

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL  
INSTITUTO DE MATEMÁTICA

ESTIMAÇÃO EM CLASSES DE PROCESSOS  
ESTOCÁSTICOS COM DECAIMENTO  
HIPERBÓLICO DA FUNÇÃO DE  
AUTOCORRELAÇÃO

Bárbara Patricia Olbermann

Tese submetida ao Programa de Pós-Graduação  
em Matemática como requisito parcial  
para a obtenção do grau de  
Doutor em Probabilidade e Estatística Matemática

Profa. Dra. Sílvia Regina Costa Lopes  
Orientadora

Porto Alegre, 22 de julho de 2002

## AGRADECIMENTOS

Agradeço à todos que torceram e colaboraram para a realização deste trabalho:

à professora Sílvia que, ao longo desta caminhada, sempre me incentivou e orientou com muita dedicação;

aos professores Artur Lopes, Francisco Cribari Neto e Valdério A. Reisen pelas contribuições e comentários para o aprimoramento deste trabalho;

à Rosane pela atenção e carinho;

aos bolsistas do LCPM, em especial Eduardo Flores e Renato Poli, pelo excelente ambiente de trabalho e inúmeros auxílios na parte computacional;

aos meus familiares e amigos pelo incentivo e apoio;

à “turma da sala da pós” pelos momentos de descontração e otimismo;

ao Sandro pelo carinho, compreensão e por compartilhar aflições e alegrias;

à Capes e à Fapergs pelo suporte financeiro e

à Universidade Pública pela oportunidade.

## RESUMO

Neste trabalho analisamos processos estocásticos com decaimento polinomial (também chamado hiperbólico) da função de autocorrelação.

Nosso estudo tem enfoque nas classes dos Processos *ARFIMA* e dos Processos obtidos à partir de iterações da transformação de Manneville-Pomeau.

Os objetivos principais são comparar diversos métodos de estimação para o parâmetro fracionário do processo *ARFIMA*, nas situações de estacionariedade e não estacionariedade e, além disso, obter resultados similares para o parâmetro do processo de Manneville-Pomeau. Entre os diversos métodos de estimação para os parâmetros destes dois processos destacamos aquele baseado na teoria de wavelets por ser aquele que teve o melhor desempenho.

## ABSTRACT

In this work we analyze stochastic processes with polynomial (also called hyperbolic) decay of the autocorrelation function.

We emphasize the class of *ARFIMA* processes and the one obtained from the Manneville-Pomeau iterated function processes. The main goal is to compare different estimation methods for the fractional parameter in *ARFIMA* process, for both stationary and non-stationary case and, moreover, to get similar results for the parameter in the Manneville-Pomeau process. Among all estimation methods for the parameters of these two processes we stress the one based on the wavelet theory since this had the best performance.

# ÍNDICE

<b>1</b>	<b>INTRODUÇÃO</b>	<b>1</b>
<b>2</b>	<b>CONCEITOS BÁSICOS</b>	<b>3</b>
2.1	Definições e Resultados na Análise de Séries Temporais . . . . .	3
2.2	Definições e Resultados da Teoria Ergódica . . . . .	11
<b>3</b>	<b>PROCESSOS ARFIMA(p,d,q)</b>	<b>15</b>
3.1	Ergodicidade . . . . .	21
3.2	Ordem da Variância da Soma Parcial . . . . .	25
3.3	Estimação do Parâmetro de Diferenciação . . . . .	28
3.4	Processos não Estacionários . . . . .	43
3.4.1	Descrição e Simulação . . . . .	43
3.4.2	Estimação dos Parâmetros . . . . .	44
3.4.3	Erros Não-Gaussianos . . . . .	49
3.4.4	Uma Aplicação . . . . .	51
3.5	Invariância para Primeira Diferença . . . . .	53
<b>4</b>	<b>PROCESSOS MANNEVILLE-POMEAU</b>	<b>60</b>
4.1	Definição e Propriedades da Transformação de Manneville-Pomeau . . . . .	60
4.2	Métodos de Estimação do Parâmetro $s$ . . . . .	70
4.3	Simulação e Resultados . . . . .	73
4.3.1	Longa Dependência . . . . .	74
4.3.2	Dependência intermediária . . . . .	78
<b>5</b>	<b>CONCLUSÃO</b>	<b>81</b>
	<b>ANEXO A</b>	<b>84</b>
	<b>ANEXO B</b>	<b>89</b>
	<b>REFERÊNCIAS</b>	<b>90</b>

# 1

## INTRODUÇÃO

Vamos considerar, neste trabalho, processos estocásticos com decaimento polinomial (também chamado hiperbólico) da função de autocorrelação.

Processos com decaimento lento da função de autocorrelação, também denominados de *processos com propriedade de longa dependência*, têm sido largamente explorados por diversos pesquisadores, pois são observados em várias áreas, como, por exemplo, genética, economia, hidrologia, astronomia, etc. O primeiro registro histórico de uma série temporal com característica de longa dependência foi a série dos níveis do rio Nilo, estudada pelo hidrologista Harold E. Hurst, que fez a primeira tentativa de modelá-la. O decaimento hiperbólico da função de autocorrelação no domínio da frequência está relacionado com a função densidade espectral ser ilimitada nas frequências próximas de zero. Para físicos e engenheiros, esta característica do processo é muitas vezes denominada  $\frac{1}{f}$ -noise. Entre os possíveis modelos para os quais ocorre o fenômeno de longa dependência, estão os incrementos estacionários de processos self-similar, em particular, os processos ruído fracionário Gaussiano e os processos autoregressivos de médias móveis com integração fracionária (*ARFIMA*).

Neste trabalho vamos apresentar, além dos processos *ARFIMA*, outros processos que também apresentam a propriedade de longa dependência: são os processos obtidos a partir de iterações da transformação de Manneville-Pomeau. Estes processos apresentam as mesmas propriedades dos processos *ARFIMA*, mas não possuem uma expressão analítica explícita para a função densidade espectral do modelo.

É bem conhecido, na análise de séries temporais, que diferentes modelos requerem também diferentes procedimentos (ou métodos) na estimação dos seus parâmetros. Não conhecemos um procedimento geral que funcione para todos os modelos. Aqui, neste trabalho, vamos concentrar nosso estudo nas classes dos Processos *ARFIMA* e dos Processos obtidos a partir de iterações da transformação de Manneville-Pomeau. Os objetivos principais deste tra-

balho são comparar diversos métodos de estimação para o parâmetro dos processos *ARFIMA*, nas situações de estacionariedade e não estacionariedade, e também comparar métodos de estimação para o parâmetro dos processos obtido a partir de iterações da transformação de Manneville-Pomeau.

Para a segunda classe de processos vamos comparar um método muito utilizado pelos físicos com outro método heurístico descrito no livro de Jan Beran e ainda com métodos que são propostos neste trabalho e que não haviam sido anteriormente explorados. Um dos métodos propostos para este caso é baseado na teoria de wavelets, que teve ótimo desempenho.

A comparação entre os métodos de estimação propostos para os processos *ARFIMA* é feita através dos vícios e dos erros quadráticos médios dos estimadores. Posteriormente, é feita a mesma análise para os processos Manneville-Pomeau.

Uma das principais contribuições originais deste trabalho é a utilização de técnicas de estimação do parâmetro de longa dependência de processos oriundos das iterações da transformação de Manneville-Pomeau, que são similares àquelas utilizadas em processos *ARFIMA*. Outra contribuição original importante foi o uso do método baseado em wavelets na estimação do parâmetro nos processos Manneville-Pomeau, no caso de longa dependência. Além disso, na situação de dependência intermediária, este trabalho contribui com um novo método de estimação baseado na propriedade de Hölder para a função densidade espectral do processo.

No Capítulo 2 apresentamos conceitos e idéias básicas sobre séries temporais e teoria ergódica. O Capítulo 3 descreve o modelo *ARFIMA* e suas propriedades, bem como métodos de estimação para o parâmetro de diferenciação e uma aplicação à dados reais. O processo Manneville-Pomeau é o enfoque do quarto Capítulo, onde também são apresentados vários métodos de estimação do parâmetro associado a este processo. Na Conclusão reunimos os resultados relevantes observados neste trabalho.

## 2

# CONCEITOS BÁSICOS

Apresentamos, neste capítulo, diversos conceitos fundamentais necessários à compreensão de modelos em séries temporais discutidos neste trabalho. Descrevemos idéias básicas de processos estocásticos, tais como as funções de autocovariância, autocorrelação e autocorrelação parcial, bem como estimadores para as mesmas e para a média de um processo. Além disso, apresentamos definições de estacionariedade, ergodicidade e idéias básicas da análise espectral e da teoria ergódica. Um estudo mais completo pode ser encontrado em Brockwell e Davis (1991) e em Durrett (1996).

## 2.1 Definições e Resultados na Análise de Séries Temporais

Nesta seção apresentamos definições importantes para a análise de processos estocásticos e séries temporais.

**DEFINIÇÃO 2.1:** Um *processo estocástico* é uma família de variáveis aleatórias  $\{X_t\}_{t \in T}$ , todas elas definidas em um mesmo espaço de probabilidade  $(\Omega, \mathcal{A}, \mathbb{P})$ , sendo  $T \neq \emptyset$  um conjunto de índices,  $\Omega$  o espaço amostral,  $\mathcal{A}$  a classe dos eventos aleatórios e  $\mathbb{P} : \mathcal{A} \rightarrow [0, 1]$  a função que associa a probabilidade de um evento qualquer.

A classe dos eventos aleatórios  $\mathcal{A}$  é uma  $\sigma$ -álgebra, i.e., uma coleção (não vazia) de subconjuntos de  $\Omega$  que satisfaz:

- i)  $\Omega \in \mathcal{A}$ ;
- ii) se  $A \in \mathcal{A}$  então  $A^c \in \mathcal{A}$ ;

iii) se  $A_i \in \mathcal{A}$  é uma seqüência contável de conjuntos em  $\mathcal{A}$  então  $\cup_i A_i \in \mathcal{A}$ .

As seguintes definições (funções de autocovariância e de autocorrelação) são muito importantes pois definem o grau de interdependência entre as variáveis dando assim informações sobre a estrutura de dependência de um processo estocástico.

Denotaremos por  $T$  o conjunto dos naturais  $\mathbb{N}$ , ou dos inteiros  $\mathbb{Z}$  ou ainda dos reais positivos  $\mathbb{R}^+$ .

**DEFINIÇÃO 2.2:** Seja  $\{X_t\}_{t \in T}$  um processo estocástico onde as variáveis  $X_t$ , para todo  $t \in T$ , possuem variância finita. A *função de autocovariância* de  $\{X_t\}_{t \in T}$  - denotada por  $\gamma_X(r, s)$  - é dada por

$$\gamma_X(r, s) = cov(X_r, X_s) = \mathbb{E}[(X_r - \mathbb{E}(X_r))(X_s - \mathbb{E}(X_s))], \quad r, s \in T,$$

onde  $\mathbb{E}(X_t) = \mu_t$  é a esperança do processo.

**Observação:** Ressaltamos que  $\gamma_X(t, t) = \mathbb{E}(X_t - \mu_t)^2 = Var(X_t)$  é a variância do processo, que será denotada por  $\sigma_X^2$ .

**DEFINIÇÃO 2.3:** Seja  $\{X_t\}_{t \in T}$  um processo estocástico onde as variáveis  $X_t$ , para todo  $t \in T$ , possuem variância finita. A *função de autocorrelação* - denotada por  $\rho_X(r, s)$  - é dada por

$$\rho_X(r, s) = \frac{\gamma_X(r, s)}{\sqrt{Var(X_r)}\sqrt{Var(X_s)}}, \quad r, s \in T.$$

Para obtermos propriedades probabilísticas e estatísticas desejáveis na utilização de modelos para descrever processos físicos, são necessárias algumas suposições. No caso de processos estocásticos, uma suposição geralmente feita é a de estacionariedade. Os conceitos, a seguir, referem-se a estacionariedade de processos estocásticos.

**DEFINIÇÃO 2.4:** A *função de distribuição n-dimensional* das variáveis aleatórias  $(X_{t_1}, X_{t_2}, \dots, X_{t_n})$  é definida por

$$F_{t_1, \dots, t_n}(x_1, \dots, x_n) = \mathbb{P}(X_{t_1} \leq x_1, X_{t_2} \leq x_2, \dots, X_{t_n} \leq x_n),$$

quaisquer que sejam  $n \in \mathbb{N} - \{0\}$ ,  $t_i \in T$ ,  $x_i \in \mathbb{R}$ ,  $1 \leq i \leq n$ .

**DEFINIÇÃO 2.5:** Um processo estocástico  $\{X_t\}_{t \in T}$  é dito ser *fortemente estacionário* se as funções de distribuição conjuntas de  $(X_{t_1}, X_{t_2}, \dots, X_{t_k})$  e

$(X_{t_1+h}, X_{t_2+h}, \dots, X_{t_k+h})$  são as mesmas, qualquer que seja  $k$  um inteiro positivo e quaisquer que sejam  $h \in T$  e  $t_i, t_{i+h} \in T$ ,  $1 \leq i \leq k$ . Se  $T = \mathbb{N}$ , a seqüência  $(X_0, X_1, \dots)$  é dita ser uma *seqüência estacionária*.

**DEFINIÇÃO 2.6:** Um processo estocástico  $\{X_t\}_{t \in T}$  é dito ser *fracamente estacionário* se

- i)  $\mathbb{E}|X_t|^2 < \infty$ , para todo  $t \in T$ ;
- ii)  $\mathbb{E}X_t = \mu$ , uma constante independente de  $t$ ;
- iii)  $\gamma_X(r, s) = \gamma_X(r + t, s + t)$ , para quaisquer  $r, s, t \in T$ .

Intuitivamente, um processo é estacionário se ele se mantém em torno da sua média e se desenvolve no tempo, de modo que a escolha de uma origem não seja importante.

Do mesmo modo como a função de autocorrelação de um processo estacionário, a função de autocorrelação parcial, definida logo a seguir, fornece informações sobre a estrutura de dependência do processo. Ela depende somente de propriedades de segunda ordem do processo  $\{X_t\}_{t \in T}$ . Deseja-se investigar a correlação entre  $X_t$  e  $X_{t+k}$  após a remoção das dependências lineares das variáveis intermediárias  $X_{t+1}, X_{t+2}, \dots, X_{t+k-1}$ .

**DEFINIÇÃO 2.7:** A *função de autocorrelação parcial*  $\phi_X(k, k)$  de ordem  $k$  de um processo estocástico  $\{X_t\}_{t \in T}$  é determinada da seguinte forma:

$$\phi_X(k, k) = \frac{\begin{vmatrix} 1 & \rho_X(1) & \rho_X(2) & \cdots & \rho_X(k-2) & \rho_X(1) \\ \rho_X(1) & 1 & \rho_X(1) & \cdots & \rho_X(k-3) & \rho_X(1) \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \rho_X(k-1) & \rho_X(k-2) & \rho_X(k-3) & \cdots & \rho_X(1) & \rho_X(k) \end{vmatrix}}{\begin{vmatrix} 1 & \rho_X(1) & \cdots & \rho_X(k-2) & \rho_X(k-1) \\ \rho_X(1) & 1 & \cdots & \rho_X(k-3) & \rho_X(k-2) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \rho_X(k-1) & \rho_X(k-2) & \cdots & \rho_X(1) & 1 \end{vmatrix}}. \quad (2.1)$$

Neste trabalho consideraremos os processos onde  $T = \mathbb{Z}$  (respectivamente,  $T = \mathbb{N}$  ou  $T = \{1, 2, \dots, n\}$ ), que serão denotados por  $\{X_t\}_{t \in \mathbb{Z}}$  (respectivamente,  $\{X_t\}_{t \in \mathbb{N}}$  ou  $\{X_t\}_{t=1}^n$ ).

Se  $T = \{1, \dots, n\}$  (ou  $T = \{0, 1, \dots, n-1\}$ ), então  $\{X_t\}_{t=1}^n$  (ou  $\{X_t\}_{t=0}^{n-1}$ ) é dito uma *série temporal*. Uma *série temporal* é um registro dos valores de certa quantidade, medida em diferentes tempos discretos. Por exemplo,

podemos ter um registro de temperaturas diárias, registro da voltagem de um circuito elétrico medida em intervalos de um segundo, índices diários da Bolsa de Valores, dados categorizados de uma cadeia de DNA, etc. Portanto, uma série temporal é um conjunto de observações de um fenômeno, onde cada observação é registrada num tempo específico  $t$ .

**Observações:**

a) Para facilitar a notação omitiremos, daqui por diante, a palavra fracamente da Definição 2.4 e chamaremos, então,  $\{X_t\}_{t \in \mathbb{Z}}$  de processo estocástico estacionário. Quando tratarmos de casos onde o processo é fortemente estacionário, ressaltaremos o fato.

b) Para um processo estocástico estacionário a função de autocovariância satisfaz  $\gamma_X(r, s) = \gamma_X(r - s, 0)$ , para quaisquer  $r, s \in \mathbb{Z}$ . Por esta razão é conveniente redefinirmos a *função de autocovariância de ordem  $k$*  de um processo estacionário como

$$\gamma_X(k) = \gamma_X(k, 0) = \text{cov}(X_{t+k}, X_t),$$

para quaisquer  $k, t \in \mathbb{Z}$ .

c) As funções de autocovariância e de autocorrelação de um processo estocástico estacionário têm as seguintes propriedades:

- $|\gamma_X(k)| \leq \gamma_X(0)$ , para todo  $k \in \mathbb{N}$ ;
- $\gamma_X(k) = \gamma_X(-k)$ , para todo  $k \in \mathbb{N}$ ;
- $\rho_X(0) = 1$ ;
- $|\rho_X(k)| \leq 1$ , para todo  $k \in \mathbb{N}$ ;
- $\rho_X(k) = \rho_X(-k)$ , para todo  $k \in \mathbb{N}$ .

A seguir, definimos estimadores para a esperança e para as funções de autocovariância, autocorrelação e autocorrelação parcial de um processo estocástico estacionário  $\{X_t\}_{t \in \mathbb{Z}}$  com

- $E(X_t) = \mu$ , para todo  $t \in \mathbb{Z}$ ;
- $Var(X_t) = \sigma_X^2$ , para todo  $t \in \mathbb{Z}$ ;
- $\text{cov}(X_t, X_{t+k}) = \gamma_X(k)$ , para todo  $t, k \in \mathbb{Z}$ .

Para a obtenção destes estimadores consideramos uma série temporal com  $n$  observações,  $X_1, X_2, \dots, X_n$ , obtida a partir do processo  $\{X_t\}_{t \in \mathbb{Z}}$ .

**DEFINIÇÃO 2.8:** Um estimador, pelo método dos momentos, para a esperança é a *média amostral* dada por

$$\bar{X} = \frac{\sum_{i=1}^n X_i}{n}.$$

O estimador  $\bar{X}$  é não-viciado ( $\mathbb{E}(\bar{X}) = \mu$ ) e consistente para  $\mu$  ( $\text{Var}(\bar{X}) \rightarrow 0$ , quando  $n \rightarrow \infty$ ).

**DEFINIÇÃO 2.9:** Um estimador para a função de autocovariância de ordem  $k$  é a *função de autocovariância amostral* dada por

$$\hat{\gamma}_X(k) = \frac{1}{n} \sum_{i=1}^{n-k} (X_{i+k} - \bar{X})(X_i - \bar{X}), \text{ para todo } |k| \leq n,$$

onde  $\bar{X}$  é a média amostral.

**Observação:** O quociente  $n$  na expressão de  $\hat{\gamma}_X(\cdot)$  é preferível ao quociente  $n-k$  porque, desta forma,  $\hat{\gamma}_X(\cdot)$  é função definida não-negativa. Geralmente, o estimador baseado na divisão por  $n-k$  tem menor vício mas maior variância quando comparado com o estimador definido acima (ver Wei (1990)).

**DEFINIÇÃO 2.10:** Um estimador para a função de autocorrelação de ordem  $k$  é a *função de autocorrelação amostral* dada por

$$\hat{\rho}_X(k) = \frac{\hat{\gamma}_X(k)}{\hat{\gamma}_X(0)}, \text{ para todo } |k| \leq n,$$

onde  $\hat{\gamma}_X(0) = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$  é o estimador para a variância  $\sigma_X^2$  do processo  $\{X_t\}_{t \in \mathbb{Z}}$ .

**DEFINIÇÃO 2.11:** A *função de autocorrelação parcial amostral*,  $\hat{\phi}_X(k, k)$ , é obtida substituindo  $\rho_X(i)$  por  $\hat{\rho}_X(i)$  na expressão (2.1). Ao invés de calcular os complicados determinantes para valores grandes de  $k$  na expressão (2.1), um método recursivo para obter  $\hat{\phi}_X(k, k)$  é sugerido em Box et al. (1995) através de

$$\hat{\phi}_X(k+1, k+1) = \frac{\hat{\rho}_X(k+1) - \sum_{j=1}^k \hat{\phi}_X(k, j) \hat{\rho}_X(k+1-j)}{1 - \sum_{j=1}^k \hat{\phi}_X(k, j) \hat{\rho}_X(j)}$$

e

$$\hat{\phi}_X(k+1, j) = \hat{\phi}_X(k, j) - \hat{\phi}_X(k+1, k+1) \hat{\phi}_X(k, k+1-j),$$

para todo  $j = 1, 2, \dots, k$  e  $k \in \{1, 2, \dots, n-1\}$ , onde  $\hat{\phi}_X(1, 1) = \hat{\rho}_X(1)$ .

O método acima pode, também, ser usado para obter o valor teórico de  $\phi_X(k, k)$ .

A seguir, apresentamos alguns conceitos da análise espectral de processos estocásticos.

**DEFINIÇÃO 2.12:** Seja  $\{X_t\}_{t \in \mathbb{Z}}$  um processo estocástico estacionário com função de autocovariância  $\gamma_X(\cdot)$  absolutamente convergente, i.e.,  $\sum_{k \in \mathbb{Z}} |\gamma_X(k)| < \infty$ . A *função densidade espectral* de  $\{X_t\}_{t \in \mathbb{Z}}$  é dada por

$$\begin{aligned} f_X(w) &= \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} \gamma_X(k) e^{-iwk} = \\ &= \frac{1}{2\pi} [\gamma_X(0) + 2 \sum_{k=1}^{\infty} \gamma_X(k) \cos(wk)], \text{ para } w \in [-\pi, \pi], \end{aligned} \quad (2.2)$$

onde foram consideradas as propriedades

- $\gamma_X(k) = \gamma_X(-k)$ ,
- $\text{sen}(-wk) = -\text{sen}(wk)$  e
- $\text{cos}(-wk) = \text{cos}(wk)$ .

O Teorema 2.1 abaixo fornece propriedades da função densidade espectral e sua demonstração pode ser encontrada em Wei (1990).

**TEOREMA 2.1:** *Seja  $\{X_t\}_{t \in \mathbb{Z}}$  um processo estocástico estacionário com função de autocovariância  $\gamma_X(\cdot)$  absolutamente convergente. A função densidade espectral, dada pela expressão (2.2), tem as seguintes propriedades:*

- $f_X(w)$  é uma função real contínua;
- $f_X(w) = f_X(-w)$ , para todo  $w \in [-\pi, \pi]$ ;
- $f_X(w) \geq 0$ , para todo  $w \in [-\pi, \pi]$ .

A seguir definimos dois estimadores da função densidade espectral (ver Definições 2.13 e 2.14) que, posteriormente, serão utilizados para a estimação de parâmetros (ver Seção 3.3 do Capítulo 3 e Seção 4.2 Capítulo 4).

**DEFINIÇÃO 2.13:** Para uma série temporal  $\{X_t\}_{t=1}^n$  com  $n$  observações, obtida de um processo  $\{X_t\}_{t \in \mathbb{Z}}$ , a *função periodograma* é definida por

$$I(w) = 2[\hat{\gamma}_X(0) + 2 \sum_{k=1}^{n-1} \hat{\gamma}_X(k) \cos(wk)], \text{ para todo } w \in [-\pi, \pi],$$

onde  $\hat{\gamma}_X(\cdot)$  é a função de autocovariância amostral do processo  $\{X_t\}_{t \in \mathbb{Z}}$  (ver Observação b) após a Definição 2.7).

Seja

$$I^*(w) = \frac{I(w)}{4\pi} = \frac{1}{2\pi} [\hat{\gamma}_X(0) + 2 \sum_{k=1}^{n-1} \hat{\gamma}_X(k) \cos(wk)], \quad w \in [-\pi, \pi]. \quad (2.3)$$

Dizemos que  $I^*(\cdot)$  é o estimador da função densidade espectral  $f_X(\cdot)$ , definida pela expressão (2.2). Consideraremos, daqui por diante,  $I^*(\cdot)$  como a função periodograma a qual passará a ser denotada por  $I(\cdot)$ .

Referimos o leitor a Lopes e Lopes (2002) para uma demonstração da convergência no sentido de distribuição do periodograma, que cobre os casos analisados no presente trabalho.

Como a função periodograma não é um estimador consistente da função densidade espectral  $f_X(\cdot)$ , consideraremos estimadores alternativos que possuam esta propriedade (uma abordagem completa pode ser encontrada em Priestley (1981)). Na próxima definição apresentamos um estimador consistente para a função densidade espectral.

**DEFINIÇÃO 2.14:** A *função periodograma suavizado*, denotada por  $f_s(\cdot)$ , é um estimador consistente da função densidade espectral  $f_X(\cdot)$  e é dada por

$$f_s(w) = \frac{1}{2\pi} \sum_{k=-(n-1)}^{n-1} \lambda(k) \hat{\gamma}_X(k) \cos(kw), \quad w \in [-\pi, \pi], \quad (2.4)$$

onde  $\lambda(\cdot)$  é uma função de ponderação conhecida como *janela* e  $\hat{\gamma}_X(\cdot)$  é a função de autocovariância amostral do processo.

Diferentes formas da função  $\lambda(\cdot)$  são sugeridas na literatura de análise de séries temporais (ver Priestley (1981)). Neste trabalho, usaremos a *janela de Parzen* (ver Seção 3.3 do Capítulo 3), que é dada por

$$\lambda(k) = \begin{cases} 1 - 6\left(\frac{k}{m}\right)^2 + 6\left(\frac{|k|}{m}\right)^3, & \text{se } |k| \leq \frac{m}{2}, \\ 2\left(1 - \frac{|k|}{m}\right)^3, & \text{se } \frac{m}{2} < |k| \leq m, \\ 0, & \text{caso contrário,} \end{cases} \quad (2.5)$$

onde  $m$  é função de  $n$ , o número de observações, e é chamado de *ponto de truncamento*.

Usaremos, também, a janela “cosine-bell” (ver Seção 3.3 do Capítulo 3) dada por

$$\lambda(k) = \frac{1}{2} \left\{ 1 - \cos \left[ \frac{2\pi(k + 0.5)}{n} \right] \right\}.$$

O Teorema 2.2, abaixo, é conhecido como o *Teorema da Representação Espectral* ou *Teorema de Herglotz* e sua demonstração pode ser encontrada em Brockwell e Davis (1991), página 117.

**TEOREMA 2.2** (Herglotz): *Uma função de autocovariância  $\gamma_X(\cdot)$ , de valor complexo definida em  $\mathbb{Z}$ , de um processo estocástico estacionário  $\{X_t\}_{t \in \mathbb{Z}}$ , é função definida não negativa se e somente se*

$$\gamma_X(h) = \int_{-\pi}^{\pi} e^{ihw} dF_X(w), \text{ para todo } h \in \mathbb{Z},$$

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 chamada função de distribuição espectral de  $\gamma_X(\cdot)$  (ou de  $X$ ) e se  $F_X(w) = \int_{-\pi}^w f_X(w) dw$ ,  $-\pi \leq w \leq \pi$ , então  $f_X(\cdot)$  é a função densidade espectral de  $\gamma_X(\cdot)$  (ou de  $X$ ).

As seguintes notações serão utilizadas ao longo deste trabalho.

**Notações:**

a) Se, para a seqüência  $\{a_n\}_{n \in \mathbb{N}}$ , existe um número real  $u \in \mathbb{R}$  e constantes  $c_1, c_2 > 0$  tais que, para todo  $n \in \mathbb{N}$ , vale

$$c_1 \leq \left| \frac{a_n}{n^{-u}} \right| \leq c_2,$$

então, denotaremos tal fato por  $a_n \approx n^{-u}$ .

b) Se, para a função  $g(\cdot)$ , existe um número real  $b \in \mathbb{R}$  e constantes  $d_1, d_2 > 0$  tais que, para todo  $x$ , vale

$$d_1 \leq \left| \frac{g(x)}{x^b} \right| \leq d_2,$$

então, denotaremos tal fato por  $g(x) \approx x^b$ .

A definição seguinte caracteriza, através de duas formas, os processos estocásticos que apresentam propriedade de longa dependência.

**DEFINIÇÃO 2.15:** Seja  $\{X_t\}_{t \in \mathbb{Z}}$  um processo estocástico estacionário. Se existe um número real  $u \in (0, 1)$  tal que

$$\rho_X(k) \approx k^{-u}, \tag{2.6}$$

onde  $\rho_X(\cdot)$  é a função de autocorrelação do processo  $\{X_t\}_{t \in \mathbb{Z}}$ , ou equivalentemente (ver minuciosas considerações no Anexo A), se existe um número real  $b \in (0, 1)$  tal que

$$f_X(w) \approx w^b, \tag{2.7}$$

onde  $f_X(\cdot)$  é a função densidade espectral do processo  $\{X_t\}_{t \in \mathbb{Z}}$ , então dizemos que  $\{X_t\}_{t \in \mathbb{Z}}$  é um *processo estocástico estacionário com longa memória ou longa dependência*. Neste caso, é verdadeiro que  $u = 1 + b$  (ver Anexo A).

Dizemos que o processo possui *dependência intermediária* se vale a expressão (2.6) com  $u \in (1, 2)$ .

**Observação:** É importante notar que a Definição 2.15, de longa dependência, é uma definição assintótica. Apenas mostra o comportamento da função de autocorrelação quando a ordem  $k$  tende ao infinito e da função densidade espectral quando a frequência tende a zero. Geralmente, não especifica explicitamente a função de autocorrelação para nenhuma ordem  $k$  finita fixada e nem especifica a função densidade espectral para nenhuma frequência  $w$  fixa. A Definição 2.15 determina, somente, a velocidade de convergência destas funções.

## 2.2 Definições e Resultados da Teoria Ergódica

Nesta seção, veremos alguns conceitos básicos da Teoria Ergódica e estaremos considerando sempre processos fortemente estacionários. Apresentamos várias definições importantes como ergodicidade e “mixing”. Os conceitos e propriedades apresentados nesta seção serão utilizados nas Seções 3.1 e 4.1 dos Capítulos 3 e 4, respectivamente.

Seja  $\{X_t\}_{t \in \mathbb{N}}$  um processo estocástico tal que  $X_t \in S$ , para todo  $t \in \mathbb{N}$ , onde  $S$  é um conjunto qualquer munido de uma  $\sigma$ -álgebra  $\mathcal{F}$ . Seja  $\Omega = S^{\mathbb{N}}$  o espaço amostral dos caminhos. Chamamos de cilindro em  $\Omega$  um conjunto da forma

$$\{X_{t_1} \in A_1, X_{t_2} \in A_2, \dots, X_{t_k} \in A_k\} \subset \Omega,$$

onde  $A_i$  são subconjuntos de  $\mathcal{F}$ ,  $X_{t_i} \in S$  e  $t_i \in \mathbb{N}$  para  $1 \leq i \leq k$ .

Seja  $\mathcal{A}$  a  $\sigma$ -álgebra gerada pelos cilindros em  $\Omega$ . Vamos definir  $\mathbb{P}$  sobre  $\mathcal{A}$ , que se denominará *probabilidade associada ao processo estocástico*  $\{X_t\}_{t \in \mathbb{N}}$ . O processo  $\{X_t\}_{t \in \mathbb{N}}$  define  $\mathbb{P}$  sobre cilindros:

$$\mathbb{P}(X_{t_1} \in A_1, X_{t_2} \in A_2, \dots, X_{t_k} \in A_k), \text{ para } k, t_i \in \mathbb{N},$$

onde  $t_1 < t_2 < \dots < t_k$ ,  $A_i$  subconjuntos de  $S$ ,  $1 \leq i \leq k$ .

Pelo Teorema da Extensão (ver Billingsley (1995)), fica definida então, de maneira única, uma probabilidade  $\mathbb{P}$  sobre  $\mathcal{A}$ .

**DEFINIÇÃO 2.16:** Seja  $(\Omega, \mathcal{A})$  um espaço mensurável. Uma transformação  $\varphi : \Omega \rightarrow \Omega$  é chamada *mensurável* se e só se para todo  $A \in \mathcal{A}$ ,  $\varphi^{-1}(A) \in \mathcal{A}$ .

$\mathcal{A}$ .

Como é usual, denota-se

$$\varphi^i = \underbrace{\varphi \circ \cdots \circ \varphi}_{i\text{-vezes}}.$$

**DEFINIÇÃO 2.17:** Seja  $(\Omega, \mathcal{A}, \mathbb{P})$  um espaço de probabilidade. Dizemos que a transformação  $\varphi : \Omega \rightarrow \Omega$  *preserva a probabilidade  $\mathbb{P}$*  se e só se  $\mathbb{P}(\varphi^{-1}(A)) = \mathbb{P}(A)$ , para todo  $A \in \mathcal{A}$ . Dizemos também que  $\mathbb{P}$  é invariante para  $\varphi$ .

**DEFINIÇÃO 2.18:** Seja  $\Omega = S^{\mathbb{N}}$  e  $\mathcal{A}$  a  $\sigma$ -álgebra gerada pelos cilindros. A transformação  $\tau : \Omega \rightarrow \Omega$  é chamada *operador shift* se ela é definida por  $\tau(\omega) = (\omega_2, \omega_3, \omega_4, \dots)$ , para todo  $\omega = (\omega_1, \omega_2, \dots) \in \Omega$ .

**Observações:**

a) É fácil ver que o operador shift  $\tau$  é mensurável no sentido da Definição 2.16.

b) Considere  $\Omega$  o espaço de Bernoulli com  $d = 2$ , isto é,  $\Omega = \{H, T\}^{\mathbb{N}}$  onde  $H$  é o evento “sair cara” e  $T$  é o evento “sair coroa” no lançamento de uma moeda. Considere a probabilidade  $\mathbb{P}$  tal que  $\mathbb{P}(H) = p$  e  $\mathbb{P}(T) = 1 - p$ , com  $0 < p < 1$ . Então, a probabilidade produto  $\mathbb{P}$ , definida no processo estocástico  $\{X_t\}_{t \in \mathbb{N}}$ , onde as variáveis aleatórias estão definidas no espaço de probabilidade  $(\Omega, \mathcal{A}, \mathbb{P})$ , com  $\mathcal{A}$  a  $\sigma$ -álgebra gerada pelos cilindros, é *invariante para o operador shift  $\tau$*  (ver Durrett (1996)). *Tal processo  $\{X_t\}_{t \in \mathbb{N}}$  é independente.*

c) É fácil ver que, quando consideramos o operador shift  $\tau$  agindo no espaço  $\Omega = S^{\mathbb{N}}$  e uma probabilidade  $\mathbb{P}$  qualquer sobre  $(\Omega, \mathcal{A})$  (não necessariamente independente), obtida a partir do processo estocástico  $\{X_t\}_{t \in \mathbb{N}}$ , então *este processo é fortemente estacionário se e só se o operador shift  $\tau$  preserva a probabilidade  $\mathbb{P}$ .*

d) Dizemos que o conjunto  $A \in \mathcal{A}$  é *shift-invariante* se  $\tau^{-1}(A) = A$ .

**DEFINIÇÃO 2.19:** Considere  $\{X_t\}_{t \in \mathbb{N}}$  um processo estocástico (fortemente) estacionário, definido em  $(\Omega, \mathcal{A}, \mathbb{P})$ , onde  $\Omega = S^{\mathbb{N}}$ ,  $\mathcal{A}$  é a  $\sigma$ -álgebra gerada pelos cilindros e  $\mathbb{P}$  é a probabilidade definida por este processo. Então, a probabilidade  $\mathbb{P}$  é *ergódica (para o operador shift  $\tau$ )* se e só se  $\mathbb{P}\{(X_1, X_2, \dots) \in A\} = 0$  ou  $\mathbb{P}\{(X_1, X_2, \dots) \in A\} = 1$  sempre que  $A$  é shift invariante.

A seguir, daremos uma definição mais geral de probabilidade ergódica.

**DEFINIÇÃO 2.20:** Seja  $\varphi : \Omega \rightarrow \Omega$  uma transformação que preserva a probabilidade  $\mathbb{P}$ . Dizemos que a probabilidade  $\mathbb{P}$  é *ergódica* se e só se para todo  $A \in \mathcal{A}$  tal que  $\varphi^{-1}(A) = A$  tem-se  $\mathbb{P}(A) = 0$  ou  $\mathbb{P}(A) = 1$ .

**Observação:** Pode-se mostrar que todo processo estocástico cujas variáveis aleatórias são *independentes entre si* (isto é, um processo estocástico independente) é *ergódico* (ver Durrett (1996)).

Ao leitor interessado em mais exemplos e detalhes referenciamos o interessante artigo de Borovkova et al. (2001). Este artigo trata de sistemas dinâmicos caóticos e apresenta resultados que podem ser utilizados para o estudo de aspectos probabilísticos destes sistemas.

Enunciaremos agora um teorema que estabelece equivalências para a condição de ergodicidade de um processo estocástico (ver Karlin e Taylor (1975)). Denotamos por  $\mathbb{I}_A$  a função indicadora do conjunto  $A$ , isto é,  $\mathbb{I}_A = 1$ , se  $z \in A$ , e  $\mathbb{I}_A = 0$ , se  $z \notin A$ .

**TEOREMA 2.3:** *Seja  $\{X_t\}_{t \in \mathbb{N}}$  um processo (fortemente) estacionário para a probabilidade  $\mathbb{P}$ . As seguintes condições são equivalentes:*

- (a)  $\{X_t\}_{t \in \mathbb{N}}$  é ergódico;
- (b) Para todo conjunto  $A \in \mathcal{A}$  e para todo  $(X_0, X_1, \dots)$   $\mathbb{P}$ -quase toda parte, temos

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=0}^{n-1} \mathbb{I}_A\{(X_i, X_{i+1}, \dots)\} = \mathbb{P}\{A\};$$

- (c) Para todo  $k \in \mathbb{N} - \{0\}$  e toda função  $g : \mathbb{R}^{k+1} \rightarrow \mathbb{R}$  mensurável e para todo  $(X_0, X_1, \dots)$   $\mathbb{P}$ -quase toda parte, temos

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=0}^{n-1} g(X_i, X_{i+1}, \dots, X_{i+k}) = \mathbb{E}[g],$$

sempre que a esperança existir;

- (d) Para toda função  $g : \Omega \rightarrow \mathbb{R}$  mensurável e para todo  $(X_0, X_1, \dots)$   $\mathbb{P}$ -quase toda parte, temos

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=0}^{n-1} g(\tau^i(X_0, X_1, \dots)) = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=0}^{n-1} g(X_i, X_{i+1}, \dots) = \mathbb{E}[g],$$

sempre que a esperança existir, onde  $\tau$  é o operador shift.

A caracterização no item (d) acima indica que um processo ergódico é aquele para o qual a média de uma única série temporal, extraída deste processo, permite estimar o seu valor esperado.

A seguir daremos a definição para uma transformação  $\varphi$  qualquer, que preserva uma probabilidade  $\mathbb{P}$ , ser “*mixing*”. Esta é uma propriedade de independência assintótica mais forte do que a de ergodicidade.

**DEFINIÇÃO 2.21:** Seja  $\varphi : (\Omega, \mathcal{A}) \rightarrow (\Omega, \mathcal{A})$  uma transformação que preserva a probabilidade  $\mathbb{P}$ . A transformação  $\varphi$  é chamada *mixing* se, para todos os conjuntos  $A, B \in \mathcal{A}$ ,

$$\lim_{n \rightarrow \infty} \mathbb{P}(A \cap \varphi^{-n}(B)) = \mathbb{P}(A) \mathbb{P}(B). \quad (2.8)$$

O processo  $\{X_t\}_{t \in \mathbb{N}}$  é dito ser *mixing* se o correspondente *operador shift*  $\varphi = \tau$  no espaço das seqüências  $S^{\mathbb{N}}$  é *mixing*.

Outra maneira de expressar (2.8) é

$$\lim_{n \rightarrow \infty} \int [\mathbb{I}_A(x) \mathbb{I}_B(\varphi^n(x))] d\mathbb{P}(x) = \int \mathbb{I}_A(x) d\mathbb{P}(x) \int \mathbb{I}_B(x) d\mathbb{P}(x),$$

que descreve o decaimento da função de autocorrelação das variáveis  $\mathbb{I}_A$  e  $\mathbb{I}_B$ .

**Observação:** A propriedade *mixing* é uma forma de independência assintótica, como vemos pela expressão (2.8) acima. Observe também que *todo processo mixing é ergódico*. De fato: seja  $\varphi : \Omega \rightarrow \Omega$  uma transformação que preserva a probabilidade  $\mathbb{P}$  e seja  $A \in \mathcal{A}$  tal que  $\varphi^{-1}(A) = A$ . Então, para todo  $n \in \mathbb{N}$ , temos, por indução, que

$$\varphi^{-n}(A) = \varphi^{-1}(\varphi^{-n+1}(A)) = \varphi^{-n+1}(A) = \dots = A.$$

Da expressão (2.8) segue que

$$\mathbb{P}(A) = \lim_{n \rightarrow \infty} \mathbb{P}(A \cap A) = \lim_{n \rightarrow \infty} \mathbb{P}(A \cap \varphi^{-n}(A)) = \mathbb{P}(A) \mathbb{P}(A).$$

Mas  $\mathbb{P}(A) = [\mathbb{P}(A)]^2$  implica em  $\mathbb{P}(A) = 0$  ou  $\mathbb{P}(A) = 1$ . E, portanto, o processo estocástico  $\{X_t\}_{t \in \mathbb{N}}$  *mixing*, associado à probabilidade  $\mathbb{P}$ , é ergódico conforme a Definição 2.20.

**Observação Importante:** Se o processo  $\{X_t\}_{t \in \mathbb{Z}}$  estiver definido para  $t \in \mathbb{Z}$  então todas as definições e resultados vistos nesta seção são análogos, bastando considerar  $\Omega = S^{\mathbb{Z}}$ .

# 3

## PROCESSOS ARFIMA(p,d,q)

Neste capítulo apresentamos uma importante classe de processos estocásticos, os *processos autoregressivos de médias móveis (ARMA)* introduzidos na análise de séries temporais por G.E.P. Box e G.M. Jenkins na década de 70. Mandelbrot (1965) utilizou os conceitos de integração e diferenciação fracionária para criar processos em tempo contínuo de características anteriormente inexplicáveis pelos processos existentes. Mandelbrot (1965) e Mandelbrot e van Ness (1968) definiram o movimento Browniano fracionário para explicar o fenômeno observado por Hurst (1951) na série temporal de níveis do rio Nilo. O movimento Browniano fracionário é um processo estocástico estacionário e contínuo no tempo, cuja função de autocorrelação decai hiperbolicamente, isto é, é um processo com característica de longa dependência. Mandelbrot e Wallis (1969) definiram o ruído Gaussiano fracionário, que é uma versão análoga ao movimento Browniano fracionário mas agora para um tempo discreto, mostrando que este processo exibe o fenômeno observado por H. E. Hurst. Posteriormente, de forma semelhante e independente, Granger e Joyeux (1980) e Hosking (1981) criaram processos em tempo discreto com base na integração e diferenciação fracionária. Estes modelos são uma generalização dos modelos *ARMA* e *ARIMA*, definidos em Box et al. (1995), e são conhecidos como *processos autoregressivos de médias móveis com integração fracionária*, denotados por *ARFIMA(p, d, q)*. Estes processos apresentam a propriedade de longa dependência dada pela Definição 2.15, nosso maior interesse neste trabalho. Maiores detalhes e teoremas para a melhor compreensão dos modelos *ARFIMA(p, d, q)* são apresentados em Beran (1994), Brockwell e Davis (1991), Hosking (1981), Reisen (1994) e Yajima (1985).

Nosso objetivo é definir os processos *ARFIMA(p, d, q)*. Para isso, precisamos antes das definições seguintes.

**DEFINIÇÃO 3.1:** O processo  $\{X_t\}_{t \in \mathbb{Z}}$  é um processo *Gaussiano* se e somente se as funções de distribuição  $n$ -dimensional de  $(X_{t_1}, X_{t_2}, \dots, X_{t_n})$  para

todo  $n \in \mathbb{N} - \{0\}$  e todo  $t_i \in \mathbb{Z}$ , para  $1 \leq i \leq n$ , são todas normais multivariadas.

**DEFINIÇÃO 3.2:** Uma família de variáveis aleatórias  $\{\epsilon_t\}_{t \in \mathbb{Z}}$  é dita ser um processo estocástico *ruído branco* de média zero se

- i)  $\mathbb{E}(\epsilon_t) = 0$ , para todo  $t \in \mathbb{Z}$ ,
- ii)  $\gamma_\epsilon(h) = \begin{cases} \sigma_\epsilon^2, & \text{se } h = 0, \\ 0, & \text{caso contrário,} \end{cases}$

ou seja, sua esperança é zero, sua variância é constante e as variáveis aleatórias deste processo são não correlacionadas.

Apresentamos, a seguir, processos estocásticos definidos em termos de equações diferenças lineares com coeficientes constantes, chamados *processos autoregressivo média móvel de ordem p e q*, denotados por  $ARMA(p, q)$ . Para maiores detalhes, ver Brockwell e Davis (1991) e Box et al. (1995).

**DEFINIÇÃO 3.3:** O processo estocástico  $\{X_t\}_{t \in \mathbb{Z}}$  é dito ser um processo *autoregressivo média móvel de ordem p e q*, com média zero, denotado por  $ARMA(p, q)$ , se é um processo estacionário tal que

$$X_t - \phi_1 X_{t-1} - \dots - \phi_p X_{t-p} = \epsilon_t + \theta_1 \epsilon_{t-1} + \dots + \theta_q \epsilon_{t-q}, \quad (3.1)$$

para qualquer  $t \in \mathbb{Z}$ , onde  $\{\epsilon_t\}_{t \in \mathbb{Z}}$  é um processo estocástico ruído branco,  $\phi_i$ ,  $1 \leq i \leq p$  e  $\theta_j$ ,  $1 \leq j \leq q$ , são constantes reais.

A expressão (3.1) pode ser reescrita por

$$\Phi(\mathcal{B})X_t = \Theta(\mathcal{B})\epsilon_t, \quad t \in \mathbb{Z},$$

onde  $\Phi(\mathcal{B})$  e  $\Theta(\mathcal{B})$  são polinômios de ordens  $p$  e  $q$ , respectivamente, dados por

$$\Phi(z) = \sum_{i=0}^p (-1)\phi_i z^i, \quad \text{para todo } p \in \mathbb{N},$$

onde  $\phi_0 \equiv -1$ , e

$$\Theta(z) = \sum_{i=0}^q \theta_i z^i, \quad \text{para todo } q \in \mathbb{N},$$

onde  $\theta_0 \equiv 1$ ; e  $\mathcal{B}$  é o operador defasagem, i.e.,  $\mathcal{B}X_t = X_{t-1}$ . Se  $\Theta(\mathcal{B}) \equiv 1$  o processo  $\Phi(\mathcal{B})X_t = \epsilon_t$  é dito ser um *processo autoregressivo de ordem p*, denotado por  $AR(p)$ . Da mesma forma, se  $\Phi(\mathcal{B}) \equiv 1$ , o processo  $X_t = \Theta(\mathcal{B})\epsilon_t$  é dito ser um *processo média móvel de ordem q*, denotado por  $MA(q)$ .

Para o processo  $ARMA(p, q)$  ser *invertível*, i.e., para existir uma seqüência  $\{\pi_k\}_{k \in \mathbb{N}}$  tal que  $\sum_{k \geq 0} \pi_k < \infty$  e

$$\epsilon_t = \sum_{k \geq 0} |\pi_k| X_{t-k} \text{ para todo } t \in \mathbb{Z},$$

as raízes da equação  $\Theta(\mathcal{B}) = 0$  devem estar fora do círculo unitário; para que o processo  $ARMA(p, q)$  seja *estacionário (causal)*, as raízes da equação  $\Phi(\mathcal{B}) = 0$  devem estar fora do círculo unitário. Assume-se que as equações  $\Phi(\mathcal{B}) = 0$  e  $\Theta(\mathcal{B}) = 0$  não possuem raízes em comum. No teorema, a seguir, é dada a função densidade espectral de um processo  $ARMA(p, q)$  e a demonstração pode ser encontrada em Brockwell e Davis (1991), página 123.

**TEOREMA 3.1:** *Sejam  $\{X_t\}_{t \in \mathbb{Z}}$  um processo estocástico estacionário  $ARMA(p, q)$  e  $\{\epsilon_t\}_{t \in \mathbb{Z}}$  o processo ruído branco com média zero e variância  $\sigma_\epsilon^2$ . A função densidade espectral de  $\{X_t\}_{t \in \mathbb{Z}}$  é dada por*

$$f_X(w) = \frac{\sigma_\epsilon^2}{2\pi} \left| \frac{\Theta(e^{-iw})}{\Phi(e^{-iw})} \right|^2, \text{ para todo } w \in [-\pi, \pi],$$

onde  $\Phi(\cdot)$  e  $\Theta(\cdot)$  são os polinômios do processo  $ARMA(p, q)$ .

Definimos agora o processo  $ARIMA(p, d, q)$ .

**DEFINIÇÃO 3.4:** Um processo estocástico  $\{X_t\}_{t \in \mathbb{Z}}$  é dito ser um *processo autoregressivo integrado média móvel* se a diferenciação  $\nabla^d X_t$  ( $d \in \mathbb{N} - \{0\}$  e  $\nabla = 1 - \mathcal{B}$ ) resultar em um processo  $ARMA(p, q)$ .

O modelo é expresso na forma

$$\Phi(\mathcal{B})(1 - \mathcal{B})^d X_t = \Theta(\mathcal{B})\epsilon_t, \text{ para todo } t \in \mathbb{Z}, \quad (3.2)$$

onde  $\Phi(\cdot)$  e  $\Theta(\cdot)$  são os polinômios dados na Definição 3.3 e  $\{\epsilon_t\}_{t \in \mathbb{Z}}$  um processo ruído branco com média zero e variância  $\sigma_\epsilon^2$ . O parâmetro  $d$ , um natural não nulo, é chamado *parâmetro ou grau de diferenciação*. O processo (3.2) pode ser reescrito na forma

$$\Phi(\mathcal{B})U_t = \Theta(\mathcal{B})\epsilon_t, \text{ para todo } t \in \mathbb{Z},$$

onde  $U_t \equiv \nabla^d X_t$  é um processo estacionário  $ARMA(p, q)$ . Denotamos o processo, dado pela expressão (3.2), por  $ARIMA(p, d, q)$ .

## O Processo ARFIMA(p,d,q)

Como os modelos  $ARIMA(p, d, q)$  não são flexíveis na modelagem de estruturas da função de autocorrelação de séries temporais que apresentam comportamento de longa dependência, estes modelos foram extendidos para os *modelos autoregressivos de médias móveis com integração fracionária* ver Granger e Joyeux (1980), Hosking (1981) e Beran (1994)). A generalização natural do processo  $ARIMA(p, d, q)$  é permitir que o parâmetro  $d$  assumia valores fracionários.

**DEFINIÇÃO 3.5:** Seja  $\{X_t\}_{t \in \mathbb{Z}}$  um processo estocástico dado por

$$\Phi(\mathcal{B})\nabla^d X_t = \Theta(\mathcal{B})\epsilon_t, \text{ para todo } t \in \mathbb{Z}, \quad (3.3)$$

onde  $\nabla^d \equiv (1 - \mathcal{B})^d$  é definido pela expansão binomial

$$\nabla^d = \sum_{k=0}^{\infty} \binom{d}{k} (-\mathcal{B})^k = 1 - d\mathcal{B} - \frac{d}{2!}(1-d)\mathcal{B}^2 \dots$$

e  $d \in \mathbb{R}$  é o *parâmetro ou grau de diferenciação*. O processo  $\{X_t\}_{t \in \mathbb{Z}}$ , dado pela expressão acima, é então chamado de *processo autoregressivo de médias móveis com integração fracionária*, com média zero, denotado por  $ARFIMA(p, d, q)$ .

O teorema a seguir (ver Hosking (1981)) apresenta várias propriedades dos processos  $\{X_t\}_{t \in \mathbb{Z}}$  dados pela expressão (3.3).

**TEOREMA 3.2:** *Seja  $\{X_t\}_{t \in \mathbb{Z}}$  um processo ARFIMA(p,d,q) dado pela expressão (3.3). Então,*

- (a)  $\{X_t\}_{t \in \mathbb{Z}}$  é estacionário se  $d < 0,5$  e todas as raízes da equação  $\Phi(z) = 0$  estão fora do círculo unitário;
- (b)  $\{X_t\}_{t \in \mathbb{Z}}$  é invertível se  $d > -0,5$  e todas as raízes da equação  $\Theta(z) = 0$  estão fora do círculo unitário;

*Se  $\{X_t\}_{t \in \mathbb{Z}}$  é estacionário e invertível, isto é,  $d \in (-0,5; 0,5)$ , com função densidade espectral  $f_X(\cdot)$  e função de autocorrelação  $\rho_X(\cdot)$ , então*

- (c)  $\lim_{\lambda \rightarrow 0} \lambda^{2d} f_X(\cdot)$  existe e é finito;
- (d)  $\lim_{k \rightarrow \infty} k^{1-2d} \rho_X(k)$  existe e é finito.

Consideremos agora  $p$  e  $q$  iguais a zero, de maneira que  $\Phi(z) \equiv 1 \equiv \Theta(z)$ . Temos então o processo  $ARFIMA(0, d, 0)$  com a seguinte expressão

$$\nabla^d X_t \equiv (1 - \mathcal{B})^d X_t = \epsilon_t, \text{ para todo } t \in \mathbb{Z}. \quad (3.4)$$

O teorema a seguir (ver Hosking (1981)) apresenta várias propriedades dos processos  $\{X_t\}_{t \in \mathbb{Z}}$ , dados pela expressão (3.4), onde, por conveniência, assumimos  $\sigma_\epsilon^2 = 1$ .

**TEOREMA 3.3:** *Seja  $\{X_t\}_{t \in \mathbb{Z}}$  um processo ARFIMA(0, d, 0) dado pela expressão (3.4).*

(a) *Quando  $d < 0,5$ ,  $\{X_t\}_{t \in \mathbb{Z}}$  é um processo estacionário e tem representação média móvel infinita dada por*

$$X_t = \psi(\mathcal{B})\epsilon_t = \sum_{k=0}^{\infty} \psi_k \epsilon_{t-k},$$

onde

$$\psi_k = \frac{d(1+d) \cdots (k-1+d)}{k!} = \frac{(k+d-1)!}{k!(d-1)!}.$$

Quando  $k \rightarrow \infty$ ,  $\psi_k \sim \frac{k^{d-1}}{(d-1)!}$ .

(b) *Quando  $d > -0,5$ ,  $\{X_t\}_{t \in \mathbb{Z}}$  é um processo invertível e tem representação autoregressiva infinita dada por*

$$\pi(\mathcal{B})X_t = \sum_{k=0}^{\infty} \pi_k X_{t-k} = \epsilon_t,$$

onde

$$\pi_k = \frac{-d(1-d) \cdots (k-1-d)}{k!} = \frac{(k-d-1)!}{k!(-d-1)!}.$$

Quando  $k \rightarrow \infty$ ,  $\pi_k \sim \frac{k^{-d-1}}{(-d-1)!}$ .

Nos itens (c), (d) e (e) abaixo, assumimos que  $d \in (-0,5; 0,5)$ .

(c) *A função densidade espectral de  $\{X_t\}_{t \in \mathbb{Z}}$  é dada por*

$$f_X(w) = \left[ 2 \operatorname{sen}\left(\frac{w}{2}\right) \right]^{-2d}, \text{ para } 0 < w \leq \pi. \quad (3.5)$$

Sendo assim, quando  $w$  está próximo de 0,  $f_X(w) \sim w^{-2d}$ .

(d) A função de autocovariância de ordem  $k$  de  $\{X_t\}_{t \in \mathbb{Z}}$  é dada por

$$\gamma_X(k) = \frac{(-1)^k (-2d)!}{(k-d)! (-k-d)!}, \text{ para todo } k \in \mathbb{N},$$

e a função de autocorrelação de ordem  $k$  é dada por

$$\rho_X(k) = \frac{(-d)! (k+d-1)!}{(d-1)! (k-d)!}, \text{ para todo } k \in \mathbb{N}.$$

Quando  $k \rightarrow \infty$ ,  $\rho_X(k) \sim \frac{(-d)!}{(d-1)!} k^{2d-1}$ .

(e) A função de autocorrelação parcial de ordem  $k$  de  $\{X_t\}_{t \in \mathbb{Z}}$  é dada por

$$\phi_X(k, k) = \frac{d}{k-d}, \text{ para todo } k \in \mathbb{N} - \{0\}.$$

O teorema a seguir mostra as propriedades do estimador da média do processo  $\{X_t\}_{t \in \mathbb{Z}}$  dado pela Definição 2.8. Isto é, ele mostra que o estimador, pelo método dos momentos para a média  $\mu$  do processo  $\{X_t\}_{t \in \mathbb{Z}}$ , é não viciado e consistente quando  $d \in (-0, 5; 0, 5)$ .

**TEOREMA 3.4:** *Seja  $\{X_t\}_{t \in \mathbb{Z}}$  um processo ARFIMA(0,  $d$ , 0) dado pela expressão (3.4), onde  $d \in (-0, 5; 0, 5)$ . Considere  $\text{Var}(X_t) = \mathbb{E}(X_t^2) = \gamma_X(0)$  e seja  $\bar{X}$  a média amostral do processo  $\{X_t\}_{t \in \mathbb{Z}}$  dada pela Definição 2.8. Então,*

(a)  $\mathbb{E}(\bar{X}) = 0$ .

(b)  $\text{Var}(\bar{X}) = \frac{\gamma_X(0)d}{n^2(1+2d)} \left(1 - \frac{(-d-1)!(n+d)!}{d!(n-d-1)!}\right) \sim \frac{\gamma_X(0)(-d)!}{(1+2d)d!} n^{2d-1}$ ,  
quando  $n \rightarrow \infty$ .

O Teorema 3.4 pode ser encontrado em Hosking (1982). O resultado (b) do Teorema 3.4 mostra que, para  $d \in (0, 0; 0, 5)$ , isto é, quando estamos no caso de *longa dependência*, a variância da média amostral decresce menos rapidamente do que  $n^{-1}$ . Já quando  $d \in (-0, 5; 0, 0)$ , isto é, quando estamos no caso de *dependência intermediária*, a variância da média amostral decresce mais rapidamente do que  $n^{-1}$ .

#### Observações:

a) Quando  $d = 0$  no processo ARFIMA( $p, d, q$ ) recaímos no processo ARMA( $p, q$ ), ou seja, processo de curta dependência.

b) Os itens (c) e (d) do Teorema 3.3 refletem a propriedade de longa dependência do processo, ou seja, o decaimento da função de autocorrelação é hiperbólico e a função densidade espectral é ilimitada nas frequências próximas de zero. O decaimento hiperbólico é mais lento em relação ao decaimento exponencial do processo  $ARMA(p, q)$ .

c) Se  $\{X_t\}_{t \in \mathbb{Z}}$  for um processo estocástico estacionário  $ARFIMA(p, d, q)$ ,  $d \in (-0, 5; 0, 5)$ , a função densidade espectral do processo é dada por

$$f_X(w) = f_U(w) \left[ 2 \operatorname{sen}\left(\frac{w}{2}\right) \right]^{-2d} = \frac{\sigma^2 |\Theta(e^{-iw})|^2}{2\pi |\Phi(e^{-iw})|^2} \left[ 2 \operatorname{sen}\left(\frac{w}{2}\right) \right]^{-2d}, \quad (3.6)$$

para todo  $0 < w \leq \pi$ , onde  $f_U(\cdot)$  denota a função densidade espectral do processo  $ARMA(p, q)$ , dado por  $U_t \equiv \nabla^d X_t$ , com

$$\Phi(\mathcal{B})U_t = \Theta(\mathcal{B})\epsilon_t, \text{ para todo } t \in \mathbb{Z}.$$

### 3.1 Ergodicidade

Nesta seção analisaremos a ergodicidade dos processos  $ARFIMA(p, d, q)$ . Primeiramente estudaremos a ergodicidade dos processos  $MA(q)$ ,  $MA(\infty)$  e  $AR(p)$ . Para isto necessitaremos dos teoremas a seguir (ver Durrett (1996)).

**TEOREMA 3.5:** *Sejam  $g : \Omega \rightarrow \mathbb{R}$  função mensurável e  $(X_0, X_1, \dots)$  uma seqüência estacionária. Então,  $Y_k = g(X_k, X_{k+1}, \dots)$  é uma seqüência estacionária.*

**TEOREMA 3.6:** *Seja  $g : \Omega \rightarrow \mathbb{R}$  função mensurável. Se  $(X_0, X_1, \dots)$ , é uma seqüência estacionária ergódica, então  $Y_k = g(X_k, X_{k+1}, \dots)$  é ergódica.*

As provas dos Teoremas 3.5 e 3.6 podem ser encontradas em Durrett (1996), página 340.

O lema, a seguir, determina a ergodicidade dos processos médias móveis de ordem  $q$ , denotado por  $MA(q)$ , e de ordem infinita, denotado por  $MA(\infty)$ .

**LEMA 3.1:** *Seja  $\{\epsilon_t\}_{t \in \mathbb{Z}}$  um processo estocástico onde as variáveis aleatórias  $\epsilon_t$ , com  $t \in \mathbb{Z}$ , são independentes e identicamente distribuídas com média zero e variância 1. O processo  $\{X_t\}_{t \in \mathbb{Z}}$  dado por*

$$X_t = \sum_{k=0}^q a_k \epsilon_{t-k}, \text{ para todo } t \in \mathbb{Z},$$

é estacionário e ainda ergódico. Mais geralmente, se ainda  $\sum_{k \in \mathbb{N}} a_k^2 < \infty$  então o processo  $Y_t = \sum_{k \in \mathbb{N}} a_k \epsilon_{t-k}$  também é ergódico.

**Prova:** Seja  $(\Omega, \mathcal{A}, \mathbb{P})$  um espaço de probabilidade. Sabemos que  $\epsilon_t$  são variáveis aleatórias definidas em  $(\Omega, \mathcal{A}, \mathbb{P})$  tais que  $\mathbb{E}(\epsilon_t) = 0$  e  $\text{Var}(\epsilon_t) = 1$ , para todo  $t \in \mathbb{Z}$ .

Consideremos  $\tau : \Omega \rightarrow \Omega$  o operador shift (sabemos, pelo Teorema 3.5, que  $\mathbb{P}$  é invariante sob  $\tau$ ). Vamos mostrar que  $\tau$  é mixing (portanto, é ergódica pela observação feita após a Definição 2.21).

Sejam  $\mathcal{C}$  o conjunto dos cilindros e  $\mathcal{A}$  a  $\sigma$ -álgebra gerada por  $\mathcal{C}$ . Sejam  $A$  e  $B$  em  $\mathcal{C}$ . Vamos mostrar que

$$\lim_{i \rightarrow \infty} \mathbb{P}(\tau^{-i}(A) \cap B) = \mathbb{P}(A)\mathbb{P}(B).$$

Para isto vamos mostrar que  $\mathbb{P}(\tau^{-i}(A) \cap B) = \mathbb{P}(A)\mathbb{P}(B)$ , para todo  $i$  suficientemente grande.

Suponha que  $A$  e  $B$  sejam dois cilindros da forma

$$A = A_1 \times A_2 \times \cdots \times A_m \times \mathbb{R} \times \mathbb{R} \times \cdots$$

e

$$B = B_1 \times B_2 \times \cdots \times B_{m'} \times \mathbb{R} \times \mathbb{R} \times \cdots,$$

para  $m$  e  $m'$  naturais fixos. Para  $i > m + m'$  temos

$$\tau^{-i}(A) = \underbrace{\mathbb{R} \times \mathbb{R} \times \cdots \times \mathbb{R}}_{i\text{-vezes}} \times \underbrace{A_1 \times A_2 \times \cdots \times A_m}_{m\text{-vezes}} \times \mathbb{R} \times \mathbb{R} \times \cdots.$$

Assim, para  $i > m + m'$  temos

$$\begin{aligned} \tau^{-i}(A) \cap B &= \underbrace{B_1 \times B_2 \times \cdots \times B_{m'}}_{m'\text{-vezes}} \times \underbrace{\mathbb{R} \times \mathbb{R} \times \cdots \times \mathbb{R}}_{i-m'\text{-vezes}} \\ &\quad \times \underbrace{A_1 \times A_2 \times \cdots \times A_m}_{m\text{-vezes}} \times \mathbb{R} \times \mathbb{R} \times \cdots. \end{aligned} \quad (3.7)$$

Considerando a probabilidade do evento dado pela expressão (3.7) acima, temos que

$$\begin{aligned} \mathbb{P}(\tau^{-i}(A) \cap B) &= \mathbb{P}(X_1 \in B_1, X_2 \in B_2, \cdots, X_{m'} \in B_{m'}, \\ &\quad X_{i+1} \in A_1, X_{i+2} \in A_2, \cdots, X_{i+m} \in A_m). \end{aligned}$$

Sejam os eventos

$$\tau^{-i}(A) = C = [X_{i+1} \in A_1, X_{i+2} \in A_2, \dots, X_{i+m} \in A_m] \in \sigma(\epsilon_{i+1-q}, \dots, \epsilon_{i+m})$$

e

$$D = [X_1 \in B_1, X_2 \in B_2, \dots, X_{m'} \in B_{m'}] \in \sigma(\epsilon_1, \dots, \epsilon_{m'}).$$

Desde que  $m' < i+1-q$  (isto é,  $i > m'+q-1$ ), os eventos  $C$  e  $D$  pertencem a  $\sigma$ -álgebras independentes e são, portanto, eventos independentes. Então,

$$\mathbb{P}(\tau^{-i}(A) \cap B) = \mathbb{P}(C \cap D) = \mathbb{P}(C)\mathbb{P}(D) = \mathbb{P}(\tau^{-i}(A))\mathbb{P}(B) \quad (3.8)$$

para todo  $i > m' + q - 1$ . Como  $\mathbb{P}(\tau^{-i}(A)) = \mathbb{P}(A)$ , pois  $\mathbb{P}$  é invariante sob  $\tau$ , pela igualdade (3.8) temos que

$$\lim_{i \rightarrow \infty} \mathbb{P}(\tau^{-i}(A) \cap B) = \mathbb{P}(A)\mathbb{P}(B).$$

Como vale a igualdade (2.8), concluímos que a transformação  $\tau$  é “mixing”. Logo,  $\tau$  é ergódica e concluímos, portanto, que o processo  $\{X_t\}_{t \in \mathbb{Z}}$  é ergódico, ou seja, todo processo  $MA(q)$  é ergódico.

Suponhamos agora que  $\sum_{k \in \mathbb{N}} a_k^2 < \infty$ . Segue de Durrett (1996) que  $Y_t = \sum_{k \in \mathbb{N}} a_k \epsilon_{t-k}$  converge quase certamente. Pelos Teoremas 3.5 e 3.6,  $\{Y_t\}_{t \in \mathbb{Z}}$  é estacionário e ergódico. Portanto, todo processo média móvel de ordem infinita é ergódico.

□

**COROLÁRIO 3.1:** *Seja  $\{X_t\}_{t \in \mathbb{Z}}$  um processo estocástico autoregressivo de ordem  $p$ , denotado por  $AR(p)$ , tal que as raízes da equação  $\Phi(\mathcal{B})$  estejam fora do círculo unitário. Então, o processo  $\{X_t\}_{t \in \mathbb{Z}}$  é ergódico.*

**Prova:** Como todo processo autoregressivo tem uma representação média móvel (ver Brockwell e Davis (1991)), pelo Lema 3.1 segue que  $AR(p)$  é também um processo ergódico.

□

Queremos agora mostrar que todo processo  $ARFIMA(0, d, 0)$ , dado pela expressão (3.4), é ergódico.

**PROPOSIÇÃO 3.1:** *Seja  $\{X_t\}_{t \in \mathbb{Z}}$  um processo  $ARFIMA(0, d, 0)$ , dado pela expressão (3.4), com  $d < 0,5$ . Então,  $\{X_t\}_{t \in \mathbb{Z}}$  é ergódico.*

**Prova:** Pelo Teorema 3.3, parte (a), o processo  $\{X_t\}_{t \in \mathbb{Z}}$  pode ser escrito na forma

$$X_t = \sum_{k=0}^{\infty} \psi_k \epsilon_{t-k}, \text{ para todo } t \in \mathbb{Z}, \quad (3.9)$$

onde  $\psi_k = \frac{d(1+d)\cdots(k-1+d)}{k!} = \frac{(k+d-1)!}{k!(d-1)!}$ . Pelo Lema 3.1 basta mostrar que  $\sum_{k \in \mathbb{N}} \psi_k^2 < \infty$  para que o processo  $X_t = \sum_{k \in \mathbb{N}} \psi_k \epsilon_{t-k}$ , para todo  $t \in \mathbb{Z}$ , convirja quase certamente.

Observe que

$$\begin{aligned} \sum_{k=0}^{\infty} \psi_k^2 &= \sum_{k=0}^{\infty} \left[ \frac{(k+d-1)!}{k!(d-1)!} \right]^2 = \sum_{k=0}^{\infty} \left[ \frac{\Gamma(k+d)}{\Gamma(k+1)\Gamma(d)} \right]^2 = \\ &= \frac{1}{\Gamma^2(d)} \sum_{k=0}^{\infty} \left[ \frac{\Gamma(k+d)}{\Gamma(k+1)} \right]^2, \end{aligned} \quad (3.10)$$

onde  $\Gamma(\cdot)$  é a função Gama definida por

- i)  $\Gamma(x) = \int_0^{\infty} t^{x-1} e^{-t} dt$ ,  $x > 0$ ;
- ii)  $\Gamma(x+1) = x\Gamma(x)$ .

Pela fórmula de Stirling (ver Brockwell e Davis (1991)) segue que

$$\Gamma(x) \sim \sqrt{2\pi} \exp(-x+1)(x-1)^{x-0,5}, \text{ quando } x \rightarrow \infty,$$

onde “ $\sim$ ” indica que a razão dos dois termos tende a 1 quando  $x \rightarrow \infty$ . Desta forma,  $\frac{\Gamma(x+a)}{\Gamma(x+b)}$  é aproximadamente igual a  $x^{a-b}$ , quando  $x \rightarrow \infty$ . Portanto, da expressão (3.10) temos que

$$\sum_{k \in \mathbb{N}} \psi_k^2 \sim C + \frac{1}{\Gamma^2(d)} \sum_{k=N}^{\infty} [k^{d-1}]^2 = C + \frac{1}{\Gamma^2(d)} \sum_{k=N}^{\infty} k^{2d-2},$$

onde  $C = \frac{1}{\Gamma^2(d)} \sum_{k=0}^{N-1} \left[ \frac{\Gamma(k+d)}{\Gamma(k+1)} \right]^2$  e  $N$  é um inteiro suficientemente grande. A série  $\sum_{k=N}^{\infty} k^{2d-2}$  converge se e somente se  $2d-2 < -1$ . Como temos  $d < 0,5$ , segue que  $\sum_{k \in \mathbb{N}} \psi_k^2 < \infty$ .

Portanto, pelo Lema 3.1, concluímos que o processo  $ARFIMA(0, d, 0)$  é um processo ergódico. □

**Conclusão:** Podemos concluir que, além do processo  $ARFIMA(0, d, 0)$ , o processo  $ARFIMA(p, d, q)$  também é um processo ergódico se o processo  $ARMA(p, q)$  correspondente é estacionário. De fato, neste caso, a ordem dos coeficientes  $\psi_k$  é a mesma que para o processo  $ARFIMA(0, d, 0)$  (ver Hosking (1981)) e, analogamente, temos a convergência da série  $\sum_{k \in \mathbb{N}} \psi_k^2$ .

Na seção a seguir queremos mostrar a ordem de magnitude da variância da soma parcial de um processo  $\{X_t\}_{t \in \mathbb{Z}}$  dado pela expressão (3.3).

## 3.2 Ordem da Variância da Soma Parcial

Considerando a velocidade com que a função de autocorrelação converge a zero podemos ter uma idéia da ordem de magnitude da variância de  $S_n = \sum_{i=0}^{n-1} X_i$  para uma amostra  $X_0, X_1, \dots, X_{n-1}$  de um processo estocástico *ARFIMA*. Beran (1994) apresenta um resultado análogo para a variância de  $\overline{X}_n = n^{-1}S_n$ . A demonstração apresentada abaixo é diferente e mais geral.

**PROPOSIÇÃO 3.2:** *Seja  $\{X_t\}_{t \in \mathbb{Z}}$  um processo estacionário qualquer. Considere a variável aleatória soma parcial  $S_n = \sum_{i=0}^{n-1} X_i$  de uma série temporal  $X_0, X_1, \dots, X_{n-1}$  do processo  $\{X_t\}_{t \in \mathbb{Z}}$ . Então,*

$$\text{Var}(S_n) = 2n \left[ \frac{\gamma_X(0)}{2} + \frac{1}{n} \sum_{j=1}^{n-1} (n-j) \gamma_X(j) \right],$$

onde  $\gamma_X(\cdot)$  é a função de autocovariância do processo  $\{X_t\}_{t \in \mathbb{Z}}$ .

**Prova:** Como o processo  $\{X_t\}_{t \in \mathbb{Z}}$  é estacionário, observe que

$$\begin{aligned} \text{Var}(S_n) &= \text{Var} \left( \sum_{i=0}^{n-1} X_i \right) \\ &= \sum_{i=0}^{n-1} \text{Var}(X_i) + \sum_{j=0}^{n-1} \sum_{\ell=0}^{n-1} \text{cov}(X_j, X_\ell) \\ &= n \text{Var}(X_0) + \sum_{j=0}^{n-1} \sum_{\ell=0}^{n-1} (\mathbb{E}(X_j X_\ell) - [\mathbb{E}(X_0)]^2) \\ &= n \gamma_X(0) + 2 \sum_{\substack{j, \ell=0 \\ j < \ell}}^{n-1} \gamma_X(j - \ell). \end{aligned} \tag{3.11}$$

Segue da expressão (3.11) que

$$\begin{aligned} \text{Var}(S_n) &= n \gamma_X(0) + 2 \sum_{\substack{j, \ell=0 \\ j < \ell}}^{n-1} \gamma_X(j - \ell) = \\ &= n \gamma_X(0) + 2 \left( \underbrace{\gamma_X(-1) + \gamma_X(-2) + \gamma_X(-3) + \dots + \gamma_X(-n+1)}_{j=0} \right) + \\ &+ \underbrace{\gamma_X(-1) + \gamma_X(-2) + \dots + \gamma_X(1 - (n-1))}_{j=1} \end{aligned}$$

$$\begin{aligned}
& + \underbrace{\gamma_X(-1) + \gamma_X(-2) + \cdots + \gamma_X(2 - (n-1))}_{j=2} + \\
& + \underbrace{\gamma_X(-1) + \gamma_X(-2) + \cdots + \gamma_X(3 - (n-1))}_{j=3} + \cdots + \underbrace{\gamma_X(-1)}_{j=n-2} \Big) = \\
& = n \gamma_X(0) + 2 [(n-1)\gamma_X(-1) + (n-2)\gamma_X(-2) + (n-3)\gamma_X(-3) + \\
& + \cdots + 3\gamma_X(-(n-3)) + 2\gamma_X(-(n-2)) + \gamma_X(-(n-1))] = \\
& = n \gamma_X(0) + 2 \sum_{j=1}^{n-1} (n-j) \gamma_X(-j) = \\
& = n \gamma_X(0) + 2 \sum_{j=1}^{n-1} (n-j) \gamma_X(j). \tag{3.12}
\end{aligned}$$

A última igualdade segue do fato de estarmos considerando o caso em que o processo é estacionário, o que implica que  $\gamma_X(j) = \gamma_X(-j)$ .

Portanto,

$$Var(S_n) = n \gamma_X(0) + 2 \sum_{j=1}^{n-1} (n-j) \gamma_X(j),$$

o que prova a Proposição 3.2. □

Na proposição, a seguir, apresentamos a ordem de  $Var(S_n)$  para processos  $ARFIMA(p, d, q)$ .

**PROPOSIÇÃO 3.3:** *Seja  $\{X_t\}_{t \in \mathbb{Z}}$  um processo  $ARFIMA(p, d, q)$  onde  $d \in (0, 0.5)$  dado pela expressão (3.3). Considere a soma parcial  $S_n = \sum_{i=0}^{n-1} X_i$  de uma série temporal  $X_0, X_1, \dots, X_{n-1}$  do processo  $\{X_t\}_{t \in \mathbb{Z}}$ . Então,*

$$Var(S_n) \approx n^{2d+1}.$$

**Prova:** Observe que, pela parte (d) do Teorema 3.2, a função de autocovariância  $\gamma_X(k)$  do processo  $\{X_t\}_{t \in \mathbb{Z}}$  é da ordem de  $k^{2d-1}$ .

Para  $0 < d < 0.5$ , a integral

$$I = \int_0^1 (1-x)x^{2d-1} dx$$

é finita. Podemos considerar, então, as somas de Riemann associadas à partição

$$\left\{ 0, \frac{1}{n}, \frac{2}{n}, \dots, \frac{n-1}{n}, 1 \right\},$$

obtendo a aproximação

$$\sum_{j=1}^n \left(1 - \frac{j}{n}\right) \left(\frac{j}{n}\right)^{2d-1} \frac{1}{n},$$

que converge para  $I$ , quando  $n \rightarrow \infty$ .

De maneira análoga ao Lema 8.1 de Fisher e Lopes (2001), considere

$$\begin{aligned} c_n &= \sum_{j=1}^n \left(1 - \frac{j}{n}\right) \left(\frac{j}{n}\right)^{2d-1} \\ &= \sum_{j=1}^n \left(\frac{n-j}{n}\right) \left(\frac{j}{n}\right)^{2d-1} \\ &= \sum_{j=1}^n (n-j) j^{2d-1} \left(\frac{1}{n}\right)^{2d}. \end{aligned} \quad (3.13)$$

Dado  $\varepsilon > 0$ , para  $n$  suficientemente grande, temos que

$$I - \varepsilon \leq \frac{1}{n} c_n \leq I + \varepsilon.$$

Usando a expressão (3.13), a desigualdade acima é dada por

$$(I - \varepsilon) n^{2d} \leq \frac{1}{n} \sum_{j=1}^n (n-j) j^{2d-1} \leq (I + \varepsilon) n^{2d}, \quad (3.14)$$

para  $n$  suficientemente grande.

Portanto,  $\frac{1}{n} \sum_{j=1}^n (n-j) j^{2d-1}$  é da ordem de  $n^{2d}$ . A partir das expressões (3.12) e (3.14) temos que

$$\text{Var}(S_n) = 2n \left[ \frac{\gamma_X(0)}{2} + \frac{1}{n} \sum_{j=1}^{n-1} (n-j) \gamma_X(j) \right] \approx n^{2d+1}, \quad (3.15)$$

o que prova a Proposição 3.3. □

**Observação:** O resultado acima vale, em particular, para processos  $ARFIMA(0, d, 0)$ ,  $d \in (0, 0; 0, 5)$ , onde  $\gamma_X(0) = \sigma_\epsilon^2 \frac{\Gamma(1-2d)}{\Gamma^2(1-2d)}$  é a variância do processo,  $\sigma_\epsilon^2$  é a variância do processo ruído branco  $\{\epsilon_t\}_{t \in \mathbb{Z}}$  e  $\gamma_X(k) = \sigma_\epsilon^2 \frac{(-1)^k (-2d)!}{(k-d)! (-k-d)!}$  para todo  $k \in \mathbb{N} - \{0\}$  (ver parte (d) do Teorema 3.3) é a função de autocovariância do processo  $\{X_t\}_{t \in \mathbb{Z}}$  dado pela expressão (3.4). Na seção seguinte vamos usar o resultado da ordem da variância de  $S_n$ , dada pela Proposição 3.3, para obter um estimador para o parâmetro  $d$ .

Mais geralmente, vale a seguinte proposição:

**PROPOSIÇÃO 3.4:** *Seja  $\{X_t\}_{t \in \mathbb{Z}}$  um processo estocástico estacionário qualquer. Considere a soma parcial  $S_n = \sum_{i=0}^{n-1} X_i$  de uma série temporal  $X_0, X_1, \dots, X_{n-1}$  do processo  $\{X_t\}_{t \in \mathbb{Z}}$ . Se existe  $u \in (0, 1)$  tal que  $\gamma_X(k) \approx k^{-u}$ , então vale que*

$$\text{Var}(S_n) \approx n^{2-u},$$

para todo  $n \in \mathbb{N}$ .

A demonstração é similar a prova da Proposição 3.3.

### 3.3 Estimação do Parâmetro de Diferenciação

Veremos, a seguir, a estimação do *parâmetro ou grau de diferenciação* através de diversos métodos: métodos de regressão utilizando a função periodograma e a função periodograma suavizado (ver Definições 2.13 e 2.14); algumas versões destes estimadores propostos por Robinson (1995), Velasco (1999a, 1999b); o método da máxima verossimilhança (proposto por Dahlhaus (1989) e Fox e Taquq (1986)); o método baseado na variância da soma parcial (proposto a partir da Proposição 3.3); o método que sugere obter um estimador para o grau de diferenciação a partir do gráfico do logaritmo da variância de  $\bar{X}_n$  (proposto por Beran (1994)) e apresentamos, ainda, o método baseado na Teoria de Wavelets proposto por Jensen (1999).

#### Estimador *GPH*

O método de regressão utilizando a função periodograma foi primeiramente estudado por Geweke e Porter-Hudak (1983) e tem sido largamente utilizado por pesquisadores. Usaremos a notação *GPH* para este estimador. Consideremos  $\{X_t\}_{t \in \mathbb{Z}}$  um processo *ARFIMA*( $p, d, q$ ), com  $d \in (-0, 5; 0, 5)$ , representado pela expressão (3.2). A função densidade espectral de  $\{X_t\}_{t \in \mathbb{Z}}$  é dada pela expressão (3.6) e tomando o logaritmo daquela expressão temos

$$\ln f_X(w) = \ln f_U(w) - d \ln \left[ 2 \operatorname{sen} \left( \frac{w}{2} \right) \right]^2,$$

ou, escrevendo de outra maneira,

$$\ln f_X(w) = \ln f_U(0) - d \ln \left[ 2 \operatorname{sen} \left( \frac{w}{2} \right) \right]^2 + \ln \left\{ \frac{f_U(w)}{f_U(0)} \right\}. \quad (3.16)$$

Substituindo  $w$  por  $w_j = \frac{2\pi j}{n}$  e adicionando  $\ln I(w_j)$  em ambos os lados da expressão (3.16), onde  $I(\cdot)$  é a função periodograma, dada pela expressão (2.3), temos

$$\ln I(w_j) = \ln f_U(0) - d \ln \left[ 2 \operatorname{sen} \left( \frac{w_j}{2} \right) \right]^2 + \ln \left\{ \frac{f_U(w_j)}{f_U(0)} \right\} + \ln \left\{ \frac{I(w_j)}{f_X(w_j)} \right\}. \quad (3.17)$$

Se o valor máximo para  $j$ , digamos  $g(n)$ , é escolhido tal que  $\frac{g(n)}{n} \rightarrow 0$  quando  $n \rightarrow \infty$  e se  $w_j$  é próximo de zero, digamos,  $w_j \leq w_{g(n)}$  onde  $w_{g(n)}$  é pequeno (neste trabalho consideramos  $g(n) = n^\alpha$ , onde  $\alpha \in (0, 1)$ ), então o termo  $\ln \left\{ \frac{f_U(w_j)}{f_U(0)} \right\}$  é desprezível comparado com os outros termos do membro da direita na igualdade (3.17) (ver Reisen (1994)), e obtemos uma equação aproximada dada por

$$\ln I(w_j) \sim \ln f_U(0) - d \ln \left[ 2 \operatorname{sen} \left( \frac{w_j}{2} \right) \right]^2 + \ln \left\{ \frac{I(w_j)}{f_X(w_j)} \right\}. \quad (3.18)$$

A equação (3.18) acima é da forma de uma equação de regressão simples dada por

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

onde  $y_j = \ln I(w_j)$ ,  $x_j = \ln \left[ 2 \operatorname{sen} \left( \frac{w_j}{2} \right) \right]^2$ ,  $e_j = \ln \left\{ \frac{I(w_j)}{f_X(w_j)} \right\} + c$ ,  $b = -d$ ,  $a = \ln f_U(0) - c$  e  $c = \mathbb{E} \left\{ - \ln \frac{I(w_j)}{f_X(w_j)} \right\}$ .

O estimador de  $d$  pelo método de mínimos quadrados da regressão de  $y_1, y_2, \dots, y_{g(n)}$  em  $x_1, x_2, \dots, x_{g(n)}$ , onde  $g(n)$  é escolhido como acima, é dado por

$$GPH = - \frac{\sum_{j=1}^{g(n)} (x_j - \bar{x}) y_j}{\sum_{j=1}^{g(n)} (x_j - \bar{x})^2}. \quad (3.19)$$

Temos que

$$\mathbb{E}(GPH) = d \text{ e } \operatorname{Var}(GPH) = \frac{\pi^2}{6 \sum_{j=1}^{g(n)} (x_j - \bar{x})^2},$$

com  $\bar{x} = \frac{1}{g(n)} \sum_{j=1}^{g(n)} x_j$ .

Apresentamos, a seguir, o teorema que garante a convergência quase certa do estimador  $GPH$  (ver Geweke e Porter Hudak (1983)).

**TEOREMA 3.7:** *Sejam  $\{X_t\}_{t \in \mathbb{Z}}$  um processo ARFIMA( $p, d, q$ ) com  $d < 0$  e  $I(\cdot)$  a função periodograma do processo sobre as frequências de Fourier*

$w_j = \frac{2\pi j}{n}$ ,  $j \in \{1, 2, \dots, g(n)\}$ . Seja  $GPH$ , dado pela expressão (3.19), o estimador de  $d$  e suponha que  $g(n)$  satisfaça  $g(n) \rightarrow \infty$  e  $\frac{g(n)}{n} \rightarrow 0$ , quando  $n \rightarrow \infty$ . Então,  $GPH \rightarrow d$ .

**Observações:**

- a) O Teorema 3.7 sugere  $g(n) = n^\alpha$ , para  $0 < \alpha < 1$ .
- b) Se  $\lim_{n \rightarrow \infty} \frac{(\ln n)^2}{g(n)} = 0$  então  $\frac{GPH-d}{\sqrt{Var(GPH)}}$  tem distribuição assintótica normal (para mais detalhes sobre a distribuição assintótica de  $GPH$ , ver Geweke e Porter Hudak (1983)).

**Estimador  $SPR$**

Como a função periodograma não é um estimador consistente da função densidade espectral (ver Priestley (1981)), Reisen (1994) propõe uma forma modificada do método de regressão baseada em um estimador consistente para  $f_X(\cdot)$ . Este estimador, baseado na versão suavizada da função periodograma, será denotado por  $SPR$ .

Consideremos a função periodograma suavizado dada pela expressão (2.4). Podemos reescrever (2.4) em termos da função periodograma  $I(\cdot)$ :

$$f_s(w) = \int_{-\pi}^{\pi} W_n(\lambda) I(w - \lambda) d\lambda = \int_{-\pi}^{\pi} W_n(w - \lambda) I(\lambda) d\lambda, \quad (3.20)$$

onde  $W_n(\lambda) = \frac{1}{2\pi} \sum_{k=-(n-1)}^{n-1} \lambda(k) e^{-ik\lambda}$  é a transformada de Fourier da janela  $\lambda(k)$ , chamada *janela espectral*.

A função  $\lambda(\cdot)$  é escolhida de tal forma que as seguintes condições para  $W_n(\lambda)$ , estejam satisfeitas:

- a)  $\int W_n(\lambda) d\lambda = 1$ , para todo  $\lambda \in (-\pi, \pi)$ ;
- b)  $W_n(\lambda) = W_n(-\lambda)$ ,  $W_n(\lambda) \geq 0$ , para todo  $\lambda \in (-\pi, \pi)$ ;
- c)  $\lim_{n \rightarrow \infty} \frac{1}{n} \int W_n^2(\lambda) d\lambda = 0$  para todo  $\lambda \in (-\pi, \pi)$ .

A expressão (3.20) pode ser aproximada pela soma discreta

$$f_s(w) \sim \frac{2\pi}{n} \sum_{j=-\frac{n}{2}}^{\frac{n}{2}} W_n(w - w_j) I(w_j) \text{ para todo } w \in [-\pi, \pi],$$

onde  $w_j = \frac{2\pi j}{n}$ .

A função  $f_s(\cdot)$  tem as seguintes propriedades, quando  $n \rightarrow \infty$  (ver Reisen (1994)):

- $\mathbb{E}(f_s(w)) \sim f_X(w)$ , para todo  $w \in [-\pi, \pi]$ .
- $Var(f_s(w)) \sim (1 + \delta) \frac{2\pi}{n} f_X^2(w) \int W_n^2(\lambda) d\lambda$  ou  
 $Var(f_s(w)) \sim (1 + \delta) f_X^2(w) \frac{1}{n} \sum_{k=-(n-1)}^{n-1} \lambda_n^2(k)$ ,  $w \in [-\pi, \pi]$ ,  
 onde  $\delta = 1$  para  $w = 0, |\pi|$  e  $\delta = 0$  para  $w \neq 0, |\pi|$ .

Pela condição c) acima temos que  $Var(f_s(w)) \rightarrow 0$ , quando  $n \rightarrow \infty$ . Portanto,  $f_s(\cdot)$  é um estimador consistente e não viciado de  $f_X(\cdot)$ .

- $cov(f_s(w_1), f_s(w_2)) = 0$ , para todo  $w_1 \neq w_2$ , i.e., o estimador da função densidade espectral é assintoticamente não correlacionado em diferentes frequências  $w_1$  e  $w_2$ .

Retornemos, agora, à *janela de Parzen* definida pela expressão (2.5). Podemos reescrever esta janela na forma de parâmetro de escala

$$\lambda(k) = l\left(\frac{k}{m}\right),$$

onde  $l(u)$  é dita ser “lag window generator”, que é uma função real contínua no domínio  $u \in (-1, 1)$  com  $l(0) = 1$  e  $l(-u) = l(u)$ . O parâmetro  $m$  (conhecido como *ponto de truncamento*) é função do tamanho amostral  $n$  e é escolhido satisfazendo  $\frac{m}{n} \rightarrow 0$  quando  $n \rightarrow \infty$  e  $m \rightarrow \infty$ . Então escolhe-se  $m = n^\beta$ ,  $0 < \beta < 1$ . Neste trabalho consideramos  $\beta = 0,9$ , pois, segundo Reisen (1994), resultados experimentais sugerem que quando  $m$  aumenta o vício de *SPR* diminui.

A janela de Parzen na forma de parâmetro de escala é dada por

$$\lambda(k) = l\left(\frac{k}{m}\right),$$

onde

$$l(u) = \begin{cases} 1 - 6u^2 + 6|u|^3, & \text{se } |u| \leq \frac{1}{2}, \\ 2(1 - |u|)^3, & \text{se } \frac{1}{2} < |u| \leq 1, \\ 0, & \text{caso contrário.} \end{cases}$$

Podemos então reescrever (2.4) na forma

$$f_s(w) = \frac{1}{2\pi} \sum_{k=-m}^m l\left(\frac{k}{m}\right) \hat{\gamma}_X(k) \cos(kw), \text{ para todo } w \in [-\pi, \pi].$$

Neste caso,

$$\frac{n}{m} \text{Var}(f_s(w)) \sim (1 + \delta) f^2(w) \int l^2(u) \, du,$$

quando  $n \rightarrow \infty$  e

$$\lim_{n \rightarrow \infty} \frac{n}{m} \text{Cov}(f_s(w_1), f_s(w_2)) = 0, \quad w_1 \neq |w_2|.$$

A variância assintótica de  $f_s(w)$  é dada por

$$\text{Var}(f_s(w)) = \begin{cases} 0,539285 \frac{m}{n} f_X^2(w), & \text{se } w \neq 0, \pi, \\ 1,07856 \frac{m}{n} f_X^2(w), & \text{se } w = 0, \pi, \end{cases}$$

onde  $\int_{-1}^1 l^2(u) \, du = 0,539285$ .

Anderson (1971, Capítulo 9) afirma que, sob condições gerais, pode ser mostrado que tanto  $f_s(\cdot)$  quanto  $\ln[f_s(\cdot)]$  têm distribuição normal assintótica.

O lema a seguir é apresentado em Reisen (1994).

**LEMA 3.2:** *Seja  $f_X(\cdot)$  a função densidade espectral do processo ARFIMA( $p, d, q$ ) com  $d \in (-0,5; 0,5)$ , dada pela expressão (3.6). Seja  $f_s(\cdot)$  a função periodograma suavizado dada pela expressão (2.4), onde  $\lambda(\cdot)$  é a janela de Parzen. Então, as variável  $\frac{f_s(\cdot)}{f_X(\cdot)}$  possui distribuição Qui-quadrado com graus de liberdade  $\nu = [3,708617 \frac{m}{n}]$ . Além disso, a variável  $\ln\{\frac{f_s(\cdot)}{f_X(\cdot)}\}$  possui distribuição normal assintótica com média zero e variância dada por*

$$\text{Var} \ln\left\{\frac{f_s(w)}{f_X(w)}\right\} = \begin{cases} 0,539285 \frac{m}{n}, & \text{se } w \neq 0, \pi, \\ 1,07856 \frac{m}{n}, & \text{se } w = 0, \pi. \end{cases}$$

Voltando à equação da expressão (3.17) com  $f_s(w_j)$  no lugar de  $I(w_j)$ , temos

$$\ln f_s(w_j) = \ln f_U(0) - d \ln \left[2 \sin\left(\frac{w_j}{2}\right)\right]^2 + \ln\left\{\frac{f_s(w_j)}{f_X(w_j)}\right\} + \ln\left\{\frac{f_U(w_j)}{f_U(0)}\right\} \quad (3.21)$$

Restringindo o domínio de  $j$  para  $1 \leq j \leq g(n)$ , onde  $g(n)$  é escolhido como anteriormente, podemos reescrever a equação (3.21) para todo  $w_j = \frac{2\pi j}{n}$  na forma

$$\ln f_s(w_j) \sim \ln f_U(0) - d \ln \left[2 \sin\left(\frac{w_j}{2}\right)\right]^2 + \ln\left\{\frac{f_s(w_j)}{f_X(w)}\right\}. \quad (3.22)$$

A equação (3.22) também é uma forma aproximada da equação de regressão linear simples, i.e.,

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

onde  $y_j = \ln f_s(w_j)$ ,  $x_j = \ln \left[ 2 \operatorname{sen} \left( \frac{w_j}{2} \right) \right]^2$ ,  $e_j = \ln \left\{ \frac{f_s(w_j)}{f_X(w_j)} \right\} + c$ ,  $b = -d$ ,  $a = \ln f_U(0) - c$  e  $c = \mathbb{E} \left\{ -\ln \frac{f_s(w_j)}{f_X(w_j)} \right\}$ .

O estimador de  $d$ , obtido pelo método de regressão utilizando a função periodograma suavizado, é dado por

$$SPR = -\frac{\sum_{j=1}^{g(n)} (x_j - \bar{x}) y_j}{\sum_{j=1}^{g(n)} (x_j - \bar{x})^2}. \quad (3.23)$$

Temos que

$$\mathbb{E}(SPR) = d \text{ e } \operatorname{Var}(SPR) = 0,539285 \frac{m}{n \sum_{j=1}^{g(n)} (x_j - \bar{x})^2},$$

com  $\bar{x} = \frac{1}{g(n)} \sum_{j=1}^{g(n)} x_j$ . Observe que o valor  $0,539285 \frac{m}{n}$  na expressão da variância do estimador  $SPR$  é a variância assintótica de  $e_j = \ln \left\{ \frac{f_s(w_j)}{f_X(w_j)} \right\}$  (ver Reisen (1994)).

O estimador  $SPR$  é assintoticamente normal com  $\mathbb{E}(SPR) = d$  e variância dada pela expressão acima. Um estudo mais detalhado do estimador  $SPR$  pode ser encontrado em Reisen (1994).

### Estimador $GPHT$

O próximo estimador é uma forma modificada do estimador  $GPH$ , denotado aqui por  $GPHT$ . Robinson (1995), fazendo suaves modificações no estimador  $GPH$ , provou propriedades estatísticas para este novo estimador, que valem tanto para  $d$  negativo como para  $d$  positivo, o que não acontece com a forma originalmente proposta para o estimador  $GPH$ . Este estimador regressa  $\ln \{I(w_j)\}$  em  $\ln \{2 \operatorname{sen}(w_j/2)\}^2$ , para  $j = l, l+1, \dots, g(n)$ , onde  $l$  é o regressor inicial. Na teoria assintótica, ambos  $l$  e  $g(n)$  tendem para infinito com  $n$ , mas mais lentamente, enquanto  $\frac{l}{g(n)}$  vai a zero.

O teorema a seguir é uma versão simplificada de resultado dado em Robinson (1995) e garante a distribuição assintótica de  $GPHT$  (ver Beran (1994)).

**TEOREMA 3.8:** *Seja  $\{X_t\}_{t \in \mathbb{Z}}$  um processo estacionário ARFIMA com longa dependência, isto é, um processo definido de acordo com a Definição 2.15. Seja  $GPHT$  o estimador de mínimos quadrados de  $d$ , baseado nas*

freqüências de Fourier  $w_j$ , para todo  $l \leq j \leq g(n)$ , onde  $l$  e  $g(n)$  são tais que, quando  $n \rightarrow \infty$ ,

$$g(n) \rightarrow \infty, \quad l \rightarrow \infty,$$

$$\frac{g(n)^5}{n^4} \rightarrow 0, \quad \frac{(\ln n)^2}{g(n)} \rightarrow 0,$$

e

$$\frac{l}{g(n)} \rightarrow 0, \quad \frac{\sqrt{g(n)} \ln n}{l} \rightarrow 0.$$

Então, considerando algumas condições de regularidade para a função densidade espectral  $f_X(\cdot)$ ,

$$\sqrt{g(n)}(GPHT - d) \rightarrow \sigma_0^2 Z,$$

onde  $Z$  é variável aleatória com distribuição normal padrão e  $\sigma_0^2 = \frac{\pi^2}{24}$ .

### Estimador *MGPH*

Outra forma modificada do método *GPH*, denotada aqui por *MGPH*, é obtida por Velasco (1999a) trocando, na equação de regressão (3.17),  $2 \operatorname{sen}(\frac{w_j}{2})$  por  $j$  (lembre que  $\operatorname{sen}(w) \sim w$  para  $w \sim 0$ ). Temos, então

$$MGPH = -\frac{1}{2} \frac{\sum_{j=1}^{g(n)} \ln I(w_j) \{ \ln j - \frac{1}{g(n)} \sum_{l=1}^{g(n)} \ln l \}}{\sum_{j=1}^{g(n)} \ln j \{ \ln j - \frac{1}{g(n)} \sum_{l=1}^{g(n)} \ln l \}},$$

onde  $I(\cdot)$  é o periodograma dado pela expressão (2.3) e  $w_j$  as freqüências de Fourier para  $j \in \{1, \dots, g(n)\}$ .

### Estimador *GPHT*

Este estimador também é baseado no método de regressão do periodograma. Utiliza-se a função cosine-bell como janela espectral para reduzir o vício do periodograma, dada pela expressão

$$\lambda(t) = \frac{1}{2} \left[ 1 - \cos \left( \frac{2\pi(t + 0,5)}{n} \right) \right].$$

O estimador é obtido de maneira análoga ao estimador *GPH*, isto é, regressa-se  $\ln\{I_n(w_j)\}$  em  $\ln\{2 \operatorname{sen}(\frac{w_j}{2})\}$ , para  $j = 2, \dots, g(n)$ .

## Estimador $FT$

O estudo do estimador de máxima verossimilhança para observações dependentes é vasto e reúne diversos resultados na literatura. Para modelos  $ARFIMA(p, d, q)$  o trabalho de Sowell (1991) apresenta resultados para o *estimador de máxima verossimilhança exato* e Dahlhaus (1989) apresenta resultados para o *estimador de máxima verossimilhança aproximado*, utilizando a aproximação sugerida por Whittle (1951). Fox e Taqqu (1986) apresentam condições que permitem que o estimador de máxima verossimilhança aproximado para seqüências com forte dependência seja consistente e tenha distribuição assintótica normal. Tais condições são satisfeitas pelos processos  $ARFIMA(p, d, q)$ , segundo Dahlhaus (1989). Consideraremos aqui o estimador sugerido por Fox e Taqqu (1986), que será denotado por  $FT$  e é o estimador de máxima verossimilhança aproximado.

Seja  $\{X_t\}_{t \in \mathbb{Z}}$  um processo estacionário com média  $\mu$  e variância  $\sigma_X^2$ . Seja  $f_X(\cdot)$  a função densidade espectral caracterizada por um vetor de parâmetros finito-dimensional desconhecido

$$v = (\sigma_X^2, d, \phi_1, \phi_2, \dots, \phi_p, \theta_1, \theta_2, \dots, \theta_q).$$

Assumimos que a função densidade espectral de uma família paramétrica de densidades é dada por  $f_X(\cdot) = f_X(\cdot; v)$ , onde  $v \in \Upsilon \subset \mathbb{R}^a$ ,  $a = p + q + 2$  é a dimensão do vetor  $v$  e  $\Upsilon$  é o espaço de parâmetros. Estimativas para o vetor  $v$  são obtidas através de uma série temporal  $X_0, X_1, \dots, X_{n-1}$ .

Suponha que  $\{X_t\}_{t \in \mathbb{Z}}$  é um processo Gaussiano linear causal e invertível. Então, a função de distribuição conjunta de  $X = (X_0, X_1, \dots, X_{n-1})$  é dada por

$$h(x; v) = (2\pi)^{-\frac{n}{2}} |\Sigma_n(v)|^{-\frac{1}{2}} e^{-\frac{1}{2} x' [\Sigma_n(v)]^{-1} x}, \quad (3.24)$$

com  $x' = (x_0, x_1, \dots, x_{n-1})' \in \mathbb{R}^n$ , onde  $x'$  denota o vetor transposto do vetor  $x$  e  $\Sigma_n(v)$  é a matriz quadrada  $n \times n$ , cujos elementos são a função de autocovariância, dada por

$$\Sigma_n(v) = [\gamma_X(j-l)]_{j,l=0}^{n-1}.$$

O logaritmo da função de verossimilhança  $h(x; v)$ , dada pela expressão (3.24), é dado por

$$\begin{aligned} \mathcal{L}(x; v) &= \ln h(x; v) \\ &= -\frac{n}{2} \ln(2\pi) - \frac{1}{2} \ln |\Sigma_n(v)| - \frac{1}{2} x' [\Sigma_n(v)]^{-1} x. \end{aligned} \quad (3.25)$$

O estimador de máxima verossimilhança de  $v$  é obtido maximizando a função  $\mathcal{L}(x; v)$  com respeito ao vetor de parâmetros  $v$ . Fox e Taqu (1986) aplicam o método da máxima verossimilhança, com a aproximação sugerida por Whittle (1951), para estimar  $v$  e  $\sigma_X^2$  maximizando a função

$$h(x; v; \sigma_X^2) \simeq \left( \frac{1}{\sqrt{2\pi\sigma_X}} \right)^n \exp \left\{ -\frac{Z' A_n(v) Z}{2n\sigma_X} \right\},$$

onde  $Z'$  é o vetor transposto do vetor  $Z = (X_0 - \bar{X}, \dots, X_{n-1} - \bar{X})$  e  $\bar{X}$  é a média amostral. Maximizar a função  $h(x; v; \sigma_X^2)$  é equivalente a escolher  $\hat{v}$  que minimiza

$$\sigma_{X,n}^2 = \frac{Z' A_n(v) Z}{n}.$$

Uma prova para a consistência e distribuição assintótica normal do estimador  $\hat{v}$ , onde  $\sigma_X^2$  é a primeira componente do vetor  $v$ , pode ser encontrada em Dahlhaus (1989). Dahlhaus (1989) mostra, também, que este estimador é assintoticamente eficiente no sentido de Fisher.

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

$$\mathcal{L}_n(v) = \frac{1}{2n} \sum_{j=1}^{n-1} \left( \ln f_X(w_j; v) + \frac{I(w_j)}{f_X(w_j; v)} \right),$$

onde  $v$  denota o vetor de parâmetros desconhecidos.

## Estimador $V$

Podemos, a partir do resultado da ordem da variância de  $S_n$  (ver Proposição 3.3), obter um estimador de  $d$  dado por

$$V = \frac{\ln \widehat{Var}(S_n)}{2 \ln n} - \frac{1}{2}.$$

A proposição a seguir apresenta mais detalhadamente este estimador.

**PROPOSIÇÃO 3.5:** *Seja  $\{X_t\}_{t \in \mathbb{Z}}$  um processo ARFIMA( $p, d, q$ ), com  $d \in (0, 0; 0, 5)$ , dado pela expressão (3.3). Considere  $\{X_t\}_{t=0}^{n-1}$  uma série temporal obtida do processo  $\{X_t\}_{t \in \mathbb{Z}}$ . Então, o estimador para  $d$ , baseado no resultado da Proposição 3.3 é dado por*

$$V = \frac{\ln \{2n[\frac{\hat{\gamma}_X(0)}{2} + \frac{1}{n} \sum_{j=1}^{n-1} (n-j) \hat{\gamma}_X(j)]\}}{2 \ln n} - \frac{1}{2}.$$

*Este estimador possui cota inferior para o seu valor esperado dada por*

$$\mathbb{E}(V) \geq \frac{1}{2 \ln n} \{n(\gamma_X(0) - Var \bar{X}) + \frac{2}{n} \sum_{j=1}^{n-1} (n-j)^2 (\gamma_X(j) - Var \bar{X})\} - \frac{1}{2},$$

onde  $\gamma_X(\cdot)$  é a função de autocovariância e  $\hat{\gamma}_X(\cdot)$  é a autocovariância amostral do processo  $\{X_t\}_{t \in \mathbb{Z}}$ .

**Prova:** Pela expressão (3.15) sabemos que

$$\text{Var}(S_n) = 2n \left[ \frac{\gamma_X(0)}{2} + \frac{1}{n} \sum_{j=1}^{n-1} (n-j) \gamma_X(j) \right] \approx n^{2d+1},$$

onde  $S_n = \sum_{i=0}^{n-1} X_i$  é a soma parcial de  $\{X_t\}_{t=0}^{n-1}$ .

Portanto,  $\text{Var}(S_n) \approx n^{2d+1}$ . Aplicando logaritmo a esta expressão, temos que

$$\begin{aligned} \text{Var}(S_n) &\approx n^{2d+1} \Leftrightarrow \ln \text{Var}(S_n) \approx \ln(n^{2d+1}) \Leftrightarrow \\ \ln \text{Var}(S_n) &\approx (2d+1) \ln n \Leftrightarrow \frac{\ln \text{Var}(S_n)}{\ln n} \approx 2d+1, \end{aligned}$$

ou seja,

$$V \equiv \hat{d} = \frac{\ln \widehat{\text{Var}}(S_n)}{2 \ln n} - \frac{1}{2}.$$

A partir da expressão acima obtemos o estimador de  $d$  dado por

$$\begin{aligned} V &= \frac{\ln \widehat{\text{Var}}(S_n)}{2 \ln n} - \frac{1}{2} = \frac{\ln \left\{ 2n \left[ \frac{\hat{\gamma}_X(0)}{2} + \frac{1}{n} \sum_{j=1}^{n-1} (n-j) \hat{\gamma}_X(j) \right] \right\}}{2 \ln n} - \frac{1}{2} \\ &= \frac{1}{2} \left\{ \frac{\ln [n \hat{\gamma}_X(0) + 2 \sum_{j=1}^{n-1} (n-j) \hat{\gamma}_X(j)]}{\ln n} \right\} - \frac{1}{2}, \end{aligned} \quad (3.26)$$

onde  $\hat{\gamma}_X(\cdot)$  é o estimador da função de autocovariância do processo dado pela Definição 2.9. Este estimador é viciado (ver Wei (1990)), pois

$$\mathbb{E}[\hat{\gamma}_X(j)] = \gamma_X(j) - \frac{j}{n} \gamma_X(j) - \frac{(n-j)}{n} \text{Var} \bar{X} = \frac{(n-j)}{n} [\gamma_X(j) - \text{Var} \bar{X}].$$

O estimador  $V$  é também viciado e queremos obter a cota inferior para seu valor esperado. Observe que

$$\begin{aligned} \mathbb{E}(V) &= \mathbb{E} \left( \frac{1}{2} \left\{ \frac{\ln [n \hat{\gamma}_X(0) + 2 \sum_{j=1}^{n-1} (n-j) \hat{\gamma}_X(j)]}{\ln n} \right\} - \frac{1}{2} \right) \\ &= \frac{1}{2 \ln n} \mathbb{E} \left( \left\{ \ln [n \hat{\gamma}_X(0) + 2 \sum_{j=1}^{n-1} (n-j) \hat{\gamma}_X(j)] \right\} \right) - \frac{1}{2}. \end{aligned} \quad (3.27)$$

Como a função logaritmo é côncava, aplicando a Desigualdade de Jensen na expressão (3.27) segue que

$$\begin{aligned}
\mathbb{E}(V) &\geq \frac{1}{2 \ln n} \ln \left\{ \mathbb{E} \left[ n \hat{\gamma}_X(0) + 2 \sum_{j=1}^{n-1} (n-j) \hat{\gamma}_X(j) \right] \right\} - \frac{1}{2} = \\
&= \frac{1}{2 \ln n} \ln \left\{ n \mathbb{E}(\hat{\gamma}_X(0)) + 2 \sum_{j=1}^{n-1} (n-j) \mathbb{E}(\hat{\gamma}_X(j)) \right\} - \frac{1}{2} \\
&= \frac{1}{2 \ln n} \ln \left\{ n(\gamma_X(0) - \text{Var}(\bar{X})) + 2 \sum_{j=1}^{n-1} (n-j) \frac{(n-j)}{n} (\gamma_X(j) - \text{Var}(\bar{X})) \right\} - \frac{1}{2} \\
&= \frac{1}{2 \ln n} \ln \left\{ n(\gamma_X(0) - \text{Var}(\bar{X})) + \frac{2}{n} \sum_{j=1}^{n-1} (n-j)^2 (\gamma_X(j) - \text{Var}(\bar{X})) \right\} - \frac{1}{2}.
\end{aligned}$$

□

**Observação:** O mesmo estimador será obtido no Capítulo 4 quando utilizarmos a relação entre os parâmetros  $d$  e  $s$ , onde  $d$  é o parâmetro fracionário do modelo *ARFIMA* e  $s$  é o parâmetro da transformação Manneville-Pomeau.

### Estimador *VP*

Este estimador é sugerido em Beran (1994) (chamado de “variance plot”) e é obtido a partir da ordem da variância de  $\bar{X}_n = \frac{S_n}{n}$ , ou seja,  $n^{2d-1}$ . O método consiste em obter o estimador através do gráfico de  $\ln s^2(k)$  versus  $\ln(k)$ , onde  $k$  é um inteiro,  $2 \leq k \leq [\frac{n}{2}]$  e  $s^2(k)$  é a variância das médias amostrais  $\overline{X_j(k)}$ , com  $1 \leq j \leq m_k$ , para  $m_k$  suficientemente grande. Ou seja, calculam-se as médias amostrais  $\overline{X_1(k)}, \overline{X_2(k)}, \dots, \overline{X_{m_k}(k)}$  e a média amostral geral

$$\overline{X(k)} = \frac{1}{m_k} \sum_{j=1}^{m_k} \overline{X_j(k)},$$

para diferentes valores de  $k$  entre 2 e  $[\frac{n}{2}]$ . Para cada  $k$ , calcula-se a variância amostral das médias  $\overline{X_j(k)}$ ,  $1 \leq j \leq m_k$ , da seguinte maneira:

$$s^2(k) = \frac{1}{m_k - 1} \sum_{j=1}^{m_k} (\overline{X_j(k)} - \overline{X(k)})^2.$$

A inclinação da reta ajustada aos valores de  $\ln s^2(k)$  versus  $\ln(k)$  é aproximadamente  $2d - 1$  e então obtemos um estimador para  $d$ . No entanto,

este método não fornece bons resultados, sendo indicado apenas como uma primeira ferramenta para verificação sobre a propriedade de longa dependência na série temporal, assim como vários outros métodos heurísticos descritos em Beran (1994).

## Estimador $W$

Apresentamos, a seguir, um estimador para o parâmetro de longa dependência obtido a partir da *transformada de wavelets (ondaletas)*. Primeiramente vamos dar uma introdução às wavelets (um estudo detalhado pode ser encontrado em Meyer (1993), Morettin (1997, 1999), Percival e Walden (2000) e Zandonade (1999)).

A teoria de wavelets, pela sua forma, é muito útil por localizar o processo não só na frequência, como também no tempo, sendo, segundo Meyer (1993), uma alternativa a outros sistemas de funções usados como bases para representação de funções pertencentes a certos espaços, como senos e cossenos na transformada de Fourier.

Atualmente as wavelets são assunto de estudo em várias áreas como processamento de sinais, codificação de imagens e também como ferramenta para a análise de séries temporais. A teoria de wavelets vem revelando-se uma ótima alternativa à transformada de Fourier na análise de sinais não estacionários.

Na análise de Fourier, toda função periódica, de período  $2\pi$ , de quadrado integrável, ou seja, função no espaço  $\mathcal{L}^2(0, 2\pi)$ , é gerada por uma superposição de exponenciais complexas,  $w_n(t) = e^{int}$ ,  $n \in \mathbb{Z}$ , obtidas por dilatações da função  $w(t) = e^{it}$  :  $w_n(t) = w(nt)$ . O objetivo é estender essa idéia para  $\mathcal{L}^2(\mathbb{R})$ , isto é, gerar esse espaço a partir de uma única função.

Uma *wavelet* é definida como qualquer função  $\psi(t)$  contínua que decaia rapidamente a zero quando  $|t| \rightarrow \infty$  e oscile de tal forma que  $\int_{-\infty}^{\infty} \psi(t) dt = 0$ . A idéia é considerar dilatações e translações dessa função (denominada wavelet-mãe), de modo a gerar  $\mathcal{L}^2(\mathbb{R})$ . Consideramos, para tanto,

$$\psi_{a,b}(t) = |a|^{-1/2} \psi\left(\frac{t-b}{a}\right), \quad a \neq 0,$$

sendo  $a$  e  $b$  parâmetros reais que definem, respectivamente, as dilatações e translações da wavelet-mãe, e  $|a|^{-1/2}$  um termo normalizador. Uma particularização é obtida quando os parâmetros  $a$  e  $b$  assumem valores  $a = 2^{-j}$  e  $b = k 2^{-j}$ , onde  $j, k \in \mathbb{Z}$ . Desta forma, as wavelets consideradas são dadas por

$$\psi_{j,k}(t) = 2^{j/2} \psi(2^j t - k), \quad j, k \in \mathbb{Z},$$

que constitui uma base ortonormal de  $\mathcal{L}^2(\mathbb{R})$ .

Portanto, as funções  $\{\psi_{j,k}(t), j, k \in \mathbb{Z}\}$  são obtidas a partir da função  $\psi(t)$  através de uma translação diádica  $k2^{-j}$  e uma dilatação binária  $2^j$ , onde o fator escala  $2^j$  é chamado de *fator de dilatação*. Esta base ortonormal permite uma reconstrução perfeita de um sinal a partir dos coeficientes da transformada (ver Definição 3.6 a seguir), isto é, cada coeficiente é obtido como o produto escalar do sinal e da função base (aqui as funções  $\psi_{j,k}(\cdot)$  e, no caso da análise de Fourier, as funções  $w_n(\cdot)$ ).

Existem muitos tipos de wavelets das mais diversas formas. Wavelets podem ser suaves ou não, de suporte compacto ou não, com expressões matemáticas simples ou não. Na literatura da área, são encontradas diversas bases de wavelets tais como, por exemplo, as bases Haar, Chapéu Mexicano, Shannon e Morlet. Além dessas, outras que se destacam são as *Daubechies*, *Symmlets* e *Coiflets*, todas de suporte compacto e introduzidas por I. Daubechies.

Apresentamos abaixo duas bases de wavelets que têm expressão analítica:

**Haar:** A mais simples e antiga wavelet, uma função escada:

$$\psi(t) = \begin{cases} 1, & 0 \leq t < 1/2, \\ -1, & 1/2 \leq t < 1, \\ 0, & \text{caso contrário,} \end{cases}$$

e

$$\psi_{j,k}(t) = \begin{cases} 2^{\frac{j}{2}}, & 2^{-j}k \leq t < 2^{-j}(k + 1/2) \\ -2^{-j}, & 2^{-j}(k + 1/2) \leq t < 2^{-j}(k + 1) \\ 0, & \text{caso contrário,} \end{cases} \quad (3.28)$$

$j = 0, 1, \dots, m - 1$  e  $k = 0, 1, \dots, 2^j - 1$ .

**Chapéu Mexicano:** Obtida a partir da segunda derivada da curva Gaussiana:

$$\psi(t) = (1 - t^2) \exp(-t^2/2), \quad t \in \mathbb{R},$$

e

$$\psi_{j,k}(t) = 2^{j/2} [1 - (2^j t - k)^2] \exp[-((2^j t - k)^2/2)]. \quad (3.29)$$

A partir de agora vamos considerar  $t \in \mathbb{Z}$  ou  $t \in \mathbb{N}$ .

**DEFINIÇÃO 3.6:** Considere uma série temporal obtida a partir do processo estocástico  $\{X_t\}_{t \in \mathbb{Z}}$ . Os *coeficientes de wavelets (ou transformada de wavelets)* para  $X_t$  são uma função do parâmetro de escala,  $j$ , e do parâmetro

de translação,  $k$ , e são dados pelo produto interno

$$\omega_{j,k} = \langle X_t, \psi_{j,k}(t) \rangle = \sum_{t=-\infty}^{\infty} X_t \omega_{j,k}(t) = 2^{j/2} \sum_{t=-\infty}^{\infty} X_t \psi(2^j t - k).$$

**Observação:** Na prática, temos uma série temporal  $\{X_t\}_{t=0}^{n-1}$  com  $n$  observações. Assim, a transformada finita de wavelets é dada por

$$\omega_{j,k} = 2^{j/2} \sum_{t=0}^{n-1} X_t \psi(2^j t - k), \quad (3.30)$$

com  $0 \leq j \leq m-1$ , sendo  $m$  inteiro tal que  $n = 2^m$  e  $0 \leq k \leq 2^j - 1$ . É intuitivo pensarmos que a variação de  $k$  dependa de  $j$ , sabendo que, quanto maior o valor de  $j$ , mais compacta é a wavelet e mais curto é o intervalo de translação. Assim, serão necessárias mais translações para percorrer toda a série temporal em questão.

O teorema, a seguir, descreve a propriedade assintótica da transformada finita de wavelets, dada pela expressão (3.30). Para maiores detalhes ver Jensen (1999).

**TEOREMA 3.10:** *Seja  $\{X_t\}_{t \in \mathbb{Z}}$  um processo ARFIMA(0,  $d$ , 0) de média zero, com  $|d| < 0,5$ . A transformada finita de wavelets,  $\omega_{j,k}$ , definida pela expressão (3.30), tem distribuição normal com média zero e variância aproximada  $\sigma^2 |2^{-j}|^{2d}$ , onde  $\sigma^2$  é uma constante finita.*

Considere  $\{X_t\}_{t \in \mathbb{Z}}$  um processo ARFIMA(0,  $d$ , 0) de média zero com  $|d| < 0,5$ . Definimos a variância dos coeficientes de wavelets como

$$R(j) = \mathbb{E}[(\omega_{j,k})^2], \text{ para todo } j = 0, \dots, m-1. \quad (3.31)$$

Assim, pelo Teorema 3.10,  $R(j) = \sigma^2 |2^{-j}|^{2d}$ . Aplicando o logaritmo à expressão (3.31), obtemos

$$\ln R(j) = \ln \sigma^2 + d \ln |2^{-j}|^2. \quad (3.32)$$

Por outro lado, podemos utilizar a variância amostral de  $\omega_{j,k}$  para estimar  $R(j)$ , obtendo o estimador

$$\hat{R}(j) = \frac{1}{2^j} \sum_{k=0}^{2^j-1} (\omega_{j,k})^2, \text{ para todo } j = 0, 1, \dots, m-1,$$

onde  $m$  é tal que  $n = 2^m$ . Adicionando  $\ln \hat{R}(j)$  em ambos os lados da expressão (3.32) e fazendo  $e_j = \ln \hat{R}(j) - \ln R(j)$ , segue que

$$\ln \hat{R}(j) = \ln \sigma^2 + d \ln |2^{-j}|^2 + e_j, \text{ para todo } j = 0, 1, \dots, m-1. \quad (3.33)$$

A partir da expressão (3.33), Jensen (1999) sugere estimar o parâmetro de diferenciação fracionária  $d$  através do método de mínimos quadrados ordinários. Pode-se demonstrar que  $E(\hat{R}(j)) = R(j)$  e  $Var(\hat{R}(j)) \rightarrow 0$ , quando  $j \rightarrow \infty$ . Pela lei dos grandes números de Markov,  $\hat{R}(j)$  converge em probabilidade para  $R(j)$ , quando  $j \rightarrow \infty$ . Portanto,

$$\ln \hat{R}(j) = \ln R(j) + o_p(1), \text{ para todo } j = 0, 1, \dots, m-1.$$

Substituindo  $R(j)$  por  $\sigma^2|2^{-j}|^{2d}$ , temos

$$\ln \hat{R}(j) = \ln \sigma^2 + d \ln |2^{-j}|^2 + o_p(1), \text{ para todo } j = 0, 1, \dots, m-1.$$

O estimador de  $d$  baseado em wavelets é dado por

$$W = \left[ \sum_{j=0}^{m-1} x_j^2 \right]^{-1} \left[ \sum_{j=0}^{m-1} x_j \ln \hat{R}(j) \right], \quad (3.34)$$

onde  $x_j$  é definido por

$$x_j = \ln 2^{-2j} - \frac{1}{m} \sum_{j=0}^{m-1} \ln 2^{-2j}.$$

Ou seja, quando  $j \rightarrow \infty$ , o estimador de mínimos quadrados ordinários,  $W$ , obtido através da relação acima, oferece uma estimativa para o parâmetro de diferenciação fracionária  $d$  do processo  $ARFIMA(0, d, 0)$  sendo obtido através de uma equação de regressão linear simples.

### Observações:

a) Na prática, não se utiliza  $\omega_{j,k}$  para  $j$  muito pequeno, pois as respectivas wavelets não convergem a 0 no intervalo da série. Sendo assim, a série não ocupa toda a onda e, portanto, há uma influência negativa desses na estimativa. Poli e Lopes (2002) consideraram  $l = 4$  como sendo o ponto de truncamento inferior.

b) Através de simulações, análise de dados reais e comparações com outros estimadores, Poli e Lopes (2002) observaram que o estimador  $W$  utilizando as bases Haar, Chapéu Mexicano, Shannon e Morlet se comportou muito bem, com capacidade de analisar séries de dados estacionários ( $d < 0,5$ ) e não estacionários ( $0,5 < d^* < 1,5$ ). O uso da transformada de wavelets comprovou ser muito eficiente e robusto, tendo em vista os baixos vício e erro quadrático médio apresentados nas simulações (ver Poli e Lopes (2002)).

c) Ressaltamos o fato de que o tamanho amostral da série temporal deve ser uma potência de dois para que possamos utilizar o estimador de wavelets.

No Capítulo 4, adaptaremos alguns dos estimadores apresentados nesta seção para a classe de processos Manneville-Pomeau, que apresentam características comuns aos processos *ARFIMA*.

## 3.4 Processos não Estacionários

### 3.4.1 Descrição e Simulação

Seja  $\{\tilde{X}_t\}_{t \in \mathbb{Z}}$  um processo *ARFIMA*(0,  $d^*$ , 0) onde  $d^* = d + r$ , com  $d \in (0, 0; 0, 5)$  e  $r > 0$  tal que  $d^* \in (0, 5; 1, 5)$ . Da expressão (3.4) temos que

$$(1 - \mathcal{B})^{d^*} \tilde{X}_t = \epsilon_t, \text{ para todo } t \in \mathbb{Z}. \quad (3.35)$$

Observe que

$$(1 - \mathcal{B})^{d^*} \tilde{X}_t = \epsilon_t \Leftrightarrow (1 - \mathcal{B})^d (1 - \mathcal{B})^r \tilde{X}_t = \epsilon_t \Leftrightarrow (1 - \mathcal{B})^d Y_t = \epsilon_t.$$

Portanto, o processo  $\{\tilde{X}_t\}_{t \in \mathbb{Z}}$  é um *ARFIMA*(0,  $d^*$ , 0) enquanto que o processo  $\{Y_t\}_{t \in \mathbb{Z}}$ , definido por

$$Y_t = (1 - \mathcal{B})^r \tilde{X}_t, \text{ para todo } t \in \mathbb{Z},$$

é um *ARFIMA*(0,  $d$ , 0).

Observe que o processo  $\{\tilde{X}_t\}_{t \in \mathbb{Z}}$ , dado pela expressão (3.35), não possui uma função densidade espectral, visto ser não estacionário. No entanto, a função  $f_{\tilde{X}}(\cdot)$ , dada pela expressão (3.5), mas agora para o processo  $\{\tilde{X}_t\}_{t \in \mathbb{Z}}$ , cumpre as características desempenhadas pela função densidade espectral (ver Hurvich e Ray (1995) e Velasco (1999b)).

Quando  $d^* \in (0, 5; 1, 0)$  o processo  $\{\tilde{X}_t\}_{t \in \mathbb{Z}}$  é não estacionário mas possui a propriedade de reversão do nível, o que influencia positivamente nos resultados obtidos a partir de métodos de estimação descritos para o caso estacionário. Quando  $d^* \in (1, 0; 1, 5)$  o processo  $\{\tilde{X}_t\}_{t \in \mathbb{Z}}$  é não estacionário e não vale a propriedade de reversão do nível. Veremos nas Seções 3.4.2 e 3.5 deste Capítulo que a falta desta propriedade influencia, de forma bastante negativa, as estimativas do parâmetro  $d$  obtidas através dos diversos métodos descritos na Seção 3.3. A única exceção ocorre para o método descrito através de wavelets (ver Poli e Lopes (2002)).

A série temporal  $\{Y_t\}_{t=1}^n$  é simulada através do algoritmo sugerido no artigo de Hosking (1981) onde o processo  $\{\epsilon_t\}_{t \in \mathbb{Z}}$  é um ruído branco Gaussiano com média zero e  $\sigma_\epsilon^2 = 1.0$  e este processo é gerado através da subrotina

RNNOR do IMSL. O processo  $\{\tilde{X}_t\}_{t \in \mathbb{Z}}$  é obtido através da transformação algébrica

$$\tilde{X}_t = (1 - \mathcal{B})^{-r} Y_t, \text{ para } t \in \mathbb{N} - \{0\},$$

com  $\tilde{X}_1 = Y_1$ .

No caso de processos  $ARFIMA(p, d^*, q)$  simulamos primeiramente processos  $ARFIMA(0, d^*, 0)$ , conforme descrito acima, e depois incluímos as componentes autoregressivas (*AR*) e médias móveis (*MA*) a partir de subrotinas escritas na linguagem FORTRAN.

Aplicaremos alguns dos métodos de estimação do parâmetro  $d$ , descritos na Seção 3.3 para processos não estacionários, ou seja, quando  $d^* > 0,5$ . Para o caso estacionário alguns destes estimadores já foram bastante explorados (veja em cada estimador as referências citadas) e, por isso, na próxima seção, vamos nos deter, principalmente, ao caso não estacionário.

### 3.4.2 Estimação dos Parâmetros

Apresentamos, a seguir, alguns resultados obtidos a partir de séries temporais com tamanhos amostrais  $n = 100$  e  $300$  baseados em 800 replicações do processo, utilizando os métodos de estimação *GPH*, *SPR*, *GPht*, *MGPH* e *FT* descritos na Seção 3.3. Foram calculados os valores empíricos da média, do vício, do desvio padrão (*sd*) e do erro quadrático médio (*eqm*). Estes valores estão nas tabelas a seguir. O vício e o erro quadrático médio de menor valor absoluto estão em negrito. O valor de  $g(n) = n^\alpha$  nos métodos semi-paramétricos *GPH* e *SPR* está fixado com  $\alpha = 0,5$  (valor amplamente utilizado na literatura). O ponto de truncamento na janela de Parzen para o estimador *SPR* é  $m = n^\beta$ , com  $\beta = 0,9$  (ver Reisen (1994) e Olbermann (1998) para uma discussão sobre o valor de  $\beta$ ). Para os estimadores *GPht* e *MGPH* foram utilizados dois diferentes valores para o limite  $g(n) = n^{\alpha_i}$ : *GPht*(1) e *MGPH*(1) referem-se a  $\alpha_1 = 0,6$  e *GPht*(2) e *MGPH*(2) referem-se a  $\alpha_2 = 0,7$  (os valores de  $g(n)$  correspondente aparecem nas tabelas). O regressor inicial  $l$  nos métodos de regressão é  $l = 1$  exceto para *GPht*, onde consideramos  $l = 2$ .

#### Caso $ARFIMA(0, d^*, 0)$

A Tabela 3.4.1 apresenta o caso  $d^* = 0,6$ . O método *FT* é, em geral, o mais apropriado (menores valores para o vício e erro quadrático médio) do que os outros estimadores que, no entanto, também apresentam bons resultados. A partir da média de cada estimador podemos observar que a tendência é superestimar o verdadeiro valor do parâmetro, exceto no método

*SPR*.

Note que, aumentando o limite  $g(n)$ , em ambos os métodos *GPht* e *MGPH*, o vício e o erro quadrático médio decrescem consideravelmente (observe que o erro quadrático médio decresce, geralmente, à metade). O estimador *SPR* tem um melhor desempenho com relação ao estimador *GPH*, no sentido de menores valores para o erro quadrático médio. Isto já era esperado pois o estimador *SPR* utiliza a função periodograma suavizado para estimar a função densidade espectral.

Quando  $d^* = 1, 0$ , isto é, quando o processo é um passeio aleatório, todos os métodos funcionam bem, são bastante competitivos e tendem a subestimar o parâmetro (ver Tabela 3.5.2).

**Tabela 3.4.1:** Estimativas para o modelo *ARFIMA*(0; 0, 6; 0).

$n$	Método	$g(n)$	média( $\hat{d}^*$ )	vício( $\hat{d}^*$ )	$sd(\hat{d}^*)$	$eqm(\hat{d}^*)$
100	<i>GPH</i>	10	0,6302	0,0302	0,2944	0,0874
	<i>SPR</i>	10	0,5175	-0,0825	0,2345	0,0617
	<i>GPht</i> (1)	16	0,6135	0,0135	0,3287	0,1080
	<i>GPht</i> (2)	25	0,6151	0,0151	0,2272	0,0517
	<i>MGPH</i> (1)	16	0,6103	0,0103	0,2080	0,0433
	<i>MGPH</i> (2)	25	0,6006	<b>0,0006</b>	0,1557	0,0242
	<i>FT</i>	-	0,6039	0,0039	0,0926	<b>0,0086</b>
300	<i>GPH</i>	17	0,6194	0,0194	0,1976	0,0393
	<i>SPR</i>	17	0,5630	-0,0370	0,1674	0,0293
	<i>GPht</i> (1)	31	0,6041	<b>0,0041</b>	0,1897	0,0359
	<i>GPht</i> (2)	54	0,6110	0,0110	0,1279	0,0164
	<i>MGPH</i> (1)	31	0,6071	0,0071	0,1413	0,0200
	<i>MGPH</i> (2)	54	0,6045	0,0045	0,1031	0,0106
	<i>FT</i>	-	0,6080	0,0080	0,0517	<b>0,0027</b>

Quando consideramos  $d^* > 1, 0$  (ver Tabela 3.4.2 para o caso  $d^* = 1, 2$ ), os estimadores não são bons. Isto deve-se ao fato de que, neste caso, não vale a propriedade de reversão do nível. Observe que todos os estimadores apresentados são muito viciados. Quanto mais  $d^*$  se afasta do valor de um, maior é o vício resultante dos estimadores. Uma possível solução seria aplicar uma primeira diferença e então estimar o parâmetro de diferenciação. Analisaremos este problema na Seção 3.5.

**Tabela 3.4.2:** Estimativas para o modelo  $ARFIMA(0; 1, 2; 0)$ .

$n$	Método	$g(n)$	média( $\hat{d}^*$ )	vício( $\hat{d}^*$ )	$sd(\hat{d}^*)$	$eqm(\hat{d}^*)$
100	<i>GPH</i>	10	1,0717	<b>-0,1283</b>	0,2494	0,0763
	<i>SPR</i>	10	1,0466	-0,1534	0,1775	0,0550
	<i>GPht(1)</i>	16	1,0700	-0,1300	0,2570	0,0828
	<i>GPht(2)</i>	25	1,0637	-0,1363	0,1668	0,0463
	<i>MGPH(1)</i>	16	1,0538	-0,1462	0,1766	0,0525
	<i>MGPH(2)</i>	25	1,0332	-0,1668	0,1313	0,0450
	<i>FT</i>	-	1,0568	-0,1432	0,0951	<b>0,0295</b>
300	<i>GPH</i>	17	1,0748	-0,1252	0,1663	0,0433
	<i>SPR</i>	17	1,0883	<b>-0,1117</b>	0,1235	0,0277
	<i>GPht(1)</i>	31	1,0613	-0,1387	0,1519	0,0423
	<i>GPht(2)</i>	54	1,0591	-0,1409	0,1075	0,0314
	<i>MGPH(1)</i>	31	1,0651	-0,1349	0,1263	0,0341
	<i>MGPH(2)</i>	54	1,0503	-0,1497	0,0972	0,0318
	<i>FT</i>	-	1,0636	-0,1364	0,0735	<b>0,0240</b>

**Caso  $ARFIMA(p, d^*, q)$**

A Tabela 3.4.3 apresenta o caso em que  $p = 1$ ,  $q = 0$  e  $d^* = 0,6$ . Além dos resultados da estimação de  $d^*$ , esta tabela também mostra resultados da estimação do parâmetro de curta memória (parte autoregressiva do modelo).

No método *FT*, os parâmetros são estimados simultaneamente através da subrotina BCONF do IMSL. Para os métodos semiparamétricos, os parâmetros de curta memória (partes autoregressiva ou média móvel) são estimados utilizando a subrotina NSLSE depois da série ter sido diferenciada da estimativa de  $d^*$ .

A Tabela 3.4.3 revela o impacto nas estimativas quando a parte *AR* é incluída no modelo. Para valores negativos de  $\phi$ , os métodos ainda funcionam bem e são competitivos. O estimador *FT* fornece, em geral, melhores estimativas. No entanto, para valores positivos de  $\phi$ , *FT* perde sua superioridade mostrando maiores vício e erro quadrático médio. Nesta situação, o vício é predominantemente positivo para  $d^*$  e é negativo para o parâmetro  $\phi$ .

O método *SPR* é o mais adequado, fornecendo menores valores para o vício e para o erro quadrático médio. Este estimador é o que apresenta menor impacto quando parâmetros de curta memória estão no modelo. Isto é mais evidente quando  $\phi$  varia de 0,5 a 0,7. O vício e o erro quadrático médio aumentam para este estimador, mas não significativamente, como acontece nos outros métodos.

**Tabela 3.4.3:** Estimativas para o modelo  $ARFIMA(1; 0, 6; 0)$  quando  $n = 300$ .

$\phi$		<i>GPH</i>	<i>SPR</i>	<i>GPht</i> (1)	<i>GPht</i> (2)	<i>MGPH</i> (1)	<i>MGPH</i> (2)	<i>FT</i>
-0,5	<i>média</i> ( $\hat{d}^*$ )	0,6234	0,5613	0,5996	0,5487	0,6093	0,5613	0,6063
	<i>vício</i> ( $\hat{d}^*$ )	0,0234	-0,0387	<b>-0,0004</b>	-0,0513	0,0093	-0,0387	0,0063
	<i>sd</i> ( $\hat{d}^*$ )	0,2049	0,1668	0,1873	0,1219	0,1326	0,0918	0,0661
	<i>eqm</i> ( $\hat{d}^*$ )	0,0424	0,0292	0,0350	0,0174	0,0176	0,0099	<b>0,0044</b>
	<i>média</i> ( $\hat{\phi}$ )	-0,4834	-0,4554	-0,4576	-0,4738	-0,4947	-0,4728	-0,5104
	<i>sd</i> ( $\hat{\phi}$ )	0,1537	0,1400	0,1004	0,1509	0,1008	0,0795	0,1146
	<i>eqm</i> ( $\hat{\phi}$ )	0,0238	0,0215	0,0118	0,0234	0,0101	<b>0,0070</b>	0,0132
-0,2	<i>média</i> ( $\hat{d}^*$ )	0,6171	0,5685	0,5914	0,5829	0,6028	0,5854	0,6080
	<i>vício</i> ( $\hat{d}^*$ )	0,0171	-0,0315	-0,0086	-0,0171	<b>0,0028</b>	-0,0146	0,0080
	<i>sd</i> ( $\hat{d}^*$ )	0,2058	0,1646	0,1922	0,1229	0,1356	0,0958	0,0685
	<i>eqm</i> ( $\hat{d}^*$ )	0,0426	0,0280	0,0369	0,0154	0,0184	0,0094	<b>0,0047</b>
	<i>média</i> ( $\hat{\phi}$ )	-0,1847	-0,1548	-0,1762	-0,1661	-0,1920	-0,1844	-0,1960
	<i>sd</i> ( $\hat{\phi}$ )	0,1939	0,1647	0,1261	0,1907	0,1319	0,1002	0,0822
	<i>eqm</i> ( $\hat{\phi}$ )	0,0377	0,0291	0,0164	0,0374	0,0174	0,0102	<b>0,0067</b>
0,2	<i>média</i> ( $\hat{d}^*$ )	0,6287	0,5721	0,6304	0,6625	0,6304	0,6465	0,7494
	<i>vício</i> ( $\hat{d}^*$ )	0,0287	<b>-0,0279</b>	0,0304	0,0625	0,0304	0,0465	0,1494
	<i>sd</i> ( $\hat{d}^*$ )	0,2053	0,1623	0,1983	0,1272	0,1462	0,0992	0,4967
	<i>eqm</i> ( $\hat{d}^*$ )	0,0429	0,0270	0,0401	0,0201	0,0222	<b>0,0120</b>	0,2684
	<i>média</i> ( $\hat{\phi}$ )	0,1788	0,2272	0,1369	0,1777	0,1685	0,1483	0,1357
	<i>sd</i> ( $\hat{\phi}$ )	0,2071	0,1743	0,1280	0,2009	0,1536	0,1072	0,2994
	<i>eqm</i> ( $\hat{\phi}$ )	0,0432	0,0310	0,0203	0,0407	0,0245	<b>0,0141</b>	0,0935
0,5	<i>média</i> ( $\hat{d}^*$ )	0,6573	0,6018	0,7468	0,8499	0,7032	0,7869	1,2547
	<i>vício</i> ( $\hat{d}^*$ )	0,0573	<b>0,0018</b>	0,1468	0,2499	0,1032	0,1869	0,6547
	<i>sd</i> ( $\hat{d}^*$ )	0,2070	0,1683	0,1891	0,1259	0,1366	0,0966	0,6880
	<i>eqm</i> ( $\hat{d}^*$ )	0,0460	<b>0,0283</b>	0,0572	0,0783	0,0292	0,0443	0,9090
	<i>média</i> ( $\hat{\phi}$ )	0,4313	0,4822	0,2499	0,3478	0,3867	0,3067	0,0502
	<i>sd</i> ( $\hat{\phi}$ )	0,1928	0,1595	0,1193	0,1761	0,1323	0,0973	0,4657
	<i>eqm</i> ( $\hat{\phi}$ )	0,0418	<b>0,0257</b>	0,0767	0,0541	0,0303	0,0467	0,4186
0,7	<i>média</i> ( $\hat{d}^*$ )	0,7328	0,6644	0,9232	1,0652	0,8305	0,9598	1,2487
	<i>vício</i> ( $\hat{d}^*$ )	0,1328	<b>0,0644</b>	0,3232	0,4652	0,2305	0,3598	0,6487
	<i>sd</i> ( $\hat{d}^*$ )	0,2145	0,1803	0,1871	0,1223	0,1469	0,1017	0,6514
	<i>eqm</i> ( $\hat{d}^*$ )	0,0635	<b>0,0366</b>	0,1394	0,2314	0,0747	0,1397	0,8441
	<i>média</i> ( $\hat{\phi}$ )	0,5600	0,6218	0,2591	0,3905	0,4769	0,3570	0,1621
	<i>sd</i> ( $\hat{\phi}$ )	0,1840	0,1518	0,1229	0,1797	0,1421	0,1069	0,4958
	<i>eqm</i> ( $\hat{\phi}$ )	0,0533	<b>0,0291</b>	0,2094	0,1280	0,0699	0,1290	0,5344

Nos métodos *GPht* e *MGPH* observamos que  $g(n)$  produz efeito contrário ao apresentado para os modelos  $ARFIMA(0, d^*, 0)$ . Agora os valores do vício e do erro quadrático médio aumentam com  $g(n)$ , especialmente para  $\phi$  positivo. Isto não surpreende pois a presença de componentes *AR* no modelo contribui fortemente na função densidade espectral do processo em certas frequências (não próximas de zero mas frequências ainda envolvidas na equação de regressão).

**Observação:** No método *FT*, ao utilizarmos a subrotina *BCONF* do *IMSL*, é necessário que se limite o intervalo de estimação, que consideramos

$[-1, 9; 1, 9]$ . Esta subrotina nem sempre é eficiente e em algumas situações define o máximo do intervalo como sendo a estimativa que, no nosso caso, é 1,9.

As estimativas de  $d^*$  são muito afetadas quando incluímos parâmetros de curta dependência autoregressivos ou médias móveis. O método *FT* apresenta maiores vício e erro quadrático médio, principalmente quando  $\phi > 0$  ou  $\theta < 0$ . O vício de  $d^*$  é predominantemente positivo e para os parâmetros *AR* e *MA* é predominantemente negativo.

Os vícios na estimação dos parâmetros de curta e longa dependência são maiores para o caso  $ARFIMA(0, d^*, 1)$  do que para o caso  $ARFIMA(1, d^*, 0)$ .

No modelo  $ARFIMA(1, d^*, 1)$  o vício na estimação do parâmetro de longa dependência é relativamente pequeno para  $\phi = -0.6$  comparado com  $\phi = 0.2$  para combinações de valores de  $\theta$ . O método *FT* sempre superestima o parâmetro de diferenciação enquanto os métodos semiparamétricos superestimam e subestimam para valores de  $\phi$  negativos e positivos, respectivamente.

Uma análise detalhada para modelos  $ARFIMA(0, d^*, 1)$  e  $ARFIMA(1, d^*, 1)$  para vários valores de  $\phi$  e  $\theta$  é apresentada em Lopes et al. (2002a, 2002b).

Foram feitas várias análises utilizando também os estimadores *V* e *VP* para o parâmetro de longa dependência no modelo  $ARFIMA$ , mas os resultados não foram satisfatórios. Os vícios gerados por estes estimadores foram muito altos tanto para o caso estacionário como para o caso não estacionário. Beran (1994) afirma que o estimador *VP* não é um estimador adequado. No caso do método *V*, o alto vício pode ser explicado pela Proposição 3.5, que apresenta uma cota inferior para o valor esperado do estimador. Apresentamos os resultados do método *V* na Tabela 3.4.4, com base em 100 replicações.

**Tabela 3.4.4:** Estimativas para o modelo  $ARFIMA(0, d, 0)$ .

$d$	$n$	$V$	$sd(\hat{d})$	$eqm(\hat{d})$
0,3	500	0,0697	0,0846	0,0601
	1000	0,0780	0,0699	0,0541
	2000	0,1072	0,0637	0,0412
	3000	0,1151	0,0621	0,0379
0,8	500	0,5210	0,0486	0,0801
	1000	0,5340	0,0455	0,0727
	2000	0,5555	0,0698	0,0646
	3000	0,5640	0,0567	0,0588

Podemos observar, através desta tabela, que o aumento do tamanho amostral melhora um pouco os valores das estimativas tanto no caso esta-

cionário ( $d = 0,3$ ) como no caso não estacionário ( $d = 0,8$ ). No entanto, observamos que nos dois casos o vício continua grande mesmo com o aumento do tamanho amostral.

**Conclusão:** Apresentamos resultados de simulação para avaliar o comportamento de alguns métodos de estimação do parâmetro  $d^*$  no caso de processos *ARFIMA* não estacionários. Quando o parâmetro de diferenciação  $d^*$  está no intervalo  $(0,5;1,0)$  os estimadores são bons e o método *FT* fornece, em geral, melhores resultados. Este estimador apresenta piores resultados que os outros métodos quando são incluídos parâmetros de curta dependência no modelo, tanto autoregressivos como médias móveis. Os estimadores *GPH*, *SPR*, *GPht*, *MGPH* e *FT* melhoram sempre que o tamanho amostral aumenta, especialmente no modelo *ARFIMA*(0,  $d^*$ , 0).

No caso não estacionário sem propriedade de reversão do nível (isto é, quando  $d^* \in (1,0;1.5)$ ), observamos que todos os métodos analisados nesta seção são muito viciados e subestimam o verdadeiro valor do parâmetro. Uma alternativa para este caso é utilizar o método de estimação *W*, baseado na teoria de wavelets. Poli e Lopes (2002) apresentam resultados satisfatórios obtidos através de diferentes bases de wavelets.

Os métodos *V* e *VP* são muito viciados e não são recomendados para nenhum dos casos: estacionário ou não estacionário.

### 3.4.3 Erros Não-Gaussianos

Nesta seção analisamos o impacto dos estimadores do parâmetro fracionário em modelos *ARFIMA* quando a distribuição do processo de inovação, isto é, do processo  $\{\epsilon_t\}_{t \in \mathbb{Z}}$ , é não-Gaussiana.

Para esta análise, consideramos inovações com as seguintes distribuições: uniforme  $\mathcal{U}(-\sqrt{3}, \sqrt{3})$ , exponencial com parâmetro  $\lambda = 1$  (com média recentralizada em zero), *t*-Student com 3 graus de liberdade e  $\chi_1^2$  (com média recentralizada em zero). Como os resultados foram muito semelhantes à situação onde o processo  $\{\epsilon_t\}_{t \in \mathbb{Z}}$  tem distribuição Gaussiana, apresentamos somente dois casos: o modelo *ARFIMA*(0; 0, 6; 0) considerando o processo de inovação com distribuição exponencial com parâmetro  $\lambda = 1$  e média recentralizada em zero, e o modelo *ARFIMA*(0; 1, 2; 0) considerando o processo de inovação com distribuição  $\mathcal{U}(-\sqrt{3}, \sqrt{3})$ .

**Tabela 3.4.5:** Estimativas para o modelo  $ARFIMA(0; 0, 6; 0)$  quando o processo de inovação tem distribuição exponencial  $\mathcal{E}(1)$ .

$n$	Método	$g(n)$	média( $d^*$ )	vício( $d^*$ )	$sd(d^*)$	$eqm(d^*)$
100	<i>GPH</i>	10	0,6470	0,0470	0,2830	0,0821
	<i>SPR</i>	10	0,5291	-0,0709	0,2239	0,0550
	<i>GPht(1)</i>	16	0,6361	0,0361	0,3702	0,1380
	<i>GPht(2)</i>	25	0,6138	0,0138	0,2261	0,0512
	<i>MGPH(1)</i>	16	0,6330	0,0330	0,2178	0,0484
	<i>MGPH(2)</i>	25	0,6070	0,0070	0,1537	0,0236
	<i>FT</i>	-	0,6053	<b>0,0053</b>	0,0914	<b>0,0084</b>
300	<i>GPH</i>	17	0,6339	0,0339	0,1897	0,0371
	<i>SPR</i>	17	0,5702	-0,0298	0,1640	0,0277
	<i>GPht(1)</i>	31	0,6132	0,0132	0,1766	0,0313
	<i>GPht(2)</i>	54	0,6209	0,0209	0,1196	0,0147
	<i>MGPH(1)</i>	31	0,6158	0,0158	0,1228	0,0153
	<i>MGPH(2)</i>	54	0,6137	0,0137	0,0914	0,0085
	<i>FT</i>	-	0,6067	<b>0,0067</b>	0,0486	<b>0,0024</b>

**Tabela 3.4.6:** Estimativas para o modelo  $ARFIMA(0; 1, 2; 0)$  quando o processo de inovação tem distribuição uniforme  $\mathcal{U}(-\sqrt{3}, \sqrt{3})$ .

$n$	Método	$g(n)$	média( $d^*$ )	vício( $d^*$ )	$sd(d^*)$	$eqm(d^*)$
100	<i>GPH</i>	10	1,0850	<b>-0,1150</b>	0,2553	0,0783
	<i>SPR</i>	10	1,0484	-0,1516	0,1798	0,0553
	<i>GPht(1)</i>	16	1,0778	-0,1222	0,2677	0,0865
	<i>GPht(2)</i>	25	1,0787	-0,1213	0,1820	0,0478
	<i>MGPH(1)</i>	16	1,0660	-0,1340	0,1842	0,0518
	<i>MGPH(2)</i>	25	1,0483	-0,1517	0,1407	0,0428
	<i>FT</i>	-	1,0657	-0,1343	0,0932	<b>0,0267</b>
300	<i>GPH</i>	17	1,0684	<b>-0,1316</b>	0,1692	0,0459
	<i>SPR</i>	17	1,0684	<b>-0,1316</b>	0,1228	0,0270
	<i>GPht(1)</i>	31	1,0532	-0,1468	0,1475	0,0433
	<i>GPht(2)</i>	54	1,0492	-0,1508	0,1080	0,0344
	<i>MGPH(1)</i>	31	1,0554	-0,1446	0,1209	0,0355
	<i>MGPH(2)</i>	54	1,0400	-0,1600	0,0959	0,0348
	<i>FT</i>	-	1,0585	-0,1415	0,0684	<b>0,0247</b>

Ambas as análises são baseadas em 800 replicações para situações onde  $n = 100$  e  $300$ . As Tabelas 3.4.5 e 3.4.6 mostram que os estimadores apresentam o mesmo comportamento observado para a distribuição de inovação Gaussiana (compare as Tabelas 3.4.1 e 3.4.5 com as Tabelas 3.4.2 e 3.4.6). Velasco (2000) apresenta uma análise quando o processo é estacionário e o processo de inovação tem distribuição não-Gaussiana.

**Conclusão:** Nesta seção, observamos que as estimativas do parâmetro de diferenciação  $d^*$  em processos *ARFIMA*, quando as inovações são não-Gaussianas, mantêm o mesmo comportamento dos resultados obtidos, no caso Gaussiano, para os métodos analisados.

### 3.4.4 Uma Aplicação

Vamos analisar, nesta seção, a série temporal “*taxa de juros a curto prazo da Inglaterra*” quanto à longa dependência e à presença de raiz unitária, aplicando alguns métodos de estimação descritos na Seção 3.3. A série temporal consiste de 148 observações medidas trimestralmente de 1952 a 1988 (ver Anexo B). Esta série temporal é apresentada e analisada em Mills (1997) com especial interesse na teoria de raízes unitárias. A Figura 3.4.1 apresenta o gráfico da série e as suas funções de autocorrelação amostral e periodograma. Podemos observar o comportamento de não estacionariedade e característica de longa dependência presentes nesta série temporal, através dos gráficos (a), (b) e (c), respectivamente.

O teste de raiz unitária não fornece evidências contra a hipótese de que a série temporal contém uma raiz unitária, isto é,  $d^* = 1,0$  (ver Mills (1997)). Os resultados de estimação e identificação do modelo para esta série temporal estão apresentados na Tabela 3.4.7.

O estimador *GPH* indica que o processo é um *ARFIMA*(0, 1, 0), onde  $GPH = 1,0115$  com  $\sigma(GPH) = 0,2560$ . Os métodos *GPHt* e *MGPH* fornecem estimativas similares e seus valores estão entre as estimativas obtidas pelos métodos *SPR* e *FT*. Poli e Lopes (2002) analisaram este conjunto de dados reais através do estimador baseado em wavelets utilizando as 128 observações iniciais (foram excluídas as 20 últimas observações para que o número de dados fosse uma potência de 2). Os resultados obtidos foram: 0,8987 com a base Haar e 0,8946 com a base de wavelets Chapéu Mexicano.

Ressaltamos que, para  $g(n) = 33$ ,  $GPHt(2) = 0,8014$  ( $\sigma(GPHt(2)) = 0,1479$ ) e  $MGPH(2) = 0,8362$  ( $\sigma(MGPH(2)) = 0,1116$ ). Para o método *FT* precisamos das ordens *AR* e *MA* do processo. Diferentes valores de  $p$  e  $q$  foram testados e os resultados sugerem um modelo *ARFIMA*(0,  $d^*$ , 0). Para

os métodos semiparamétricos a ordem do processo *ARMA* foi identificada após a diferenciação da série da estimativa de  $d^*$ .

A escolha do modelo foi feita testando as estimativas das partes autoregressiva e média móvel e pelo critério *AIC* (Akaike Information Criterion).

**Figura 3.4.1:** (a) série temporal “*taxa de juros a curto prazo da Inglaterra*”; (b) função de autocorrelação amostral; (c) função periodograma.

As estimativas dos parâmetros de curta memória e as demais estatísticas apresentadas na Tabela 3.4.2 foram obtidas através do pacote estatístico MINITAB.

Exceto para o estimador obtido pelo método *GPH*, todos os outros métodos indicam que a série temporal é não estacionária com  $d^* < 1, 0$ .

A análise dos resíduos para os modelos ajustados indicou que os erros são aproximadamente um ruído branco gaussiano.

Baseado no teste modificado de Ljung-Box (*MBP*), a hipótese de adequação do modelo não foi rejeitada para todos os casos ao nível de significância de 5%.

A análise de previsão também foi considerada calculando-se o erro quadrático médio e o erro percentual absoluto médio (*MAPE*) (ver Wei (1990), página 179) com valor inicial para previsão dado por  $t = 135$ .

**Tabela 3.4.7:** Identificação, estimação e resultados de previsão.

Estimador	<i>ARFIMA</i> (0, <i>FT</i> , 0)	<i>ARFIMA</i> (0, <i>SPR</i> , 1)	<i>ARFIMA</i> (0, 1, 0)
$\hat{d}^*$	0,8514	0,7675	-
$\sigma(\hat{d}^*)$	0,0640	0,1140	-
$\hat{\theta}_1$	-	-0,1936	-
$\sigma(\hat{\theta}_1)$	-	0,0836	-
$\hat{\sigma}_\epsilon^2$	1,4671	1,4553	1,5089
<i>AIC</i>	53,30	56,12	57,18
<i>MBP</i>	0,108	0,107	0,064
<i>eqm</i>	1,126	1,041	1,171
<i>MAPE</i> (%)	33,70	32,20	34,60

*MBP*- Estatística Ljung-Box ( $\chi^2$  com 11 graus de liberdade)

**Conclusão:** A identificação e a previsão (ver Reisen e Lopes (1999) para mais detalhes sobre análise de previsão em processos *ARFIMA*) indicam que a série temporal é não estacionária com longa dependência onde o parâmetro de diferenciação é estimado por  $\hat{d}^* < 1, 0$  (ver Tabela 3.4.7 e Poli e Lopes (2002)). Portanto, considerar  $d^* = 1, 0$  para a série temporal “*taxa de juros a curto prazo da Inglaterra*”, como Mills (1997) sugere, acarreta em uma super diferenciação da série temporal.

### 3.5 Invariância para Primeira Diferença

Nesta seção vamos analisar o comportamento do estimador do parâmetro de diferenciação após aplicar primeira diferença na série original. Seja  $\{X_t\}_{t \in \mathbb{Z}}$

um processo  $ARFIMA(0, d^*, 0)$ , onde  $d^* = d + r$ , dado pela expressão (3.28). Seja  $\hat{d}^*$  o estimador de  $d^*$  através de alguns dos métodos descritos anteriormente na Seção 3.3. Diferenciamos a série de 1 e estimamos novamente o parâmetro de diferenciação obtendo  $\hat{d}$ . O objetivo é verificar se vale a igualdade  $\hat{d}^* = \hat{d} + 1$ .

Hurvich e Ray (1995) analisam o problema exposto acima considerando apenas o método  $GPH$ , descrito na Seção 3.3, e um método utilizando “tapering”. Aqui, neste trabalho, fazemos um estudo através de simulação de Monte Carlo considerando os estimadores  $SPR$ ,  $GPHT$ ,  $MGPH$  e  $FT$ .

Apresentamos a seguir alguns resultados, obtidos a partir de séries temporais com tamanhos amostrais  $n = 256, 512$  e  $1024$  baseados em 2000 replicações do processo.

Foram calculados os valores empíricos da média, do desvio padrão e do erro quadrático médio. Estes valores estão nas tabelas a seguir. O erro quadrático médio de menor valor está em negrito.

O valor de  $g(n) = n^\alpha$  nos métodos semiparamétricos  $GPH$  e  $SPR$  está fixado em  $\alpha = 0,5$  (valor amplamente utilizado na literatura). O ponto de truncamento na janela de Parzen para o estimador  $SPR$  é  $m = n^\beta$ , com  $\beta = 0,9$  (ver Reisen (1994) e Olbermann (1998) para uma discussão sobre o valor de  $\beta$ ). Para os estimadores  $GPHT$  e  $MGPH$ , dois diferentes valores para o limite  $g(n) = n^{\alpha_i}$  foram usados: consideramos  $\alpha_1 = 0,6$  ( $GPHT(1)$  e  $MGPH(1)$ ) e  $\alpha_2 = 0,7$  ( $GPHT(2)$  e  $MGPH(2)$ ). Para o estimador  $GPHT$  também foram usados dois diferentes valores para o limite  $g(n) = n^{\alpha_i}$ : neste estimador consideramos  $\alpha_1 = 0,5$  ( $GPHT(1)$ ) e  $\alpha_2 = 0,7$  ( $GPHT(2)$ ). O regressor inicial  $l$  nos métodos de regressão é  $l = 1$ , exceto para  $GPHT$  e  $GPHT$ , onde consideramos  $l = 2$ .

As tabelas a seguir apresentam os resultados dos estimadores do parâmetro de diferenciação de modelos  $ARFIMA$  não estacionários, antes e após a primeira diferença da série. Para o cálculo do erro quadrático médio depois de diferenciar a série, consideramos o verdadeiro valor do parâmetro igual a  $1 - d^*$ .

### Caso $ARFIMA(0, d^*, 0)$

Primeiramente analisamos os modelos  $ARFIMA(0, d^*, 0)$  para  $d^* \in (0, 5; 1, 0]$  e  $d^* \in (1, 0; 1, 5)$ .

Através das simulações apresentadas nas Tabelas 3.5.1 e 3.5.2, quando  $d^* \in (0, 5; 1, 0)$  todos os métodos analisados comprovam a relação  $\hat{d}^* \simeq \hat{d} + 1$ .

**Tabela 3.5.1:** Estimação para o modelo  $ARFIMA(0; 0, 6; 0)$ , com tamanhos amostrais 256 e 512.

$n$	Método	média( $\hat{d}^*$ )	sd( $\hat{d}^*$ )	eqm( $\hat{d}^*$ )	média( $\hat{d}$ )	sd( $\hat{d}$ )	eqm( $\hat{d}$ )
256	<i>GPH</i>	0,6183	0,2222	0,0497	-0,3782	0,2143	0,0464
	<i>SPR</i>	0,5566	0,1773	0,0333	-0,3860	0,1525	0,0234
	<i>MGPH</i> (1)	0,6106	0,1499	0,0226	-0,3834	0,1470	0,0219
	<i>MGPH</i> (2)	0,6050	0,1074	0,0115	-0,3809	0,1042	0,0112
	<i>FT</i>	0,6076	0,0552	<b>0,0031</b>	-0,3965	0,0544	<b>0,0030</b>
	<i>GPht</i> (1)	0,6136	0,2004	0,0403	-0,3927	0,2007	0,0403
	<i>GPht</i> (2)	0,6147	0,1321	0,0176	-0,3909	0,1320	0,0175
	<i>GPHT</i> (1)	0,6041	0,6568	0,4310	-0,4214	0,5836	0,3408
	<i>GPHT</i> (2)	0,6065	0,2473	0,0611	-0,3936	0,2355	0,0555
512	<i>GPH</i>	0,6186	0,1730	0,0303	-0,3738	0,1721	0,0303
	<i>SPR</i>	0,5712	0,1443	0,0216	-0,3824	0,1267	0,0163
	<i>MGPH</i> (1)	0,6180	0,1196	0,0146	-0,3791	0,1149	0,0136
	<i>MGPH</i> (2)	0,6080	0,0827	0,0069	-0,3839	0,0799	0,0066
	<i>FT</i>	0,6039	0,0405	<b>0,0017</b>	-0,3992	0,0385	<b>0,0015</b>
	<i>GPht</i> (1)	0,6232	0,1508	0,0233	-0,3860	0,1454	0,0213
	<i>GPht</i> (2)	0,6143	0,0958	0,0094	-0,3925	0,0938	0,0089
	<i>GPHT</i> (1)	0,6078	0,4836	0,2337	-0,4077	0,4334	0,1877
	<i>GPHT</i> (2)	0,6059	0,1673	0,0280	-0,4034	0,1611	0,0259

**Tabela 3.5.2:** Estimação para o modelo  $ARFIMA(0, d^*, 0)$ , com tamanho amostral 512.

$d^*$	Método	média( $\hat{d}^*$ )	sd( $\hat{d}^*$ )	eqm( $\hat{d}^*$ )	média( $\hat{d}$ )	sd( $\hat{d}$ )	eqm( $\hat{d}$ )
0,8	<i>GPH</i>	0,8332	0,1651	0,0283	-0,1901	0,1734	0,0301
	<i>SPR</i>	0,7975	0,1419	0,0201	-0,2082	0,1218	0,0149
	<i>MGPH</i> (1)	0,8338	0,1179	0,0150	-0,1889	0,1138	0,0131
	<i>MGPH</i> (2)	0,8232	0,0850	0,0077	-0,1932	0,0797	0,0064
	<i>FT</i>	0,8231	0,0439	<b>0,0025</b>	-0,1965	0,0358	<b>0,0013</b>
	<i>GPht</i> (1)	0,8373	0,1510	0,0242	-0,1892	0,1472	0,0218
	<i>GPht</i> (2)	0,8306	0,1005	0,0110	-0,1964	0,0968	0,0094
	<i>TAPE</i> (1)	0,8254	0,5118	0,2624	-0,2071	0,4462	0,1989
	<i>TAPE</i> (2)	0,8236	0,1714	0,0299	-0,1887	0,1665	0,0278
1,0	<i>GPH</i>	0,9775	0,1522	0,0236	0,0160	0,1712	0,0295
	<i>SPR</i>	0,9761	0,1308	0,0176	-0,0204	0,1262	0,0163
	<i>MGPH</i> (1)	0,9846	0,1067	0,0116	0,0146	0,1157	0,0136
	<i>MGPH</i> (2)	0,9824	0,0674	0,0048	0,0115	0,0786	0,0063
	<i>FT</i>	0,9964	0,0314	<b>0,0010</b>	0,0131	0,0365	<b>0,0015</b>
	<i>GPht</i> (1)	0,9936	0,1325	0,0175	0,0123	0,1452	0,0212
	<i>GPht</i> (2)	0,9964	0,0785	0,0062	0,0098	0,0937	0,0088
	<i>GPHT</i> (1)	0,9925	0,4916	0,2412	0,0067	0,4173	0,1738
	<i>GPHT</i> (2)	1,0111	0,1574	0,0248	-0,0017	0,1569	0,0246

**Tabela 3.5.3:** Estimação para  $d^* > 1, 0$ , com tamanho amostral 512.

$d^*$	Método	média( $\hat{d}^*$ )	sd( $\hat{d}^*$ )	eqm( $\hat{d}^*$ )	média( $\hat{d}$ )	sd( $\hat{d}$ )	eqm( $\hat{d}$ )
1,1	<i>GPH</i>	1,0449	0,1490	0,0252	0,1064	0,1714	0,0294
	<i>SPR</i>	1,0577	0,1212	0,0165	0,0672	0,1318	0,0184
	<i>MGPH</i> (1)	1,0408	0,1059	0,0147	0,1013	0,1158	0,0134
	<i>MGPH</i> (2)	1,0310	0,0770	0,0107	0,0974	0,0798	0,0064
	<i>FT</i>	1,0453	0,0437	<b>0,0049</b>	0,0978	0,0364	<b>0,0013</b>
	<i>GPHT</i> (1)	1,0409	0,1385	0,0227	0,1002	0,1533	0,0235
	<i>GPHT</i> (2)	1,0390	0,0913	0,0121	0,0968	0,0957	0,0092
	<i>GPHT</i> (1)	1,0704	0,4885	0,2392	0,0816	0,4495	0,2022
	<i>GPHT</i> (2)	1,0803	0,1591	0,0257	0,0989	0,1646	0,0271
1,3	<i>GPH</i>	1,0822	0,1550	0,0714	0,2982	0,1682	0,0282
	<i>SPR</i>	1,1308	0,1014	<b>0,0389</b>	0,2624	0,1313	0,0186
	<i>MGPH</i> (1)	1,0603	0,1156	0,0708	0,3015	0,1184	0,0140
	<i>MGPH</i> (2)	1,0436	0,0975	0,0752	0,3000	0,0799	0,0064
	<i>FT</i>	1,0558	0,0831	0,0665	0,3033	0,0342	<b>0,0012</b>
	<i>GPHT</i> (1)	1,0528	0,1254	0,0768	0,3011	0,1569	0,0246
	<i>GPHT</i> (2)	1,0461	0,1004	0,0745	0,3025	0,0991	0,0098
	<i>GPHT</i> (1)	1,1476	0,5084	0,2812	0,3031	0,4477	0,2000
	<i>GPHT</i> (2)	1,1718	0,1624	0,0428	0,3098	0,1596	0,0255
1,45	<i>GPH</i>	1,0462	0,1239	0,1783	0,4647	0,1815	0,0329
	<i>SPR</i>	1,1254	0,0756	0,1110	0,4189	0,1407	0,0206
	<i>MGPH</i> (1)	1,0342	0,1064	0,1841	0,4581	0,1223	0,0149
	<i>MGPH</i> (2)	1,0194	0,0954	0,1945	0,4512	0,0829	0,0068
	<i>FT</i>	1,0255	0,0886	0,1880	0,4721	0,0352	<b>0,0017</b>
	<i>GPHT</i> (1)	1,0283	0,1066	0,1891	0,4513	0,1580	0,0248
	<i>GPHT</i> (2)	1,0230	0,0947	0,1912	0,4509	0,0954	0,0090
	<i>GPHT</i> (1)	1,1563	0,6455	0,4996	0,5398	0,4554	0,2138
	<i>GPHT</i> (2)	1,2452	0,1774	<b>0,0732</b>	0,4645	0,1514	0,0229

Para o caso  $d^* = 1, 0$  (observe Tabela 3.5.2), após diferenciar a série de 1 os estimadores  $\hat{d}$  ficam muito próximos de zero. Isto significa que  $\hat{d}^* \simeq \hat{d} + 1$ .

Para  $d^* > 1$  os estimadores não são bons, a relação não é tão evidente. Na Tabela 3.5.3 temos o caso  $d^* > 1, 0$ , como os estimadores são muito viciados nessa faixa, não vale a invariância para primeira diferença.

### Caso $ARFIMA(p, d^*, q)$

Nas tabelas, a seguir, vamos apresentar as simulações para analisar a invariância para a primeira diferença nos modelos  $ARFIMA(1, d^*, 0)$ ,  $ARFIMA(0, d^*, 1)$  e  $ARFIMA(1, d^*, 1)$ .

Na Tabela 3.5.4 apresentamos os resultados da estimação de  $d^*$  para o modelo  $ARFIMA(1, d^*, 0)$  quando  $d^* = 0, 8$  e  $\phi = -0, 6$ . Podemos observar que o comportamento das estimativas de  $d^*$  segue como já descrevemos na Seção 3.4 deste Capítulo, onde estudamos os casos não estacionários. Como

$\phi = -0,6$ , os estimadores não sofrem muito impacto pela presença da parte autoregressiva, o que não se mantém quando  $\phi > 0$ , ou seja, a estimação de  $\hat{d}^*$  piora neste caso (ver Olbermann et. al. (2002)). A estimação de  $\phi$  segue o mesmo comportamento descrito anteriormente. Depois de diferenciar a série de 1 observamos que, neste caso, também vale a invariância dos estimadores. O estimador pelo método *GPHT*(1) apresenta grande desvio padrão, como já observado para o caso não estacionário, mesmo para tamanhos amostrais grandes. Vale ressaltar que depois de diferenciar a série de 1 significa que estamos no caso estacionário.

**Tabela 3.5.4:** Estimação para o modelo *ARFIMA*(1; 0, 8; 0) quando  $\phi = -0,6$  e tamanhos amostrais 256 e 1024.

$n$	Método	média( $\hat{d}^*$ )	sd( $\hat{d}^*$ )	eqm( $\hat{d}^*$ )	média( $\hat{d}$ )	sd( $\hat{d}$ )	eqm( $\hat{d}$ )
256	<i>GPHT</i>	0,8380	0,2058	0,0438	-0,1813	0,2025	0,0413
	<i>SPR</i>	0,7788	0,1770	0,0318	-0,2080	0,1482	0,0220
	<i>MGPHT</i> (1)	0,8125	0,1490	0,0223	-0,2014	0,1398	0,0195
	<i>MGPHT</i> (2)	0,7778	0,1133	0,0133	-0,2322	0,0992	0,0109
	<i>FT</i>	0,8474	0,0979	<b>0,0118</b>	-0,1967	0,0611	<b>0,0037</b>
	<i>GPHT</i> (1)	0,8035	0,2127	0,0452	-0,2207	0,2107	0,0448
	<i>GPHT</i> (2)	0,7735	0,1470	0,0223	-0,2578	0,1375	0,0222
	<i>GPHT</i> (1)	0,8178	0,6697	0,4483	-0,2318	0,6569	0,4322
	<i>GPHT</i> (2)	0,7554	0,2520	0,0654	-0,2886	0,2355	0,0633
1024	<i>GPHT</i>	0,8292	0,1372	0,0197	-0,1922	0,1298	0,0169
	<i>SPR</i>	0,8071	0,1159	0,0135	-0,2081	0,0995	0,0100
	<i>MGPHT</i> (1)	0,8256	0,0931	0,0093	-0,1943	0,0858	0,0074
	<i>MGPHT</i> (2)	0,8074	0,0689	<b>0,0048</b>	-0,2099	0,0585	0,0035
	<i>FT</i>	0,9047	0,1131	0,0237	-0,1956	0,0284	<b>0,0008</b>
	<i>GPHT</i> (1)	0,8241	0,1119	0,0131	-0,2005	0,1066	0,0113
	<i>GPHT</i> (2)	0,8082	0,0789	0,0063	-0,2173	0,0691	0,0051
	<i>GPHT</i> (1)	0,7994	0,3483	0,1212	-0,2059	0,3306	0,1092
	<i>GPHT</i> (2)	0,8047	0,1208	0,0146	-0,2201	0,1162	0,0139

A Tabela 3.5.5 apresenta os resultados para *ARFIMA*(0; 0, 8; 1) quando  $\theta = 0,2$ . Podemos observar que o estimador *FT* gera vício alto (devido a componente *MA*) e não vale a invariância após primeira diferença da série para este método. Os demais estimadores fornecem bons resultados neste caso e vale a relação  $\hat{d}^* \simeq \hat{d} + 1$ . Analisamos também a situação  $\theta = -0,2$  e observamos um comportamento análogo ao descrito para  $\theta = 0,2$ .

A Tabela 3.5.6 apresenta os resultados para tamanho amostral 256 com duas combinações para os valores de  $\phi$  e  $\theta$ . Neste caso o método *FT* não fornece boas estimativas apresentando alto erro quadrático médio e o método *MGPHT* é o que apresenta os melhores resultados. O método *GPHT* tem o maior desvio padrão para  $\hat{d}^*$  e para  $\hat{d}$ .

**Tabela 3.5.5:** Estimação para o modelo  $ARFIMA(0; 0, 8; 1)$  quando  $\theta = 0, 2$  e tamanhos amostrais 256 e 1024.

$n$	Método	média( $\hat{d}^*$ )	sd( $\hat{d}^*$ )	eqm( $\hat{d}^*$ )	média( $\hat{d}$ )	sd( $\hat{d}$ )	eqm( $\hat{d}$ )
256	<i>GPH</i>	0,8260	0,2145	0,0466	-0,2010	0,2160	0,0466
	<i>SPR</i>	0,7663	0,1734	0,0312	-0,2277	0,1565	0,0252
	<i>MGPH</i> (1)	0,8111	0,1503	0,0227	-0,2099	0,1443	0,0209
	<i>MGPH</i> (2)	0,7701	0,1129	<b>0,0136</b>	-0,2431	0,1053	<b>0,0129</b>
	<i>FT</i>	1,0172	0,3746	0,1874	-0,1693	0,1499	0,0234
	<i>GPHT</i> (1)	0,8100	0,2104	0,0443	-0,2202	0,2072	0,0433
	<i>GPHT</i> (2)	0,7677	0,1413	0,0210	-0,2649	0,1331	0,0219
	<i>GPHT</i> (1)	0,7181	0,5684	0,3294	-0,2355	0,6154	0,3796
	<i>GPHT</i> (2)	0,7450	0,2433	0,0622	-0,2783	0,2306	0,0593
1024	<i>GPH</i>	0,8183	0,1313	0,0175	-0,1983	0,1312	0,0172
	<i>SPR</i>	0,7951	0,1136	0,0129	-0,2157	0,1019	0,0106
	<i>MGPH</i> (1)	0,8096	0,0911	0,0084	-0,2076	0,0879	0,0078
	<i>MGPH</i> (2)	0,7897	0,0608	<b>0,0038</b>	-0,2215	0,0600	0,0041
	<i>FT</i>	1,5340	0,3119	0,6359	-0,2012	0,0474	<b>0,0022</b>
	<i>GPHT</i> (1)	0,8046	0,1088	0,0118	-0,2171	0,1071	0,0117
	<i>GPHT</i> (2)	0,7884	0,0690	0,0049	-0,2303	0,0677	0,0055
	<i>GPHT</i> (1)	0,7900	0,3544	0,1255	-0,2318	0,3211	0,1040
	<i>GPHT</i> (2)	0,7852	0,1171	0,0139	-0,2341	0,1172	0,0149

**Tabela 3.5.6:** Estimação para o modelo  $ARFIMA(1; 0, 8; 1)$  com tamanho amostral 256.

$\phi$	$\theta$	Método	média( $\hat{d}^*$ )	sd( $\hat{d}^*$ )	eqm( $\hat{d}^*$ )	média( $\hat{d}$ )	sd( $\hat{d}$ )	eqm( $\hat{d}$ )
-0,6	0,2	<i>GPH</i>	0,8223	0,1975	0,0395	-0,2029	0,2163	0,0467
		<i>SPR</i>	0,7595	0,1699	0,0306	-0,2270	0,1536	0,0243
		<i>MGPH</i> (1)	0,8001	0,1470	0,0216	-0,2300	0,1463	0,0223
		<i>MGPH</i> (2)	0,7291	0,1085	<b>0,0168</b>	-0,2893	0,1027	<b>0,0185</b>
		<i>FT</i>	0,8605	0,1712	0,0329	-0,1559	0,1665	0,0296
		<i>GPHT</i> (1)	0,7947	0,2066	0,0427	-0,2474	0,2089	0,0459
		<i>GPHT</i> (2)	0,7125	0,1385	0,0268	-0,3255	0,1352	0,0340
		<i>GPHT</i> (1)	0,7828	0,6333	0,4010	-0,2356	0,6318	0,4000
		<i>GPHT</i> (2)	0,6894	0,2515	0,0754	-0,3390	0,2294	0,0719
0,6	0,2	<i>GPH</i>	0,7581	0,2080	0,0450	-0,2635	0,2117	0,0488
		<i>SPR</i>	0,7033	0,1711	0,0386	-0,2810	0,1493	<b>0,0288</b>
		<i>MGPH</i> (1)	0,6998	0,1559	<b>0,0343</b>	-0,3284	0,1499	0,0389
		<i>MGPH</i> (2)	0,6023	0,1145	0,0522	-0,4173	0,1038	0,0580
		<i>FT</i>	0,8888	0,2690	0,0801	-0,1190	0,2716	0,0802
		<i>GPHT</i> (1)	0,6563	0,2161	0,0673	-0,3877	0,2153	0,0815
		<i>GPHT</i> (2)	0,5548	0,1457	0,0813	-0,4873	0,1384	0,1017
		<i>GPHT</i> (1)	0,6076	0,5375	0,3256	-0,3335	0,6096	0,3890
		<i>GPHT</i> (2)	0,5128	0,2490	0,1444	-0,5196	0,2292	0,1546

**Conclusão:** Para o modelo  $ARFIMA(0, d^*, 0)$ , quando  $d^* \in (0, 5; 1, 0]$ ,  $FT$  é o melhor método de estimação com a propriedade de invariância para primeira diferença, seguido por  $MGPH(2)$ . Para o modelo  $ARFIMA(0, d^*, 0)$ , quando  $d^* \in (1, 0; 1, 5)$ , nenhum dos métodos considerados é satisfatório para o caso não estacionário e, portanto, não vale a propriedade de invariância para primeira diferença.

No modelo  $ARFIMA(0, d^*, 1)$  o método  $FT$  apresenta resultados ruins para a estimação de  $d^*$ , mas as estimativas de  $d$  após a primeira diferença, apresentam baixo erro quadrático médio. Para este estimador não se verifica a propriedade de invariância para primeira diferença, nesta situação; mas para os outros métodos temos, em geral,  $\hat{d}^* \simeq \hat{d} + 1$ . Em ambos os modelos  $ARFIMA(1, d^*, 0)$  e  $ARFIMA(0, d^*, 1)$  foram obtidos bons resultados através do método  $MGPH(2)$ . Embora o método  $SPR$  apresente grande erro quadrático médio, ele foi satisfatório nas estimações simultâneas de  $d^*$  e  $d$ .

Para o modelo  $ARFIMA(1, d^*, 1)$  quando  $\phi$  e  $\theta$  são positivos, todos os métodos, exceto  $FT$ , subestimam  $d^*$  e, em geral, superestimam  $d$ . Para  $\phi$  negativo e  $\theta$  positivo,  $MGPH$  apresenta os melhores resultados para estimar  $d^*$ . Nesta situação, a relação  $\hat{d}^* \simeq \hat{d} + 1$  é mais evidente para o método  $SPR$ .

# 4

## PROCESSOS MANNEVILLE-POMEAU

Neste capítulo estudamos processos estocásticos  $\{X_t\}_{t \in \mathbb{Z}}$  que, apesar de apresentarem decaimento hiperbólico para a função de autocorrelação, não podem ser descritos através de processos  $ARFIMA(p, d, q)$  dados pela expressão (3.3). Consideraremos o processo estocástico obtido através de iterações da transformação de Manneville-Pomeau (ver Manneville (1980)), denotada por  $T_s$ . Zebrowski (2000) mostra que a propriedade de longa dependência, verificada nos batimentos cardíacos de pacientes com arritmia, está relacionada com a intermitência apresentada pelas transformações de Manneville-Pomeau. Utilizamos técnicas de estimação do parâmetro de longa dependência de processos Manneville-Pomeau, que são similares àquelas utilizadas em processos  $ARFIMA$ . Introduzimos um método baseado em wavelets na estimação do parâmetro nos processos Manneville-Pomeau, no caso de longa dependência. Além disso, na situação de dependência intermediária, apresentamos um novo método de estimação baseado na propriedade de Hölder para a função densidade espectral. Também consideramos outros exemplos de processos estacionários cujas análises são similares àquela para o processo obtido pela transformação de Manneville-Pomeau, a ser definida a seguir. Em todos os casos considerados nesta seção não existe expressão analítica conhecida para a função densidade espectral.

### 4.1 Definição e Propriedades da Transformação de Manneville-Pomeau

**DEFINIÇÃO 4.1:** A aplicação  $T_s : [0, 1] \rightarrow [0, 1]$  é dita *transformação de Manneville-Pomeau* (ver Figura 4.1.1 (a)) se é dada por

$$T_s(x) = x + x^{1+s} \pmod{1} = \begin{cases} x + x^{1+s}, & \text{se } x + x^{1+s} \leq 1, \\ x + x^{1+s} - 1, & \text{se } x + x^{1+s} > 1, \end{cases} \quad (4.1)$$

onde  $s$  é uma constante real positiva.

Como é usual, denotamos

$$T_s^t = \underbrace{T_s \circ \dots \circ T_s}_{t\text{-vezes}}.$$

**Propriedades.** A aplicação  $T_s$ , dada pela expressão (4.1), possui as seguintes propriedades:

- $T_s$  é monótona por partes com dois ramos completos, isto é, existe  $p \in \mathbb{N} - \{0\}$  tal que  $T_s|_{(0,p)}$  e  $T_s|_{(p,1)}$  são estritamente monótonas, contínuas e  $T_s((0,p)) = (0,1) = T_s((p,1))$ , onde  $p + p^{1+s} = 1$ .
- Os ramos  $T_s|_{(0,p)}$  e  $T_s|_{(p,1)}$  são  $C^1$ .
- $T_s'(x) = \frac{dT_s(x)}{dx} > 1$ , para todo  $x > 0$ , e  $T_s'(x) \geq \lambda > 1$ , para  $x \in (p,1)$ .
- $T_s$  tem um único ponto fixo indiferente. De fato, dizemos que  $x_0 \in [0,1]$  é *ponto fixo indiferente* de  $T$  se  $T(x_0) = x_0$  e  $|T'(x_0)| = 1$ . Por esta definição, 0 é o único ponto fixo indiferente da transformação  $T_s$ , dada pela expressão (4.1), pois

$$T_s(0) = 0 + 0^{1+s} = 0$$

e

$$T_s'(x) = 1 + (1+s)x^s.$$

Em resumo,  $T_s(0) = 0$  e  $|T_s'(0)| = 1$ .

- Existe uma medida invariante  $\mu_s$  para a transformação Manneville-Pomeau  $T_s$ , que é absolutamente contínua em relação à medida de Lebesgue. Pianigiani (1980), usando transformação de primeiro retorno, estabelece a existência da medida  $\mu_s$  para a transformação  $T_s$ . A medida invariante  $\mu_s$  é uma medida Sinai-Ruelle-Bowen, isto é, para quase todo  $x \in [0,1]$ , com respeito à medida de Lebesgue, vale a convergência fraca

$$\frac{1}{n} \sum_{i=0}^{n-1} \delta_{T_s^i(x)} \rightarrow \mu_s,$$

onde  $\delta_y$  denota a medida delta de Dirac em  $y$ .

- Estimativas da densidade de  $\mu_s$ . Thaler (1980), usando as propriedades da transformação de Manneville-Pomeau, mostrou que, para a medida  $d\mu_s(x) \equiv h_s(x) dx$ , existe  $s \in (0, 1)$  tal que

$$h_s(x) \approx x^{-s}, \text{ para todo } x \in (0, 1).$$

Esta medida  $d\mu_s(x)$  é “*mixing*” para  $T_s : [0, 1] \rightarrow [0, 1]$  (ver Young (1999) e Fisher e Lopes (2001)).

O sistema dinâmico  $([0, 1], T_s, \mu_s)$  é *exato*, isto é,

$$\lim_{n \rightarrow \infty} \mu_s(T_s^n(A)) = 1,$$

para todo conjunto mensurável  $A$  com  $\mu_s(A) > 0$ , implicando assim *ergodicidade* e *mixing*. Referimos o leitor a Maes et al. (1999) para maiores informações sobre o sistema dinâmico acima.

Quando  $s < 1$ ,  $d\mu_s(x) = h_s(x)dx$  é uma medida de probabilidade. No entanto, quando  $s \geq 1$ ,  $d\mu_s(x) = h_s(x)dx$  é uma medida invariante infinita (ver Fisher e Lopes (2001)).

A partir da transformação Manneville-Pomeau, dada pela expressão (4.1), vamos definir o processo estocástico  $\{X_t\}_{t \in \mathbb{N}}$  associado a esta transformação.

**DEFINIÇÃO 4.2:** Sejam  $\varphi : [0, 1] \rightarrow \mathbb{R}$  uma função integrável em relação a  $\mu_s$  e  $T_s(\cdot)$  a transformação Manneville-Pomeau, dada pela expressão (4.1). Fixada  $\varphi$ , o *processo estocástico Manneville-Pomeau*  $\{X_t\}_{t \in \mathbb{N}}$  é dado por

$$X_t = (\varphi \circ T_s^t)X_0 = \varphi(T_s^t(X_0)) = \varphi(T_s(X_{t-1})) = (\varphi \circ T_s)(X_{t-1}), \quad (4.2)$$

para todo  $t \in \mathbb{N}$ , onde  $X_0$  é uma variável aleatória distribuída de acordo com a medida  $\mu_s$ . Ou seja, o processo Manneville-Pomeau  $\{X_t\}_{t \in \mathbb{N}}$  é obtido pela aplicação da função  $\varphi$  às iterações da transformação  $T_s$ , mais explicitamente,  $X_t = \varphi \circ T_s^t$ , para  $s$  fixo e  $t \in \mathbb{N}$ .

### Observações:

a) Aqui vamos considerar apenas o caso em que as funções  $\varphi$  são obtidas a partir de funções indicadoras, isto é,  $\varphi \equiv \mathbb{I}_A - \mu_s(A)$ , ou  $\varphi \equiv \mathbb{I}_A$  onde  $A$  é um intervalo em  $[0, 1]$ .

b) Dado um valor inicial  $x_0$  da variável aleatória  $X_0$  no processo  $\{X_t\}_{t \in \mathbb{N}}$ , obtemos o caminho amostral  $w = (\varphi(x_0), \varphi(T_s(x_0)), \varphi(T_s^2(x_0)), \dots, \varphi(T_s^n(x_0)), \dots) \in \Omega = [0, 1]^{\mathbb{N}}$ . O processo estocástico  $\{X_t\}_{t \in \mathbb{N}}$ , definido pela expressão (4.2), é *estacionário* pois a transformação  $T_s$  deixa invariante a medida  $\mu_s$ , absolutamente contínua com respeito à medida de Lebesgue. Mais do que isto,  $\{X_t\}_{t \in \mathbb{N}}$  é *ergódico* pois  $\mu_s$  é *ergódica* para  $T_s$  (ver Durrett (1996)).

O maior problema na análise das propriedades ergódicas e na taxa de mixing do processo  $\{X_t\}_{t \in \mathbb{N}}$  é que a transformação  $T_s$  não é uniformemente expansiva (ver Pollicott e Yuri (1998) e Parry e Pollicott (1990)), devido ao ponto fixo indiferente desta transformação. Vamos analisar aqui o processo estocástico estacionário  $X_t$  associado a  $T_s$  e  $\varphi$  sobre  $(\Omega, \mathcal{A}, \mathbb{P})$ , onde  $\mathbb{P}$  é definida a partir de  $\mu_s$ , por exemplo,

$$\mathbb{P}(\{X_1 \in A_1, X_3 \in A_3\}) = \mu_s(x \in [0, 1] \mid \varphi(T_s(x)) \in A_1, \varphi(T_s^3(x)) \in A_3),$$

onde  $A_1$  e  $A_3$  são subconjuntos de  $[0, 1]$  na  $\sigma$ -álgebra de Borel.

A determinação da taxa de “mixing” (ou, em outras palavras, a velocidade do decaimento da função de autocorrelação) para transformações do tipo Manneville-Pomeau foi objeto de atenção de diversos pesquisadores. Este problema foi estudado por Lopes (1993), Isola (1995), Fisher e Lopes (2001), Hu (1998), Liverani et al. (1998) e Maes et al. (1999). É conhecido que transformações do tipo Manneville-Pomeau  $T_s$  possuem decaimento polinomial para a função de autocorrelação, isto é, para funções  $f$  e  $g$  (funções contínuas Hölder)

$$|\rho_X(k)| = \left| \int f(x)g(T_s^k(x))d\mu_s(x) - \int f(x)d\mu_s(x) \int g(x)d\mu_s(x) \right| \approx k^{1-\frac{1}{s}},$$

para  $k \in \mathbb{N}$ .

Ressaltamos ainda que uma aplicação da propriedade de decaimento polinomial da função de autocorrelação na transformação de Manneville-Pomeau pode ser encontrada no artigo Zebrowski (2000). Neste artigo, a propriedade de longa dependência, que se verifica nos batimentos cardíacos de pacientes com arritmia, é relacionada com a intermitência apresentada pelas transformações de Manneville-Pomeau.

Vamos assumir, sem perda de generalidade, que  $\varphi = \mathbb{I}_A - \mu_s(A)$  é tal que

$$\mathbb{E}_{\mu_s}(X_t) = \int \varphi(x)d\mu_s(x) = 0,$$

isto é, a esperança do processo estacionário é zero.

Observe que, neste caso, o processo  $\{X_t\}_{t \in \mathbb{N}}$ , dado pela expressão (4.2), tem função de autocovariância dada por

$$\gamma_X(k) = \mathbb{E}_{\mu_s}(X_0 X_k) - \mathbb{E}_{\mu_s}(X_0)\mathbb{E}_{\mu_s}(X_k) = \int \varphi(x)(\varphi \circ T_s^k)(x)d\mu_s(x), \quad (4.3)$$

para todo  $k \in \mathbb{N}$ .

**Observações:**

a) O processo  $\{X_t\}_{t \in \mathbb{N}}$ , dado pela expressão (4.2), pode ser definido para  $t \in \mathbb{Z}$  tomando-se uma transformação bijetiva  $F_s : [0, 1] \times [0, 1] \rightarrow [0, 1] \times [0, 1]$ , chamada de *extensão natural* (ver Lopes e Lopes (1998)), dada por

$$F_s(x, y) = (T_s(x), G(x, y)), \text{ para todo } (x, y) \in [0, 1] \times [0, 1].$$

Tomando  $\varphi(x, y) = \varphi(x)$ , então

$$\varphi(F_s^t(x, y)) = \varphi(T_s^t(x))$$

e podemos considerar assim que  $t \in \mathbb{Z}$ .

Uma outra maneira canônica de passar de processos  $\{X_t\}_{t \in \mathbb{N}}$  para processos  $\{Y_t\}_{t \in \mathbb{Z}}$  tais que  $\rho_X(t) = \rho_Y(t) = \rho_Y(-t)$ , para  $t \in \mathbb{N}$ , é apresentada através do Teorema 1.2, Seção 6.1, página 337, em Durrett (1996).

b) Para o processo  $\{X_t\}_{t \in \mathbb{N}}$ , descrito na Definição 4.2, consideraremos a forma da função densidade espectral dada em termos de cossenos através da expressão (2.2).

Consideremos  $\frac{1}{2} < s < 1$ . Do trabalho de Young (1999), para qualquer função  $\varphi$  Hölder, tal que  $\int \varphi(x) d\mu_s(x) = 0$ , existe uma constante positiva  $C_2$  tal que, para  $k$  positivo

$$|\mathbb{E}_{\mu_s}(X_0 X_k)| = \left| \int \varphi(x) (\varphi \circ T_s^k)(x) d\mu_s(x) \right| < C_2 k^{1-\frac{1}{s}}, \text{ para todo } k \in \mathbb{N}.$$

Apresentamos, a seguir, a definição de  $T_\gamma$ , a versão linear por partes da transformação Manneville-Pomeau.

**DEFINIÇÃO 4.3:** Seja  $\zeta(\gamma) = \sum_{n \geq 1} n^{-\gamma}$ , a função zeta de Riemann. Sejam os intervalos

$$M_0 = \left(1 - \frac{1}{\zeta(\gamma)}, 1\right) \text{ e } M_k = \left(1 - \frac{1}{\zeta(\gamma)} \sum_{n=1}^{k-1} n^{-\gamma}, 1 - \frac{1}{\zeta(\gamma)} \sum_{n=1}^k n^{-\gamma}\right),$$

para  $k \geq 1$ . Seja  $\gamma > 2$ . Definimos a transformação linear por partes  $T_\gamma : [0, 1] \rightarrow [0, 1]$  tal que, sobre os intervalos  $M_k$ ,  $k \geq 1$ ,  $T_\gamma$  tem inclinação  $((k+1)k^{-1})^\gamma$  e sobre o intervalo  $M_0$  tem inclinação  $\zeta(\gamma)$ . Assumindo que os ramos

$$T_\gamma|_{(0, 1 - \frac{1}{\zeta(\gamma)})} \text{ e } T_\gamma|_{(1 - \frac{1}{\zeta(\gamma)}, 1)}$$

são contínuos, então a transformação  $T_\gamma$  fica definida de maneira única (ver Figura 4.1.1 (b)). Esta transformação  $T_\gamma$  é chamada *aproximação linear por partes da transformação Manneville-Pomeau*.

Ambos os modelos, isto é, transformação Manneville-Pomeau  $T_s$  e versão linear por partes da transformação Manneville-Pomeau  $T_\gamma$ , são muito similares em termos da velocidade do decaimento da função de autocorrelação.

A transformação linear por partes  $T_\gamma$  é indexada por um parâmetro  $\gamma > 2$ . O regime de longa dependência fica então na região  $2 < \gamma < 3$  como veremos mais adiante. A transformação  $T_\gamma$  possui, basicamente, as mesmas propriedades de  $T_s$  (ver Fisher e Lopes (2001)). A relação do correspondente  $s$  ao valor de  $\gamma$  é dada por  $\gamma = 1 + \frac{1}{s}$ , no sentido de determinarem o mesmo  $u$  na expressão (2.6). Na descrição abaixo, sobre a equivalência dos parâmetros, apresentamos mais detalhes sobre esta relação.

**Figura 4.1.1:** (a) transformação Manneville-Pomeau  $T_s$  e (b) sua aproximação linear por partes  $T_\gamma$ .

Fisher e Lopes (2001) mostram o assintótico exato da velocidade de decaimento da função de autocorrelação para a transformação  $T_\gamma$ , ou seja, para o processo  $X_t = \varphi \circ T_\gamma^t$ , onde  $\varphi = \mathbb{I}_{M_0}$  e  $X_0$  é a variável aleatória distribuída de acordo com  $\mu_\gamma$ , a medida invariante para  $T_\gamma$ , absolutamente contínua com relação à medida de Lebesgue (ver Teorema 4.1).

O Teorema 4.1 abaixo revela que o processo  $\{X_t\}_{t \in \mathbb{N}}$  é de longa dependência no sentido da Definição 2.15.

**TEOREMA 4.1:** *Seja a função  $\varphi$  dada por  $\varphi(x) = \mathbb{I}_{M_0}(x) - \mu_\gamma(M_0)$ , seja  $\{X_t\}_{t \in \mathbb{N}}$  o processo dado pela expressão (4.2) com  $T_\gamma$  no lugar de  $T_s$  e  $\gamma \in (2, \infty)$  fixado. Então, a função de autocovariância de ordem  $k$  do processo  $\{X_t\}_{t \in \mathbb{N}}$ , dada por*

$$\gamma_X(k) \equiv \mathbb{E}_{\mu_\gamma}(X_0 X_k) = \int \varphi(x)(\varphi \circ T_\gamma^k)(x) d\mu_\gamma(x),$$

*é uma função monótona decrescente em  $k$  tal que*

$$\mathbb{E}_{\mu_\gamma}(X_0 X_k) \approx k^{2-\gamma}, \text{ para todo } k \in \mathbb{N}. \quad (4.4)$$

O processo  $\{X_t\}_{t \in \mathbb{N}}$ , definido pela expressão (4.2), é um processo estacionário e ergódico mas não é Gaussiano e nem pode ser descrito através de um modelo do tipo  $ARFIMA(p, d, q)$ . No entanto, conforme a expressão (4.4), ele possui a característica de longa dependência quando  $\gamma \in (2, 3)$ .

A seguir, estabelecemos a correspondência da velocidade de decaimento da função de autocorrelação entre os processos  $ARFIMA(p, d, q)$  e os processos  $\{X_t\}_{t \in \mathbb{N}}$ , obtidos através de  $\varphi \circ T_s^t$  e  $\varphi \circ T_\gamma^t$ .

**Equivalência dos Parâmetros:** Sejam  $d = 1 - \frac{1}{2s}$  o parâmetro de diferenciação de um processo  $ARFIMA(p, d, q)$ ,  $\gamma = 1 + \frac{1}{s}$ , onde  $s$  é o parâmetro da transformação Manneville-Pomeau  $T_s$  e  $\gamma$  é o parâmetro da aproximação linear por partes  $T_\gamma$ . Então, a velocidade de convergência da função de autocorrelação pode ser descrita por

- (i)  $1 < \gamma < 2 \Leftrightarrow s \in (1, +\infty) \Leftrightarrow d \in (\frac{1}{2}, 1) \Leftrightarrow d\mu_s(x) = h(x)dx$  é uma medida infinita.

Neste caso, o processo  $\{X_t\}_{t \in \mathbb{Z}}$  é não estacionário com propriedade de longa dependência.

- (ii)  $2 < \gamma < 3 \Leftrightarrow \frac{1}{2} < s < 1 \Leftrightarrow d \in (0, \frac{1}{2}) \Leftrightarrow d\mu_s(x) = h(x)dx$  é uma medida de probabilidade.

Neste caso, o processo  $\{X_t\}_{t \in \mathbb{Z}}$  é estacionário, ergódico e com propriedade de longa dependência.

- (iii)  $3 < \gamma < 4 \Leftrightarrow \frac{1}{3} < s < \frac{1}{2} \Leftrightarrow d \in (-\frac{1}{2}, 0) \Leftrightarrow d\mu_s(x) = h(x)dx$  é uma medida de probabilidade.

Neste caso, o processo  $\{X_t\}_{t \in \mathbb{Z}}$  é estacionário, ergódico e com propriedade de dependência intermediária. Observamos que as propriedades obtidas na Sub-seção 4.3.2 continuam válidas para  $d \in (-\infty, 0)$ .

Note que

$$\begin{aligned} 1 < \gamma < 2 &\Leftrightarrow 1 < 1 + \frac{1}{s} < 2 \Leftrightarrow 0 < \frac{1}{s} < 1 \Leftrightarrow s > 1, \\ 2 < \gamma < 3 &\Leftrightarrow 2 < 1 + \frac{1}{s} < 3 \Leftrightarrow 1 < \frac{1}{s} < 2 \Leftrightarrow \frac{1}{2} < s < 1, \\ 3 < \gamma < 4 &\Leftrightarrow 3 < 1 + \frac{1}{s} < 4 \Leftrightarrow 2 < \frac{1}{s} < 3 \Leftrightarrow \frac{1}{3} < s < \frac{1}{2}. \end{aligned}$$

Temos, pela relação entre  $s$  e  $\gamma$ , que

$1 < \gamma < 2 \Leftrightarrow s \in (1, +\infty) \Leftrightarrow d\mu_s(x) = h(x)dx$  é uma medida infinita

e

$2 < \gamma < 4 \Leftrightarrow \frac{1}{3} < s < 1 \Leftrightarrow d\mu_s(x) = h(x)dx$  é uma medida de probabilidade.

Para confirmar a relação entre  $\gamma$  e  $d$ , observe que do item (d) do Teorema 3.3, temos que  $\rho_X(k) \sim k^{2d-1}$ . Então,

$$2d - 1 = 2 - \gamma.$$

Portanto,

$$\begin{aligned} d = \frac{3 - \gamma}{2} &\Leftrightarrow d = \frac{3}{2} - \frac{1}{2}\gamma \\ &\Leftrightarrow d = \frac{3}{2} - \frac{1}{2} \left(1 + \frac{1}{s}\right) \\ &\Leftrightarrow d = \frac{3}{2} - \frac{1}{2} - \frac{1}{2s} \\ &\Leftrightarrow d = 1 - \frac{1}{2s}. \end{aligned}$$

Temos que

$$1 < \gamma < 2 \Leftrightarrow d = 1 - \frac{1}{2s} \text{ e } s > 1 \Leftrightarrow \frac{1}{2} < d < 1,$$

$$2 < \gamma < 3 \Leftrightarrow d = 1 - \frac{1}{2s} \text{ e } \frac{1}{2} < s < 1 \Leftrightarrow 0 < d < \frac{1}{2}$$

e

$$3 < \gamma < 4 \Leftrightarrow d = 1 - \frac{1}{2s} \text{ e } \frac{1}{3} < s < \frac{1}{2} \Leftrightarrow -\frac{1}{2} < d < 0.$$

A equivalência dos parâmetros será utilizada na Seção 4.2 como motivação para obter um estimador para o parâmetro  $s$  de maneira análoga ao que foi feito para modelos *ARFIMA*.

Para esclarecer melhor o leitor consideremos o exemplo abaixo:

**Exemplo 4.1:** Consideremos os processos  $\{X_t\}_{t \in \mathbb{N}}$  obtidos, respectivamente, através de  $\varphi \circ T_s^t$ , com  $s = \frac{2}{3}$  e  $\varphi \circ T_\gamma^t$ , com  $\gamma = 1 + \frac{3}{2}$ . Seja  $\varphi = \mathbb{I}_A - \mu_s(A)$ , com  $A = (0, 1; 0, 9)$ . Então, pelo Teorema 4.1, a função de autocorrelação de ordem  $k$  para os processos  $X_t = \varphi \circ T_s^t$  e  $X_t = \varphi \circ T_\gamma^t$ , para todo  $t \in \mathbb{N}$ , é tal que

$$\rho_X(k) \approx k^{-\frac{1}{2}},$$

conforme Definição 2.15. Observamos que a função de autocorrelação de ordem  $k$  do processo  $ARFIMA(0, d, 0)$ , quando  $d = \frac{1}{4}$ , possui também decaimento da ordem de  $k^{-\frac{1}{2}}$ .

**COROLÁRIO 4.1:** *Seja  $\{X_t\}_{t \in \mathbb{N}}$  o processo obtido a partir da transformação  $T_\gamma$ , dada na Definição 4.3. Seja  $\{X_t\}_{t=0}^{n-1}$  uma série temporal obtida a partir do processo  $X_t = \mathbb{I}_A \circ T_\gamma^t$ , onde  $A = (0, 1; 0, 9)$ . Considere a variável aleatória  $N_n$ , que define o número de 1's da série temporal  $\{X_t\}_{t=0}^{n-1}$ . Então,*

$$\text{Var}(N_n) \approx n^{4-\gamma}.$$

**Prova:** Observe que, como

$$N_n = \text{total de números 1's da série temporal } \{X_t\}_{t=0}^{n-1} = \sum_{i=0}^{n-1} X_i = S_n,$$

então,  $N_n$  corresponde à variável aleatória  $S_n = \sum_{i=0}^{n-1} X_i$ . O resultado segue da Proposição 3.4. □

Outro exemplo de processo com decaimento lento da função de autocorrelação é aquele obtido a partir de cadeias de Markov sobre o conjunto de estados  $\mathbb{N}$ . Este modelo é basicamente igual (em termos de probabilidade) à transformação  $T_\gamma$ .

**DEFINIÇÃO 4.4:** Considere a cadeia de Markov  $\mathbf{P}$  dada pela matriz infinita  $\mathbf{P} = (\mathbb{P}(i, j))_{i, j \in \mathbb{N}}$  (ver Lopes (1993) e Feller (1949)) onde as probabilidades de transição são dadas por

$$\begin{aligned} \mathbb{P}(n, n-1) &= 1, \text{ para todo } n \in \mathbb{N} - \{0\}, \\ \mathbb{P}(n, j) &= 0, \text{ para } j \neq n-1, \\ \mathbb{P}(0, n) &= \frac{(n+1)^{-\gamma}}{\zeta(\gamma)}, \end{aligned}$$

onde  $\zeta(\gamma)$  é a função zeta de Riemann e  $\gamma \in (2, 3)$ .

Seja  $\{Z_t\}_{t \in \mathbb{N}}$  o processo estocástico estacionário de Markov obtido a partir da matriz  $\mathbf{P}$  acima e da distribuição inicial  $\pi_0$  estacionária. Seja  $\mathbb{I}_0$  a função indicadora do conjunto  $A = \{0\}$  sobre o conjunto  $\mathbb{N}$ . Seja agora  $X_t = 1 - \mathbb{I}_0(Z_t)$ , para todo  $t \in \mathbb{N}$ . Então,  $\{X_t\}_{t \in \mathbb{N}}$  é um processo cujas variáveis aleatórias assumem apenas os valores 0 e 1.

**Exemplo 4.2:** Dado um caminho amostral  $w \in \mathbb{N}^{\mathbb{N}}$  do processo  $\{Z_t\}_{t \in \mathbb{N}}$ , por exemplo,  $w = \{032109876543210543210 \dots\}$ , associamos outro caminho referente ao processo  $\{X_t\}_{t \in \mathbb{N}}$ , dado por

$$\tilde{w} = \{0 \underbrace{111}_3 0 \underbrace{111111111}_9 0 \underbrace{1111110}_5 \dots\}$$

de  $\{X_t\}_{t \in \mathbb{N}}$ , ou seja, aplicamos uma mudança de coordenadas,  $Z_t \rightarrow X_t$ , associando números naturais a blocos de 1's intercalados por 0, o que mantém a estrutura do processo.

Pode-se mostrar que as seqüências de símbolos em  $\mathbb{N}$  obtidos de acordo com a probabilidade de transição  $\mathbb{P}(0, n) = \frac{(n+1)^{-\gamma}}{\zeta(\gamma)}$  acima têm as mesmas propriedades que aquelas do processo estocástico  $\{X_t\}_{t \in \mathbb{N}}$  obtido a partir de  $X_t = \mathbb{I}_{M_0} \circ T_\gamma^t$  (outras propriedades deste modelo podem ser encontradas em Feller (1949) e Lopes (1993)). Mais exatamente, o processo  $\{X_t\}_{t \in \mathbb{N}}$  obtido de  $\{Z_t\}_{t \in \mathbb{N}}$  é o mesmo processo originado por  $\mathbb{I}_{M_0} \circ T_\gamma^t = X_t$ , onde  $\mu_\gamma$  é uma medida invariante para  $T_\gamma$ , absolutamente contínua com relação à medida de Lebesgue e  $M_0 = \left(1 - \frac{1}{\zeta(\gamma)}, 1\right)$ .

Esta cadeia de Markov,  $\{Z_t\}_{t \in \mathbb{N}}$  ou  $\{X_t\}_{t \in \mathbb{N}}$ , é a mesma considerada em Feller (1949) que mostra que, para  $\gamma > 3$ , vale o teorema central do limite (convergindo para a distribuição Gaussiana) para a variável aleatória  $N_n$ , que descreve o número de ocorrências do símbolo 0. Para  $2 < \gamma < 3$ , o teorema do limite central é verdadeiro, mas converge para uma distribuição de Lévy, também denominada distribuição estável, com parâmetro  $\alpha = \gamma - 1$ , apresentada na definição abaixo (para maiores detalhes sobre estas distribuições ver Samorodnitsky e Taqqu (1994)).

**DEFINIÇÃO 4.5:** A *distribuição estável* de uma variável aleatória  $X$ , com parâmetro  $\alpha$ , denotada por  $G_\alpha(x)$ , é aquela cuja função característica é dada por

$$\phi_X^\alpha(z) = \exp \left\{ - \left[ |z|^\alpha \left( \cos\left(\frac{\pi\alpha}{2}\right) - i \operatorname{sen}\left(\frac{\pi\alpha}{2}\right) \operatorname{sgn} z \right) \right] \Gamma(1 - \alpha) \right\},$$

onde  $\Gamma(\cdot)$  é a função Gama e  $\operatorname{sgn}(\cdot)$  é a função sinal.

Descrevemos acima, nesta seção, a relação entre o decaimento hiperbólico da função de autocorrelação nos vários casos: *ARFIMA*, processo Manneville-Pomeau obtido a partir de  $T_s$ , de  $T_\gamma$  e da cadeia de Markov  $\mathbf{P}$ . A partir de agora vamos considerar apenas o processo dado pela expressão (4.2) definido a partir de  $T_s$ .

O objetivo da próxima seção é estimar o parâmetro  $s$  utilizando diferentes métodos de estimação. Os resultados das simulações estão na Seção 4.3.

**Observação:** Na Seção 4.3 utilizamos a transformação Manneville-Pomeau para as simulações tendo em vista a facilidade em definir a transformação  $T_s(\cdot)$  em uma forma fechada, isto é,  $T_s(x) = x + x^{s+1} \pmod{1}$ . Os resultados são igualmente válidos para estimar  $\gamma$  para estimar séries temporais de dados obtidos a partir da cadeia de Markov  $\mathbf{P}$  ou mesmo de  $\varphi \circ T_\gamma^t = X_t$ .

## 4.2 Métodos de Estimação do Parâmetro $s$

O principal objetivo desta seção é estimar a transformação  $T_s$ , ou equivalentemente, estimar o parâmetro  $\frac{1}{2} < s < 1$ , a partir de uma série temporal  $\{X_t\}_{t=0}^{n-1}$  obtida do processo  $\{X_t\}_{t \in \mathbb{N}}$  definido através da expressão (4.2).

Queremos comparar o método amplamente utilizado pelos físicos para estimar  $s$  (ver o primeiro método a seguir) com outros métodos. Desejamos saber se o método proposto pelos físicos (ver Schuster (1984)) é satisfatório ou se podemos obter outro método que melhor estima o parâmetro  $s$ . O processo  $\{X_t\}_{t \in \mathbb{N}}$ , definido através da expressão (4.2), onde  $X_t = \mathbb{I}_A \circ T_s^t$  com  $A = (0, 1; 0, 9)$ , é estacionário e ergódico (ver Lopes e Lopes (1998)). A ergodicidade do sistema justifica o uso da função de autocovariância amostral como base para a estimação de  $s$ .

Suponha que existe  $c$  tal que  $f_X(w) \approx w^c$ , para  $w$  próximo da frequência zero. A idéia básica para a estimação do parâmetro  $s$ , no caso de longa dependência, é primeiro calcular aproximadamente  $f_X(\cdot)$  através da função periodograma  $I(\cdot)$ , dada pela expressão (2.3). Após, colocar  $f_X(w)$  em escala logarítmica para  $w$  próximo da origem e então obtemos

$$\frac{\ln f_X(w)}{\ln w} \approx c.$$

A partir de  $c$  se descobre  $s$ .

Veremos, a seguir, a estimação do parâmetro  $s$ . Os diferentes métodos a serem descritos nesta seção e, posteriormente, utilizados na Seção 4.3 são baseados nas funções periodograma e periodograma suavizado (ver Definições 2.13 e 2.14); na variância da soma parcial (proposto a partir da Proposição 3.4); no gráfico do logaritmo da variância de  $\bar{X}_n$  e, ainda, na teoria de wavelets.

### Estimador *Perio*

Este método é baseado na função periodograma de uma série temporal e é amplamente utilizado pelos físicos (ver Schuster (1984)).

O estimador de  $s$  é obtido pelo método de mínimos quadrados a partir da regressão linear de  $y_1, y_2, \dots, y_{g(n)}$  em  $x_1, x_2, \dots, x_{g(n)}$ , onde  $y_j = \ln I(w_j)$ ,  $x_j = \ln j$  e  $g(n) = n^{0.5}$ , onde  $I(\cdot)$  é a função periodograma dada pela expressão (2.3) e  $w_j$  são as freqüências de Fourier. Seja  $c$  o coeficiente de inclinação da reta de regressão na escala logarítmica. O coeficiente  $c$  irá permitir estimar  $s$  através da igualdade

$$s = \frac{1}{c + 2},$$

pois, pela Conclusão 2 do Anexo A,

$$f_X(w) \approx w^{\frac{1}{s}-2}, \text{ para } w \text{ próximo da origem.}$$

Logo,  $\hat{c} = \frac{1}{\hat{s}} - 2$  e assim  $\hat{s} = (c + 2)^{-1}$ . Denotaremos este estimador por *Perio*.

Assim como nos processos *ARFIMA* podemos utilizar a função periodograma suavizado no lugar da função periodograma obtendo, desta forma, outros estimadores para o parâmetro  $s$ . Consideremos, nos métodos a seguir, novamente a janela espectral de Parzen e a função “cosine bell” dadas na Definição 2.14.

### Estimador *Parzen*

Consideremos a função periodograma suavizado com a janela de Parzen. O estimador de  $s$  é obtido analogamente ao método descrito acima, substituindo  $I(\cdot)$  por  $f_s(\cdot)$ , dada pela expressão (2.4), com a janela de Parzen e será denotado por *Parzen*. Neste método, também consideramos  $g(n) = n^{0.5}$ . Para o ponto de truncamento da janela de Parzen consideramos  $m = n^{0.9}$ .

### Estimador *Cos*

Neste método, análogo ao estimador *Perio*, utilizamos a função “cosine bell” como janela espectral para suavizar o periodograma (ver Definição 2.14). Novamente, pelo método de regressão, obtemos um estimador para  $s$ , denotado aqui por *Cos*. Neste método, consideramos diferentes limites para  $g(n) = n^{\alpha_i}$ : utilizamos  $\alpha_i \in \{0, 5; 0, 7\}$  e denotamos o estimador por  $Cos(i)$ ,  $i = 1, 2$ .

**Observação:** Os métodos *Perio*, *Parzen* e *Cos*, acima definidos, são análogos aos métodos semiparamétricos descritos na Seção 3.3 do Capítulo 3 para o modelo *ARFIMA*. Devemos ressaltar que não temos uma expressão exata

para a função densidade espectral  $f_X(\cdot)$  no caso de processos Manneville-Pomeau.

### Estimador $Varm_p$

Este método, denotado por  $Varm_p$ , que tem natureza diferente dos estimadores acima, considera a ordem da variância de  $N_n$ , que é o número de 1's da série temporal  $\{X_t\}_{t=0}^{n-1}$ . Pelo Corolário 4.1 temos que

$$Var(N_n) \approx n^{4-\gamma} = n^{3-\frac{1}{s}}.$$

Se aplicarmos o logaritmo na expressão acima obtemos

$$Varm_p = \frac{1}{3 - \frac{\ln(Var(N_n))}{\ln(n)}} = \hat{s}.$$

**Observação:** Como no processo *ARFIMA* (veja o estimador  $V$  dado na Seção 3.3), vamos observar nos resultados das simulações que, para o processo obtido através das iterações da transformação de Manneville-Pomeau, dado pela expressão (4.2), este método fornece um estimador bastante viciado para o parâmetro  $s$ .

### Estimador $Vpmp$

Este estimador baseia-se também na variância de  $N_n$  de maneira análoga ao estimador VP, descrito na Seção 3.3 para modelos *ARFIMA*. Para o caso dos processos obtidos a partir das iterações da transformação de Manneville-Pomeau, basta considerar as relações entre  $N_n$  e  $S_n$  e entre  $s$  e  $d$ . Usaremos a notação  $Vpmp$  para este estimador.

### Estimador $Wmp$

Este método obtém estimativas para o parâmetro  $s$  a partir de uma adequação do estimador de wavelets, descrito na Seção 3.3, para o caso em que o processo é obtido a partir de iterações da transformação Manneville-Pomeau considerando a relação entre  $s$  e  $d$ . Será denotado aqui por  $Wmp$ .

Consideramos aqui as bases de wavelets Haar e Chapéu Mexicano, dadas pelas expressões (3.28) e (3.29), respectivamente, pois as mesmas apresentaram bons resultados no trabalho de Poli e Lopes (2002).

Este método foi também testado para o caso dos processos  $\{X_t\}_{t \in \mathbb{N}}$  obtidos a partir das iterações da transformação de Manneville-Pomeau quando  $s \geq 1$ . Este caso corresponde a medida invariante  $\mu_s$  não ser uma medida de probabilidade. Comparamos com as estimativas obtidas para o caso *ARFIMA* não estacionário testados por Poli e Lopes (2002) considerando  $l = 4$  como sendo o ponto de truncamento inferior.

Neste trabalho, utilizamos  $l = 4$  como sendo o ponto de truncamento inferior. Então, a partir da expressão (3.4) e considerando a relação  $d = 1 - \frac{1}{2s}$ , o estimador de  $s$  baseado em wavelets é dado por

$$Wmp = \frac{\sum_{j=4}^{m-1} x_j^2}{2 \left[ \left( \sum_{j=4}^{m-1} x_j \ln \hat{R}(j) \right) + \left( \sum_{j=4}^{m-1} x_j^2 \right) \right]},$$

onde  $x_j$  é definido por

$$x_j = \ln 2^{-2j} - \frac{1}{m-4} \sum_{j=4}^{m-1} \ln 2^{-2j}$$

e  $\hat{R}(j)$  é obtido a partir da variância amostral dos coeficientes de wavelets.

### 4.3 Simulação e Resultados

Seja  $\{X_t\}_{t \in \mathbb{N}}$  o processo obtido a partir de iterações da transformação Manneville-Pomeau  $T_s$ , dado pela expressão (4.2). Fixamos  $\varphi = \mathbb{I}_A$ , onde  $A = (0, 1; 0, 9)$  de modo que  $X_t = \mathbb{I}_A \circ T_s^t$ . Escolhemos, ao acaso, um valor  $x_0$  de acordo com a distribuição uniforme (isto resulta ser o mesmo que escolher  $x_0$  ao acaso de acordo com a distribuição  $\mu_s$  porque os conjuntos de medida nula são os mesmos). Seja  $\{X_t\}_{t=0}^{n-1}$  uma série temporal com  $n$  observações, obtida a partir do processo  $\{X_t\}_{t \in \mathbb{N}}$ . Então, esta série temporal é dada por

$$X_t = \mathbb{I}_A(T_s^t(x_0)), \text{ para todo } t = 0, \dots, n-1. \quad (4.5)$$

Os resultados foram obtidos através de simulações utilizando programas na linguagem FORTRAN e subrotinas do IMSL. Os procedimentos de estimação para processos de longa dependência envolvem uma quantidade grande de dados e que tomam bastante tempo de processamento computacional.

### 4.3.1 Longa Dependência

A Figura 4.3.1 apresenta as funções de autocorrelação amostral e periodograma de uma série temporal, obtida a partir da expressão (4.5), com 10.000 observações e  $s = 0,8$ . Podemos observar a propriedade de longa dependência tanto no domínio do tempo como no da frequência.

**Figura 4.3.1:** (a) função de autocorrelação amostral e (b) função periodograma de uma série temporal de tamanho amostral 10.000 obtida a partir de iterações da transformação  $T_s$  com  $s = 0,8$ .

Apresentamos, a seguir, algumas tabelas com resultados da estimação do parâmetro  $s$  para diferentes tamanhos amostrais.

Tabela 4.3.1: Estimação para os parâmetros  $s = 0,6$  e  $s = 0,65$ .

$s$	$n$	Método	média( $\hat{s}$ )	sd( $\hat{s}$ )	eqm( $\hat{s}$ )
0,6	10.000	<i>Perio</i>	0,6545	0,1394	0,0223
		<i>Parzen</i>	0,6313	0,1125	0,0136
		<i>Cos(1)</i>	0,5531	0,0572	0,0054
		<i>Cos(2)</i>	0,5993	0,0220	<b>0,0005</b>
		<i>Varmp</i>	0,5309	0,0396	0,0063
		<i>Vpmp</i>	0,5598	0,0718	0,0067
	20.000	<i>Perio</i>	0,6364	0,1094	0,0130
		<i>Parzen</i>	0,6147	0,0086	0,0070
		<i>Cos(1)</i>	0,5488	0,0535	0,0054
		<i>Cos(2)</i>	0,5979	0,0264	<b>0,0007</b>
		<i>Varmp</i>	0,5241	0,0303	0,0067
		<i>Vpmp</i>	0,5513	0,0583	0,0057
	30.000	<i>Perio</i>	0,6004	0,1051	0,0110
		<i>Parzen</i>	0,5865	0,0736	0,0056
		<i>Cos(1)</i>	0,5275	0,0508	0,0078
		<i>Cos(2)</i>	0,5933	0,0316	<b>0,0010</b>
		<i>Varmp</i>	0,5204	0,0264	0,0070
		<i>Vpmp</i>	0,5144	0,0608	0,0110
0,65	10.000	<i>Perio</i>	0,7539	0,1518	0,0337
		<i>Parzen</i>	0,7107	0,0107	0,0151
		<i>Cos(1)</i>	0,6145	0,0614	0,0050
		<i>Cos(2)</i>	0,6129	0,0198	<b>0,0017</b>
		<i>Varmp</i>	0,5293	0,0332	0,0156
		<i>Vpmp</i>	0,5461	0,0763	0,0166
	20.000	<i>Perio</i>	0,7113	0,0779	0,0098
		<i>Parzen</i>	0,6927	0,0706	0,0068
		<i>Cos(1)</i>	0,6035	0,0472	0,0044
		<i>Cos(2)</i>	0,6076	0,0181	<b>0,0021</b>
		<i>Varmp</i>	0,5251	0,0246	0,0162
		<i>Vpmp</i>	0,5257	0,0630	0,0194
	30.000	<i>Perio</i>	0,6806	0,0445	0,0029
		<i>Parzen</i>	0,6910	0,0392	0,0032
		<i>Cos(1)</i>	0,6141	0,0552	0,0043
		<i>Cos(2)</i>	0,6090	0,0147	<b>0,0019</b>
		<i>Varmp</i>	0,5419	0,0262	0,0123
		<i>Vpmp</i>	0,5451	0,0562	0,0141

Foram calculados os valores empíricos da média, do desvio padrão (*sd*) e do erro quadrático médio (*eqm*) dos estimadores de *s*. O erro quadrático médio de menor valor está em negrito nas tabelas. As estimativas nas Tabelas 4.3.1 e 4.3.2 estão baseadas em 200 replicações. Para o estimador *Cos* foram utilizados dois diferentes valores para o limite  $g(n) = n^{\alpha_i}$ : *Cos*(1) refere-se a  $\alpha_1 = 0,5$  e *Cos*(2) refere-se a  $\alpha_2 = 0,7$ .

Observe na Tabela 4.3.1 que os estimadores *Varmp* e *Vpmp* são bastante viciados. O mesmo já tinha sido constatado para estes métodos quando o processo é um *ARFIMA*. O estimador *Cos*(2) forneceu as melhores estimativas tanto para  $s = 0,6$  como para  $s = 0,65$ , apresentando menor erro quadrático médio para todos os tamanhos amostrais analisados. Para o caso  $s = 0,8$  (ver Tabela 4.3.2), considerando o tamanho amostral 10.000, os melhores resultados foram obtidos através dos métodos *Perio* e *Parzen*. Observe que, para os demais tamanhos amostrais, todos os estimadores apresentaram vício alto.

**Tabela 4.3.2:** Estimação para o parâmetro  $s = 0,8$ .

<i>n</i>	Método	média( $\hat{s}$ )	sd( $\hat{s}$ )	eqm( $\hat{s}$ )
10.000	<i>Perio</i>	0,7773	0,1648	0,0275
	<i>Parzen</i>	0,7607	0,1444	<b>0,0222</b>
	<i>Cos</i> (1)	0,6286	0,2507	0,0919
	<i>Cos</i> (2)	0,6626	0,0822	0,0256
	<i>Varmp</i>	0,5472	0,0426	0,0657
	<i>Vpmp</i>	0,5781	0,0806	0,0557
20.000	<i>Perio</i>	0,6921	0,1220	<b>0,0264</b>
	<i>Parzen</i>	0,6740	0,1127	0,2849
	<i>Cos</i> (1)	0,5731	0,0699	0,0563
	<i>Cos</i> (2)	0,6434	0,0437	<b>0,0264</b>
	<i>Varmp</i>	0,5292	0,0434	0,0752
	<i>Vpmp</i>	0,5416	0,0848	0,0739
30.000	<i>Perio</i>	0,6559	0,1164	0,0342
	<i>Parzen</i>	0,6150	0,1044	0,0456
	<i>Cos</i> (1)	0,5335	0,1713	0,1002
	<i>Cos</i> (2)	0,6382	0,0337	<b>0,0273</b>
	<i>Varmp</i>	0,5354	0,0524	0,0727
	<i>Vpmp</i>	0,5434	0,0861	0,0732

Quando *s* se aproxima de 1, a série temporal  $\{X_t\}_{t=0}^{n-1}$  fica cada vez mais tempo no zero, não sendo possível detectar uma boa informação. Por esta razão as estimativas não ficam boas (ver Tabela 4.3.2). Os estimadores

$Varm\tilde{p}$  e  $Vpmp$  não são recomendados aqui, pois não fornecem boas estimativas, apresentando vício alto, assim como ocorre quando utilizamos estes dois métodos nos processos *ARFIMA*.

Apresentamos agora os resultados obtidos através do estimador  $Wmp$  considerando as bases de wavelets Haar e Chapéu Mexicano. Ressaltamos que para este estimador devemos ter o tamanho amostral como uma potência de 2. Os resultados nas Tabelas 4.3.3 e 4.3.4 a seguir são baseados em 50 replicações.

**Tabela 4.3.3:** Estimação para os parâmetros  $s = 0,65$  e  $s = 0,8$ .

$s$	$n$	$Base$	$média(\hat{s})$	$sd(\hat{s})$	$eqm(\hat{s})$
0,65	8.192	Haar	0,8531	0,0470	0,0434
		Chapéu Mex.	0,8022	0,0480	<b>0,0254</b>
	16.384	Haar	0,8311	0,0446	0,0347
		Chapéu Mex.	0,7882	0,0472	<b>0,0213</b>
	32.768	Haar	0,8283	0,0619	0,0355
		Chapéu Mex.	0,7864	0,0451	<b>0,0206</b>
0,8	8.192	Haar	0,9839	0,0619	0,0376
		Chapéu Mex.	0,8873	0,0670	<b>0,0120</b>
	16.384	Haar	0,9321	0,0659	0,0217
		Chapéu Mex.	0,8237	0,0675	<b>0,0050</b>
	32.768	Haar	0,8639	0,0915	0,0120
		Chapéu Mex.	0,7747	0,0464	<b>0,0027</b>

Podemos observar, através da Tabela 4.3.3, que as estimativas obtidas para  $s \in (0,5; 1,0)$ , a partir da base de wavelets Chapéu Mexicano, apresentam menores vício e erro quadrático médio quando comparadas com aquelas obtidas a partir da base de wavelets Haar. Resultados semelhantes foram obtidos por Poli e Lopes (2002) para processos *ARFIMA* estacionários com longa dependência.

A Tabela 4.3.4 apresenta resultados para o caso dos processos obtidos a partir das iterações da transformação de Manneville-Pomeau  $T_s$  quando  $s \geq 1$ , o que corresponde a medida invariante  $\mu_s$  não ser medida de probabilidade. Conforme os resultados desta tabela, onde  $s \in \{1,0; 1,1; 1,2; 1,3\}$  e tamanho amostral  $n = 32.768$ , as melhores estimativas são obtidas através da base de wavelets Haar. Note que, no caso em que  $s \geq 1$ , qualquer método baseado na função periodograma não faz sentido quando os processos são obtidos a partir das iterações da transformação de Manneville-Pomeau  $T_s$ .

Observe que para o parâmetro  $s = 0,65$  o método *Perio* funciona melhor que aquele baseado em wavelets,  $Wmp$ . Já para o parâmetro  $s = 0,8$  ocorre o contrário.

**Tabela 4.3.4:** Estimação para o parâmetro  $s \geq 1$ , quando  $n = 32.768$ .

$s$	<i>Base</i>	$média(\hat{s})$	$sd(\hat{s})$	$eqm(\hat{s})$
1,0	Haar	0,9461	0,1090	0,0145
	Chapéu Mex.	0,8931	0,1148	0,0243
1,1	Haar	1,0924	0,0589	0,0034
	Chapéu Mex.	0,9943	0,0461	0,0132
1,2	Haar	1,0825	0,0729	0,0190
	Chapéu Mex.	0,9642	0,0939	0,0642
1,3	Haar	1,1422	0,0638	0,0288
	Chapéu Mex.	1,0064	0,0703	0,0910

### 4.3.2 Dependência intermediária

Faremos agora uma breve análise para o caso de dependência intermediária em processos Manneville-Pomeau, dados pela expressão (4.2), embora este não seja o enfoque principal deste trabalho. Neste caso, temos também o decaimento polinomial da função de autocorrelação do processo  $\{X_t\}_{t \in \mathbb{N}}$ .

Para o modelo *ARFIMA* a dependência intermediária ocorre quando  $d \in (-0, 5; 0, 0)$  e existem análises para a estimação do parâmetro fracionário para este caso (ver Geweke e Porter-Hudak (1983), Robinson (1995), Reisen (1994) e Fox e Taqqu (1986)). Ressaltamos que os métodos de estimação *GPH*, *SPR*, *GPht*, *MGPH*, *FT*, etc., descritos na Seção 3.3, podem ser utilizados nesta situação. Nos processos Manneville-Pomeau a dependência intermediária corresponde a  $s \in (1/3, 1/2)$  e os métodos *Perio*, *Parzen* e *Cos* não podem ser utilizados pois estes métodos são descritos através do coeficiente de inclinação.

No caso de dependência intermediária nos processos Manneville-Pomeau, introduzimos aqui um método semelhante, embora diferente, ao estimador *Perio* descrito na Seção 4.2.

Seja  $\{X_t\}_{t \in \mathbb{N}}$  o processo obtido a partir da transformação Manneville-Pomeau  $T_s$ , dado pela expressão (4.2), isto é,  $X_t = \varphi \circ T_s^t$  onde  $\varphi = \mathbb{I}_A$ , com  $A = (0, 1; 0, 9)$ . Seja  $\{X_t\}_{t=0}^{n-1}$  uma série temporal obtida a partir do processo  $\{X_t\}_{t \in \mathbb{N}}$ , quando  $s \in (1/3, 1/2)$ , conforme a expressão (4.5). A idéia básica para estimar  $s$  baseia-se no Teorema A.2 (ver Anexo A) pois existe  $a > 0$  tal que

$$a \approx \frac{\ln |f_X(x) - f_X(y)|}{\ln |x - y|}, \text{ para } x, y \in (-\pi, \pi), \text{ bem próximos,}$$

onde  $f_X(\cdot)$  é a função densidade espectral do processo  $\{X_t\}_{t \in \mathbb{N}}$ . Definimos o estimador por

$$\hat{s} = \frac{1}{a+2} \quad \text{onde} \quad a = \frac{\ln |I(w_j) - I(w_k)|}{\ln |w_j - w_k|}, \quad (4.6)$$

sendo  $I(\cdot)$  a função periodograma, dada pela expressão (2.3), e  $w_j, w_k$  frequências de Fourier que estão bem próximas.

Consideramos em nosso estudo além da função periodograma também a função periodograma suavizado com a janela de Parzen (ver Definição 2.14) e denotamos os estimadores por  $P$  e  $SP$ , respectivamente.

O método acima vale para estimar qualquer  $s \in (0, \frac{1}{2})$  e não apenas para  $s \in (\frac{1}{2}, \frac{1}{3})$

A Tabela 4.3.5 apresenta resultados para a estimação de  $s$  no caso de dependência intermediária, considerando 200 replicações e  $j = 4$  e  $k = 7$  na expressão (4.6) pois, deste modo,  $w_j$  está próximo de  $w_k$ .

**Observação:** Outros valores para  $j$  e  $k$  poderiam ter sido utilizados pois a relação vale para todo  $x, y \in (-\pi, \pi)$  suficientemente próximos.

**Tabela 4.3.5:** Estimação para o parâmetro  $s$ .

$s$	$n$		$P$	$SP$
0,35	10.000	$média(\hat{s})$	0,4078	0,3970
		$sd(\hat{s})$	0,0374	0,0255
		$eqm(\hat{s})$	0,0047	<b>0,0028</b>
	30.000	$média(\hat{s})$	0,3870	0,4136
		$sd(\hat{s})$	0,0298	0,0208
		$eqm(\hat{s})$	<b>0,0022</b>	0,0044
0,4	10.000	$média(\hat{s})$	0,4210	0,4024
		$sd(\hat{s})$	0,0378	0,0258
		$eqm(\hat{s})$	0,0018	<b>0,0006</b>
	30.000	$média(\hat{s})$	0,4397	0,4046
		$sd(\hat{s})$	0,0432	0,0405
		$eqm(\hat{s})$	0,0034	<b>0,0016</b>
0,45	10.000	$média(\hat{s})$	0,4652	0,4359
		$sd(\hat{s})$	0,0312	0,0285
		$eqm(\hat{s})$	<b>0,0012</b>	0,0050
	30.000	$média(\hat{s})$	0,5218	0,4808
		$sd(\hat{s})$	0,0800	0,0619
		$eqm(\hat{s})$	0,0115	<b>0,0047</b>

Podemos observar (ver Tabela 4.3.5) que os resultados obtidos foram bastante satisfatórios, as estimativas apresentam baixos vício e erro quadrático

médio. Em geral, o estimador  $SP$ , ou seja, o estimador baseado na função periodograma suavizado com a janela de Parzen, fornece melhores resultados.

**Conclusão:** Analisamos neste capítulo processos que, assim como os processos  $ARFIMA$ , apresentam propriedade de longa dependência. Foram propostos vários métodos de estimação para o parâmetro  $s$ . No caso dos processos Manneville-Pomeau são necessárias séries temporais de tamanho amostral grande, implicando em tempo computacional alto para os procedimentos de estimação. Analisamos vários métodos para estimar o parâmetro  $s$ . O estimador  $Perio$  (utilizado pelos físicos) nem sempre é o método mais indicado, como por exemplo para  $s = 0,8$ , em que obtivemos melhores resultados através do estimador  $Wmp$  a partir da base de wavelets Chapéu Mexicano. Os estimadores  $Varmp$  e  $Vpmp$  apresentam vício alto. O estimador  $Cos(2)$  forneceu as melhores estimativas tanto para  $s = 0,6$  como para  $s = 0,65$ , apresentando menor erro quadrático médio para os tamanhos amostrais analisados.

Avaliamos o comportamento do estimador baseado na teoria de wavelets  $Wmp$  para o caso dos processos obtidos a partir das iterações da transformação de Manneville-Pomeau  $T_s$  quando  $s \geq 1$ , o que corresponde à medida invariante  $\mu_s$  não ser medida de probabilidade. Neste caso, as melhores estimativas foram obtidas através da base de wavelets Haar.

Os estimadores propostos para processos Manneville-Pomeau, baseados nas funções periodograma e periodograma suavizado com a janela de Parzen, no caso de dependência intermediária, forneceram estimativas muito boas com vício e erro quadrático médio baixos.

## 5

# CONCLUSÃO

Analizamos, neste trabalho, processos estocásticos com decaimento polinomial (também chamado hiperbólico) da função de autocorrelação, também denominados de *processos com propriedade de longa dependência*.

Concentramos nosso estudo nas classes dos processos *ARFIMA* e dos processos obtidos à partir de iterações da transformação de Manneville-Pomeau.

Verificamos a ergodicidade dos processos *ARFIMA* e a ordem de magnitude da variância de  $S_n = \sum_{i=0}^{n-1} X_i$  para uma amostra  $\{X_t\}_{t=0}^{n-1}$  de um processo estocástico *ARFIMA*. A partir desta ordem definimos um estimador para o parâmetro fracionário  $d$  dado por

$$V = \frac{\ln\{2n[\frac{\hat{\gamma}_X(0)}{2} + \frac{1}{n} \sum_{j=1}^{n-1} (n-j)\hat{\gamma}_X(j)]\}}{2 \ln n} - \frac{1}{2},$$

onde  $\hat{\gamma}_X(\cdot)$  é a função de autocovariância amostral do processo.

Avaliamos o comportamento de alguns métodos de estimação do parâmetro  $d^*$  no caso de processos *ARFIMA*( $p, d^*, q$ ) não estacionários. Quando o parâmetro de diferenciação  $d^*$  está no intervalo  $(0, 5; 1, 0)$  os estimadores são bons e o método *FT* fornece, em geral, melhores resultados. Este estimador apresenta piores resultados do que os métodos semiparamétricos quando são incluídos no modelo parâmetros de curta dependência, tanto autoregressivos como médias móveis. Os estimadores melhoram sempre que o tamanho amostral aumenta, especialmente no modelo *ARFIMA*( $0, d^*, 0$ ).

No caso não estacionário, sem propriedade de reversão do nível (isto é, quando  $d^* \in (1, 0; 1.5)$ ), observamos que todos os métodos analisados são muito viciados e subestimam o verdadeiro valor do parâmetro. Utilizando, por exemplo, a função cosine-bell observamos, nas simulações, que as estimativas melhoram, no sentido de redução do vício, ocorrendo, no entanto, aumento na variância. Uma alternativa para este caso é utilizar o método de estimação *W*, baseado na teoria de wavelets.

Os métodos  $V$  e  $VP$ , apresentados na Seção 3.3, resultam ser muito viciados e não são recomendados para nenhum dos casos analisados: estacionário ou não estacionário.

Analisamos a série temporal “*taxa de juros a curto prazo da Inglaterra*” quanto à longa dependência e à presença de raiz unitária, aplicando alguns métodos de estimação. Concluimos que a série temporal é não estacionária com longa dependência onde o parâmetro de diferenciação é estimado por  $\hat{d}^* = 0,8514$ . Considerar  $d^* = 1,0$  para esta série temporal, como Mills (1997) sugere, acarreta em uma super diferenciação desta série.

Observamos que as estimativas do parâmetro de diferenciação  $d^*$  em processos  $ARFIMA$  quando as inovações são não Gaussianas, mantém o mesmo comportamento dos resultados obtidos, no caso Gaussiano, para os métodos analisados.

Estudamos o comportamento do estimador do parâmetro de diferenciação após aplicar primeira diferença na série original, com o objetivo de verificar se vale a igualdade  $\hat{d}^* = \hat{d} + 1$ . Para o modelo  $ARFIMA(0, d^*, 0)$ , quando  $d^* \in (0, 5; 1, 0]$ ,  $FT$  é o melhor método de estimação com a propriedade de invariância para a primeira diferença, seguido por  $MGPH(2)$ . Para o modelo  $ARFIMA(0, d^*, 0)$ , quando  $d^* \in (1, 0; 1, 5)$ , nenhum dos métodos considerados é satisfatório para o caso não estacionário e, portanto, não vale a propriedade de invariância para a primeira diferença. No modelo  $ARFIMA(0, d^*, 1)$  o método  $FT$  apresenta resultados ruins para a estimação de  $d^*$ , mas as estimativas de  $d$  após a primeira diferença, apresentam baixo erro quadrático médio. Para este estimador não se verifica a propriedade de invariância para primeira diferença, nesta situação; mas para os outros métodos temos, em geral,  $\hat{d}^* \simeq \hat{d} + 1$ . Em ambos os modelos  $ARFIMA(1, d^*, 0)$  e  $ARFIMA(0, d^*, 1)$  foram obtidos bons resultados através do método  $MGPH(2)$ . Embora o método  $SPR$  apresente grande erro quadrático médio, ele foi satisfatório nas estimações simultâneas de  $d^*$  e  $d$ . Para o modelo  $ARFIMA(1, d^*, 1)$  quando  $\phi$  e  $\theta$  são positivos, todos os métodos, exceto  $FT$ , subestimam  $d^*$  e, em geral, superestimam  $d$ . Para  $\phi$  negativo e  $\theta$  positivo,  $MGPH$  apresenta os melhores resultados para estimar  $d^*$ . Nesta situação, a relação  $\hat{d}^* \simeq \hat{d} + 1$  é mais evidente para o método  $SPR$ .

Estudamos os processos obtidos a partir de iterações da transformação Manneville-Pomeau  $T_s$  e observamos que estes processos apresentam característica de longa dependência (denominada, em outras áreas, de intermitência). Apresentamos a equivalência dos parâmetros entre os processos  $ARFIMA$  e processos Manneville-Pomeau obtidos a partir das iterações da transformação  $T_s$  e das iterações da transformação aproximação linear por partes  $T_\gamma$  da aplicação Manneville-Pomeau. Implementamos vários estimadores para processos Manneville-Pomeau obtidos a partir das iterações da transformação  $T_s$ .

No caso dos processos Manneville-Pomeau são necessárias séries temporais de tamanho amostral grande, implicando em muito tempo computacional para os procedimentos de estimação. O estimador *Perio* (utilizado pelos físicos) nem sempre é o método mais indicado, como por exemplo para  $s = 0,8$ , em que obtivemos melhores resultados através do estimador *Wmp* a partir da base de wavelets Chapéu Mexicano. Os estimadores *Varmp* e *Vpmp* apresentam vício alto. O estimador *Cos(2)* forneceu as melhores estimativas para  $s = 0,6$  e para  $s = 0,65$ , apresentando menor erro quadrático médio para os tamanhos amostrais analisados.

Avaliamos o comportamento do estimador baseado na teoria de wavelets *Wmp* para o caso dos processos obtidos a partir das iterações da transformação de Manneville-Pomeau  $T_s$  quando  $s \geq 1$ , o que corresponde à medida invariante  $\mu_s$  não ser medida de probabilidade. Neste caso, as melhores estimativas foram obtidas através da base de wavelets Haar.

Os estimadores propostos para processos Manneville-Pomeau, baseados nas funções periodograma e periodograma suavizado com a janela de Parzen, no caso de dependência intermediária, forneceram estimativas muito boas com baixos vício e erro quadrático médio.

Finalmente, ressaltamos que os processos Manneville-Pomeau apresentam característica de longa dependência, assim como os processos *ARFIMA*. Os processos Manneville-Pomeau obtidos a partir de iterações da transformação  $T_s$  apresentam uma correspondência com os processos *ARFIMA* através da relação  $d = 1 - \frac{1}{2s}$ , onde  $d$  é o parâmetro fracionário destes modelos. Para o parâmetro fracionário  $d$  temos vários estudos de procedimentos de estimação. O mesmo não ocorre para o parâmetro  $s$  nos processos Manneville-Pomeau: neste caso, os pesquisadores (em especial, físicos) utilizam o método *Perio*. Mas, como ficou evidenciado neste trabalho, existem outros métodos mais eficazes para a estimação do parâmetro  $s$ .

# ANEXO A

Considere os processos *ARFIMA*( $p, d, q$ ), dados pela expressão (3.3), ou os processos Manneville-Pomeau, dados pela expressão (4.2). Sejam  $\rho_X(\cdot)$  e  $f_X(\cdot)$  as funções de autocorrelação e densidade espectral, respectivamente, destes processos.

Neste anexo, explicamos porque o decaimento hiperbólico (ou polinomial) da função de autocorrelação, isto é,

$$\rho_X(k) \approx k^{-(1-2d)},$$

corresponde ao comportamento

$$f_X(w) \approx w^{-2d},$$

para a função densidade espectral nos processos *ARFIMA* (ver Conclusão 1 a seguir).

Os processos Manneville-Pomeau, dados pela expressão (4.2), satisfazem propriedades análogas, conforme a Conclusão 2.

Para esclarecer esta e outras relações enunciamos a definição e os seguintes teoremas.

**DEFINIÇÃO A.1:** Uma função  $g : (-\pi, \pi) \rightarrow \mathbb{R}$  é dita Hölder de ordem  $a$  se existe constante  $K > 0$  tal que

$$|g(x) - g(y)| \leq K|x - y|^a,$$

para todo  $x, y \in (-\pi, \pi)$ .

A relação entre decaimento hiperbólico da função de autocorrelação e a expressão da função densidade espectral para frequência próxima do ponto 0 é uma questão que envolve apenas séries de Fourier (ver Bary (1964)).

**TEOREMA A.1:** *Suponha que  $g(\theta) = \sum_{n=1}^{\infty} b_n \cos(n\theta)$ , onde  $b_n \in \mathbb{R}$ , converge a zero. Se  $a > 0$  e  $g(\cdot)$  é função Hölder de ordem  $a$ , então existe uma constante  $c$  tal que  $b_n < cn^{-(1+a)}$  para todo  $n \in \mathbb{N} - \{0\}$ .*

**TEOREMA A.2:** *Suponha que  $g(\theta) = \sum_{n=1}^{\infty} b_n \cos(n\theta)$ , onde  $b_n \in \mathbb{R}$ , decresce monotonamente a zero. Se  $a > 0$  e existe uma constante  $c$  maior do que zero tal que  $b_n < cn^{-(1+a)}$ , então  $g(\cdot)$  é função Hölder de ordem  $a$ .*

Os Teoremas A.1 e A.2 tratam do caso de dependência intermediária.

Um outro resultado interessante, e que pode ser aplicado ao caso de longa dependência, é descrito a seguir.

**TEOREMA A.3 (Riesz):** *Suponha que  $g(\theta) = \sum_{n=1}^{\infty} b_n \cos(n\theta)$ , para todo  $\theta \in (-\pi, \pi)$  e que  $b_n \in \mathbb{R}$ , é tal que a seqüência  $\{b_n\}_{n \in \mathbb{N}}$  tende a zero quando  $n$  tende a infinito. Suponha que existam  $c_1, c_2$  e  $u$  constantes reais não negativas, tais que  $c_1 n^{-u} < |b_n| < c_2 n^{-u}$ , ou seja,  $b_n \approx n^{-u}$  para alguma constante  $u > 0$ . Suponha ainda que existam  $d_1, d_2$  constantes reais não negativas e  $b \in (-1, 0)$  tais que*

$$d_1 |\theta|^b \leq |g(\theta)| \leq d_2 |\theta|^b,$$

ou seja,  $g(\theta) \approx \theta^b$ , para  $\theta$  próximo de zero.

(a) Se existem  $a \in (-1, 0)$  e  $k$  constante real não negativa tais que

$$\left| \frac{g(\theta)}{\theta^a} \right| \leq k,$$

então  $u \geq 1 + a$ . Ou seja, o decrescimento de  $|b_n|$  é pelo menos da ordem de  $n^{-(1+a)}$ , quando  $n$  vai a infinito.

(b) Se existem  $a \in (-1, 0)$  e  $v \in \mathbb{R}$ ,  $v > 0$ , tais que  $|b_n| < v n^{-(1+a)}$ , então  $b \leq a$ . Ou seja,  $g(\theta)$  é pelo menos da ordem de  $|\theta|^a$ , quando  $\theta$  vai a zero.

Assim, dos itens (a) e (b) concluímos que  $u = 1 + b$ .

### Observações:

a) No caso de processos  $ARFIMA(p, d, q)$ , quando  $0 < d < 0,5$ , concluímos que  $u = 1 + b$  onde  $b = -2d$  a partir das considerações acima. Deste modo se obtém o resultado da Conclusão 1. Portanto, obtemos  $u = 1 - 2d$  (ver também Exemplo A.1).

b) No caso geral, observamos que existem seqüências  $\{b_n\}_{n \in \mathbb{N}}$  que satisfazem  $c_1 n^{-u} < |b_n| < c_2 n^{-u}$  para algum  $u$ , mas  $g(\theta)$  não satisfaz  $c_3 |\theta|^b \leq |g(\theta)| \leq c_4 |\theta|^b$  para nenhum  $b$ .

O espaço  $\mathcal{L}^p$  é o conjunto das funções  $f$  tais que  $\int_{-\pi}^{\pi} |f(x)|^p dx < \infty$ .

O resultado acima é consequência do seguinte Teorema (ver Bary (1964)).

**TEOREMA A.4 (Riesz):** Suponha que  $g(\theta) = \sum_{n=1}^{\infty} b_n \cos(n\theta)$  para todo  $\theta \in (-\pi, \pi)$  e que  $\{b_n\}_{n \in \mathbb{N}} \in \mathbb{R}$  tende a zero. Sejam  $p > 1$  e  $q > 1$  tais que  $\frac{1}{p} + \frac{1}{q} = 1$ .

(a) Se  $g \in \mathcal{L}^p$ , então  $\sum_{n=1}^{\infty} |b_n|^q < \infty$ .

(b) Se  $\sum_{n=1}^{\infty} |b_n|^q < \infty$ , então  $g \in \mathcal{L}^p$ .

**Observação:** O Teorema A.3 segue do Teorema A.4 usando que

a) para  $f$  de ordem  $x^\alpha$  ( $x$  próximo de zero), então  $f \in \mathcal{L}^1 \Leftrightarrow \alpha > -1$

e

b) para  $b_n$  da ordem de  $n^{-\beta}$ , ( $n$  próximo de infinito) então  $\sum_{n=1}^{\infty} |b_n| < \infty \Leftrightarrow \beta > 1$ .

Os detalhes da demonstração de todos os resultados acima podem ser encontrados em Gioveli (1999).

Em processos  $ARFIMA(p, d, q)$  as hipóteses do Teorema A.3 estão satisfeitas. Isto é visto através do exemplo a seguir:

**Exemplo A.1:** Seja  $\{X_t\}_{t \in \mathbb{Z}}$  um processo  $ARFIMA(p, d, q)$ , dado pela expressão (3.3). Seja  $d \in (0, 0; 0, 5)$ . Seja  $f_X(\cdot)$  a função densidade espectral do processo  $\{X_t\}_{t \in \mathbb{Z}}$  dada por  $f_X(w) = \frac{\sigma_\epsilon^2}{2\pi} [2 \sin \frac{w}{2}]^{-2d} \sim C w^{-2d}$ , para  $w$  próximo da origem, onde  $C$  e  $\sigma_\epsilon^2$  são constantes. A função de autocovariância do processo  $\{X_t\}_{t \in \mathbb{Z}}$  é dada por  $\gamma_X(k) = \int_{-\pi}^{\pi} e^{-ikw} f_X(w) dw$ , que é o coeficiente de Fourier da  $f_X(\cdot)$  na forma complexa.

É possível mostrar que

$$\gamma_X(k) = C \frac{\Gamma(k+d)}{\Gamma(k+1-d)}, \text{ para todo } k \in \mathbb{N}$$

onde  $C = \frac{\sigma_\epsilon^2 \Gamma(1-2d)}{\Gamma(d)\Gamma(1-d)} > 0$  e  $\Gamma(\cdot)$  é a função Gama.

Vamos mostrar que  $\gamma_X(k) \approx k^{2d-1}$ , para todo  $k \in \mathbb{N}$ . De fato, mostraremos que  $\lim_{k \rightarrow \infty} \frac{\Gamma(k+a)}{\Gamma(k+b)k^{a-b}} = L$ , onde  $L$  é uma constante positiva.

Pela fórmula de Stirling (ver Brockwell e Davis (1991)) segue que

$$\Gamma(x) \sim \sqrt{2\pi} \exp(-x+1)(x-1)^{x-0,5}, \text{ quando } x \rightarrow \infty,$$

onde “ $\sim$ ” indica que a razão dos dois termos tende a 1 quando  $x \rightarrow \infty$ .

Portanto,

$$\begin{aligned} & \lim_{k \rightarrow \infty} \frac{\Gamma(k+a)}{\Gamma(k+b)k^{a-b}} = \\ & = \lim_{k \rightarrow \infty} \frac{\sqrt{2\pi} e^{(1-k-a)} (k+a-1)^{k+a-\frac{1}{2}}}{\sqrt{2\pi} e^{(1-k-b)} (k+b-1)^{k+b-\frac{1}{2}}} \frac{1}{k^{a-b}} = \end{aligned}$$

$$\begin{aligned}
&= \lim_{k \rightarrow \infty} e^{b-a} \left( \frac{k+a-1}{k+b-1} \right)^k \left( \frac{k+a-1}{k+b-1} \right)^{-\frac{1}{2}} \left( \frac{k+a-1}{k} \right)^a \left( \frac{k}{k+b-1} \right)^b = \\
&= e^{b-a} \lim_{k \rightarrow \infty} \left( \frac{k+a-1}{k+b-1} \right)^k \left( \frac{k+a-1}{k+b-1} \right)^{-\frac{1}{2}} \left( \frac{k+a-1}{k} \right)^a \left( \frac{k}{k+b-1} \right)^b = \\
&= e^{b-a} = L.
\end{aligned}$$

Isto é, dado  $\varepsilon > 0$ , existe  $K > 0$  tal que, para todo  $k > K$ , implica

$$\left| \frac{\Gamma(k+a)}{\Gamma(k+b)k^{a-b}} - L \right| < \varepsilon.$$

Ou seja,

$$(L - \varepsilon)k^{a-b} < \frac{\Gamma(k+a)}{\Gamma(k+b)} < (L + \varepsilon)k^{a-b}.$$

Podemos escolher  $\varepsilon$  de modo que  $L - \varepsilon > 0$ .

Fazendo  $a = d$  e  $b = 1 - d$  obtemos

$$(L - \varepsilon)k^{2d-1} < \frac{\Gamma(k+d)}{\Gamma(k+1-d)} < (L + \varepsilon)k^{2d-1}.$$

Então,

$$(L - \varepsilon)C k^{2d-1} < \gamma_X(k) < (L + \varepsilon)C k^{2d-1}.$$

Ou seja, existem  $c_1, c_2$  positivas tais que  $c_1 k^{2d-1} \leq \gamma_X(k) \leq c_2 k^{2d-1}$ , isto é,  $\gamma_X(k) \approx k^{2d-1}$  para todo  $k \in \mathbb{N}$ .

**Conclusão 1:** Seja  $\{X_t\}_{t \in \mathbb{Z}}$  um processo *ARFIMA*( $p, d, q$ ), dado pela expressão (3.3). Pelo Teorema A.3, temos que a relação entre  $b$  e  $u$  determina que

$$\rho_X(k) \approx k^{2d-1}, \text{ para todo } k \in \mathbb{Z},$$

para a função de autocorrelação do processo  $\{X_t\}_{t \in \mathbb{Z}}$ , é equivalente à relação

$$f_X(w) \approx w^{-2d}, \text{ para } w \text{ próximo da origem,}$$

para a função densidade espectral do mesmo processo. O Teorema A.3 confirma a estimativa do Exemplo A.1.

**Conclusão 2:** Seja  $\{X_t\}_{t \in \mathbb{N}}$  um processo Manneville-Pomeau, dado pela expressão (4.2). Pelo Teorema A.3, temos que a relação entre  $b$  e  $u$  determina que

$$\rho_X(k) \approx k^{1-\frac{1}{s}}, \text{ para todo } k \in \mathbb{N}, \tag{A.1}$$

para a função de autocorrelação do processo  $\{X_t\}_{t \in \mathbb{N}}$ , é equivalente à relação

$$f_X(w) \approx w^{\frac{1}{s}-2}, \text{ para } w \text{ próximo da origem,} \quad (\text{A.2})$$

para a função densidade espectral do mesmo processo.

A expressão (A.1), para processos Manneville-Pomeau, segue de Young (1999) e Hu (1998). As expressões (A.1) e (A.2), para processos Manneville-Pomeau por partes, seguem de Fisher e Lopes (2001) que também mostram que  $\gamma_X(k)$  é monótona em  $k \in \mathbb{N}$ . Este resultado é importante para utilizar o Teorema A.2 e também para mostrar que  $f_X(w) \approx w^{\gamma-3}$ , para  $2 < \gamma < 3$ .

## ANEXO B

Apresentamos, a seguir, os dados da série temporal “*taxa de juros a curto prazo da Inglaterra*”, que consiste de 148 observações medidas trimestralmente de 1952 a 1988. Esta série temporal é apresentada em Mills (1997) e foi analisada na Seção 3.4.4. Os dados estão listados em seqüência cronológica por linha.

0,99	2,32	2,47	2,42	2,40	2,38	2,35	2,10
2,10	2,05	1,56	1,59	2,36	3,90	3,97	4,07
4,07	5,01	4,98	4,99	4,55	3,91	3,85	6,60
6,13	5,18	4,16	3,58	3,11	3,28	3,48	3,39
4,54	4,65	5,55	5,08	4,17	4,40	6,69	5,73
5,25	4,05	3,89	3,86	3,49	3,71	3,76	3,75
3,75	4,30	4,65	4,70	6,55	6,38	5,63	5,44
5,49	5,64	6,68	6,51	6,08	5,41	5,35	5,73
7,52	7,08	7,06	6,50	6,73	7,80	7,83	7,75
7,51	6,77	6,82	6,81	6,77	5,68	5,56	4,56
4,35	4,27	5,76	6,89	8,13	7,67	10,89	10,67
12,03	11,48	11,19	10,89	10,96	9,24	10,44	11,41
9,30	9,94	10,87	14,43	11,74	7,50	7,30	4,48
5,77	6,99	9,11	10,98	12,09	11,29	13,35	13,47
15,74	16,06	14,44	14,36	12,61	11,24	13,8	15,66
13,51	12,98	11,08	8,83	10,94	9,68	9,4	8,83
8,86	8,36	11,36	9,88	11,52	11,90	10,99	11,13
12,10	9,85	9,53	10,56	10,56	9,17	8,94	8,87
8,21	7,83	10,26	11,50				

# REFERÊNCIAS

- [1] Anderson, T.W. (1971). *The Statistical Analysis of Time Series*. John Wiley, New York.
- [2] Bary, N.K. (1964). *A Treatise on Trigonometric series*, Vol. II. Pergamon Press, New York.
- [3] Beran, J. (1994). *Statistics for Long-Memory Processes*. Chapman & Hall, New York.
- [4] Billingsley, P. (1995). *Probability and Measure*. John Wiley, New York.
- [5] Box, G.E.P., G.M. Jenkins e G.C. Reinsel (1995). *Time Series Analysis: Forecasting and Control*. Prentice Hall, New Jersey.
- [6] Borovkova, S., R. Burton e H. Dehling (2001). “Limit Theorems for Functionals of Mixing Processes with Applications to U-Statistics and Dimension Estimation”. *Transactions of the American Mathematical Society*, Vol. **353**(11), pp. 4261-4318.
- [7] Brockwell, P.J. e R.A. Davis (1991). *Time Series: Theory and Methods*. Springer-Verlag, New York.
- [8] Dahlhaus, R. (1989). “Efficient parameter estimation for self-similar processes”. *Annals of Statistics*, Vol. **17**(4), pp. 1749-1766.
- [9] Durrett, R. (1996). *Probability: Theory and Examples*. Duxbury Press, Belmont.
- [10] Feller, W. (1949). “Fluctuation Theory of Recurrent Events”. *Trans. Amer. Math. Soc.*, Vol. **67**, pp. 98-119.
- [11] Fisher, A. e A. Lopes (2001). “Exact bounds for the polynomial decay of correlation, 1/f noise and the CLT for the equilibrium state of a non-Hölder potential”. *Nonlinearity*, Vol. **14**, pp. 1071-1104.

- [12] Fox, R. and M.S. Taqqu (1986). “Large-sample properties of parameter estimates for strongly dependent stationary Gaussian time series”. *The Annals of Statistics*, Vol. **14**, pp. 517-532.
- [13] 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), pp. 221-238.
- [14] Gioveli, I. (1999). *Teoremas Tauberianos para Séries de Fourier*. Dissertação de Mestrado. Instituto de Matemática da UFRGS, Porto Alegre.
- [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), pp. 15-29.
- [16] Hosking, J.R.M. (1981). “Fractional Differencing”. *Biometrika*, Vol. **68**(1), pp. 165-167.
- [17] Hosking, J. (1982). “Some models of persistence in time series analysis: theory and practice”. *Time Series Analysis: Theory and Practice*, Vol. **1**, pp. 641-653.
- [18] Hu, H. (1998). “Decay of correlations for piecewise smooth maps with indifferent fixed points”. Preprint.
- [19] Hurst, H.E. (1951). “Long-term storage capacity of reservoirs”. *Transactions of American Society in Civil Engineers*, Vol. **116**, pp. 165-167.
- [20] Hurvich, C.M. e B.K. Ray (1995). “Estimation of the Memory Parameter for Nonstationary or Noninvertible Fractionally Integrated Processes”. *Journal of Time Series Analysis*, Vol. **16**(1), pp. 017-042.
- [21] Isola, S. (1995). “Dynamical zeta functions and correlation functions for non-uniformly hyperbolic transformations”. Preprint.
- [22] Jensen, M.J. (1999). “Using Wavelets to Obtain a Consistent Ordinary Least Squares Estimator”. *Journal of Forecasting*, Vol. **18**, pp. 17-32.
- [23] Karlin, S. e H.M. Taylor (1975). *A First Course in Stochastic Processes*. Academic Press, San Diego.

- [24] Liverani, C., B. Saussol e S. Vaienti (1998). “Conformal measure and decay of correlation for covering weighted systems”. *Ergodic Theory and Dynamical Systems*, Vol. **18**(6), pp. 1399-1420.
- [25] Lopes, A. (1993). “The zeta function, non-differentiability of pressure and the critical exponent of transition”. *Advances in Mathematics*, Vol. **101**, pp. 133-165.
- [26] Lopes, A. e S.R.C. Lopes (1998). “Parametric Estimation and Spectral Analysis of Piecewise Linear Maps of the Interval”. *Advances on Applied Probability*, Vol. **30**, pp. 757-776.
- [27] Lopes, A. e S.R.C. Lopes (2002). “Convergence in Distribution for the Periodogram of Chaotic Processes”. Submetido.
- [28] Lopes, S.R.C., B.P. Olbermann e V.A. Reisen (2002a). “A Comparison of Estimation Methods in Non-stationary ARFIMA Processes”. Em fase de revisão.
- [29] Lopes, S.R.C., B.P. Olbermann e V.A. Reisen (2002b). “Non-stationary Gaussian ARFIMA Processes: Estimation and Application”. Em fase de revisão.
- [30] Maes, C., F. Redig, F. Takens, A.V. Moffaert e E. Verbitski (1999). “Intermittency and weak Gibbs states”. Institut voor Theoretisch Fysica K.U. Leuven, Bélgica.
- [31] Mandelbrot, B.B. (1965). “Une classe de processus stochastiques homothétiques à soi; application a la loi climatologique de E. Hurst.” *Comptes Rendus Academ. Sci. Paris*, Vol. **260**, pp. 3274-3277.
- [32] Mandelbrot, B.B. e J.W. van Ness (1968). “Fractional Brownian motions, fractional noises and applications”. *SIAM Review*, Vol. **10**(4), pp. 422-437.
- [33] Mandelbrot, B.B. e J.R. Wallis (1969). “Computer experiments with fractional Gaussian noises. Part One, Averages and variances”. *Water Resources Research*, Vol. **5**(1), pp. 228-241.
- [34] Manneville, P. (1980). “Intermittency, self-similarity and  $1/f$  spectrum in dissipative dynamical systems”. *J. Physique*, Vol. **41**, pp. 1235-1243.
- [35] Meyer, Y. (1993). *Wavelets: Algorithms and Applications*. Philadelphia: SIAM.

- [36] Mills, T.C. (1997). *The Econometric Modelling of Financial Time Series*. Cambridge University Press, London.
- [37] Morettin, P.A. (1997). “Wavelets in Statistics”. *Resenhas IME-USP*, Vol. **3**(2), pp. 211-272.
- [38] Morettin, P.A. (1999). *Ondas e Ondaletas: Da Análise de Fourier à Análise de Ondaletas*. Editora da USP, São Paulo.
- [39] Olbermann, B.P., S.R.C. Lopes e V.A. Reisen (2002). “Invariance of the first difference in ARFIMA models”. Em preparação.
- [40] Olbermann, B.P. (1998). *Estimação do Parâmetro de Diferenciação em Séries Temporais Não Estacionárias com Características de Longa Dependência*. Dissertação de Mestrado. Instituto de Matemática da UFRGS, Porto Alegre.
- [41] Parry, W. e M. Pollicot (1990). “Zeta functions and the periodic orbit structure of hyperbolic dynamics”. *Asterisque*, Vol. **187-188**, pp. 1-226.
- [42] Percival, D.B. e A.T. Walden (1993). *Spectral Analysis for Physical Applications: Multitaper and Conventional Univariate Techniques*. Cambridge University Press, Cambridge.
- [43] Pianigiani, G. (1980). “First return maps and invariant measures”. *Israel Journal of Mathematics*, Vol. **35**(1-2), pp. 32-48.
- [44] Poli, R. e S.R.C. Lopes (2002). “O Uso de Wavelets em Processos de Longa Dependência”. *Relatório Técnico N° 57, Série A*, Instituto de Matemática, UFRGS, Porto Alegre.
- [45] Pollicott, M. e M. Yuri (1998). *Dynamical Systems and Ergodic Theory*. Cambridge University Press, Cambridge.
- [46] Priestley, M.B. (1981). *Spectral Analysis in Time Series*. Academic Press, New York.
- [47] Reisen, V.A. (1994). “Estimation of the Fractional Difference Parameter in the ARIMA(p,d,q) model using the Smoothed Periodogram”. *Journal of Time Series Analysis*, Vol. **15**(3), pp. 335-350.
- [48] Reisen, V.A. e S.R.C. Lopes (1999). “Some Simulations and Applications of Forecasting Long-Memory Time Series Models”. *Journal of Statistical Planning and Inference*, Vol. **80**(2), pp. 269-287.

- [49] Robinson, P.M. (1995). “Log-periodogram regression of time series with long range dependence”. *Annals of Statistics*, Vol. **23**(3), pp. 1630-1661.
- [50] Samorodnitsky, G. e M.S. Taqqu (1994). *Stable Non-Gaussian Random Processes: Stochastic Models with Infinite Variance*. Chapman & Hall, New York.
- [51] Sowell, F. (1992). “Maximum Likelihood estimation of stationary univariate fractionally integrated time series models”. *Journal of Econometrics*, Vol. **53**, pp. 165-188.
- [52] Schuster, H.G. (1984). *Deterministic Chaos - an Introduction*. Physik-Verlag, Weinheim.
- [53] Thaler, M. (1980). “Estimates of the invariant densities of endomorphism with indifferent fixed points”. *Israel Journal of Mathematics*, Vol. **37**(4), pp. 303-313.
- [54] Velasco, C. (1999a). “Gaussian Semiparametric Estimation of Non-stationary Time Series”. *Journal of Time Series Analysis*, Vol. **20**(1) pp. 87-127.
- [55] Velasco, C. (1999b). “Non-stationary log-periodogram regression”. *Journal of Econometrics*, Vol. **91** pp. 325-371.
- [56] Velasco, C. (2000). “Non-Gaussian log-periodogram regression”. *Econometric Theory*, Vol. **16**, pp. 44-79.
- [57] Wei, W. (1990). *Time Series Analysis. Univariate and Multivariate Methods*. Addison Wesley, New York.
- [58] Whittle, P. (1951). *Hypothesis Testing in Time Series Analysis*. Hafner, New York.
- [59] Yajima, Y. (1985). “On Estimation of Long Memory Time Series Models”. *Austral. J. Statist.*, Vol. **27**(3), pp. 303-320.
- [60] Young, L.-S. (1999). “Recurrence times and rates of mixing”. *Israel Journal of Mathematics*, Vol. 110, pp. 153-188.
- [61] Zandonade, E. (1999). *Utilização de Ondaletas em Modelos de Espaços de Estado*. Tese de Doutorado, Departamento de Estatística, IME, USP, São Paulo.
- [62] Zebrowski, J.J. (2000). “Intermittency in Human Heart Rate Variability”. Faculty of Physics, Warsaw University of Technology, Polônia.