,  $X|Z = z \sim U\left(0, (\ln(2z))^2\right)$ .

Utilizando o algoritmo Amostrador de Gibbs, foram gerados 5000 valores da variável  $X$ . As Figuras 4.1 e 4.2 mostram o histograma e a função de autocorrelação amostral, respectivamente. A forte autocorrelação pode ser percebida, já que essa é uma característica desse método de amostragem.



**Figura 4.1:** Histograma dos 5000 valores gerados através da variável  $X$ .

**Exemplo 4.2.** Considere  $Y \sim \mathcal{N}(0, 1)$ . Sabemos que sua função densidade é dada por  $g_Y(y) = \frac{1}{\sqrt{2\pi}}e^{-\frac{y^2}{2}}$ , para todo  $y \in \mathbb{R}$ . Então,  $Z|Y = y \sim U\left(0, \frac{1}{\sqrt{2\pi}}e^{-\frac{y^2}{2}}\right)$  e

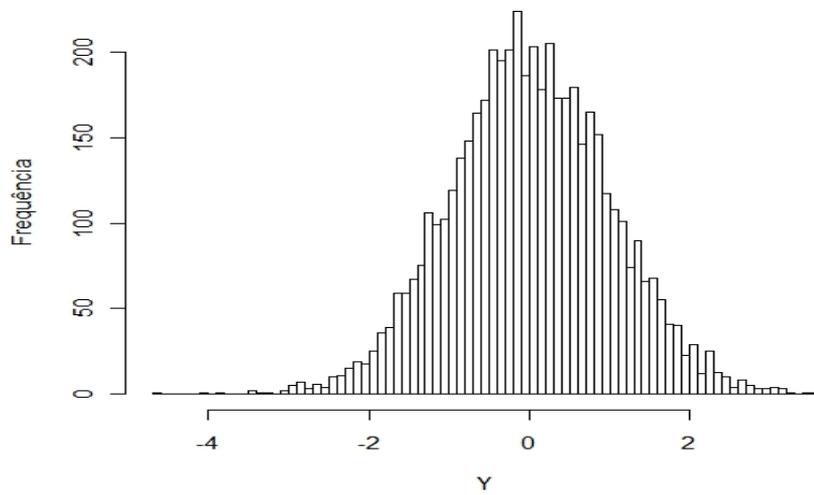


**Figura 4.2:** Função de Autocorrelação Amostral dos 5000 valores gerados através da variável  $X$ .

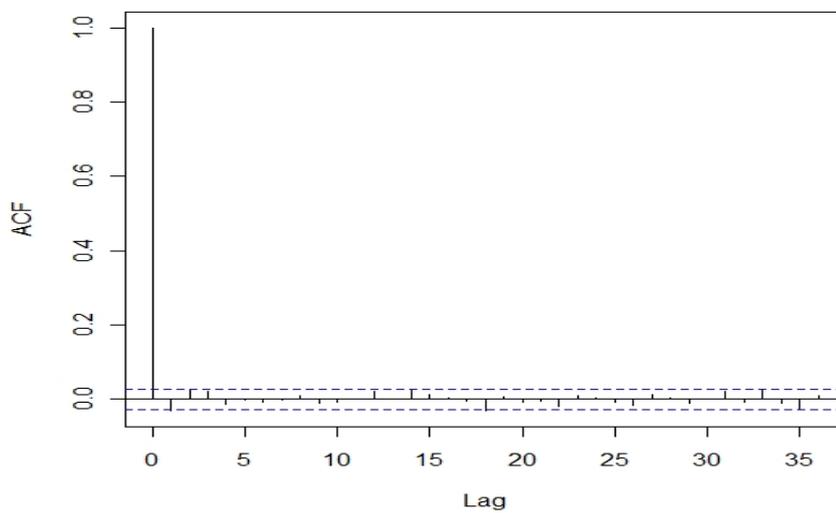
$$\begin{aligned}
 A(z) &= \{y : g_Y(y) \geq z\} = \left\{y : \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} \geq z\right\} \\
 &= \left\{y : -\sqrt{-2 \ln(\sqrt{2\pi}z)} \leq y \leq \sqrt{-2 \ln(\sqrt{2\pi}z)}\right\}.
 \end{aligned} \tag{4.13}$$

Logo  $Y|Z = z \sim U\left(-\sqrt{-2 \ln(\sqrt{2\pi}z)}, \sqrt{-2 \ln(\sqrt{2\pi}z)}\right)$ .

Foram gerados 5000 valores da variável  $Y$ . As Figuras 4.3 e 4.4 mostram o histograma e a função de autocorrelação amostral, respectivamente. Podemos verificar que o efeito da autocorrelação é bem menor do que no Exemplo 4.1.



**Figura 4.3:** Histograma dos 5000 Valores Gerados da Variável  $Y$ .



**Figura 4.4:** Função de Autocorrelação Amostral dos 5000 Valores Gerados da Variável  $Y$ .

# Capítulo 5

## Conclusão

Neste trabalho realizamos um estudo sobre a estimação dos parâmetros dos processos *autoregressivo de médias móveis*, denotado por  $\text{ARMA}(p, q)$  e *autoregressivo fracionalmente integrado de médias móveis*, denotado por  $\text{ARFIMA}(p, d, q)$ . Os processos  $\text{ARFIMA}(p, d, q)$  são importantes no estudo de séries temporais, já que o parâmetro  $d$  permite caracterizar a existência ou não de longa dependência. Aqui apresentamos um estudo dos processos  $\text{ARFIMA}(0, d, 0)$  sob o enfoque Bayesiano.

Apresentamos uma abordagem Bayesiana para a estimação dos parâmetros dos processos  $\text{ARMA}$  e  $\text{ARFIMA}$ . Com relação ao modelo autoregressivo de ordem  $p$  ( $\text{AR}(p)$ ), o estimador proposto através da abordagem Bayesiana apresentou resultados ligeiramente melhores quando comparado com o método clássico da máxima verossimilhança. As simulações apresentadas na Seção 3.2 mostram que houve uma melhor estimação para os parâmetros dos modelos  $\text{AR}(1)$  e  $\text{AR}(2)$ . Nesse caso, a variância e o erro quadrático médio apresentaram valores um pouco menores em ambos os modelos. A desvantagem do estimador proposto através da abordagem Bayesiana deve-se ao tempo computacional, já que o tempo de estimação dos parâmetros do modelo ficou próximo de 3 minutos. Utilizando o método clássico, os parâmetros foram estimados em poucos segundos.

No processo  $\text{ARFIMA}(0, d, 0)$ , analisado na Seção 3.4, estimamos os parâmetros  $d$  e  $\sigma_\varepsilon^2$ . Nesse caso, o modelo  $\text{ARFIMA}(0, d, 0)$  foi aproximado através da sua forma autoregressiva infinita. Na prática, consideramos um modelo  $\text{AR}(m)$ , com  $m$  finito. Observamos que, em geral, os valores mais altos de  $m$  produzem melhores resultados. O erro quadrático médio encontrado foi muito menor para parâmetro  $d$  do que para  $\sigma_\varepsilon^2$ , para os diferentes valores de  $m$ . Resultados melhores exigem que esta ordem de aproximação assumam valores bem mais elevados, como mostra a Tabela B.1 no Anexo B. Na Tabela B.1, podemos ver que valores de  $m$  maiores que 10000 representam uma boa aproximação quando  $d$  está próximo de 0.5. Nesse caso, o tempo computacional necessário para a estimação dos parâmetros será bem maior.

O final do Capítulo 3 apresenta um estudo do processo  $\text{ARFIMA}(0, d, 0)$

através da abordagem Bayesiana. São apresentados os cálculos que definem as funções de distribuição *a priori*, de verossimilhança e a função de distribuição *a posteriori*. As funções de distribuição condicionais para os parâmetros do modelo também foram calculadas e podem ser utilizadas com o objetivo de obter amostras para os parâmetros.

O Capítulo 4 descreve alguns algoritmos computacionais que permitem amostrar valores da função de distribuição *a posteriori*. O algoritmo Metropolis-Hastings tem a característica importante de poder ser utilizado para amostrar valores da função de distribuição *a posteriori* mesmo quando as funções de distribuição condicionais dos parâmetros do modelo não tem uma forma fechada conhecida.

Nesse trabalho, o processo de inovações  $\{\varepsilon_t\}_{t \in \mathbb{Z}}$  foi considerado ser um processo de ruído branco. Para trabalhos futuros, outros tipos de ruídos podem ser utilizados, como por exemplo, aqueles com distribuição t-Student, misturas de normais, chi-quadrado ou exponencial (ver artigo Sena, Reisen e Lopes, 2002). Além disso, o estudo do processo ARFIMA( $p, d, q$ ) pode ser estendido para valores de  $p$  e  $q$  não necessariamente nulos.

# Referências

- [1] Achcar, J.A. (2008). *Inferência Bayesiana: Análise Estatística de Dados Usando Métodos Bayesianos*. VI Jornada de Estatística, em Maringá, PR.
- [2] Albert, J. (2008). *Bayesian Computation with R*. New York: Springer-Verlag.
- [3] Ash, R.B. (1972). *Real Analysis and Probability*. New York: Academic Press.
- [4] Beran, J. (1994). *Statistics for Long Memory Processes*. New York: Chapman & Hall.
- [5] Bisognin, C. e S.R.C. Lopes (2007). “Estimating and Forecasting the Long Memory Parameter in the Presence of Periodicity”. *Journal of Forecasting*, Vol. **26**, 405-427.
- [6] Bloomfield, P. (1975). *Fourier Analysis of Time Series: An Introduction*. New York: Wiley.
- [7] Bolfarine, H. e Sandoval, M.C. (2000). *Introdução à Inferência Estatística*. Rio de Janeiro: SBM.
- [8] Brockwell, P.J. e Davis, R.A. (1991). *Time Series: Theory and Methods*. New York: Springer-Verlag.
- [9] Fox, R. e M.S. Taqqu (1986). “Large-sample Properties of Parameter Estimates for Strongly Dependent Stationary Gaussian Time Series”. *The Annals of Statistics*, Vol. **14**, 517-532.
- [10] Gamerman, D. (1997). *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference*. London: Chapman & Hall.
- [11] Gelfand, A.E. e Smith, A.F.M. (1990). “Journal of the American Statistical Association”. *Sampling-Based Approaches to Calculating Marginal Densities*, Vol. **85**, 398-409.

- [12] Gelman, A., Carlin, J.B., M.C., Stern, H.S. e Rubin, D.B. (2004). *Bayesian Data Analysis*. London: Chapman & Hall.
- [13] Geman, S. e Geman, D. (1984). “Stochastic Relaxation, Gibbs Distributions and the Bayesian Restoration of Images”. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. **6**, 721-741.
- [14] Geweke, J. e S. Porter-Hudak (1983). “The Estimation and Application of Long Memory Time Series Model”. *Journal of Time Series Analysis*, Vol. **4**(4), 221-238.
- [15] Granger, C.W.J. e R. Joyeux (1980). “An Introduction to Long Memory Time Series Models and Fractional Differencing”. *Journal of Time Series Analysis*, Vol. **1**(1), 15-29.
- [16] Hamilton, J. D. (1994). *Time Series Analysis*. New Jersey: Princeton University Press.
- [17] Hosking, J. (1981). “Fractional Differencing”. *Biometrika*, Vol. **68**, 165-167.
- [18] Hosking, J. (1984). “Modelling Persistence in Hydrological Time Series using Fractional Differencing”. *Water Resources Research*, Vol. **20**(12), 1898-1908.
- [19] Karlin, S. e Taylor, H.M. (1975). *A First Course in Stochastic Process*. London: Academic Press.
- [20] Kinas, P.G. e Andrade, H.A. (2010). *Introdução à Análise Bayesiana (com R)*. Porto Alegre: mQn.
- [21] Lopes, S.R.C. (2008). “Long-Range Dependence in Mean and Volatility: Models, Estimation and Forecasting”. *Progress in Probability*, Vol. **60**, 497-525.
- [22] Migon, H.S. e Lopes, H.F. (2002). *Análise Bayesiana de Decisões: Aspectos Práticos*. São Paulo: ABE.
- [23] Migon, H.S., Souza, A.D.P. e Schmidt A.M. (2008). *Modelos Hierárquicos e Aplicações*. Rio de Janeiro: ABE.
- [24] Neal, R. (1997). “Markov Chain Monte Carlo Methods Based on ‘Slicing’ the Density Function”. Toronto, Canada.
- [25] Palma, W. (2007). *Long-Memory Time Series: Theory and Methods*. New Jersey: Wiley.

- [26] Pole, A., West, M. e Harrison, J. (1994). *Applied Bayesian Forecasting and Time Series Analysis*. Boca Raton: Chapman & Hall.
- [27] Robert, C.P. e Casella, G. (1994). *Introducing Monte Carlo Methods with R*. New York: Springer-Verlag.
- [28] Sáfadi, T. e Andrade, M. (2007). *Abordagem Bayesiana de Modelos de Séries Temporais*. São Paulo: ABE.
- [29] Schmidt, A.M. e Sansó, B. (2006). *Modelagem Bayesiana da Estrutura de Covariância de Processos Espaciais e Espaço-Temporais*. São Paulo: ABE.
- [30] Sena Júnior, M.R., V.A. Reisen e S.R.C. Lopes (2002). “Error and Model Misspecification in ARFIMA Process”. *Brazilian Review of Econometrics*, Vol. **21**, 101-135.
- [31] Sowell, F. (1992). “Maximum Likelihood Estimation of Stationary Univariate Fractionally Integrated Time Series Models”. *Journal of Econometrics*, Vol. **53**, 165-188.
- [32] Verzani, J. (2005). *Using R for Introductory Statistics*. Boca Raton: Chapman & Hall.
- [33] Wei, W.W.S. (1990). *Time Series Analysis: Univariate and Multivariate Methods*. New York: Addison Wesley.
- [34] Whittle, P. (1953). *Hypothesis Testing in Time Series Analysis*. New York: Hafner.

# Anexo A

Ao estudar o processo ARFIMA( $p, q, d$ ) vimos que este modelo é utilizado para séries temporais que tenham a característica da longa dependência, ou de longa memória. Nesse caso, as variáveis aleatórias do processo  $\{X_t\}_{t \in \mathbb{Z}}$  são correlacionadas. Então a função de distribuição conjunta do processo pode ser calculada através das funções de distribuição condicionais, dadas pela expressão (3.47). Em muitas situações também assumimos que o processo é Gaussiano, ou seja, as funções de distribuição do processo  $\{X_t\}_{t \in \mathbb{Z}}$  são normais multivariadas. Se conhecemos a função de distribuição conjunta do processo, essas funções de distribuição condicionais podem ser calculadas.

Como um exercício, apresentamos a função de distribuição condicional  $X$  dado  $Y = y$ , onde  $(X, Y)$  é um vetor de variáveis aleatórias correlacionadas cuja função de distribuição conjunta é dada através de uma normal bivariada  $\mathcal{N}(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$ .

**Proposição A.1** (Densidade Condicional). *Seja  $(X, Y)$  um vetor de variáveis aleatórias correlacionadas cuja função de distribuição conjunta é uma normal bivariada  $\mathcal{N}(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$ . Então, a distribuição condicional de  $X$  dado  $Y = y$  é dada por*

$$X|Y = y \sim \mathcal{N}\left(\mu_1 + \rho \frac{\sigma_1}{\sigma_2}(y - \mu_2), \sigma_1^2(1 - \rho^2)\right).$$

**Demonstração:** Como  $Y$  tem distribuição  $\mathcal{N}(\mu_2, \sigma_2^2)$ , podemos calcular a função de distribuição condicional  $X|Y$  através de

$$\begin{aligned} f(x|y) &= \frac{f_{(X,Y)}(x, y)}{f_Y(y)} = \\ &= \frac{1}{\sigma_1 \sqrt{2\pi(1 - (1 - \rho^2))}} \exp\left\{ \frac{1}{2(1 - \rho^2)} \left[ \frac{(x - \mu_1)^2}{\sigma_1^2} - \frac{2\rho}{\sigma_1 \sigma_2} (x - \mu_1) \right. \right. \\ &\quad \left. \left. (y - \mu_2) + [1 - (1 - \rho^2)] \frac{(y - \mu_2)^2}{\sigma_2^2} \right] \right\} = \\ &= \frac{1}{\sigma_1 \sqrt{2\pi(1 - \rho^2)}} \exp\left\{ \frac{1}{2(1 - \rho^2)} \left[ \frac{(x - \mu_1)}{\sigma_1} - \rho \frac{(y - \mu_2)}{\sigma_2} \right]^2 \right\} = \\ &= \frac{1}{\sigma_1 \sqrt{2\pi(1 - \rho^2)}} \exp\left\{ \frac{1}{2(1 - \rho^2)} \left[ \frac{\sigma_2(x - \mu_1) - \rho\sigma_1(y - \mu_2)}{\sigma_1 \sigma_2} \right]^2 \right\} = \end{aligned}$$

$$\begin{aligned}
&= \frac{1}{\sigma_1 \sqrt{2\pi(1-\rho^2)}} \exp\left\{ \frac{1}{2(1-\rho^2)} \left[ \frac{\sigma_2 x - \sigma_2 \mu_1 - \rho \sigma_1 y + \rho \sigma_1 \mu_2}{\sigma_1 \sigma_2} \right]^2 \right\} = \\
&= \frac{1}{\sigma_1 \sqrt{2\pi(1-\rho^2)}} \exp\left\{ \frac{1}{2(1-\rho^2)} \left[ \frac{x}{\sigma_1} - \frac{\mu_1}{\sigma_1} - \frac{\rho \sigma_1 y + \rho \sigma_1 \mu_2}{\sigma_1} \frac{1}{\sigma_2} \right]^2 \right\} = \\
&= \frac{1}{\sigma_1 \sqrt{2\pi(1-\rho^2)}} \exp\left\{ \frac{1}{2(1-\rho^2)} \left[ \frac{x - \mu_1 - \rho \frac{\sigma_1}{\sigma_2} (y - \mu_2)}{\sigma_1} \right]^2 \right\} = \\
&= \frac{1}{\sigma_1 \sqrt{2\pi(1-\rho^2)}} \exp\left\{ \frac{1}{2(1-\rho^2)} \left[ \frac{x - (\mu_1 + \rho \frac{\sigma_1}{\sigma_2} (y - \mu_2))}{\sigma_1} \right]^2 \right\}
\end{aligned}$$

Assim a distribuição de  $X$  dado  $Y = y$  é  $\mathcal{N}\left(\mu_1 + \rho \frac{\sigma_1}{\sigma_2} (y - \mu_2), \sigma_1^2 (1 - \rho^2)\right)$ .

## Anexo B

O Teorema 2.3, na Seção 2.3, mostra que, quando  $d > 0.5$ , o processo ARFIMA(0,  $d$ , 0) apresenta-se como um processo autoregressivo de ordem infinita, dado pela expressão (2.23). Truncando essa série para um certo valor  $m$ , é possível analisar o processo ARFIMA(0,  $d$ , 0) através da sua aproximação autorregressiva de ordem  $m$ . O objetivo é fazer uma análise do modelo ARFIMA(0,  $d$ , 0) no sentido de avaliar os valores do parâmetro  $d$  que tornam a convergência da série mais rápida ou mais lenta.

Consideremos o processo ARFIMA(0,  $d$ , 0) definido em (2.22) e dado por

$$(1 - \mathcal{B})^d X_t = \varepsilon_t = \sum_{j=0}^{\infty} \varphi_j(d) X_{t-j},$$

onde  $d \in (-0.5, 0.5)$ . Os coeficientes  $\varphi_j(d)$  são obtidos de acordo com a expressão (2.24). Observando a expressão

$$(1 - z)^d = \sum_{j=0}^{\infty} \varphi_j(d) z^j,$$

vemos que se  $z = 1$  então  $\sum_{j=0}^{\infty} \varphi_j(d) = 0 \Rightarrow 1 + \sum_{j=1}^{\infty} \varphi_j(d) = 0 \Rightarrow \sum_{j=1}^{\infty} \varphi_j(d) = -1$ .

Assim  $-\sum_{j=1}^m \varphi_j(d) \rightarrow 1$ , quando  $m \rightarrow +\infty$ .

Pela Observação 2.7 temos a aproximação

$$\varphi_j(d) \sim \frac{j^{-d-1}}{\Gamma(-d)}, \quad \text{quando } j \rightarrow \infty.$$

Podemos utilizar a aproximação  $\varphi_m(d) \sim \frac{1}{\Gamma(-d)} m^{-d-1}$ , já que estamos considerando  $m$  grande. Então

$$|\varphi_m(d)| \sim \frac{1}{\Gamma(-d)} m^{-d-1} < \epsilon \Leftrightarrow \frac{-C}{m^{1+d}} < \epsilon \Leftrightarrow \frac{-C}{\epsilon} < m^{1+d} \Leftrightarrow \left(\frac{-C}{\epsilon}\right)^{\frac{1}{1+d}} < m,$$

onde  $C = \frac{1}{\Gamma(-d)} < 0$ .

Então, para  $0 < d_1 < d_2$  valem as desigualdades

$$m_{d_2} = \left(\frac{-C}{\epsilon}\right)^{\frac{1}{1+d_2}} < \left(\frac{-C}{\epsilon}\right)^{\frac{1}{1+d_1}} = m_{d_1}.$$

Isso mostra que  $-\sum_{j=0}^m \varphi_j(d)$  se aproxima do valor 1 mais rápido para valores de  $d$  mais próximos de 0.5, como mostram os valores da Tabela B.1.

**Tabela B.1:** Valores da Soma  $-S_1 = -\sum_{j=0}^m \varphi_j(d)$ , para  $d \in \{0.05, \dots, 0.5\}$ .

$d \backslash m$	10000	20000	30000	40000	50000	60000	70000	80000	90000	1e+05
0.05	0.3883	0.4091	0.4210	0.4292	0.4356	0.4407	0.4450	0.4487	0.4519	0.4548
0.10	0.6275	0.6524	0.6662	0.6757	0.6828	0.6886	0.6933	0.6974	0.7009	0.7041
0.15	0.7742	0.7965	0.8085	0.8166	0.8226	0.8274	0.8314	0.8347	0.8376	0.8402
0.20	0.8639	0.8815	0.8907	0.8968	0.9013	0.9049	0.9078	0.9102	0.9123	0.9141
0.25	0.9184	0.9314	0.9380	0.9423	0.9454	0.9479	0.9498	0.9515	0.9529	0.9541
0.30	0.9514	0.9605	0.9650	0.9679	0.9700	0.9716	0.9729	0.9740	0.9749	0.9756
0.35	0.9713	0.9774	0.9804	0.9823	0.9836	0.9846	0.9855	0.9861	0.9867	0.9872
0.40	0.9831	0.9872	0.9891	0.9903	0.9911	0.9918	0.9923	0.9927	0.9930	0.9933
0.45	0.9902	0.9928	0.9940	0.9947	0.9952	0.9956	0.9959	0.9962	0.9964	0.9965
0.50	0.9944	0.9960	0.9967	0.9972	0.9975	0.9977	0.9979	0.9980	0.9981	0.9982