



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

# **Teoria dos Grandes Desvios na Estimação de Processos Autorregressivos de Primeira Ordem**

Dissertação de Mestrado

**Maicon Josué Karling**

Porto Alegre, Agosto de 2017



Dissertação submetida por Maicon Josué Karling<sup>1</sup> como requisito parcial para a obtenção do grau de Mestre em Ciência Probabilidade e Estatística pelo Programa de Pós-Graduação em Matemática do Instituto de Matemática e Estatística da Universidade Federal do Rio Grande do Sul.

Professora Orientadora:

Dr<sup>a</sup>. Sílvia Regina Costa Lopes - (UFRGS)

Banca Examinadora:

Dr. Ary Vasconcelos Medino - (UnB)

Dr. Rafael Rigão Souza (UFRGS)

Dr. Rogério Ricardo Steffenon (UNISINOS)

Data da Defesa: 03 de Agosto de 2017

---

<sup>1</sup> Bolsista do Conselho Nacional de Desenvolvimento Científico e Tecnológico - CNPq

Karling, Maicon Josué

Teoria dos grandes desvios na estimação de processos autorregressivos de primeira ordem / Maicon Josué Karling. - - Agosto de 2017.  
136 p. : il. (algumas color.)

Orientadora: Dr<sup>a</sup> Sílvia Regina Costa Lopes

Dissertação (Mestrado) - - Universidade Federal do Rio Grande do Sul, Instituto de Matemática, Programa de Pós-Graduação em Matemática, Porto Alegre, BR-RS, Agosto de 2017.

1. Grandes desvios. 2. Processos autorregressivos de primeira ordem. 3. Estimador de Yule-Walker. 4. Estimador de mínimos quadrados. 5. Distribuições  $\alpha$ -estáveis. I. Lopes, Sílvia Regina Costa, orient. II. Título.

# Agradecimentos

Sou grato àqueles que, com pouca ou grande influência, direta ou indiretamente, deixaram seus traços presentes ao longo deste trabalho. Quero agradecer especialmente aos meus pais, José e Mariléia, pelo amor incondicional. Ao professor Rogério, pelo incentivo em prosseguir na vida acadêmica. À professora Sílvia, por acreditar no meu potencial e me envolver com sabedoria e inspiração. Aos demais componentes da banca, professores Ary e Rafael, por demonstrarem grande interesse na leitura deste trabalho. À UFRGS pela sua excelência como universidade. Ao CNPq, pelo apoio financeiro. E aos meus colegas e amigos, pelo companheirismo e vivências durante esta jornada acadêmica.



# Resumo

Neste trabalho estudamos princípios de grandes desvios aplicados à estimação do parâmetro de um processo autorregressivo Gaussiano de primeira ordem. Levamos em consideração o estimador de Yule-Walker e o estimador de mínimos quadrados do parâmetro deste processo. O método utilizado para obter a função taxa, consiste na decomposição dos estimadores em uma combinação linear entre variáveis aleatórias independentes e identicamente distribuídas, com distribuição  $\chi_1^2$ . Os coeficientes dessa combinação linear são os autovalores do produto de duas matrizes de Toeplitz. Também estudamos o processo autorregressivo de primeira ordem com inovações advindas de um processo  $\alpha$ -estável não-Gaussiano. Mostramos que tal processo é estacionário, mixing e ergódico. Além disso, provamos que a matriz de codiferença deste processo pode ser representada através de uma matriz de Toeplitz.

**Palavras-chaves:** Grandes desvios; Processos autorregressivos de primeira ordem; Estimador de Yule-Walker; Estimador de mínimos quadrados; Distribuições  $\alpha$ -estáveis.

# Abstract

In the present work, we apply the large deviation principles to the parameter estimation of a first order autoregressive Gaussian process. We consider the Yule-Walker and the least squares estimators for the parameter of this process. The method used here consists in the decomposition of the estimators into a linear combination of independent and identically distributed random variables, with a  $\chi_1^2$  distribution. The coefficients of this linear combination are the eigenvalues of a two Toeplitz matrices product. We also study the first order autoregressive process with non-Gaussian  $\alpha$ -stable innovations. We show that this process has stationarity, mixing and ergodic properties. Moreover, we prove that the codifference matrix of this process can be represented as a Toeplitz matrix.

**Keywords:** Large deviations; First order autoregressive processes; Yule-Walker estimator; Least squares estimator;  $\alpha$ -stable distributions.





# Lista de abreviaturas e notações

$\mathbb{N}$	conjunto dos números naturais, dado por $\{1, 2, 3, \dots\}$
$AR(p)$	processo autorregressivo de ordem $p \in \mathbb{N}$
i.i.d.	variáveis aleatórias independentes e identicamente distribuídas
LFGN	Lei Forte dos Grandes Números
$\bar{X}_n$	média amostral das variáveis aleatórias $X_1, \dots, X_n$
$\xrightarrow{q.c.}$	convergência quase certa
$\xrightarrow{D}$	convergência em distribuição
$\stackrel{d}{=}$	variáveis aleatórias idênticas em distribuição
LDP	princípio de grandes desvios
SLDP	princípio forte de grandes desvios
$\mathcal{N}(\mu, \sigma^2)$	distribuição normal de média $\mu$ e variância $\sigma^2$
$\mathcal{B}(\mathcal{X})$	$\sigma$ -álgebra de Borel do espaço topológico $\mathcal{X}$
$\log(\cdot)$	logaritmo natural
$\bar{A}$	fecho do conjunto $A$
$\mathcal{L}_X(\cdot)$	função de log-Laplace da variável aleatória $X$
$M_X(\cdot)$	função geradora de momentos da variável aleatória $X$
$\mathbb{Z}^+$	conjunto dos inteiros não-negativos
$\Gamma(a)$	função gama, calculada em $a$
$\chi_n^2$	distribuição qui-quadrado com $n$ graus de liberdade
$D_h$	domínio da função $h(\cdot)$
$L_n(\lambda, X_n)$	função geradora de cumulantes normalizada da variável aleatória $X_n$
$h^*(\cdot)$	transformada de Fenchel-Legendre da função $h(\cdot)$
$A^\circ$	interior do conjunto $A$
$\tilde{\theta}_n$	estimador de Yule-Walker do parâmetro $\theta$

$\hat{\theta}_n$	estimador de mínimos quadrados do parâmetro $\theta$
$\mathcal{W}_n(f)$	estatística definida na equação (3.9)
$\mathbb{T}$	subconjunto da reta, dado por $[-\pi, \pi)$
$T_n(h)$	matriz de Toeplitz de ordem $n$ , associada à função $h(\cdot)$
$T_n^{1/2}(h)$	raiz quadrada da matriz de Toeplitz $T_n(h)$
$T_n^{-1}(h)$	matriz inversa da matriz de Toeplitz $T_n(h)$
$\ h\ _\infty$	norma da função $h(\cdot)$ no espaço $L^\infty$
$\gamma_X(\cdot)$	função de autocovariância do processo $(X_n)_{n \in \mathbb{N}}$
$M_{n \times n}(\mathbb{R})$	matrizes reais quadradas de ordem $n$
$\underline{a}_n(f, g)$	menor dos $n$ autovalores da matriz $T_n(f)T_n(g)$
$\bar{a}_n(f, g)$	maior dos $n$ autovalores da matriz $T_n(f)T_n(g)$
$M_h$	supremo essencial da função $h(\cdot)$
$m_h$	ínfimo essencial da função $h(\cdot)$
$\mathcal{A}(f, g)$	conjunto de todos os pontos de acumulação de sequências da forma $(\underline{a}_n(f, g), \bar{a}_n(f, g)) \in \mathbb{R}^2$
$L(fg, \cdot)$	ver equação (3.32)
$K_{fg}(\cdot)$	ver equação (3.33)
$P_h$	medida pushforward da medida de Lebesgue normalizada, associada à função $h(\cdot)$
$\delta_\lambda$	distribuição delta de Dirac concentrada em $\lambda$
$\mathbb{1}_A$	função indicadora do conjunto $A$
$\mu_n \Rightarrow \mu$	$(\mu_n)_{n \in \mathbb{N}}$ converge fracamente para $\mu$
$\text{Tr}(A)$	traço da matriz $A$
$G_a(\cdot)$	ver equação (3.48)
$h^+(\cdot)$	função parte positiva da função $h(\cdot)$
$h^-(\cdot)$	função parte negativa da função $h(\cdot)$
$(f_1 \diamond \dots \diamond f_n)(\cdot)$	convolução ínfima das funções $f_1(\cdot), \dots, f_n(\cdot)$

$Z_n(c)$	estatística definida na equação (3.68)
$M_n$	matriz definida pela expressão (3.70)
$\mathcal{D}$	domínio da função $L_n(\lambda, Z_n(c))$ (ver Lema 3.29)
$\rho$	ver equação (3.79)
$\tau$	ver equação (3.79)
$\varphi_X(\cdot)$	função característica da variável aleatória $X$
$\text{sign}(\cdot)$	função sinal
$S_\alpha(\sigma, \beta, \mu)$	distribuição $\alpha$ -estável com $\alpha$ , $\sigma$ , $\beta$ e $\mu$ indicando respectivamente, o índice de estabilidade e os parâmetros de escala, simetria e locação
$S\alpha S$	distribuição simetricamente $\alpha$ -estável
$\mathcal{C}(\mu, \sigma)$	distribuição de Cauchy com parâmetro de locação $\mu$ e parâmetro de escala $\sigma$
$\tau_{X,Y}$	função codiferença das variáveis $\alpha$ -estáveis simétricas $X$ e $Y$
$\text{Cov}(X, Y)$	covariância entre as variáveis aleatórias $X$ e $Y$
$\Upsilon_{\mathbf{X}_n}$	matriz de codiferença do vetor aleatório $S\alpha S$ , $\mathbf{X}_n$
$\tau_X(k; \alpha)$	função codiferença do processo estacionário $(X_n)_{n \in \mathbb{N}}$ , com índice de estabilidade $\alpha$ , considerada no lag $k \in \mathbb{Z}$
$\vec{T}(\cdot)$	operador <i>shift</i>
$g_\alpha(\cdot)$	ver equação (4.69)
$\ A\ _\infty$	norma da matriz $A$ (ver expressão (B.2))
$A^*$	operador adjunto de $A$



# Sumário

<b>1</b>	<b>INTRODUÇÃO</b>	<b>13</b>
<b>2</b>	<b>PRINCÍPIO DE GRANDES DESVIOS</b>	<b>15</b>
2.1	Lei Forte dos Grandes Números e Teorema Central do Limite	15
2.2	Princípio de Grandes Desvios e suas Propriedades	16
2.2.1	Teorema Central do Limite como Consequência do Princípio de Grandes Desvios	18
2.2.2	Teorema de Cramér-Chernoff	19
2.2.3	Teorema de Gärtner-Ellis	26
2.2.4	Princípio da Contração	28
2.2.5	Teorema de Bahadur-Rao	29
<b>3</b>	<b>PROCESSO AR(1) COM INOVAÇÕES GAUSSIANAS</b>	<b>31</b>
3.1	Grandes Desvios e o Processo AR(1) com Inovações Gaussianas	31
3.2	LDP Para o Estimador de Yule-Walker	33
3.2.1	Matrizes de Toeplitz	34
3.2.2	Decomposição Linear da Estatística $\mathcal{W}_n(f)$	38
3.2.3	Resultados Principais	40
3.3	LDP Para o Estimador de Mínimos Quadrados	53
3.3.1	Decomposição Linear da Estatística $Z_n(c)$	54
3.3.2	Demonstração do Teorema 3.5	58
3.4	Comparação Entre os Estimadores	64
3.5	Princípio Forte de Grandes Desvios e o Processo AR(1) com Inovações Gaussianas	67
<b>4</b>	<b>PROCESSOS AR(1) COM INOVAÇÕES <math>\alpha</math>-ESTÁVEIS</b>	<b>69</b>
4.1	Distribuições $\alpha$ -Estáveis	69
4.1.1	Definições Equivalentes de Estabilidade	70
4.1.2	Funções de Densidade das Distribuições $\alpha$ -Estáveis e Grandes Desvios para a Média Amostral	72
4.2	Processo AR(1) com Inovações $\alpha$ -Estáveis Simétricas	78
4.2.1	Codiferença	82
4.2.2	Matriz de Codiferença de um Processo AR(1)	84
4.2.3	Comportamento Assintótico da Codiferença de um Processo AR(1)	87
4.2.4	Processos Mixing e Ergódicos	88

4.3	Função Densidade Espectral para o Processo AR(1) com Inovações $\alpha$ -Estáveis Simétricas . . . . .	93
4.3.1	Relação Entre a Codiferença e Função Densidade Espectral . . . . .	95
4.3.2	Caso $\alpha$ Geral . . . . .	97
4.3.3	Cálculo da Função $g_\alpha(\cdot)$ . . . . .	101
4.4	Grandes Desvios e o Processo AR(1) com Inovações $\alpha$ -Estáveis Não-Gaussianas . . . . .	104
4.4.1	Princípio de Grandes Desvios Para o Estimador de Yule-Walker na Versão $\alpha$ -Estável . . . . .	106
<b>5</b>	<b>CONCLUSÃO . . . . .</b>	<b>109</b>
	<b>REFERÊNCIAS . . . . .</b>	<b>111</b>
	<b>APÊNDICES . . . . .</b>	<b>115</b>
	<b>A – PROCESSOS AUTORREGRESSIVOS . . . . .</b>	<b>117</b>
	<b>B – ÁLGEBRA LINEAR . . . . .</b>	<b>121</b>
	<b>C – DISTRIBUIÇÕES <math>\alpha</math>-ESTÁVEIS ESPECIAIS E SUAS PROPRIEDADES . . . . .</b>	<b>129</b>
	<b>D – GRÁFICOS DA FUNÇÃO CODIFERENÇA . . . . .</b>	<b>133</b>

# 1 Introdução

O objetivo principal deste trabalho consiste no estudo do princípio de grandes desvios e de uma aplicação destes estudos à estimação do parâmetro de um processo autorregressivo de primeira ordem, com inovações  $\alpha$ -estáveis. Os processos autorregressivos desempenham um papel importante na modelagem de séries temporais e análise de dados. Esta classe especial de processos estocásticos tem como característica relevante a possibilidade de escrever o valor atual da série, por exemplo  $X_n$  em tempo  $n \in \mathbb{N}$ , como uma combinação linear dos seus próprios  $p$  valores passados  $X_{n-1}, X_{n-2}, \dots, X_{n-p}$ , no caso de um processo autorregressivo de ordem  $p$  (AR( $p$ )), conforme Definição A.2, adicionado de um processo de ruído. Utilizamos, com frequência, os resultados expostos em Brockwell e Davis (1991).

Os modelos autorregressivos de ordem  $p$ , com inovações  $\alpha$ -estáveis, são frequentemente utilizados para descrever séries temporais de caudas pesadas, que apresentam variância infinita. Por serem simples e capazes de descrever diferentes comportamentos, foram extensivamente estudados por diversos autores. Os textos clássicos que compreendem uma discussão detalhada desta teoria podem ser encontrados em Gnedenko et al. (1954), Feller (1970) e Samorodnitsky e Taqqu (2000). Dentre as aplicações de modelos autorregressivos podemos citar: tempo em redes (Resnick, 1997), temperatura da superfície do mar (Gallagher, 2001) e log-retorno do mercado de ações (Ling, 2005 e Andrews e Davis, 2013).

Este trabalho encontra-se organizado da seguinte forma: o Capítulo 2 inicia apresentando a Lei Forte dos Grandes Números e o Teorema Central do Limite. Em seguida, exibimos uma comparação entre estes dois teoremas e apresentamos o princípio de grandes desvios. Mostramos condições em que o resultado fornecido pelo Teorema Central do Limite pode ser obtido a partir do princípio de grandes desvios. Também apresentamos uma discussão sobre os resultados clássicos envolvendo o princípio de grandes desvios, como por exemplo: o Teorema de Cramér-Chernoff, o Teorema de Gärtner-Ellis e o princípio da contração. Todos estes tópicos podem ser consultados em Ellis (1985), Bucklew (1990) ou Dembo e Zeitouni (1993). Finalizamos o capítulo com uma breve discussão referente ao princípio forte de grandes desvios, introduzido por Bahadur e Rao (1960).

No Capítulo 3 estudamos o processo autorregressivo Gaussiano de primeira ordem. Consideramos dois estimadores distintos do parâmetro deste processo: o estimador de Yule-Walker e o estimador de mínimos quadrados. Provamos que ambos satisfazem um princípio de grandes desvios e exibimos a sua respectiva função taxa. Este estudo foi realizado primeiramente por Avram (1988) e Bucklew (1990). No entanto, os resultados encontrados

por estes dois autores não são consistentes. Mais tarde, Bryc e Dembo (1997) e Bercu et al. (1997) apresentaram um método diferente dos dois primeiros autores, realizando um estudo aprofundado dos autovalores do produto de duas matrizes de Toeplitz. Direcionamos nossos estudos para o método empregado por Bercu et al. (1997). Ao fim deste capítulo, apresentamos gráficos das funções taxa encontradas para os estimadores de Yule-Walker e de mínimos quadrados. Realizamos uma pequena análise destes gráficos e fazemos uma comparação entre os estimadores quanto à sua eficácia na estimação do respectivo parâmetro.

A fim de estender a abordagem utilizada por Bercu et al. (1997), no Capítulo 4 estudamos algumas propriedades do processo autorregressivo de primeira ordem, com inovações advindas de um processo  $\alpha$ -estável não-Gaussiano. Até o momento, desconhecemos qualquer estudo relacionando o princípio de grandes desvios à estimadores deste processo. As primeiras pesquisas envolvendo distribuições  $\alpha$ -estáveis e grandes desvios iniciaram com Heyde (1967), mas este autor providenciou apenas resultados para a soma de variáveis independentes e identicamente distribuídas, pertencendo ao domínio de atração de uma variável com distribuição  $\alpha$ -estável. Posteriormente, outros autores apresentaram soluções para problemas dentro deste contexto, entre eles podemos citar os mais recentes, Zaigraev (1999), Rozovskii (1999), Mikosch (2013) e referências dentro destes. Para sermos mais exatos, Mikosch (2013) apresentou resultados envolvendo o conceito de *precise large deviations* para sequências de variáveis aleatórias estacionárias com variação regular.

Ainda no Capítulo 4, mostramos que o processo autorregressivo de primeira ordem, com inovações advindas de um processo  $\alpha$ -estável não-Gaussiano, é estacionário, mixing e ergódico. Além disso, impedidos de definir a função de autocovariância para este processo, devido à falta de momentos finitos de segunda ordem, estudamos uma medida de dependência entre variáveis, denominada de função codiferença. Embora a primeira noção de codiferença tenha sido introduzida por Astrauskas (1983), preferimos adotar uma versão mais simples introduzida por Samorodnitsky e Taqqu (2000). Deduzimos a função codiferença do processo autorregressivo de primeira ordem para o caso de inovações  $\alpha$ -estáveis não-Gaussianas e mostramos que esta função converge à zero, à medida que seu *lag* converge para infinito. Ainda provamos que a matriz de codiferença deste processo pode ser reescrita através de uma matriz de Toeplitz, no sentido apresentado por Grenander e Szegö (1958). Por fim, fornecemos uma alternativa para deduzir as propriedades de grandes desvios do processo autorregressivo estudado no Capítulo 4.

As conclusões finais são apresentadas no Capítulo 5.



## 2 Princípio de Grandes Desvios

Neste capítulo apresentamos algumas relações entre a Lei dos Grandes Números, o Teorema Central do Limite e o Princípio de Grandes Desvios. Como base para os dois primeiros tópicos, sugerimos a consulta dos capítulos 5 e 7 de James (2015) àqueles leitores que não possuem tanta familiaridade com o assunto. Ainda que possam ser encontrados aqui generalizações de alguns dos exemplos, com acréscimo de outros, para o terceiro tópico seguimos de perto o texto de Bercu et al. (2015, p. 1-6). Com relação a resultados mais abrangentes sobre o Princípio de Grandes Desvios, aconselhamos a consulta de Ellis (1985), Bucklew (1990) e Dembo e Zeitouni (1993).

### 2.1 Lei Forte dos Grandes Números e Teorema Central do Limite

Considere uma sequência  $(X_n)_{n \in \mathbb{N}}$  de variáveis aleatórias independentes e identicamente distribuídas (i.i.d.), definidas no espaço de probabilidades  $(\Omega, \mathcal{A}, \mathbb{P})$ . Assuma que, para todo  $n$  natural (convencionamos em todo o trabalho que  $\mathbb{N} = \{1, 2, 3, \dots\}$ ), a variável  $X_n$  possui média  $\mu \in \mathbb{R}$  e variância  $\sigma^2 > 0$ , ambas desconhecidas. Considerado um dos teoremas mais importantes dentro da Teoria das Probabilidades, a Lei Forte dos Grandes Números (LFGN) concede uma estimativa para a média destas variáveis aleatórias.

**Teorema 2.1 (Lei Forte dos Grandes Números).** *Seja  $(X_n)_{n \in \mathbb{N}}$  uma sequência de variáveis aleatórias i.i.d., integráveis no espaço de probabilidade  $(\Omega, \mathcal{A}, \mathbb{P})$  e com média  $\mathbb{E}(X_1) = \mu$ . Considere  $\bar{X}_n = n^{-1} \sum_{k=1}^n X_k$ , a média amostral das variáveis aleatórias  $X_1, \dots, X_n$ . Então*

$$\bar{X}_n \xrightarrow{q.c.} \mu, \quad (2.1)$$

onde  $\xrightarrow{q.c.}$  significa convergência quase certa.

Embora a Lei Forte dos Grandes Números informe sobre o tipo de convergência da média amostral para o parâmetro  $\mu$ , ela infelizmente não traz muitos detalhes acerca do comportamento assintótico da sequência  $(\bar{X}_n)_{n \in \mathbb{N}}$ . Além disso, em Estatística Matemática constantemente se está interessado na determinação do intervalo de confiança do parâmetro em estimação, o que também a LFGN não é capaz de estabelecer. Em contrapartida, uma das alternativas é recorrer ao Teorema Central do Limite.

**Teorema 2.2 (Teorema Central do Limite).** *Se  $(X_n)_{n \in \mathbb{N}}$  é uma sequência de variáveis aleatórias i.i.d. com média  $\mu$  e variância  $\sigma^2 < \infty$ , então*

$$\sqrt{n}(\bar{X}_n - \mu) \xrightarrow{D} \mathcal{N}(0, \sigma^2), \quad (2.2)$$

onde  $\xrightarrow{D}$  significa convergência em distribuição.

É bem conhecido que a convergência em (2.2) pode ser usada para se obter um intervalo de confiança para a média  $\mu$  da variável aleatória  $X_n$ . De fato, para  $0 < \gamma < 1$ , um intervalo de confiança com nível de significância  $(1 - \gamma)\%$  é dado por

$$C_n = \left[ \bar{X}_n - \frac{z_{1-\gamma/2}}{\sqrt{n}} \hat{\sigma}_n, \bar{X}_n + \frac{z_{1-\gamma/2}}{\sqrt{n}} \hat{\sigma}_n \right], \quad (2.3)$$

onde  $z_{1-\gamma/2}$  representa o quantil de nível  $(1 - \gamma/2)\%$  de uma variável  $\mathcal{N}(0, 1)$  e  $\hat{\sigma}_n^2$  representa uma estimativa da variância amostral de ordem  $n$ , definida pela expressão

$$\hat{\sigma}_n^2 = \sum_{k=1}^n \frac{(X_k - \bar{X}_n)^2}{n-1}. \quad (2.4)$$

Com efeito, para  $n$  suficientemente grande, a probabilidade da estimativa do parâmetro  $\mu$  pertencer ao intervalo  $C_n$  é dada por

$$\begin{aligned} \mathbb{P} \left( \bar{X}_n - \frac{z_{1-\gamma/2}}{\sqrt{n}} \hat{\sigma}_n \leq \mu \leq \bar{X}_n + \frac{z_{1-\gamma/2}}{\sqrt{n}} \hat{\sigma}_n \right) &= \mathbb{P} \left( -z_{1-\gamma/2} \leq \sqrt{n} \left( \frac{\bar{X}_n - \mu}{\hat{\sigma}_n} \right) \leq z_{1-\gamma/2} \right) \\ &\approx 1 - \gamma, \end{aligned}$$

demonstrando que  $C_n$  é um intervalo de confiança de nível de significância  $(1 - \gamma)\%$ .

## 2.2 Princípio de Grandes Desvios e suas Propriedades

A teoria que engloba Grandes Desvios é bastante ampla. Recorremos com frequência às definições apresentadas por Bucklew (1990), Dembo e Zeitouni (1993) e Ellis (1985). Antes de mais nada, é preciso estabelecer o que significa uma função ser semi-contínua inferiormente. Adotamos abaixo a definição de Dembo e Zeitouni (1993, p. 308).

**Definição 2.3.** Considere um espaço topológico  $\mathcal{X}$ . Dizemos que uma função  $I : \mathcal{X} \rightarrow \mathbb{R}$  é *semi-contínua inferiormente* se, para qualquer  $c \in \mathbb{R}$ , o conjunto de nível  $\Psi_I(c) \equiv \{x \in \mathcal{X} : I(x) \leq c\}$  é fechado.

No contexto de grandes desvios, as funções semi-contínuas inferiormente desempenham um papel importante quando o seu contradomínio é dado pelo conjunto  $[0, \infty]$ , conforme a seguinte definição.

**Definição 2.4.** Dizemos que  $I : \mathcal{X} \rightarrow [0, \infty]$  é uma *função taxa* se for semi-contínua inferiormente. Além disso, no caso particular em que os conjuntos de níveis  $\Psi_I(c)$  forem subconjuntos compactos de  $\mathcal{X}$ , diremos que  $I(\cdot)$  é uma *boa função taxa*.

**Observação 2.5:** Alguns autores, como por exemplo Dembo e Zeitouni (1993, p. 4), Mörters (2008, p. 9) e Bercu et al. (1997, p. 76), distinguem o caso em que os conjuntos de nível  $\{x \in \mathcal{X} : I(x) \leq c\}$  da função taxa  $I(\cdot)$  são compactos, e nesta circunstância adotam a nomenclatura de *good rate function* (em português se traduziria literalmente por “boa função taxa”). Para o que pretendemos mostrar no decorrer deste trabalho, os conjuntos de nível da função  $I(\cdot)$ , acima mencionados, serão sempre compactos, e por razões de simplicidade, exigiremos que a função  $I(\cdot)$  da Definição 2.6 seja sempre uma boa função taxa, porém mantemos a nomenclatura de “função taxa”.

A vantagem de trabalharmos com boas funções taxa é de que o ínfimo sobre conjuntos fechados é sempre atingido. Além disso, se  $\mathcal{X}$  for um espaço métrico, para verificar que uma função  $I(\cdot)$  é semi-contínua inferiormente, basta que para toda sequência  $(x_n)_{n \in \mathbb{N}}$ , tal que  $\lim_{n \rightarrow \infty} x_n = x$ , se tenha

$$\liminf_{n \rightarrow \infty} I(x_n) \geq I(x),$$

para todo  $x \in \mathcal{X}$  (ver Dembo e Zeitouni, 1993, p. 4).

Como já citamos anteriormente, a LFGN afirma que, sob certas condições de regularidade, a sequência  $(\bar{X}_n)_{n \in \mathbb{N}}$  converge quase certamente para a média  $\mathbb{E}(X_1) = \mu$ . No contexto de grandes desvios, estamos preocupados com a velocidade desta convergência, e de uma certa forma, generalizar a LFGN.

Em um contexto mais geral, tal como em Bucklew (1990, p. 1), considere uma sequência de variáveis aleatórias  $(Y_n)_{n \in \mathbb{N}}$ , convergindo em probabilidade à uma constante real  $c$ , isto é,

$$\mathbb{P}(|Y_n - c| > \varsigma) \xrightarrow{n \rightarrow \infty} 0, \quad (2.5)$$

para todo  $\varsigma > 0$ . Em certos casos, a convergência em (2.5) pode ser de ordem exponencial, de tal modo que

$$\mathbb{P}(|Y_n - c| > \varsigma) \approx K(\varsigma, c, n)e^{[-nI(\varsigma, c)]}, \quad (2.6)$$

onde  $K(\cdot, \cdot, \cdot)$  é uma função de variação lenta com relação à  $n$  (veja definição da p. 6 de Bingham et al., 1987) e  $I(\cdot, \cdot)$  é uma quantidade positiva. A Teoria de Grandes Desvios se preocupa principalmente em determinar a quantidade  $I(\cdot, \cdot)$ . Se a relação (2.6) é satisfeita, dizemos que a sequência  $(Y_n)_{n \in \mathbb{N}}$  satisfaz um *princípio de grandes desvios*. A seguinte definição (ver p. 76 de Bercu et al., 1997) torna este conceito mais preciso.

**Definição 2.6.** Considere uma sequência de medidas de probabilidade  $(\mathbb{P}_n)_{n \in \mathbb{N}}$  definidas sobre o espaço de Hausdorff regular  $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$ , onde  $\mathcal{B}(\mathcal{X})$  representa a  $\sigma$ -álgebra de Borel gerada por todos os conjuntos abertos do espaço  $\mathcal{X}$ . Dizemos que a sequência  $(\mathbb{P}_n)_{n \in \mathbb{N}}$  satisfaz um *Princípio de Grandes Desvios* (que denotaremos por LDP, abreviação do inglês, *Large Deviation Principle*) com velocidade de convergência  $n$  e função taxa  $I : \mathcal{X} \mapsto [0, +\infty]$ , se  $I$  é uma função semi-contínua inferiormente tal que

a) Cota superior: para qualquer subconjunto fechado  $F \subseteq \mathcal{X}$

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}_n(F) \leq - \inf_{x \in F} I(x); \quad (2.7)$$

b) Cota inferior: para qualquer conjunto aberto  $A \subseteq \mathcal{X}$

$$- \inf_{x \in A} I(x) \leq \liminf_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}_n(A). \quad (2.8)$$

**Observação 2.7:** Na definição acima, a hipótese de que  $\mathcal{X}$  seja um espaço de Hausdorff regular tem como finalidade excluir casos triviais, tais como uma função taxa identicamente nula. Além disso, se a função taxa existe, a vantagem de trabalharmos em espaços regulares é a garantia de que tal função é única (ver lema 4.1.4 de Dembo e Zeitouni, 1993).

Existem definições bem mais abrangentes do Princípio de Grandes Desvios do que a da Definição 2.6. Apesar desta mesma definição ser uma das mais simples na respectiva literatura, ela será suficiente para o que desejamos mostrar.

Vamos trabalhar a todo instante com sequências de variáveis aleatórias, notadamente, sequências de estimadores. Portanto, quando estivermos a falar de uma sequência  $(X_n)_{n \in \mathbb{N}}$  que satisfaz um LDP, as notações  $\mathbb{P}_n(F)$  e  $\mathbb{P}(X_n \in F)$  servem para representar o mesmo conceito, uma vez que uma variável aleatória fica completamente determinada pela sua distribuição. Tentamos deixar isto evidente dentro do contexto e não faremos mais menção desta peculiaridade.

### 2.2.1 Teorema Central do Limite como Consequência do Princípio de Grandes Desvios

Sob certas condições de regularidade, é possível obter o Teorema Central do Limite como consequência do Princípio de Grandes Desvios. Bryc (1993, p. 253) mostra que, dada uma sequência de variáveis aleatórias contínuas  $\{X_t\}_{t \in [1, +\infty)}$ , então sob determinadas hipóteses, a sequência  $t^{1/2} X_t$  converge em distribuição para uma variável  $\mathcal{N}(0, \sigma^2)$ . Enunciaremos abaixo o caso em que a sequência  $\{X_t\}_{t \in [1, +\infty)}$  tem parâmetro  $t$  discreto, coincidindo assim com o resultado estabelecido pelo Teorema Central do Limite.

**Proposição 2.8.** *Suponha que  $(X_n)_{n \in \mathbb{N}}$  é uma sequência de variáveis aleatórias i.i.d. tais que  $\mathbb{E}(X_n) = 0$ ,  $\forall n \in \mathbb{N}$ . Se existe  $\varepsilon > 0$  tal que o limite*

$$L(z) = \lim_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{E} \left( e^{nzX_n} \right) \quad (2.9)$$

*existe para qualquer número complexo  $z$ , com  $|z| < \varepsilon$ , então,*

$$\sqrt{n} X_n \xrightarrow{D} \mathcal{N}(0, \sigma^2), \quad (2.10)$$

*onde  $\sigma^2 = L''(0) \geq 0$ .*

Assim como define Bryc (1993, p. 254), denote por  $Z_{e_n} \subseteq \mathbb{C}$  o conjunto das raízes da função  $\phi : z \mapsto \mathbb{E} \left( e^{nzX_n} \right)$ . Então a seguinte proposição estabelece uma condição suficiente para que o Teorema Central do Limite seja uma consequência do Princípio de Grandes Desvios.

**Proposição 2.9.** *Se para algum  $\varepsilon > 0$*

$$\sup_{n \in \mathbb{N}} [\mathbb{E} (\exp(n\varepsilon |X_n|))]^{1/n} < \infty \quad e \quad 0 \notin \overline{\bigcup_{k \in \mathbb{N}} Z_{e_k}}, \quad (2.11)$$

onde  $\bar{A}$  denota o fecho de  $A$ , então o Princípio de Grande Desvios implica o Teorema Central do Limite.

Não é corriqueiro que o Princípio de Grandes Desvios implique no Teorema Central do Limite. Para citar um contra-exemplo deste fato, referenciamos a observação 1 de Bryc (1993). Além disso, a condição exibida em (2.9), para  $z$  real, é também uma das condições exigidas para que valha o Teorema de Gärtner-Ellis, que iremos abordar na Subseção 2.2.3.

## 2.2.2 Teorema de Cramér-Chernoff

No âmbito de variáveis aleatórias independentes e identicamente distribuídas, o Teorema de Cramér-Chernoff, aqui adaptado de Bercu et al. (2015, p. 2), estabelece uma expressão para a função taxa quando a sequência avaliada é a das médias amostrais das respectivas variáveis aleatórias.

**Teorema 2.10 (Teorema de Cramér-Chernoff).** *Seja  $(X_n)_{n \in \mathbb{N}}$  uma sequência de variáveis aleatórias i.i.d.. Então a sequência  $(\bar{X}_n)_{n \in \mathbb{N}}$  satisfaz um LDP com boa função taxa*

$$I(x) = \sup_{t \in \mathbb{R}} \{xt - \mathcal{L}_{X_1}(t)\}, \quad (2.12)$$

onde  $\mathcal{L}_{X_1}(\cdot)$  é a função de log-Laplace da variável  $X_1$ , definida por

$$\mathcal{L}_{X_1}(t) = \log \left[ \mathbb{E} \left( e^{tX_1} \right) \right].$$

Informalmente, o Teorema de Cramér-Chernoff afirma que

$$\mathbb{P} (S_n \geq xn) \approx e^{nI(x)}, \quad \text{onde } S_n = \sum_{k=1}^n X_k \quad e \quad x \in \mathbb{R}.$$

**Observação 2.11:** A função  $I(\cdot)$  que aparece na expressão (2.12) é também conhecida como a transformada de Fenchel-Legendre da função  $\mathcal{L}_{X_1}(\cdot)$  e depende da familiar função geradora de momentos  $M_{X_1}(t) = \mathbb{E} \left( e^{tX_1} \right)$ . Certos autores, como por exemplo Bucklew (1990, p.183) e Rockafellar (1970, p. 104), costumam utilizar a notação  $f^*(x) = \sup_{t \in \mathbb{R}} \{xt - f(t)\}$  para denotar a transformada de Fenchel-Legendre da função real  $f(\cdot)$ .

Para ilustrar como este último teorema pode ser aplicado, os exemplos a seguir mostram casos clássicos da teoria. Bercu et al. (2015, p. 3) apenas exibem as funções taxa, sem demonstrar como elas foram obtidas.

**Exemplo 2.12. (Distribuição Bernoulli de parâmetro  $0 < p < 1$ ).** Considere uma sequência  $(X_n)_{n \in \mathbb{N}}$  de variáveis aleatórias i.i.d. seguindo uma distribuição Bernoulli de parâmetro  $p$ , isto é,  $X_n \sim \mathfrak{B}(p)$ , para  $0 < p < 1$  e  $n \in \mathbb{N}$ . Então, pelo Teorema de Cramér-Chernoff, a sequência  $(\bar{X}_n)_{n \in \mathbb{N}}$  satisfaz um LDP com função taxa

$$I(x) = \begin{cases} x \log\left(\frac{x}{p}\right) + (1-x) \log\left(\frac{1-x}{1-p}\right), & \text{se } x \in [0, 1], \\ +\infty, & \text{caso contrário.} \end{cases} \quad (2.13)$$

**Observação 2.13:** Nos casos particulares em que  $x \in \{0, 1\}$ , convencionamos que  $0 \times \log(0) = 0$ . Em adição, Müller (2016, p. 20-21) provou apenas o caso em que  $p = 1/2$ .

De fato, fixe  $\mathbb{P}(X_1 = 0) = 1 - p$  e  $\mathbb{P}(X_1 = 1) = p$ , para  $p \in (0, 1)$ . Desta forma, a função geradora de momentos da variável aleatória  $X_1$ , calculada em  $t \in \mathbb{R}$ , é dada por

$$M_{X_1}(t) = \sum_{k \in \{0,1\}} e^{kx} \mathbb{P}(X_1 = k) = \mathbb{P}(X_1 = 0) + e^t \mathbb{P}(X_1 = 1) = 1 + p(e^t - 1),$$

de maneira que a equação (2.12) pode ser escrita como

$$I(x) = \sup_{t \in \mathbb{R}} [tx - \mathcal{L}_{X_1}(t)] = \sup_{t \in \mathbb{R}} [tx - \log(1 + p(e^t - 1))]. \quad (2.14)$$

Com o intuito de simplificar a prova do resultado, consideramos  $0 < p < 1$  como uma constante fixa e seccionamos o domínio de  $I(\cdot)$  nos respectivos intervalos:

i)  $x < 0$ :

Neste primeiro caso, como a função  $-\log(1 + p(e^t - 1))$  é estritamente decrescente, temos que

$$\lim_{t \rightarrow -\infty} -\log(1 + p(e^t - 1)) = -\log(1 - p).$$

Disto, segue imediatamente que

$$\lim_{t \rightarrow -\infty} tx - \log(1 + p(e^t - 1)) = +\infty,$$

de onde podemos deduzir que  $I(x) = +\infty$ , para  $x < 0$ .

ii)  $x > 1$ :

Neste caso, observe que a função  $I(\cdot)$  em (2.14) pode ser reescrita como

$$I(x) = \sup_{t \in \mathbb{R}} \left[ \log\left(\frac{e^{tx}}{1 + p(e^t - 1)}\right) \right] = \sup_{t \in \mathbb{R}} \left[ \log\left(\frac{e^{t(x-1)}}{e^{-t}(1-p) + p}\right) \right]. \quad (2.15)$$

Uma vez que  $x > 1$ , então  $x - 1 > 0$ . Assim,

$$\lim_{t \rightarrow +\infty} e^{t(x-1)} = +\infty \quad \text{e} \quad \lim_{t \rightarrow +\infty} e^{-t}(1-p) + p = p.$$

Logo, a função

$$\log \left( \frac{e^{t(x-1)}}{e^{-t}(1-p) + p} \right)$$

é ilimitada superiormente, mostrando que também  $I(x) = +\infty$ , para  $x > 1$ .

iii)  $0 \leq x \leq 1$  :

Pela equação (2.15), para  $x = 0$  temos

$$I(0) = \sup_{t \in \mathbb{R}} \left[ \log \left( \frac{1}{1 + p(e^t - 1)} \right) \right].$$

Dado que  $\log \left( \frac{1}{1 + p(e^t - 1)} \right)$  é decrescente e

$$\lim_{t \rightarrow -\infty} 1 + p(e^t - 1) = 1 - p,$$

então segue que  $I(0) = \log \left( \frac{1}{1-p} \right)$ .

Analogamente, para  $x = 1$ , a equação (2.15) estabelece que

$$I(1) = \sup_{t \in \mathbb{R}} \left[ \log \left( \frac{1}{e^{-t}(1-p) + p} \right) \right].$$

Sabendo que  $\log \left( \frac{1}{e^{-t}(1-p) + p} \right)$  é uma função crescente e que

$$\lim_{t \rightarrow +\infty} e^{-t}(1-p) + p = p,$$

concluimos que  $I(1) = \log \left( \frac{1}{p} \right)$ .

Para o caso em que  $0 < x < 1$ , defina a função

$$\psi(t) = \log \left( \frac{e^{tx}}{1 + p(e^t - 1)} \right).$$

O cálculo da primeira e segunda derivada de  $\psi(\cdot)$  fornecem

$$\psi'(t) = x - \frac{pe^t}{1 + p(e^t - 1)} \quad \text{e} \quad \psi''(t) = -\frac{p(1-p)e^t}{(1 + p(e^t - 1))^2}.$$

Da hipótese de que  $0 < p < 1$ , é imediato que  $\psi''(t) < 0$ . Além disso, a primeira derivada de  $\psi(\cdot)$  é identicamente nula se, e somente se,

$$t = \log \left[ \frac{x(1-p)}{p(1-x)} \right].$$

Assim, é bem conhecido da Análise Infinitesimal que a função  $\psi(\cdot)$  tem um ponto de máximo em  $t_0 = \log\left(\frac{x(1-p)}{p(1-x)}\right)$ . Aproveitando este resultado, para  $0 < x < 1$ , obtemos que

$$\begin{aligned}
I(x) &= \sup_{t \in \mathbb{R}} [tx - \log(1 + p(e^t - 1))] = t_0 x - \log(1 + p(e^{t_0} - 1)) \\
&= x \log\left(\frac{x(1-p)}{p(1-x)}\right) - \log\left[1 + p\left(\frac{x(1-p)}{p(1-x)} - 1\right)\right] \\
&= x \log\left(\frac{x}{p}\right) + x \log\left(\frac{1-p}{1-x}\right) - \log\left(1 - p + \frac{x(1-p)}{1-x}\right) \\
&= x \log\left(\frac{x}{p}\right) + x \log\left(\frac{1-p}{1-x}\right) - \log\left(\frac{(1-p)(1-x) + x(1-p)}{1-x}\right) \\
&= x \log\left(\frac{x}{p}\right) + x \log\left(\frac{1-p}{1-x}\right) - \log\left(\frac{1-p}{1-x}\right) \\
&= x \log\left(\frac{x}{p}\right) - x \log\left[\left(\frac{1-p}{1-x}\right)^{-1}\right] + \log\left[\left(\frac{1-p}{1-x}\right)^{-1}\right] \\
&= x \log\left(\frac{x}{p}\right) + (1-x) \log\left(\frac{1-x}{1-p}\right).
\end{aligned}$$

□

**Exemplo 2.14. (Distribuição Poisson de parâmetro  $\lambda > 0$ ).** Seja  $(Y_n)_{n \in \mathbb{N}}$  uma sequência de variáveis aleatórias i.i.d. com distribuição Poisson de parâmetro  $\lambda > 0$ , isto é,  $Y_n \sim \mathcal{P}(\lambda)$ ,  $\forall n \in \mathbb{N}$ . Então temos

$$I(x) = \begin{cases} x \log\left(\frac{x}{\lambda}\right) - x + \lambda, & \text{se } x \in [0, +\infty), \\ +\infty, & \text{caso contrário.} \end{cases}$$

Com efeito, para cada  $n \in \mathbb{N}$  temos

$$\mathbb{P}(Y_n = k) = \begin{cases} \frac{e^{-\lambda} \lambda^k}{k!}, & \text{se } k \in \mathbb{Z}^+, \\ 0, & \text{caso contrário,} \end{cases}$$

onde  $\mathbb{Z}^+$  representa o conjunto dos inteiros não-negativos. A função geradora de momentos da variável aleatória  $Y_1$  é então calculada através de

$$M_{Y_1}(t) = \sum_{k \in \mathbb{Z}^+} e^{tk} \mathbb{P}(Y_1 = k) = e^{-\lambda} \sum_{k \in \mathbb{Z}^+} \frac{(e^t \lambda)^k}{k!} = e^{\lambda(e^t - 1)}.$$

Assim, a função taxa apresentada em (2.12) é fornecida por

$$I(x) = \sup_{t \in \mathbb{R}} \{xt - \log(e^{\lambda(e^t - 1)})\} = \sup_{t \in \mathbb{R}} \{xt - \lambda e^t + \lambda\}. \quad (2.16)$$

Repetindo os passos do Exemplo 2.12, note que, se



i)  $x < 0$  :

Observe que quando  $x < 0$  a função  $xt - \lambda e^t$  é decrescente e

$$\lim_{t \rightarrow -\infty} xt - \lambda e^t = +\infty.$$

Logo,  $I(x) = +\infty$ , para  $x < 0$ .

ii)  $x = 0$  :

Se  $x = 0$ , então  $I(x) = \sup_{t \in \mathbb{R}} \{-\lambda e^t + \lambda\}$ . Como  $-\lambda e^t$  é decrescente e  $\lim_{t \rightarrow -\infty} -\lambda e^t = 0$ , segue que  $I(0) = \lambda$ .

iii)  $x > 0$  :

Já quando  $x > 0$ , definindo  $\psi(\cdot)$  como

$$\psi(t) = xt - \lambda e^t + \lambda,$$

as primeiras derivadas desta função são dadas por

$$\psi'(t) = x - \lambda e^t \quad \text{e} \quad \psi''(t) = -\lambda e^t < 0.$$

Igualando a primeira derivada de  $\psi(\cdot)$  à 0, concluímos que  $\psi(\cdot)$  tem um ponto de máximo em  $t_0 = \log\left(\frac{x}{\lambda}\right)$ . Por fim, para  $x > 0$ , temos que

$$I(x) = \sup_{t \in \mathbb{R}} \{xt - \lambda e^t + \lambda\} = xt_0 - \lambda e^{t_0} + \lambda = x \log\left(\frac{x}{\lambda}\right) - x + \lambda.$$

□

**Exemplo 2.15. (Distribuição Gama de parâmetros  $a$  e  $b$ ).** Considere uma sequência  $(X_n)_{n \in \mathbb{N}}$  de variáveis aleatórias i.i.d., seguindo uma distribuição Gama com parâmetros  $a$  e  $b$ , ambos positivos, denotada por  $\Gamma(a, b)$ . Então, pelo Teorema de Crámer-Chernoff, a sequência  $(\bar{X}_n)_{n \in \mathbb{N}}$  satisfaz um LDP com função taxa

$$I(x) = \begin{cases} \frac{x}{b} - a \left[ 1 - \log(a) + \log\left(\frac{x}{b}\right) \right], & \text{se } x > 0, \\ +\infty, & \text{caso contrário.} \end{cases}$$

Com efeito, de acordo com a definição encontrada em Rohatgi e Saleh (2000, p. 212), a função densidade de uma variável aleatória que segue uma distribuição  $\Gamma(a, b)$  é dada por

$$f(x) = \begin{cases} \frac{1}{\Gamma(a)b^a} x^{a-1} e^{-x/b}, & \text{se } 0 < x < \infty, \\ 0, & \text{se } x \leq 0, \end{cases} \quad (2.17)$$

onde  $\Gamma(a) = \int_0^{+\infty} x^{a-1} e^{-x} dx$ , representa a função gama calculada em  $a > 0$ . A função geradora de momentos da variável aleatória  $X_1$  é dada por

$$M_{X_1}(t) = \mathbb{E}(e^{tX_1}) = \int_0^{+\infty} e^{tx} \frac{1}{\Gamma(a)b^a} x^{a-1} e^{-x/b} dx = \frac{1}{\Gamma(a)b^a} \int_0^{+\infty} x^{a-1} e^{-x(\frac{1-bt}{b})} dx. \quad (2.18)$$

Realizando uma análise do domínio da função  $M_{X_1}(\cdot)$ , obtemos dois casos distintos.

i) Se  $t \geq 1/b$ , então, para  $x > 0$  obtemos

$$0 \leq -x \left( \frac{1-bt}{b} \right) \Rightarrow 1 = e^0 \leq e^{-x(\frac{1-bt}{b})},$$

de onde segue que

$$\int_0^{+\infty} x^{a-1} dx \leq \int_0^{+\infty} x^{a-1} e^{-x(\frac{1-bt}{b})} dx. \quad (2.19)$$

Observe agora que, da hipótese de que  $a > 0$ , temos que  $1-a < 1$ , e portanto, a integral imprópria  $\int_1^{+\infty} \frac{1}{x^{1-a}} dx$  diverge. Pela expressão (2.19), concluímos que, para  $t \geq 1/b$ ,  $M_{X_1}(t) = +\infty$ .

ii) No caso em que  $t < 1/b$ , a função geradora de momentos está bem definida. Realizando a mudança de variáveis  $x \left( \frac{1-bt}{b} \right) = y$ , temos que  $dx = \left( \frac{b}{1-bt} \right) dy$ . Assim, pela equação (2.18), segue que

$$\begin{aligned} M_{X_1}(t) &= \frac{1}{\Gamma(a)b^a} \int_0^{+\infty} x^{a-1} e^{-x(\frac{1-bt}{b})} dx \\ &= \frac{1}{\Gamma(a)b^a} \int_0^{+\infty} \left( \frac{b}{1-bt} \right)^{a-1} y^{a-1} e^{-y} \frac{b}{1-bt} dy \\ &= \frac{1}{\Gamma(a)(1-bt)^a} \int_0^{+\infty} y^{a-1} e^{-y} dy = \frac{(1-bt)^{-a}}{\Gamma(a)} \Gamma(a) = (1-bt)^{-a}. \end{aligned}$$

Note agora que não é preciso considerar os casos em que  $\mathcal{L}_{X_1}(t) = \log[M_{X_1}(t)] = +\infty$ , nomeadamente, quando  $t \geq 1/b$ . De fato, dado que  $\mathcal{L}_{X_1}(0) = 0$ , temos por definição de supremo que

$$I(x) \geq 0 = 0x - \mathcal{L}_{X_1}(0).$$

Como para  $t \geq 1/b$  temos sempre que  $tx - \mathcal{L}_{X_1}(t) = -\infty$ , segue então que

$$I(x) = \sup_{t \in \mathbb{R}} [tx + a \log(1-bt)] = \sup_{t < 1/b} [tx + a \log(1-bt)]. \quad (2.20)$$

Tal como no Exemplo 2.12, é preciso separar o domínio de  $I(\cdot)$  em intervalos.

i)  $x \leq 0$ :

Se  $x = 0$ , como  $\log(1-bt)$  é decrescente e  $\lim_{t \rightarrow -\infty} \log(1-bt) = +\infty$ , concluímos que  $I(0) = +\infty$ .

Se  $x < 0$ , então  $tx$  é decrescente e também

$$\lim_{t \rightarrow -\infty} [tx + a \log(1 - bt)] = +\infty,$$

o que mostra que  $I(x) = +\infty$  para  $x \leq 0$ .

ii)  $x > 0$  :

Repare que  $I(\cdot)$  em (2.20) pode ser reescrita como

$$I(x) = \sup_{t < 1/b} \left[ \log \left( e^{tx} (1 - bt)^a \right) \right]. \quad (2.21)$$

Considere a função  $\psi : (-\infty, 1/b) \mapsto \mathbb{R}$ , definida pela lei

$$\psi(t) = \log \left( e^{tx} (1 - bt)^a \right) = tx + a \log(1 - bt).$$

Os cálculos da primeira e segunda derivada de  $\psi(\cdot)$  fornecem

$$\psi'(t) = x - \frac{ab}{1 - bt} \quad \text{e} \quad \psi''(t) = -\frac{ab^2}{(1 - bt)^2}.$$

Das hipóteses de que  $a, b > 0$  e  $t < 1/b$ , segue que  $\psi''(t) < 0$ . Além disso, a primeira derivada de  $\psi(\cdot)$  é identicamente nula se, e somente se,

$$t = \frac{x - ab}{xb}.$$

Portanto, a função  $\psi(\cdot)$  tem um ponto de máximo em  $t_0 = \frac{x - ab}{xb}$ , e conseqüentemente, para  $x > 0$  obtemos

$$\begin{aligned} I(x) &= \sup_{t < 1/b} \left\{ \log \left[ e^{tx} (1 - bt)^a \right] \right\} = \log \left[ e^{t_0 x} (1 - bt_0)^a \right] \\ &= \left( \frac{x - ab}{xb} \right) x + a \log \left[ 1 - b \left( \frac{x - ab}{xb} \right) \right] = \frac{x}{b} - a + a \log \left( 1 - \frac{xb}{xb} + \frac{ab}{x} \right) \\ &= \frac{x}{b} - a \left\{ 1 - \log \left[ a \left( \frac{x}{b} \right)^{-1} \right] \right\} = \frac{x}{b} - a \left[ 1 - \log(a) + \log \left( \frac{x}{b} \right) \right]. \end{aligned}$$

□

### Observação 2.16:

(a) Seja  $(X_n)_{n \in \mathbb{N}}$  uma sequência de variáveis aleatórias i.i.d. seguindo uma distribuição exponencial de parâmetro  $b$ , isto é,  $X_n \sim \mathcal{E}(b)$ ,  $\forall n \in \mathbb{N}$ . Adotando a definição de Rohatgi e Saleh (2000, p. 215), para cada  $n \in \mathbb{N}$ , as variáveis aleatórias  $X_n$  tem função de densidade dada por

$$f(x) = \begin{cases} \frac{1}{b} e^{-x/b}, & \text{se } x > 0, \\ 0, & \text{se } x \leq 0. \end{cases} \quad (2.22)$$

Observe que a lei da função em (2.22) apresenta semelhanças com a que é exibida em (2.17). De fato, para  $a = 1$  em (2.17), notamos que a distribuição exponencial é um caso particular de uma distribuição  $\Gamma(1, b)$ , e com esta observação concluímos que a sequência  $(\bar{X}_n)_{n \in \mathbb{N}}$  satisfaz um LDP com função taxa

$$I(x) = \begin{cases} \frac{x}{b} - 1 - \log\left(\frac{x}{b}\right), & \text{se } x > 0, \\ +\infty, & \text{caso contrário.} \end{cases}$$

A menos da parametrização, este resultado coincide com o que é mostrado em Müller (2016, p. 24).

- (b) Outra distribuição utilizada com frequência na Estatística é a distribuição qui-quadrado com  $n$  graus de liberdade, denotada por  $\mathcal{X}_n^2$ . Esta distribuição é mais um caso particular da distribuição Gama e tem como parâmetros  $a = n/2$  e  $b = 2$ , isto é,  $\mathcal{X}_n^2 \stackrel{d}{=} \Gamma(n/2, 2)$ . Como consequência deste fato, pelo que mostramos no Exemplo 2.15, dada uma sequência  $(X_n)_{n \in \mathbb{N}}$  de variáveis aleatórias i.i.d. seguindo uma distribuição  $\mathcal{X}_n^2$ , a sequência  $(\bar{X}_n)_{n \in \mathbb{N}}$  satisfaz um LDP com função taxa

$$I(x) = \begin{cases} \frac{x}{2} - \frac{n}{2} \left[ 1 - \log\left(\frac{n}{2}\right) + \log\left(\frac{x}{2}\right) \right], & \text{se } x > 0, \\ +\infty, & \text{caso contrário.} \end{cases}$$

### 2.2.3 Teorema de Gärtner-Ellis

O Teorema de Cramér-Chernoff, apresentado na Subseção 2.2.2, pode ser empregado apenas no caso em que as variáveis aleatórias forem independentes e identicamente distribuídas. Em casos mais concretos, podemos nos deparar com modelos cujas variáveis aleatórias são dependentes entre si, como é o caso de um processo autorregressivo de primeira ordem (veja o Apêndice A).

O Teorema de Gärtner-Ellis surge como alternativa ao caso em que as variáveis aleatórias não são independentes entre si. Este mesmo teorema não leva em consideração a estrutura de dependência do conjunto das variáveis e baseia-se apenas na sequência das funções geradoras de momentos. Vamos utilizar seu resultado no Capítulo 3, para mostrar propriedades de grandes desvios para o processo autorregressivo de primeira ordem com inovações Gaussianas.

Apresentamos a seguir, de acordo com a seção II.C de Bucklew (1990), a notação e as hipóteses necessárias para o Teorema de Gärtner-Ellis. Ressaltamos que a versão apresentada por Bucklew (1990) ainda não é a mais completa. Para uma versão mais sofisticada deste teorema, juntamente de sua demonstração, sugerimos a consulta da seção 2.3 de Dembo e Zeitouni (1993), mais especificamente, o teorema 2.3.6.

**Definição 2.17.** Considere uma função real estendida  $f : \mathbb{R} \mapsto \mathbb{R} \cup \{\infty\}$ .

- (a) Vamos definir o *domínio de  $f(\cdot)$*  como sendo o conjunto  $D_f = \{x \in \mathbb{R} : f(x) < \infty\}$ .
- (b) Assuma que  $f(\cdot)$  é diferenciável no seu domínio. Dizemos que  $f(\cdot)$  é uma *função íngreme* (*steep function*, no inglês) se, para toda sequência de pontos  $(x_n)_{n \in \mathbb{N}} \in D_f$  e  $x_0 \in D_f$ , tal que  $\lim_{n \rightarrow \infty} x_n = x_0$ , temos que

$$\lim_{n \rightarrow \infty} |f'(x_n)| = \infty. \quad (2.23)$$

**Definição 2.18.** Considere uma sequência de variáveis aleatórias  $(X_n)_{n \in \mathbb{N}}$ . A *função geradora de cumulantes normalizada*  $L_n(\lambda, X_n)$  da variável  $X_n$ , é definida como

$$L_n(\lambda, X_n) = \frac{1}{n} \log \mathbb{E} \left( e^{n\lambda X_n} \right). \quad (2.24)$$

**Hipótese 2.19.** Considere a sequência de variáveis aleatórias  $(X_n)_{n \in \mathbb{N}}$ . Para cada  $\lambda \in \mathbb{R}$ , assumamos que o limite

$$L_X(\lambda) \equiv \lim_{n \rightarrow \infty} L_n(\lambda, X_n) \quad (2.25)$$

existe, podendo assumir o valor  $\infty$ . Além disso, exigimos que  $L_X(\cdot)$  seja uma função convexa e semi-contínua inferiormente.

**Hipótese 2.20.** Seja  $D_{L_X} \equiv \{\lambda \in \mathbb{R} : L_X(\lambda) < \infty\}$  o domínio da função  $L_X(\cdot)$ . Denote por  $D_{L_X}^\circ$  o interior do conjunto  $D_{L_X}$ . Então, vamos assumir que  $0 \in D_{L_X}^\circ$ .

**Hipótese 2.21.**  $L_X(\cdot)$  definida em (2.25) é uma função íngreme em  $D_{L_X}$ .

Estabelecidas todas as hipóteses que embasam o Teorema de Gärtner-Ellis, estamos em condições de enunciar o mesmo.

**Teorema 2.22 (Teorema de Gärtner-Ellis).** *Seja  $(X_n)_{n \in \mathbb{N}}$  uma sequência de variáveis aleatórias definida no espaço de probabilidade  $(\mathcal{X}, \mathcal{B}(\mathcal{X}), \mathbb{P})$ , com  $L_X(\cdot)$  dada por (2.25). Além disso, denote por  $L_X^*(x) = \sup_{t \in \mathbb{R}} \{xt - L_X(t)\}$  a transformada de Fenchel-Legendre da função  $L_X(\cdot)$ .*

(a) *Suponha que vale a Hipótese 2.19, então, para todo conjunto compacto  $F \subseteq \mathcal{X}$ , temos*

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log [\mathbb{P}(X_n \in F)] \leq - \inf_{x \in F} L_X^*(x). \quad (2.26)$$

(b) *Se além da Hipótese 2.19, a Hipótese 2.20 também for válida, então o limite apresentado na expressão (2.26) também se verifica para qualquer conjunto  $F$  fechado.*

(c) *Por fim, se as hipóteses 2.19 e 2.21 forem válidas, então, para todo conjunto aberto  $G \subseteq \mathcal{X}$ , temos*

$$\liminf_{n \rightarrow \infty} \frac{1}{n} \log [\mathbb{P}(X_n \in G)] \geq - \inf_{x \in G} L_X^*(x). \quad (2.27)$$

### 2.2.4 Princípio da Contração

O teorema que vamos apresentar a seguir (ver teorema 4.2.1 de Dembo e Zeitouni, 1993) é útil quando precisamos determinar novas funções taxa a partir de “antigas”. Substancialmente, se tivermos um LDP para a sequência de variáveis aleatórias  $(X_n)_{n \in \mathbb{N}}$ , definidas no espaço  $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$ , e em seguida aplicarmos uma função contínua  $h : \mathcal{X} \mapsto \mathcal{Y}$ , seremos capazes de deduzir um novo princípio de grandes desvios para a sequência de variáveis aleatórias  $(h(X_n))_{n \in \mathbb{N}}$ , definidas no espaço  $(\mathcal{Y}, \mathcal{B}(\mathcal{Y}))$ , desde que a medida de probabilidade considerada neste segundo espaço seja a induzida pela função  $h(\cdot)$ .

**Teorema 2.23 (Princípio da Contração).** *Sejam  $\mathcal{X}$  e  $\mathcal{Y}$  espaços de Hausdorff e  $h : \mathcal{X} \mapsto \mathcal{Y}$  uma função contínua. Considere uma boa função taxa  $I : \mathcal{X} \mapsto [0, \infty]$ .*

(a) *Para cada  $y \in \mathcal{Y}$ , defina*

$$J(y) = \inf\{I(x) : x \in \mathcal{X} \text{ e } y = h(x)\}. \quad (2.28)$$

*Então  $J(\cdot)$  é uma boa função taxa em  $\mathcal{Y}$ , onde por convenção, o ínfimo sobre o conjunto vazio é tomado como  $\infty$ .*

(b) *Se  $(\mathbb{P}_n)_{n \in \mathbb{N}}$  é uma sequência de medidas de probabilidade satisfazendo um LDP com função taxa  $I(\cdot)$ , então a sequência  $(h(\mathbb{P}_n))_{n \in \mathbb{N}}$  satisfaz um LDP com função taxa  $J(\cdot)$ , dada por (2.28).*

*Demonstração.* Consulte Dembo e Zeitouni (1993, p. 110-111). □

O seguinte corolário foi extraído do exercício 4.2.7 de Dembo e Zeitouni (1993) e estende o princípio da contração.

**Corolário 2.24.** *Suponha que  $\mathcal{X}$  é um espaço regular separável e que, para todo  $n \in \mathbb{N}$ ,  $(X_n, Y_n)$  está distribuído de acordo com a medida produto  $\mathbb{P}_n \times \mathbb{Q}_n$  em  $\mathcal{B}(\mathcal{X}) \times \mathcal{B}(\mathcal{X})$  (nomeadamente,  $X_n$  é independente de  $Y_n$ ), onde  $\mathbb{P}_n$  e  $\mathbb{Q}_n$  são respectivamente, as distribuições de  $X_n$  e  $Y_n$ , para  $n \in \mathbb{N}$ . Assuma que  $(\mathbb{P}_n)_{n \in \mathbb{N}}$  satisfaz um LDP com função taxa  $I_X(\cdot)$ , enquanto que  $(\mathbb{Q}_n)_{n \in \mathbb{N}}$  satisfaz um LDP com função taxa  $I_Y(\cdot)$  e ambas  $(\mathbb{P}_n)_{n \in \mathbb{N}}$  e  $(\mathbb{Q}_n)_{n \in \mathbb{N}}$  são exponencialmente rígidas (veja p. 8 de Dembo e Zeitouni, 1993). Então, para toda função contínua  $h : \mathcal{X} \times \mathcal{X} \mapsto \mathcal{Y}$ , a família de leis induzidas em  $\mathcal{Y}$  por  $Z_n = h(X_n, Y_n)$  satisfaz um LDP com boa função taxa*

$$I_Z(z) = \inf_{\{(x,y): z=h(x,y)\}} \{I_X(x) + I_Y(y)\}. \quad (2.29)$$

*Demonstração.* Além de fazer uso do Princípio da Contração, este corolário pode ser mostrado usando o fato de que  $\mathcal{B}(\mathcal{X}) \times \mathcal{B}(\mathcal{X}) = \mathcal{B}(\mathcal{X} \times \mathcal{X})$  (teorema D.4 de Dembo e Zeitouni, 1993) e os teoremas 4.1.11 e 4.1.18 de Dembo e Zeitouni (1993). □

### 2.2.5 Teorema de Bahadur-Rao

Dada uma sequência de variáveis aleatórias  $(X_n)_{n \in \mathbb{N}}$  i.i.d. e sua respectiva sequência de médias amostrais  $(\bar{X}_n)_{n \in \mathbb{N}} = (n^{-1} \sum_{k=1}^n X_k)_{n \in \mathbb{N}}$ , se a sequência  $(\bar{X}_n)_{n \in \mathbb{N}}$  satisfaz um LDP, a estimativa através do princípio de grandes desvios da probabilidade  $\mathbb{P}(\bar{X}_n \geq c)$ , para  $c > 0$ , não é ainda a mais precisa, uma vez que as cotas superior e inferior, exibidas respectivamente nas equações (2.7) e (2.8), são dadas por um logaritmo para a probabilidade  $\mathbb{P}(\bar{X}_n \geq c)$ .

Parafraseando Bercu et al. (2000, p. 2), Bahadur e Rao (1960) provam uma expansão completa desta cauda de probabilidade. Eles estabeleceram um princípio forte de grandes desvios para a sequência  $(\bar{X}_n)_{n \in \mathbb{N}}$ , no sentido da Definição 2.26 abaixo.

**Observação 2.25:** Na Definição 2.26, usamos a notação de Landau  $o(\cdot)$ . Lembramos que, dadas duas funções inteiras  $f(\cdot)$  e  $g(\cdot)$ , então  $f(n) = o(g(n))$  equivale a dizer que

$$\lim_{n \rightarrow \infty} \frac{f(n)}{g(n)} = 0. \quad (2.30)$$

Outra notação útil que vamos usar, é a notação de Landau  $\mathcal{O}(\cdot)$ . Esta notação significa que, dadas duas funções inteiras  $f(\cdot)$  e  $g(\cdot)$ , então  $f(n) = \mathcal{O}(g(n))$  equivale a dizer que existe uma constante  $c \in \mathbb{R}$  tal que

$$\lim_{n \rightarrow \infty} \frac{f(n)}{g(n)} = c. \quad (2.31)$$

**Definição 2.26.** Seja  $(Z_n)_{n \in \mathbb{N}}$  uma sequência de variáveis aleatórias reais, convergindo quase certamente para algum número real  $z$ . Dizemos que  $(Z_n)_{n \in \mathbb{N}}$  satisfaz um *Princípio Forte de Grandes Desvios Localmente* (LSLDP, cuja abreviação surge do inglês, *Local Sharp Large Deviation Principle*) de ordem  $p \in \mathbb{N}$ , no ponto  $c \in \mathbb{R}$ , sempre que existirem uma sequência  $d_0(c), d_1(c), \dots, d_p(c) \in \mathbb{R}$  e  $b(c) \in \mathbb{R}$ , dependendo apenas de  $c$ , e tal que as seguintes expansões são satisfeitas:

$$\mathbb{P}(Z_n \geq c) = \frac{d_0(c) \exp(-nb(c))}{\sqrt{n}} \left( 1 + \frac{d_1(c)}{n} + \dots + \frac{d_p(c)}{n^p} + o\left(\frac{1}{n^p}\right) \right), \quad \text{para } c > z,$$

ou

$$\mathbb{P}(Z_n \leq c) = \frac{d_0(c) \exp(-nb(c))}{\sqrt{n}} \left( 1 + \frac{d_1(c)}{n} + \dots + \frac{d_p(c)}{n^p} + o\left(\frac{1}{n^p}\right) \right), \quad \text{para } c < z.$$

Além disso, dizemos que  $(Z_n)_{n \in \mathbb{N}}$  satisfaz um *Princípio Forte de Grandes Desvios* (SLDP) de ordem  $p \in \mathbb{N}$ , se satisfizer um LSLDP de ordem  $p$ , para todo ponto  $c \in \mathbb{R}$ .

Tendo em vista a Definição 2.26, podemos estabelecer um teorema similar ao de Cramér-Chernoff, mas desta vez com um SLDP no lugar de LDP (ver teorema 1.7 de Bercu et al., 2015).

**Teorema 2.27 (Teorema de Bahadur-Rao).** *Considere uma sequência  $(X_n)_{n \in \mathbb{N}}$  de variáveis aleatórias i.i.d.. Assuma que a função de log-Laplace  $\mathcal{L}_{X_1}(t) = \log \left[ \mathbb{E} \left( e^{tX_1} \right) \right]$  da variável  $X_1$  é finita para todo  $t \in \mathbb{R}$  e que a distribuição de  $X_1$  é absolutamente contínua. Então a sequência  $(\bar{X}_n)_{n \in \mathbb{N}}$  satisfaz um princípio forte de grandes desvios. Em particular, para todo  $x \in \mathbb{R}$ , existe uma sequência  $(d_k(x))_{k \in \mathbb{N}}$  tal que, para qualquer  $p \geq 1$  e  $n$  suficientemente grande, se  $x > \mathbb{E}(X_1)$*

$$\mathbb{P} \left( \bar{X}_n \geq x \right) = \frac{\exp(-nI(x))}{\sigma_x t_x \sqrt{2\pi n}} \left[ 1 + \sum_{k=1}^p \frac{d_k(x)}{n^k} + \mathcal{O} \left( \frac{1}{n^{p+1}} \right) \right], \quad (2.32)$$

enquanto que, para  $x < \mathbb{E}(X_1)$

$$\mathbb{P} \left( \bar{X}_n \leq x \right) = -\frac{\exp(-nI(x))}{\sigma_x t_x \sqrt{2\pi n}} \left[ 1 + \sum_{k=1}^p \frac{d_k(x)}{n^k} + \mathcal{O} \left( \frac{1}{n^{p+1}} \right) \right], \quad (2.33)$$

onde o valor de  $t_x$  é determinado por  $x = \mathcal{L}'_{X_1}(t_x)$  e  $\sigma_x^2 = \mathcal{L}''_{X_1}(t_x)$ . Finalmente, todos os coeficientes  $d_k(\cdot)$  podem ser explicitamente calculados como funções das derivadas da função  $\mathcal{L}_{X_1}(\cdot)$ , calculadas no ponto  $t_x$ .



## 3 Processo AR(1) com Inovações Gaussianas

Neste capítulo vamos estudar um caso particular de processos AR(1), cujas inovações (ou ruído) são variáveis aleatórias com distribuição normal padrão  $\mathcal{N}(0, 1)$ . A seguir, definimos os estimadores de Yule-Walker e de mínimos quadrados do parâmetro  $\theta$  deste processo (veja equação (3.1)) e estudamos as propriedades de grandes desvios de tais estimadores. Estas propriedades foram inicialmente propostas por Bercu et al. (1997), sendo posteriormente estendidas em Bercu et al. (2000) (princípio forte de grandes desvios) e Bercu (2001) (caso instável, com  $|\theta| = 1$  e caso explosivo, onde  $|\theta| > 1$ ).

### 3.1 Grandes Desvios e o Processo AR(1) com Inovações Gaussianas

Neste capítulo consideramos o processo real autorregressivo  $(X_n)_{n \in \mathbb{N}}$  de primeira ordem, satisfazendo a equação

$$X_{n+1} = \theta X_n + \varepsilon_{n+1}, \quad (3.1)$$

onde  $|\theta| < 1$  e  $(\varepsilon_n)_{n \geq 2}$  é uma sequência de variáveis aleatórias i.i.d., normalmente distribuídas com média 0 e variância 1, isto é,  $\varepsilon_n \sim \mathcal{N}(0, 1)$ , para  $n \geq 2$ . Vamos assumir que  $X_1$  é independente da sequência  $(\varepsilon_n)_{n \geq 2}$  e que  $X_1$  tem distribuição  $\mathcal{N}(0, (1 - \theta^2)^{-1})$ . Com estas condições iniciais impostas, não é difícil mostrar que  $X_n \sim \mathcal{N}(0, (1 - \theta^2)^{-1})$ , para todo  $n \in \mathbb{N}$  (semelhante a prova da Proposição 4.11). Além disso,  $(X_n)_{n \in \mathbb{N}}$  é um processo estacionário (ver prova da Proposição 4.12).

**Observação 3.1:** Para esclarecer o fato de termos adotado que  $X_1 \sim \mathcal{N}(0, (1 - \theta^2)^{-1})$ , a variância  $(1 - \theta^2)^{-1}$  é a única que torna o processo  $(X_n)_{n \in \mathbb{N}}$  identicamente distribuído. O leitor pode verificar esta afirmação, doravante às propriedades da distribuição normal.

A partir do Teorema da Densidade Espectral de um processo AR(p) (ver teorema 4.4.2 de Brockwell e Davis, 1991), concluímos que a função densidade espectral do processo  $(X_n)_{n \in \mathbb{N}}$ , definido na equação (3.1), é dada por

$$g(\omega) = \frac{1}{1 + \theta^2 - 2\theta \cos(\omega)}, \quad -\pi \leq \omega < \pi \text{ e } |\theta| < 1. \quad (3.2)$$

**Observação 3.2:** Note que a função  $g(\cdot)$  em (3.2) é sempre positiva, estando portanto bem definida. De fato, usando a relação trigonométrica  $\cos^2(\omega) + \sin^2(\omega) = 1$ , observamos

que

$$1 + \theta^2 - 2\theta \cos(\omega) = (\theta - \cos(\omega))^2 + \sin^2(\omega) > 0,$$

pois,  $\sin^2(\omega) = 0$  no intervalo  $[-\pi, \pi)$ , apenas quando  $\omega = 0$  ou  $\omega = -\pi$ , e para estes valores de  $\omega$ , temos

$$\theta - \cos(-\pi) = \theta + 1 > 0 \quad \text{e} \quad \theta - \cos(0) = \theta - 1 < 0.$$

O objetivo principal deste capítulo consiste em mostrar que, os estimadores de  $\theta$  através do método de mínimos quadrados e das equações de Yule-Walker, satisfazem um princípio de grandes desvios. Para tal, temos como base os artigos de Bercu et al. (2000), Bercu (2001) e principalmente Bercu et al. (1997). A definição seguinte determina de forma precisa quem são os estimadores de Yule-Walker e de mínimos quadrados do parâmetro  $\theta$  que aparece na expressão (3.1). Para uma demonstração de como estes estimadores podem ser obtidos, referimos o leitor ao Apêndice A.

**Definição 3.3.** Considere uma amostra aleatória  $X_1, \dots, X_n$  do processo autorregressivo  $(X_n)_{n \in \mathbb{N}}$  definido pela equação (3.1). Considere também o parâmetro  $\theta$  que figura nesta mesma equação, então:

a) o *estimador de Yule-Walker* de ordem  $n$  do parâmetro  $\theta$  é definido por

$$\tilde{\theta}_n = \frac{\sum_{k=2}^n X_k X_{k-1}}{\sum_{k=1}^n X_k^2}, \quad \text{se } n \geq 2; \quad (3.3)$$

b) o *estimador de mínimos quadrados* de ordem  $n$  do parâmetro  $\theta$  é definido por

$$\hat{\theta}_n = \frac{\sum_{k=2}^n X_k X_{k-1}}{\sum_{k=1}^{n-1} X_k^2}, \quad \text{se } n \geq 2. \quad (3.4)$$

Os estimadores apresentados nas equações (3.3) e (3.4) possuem fórmulas semelhantes, à exceção do termo que aparece no denominador. Observe que  $X_n^2$  aparece no denominador da fração dada em (3.3), enquanto que na expressão (3.4) não.

O comportamento assintótico de ambos os estimadores é bem conhecido. Para a sequência de estimadores de Yule-Walker, o teorema 8.1.1 de Brockwell e Davis (1991) mostra que

$$\sqrt{n}(\tilde{\theta}_n - \theta) \Rightarrow \mathcal{N}(0, 1 - \theta^2).$$

Analogamente, para a sequência de estimadores de mínimos quadrados, o exemplo 8.8.1 de Brockwell e Davis (1991) mostra que  $(\hat{\theta}_n)_{n \geq 2}$  e  $(\tilde{\theta}_n)_{n \geq 2}$  compartilham as mesmas propriedades assintóticas, isto é,

$$\sqrt{n}(\hat{\theta}_n - \theta) \Rightarrow \mathcal{N}(0, 1 - \theta^2).$$

Além disso, temos consistência forte para ambos os estimadores, isto é,

$$\hat{\theta}_n \xrightarrow{q.c.} \theta \quad \text{e} \quad \tilde{\theta}_n \xrightarrow{q.c.} \theta. \quad (3.5)$$

Sugerimos a consulta de Lai e Wei (1983) para a consistência de  $(\hat{\theta}_n)_{n \geq 2}$  e Mann e Wald (1943) para a consistência de  $(\tilde{\theta}_n)_{n \geq 2}$ .

Da convergência quase certa, apresentada em (3.5), podemos inferir que  $\tilde{\theta}_n$  e  $\hat{\theta}_n$  convergem ambos em probabilidade para  $\theta$  (ver proposição 5.1 de James, 2015). Assim, dentro do contexto de grandes desvios, somos indagados a responder a seguinte pergunta: “Qual a taxa de convergência das probabilidades  $\mathbb{P}(|\tilde{\theta}_n - \theta| > \varsigma)$  e  $\mathbb{P}(|\hat{\theta}_n - \theta| > \varsigma)$  para 0”? O Teorema 3.4 a seguir, apresenta um LDP para o estimador de Yule-Walker definido pela equação (3.3) (ver proposição 1 de Bercu et al., 1997).

**Teorema 3.4.** *Considere o processo  $(X_n)_{n \in \mathbb{N}}$  definido pela equação (3.1) e assumamos que  $X_1$  é independente de  $(\varepsilon_n)_{n \geq 2}$ , com  $X_1 \sim \mathcal{N}(0, (1 - \theta^2)^{-1})$ . Então, a sequência  $(\tilde{\theta}_n)_{n \geq 2}$ , para  $\tilde{\theta}_n$  definido como em (3.3), satisfaz um princípio de grandes desvios com velocidade  $n$  e função taxa*

$$S(c) = \begin{cases} \frac{1}{2} \log \left( \frac{1 + \theta^2 - 2\theta c}{1 - c^2} \right), & \text{se } c \in (-1, 1), \\ +\infty, & \text{caso contrário.} \end{cases} \quad (3.6)$$

Com condições semelhantes, é possível demonstrar um resultado análogo ao do Teorema 3.4 para o estimador de mínimos quadrados  $\hat{\theta}_n$ , apresentado na equação (3.4).

**Teorema 3.5.** *Nas condições do Teorema 3.4, a sequência  $(\hat{\theta}_n)_{n \geq 2}$ , para  $\hat{\theta}_n$  definido em (3.4), satisfaz um princípio de grandes desvios com velocidade  $n$  e função taxa*

$$R(c) = \begin{cases} \frac{1}{2} \log \left( \frac{1 + \theta^2 - 2\theta c}{1 - c^2} \right), & \text{se } c \in [a, b], \\ \log |\theta - 2c|, & \text{caso contrário,} \end{cases} \quad (3.7)$$

onde  $a$  e  $b$  satisfazem a equação  $2x^2 - \theta x - 1 = 0$ , isto é,

$$a = \frac{\theta - \sqrt{\theta^2 + 8}}{4} \quad \text{e} \quad b = \frac{\theta + \sqrt{\theta^2 + 8}}{4}. \quad (3.8)$$

## 3.2 LDP Para o Estimador de Yule-Walker

Para demonstrar os Teoremas 3.4 e 3.5, precisamos estabelecer primeiramente alguns resultados preliminares. Começamos por estudar as propriedades de grandes desvios para a probabilidade de  $\tilde{\theta}_n$  ser maior do que ou igual à uma constante real  $c$ , isto é,  $\mathbb{P}(\tilde{\theta}_n \geq c)$ . Pela expressão de  $\tilde{\theta}_n$ ,

$$\mathbb{P}(\tilde{\theta}_n \geq c) = \mathbb{P}\left(\frac{\sum_{k=2}^n X_k X_{k-1}}{\sum_{k=1}^n X_k^2} \geq c\right) = \mathbb{P}(\mathcal{W}_n(f) \geq 0),$$

onde  $\mathcal{W}_n(f)$  é a estatística definida por

$$\mathcal{W}_n(f) = \frac{1}{n} \left( \sum_{k=2}^n X_k X_{k-1} - c \sum_{k=1}^n X_k^2 \right), \quad \text{se } n \geq 2. \quad (3.9)$$

A notação  $\mathcal{W}_n(f)$  tem sua origem devido a equivalência das expressões (3.9) e

$$\mathcal{W}_n(f) = \frac{1}{2\pi n} \int_{\mathbb{T}} f(t) \left| \sum_{k=1}^n X_k e^{ikt} \right|^2 dt, \quad (3.10)$$

onde durante todo este capítulo, fixamos  $f(\cdot)$  como sendo a função definida por

$$f(t) \equiv \cos(t) - c, \quad \text{para } t \in \mathbb{T} = [-\pi, \pi]. \quad (3.11)$$

Repare que, tanto a função  $f(\cdot)$  como a estatística  $\mathcal{W}_n(f)$ , dependem do parâmetro  $c$ . A partir da desigualdade de Cauchy-Schwarz, podemos restringir  $c$  ao intervalo  $(-1, 1)$ , dado que  $|\tilde{\theta}_n| < 1$ , para todo  $n \geq 2$ . De fato, sendo  $X_1, \dots, X_n$  uma amostra aleatória, assumamos que pelo menos um dos elementos da amostra é não nulo. Sem perda de generalidade, seja  $X_1$  este elemento, então, pela desigualdade de Cauchy-Schwarz temos que

$$\begin{aligned} \left( \sum_{k=2}^n X_k X_{k-1} \right)^2 &\leq \left( \sum_{k=2}^n X_k^2 \right) \left( \sum_{k=2}^n X_{k-1}^2 \right) < \left( \sum_{k=1}^n X_k^2 \right)^2 \Rightarrow \left| \sum_{k=2}^n X_k X_{k-1} \right| < \left| \sum_{k=1}^n X_k^2 \right| \\ \Rightarrow |\tilde{\theta}_n| &= \frac{|\sum_{k=2}^n X_k X_{k-1}|}{|\sum_{k=1}^n X_k^2|} < 1. \end{aligned} \quad (3.12)$$

A mesma propriedade vale para o estimador de mínimos quadrados. Dada uma amostra aleatória  $X_1, \dots, X_n$ , suponhamos sem perda de generalidade que

$$X_1^2 \geq X_2^2 \geq \dots \geq X_n^2 \quad (3.13)$$

e que exista um  $k \in \{1, \dots, n\}$  tal que  $X_k \neq 0$ . Logo,  $X_1^2 > 0$  e da mesma forma que obtivemos (3.12), segue de (3.13) que

$$\begin{aligned} \left( \sum_{k=2}^n X_k X_{k-1} \right)^2 &\leq \left( \sum_{k=2}^n X_k^2 \right) \left( \sum_{k=1}^{n-1} X_k^2 \right) < \left( \sum_{k=1}^{n-1} X_k^2 \right)^2 \Rightarrow \left| \sum_{k=2}^n X_k X_{k-1} \right| < \left| \sum_{k=1}^{n-1} X_k^2 \right| \\ \Rightarrow |\hat{\theta}_n| &= \frac{|\sum_{k=2}^n X_k X_{k-1}|}{|\sum_{k=1}^{n-1} X_k^2|} < 1, \quad \text{para todo } n \geq 2. \end{aligned} \quad (3.14)$$

### 3.2.1 Matrizes de Toeplitz

Existe ainda uma maneira equivalente de escrever  $\mathcal{W}_n(f)$  como um produto matricial, nomeadamente

$$\mathcal{W}_n(f) = \frac{1}{n} \mathbf{X}_n^T T_n(f) \mathbf{X}_n, \quad (3.15)$$

onde  $\mathbf{X}_n = (X_1, X_2, \dots, X_n)$  é o vetor contendo como coordenadas as variáveis aleatórias  $X_1, \dots, X_n$ , que são solução da equação (3.1),  $\mathbf{X}_n^T$  é o transposto de  $\mathbf{X}_n$  e  $T_n(f)$  é a matriz de Toeplitz associada à  $f(\cdot)$ , cuja expressão é fornecida pela seguinte definição.

**Definição 3.6.** Para qualquer função real  $h \in L^\infty(\mathbb{T})$ , definida no toro  $\mathbb{T} = [-\pi, \pi)$ , a matriz de Toeplitz associada à função  $h(\cdot)$  é dada por

$$T_n(h) = [T_n(h)_{l,j}]_{1 \leq l, j \leq n} = \left[ \frac{1}{2\pi} \int_{\mathbb{T}} \exp[i(l-j)t] h(t) dt \right]_{1 \leq l, j \leq n}, \quad (3.16)$$

onde  $T_n(h)_{l,j}$  denota o elemento de  $T_n(h)$  posicionado na linha  $l$  e coluna  $j$ , para  $1 \leq l, j \leq n$ .

**Observação 3.7:** Com a finalidade de recordar o leitor, note que o espaço  $L^\infty(\mathbb{T})$ , que aparece na definição acima, é nada mais que o espaço  $L^\infty(\mathbb{T}, \mathcal{B}(\mathbb{T}), \nu)$  de todas as classes de equivalência de funções  $\mathcal{B}(\mathbb{T})$ -mensuráveis, limitadas  $\nu$ -q.t.p. (quase toda parte), duas funções sendo equivalentes quando forem iguais  $\nu$ -q.t.p. (ver p. 60 de Bartle, 1995). Observamos ainda que,  $\mathcal{B}(\mathbb{T})$  denota a  $\sigma$ -álgebra de Borel sobre o toro  $\mathbb{T}$  e que  $\nu(\cdot)$  é a medida de Lebesgue normalizada sobre  $\mathcal{B}(\mathbb{T})$  (ver Definição 3.20). Além disso, se  $h \in L^\infty(\mathbb{T})$  e  $N \in \mathcal{B}(\mathbb{T})$ , com  $\nu(N) = 0$ , definimos a norma de  $h(\cdot)$  em  $L^\infty(\mathbb{T})$  como

$$\|h\|_\infty = \inf\{ {}_h\mathbf{S}(N) : N \in \mathcal{B}(\mathbb{T}), \nu(N) = 0 \}, \quad (3.17)$$

para  ${}_h\mathbf{S}(\cdot)$  o supremo dos valores absolutos de  $h(\cdot)$  no complementar de  $N$ , isto é,

$${}_h\mathbf{S}(N) = \sup\{|h(x)| : x \notin N\}.$$

**Exemplo 3.8. (Matriz de Toeplitz da função  $f(t) = \cos(t) - c$ ).** Com base na expressão (3.16), não é difícil mostrar que a matriz de Toeplitz associada à função  $f(t) = \cos(t) - c$  é dada por

$$T_n(f) = \frac{1}{2} \begin{bmatrix} -2c & 1 & 0 & \dots \\ 1 & -2c & 1 & \dots \\ \dots & \dots & \dots & \dots \\ \dots & 1 & -2c & 1 \\ \dots & 0 & 1 & -2c \end{bmatrix}. \quad (3.18)$$

De fato, como as funções  $\cos(\cdot)$  e  $\sin(\cdot)$  são respectivamente par e ímpar, observamos primeiramente que a matriz  $T_n(f)$  é simétrica, isto é,  $T_n(f)_{l,j} = T_n(f)_{j,l}$ , para  $1 \leq l, j, \leq n$ . Portanto, se  $l = j$ , utilizando a fórmula integral dada em (3.16), obtemos que

$$T_n(f)_{l,j} = \frac{1}{2\pi} \int_{\mathbb{T}} \exp\{0\} f(t) dt = \frac{1}{2\pi} \int_{\mathbb{T}} (\cos(t) - c) dt = \frac{1}{2\pi} [\sin(t) - ct]_{-\pi}^{\pi} = -c.$$

Se  $|l - j| = 1$ , então

$$\begin{aligned} T_n(f)_{l,j} &= \frac{1}{2\pi} \int_{\mathbb{T}} \exp(it) f(t) dt = \frac{1}{2\pi} \int_{\mathbb{T}} (\cos(t) + i \sin(t)) (\cos(t) - c) dt \\ &= \frac{1}{2\pi} \int_{\mathbb{T}} (\cos^2(t) - c \cos(t) + i \sin(t) \cos(t) - ic \sin(t)) dt = \frac{1}{2\pi} \int_{\mathbb{T}} \cos^2(t) dt \\ &= \frac{1}{2\pi} \int_{\mathbb{T}} \frac{1 + \cos(2t)}{2} dt = \frac{1}{2}. \end{aligned}$$

Por fim, nos casos em que  $|l - j| = k$ , para  $2 \leq k \leq n - 1$ , obtemos que

$$\begin{aligned} T_n(f)_{l,j} &= \frac{1}{2\pi} \int_{\mathbb{T}} \exp(ikt) f(t) dt = \frac{1}{2\pi} \int_{\mathbb{T}} (\cos(kt) + i \sin(kt)) (\cos(t) - c) dt \\ &= \frac{1}{2\pi} \int_{\mathbb{T}} (\cos(kt) \cos(t) - c \cos(kt) + i \sin(kt) \cos(t) - ic \sin(kt)) dt \\ &= \frac{1}{2\pi} \int_{\mathbb{T}} \cos(kt) \cos(t) dt = \frac{1}{2\pi} \int_{\mathbb{T}} \frac{\cos[(k+1)t] + \cos[(k-1)t]}{2} dt = 0. \end{aligned}$$

□

**Exemplo 3.9. (Matriz de Toeplitz da função  $g(t) = (1 + \theta^2 - 2\theta \cos t)^{-1}$ ).** Considere agora a função  $g(t) = (1 + \theta^2 - 2\theta \cos t)^{-1}$ , onde  $|\theta| < 1$  e  $t \in \mathbb{T}$ . Já observamos que a função  $g(\cdot)$  (ver expressão (3.2)) é a função densidade espectral de um processo AR(1) Gaussiano, com inovações  $(\varepsilon_n)_{n \geq 2} \sim \mathcal{N}(0, 1)$ . Para calcular a matriz de Toeplitz associada à  $g(\cdot)$ , devemos analisar a expressão

$$T_n(g) = \left( \frac{1}{2\pi} \int_{\mathbb{T}} \frac{\exp[i(l-j)t]}{1 + \theta^2 - 2\theta \cos t} dt \right)_{1 \leq l, j \leq n}. \quad (3.19)$$

Se por acaso  $\theta = 0$ , então  $T_n(g)$  é a matriz identidade. Assumimos então que  $\theta \neq 0$ . Para calcular a integral do  $(l, j)$ -ésimo elemento da matriz na expressão (3.19), devemos analisar as integrais

$$\frac{1}{2\pi} \int_{\mathbb{T}} \frac{\cos(kt)}{1 + \theta^2 - 2\theta \cos(t)} dt \quad \text{e} \quad \frac{1}{2\pi} \int_{\mathbb{T}} \frac{\sin(kt)}{1 + \theta^2 - 2\theta \cos(t)} dt.$$

A segunda integral acima é identicamente nula, enquanto que a primeira, de acordo com a equação (3.613)(2) em Gradshteyn e Ryzhik (1965, p. 366), tem valor

$$\frac{1}{2\pi} \int_{\mathbb{T}} \frac{\cos(kt)}{1 + \theta^2 - 2\theta \cos(t)} dt = \frac{1}{2\pi} \frac{2\pi\theta^k}{1 - \theta^2} = \frac{\theta^k}{1 - \theta^2}, \quad \forall k \in \mathbb{Z}^+. \quad (3.20)$$

Como a função cosseno é par, podemos inferir que o resultado da expressão (3.20) vale para qualquer  $k$  inteiro, com

$$\frac{1}{2\pi} \int_{\mathbb{T}} \frac{\cos(kt)}{1 + \theta^2 - 2\theta \cos(t)} dt = \frac{\theta^{|k|}}{1 - \theta^2}, \quad \forall k \in \mathbb{Z}.$$

Assim, concluímos que a matriz  $T_n(g)$  é determinada pela forma

$$T_n(g) = \left( \frac{\theta^{|l-j|}}{1 - \theta^2} \right)_{1 \leq l, j \leq n}. \quad (3.21)$$

□

**Observação 3.10:** Existe ainda uma segunda maneira de se calcular a matriz  $T_n(g)$ . Observe que sendo  $g(t) = (1 + \theta^2 - 2\theta \cos t)^{-1}$  a função densidade espectral associada

ao processo  $(X_n)_{n \in \mathbb{N}}$ , através do corolário do Teorema de Herglotz (ver Teorema 4.37), concluímos que  $\gamma_X(\cdot)$  é a função de autocovariância do processo  $(X_n)_{n \in \mathbb{N}}$  se, e somente se,

$$\gamma_X(h) = \int_{(-\pi, \pi]} e^{ih\omega} g(\omega) d\omega, \quad \forall h \in \mathbb{Z}.$$

Equivalentemente,

$$T_n(g) = \left( \frac{1}{2\pi} \int_{\mathbb{T}} \exp[i(l-j)t] g(t) dt \right)_{1 \leq l, j \leq n} = (\gamma_X(l-j))_{1 \leq l, j \leq n}.$$

Desta forma,  $T_n(g)$  é completamente definida pela função de autocovariância do processo  $(X_n)_{n \in \mathbb{N}}$ . Da maneira que o processo  $(X_n)_{n \in \mathbb{N}}$  foi definido, podemos reescrever  $X_n$  como

$$X_n = \theta^{n-1} X_1 + \sum_{k=2}^n \theta^{n-k} \varepsilon_k, \quad \forall n \geq 2. \quad (3.22)$$

Se  $l = j$ , para  $1 \leq l, j \leq n$ , então é evidente que

$$\gamma_X(l-j) = \gamma_X(0) = \mathbb{E}(X_1^2) = \frac{1}{1-\theta^2}.$$

Se  $j > l$ ,  $X_j$  pode ser reescrito como

$$X_j = \theta^{j-l} X_l + \sum_{k=1}^{j-l} \theta^{j-l-k} \varepsilon_{l+k}. \quad (3.23)$$

Usando as expressões (3.22) e (3.23), o fato de que  $(\varepsilon_n)_{n \geq 2}$  é uma sequência de variáveis independentes e por hipótese,  $X_1$  é independente de  $(\varepsilon_n)_{n \geq 2}$ , a função de autocovariância do processo é dada por

$$\begin{aligned} \gamma_X(j-l) &= \mathbb{E}(X_j X_l) = \mathbb{E} \left( \left[ \theta^{j-l} X_l + \sum_{k=1}^{j-l} \theta^{j-l-k} \varepsilon_{l+k} \right] X_l \right) \\ &= \theta^{j-l} \mathbb{E}(X_l^2) + \mathbb{E} \left( X_l \sum_{k=1}^{j-l} \theta^{j-l-k} \varepsilon_{l+k} \right) \\ &= \frac{\theta^{j-l}}{1-\theta^2} + \mathbb{E} \left( \left[ \theta^{l-1} X_1 + \sum_{k=2}^l \theta^{l-k} \varepsilon_k \right] \sum_{k=1}^{j-l} \theta^{j-l-k} \varepsilon_{l+k} \right) \\ &= \frac{\theta^{j-l}}{1-\theta^2}. \end{aligned}$$

No caso em que  $j < l$ , basta lembrar que a função de autocovariância de um processo estacionário é sempre par (ver proposição 1.5.1 de Brockwell e Davis, 1991), ou seja, temos  $\gamma_X(j-l) = \gamma_X(l-j) = \frac{\theta^{l-j}}{1-\theta^2}$ .  $\square$

Ambos os métodos citados acima, tanto no Exemplo 3.9 como na Observação 3.10, produzem a matriz

$$T_n(g) = (1-\theta^2)^{-1} \begin{pmatrix} 1 & \theta & \theta^2 & \dots & \theta^{n-1} \\ \theta & 1 & \theta & \dots & \theta^{n-2} \\ \theta^2 & \theta & 1 & \dots & \theta^{n-3} \\ \dots & \dots & \dots & \dots & \dots \\ \theta^{n-1} & \theta^{n-2} & \theta^{n-3} & \dots & 1 \end{pmatrix}.$$

### 3.2.2 Decomposição Linear da Estatística $\mathcal{W}_n(f)$

Mostramos implicitamente na Subseção 3.2.1 que, as matrizes de autocovariância de ordem  $n$  do processo  $(X_n)_{n \in \mathbb{N}}$  definido pela equação (3.1), coincidem com a matriz de Toeplitz associada a função densidade espectral  $g(\cdot)$  do processo  $(X_n)_{n \in \mathbb{N}}$ , onde  $g(\cdot)$  é a função definida na expressão (3.2). Assim, como uma distribuição normal multivariada é unicamente determinada por sua média e matriz de autocovariância (ver proposição 1.6.4 de Brockwell e Davis, 1991), é imediata a seguinte definição.

**Definição 3.11.** Considere  $(X_n)_{n \in \mathbb{N}}$ , o processo definido pela equação (3.1). Então, o vetor aleatório  $\mathbf{X}_n = (X_1, \dots, X_n)$ , segue uma distribuição normal  $n$ -variada com média  $\mathbf{0}$  e matriz de autocovariância  $T_n(g)$ , isto é,  $\mathbf{X}_n \sim \mathcal{N}(\mathbf{0}, T_n(g))$ , onde  $g(\cdot)$  é a função densidade espectral do processo  $(X_n)_{n \in \mathbb{N}}$ , definida pela expressão (3.2).

Uma vez estabelecido que  $\mathbf{X}_n \sim \mathcal{N}(\mathbf{0}, T_n(g))$ , dois resultados importantes podem ser inferidos da teoria de distribuições normais multivariadas.

**Proposição 3.12.** Considere o vetor  $\mathbf{X}_n = (X_1, \dots, X_n)$  e a matriz de Toeplitz  $T_n(g)$  associada à densidade espectral  $g(\cdot)$  do processo  $(X_n)_{n \in \mathbb{N}}$ , onde  $g(\cdot)$  é dada pela expressão (3.2).

- (a) Dado que  $T_n(g)$  é a matriz de autocovariância do vetor  $\mathbf{X}_n$ , então é uma matriz simétrica e não-negativa definida.
- (b) A matriz  $T_n(g)$  pode ser utilizada para reescrever o vetor aleatório  $\mathbf{X}_n$  como um produto entre a matriz raiz quadrada  $T_n^{1/2}(g)$ , de ordem  $n$  (ver Definição B.6), e um vetor aleatório  $\mathbf{N}_n$  com distribuição normal padrão  $n$ -variada, isto é, existe  $\mathbf{N}_n \sim \mathcal{N}(\mathbf{0}, I_n)$ , onde  $I_n$  representa a matriz identidade de ordem  $n$ , tal que

$$\mathbf{X}_n = T_n^{1/2}(g)\mathbf{N}_n \quad (3.24)$$

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

*Demonstração.* Para o item (a), consulte a proposição 1.6.2 em Brockwell e Davis (1991). Enquanto que a demonstração do item (b) pode ser consultada no Apêndice B.  $\square$

Na Observação B.8 mostramos implicitamente que a matriz  $T_n^{1/2}(g)$  é simétrica. Deste resultado, da expressão (3.15) e do fato que  $\mathbf{X}_n = T_n^{1/2}(g)\mathbf{N}_n$ , obtemos que

$$\begin{aligned} \mathcal{W}_n(f) &= \frac{1}{n} \mathbf{X}_n^T T_n(f) \mathbf{X}_n = \frac{1}{n} (T_n^{1/2}(g)\mathbf{N}_n)^T T_n(f) (T_n^{1/2}(g)\mathbf{N}_n) \\ &= \frac{1}{n} \mathbf{N}_n^T T_n^{1/2}(g) T_n(f) T_n^{1/2}(g) \mathbf{N}_n, \end{aligned} \quad (3.25)$$

e com base nesta equação, o seguinte lema fornece uma forma equivalente de escrever a estatística  $\mathcal{W}_n(f)$ , como uma combinação linear entre variáveis aleatórias i.i.d., com



distribuição  $\chi_1^2$  (ver Observação 2.16 (b)) e cujos coeficientes desta combinação são os autovalores da matriz  $T_n^{1/2}(g)T_n(f)T_n^{1/2}(g)$ .

**Lema 3.13.** *Sejam  $\lambda_1^n, \dots, \lambda_n^n$  os autovalores de  $T_n^{1/2}(g)T_n(f)T_n^{1/2}(g)$ , então*

$$\mathcal{W}_n(f) = \frac{1}{n} \sum_{k=1}^n \lambda_k^n Z_k^n, \quad (3.26)$$

onde  $Z_1^n, \dots, Z_n^n$  são variáveis aleatórias i.i.d. com distribuição  $\chi_1^2$  (qui-quadrado com 1 grau de liberdade).

*Demonstração.* Uma vez que  $T_n(f)$  e  $T_n^{1/2}(g)$  são matrizes simétricas, o produto

$$T_n^{1/2}(g)T_n(f)T_n^{1/2}(g)$$

é também simétrico. Então, pela Proposição B.5, existe uma matriz  $P_n$  ortogonal tal que

$$P_n^T T_n^{1/2}(g)T_n(f)T_n^{1/2}(g)P_n = \Lambda_n \quad (3.27)$$

onde  $\Lambda_n = \text{diag}(\lambda_1^n, \dots, \lambda_n^n)$ .

Como  $P_n$  é ortogonal, então, para  $\mathbf{N}_n \sim \mathcal{N}(\mathbf{0}, I_n)$ , o vetor aleatório  $\mathbf{Y}_n = P_n^T \mathbf{N}_n = (Y_1, \dots, Y_n)$  tem distribuição normal padrão  $n$ -variada. De fato, pela linearidade da esperança, a média de  $\mathbf{Y}_n$  é igual ao vetor nulo e além disso, denotando por  $\Sigma_{\mathbf{Y}_n \mathbf{Y}_n}$  a matriz de autocovariância de  $\mathbf{Y}_n$ , obtemos que (veja proposição 1.6.1 de Brockwell e Davis, 1991)

$$\Sigma_{\mathbf{Y}_n \mathbf{Y}_n} = P_n^T \Sigma_{\mathbf{N}_n \mathbf{N}_n} (P_n^T)^T = P_n^T P_n = I_{n \times n}.$$

Assim, segue das equações (3.25) e (3.27) que

$$\begin{aligned} \mathcal{W}_n(f) &= \frac{1}{n} \mathbf{N}_n^T T_n^{1/2}(g)T_n(f)T_n^{1/2}(g) \mathbf{N}_n = \frac{1}{n} (P_n^T \mathbf{N}_n)^T \Lambda_n P_n^T \mathbf{N}_n \\ &= \frac{1}{n} \mathbf{Y}_n^T \text{diag}(\lambda_1^n, \dots, \lambda_n^n) \mathbf{Y}_n = \frac{1}{n} \sum_{k=1}^n \lambda_k^n Y_k^2 = \frac{1}{n} \sum_{k=1}^n \lambda_k^n Z_k^n, \end{aligned}$$

onde a última igualdade segue do fato que  $Z_k^n = Y_k^2$  tem distribuição  $\chi_1^2$ , para  $1 \leq k \leq n$ , dado que as variáveis  $Y_1, \dots, Y_n$  são i.i.d. com distribuição normal padrão (ver teorema 1 da página 324 de Rohatgi e Saleh, 2000).  $\square$

**Observação 3.14:** O Lema 3.13 acima, continua válido se a matriz  $T_n(f)$  for substituída por uma matriz real simétrica  $M_n$ , visto que o produto  $T_n^{1/2}(g)M_n T_n^{1/2}(g)$  ainda é uma matriz simétrica. Este resultado será necessário para a demonstração do Teorema 3.5.

### 3.2.3 Resultados Principais

O método de rescrever a estatística  $\mathcal{W}_n(f)$  como uma combinação linear de variáveis aleatórias i.i.d., seguindo distribuições qui-quadrado, é o ponto chave da abordagem apresentada por Bercu et al. (1997). Como podemos ver a seguir, vamos usar a função geradora de cumulantes normalizada da estatística  $\mathcal{W}_n(f)$  (ver Definição 2.18), para determinar a função taxa associada à sequência  $(\mathcal{W}_n(f))_{n \geq 2}$ .

**Definição 3.15.** Seja  $Z$  uma variável aleatória com distribuição qui-quadrado com um grau de liberdade, isto é,  $Z \sim \chi_1^2$ . A função geradora de momentos de  $Z$  é dada por

$$M_Z(t) = \mathbb{E} \left( e^{tZ} \right) = (1 - 2t)^{-1/2}. \quad (3.28)$$

A partir deste fato, da independência das variáveis  $Z_1^n, \dots, Z_n^n$  e do Lema 3.13, podemos concluir que a função geradora de cumulantes normalizada da estatística  $\mathcal{W}_n(f)$  é dada por

$$\begin{aligned} L_n(\lambda, \mathcal{W}_n(f)) &= \frac{1}{n} \log \mathbb{E} \left( e^{n\lambda \mathcal{W}_n(f)} \right) = \frac{1}{n} \log \mathbb{E} \left( \exp \left\{ \lambda \sum_{k=1}^n \lambda_k^n Z_k^n \right\} \right) \\ &= \frac{1}{n} \log \mathbb{E} \left( \prod_{k=1}^n \exp \{ \lambda \lambda_k^n Z_k^n \} \right) = \frac{1}{n} \sum_{k=1}^n \log \mathbb{E} \left( \exp \{ \lambda \lambda_k^n Z_k^n \} \right) \\ &= \frac{1}{n} \sum_{k=1}^n \log (1 - 2\lambda \lambda_k^n)^{-1/2} = -\frac{1}{2n} \sum_{k=1}^n \log (1 - 2\lambda \lambda_k^n), \end{aligned} \quad (3.29)$$

onde por convenção, assumimos que  $\log(z) = -\infty$ , se  $z \leq 0$ .

**Lema 3.16.** *Sejam  $f, g \in L^\infty(\mathbb{T})$ . Defina para cada  $n \in \mathbb{N}$ ,  $\underline{a}_n(f, g)$  e  $\bar{a}_n(f, g)$  como sendo o menor e o maior dos  $n$  autovalores da matriz  $T_n(f)T_n(g)$ , respectivamente. Então, as sequências  $(\underline{a}_n(f, g))_{n \in \mathbb{N}}$  e  $(\bar{a}_n(f, g))_{n \in \mathbb{N}}$  são ambas limitadas por  $\|f\|_\infty \|g\|_\infty$ .*

A demonstração do Lema 3.16 conta com o lema 1 de Avram (1988) e resultados específicos de Álgebra Linear. O leitor interessado na prova, pode consultá-la no Apêndice B. Observe ainda que os autovalores da matriz  $T_n(f)T_n(g)$  coincidem com os da matriz  $T_n^{1/2}(g)T_n(f)T_n^{1/2}(g)$  (ver Lema B.7), de tal modo que é indiferente falarmos dos autovalores de uma ou de outra, das duas matrizes citadas.

**Definição 3.17.** Considere uma função  $h \in L^\infty(\mathbb{T})$ . Dizemos que  $M_h$  é o *supremo essencial* da função  $h : \mathbb{T} \rightarrow \mathbb{R}$  se

$$M_h = \inf \{ a \in \mathbb{R} : \nu(\{x \in \mathbb{T} : h(x) > a\}) = 0 \}. \quad (3.30)$$

Da mesma forma, o *ínfimo essencial* da função  $h(\cdot)$ , denotado por  $m_h$ , pode ser definido como

$$m_h = \sup \{ a \in \mathbb{R} : \nu(\{x \in \mathbb{T} : h(x) < a\}) = 0 \}. \quad (3.31)$$

Pelo Lema 3.16, sendo  $(\underline{a}_n(f, g))_{n \in \mathbb{N}}$  e  $(\bar{a}_n(f, g))_{n \in \mathbb{N}}$  seqüências limitadas, tem sentido a seguinte definição.

**Definição 3.18.** Sejam  $f, g \in L^\infty(\mathbb{T})$ . Assim como no Lema 3.16, considere  $\underline{a}_n(f, g)$  e  $\bar{a}_n(f, g)$  como sendo o menor e o maior dos  $n$  autovalores da matriz  $T_n(f)T_n(g)$ , respectivamente. O conjunto de todos os pontos de acumulação de seqüências da forma  $(\underline{a}_n(f, g), \bar{a}_n(f, g)) \in \mathbb{R}^2$  será denotado por  $\mathcal{A}(f, g) \subseteq \mathbb{R}^2$ .

Fica estabelecido pelo Lema 3.16 que, para  $(\underline{a}, \bar{a}) \in \mathcal{A}(f, g)$ ,

$$\max(|\underline{a}|, |\bar{a}|) \leq \|f\|_\infty \|g\|_\infty,$$

de onde podemos concluir que  $\mathcal{A}(f, g)$  é um conjunto limitado. Denote por  $\mathcal{N}(\underline{a}, \bar{a})$  o conjunto de todas as seqüências inteiras crescentes  $(n_j)_{j \in \mathbb{N}}$ , tais que  $(\underline{a}_{n_j}(f, g), \bar{a}_{n_j}(f, g))$  converge para  $(\underline{a}, \bar{a}) \in \mathcal{A}(f, g)$ , à medida que  $j \rightarrow \infty$ . Com a notação básica definida, estamos preparados para enunciar o seguinte teorema.

**Teorema 3.19.** Considere as funções  $L(fg, \cdot)$  e  $K_{fg}(\cdot)$ , definidas respectivamente por

$$L(fg, t) = -\frac{1}{4\pi} \int_{\mathbb{T}} \log[1 - 2tf(x)g(x)] dx, \quad t \in \mathbb{R} \quad (3.32)$$

e

$$K_{fg}(x) = \sup_{t \in \mathbb{R}} \{xt - L(fg, t)\}. \quad (3.33)$$

Para  $(n_j)_{j \in \mathbb{N}} \in \mathcal{N}(\underline{a}, \bar{a})$ , com  $(\underline{a}, \bar{a}) \in \mathcal{A}(f, g)$ , a subsequência  $\{\mathcal{W}_{n_j}(f)\}_{j \in \mathbb{N}}$  satisfaz um LDP com função taxa

$$J(x) = \begin{cases} K_{fg}(x), & \text{se } x \in (x_1, x_2), \\ K_{fg}(x_1) + \frac{1}{2\underline{a}}(x - x_1), & \text{se } x \in (-\infty, x_1], \\ \frac{1}{2\bar{a}}(x - x_2) + K_{fg}(x_2), & \text{se } x \in [x_2, +\infty), \end{cases} \quad (3.34)$$

onde  $x_1$  e  $x_2$  são dados respectivamente por

$$x_1 = \begin{cases} L' \left( fg, \frac{1}{2\underline{a}} \right), & \text{se } \underline{a} < 0 \text{ e } \underline{a} < m_{fg}, \\ -\infty, & \text{caso contrário;} \end{cases} \quad (3.35)$$

$$x_2 = \begin{cases} L' \left( fg, \frac{1}{2\bar{a}} \right), & \text{se } \bar{a} > 0 \text{ e } \bar{a} > M_{fg}, \\ +\infty, & \text{caso contrário.} \end{cases} \quad (3.36)$$

Pela Observação 2.11, note que a função  $K_{fg}(\cdot)$ , definida pela expressão (3.33), é a transformada de Fenchel-Legendre da função  $L(fg, \cdot)$ . Para demonstrar este teorema, vamos precisar de dois lemas auxiliares e recordar a definição de medida pushforward da medida de Lebesgue normalizada (ver p. 190 de Bogachev, 2007).

**Definição 3.20.** Considere o espaço mensurável  $(\mathbb{T}, \mathcal{B}(\mathbb{T}))$ . A *medida de Lebesgue normalizada*  $\nu$ , definida neste espaço, é obtida através da medida de Lebesgue usual, dividida pelo tamanho do intervalo  $\mathbb{T}$ , de modo que  $\nu$  seja uma medida de probabilidade. Em outras palavras, se  $\tilde{\nu}$  representa a medida de Lebesgue usual em  $\mathbb{T}$ , então

$$d\nu \equiv \frac{d\tilde{\nu}}{2\pi}.$$

Considere agora a função mensurável  $h : (\mathbb{T}, \mathcal{B}(\mathbb{T}), \nu) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ . Podemos definir uma medida de probabilidade  $P_h$  no espaço  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ , através da expressão

$$P_h : A \mapsto \nu[h^{-1}(A)], \quad A \in \mathcal{B}(\mathbb{R}). \quad (3.37)$$

A medida  $P_h$  é chamada de *medida pushforward* da medida de Lebesgue normalizada  $\nu$  com relação à  $h(\cdot)$  (ou *imagem da medida*  $\nu$  com relação à  $h(\cdot)$ ).

O primeiro lema que vamos abordar, trata de relacionar os autovalores da matriz  $T_n(f)T_n(g)$  com a medida pushforward do produto das funções  $f(\cdot)$  e  $g(\cdot)$ .

**Lema 3.21.** *Seja  $P_{fg}$  a medida pushforward com relação à função  $fg(\cdot) \in L^\infty(\mathbb{T})$ . Se  $\lambda_1^n, \dots, \lambda_n^n$  são os autovalores da matriz  $T_n(f)T_n(g)$ , defina*

$$\mu_n = \frac{1}{n} \sum_{k=1}^n \delta_{\lambda_k^n}, \quad (3.38)$$

onde  $\delta_{\lambda_k^n}$  é a medida delta de Dirac concentrada em  $\lambda_k^n$ . Então  $\mu_n \Rightarrow P_{fg}$  ( $\mu_n$  converge fracamente para  $P_{fg}$ ) e existe uma sequência não-crescente de números reais positivos  $(\varsigma_n)_{n \in \mathbb{N}}$  tal que

$$(a) \quad \lim_{n \rightarrow \infty} \varsigma_n = 0;$$

$$(b) \quad \hat{\mu}_n = \mathbb{1}_{[m_{fg} - \varsigma_n, M_{fg} + \varsigma_n]} \mu_n \Rightarrow P_{fg};$$

$$(c) \quad \bar{\mu}_n = \mu_n - \hat{\mu}_n \Rightarrow 0.$$

*Demonstração.* Observe primeiramente que, da definição de integral com relação à funções simples (ver definição 4.2 de Bartle, 1995) e da expressão (3.38), temos que

$$\int_{\mathbb{R}} x^l d\mu_n(x) = \frac{1}{n} \sum_{k=1}^n (\lambda_k^n)^l, \quad \text{para todo } l \in \mathbb{N}. \quad (3.39)$$

Como  $T_n(f)T_n(g)$  é uma matriz real simétrica, existe uma matriz ortogonal  $P_n$  e uma matriz diagonal  $\Lambda_n = \text{diag}(\lambda_1^n, \dots, \lambda_n^n)$ , tais que  $T_n(f)T_n(g) = P_n \Lambda_n P_n^T$  (ver Proposição B.5). Devido à ortogonalidade da matriz  $P_n$ , segue que

$$(T_n(f)T_n(g))^l = P_n \Lambda_n^l P_n^T, \quad \text{para todo } l \in \mathbb{N}. \quad (3.40)$$

Lembramos que, dadas duas matrizes  $A$  e  $B$  quaisquer, cujos produtos  $AB$  e  $BA$  estão bem definidos, se  $\text{Tr}(A)$  representa o traço da matriz  $A$ , então temos sempre  $\text{Tr}(AB) = \text{Tr}(BA)$ . Aplicando o traço em ambos os lados de (3.40), obtemos

$$\text{Tr} \left[ (T_n(f)T_n(g))^l \right] = \text{Tr} \left( P_n \Lambda_n^l P_n^T \right) = \text{Tr} \left( \Lambda_n^l P_n^T P_n \right) = \text{Tr} \left( \Lambda_n^l \right) = \sum_{k=1}^n (\lambda_k^n)^l. \quad (3.41)$$

Para  $s = 2l$ ,  $f^{(2v-1)} = f$ ,  $f^{(2v)} = g$  e  $p_v = \infty$ , o teorema 1 de Avram (1988) mostra que

$$\lim_{n \rightarrow \infty} \frac{1}{n} \text{Tr} \left[ (T_n(f)T_n(g))^l \right] = \frac{1}{2\pi} \int_{\mathbb{T}} (f(t)g(t))^l dt. \quad (3.42)$$

Pelo teorema da mudança de variável para medidas (ver teorema 3.6.1. de Bogachev, 2007), obtemos a igualdade

$$\frac{1}{2\pi} \int_{\mathbb{T}} (f(t)g(t))^l dt = \int_{\mathbb{R}} x^l dP_{fg}(x). \quad (3.43)$$

Por fim, das equações (3.39), (3.41), (3.42) e (3.43) obtemos que

$$\lim_{n \rightarrow \infty} \int_{\mathbb{R}} x^l d\mu_n(x) = \int_{\mathbb{R}} x^l dP_{fg}(x), \quad \text{para todo } l \in \mathbb{N}. \quad (3.44)$$

A equação 3.44 mostra que os momentos de ordem  $l$  das distribuições  $\mu_n$ , convergem aos momentos de ordem  $l$  da distribuição  $P_{fg}$ . Pela definição usual de *convergência fraca* (ver seção 25 de Billingsley, 1995), para mostrar que  $\mu_n \Rightarrow P_{fg}$ , deveríamos mostrar que, para toda função  $h(\cdot)$  contínua e limitada em  $\mathbb{R}$ , temos

$$\lim_{n \rightarrow \infty} \int_{\mathbb{R}} h(x) d\mu_n(x) = \int_{\mathbb{R}} h(x) dP_{fg}(x).$$

Mas observe que, o Método dos Momentos (ver seção 30 de Billingsley, 1995) permite concluir a convergência fraca apenas da convergência dos momentos. De fato, sabemos pelo Lema 3.16 que os autovalores da matriz  $T_n(f)T_n(g)$  são limitados por  $\|f\|_{\infty}\|g\|_{\infty}$ . Segue das equações (3.41), (3.42) e (3.43) que

$$\int_{\mathbb{R}} x^l dP_{fg}(x) = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n (\lambda_k^n)^l \leq \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n (\|f\|_{\infty}\|g\|_{\infty})^l = (\|f\|_{\infty}\|g\|_{\infty})^l. \quad (3.45)$$

Seja  $(a_k)_{k \in \mathbb{N}}$  a sequência dos momentos da distribuição  $P_{fg}$ , isto é,  $a_k = \int_{\mathbb{R}} x^k dP_{fg}(x)$ , para todo  $k \in \mathbb{N}$ . É evidente pela expressão (3.45) que  $a_k$  é finito para todo  $k \in \mathbb{N}$ . Além do mais, a série de potências com termos iguais à  $a_k x^k / k!$ , converge absolutamente para todo  $x \in \mathbb{R}$ . Com efeito, esta série é majorada da seguinte forma

$$\sum_{k=1}^{\infty} \frac{a_k}{k!} x^k \leq \sum_{k=1}^{\infty} \frac{\|f\|_{\infty}^k \|g\|_{\infty}^k}{k!} x^k. \quad (3.46)$$

O raio de convergência  $R$  da série à direita nesta última igualdade, pode ser calculado através do limite

$$R = \lim_{k \rightarrow \infty} \frac{\|f\|_\infty^n \|g\|_\infty^n / n!}{\|f\|_\infty^{n+1} \|g\|_\infty^{n+1} / (n+1)!} = \lim_{n \rightarrow \infty} \frac{n+1}{\|f\|_\infty \|g\|_\infty} = \infty.$$

Segue do teorema 6, capítulo 12, de Lima (2013) que a série da direita em (3.46) converge absolutamente para todo  $x \in \mathbb{R}$ , e portanto, a série da esquerda também converge absolutamente para todo  $x \in \mathbb{R}$ .

Uma vez que a série dos momentos de  $P_{fg}$ , dada por  $\sum_{k=1}^{\infty} a_k x^k / k!$ , converge absolutamente na reta toda, podemos concluir pelo teorema 30.1 de Billingsley (1995) que,  $P_{fg}$  é a única medida de probabilidade com momentos iguais aos da sequência  $(a_k)_{k \in \mathbb{N}}$ . Em outras palavras, se existisse uma segunda medida de probabilidade  $\tilde{P}$  satisfazendo

$$\int_{\mathbb{R}} x^l d\tilde{P}(x) = \int_{\mathbb{R}} x^l dP_{fg}(x)$$

para todo  $l \in \mathbb{N}$ , então necessariamente,  $\tilde{P}$  e  $P_{fg}$  são idênticas. A medida  $P_{fg}$  é dita ser *determinada pelos seus momentos*.

Tendo em vista todos estes argumentos, da convergência exibida em (3.44) e do teorema 30.2 de Billingsley (1995), podemos concluir que  $\mu_n \Rightarrow P_{fg}$ , ficando assim demonstrada a primeira parte deste lema.

Para provar a segunda parte, observe que o suporte de  $P_{fg}$  é um subconjunto de  $[m_{fg}, M_{fg}]$  e, dado que  $\mu_n \Rightarrow P_{fg}$ , podemos achar uma sequência não-crescente de números reais positivos  $(\varsigma_n)_{n \in \mathbb{N}}$ , com  $\lim_{n \rightarrow \infty} \varsigma_n = 0$  e  $\hat{\mu}_n = \mathbf{1}_{[m_{fg}-\varsigma_n, M_{fg}+\varsigma_n]} \Rightarrow P_{fg}$ . Logo, a convergência  $\bar{\mu}_n = \mu_n - \hat{\mu}_n \Rightarrow 0$  é uma consequência imediata da convergência fraca das sequências de medidas  $(\mu_n)_{n \in \mathbb{N}}$  e  $(\hat{\mu}_n)_{n \in \mathbb{N}}$  para  $P_{fg}$ .  $\square$

O segundo lema necessário para a demonstração do Teorema 3.19, apresenta uma função taxa para uma sequência de variáveis aleatórias, que é combinação linear de distribuições i.i.d. qui-quadrados com um grau de liberdade.

**Lema 3.22.** *Seja  $(m(n))$  uma sequência crescente de números inteiros, divergindo para  $+\infty$ . Para  $n \geq 1$ , assumamos que as variáveis aleatórias  $Z_1^n, \dots, Z_{m(n)}^n$  são i.i.d., com distribuição  $\mathcal{X}_1^2$ . Considere um arranjo triangular  $a_1^n, \dots, a_{m(n)}^n$  de números reais não-negativos. Defina*

$$U_n = \frac{1}{n} \sum_{k=1}^{m(n)} a_k^n Z_k^n. \quad (3.47)$$

Assumamos também que

**(a)**  $m(n) = o(n)$ ;

$$(b) \quad \max_{1 \leq k \leq m(n)} a_k^n \rightarrow a > 0.$$

Então,  $(U_n)_{n \in \mathbb{N}}$  satisfaz um LDP com função taxa  $G_a(\cdot)$ , definida para todo  $x \in \mathbb{R}$  por

$$G_a(x) = \begin{cases} \frac{x}{2a} & \text{se } x \geq 0, \\ +\infty & \text{caso contrário.} \end{cases} \quad (3.48)$$

*Demonstração.* Assuma, sem perda de generalidade que,

$$a_1^n \equiv \max_{1 \leq k \leq m(n)} a_k^n.$$

Então segue que, para  $y > 0$  e  $1 \leq k \leq m(n)$ ,

$$\begin{aligned} a_1^n \geq a_k^n &\Rightarrow 1 - 2ya_1^n \leq 1 - 2ya_k^n \Rightarrow \log(1 - 2ya_1^n) \leq \log(1 - 2ya_k^n) \\ &\Rightarrow -\log(1 - 2ya_1^n) \geq -\log(1 - 2ya_k^n). \end{aligned} \quad (3.49)$$

Suponha que  $F$  é um conjunto fechado de  $\mathbb{R}^+$  com  $x \equiv \min F > 0$ . Seja  $i_F \equiv \inf_{z \in F} \{e^{nyz}\}$ . Uma vez que a função exponencial é estritamente crescente e  $F$  é um conjunto fechado, podemos concluir que  $i_F = e^{nyx}$ . Assim, pela desigualdade de Chebyshev (ver p. 25 de Durrett, 2013) e da monotonia da função logaritmo, obtemos para  $0 < y < (2a_1^n)^{-1}$ , que

$$\begin{aligned} \frac{1}{n} \log \mathbb{P}(U_n \in F) &\leq \frac{1}{n} \log \left[ \frac{\mathbb{E}(\exp(nyU_n))}{i_F} \right] = \frac{1}{n} \log \left[ \frac{\mathbb{E}(\exp(nyU_n))}{\exp(nyx)} \right] \\ &= \frac{1}{n} \log [\mathbb{E}(\exp(nyU_n))] - \frac{1}{n} \log(\exp(nyx)) \\ &= \frac{1}{n} \log \left[ \mathbb{E} \left( \sum_{k=1}^{m(n)} ya_k^n Z_k^n \right) \right] - xy = \frac{1}{n} \left\{ \sum_{k=1}^{m(n)} \log [\mathbb{E}(ya_k^n Z_k^n)] \right\} - xy \\ &= \frac{1}{n} \left\{ \sum_{k=1}^{m(n)} \log [(1 - 2ya_k^n)^{-1/2}] \right\} - xy = -\frac{1}{2n} \sum_{k=1}^{m(n)} \log [(1 - 2ya_k^n)] - xy \\ &\leq -\frac{m(n)}{2n} \log[(1 - 2ya_1^n)] - xy, \end{aligned}$$

onde a última desigualdade é consequência da expressão (3.49). Da hipótese de que  $m(n) = o(n)$ , isto é,  $\lim_{n \rightarrow \infty} m(n)/n = 0$ , concluimos que, para todo  $y \in (0, \frac{1}{2a})$ , vale

$$\begin{aligned} \limsup_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}(U_n \in F) &\leq \limsup_{n \rightarrow \infty} \left\{ -\frac{m(n)}{2n} \log[(1 - 2ya_1^n)] - xy \right\} \\ &= \lim_{n \rightarrow \infty} \left\{ -\frac{m(n)}{2n} \log[(1 - 2ya_1^n)] \right\} - xy = -xy \leq -\frac{x}{2a}, \end{aligned}$$

o que conclui a prova da cota superior.

Para a cota inferior, basta realizar a prova para os intervalos da forma  $(x, \infty)$ . Note que os elementos  $a_k^n$  são todos não-negativos, assim, para  $x > 0$  temos que

$$\mathbb{P}(U_n > x) = \mathbb{P}\left(\frac{1}{n} \sum_{k=1}^{m(n)} a_k^n Z_k^n > x\right) \geq \mathbb{P}(a_1^n Z_1^n > nx) = \mathbb{P}\left(Z_1^n > \frac{nx}{a_1^n}\right).$$

Se  $Z \sim \mathcal{X}_1^2$  e  $f_Z(x) = (\sqrt{2\pi})^{-1}x^{-1/2}e^{-x/2}$  é a função de densidade da variável aleatória  $Z$ , então, pelo método do Jacobiano (ver p. 82-83 de James, 2015), podemos calcular a função densidade  $f_{\sqrt{Z}}(\cdot)$ , da variável aleatória  $\sqrt{Z}$ , considerando a bijeção  $g : ]0, \infty[ \mapsto ]0, \infty[$  definida por  $g(x) = \sqrt{x}$ , da seguinte forma

$$\begin{aligned} f_{\sqrt{Z}}(y) &= f_Z(g^{-1}(y)) \left| \frac{d}{dy} (g^{-1}(y)) \right| = f_Z(y^2) \left| \frac{d}{dy} (y^2) \right| \\ &= \frac{(y^2)^{-1/2} e^{-y^2/2}}{\sqrt{2\pi}} |2y| = 2 \frac{e^{-y^2/2}}{\sqrt{2\pi}}. \end{aligned}$$

Assim, voltando ao cálculo da probabilidade de  $U_n$  ser maior do que  $x$ , sucede que

$$\begin{aligned} \mathbb{P}(U_n > x) &\geq \mathbb{P}\left(Z_1^n > \frac{nx}{a_1^n}\right) = \mathbb{P}\left(\sqrt{Z_1^n} > \sqrt{\frac{nx}{a_1^n}}\right) \\ &= \int_{\sqrt{nx/a_1^n}}^{+\infty} 2 \frac{e^{-y^2/2}}{\sqrt{2\pi}} dy = 2\mathbb{P}\left(Y > \sqrt{\frac{nx}{a_1^n}}\right), \end{aligned} \quad (3.50)$$

onde  $Y$  tem distribuição normal padrão.

Quando  $X$  segue uma distribuição normal padrão, podemos utilizar a aproximação (ver p. 228 Rohatgi e Saleh, 2000) dada por

$$\mathbb{P}(X > x) \approx \frac{1}{x\sqrt{2\pi}} e^{-x^2/2}, \quad \text{quando } x \rightarrow \infty,$$

que adaptada ao nosso caso, implica que

$$2\mathbb{P}\left(Y > \sqrt{\frac{nx}{a_1^n}}\right) \approx 2 \frac{1}{\sqrt{\frac{nx}{a_1^n}} \sqrt{2\pi}} \exp(-nx/2a_1^n) = \frac{\exp(-nx/2a_1^n)}{\sqrt{\pi} \sqrt{\frac{nx}{2a_1^n}}}. \quad (3.51)$$

Combinando portanto as expressões encontradas em (3.50) e (3.51), chegamos no seguinte limite inferior

$$\begin{aligned} \liminf_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}(U_n > x) &\geq \liminf_{n \rightarrow \infty} \frac{1}{n} \log \left[ 2\mathbb{P}\left(Y > \sqrt{\frac{nx}{a_1^n}}\right) \right] \\ &= \lim_{n \rightarrow \infty} \frac{1}{n} \log \left[ \frac{\exp(-nx/2a_1^n)}{\sqrt{\pi} \sqrt{nx/2a_1^n}} \right] \\ &= - \lim_{n \rightarrow \infty} \frac{1}{n} \frac{nx}{2a_1^n} - \lim_{n \rightarrow \infty} \frac{1}{2n} \log \left[ \pi \frac{nx}{2a_1^n} \right] \\ &= \frac{-x}{2a} - \frac{1}{2} \lim_{n \rightarrow \infty} \left\{ \frac{\log n}{n} + \frac{\log[x\pi/2a_1^n]}{n} \right\} = \frac{-x}{2a}. \end{aligned}$$

□



Com estes dois lemas estabelecidos, estamos em condições de demonstrar o Teorema 3.19. Assim como em Bercu et al. (1997, p. 81), faremos a prova do teorema apenas para o caso em que  $\underline{a} < 0$ ,  $\bar{a} > 0$ , onde  $(\underline{a}, \bar{a}) \in \mathcal{A}(f, g)$ , e ambos não pertencem ao intervalo  $[m_{fg}, M_{fg}]$ , onde  $m_{fg}$  e  $M_{fg}$  são respectivamente, o ínfimo essencial e o supremo essencial da função  $fg(\cdot)$ , introduzidos pela Definição 3.17.

**Demonstração do Teorema 3.19.** Fixe  $(\underline{a}, \bar{a}) \in \mathcal{A}(f, g)$ . Para simplificar a notação ao longo da prova, assumamos que  $(\underline{a}_n(f, g), \bar{a}_n(f, g))$  converge para  $(\underline{a}, \bar{a})$  (no caso em que esta convergência não é satisfeita, existe por definição uma subsequência  $(\underline{a}_{n_j}(f, g), \bar{a}_{n_j}(f, g))_{j \in \mathbb{N}}$  convergindo para  $(\underline{a}, \bar{a})$ , a qual pode ser usada para demonstrar o teorema).

Seja  $(\varsigma_n)_{n \in \mathbb{N}}$  a sequência definida pelo Lema 3.21. Podemos escrever  $\mathcal{W}_n(f)$  como

$$\mathcal{W}_n(f) = \mathcal{W}_n^1 + \mathcal{W}_n^2 + \mathcal{W}_n^3, \quad (3.52)$$

onde  $\mathcal{W}_n^1 = \mathcal{W}_n^+ + \mathcal{W}_n^-$  e

$$\begin{aligned} \mathcal{W}_n^+ &= \frac{1}{n} \sum_{k=1}^n \mathbf{1}_{\{\lambda_k^n \in [0, M_{fg} + \varsigma_n]\}} \lambda_k^n Z_k^n, & \mathcal{W}_n^- &= \frac{1}{n} \sum_{k=1}^n \mathbf{1}_{\{\lambda_k^n \in [m_{fg} - \varsigma_n, 0]\}} \lambda_k^n Z_k^n, \\ \mathcal{W}_n^2 &= \frac{1}{n} \sum_{k=1}^n \mathbf{1}_{\{\lambda_k^n \in (-\infty, m_{fg} - \varsigma_n]\}} \lambda_k^n Z_k^n, & \mathcal{W}_n^3 &= \frac{1}{n} \sum_{k=1}^n \mathbf{1}_{\{\lambda_k^n \in [M_{fg} + \varsigma_n, +\infty)\}} \lambda_k^n Z_k^n. \end{aligned}$$

Cada uma destas sequências satisfaz um LDP com funções taxa distintas, e como vamos provar, através de uma operação especial entre funções, chamada de *convolução ínfima* (ver p. 34 de Rockafellar, 1970), é possível fornecer uma função taxa para a sequência  $(\mathcal{W}_n(f))_{n \geq 2}$ , a partir das funções taxa encontradas para as sequências  $(\mathcal{W}_n^+)_{n \geq 2}$ ,  $(\mathcal{W}_n^-)_{n \geq 2}$ ,  $(\mathcal{W}_n^2)_{n \geq 2}$  e  $(\mathcal{W}_n^3)_{n \geq 2}$ .

**Definição 3.23.** A operação *convolução ínfima* das funções  $f_1(\cdot), \dots, f_n(\cdot)$  é definida por

$$(f_1 \diamond \dots \diamond f_n)(x) = \inf\{f_1(x_1) + \dots + f_n(x_n) \mid x_i \in \mathbb{R}^n, x_1 + \dots + x_n = x\}. \quad (3.53)$$

Vamos fazer a demonstração em quatro passos, de acordo com o seguinte roteiro:

- No Passo 1, vamos provar que a sequência  $(\mathcal{W}_n^+)_{n \geq 2}$  satisfaz um LDP com função taxa  $K_{(fg)^+}(\cdot)$ . Seguindo a mesma linha de raciocínio, podemos provar que a sequência  $(\mathcal{W}_n^-)_{n \geq 2}$  satisfaz um LDP com função taxa  $K_{(fg)^-}(\cdot)$ ;
- No Passo 2, vamos combinar os resultados encontrados no Passo 1, para mostrar que a sequência  $(\mathcal{W}_n^1)_{n \geq 2}$  satisfaz um LDP com função taxa  $K_{fg}(\cdot)$ ;
- No Passo 3, vamos usar os Lemas 3.21 e 3.22 para exibir uma função taxa para as sequências  $(\mathcal{W}_n^2)_{n \geq 2}$  e  $(\mathcal{W}_n^3)_{n \geq 2}$ ;
- No último passo, combinamos os resultados dos três primeiros e usamos a operação *convolução ínfima* para exibir uma função taxa para a sequência  $(\mathcal{W}_n(f))_{n \geq 2}$ .

Com este roteiro em mente, vamos à demonstração do Teorema.

**Passo 1:** Defina as funções parte positiva e parte negativa da função  $fg(\cdot)$  em  $\mathbb{T}$ , como sendo respectivamente,

$$(fg)^+(x) = \max[f(x)g(x), 0] \quad \text{e} \quad (fg)^-(x) = \min[-f(x)g(x), 0]. \quad (3.54)$$

Pela definição de  $\mathcal{W}_n^+$ , devemos levar em conta apenas os autovalores de  $T_n(f)T_n(g)$  que pertencem ao intervalo  $[0, M_{fg} + \varsigma_n]$ . Pelo Lema 3.21 (b), sabemos que

$$\begin{aligned} \lim_{n \rightarrow +\infty} L_n(\lambda, W_n^+) &= \lim_{n \rightarrow +\infty} \frac{1}{n} \log \mathbb{E} \left( e^{n\lambda W_n^+} \right) = \lim_{n \rightarrow +\infty} -\frac{1}{2n} \sum_{k=1}^n \log \left( 1 - 2\lambda \lambda_k^n \mathbf{1}_{\{\lambda_k^n \in [0, M_{fg} + \varsigma_n]\}} \right) \\ &= L((fg)^+, \lambda) = \begin{cases} -\frac{1}{4\pi} \int_{fg \geq 0} \log [1 - 2\lambda(fg)(t)] dt, & \text{se } \lambda < \frac{1}{2M_{fg}} \\ +\infty, & \text{caso } \lambda > \frac{1}{2M_{fg}}. \end{cases} \end{aligned} \quad (3.55)$$

Não temos certeza de como o limite acima se comporta quando  $\lambda = \frac{1}{2M_{fg}}$ . Definimos portanto,

$$L \left( (fg)^+, \frac{1}{2M_{fg}} \right) = \lim_{\lambda \uparrow \frac{1}{2M_{fg}}} L((fg)^+, \lambda).$$

Com esta escolha, a função  $L((fg)^+, \cdot)$  é semi-contínua inferiormente (ver teorema VII.2.1 de Ellis, 1985). Assim, devemos considerar dois casos distintos:

*Caso 1:* A função  $L((fg)^+, \cdot)$  é íngreme. Neste caso, o Teorema de Gärtner-Ellis (ver Subseção 2.2.3) se aplica e obtemos que  $(\mathcal{W}_n^+)_{n \geq 2}$  satisfaz um LDP com função taxa

$$K_{(fg)^+}(x) = \sup_{y \in \mathbb{R}} \left\{ xy + \frac{1}{4\pi} \int_{fg \geq 0} \log [1 - 2y(fg)(t)] dt \right\}, \quad x \in \mathbb{R}. \quad (3.56)$$

*Caso 2:* A função  $L((fg)^+, \cdot)$  não é íngreme. Defina portanto

$$x^* = \lim_{\lambda \uparrow \frac{1}{2M_{fg}}} L'((fg)^+, \lambda).$$

Apesar de não estarmos em condições de aplicar o Teorema de Gärtner-Ellis com toda a sua força, ainda é possível mostrar que  $(\mathcal{W}_n^+)_{n \geq 2}$  satisfaz um LDP com função taxa  $K_{(fg)^+}(\cdot)$ . Observe que as Hipóteses 2.19 e 2.20, apresentadas na Subseção 2.2.3, ainda são válidas, pelo que o Teorema de Gärtner-Ellis mostra que vale a cota superior do LDP, isto é, para todo fechado  $F \in \mathcal{B}(\mathbb{T})$ , temos

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P} \left( \mathcal{W}_n^+ \in F \right) \leq - \inf_{x \in F} K_{(fg)^+}(x).$$

Para mostrar que vale também a cota inferior, é suficiente mostrar o resultado para intervalos da forma  $(x, +\infty)$ . Se  $x < x^*$ , vale o Teorema de Gärtner-Ellis e a cota inferior é dada pela função taxa  $K_{(fg)^+}(\cdot)$ . No entanto, o problema se encontra no intervalo em que  $x \geq x^*$ . Se  $x \geq x^*$ , então temos que

$$K_{(fg)^+}(x) = K_{(fg)^+}(x^*) + \frac{1}{2M_{fg}}(x - x^*). \quad (3.57)$$

Vamos mostrar que, para qualquer  $x \geq x^*$ ,

$$\liminf_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}(\mathcal{W}_n^+ > x) \geq -K_{(fg)^+}(x). \quad (3.58)$$

Sendo  $\lambda_1^n, \dots, \lambda_n^n$  os autovalores de  $T_n(f)T_n(g)$ , defina

$$\lambda_{k_n}^n = \max\{\lambda_k^n : \lambda_k^n \in [0, M_{fg} + \varsigma_n]\}, \quad (3.59)$$

onde  $k_n \in \{1, \dots, n\}$  representa o índice do autovalor que satisfaz a relação (3.59). Observe que da definição de  $\lambda_{k_n}^n$ , podemos afirmar que

$$\mathbb{1}_{\{\lambda_{k_n}^n \in [0, M_{fg} + \varsigma_n]\}} \lambda_{k_n}^n Z_{\lambda_{k_n}^n} = \lambda_{k_n}^n Z_{\lambda_{k_n}^n},$$

e portanto, obtemos que

$$\mathcal{W}_n^+ - \frac{\lambda_{k_n}^n}{n} Z_{k_n}^n = \frac{1}{n} \sum_{k \in \{1, \dots, n\} \setminus \{k_n\}} \mathbb{1}_{\{\lambda_k^n \in [0, M_{fg} + \varsigma_n]\}} \lambda_k^n Z_k^n.$$

Pela independência das variáveis  $Z_1^n, \dots, Z_n^n$  e do fato que

$$\mathcal{W}_n^+ - \frac{\lambda_{k_n}^n}{n} Z_{k_n}^n > x^* - \varsigma \quad \text{e} \quad \frac{\lambda_{k_n}^n}{n} Z_{k_n}^n > x - x^* + \varsigma \quad \Rightarrow \quad \mathcal{W}_n^+ > x,$$

concluimos que, para qualquer  $0 < \varsigma < x^*$ ,

$$\mathbb{P}(\mathcal{W}_n^+ > x) \geq \mathbb{P}\left(\mathcal{W}_n^+ - \frac{\lambda_{k_n}^n}{n} Z_{k_n}^n > x^* - \varsigma\right) \mathbb{P}\left(\frac{\lambda_{k_n}^n}{n} Z_{k_n}^n > x - x^* + \varsigma\right). \quad (3.60)$$

Denote por  $A_n$  e  $B_n$ , os conjuntos definidos respectivamente por

$$A_n = \mathcal{W}_n^+ - \frac{\lambda_{k_n}^n}{n} Z_{k_n}^n > x^* - \varsigma \quad \text{e} \quad B_n = \frac{\lambda_{k_n}^n}{n} Z_{k_n}^n > x - x^* + \varsigma.$$

Por um lado, as expressões para o limite da função geradora de cumulantes normalizada, das variáveis aleatórias  $\mathcal{W}_n^+$  e  $\mathcal{W}_n^+ - \frac{\lambda_{k_n}^n}{n} Z_{k_n}^n$ , coincidem e são dadas por  $L((fg)^+, \cdot)$ . Uma vez que  $x^* - \varsigma < x^*$ , o Teorema de Gärtner-Ellis pode ser aplicado mais uma vez para deduzirmos que

$$\liminf_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}(A_n) \geq -K_{(fg)^+}(x^* - \varsigma). \quad (3.61)$$

Por outro lado, temos que  $\lim_{n \rightarrow \infty} \lambda_{k_n}^n = M_{fg}$ , de forma que o Lema 3.22 pode ser usado para obtermos

$$\liminf_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}(B_n) \geq - \left( \frac{x - x^* + \varsigma}{2M_{fg}} \right). \quad (3.62)$$

Observe que das expressões (3.57), (3.60), (3.61) e (3.62), obtemos a seguinte cota inferior

$$\begin{aligned} \liminf_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}(\mathcal{W}_n^+ > x) &\geq \liminf_{n \rightarrow \infty} \frac{1}{n} \log[\mathbb{P}(A_n) \mathbb{P}(B_n)] \\ &= \liminf_{n \rightarrow \infty} \frac{1}{n} [\log \mathbb{P}(A_n) + \log \mathbb{P}(B_n)] \\ &\geq \liminf_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}(A_n) + \liminf_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}(B_n) \\ &\geq -K_{(fg)^+}(x^* - \varsigma) - \left( \frac{x - x^* + \varsigma}{2M_{fg}} \right). \end{aligned}$$

Uma vez que  $\varsigma$  é um número real arbitrário e a função  $K_{(fg)^+}(\cdot)$  é contínua (ver corolário da p. 110 de Lima, 2013), quando  $\varsigma \rightarrow 0$ , segue que

$$\begin{aligned} \liminf_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}(\mathcal{W}_n^+ > x) &= \lim_{\varsigma \rightarrow 0} \left[ -K_{(fg)^+}(x^* - \varsigma) - \left( \frac{x - x^* + \varsigma}{2M_{fg}} \right) \right] \\ &= -K_{(fg)^+}(x^*) - \left( \frac{x - x^*}{2M_{fg}} \right) \\ &= -K_{(fg)^+}(x), \end{aligned}$$

onde a última igualdade acima decorre da expressão (3.57).

**Passo 2:** Usando argumentos análogos, podemos mostrar que a sequência  $(\mathcal{W}_n^-)_{n \geq 2}$  satisfaz um LDP com função taxa dada por

$$K_{(fg)^-}(x) = \sup_{y \in \mathbb{R}} \left\{ xy + \frac{1}{4\pi} \int_{fg \leq 0} \log[1 - 2y(fg)(t)] dt \right\}, \quad x \in \mathbb{R}. \quad (3.63)$$

Assim, podemos combinar as funções  $K_{(fg)^+}(\cdot)$  e  $K_{(fg)^-}(\cdot)$  para obter uma função taxa para a sequência  $(\mathcal{W}_n^1)_{n \geq 2} = (\mathcal{W}_n^+ + \mathcal{W}_n^-)_{n \geq 2}$ . De fato, se  $h^*(\cdot)$  representa a transformada de Fenchel-Legendre da função real  $h(\cdot)$  e  $(K_{(fg)^-} \diamond K_{(fg)^+})(\cdot)$  denota a convolução ínfima entre  $K_{(fg)^-}(\cdot)$  e  $K_{(fg)^+}(\cdot)$ , então, pelo teorema 16.4 de Rockafellar (1970) e da dualidade da transformada de Fenchel-Legendre, temos que

$$\begin{aligned} (K_{(fg)^-} \diamond K_{(fg)^+})^*(x) &= K_{(fg)^-}^*(x) + K_{(fg)^+}^*(x) \\ &= -\frac{1}{4\pi} \int_{fg \leq 0} \log[1 - 2x(fg)(t)] dt - \frac{1}{4\pi} \int_{fg \geq 0} \log[1 - 2x(fg)(t)] dt \\ &= -\frac{1}{4\pi} \int_{\mathbb{T}} \log[1 - 2x(fg)(t)] dt, \end{aligned}$$

e aplicando a transformada de Fenchel-Legendre nesta última igualdade, concluímos que

$$\left(K_{(fg)^-} \diamond K_{(fg)^+}\right)^{**}(x) = K_{fg}(x). \quad (3.64)$$

Usando o corolário 9.2.2 de Rockafellar (1970) e o teorema VI.5.3 de Ellis (1985), podemos concluir da dualidade da transformada de Fenchel-Legendre e da expressão (3.64) que

$$\left(K_{(fg)^-} \diamond K_{(fg)^+}\right)(x) = K_{fg}(x), \quad \text{para todo } x \in \mathbb{R}.$$

Dado que  $\mathcal{W}_n^1 = \mathcal{W}_n^+ + \mathcal{W}_n^-$ , segue do Corolário 2.24 que a sequência  $(\mathcal{W}_n^1)_{n \geq 2}$  satisfaz um LDP com função taxa dada por

$$\begin{aligned} I_{\mathcal{W}_n^1}(x) &= \inf_{\{(x_1, x_2) \in \mathbb{R}^2: x_1 + x_2 = x\}} \left[ K_{(fg)^-}(x_1) + K_{(fg)^+}(x_2) \right] \\ &= \left(K_{(fg)^-} \diamond K_{(fg)^+}\right)(x) = K_{fg}(x). \end{aligned}$$

**Passo 3:** Na notação do Lema 3.22, se considerarmos  $m(n)$  como sendo a sequência de índices  $k_n$ , para os quais  $\lambda_{k_n}^n \in ]M_{fg} + \varsigma_n, \infty)$ , então  $U_n$ , definida na equação (3.47), é igual à  $\mathcal{W}_n^3$ . E uma vez que  $m(n) = o(n)$  e

$$\max_{1 \leq k \leq n} \{\lambda_k : \lambda_k^n \in ]M_{fg} + \varsigma_n, \infty)\} \rightarrow \bar{a}, \quad \text{quando } n \rightarrow \infty,$$

o Lema 3.22 mostra que a sequência  $(\mathcal{W}_n^3)_{n \geq 2}$  satisfaz um LDP com função taxa  $G_{\bar{a}}(x)$ , para  $x \in \mathbb{R}$ .

De forma análoga, considerando  $m(n)$  como sendo a sequência de índices  $k_n$ , satisfazendo  $\lambda_{k_n}^n \in (-\infty, m_{fg} - \varsigma_n[$ , então  $U_n$ , definida na equação (3.47), é igual à  $-\mathcal{W}_n^2$ . Dado que  $-\underline{a} > 0$ , temos pelo Lema 3.22 que a sequência  $(-\mathcal{W}_n^2)_{n \geq 2}$  satisfaz um LDP com função taxa  $G_{-\underline{a}}(x)$ , para  $x \in \mathbb{R}$ . Deste fato, é evidente pelo Princípio da Contração que a sequência  $(\mathcal{W}_n^2)_{n \geq 2}$  satisfaz um LDP com função taxa  $G_{-\underline{a}}(-x)$ , para  $x \in \mathbb{R}$ .

**Passo 4:** Observe inicialmente que as variáveis aleatórias  $\mathcal{W}_n^1$ ,  $\mathcal{W}_n^2$  e  $\mathcal{W}_n^3$  são independentes, para todo  $n \geq 2$ , pois os intervalos  $(-\infty, m_{fg} - \varsigma_n[$ ,  $[m_{fg} - \varsigma_n, M_{fg} + \varsigma_n]$  e  $]M_{fg} + \varsigma_n, \infty)$  são disjuntos e as variáveis  $Z_1^n, \dots, Z_n^n$  são independentes entre si.

Seguindo um raciocínio semelhante ao do Passo 2, podemos mostrar que a soma das sequências  $(\mathcal{W}_n^1)_{n \geq 2}$ ,  $(\mathcal{W}_n^2)_{n \geq 2}$  e  $(\mathcal{W}_n^3)_{n \geq 2}$ , que resulta em  $(\mathcal{W}_n(f))_{n \geq 2}$ , satisfaz um LDP com função taxa dada por  $J = K_{fg} \diamond G_{-\underline{a}} \diamond G_{\bar{a}}$ , isto é, a sequência  $(\mathcal{W}_n(f))_{n \geq 2}$  satisfaz um LDP com função taxa

$$J(x) = \inf_{y_1 + y_2 + y_3 = x} \left\{ K_{fg}(y_1) + G_{-\underline{a}}(-y_2) + G_{\bar{a}}(y_3) \right\}, \quad (3.65)$$

para  $x \in \mathbb{R}$ . Por fim, a expressão apresentada em (3.34) segue da expressão (3.65). □

Da demonstração do Teorema 3.19, podemos deduzir o seguinte corolário.

**Corolário 3.24.** (a) Se  $fg \geq 0$  q.t.p. (respectivamente  $fg \leq 0$  q.t.p.) então a sequência  $(\mathcal{W}_n(f))_{n \geq 2}$  satisfaz um LDP com função taxa  $J(\cdot)$ , dada pela expressão (3.34), se, e somente se, a sequência  $(\bar{a}_n(f, g))_{n \in \mathbb{N}}$  (respectivamente  $(\underline{a}_n(f, g))_{n \in \mathbb{N}}$ ), tem apenas um único ponto de acumulação.

(b) Do contrário, a sequência  $(\mathcal{W}_n(f))_{n \geq 2}$  satisfaz um LDP com função taxa  $J(\cdot)$  se, e somente se,  $\mathcal{A}(f, g)$  é um conjunto unitário (um conjunto que contém apenas um elemento).

**Observação 3.25:** No caso particular em que  $f(\cdot)$  é uma função positiva e  $\|fg\|_\infty = \|f\|_\infty \|g\|_\infty$ , então  $(\mathcal{W}_n(f))_{n \geq 2}$  satisfaz um LDP com função taxa  $K_{fg}(\cdot)$ , dada na expressão (3.33). Por exemplo, quando  $f(\cdot)$  é a função constante  $f \equiv 1$ , então  $T_n(f) = I_n$ . Como consequência da expressão (3.15), temos que  $\mathcal{W}_n(1) = \frac{1}{n} \sum_{k=1}^n X_k^2$ , para todo  $n \in \mathbb{N}$ . Assim, a sequência  $\{\frac{1}{n} \sum_{k=1}^n X_k^2\}_{n \in \mathbb{N}}$  satisfaz um LDP com função taxa  $K_g(\cdot)$ , onde  $g(x) = (1 + \theta^2 - 2\theta \cos(x))^{-1}$ . Bryc e Smolenski (1993, p. 282) apresentam uma forma explícita para  $K_g(\cdot)$ , dada por

$$K_g(x) = \begin{cases} -\frac{1}{2} \log \left( \frac{2x}{1 + \sqrt{4\theta^2 x^2 + 1}} \right) + \frac{1}{2} (\theta^2 + 1)x - \frac{1}{2} \sqrt{4\theta^2 x^2 + 1}, & \text{se } x > 0, \\ \infty, & \text{caso contrário.} \end{cases}$$

O Corolário 3.24 (b) pode agora ser utilizado para mostrar que a sequência dos estimadores de Yule-Walker,  $(\tilde{\theta}_n)_{n \geq 2}$ , satisfaz um LDP com função taxa  $S(\cdot)$ , definida pela expressão (3.6). Mais explicitamente, podemos mostrar que o conjunto  $\mathcal{A}(f, g)$  é um conjunto unitário  $(\underline{a}, \bar{a})$  tal que

$$\underline{a}, \bar{a} \in \left\{ \frac{1}{2\lambda_1}, \frac{1}{2\lambda_2}, \frac{1}{2\lambda_3} \right\}, \quad (3.66)$$

onde

$$\lambda_1 = -2\theta(1 - \theta c), \quad \lambda_2 = -\frac{(1 + \theta)^2}{2(1 + c)}, \quad \lambda_3 = \frac{(1 - \theta)^2}{2(1 - c)}, \quad (3.67)$$

$$f(x) = \cos(x) - c \quad \text{e} \quad g(x) = (1 + \theta^2 - 2\theta \cos(x))^{-1},$$

para  $c \in (-1, 1)$  e  $x \in \mathbb{T}$ .

Como consequência, a sequência de estimadores  $(\tilde{\theta}_n)_{n \geq 2}$  satisfaz um LDP com função taxa dada pela expressão (3.34). E após uma série de cálculos e análise de casos, podemos chegar à expressão (3.6). A mesma técnica funciona para os estimadores de mínimos quadrados, no entanto, o Teorema 3.19 e Corolário 3.24 devem ser estendidos para o caso em que  $T_n(f)$  é substituída por uma matriz hermitiana  $M_n$  específica. Tendo em vista tamanha semelhança nos métodos para se chegar à função taxa, vamos nos

ocupar daqui em diante apenas com os detalhes para mostrar que a função  $R(\cdot)$ , dada na expressão (3.7), é de fato uma função taxa para a sequência dos estimadores de mínimos quadrados  $(\hat{\theta}_n)_{n \geq 2}$ .

### 3.3 LDP Para o Estimador de Mínimos Quadrados

Pretendemos calcular a probabilidade de  $\mathbb{P}(\hat{\theta}_n \geq c)$ , onde  $\hat{\theta}_n$  é dado na expressão (3.4) e  $c \in (-1, 1)$ . Pela expressão de  $\hat{\theta}_n$ , temos que

$$\mathbb{P}(\hat{\theta}_n \geq c) = \mathbb{P}(Z_n(c) \geq 0),$$

onde  $Z_n(c)$  é a estatística definida por

$$Z_n(c) = \frac{1}{n} \left( \sum_{k=2}^n X_k X_{k-1} - c \sum_{k=1}^{n-1} X_k^2 \right), \quad \text{para } n \geq 2. \quad (3.68)$$

Logo, as propriedades de grandes desvios da sequência de estimadores  $(\hat{\theta}_n)_{n \geq 2}$ , estão relacionadas com as propriedades de grandes desvios da sequência de estatísticas  $(Z_n(c))_{n \geq 2}$ .

Da mesma forma que procedemos com a estatística  $\mathcal{W}_n(f)$ , existe ainda uma maneira equivalente de escrever  $Z_n(c)$  como um produto matricial, nomeadamente,

$$Z_n(c) = \frac{1}{n} \mathbf{X}_n^T M_n \mathbf{X}_n, \quad \text{para } n \geq 2, \quad (3.69)$$

onde  $\mathbf{X}_n = (X_1, X_2, \dots, X_n)$  e  $M_n$  é a matriz quadrada tri-diagonal de ordem  $n$  dada por

$$M_n = \frac{1}{2} \begin{pmatrix} -2c & 1 & 0 & \cdots \\ 1 & -2c & 1 & \cdots \\ \cdots & \cdots & \cdots & \cdots \\ \cdots & 1 & -2c & 1 \\ \cdots & 0 & 1 & 0 \end{pmatrix}, \quad \text{para } n \geq 2. \quad (3.70)$$

A matriz  $M_n$  tem componentes idênticas às da matriz  $T_n(f)$ , apresentada na Seção 3.2, com exceção do elemento na linha  $n$  e coluna  $n$  (em  $M_n$  é nulo e em  $T_n(f)$  é igual  $-c$ ). Embora as matrizes  $M_n$  e  $T_n(f)$  sejam bastante semelhantes, o Teorema 3.19 e o Corolário 3.24 não são suficientes para mostrar que a sequência  $(Z_n(c))_{n \geq 2}$  satisfaz um LDP. Podemos estender estes dois resultados para o caso da presença de uma matriz simétrica real  $M_n$  de ordem  $n$ , para  $n \geq 2$ , através da seguinte proposição (ver proposição 3 de Bercu et al., 1997).

**Proposição 3.26.** *Seja  $M_n$  uma matriz simétrica real de ordem  $n$ . Denote por  $\lambda_1^n, \dots, \lambda_n^n$  os autovalores da matriz  $T_n^{1/2}(g)M_nT_n^{1/2}(g)$  e considere  $\underline{a}_n$  e  $\bar{a}_n$  o menor e o maior destes*

autovalores, respectivamente. Assuma que  $(\underline{a}_n)_{n \in \mathbb{N}}$  e  $(\bar{a}_n)_{n \in \mathbb{N}}$  são duas seqüências limitadas e à medida que  $n \rightarrow \infty$ , temos

$$\mu_n = \frac{1}{n} \sum_{k=1}^n \delta_{\lambda_k^n} \Rightarrow P_{fg} \quad (3.71)$$

para alguma função mensurável  $f(\cdot)$  definida em  $\mathbb{T}$ , tal que  $fg \in L^\infty(\mathbb{T})$ . Então, o Teorema 3.19 e o Corolário 3.24 valem se  $\mathcal{W}_n(f)$  for substituído por  $Z_n(c)$ , a estatística definida pela equação (3.69).

*Demonstração.* A Proposição 3.26 pode ser demonstrada de maneira análoga ao Teorema 3.19, colocando no lugar da matriz  $T_n(f)$  a matriz tri-diagonal  $M_n$ . A única observação que fazemos é que, a convergência fraca de  $\mu_n$  para  $P_{fg}$ , que foi obtida no Lema 3.21, já é satisfeita por hipótese.  $\square$

### 3.3.1 Decomposição Linear da Estatística $Z_n(c)$

Para mostrar o Teorema 3.5, vamos novamente fazer uso das funções

$$f(t) \equiv \cos(t) - c \quad \text{e} \quad g(t) \equiv (1 + \theta^2 + 2\theta \cos(t))^{-1},$$

para  $t \in \mathbb{T}$  e  $c \in (-1, 1)$ .

Sem perda de generalidade, sejam  $\lambda_1^n \leq \dots \leq \lambda_n^n$  os autovalores ordenados da matriz  $M_n T_n(g)$ . É simples mostrar que  $\lambda_1^n, \dots, \lambda_n^n$  são também os autovalores da matriz  $T_n^{1/2}(g) M_n T_n^{1/2}(g)$  (ver Lema B.7 (a)). Seja  $L_n(\lambda, Z_n(c)) = \frac{1}{n} \log \mathbb{E}(\exp[n\lambda Z_n(c)])$  a função geradora de cumulantes normalizada da estatística  $Z_n(c)$ . Pela Observação 3.14, podemos reescrever  $Z_n(c)$  como

$$Z_n(c) = \frac{1}{n} \sum_{k=1}^n \lambda_k^n Z_k^n, \quad (3.72)$$

onde  $Z_1^n, \dots, Z_n^n$  são variáveis aleatórias i.i.d. com distribuição  $\chi_1^2$ .

Tendo em vista a decomposição apresentada na equação (3.72), podemos deduzir com o uso do Lema B.7 (b) que, a função geradora de cumulantes normalizada da estatística  $Z_n(c)$  é dada pela expressão

$$L_n(\lambda, Z_n(c)) = \frac{1}{n} \log \mathbb{E}(e^{n\lambda Z_n(c)}) = \begin{cases} -\frac{1}{2n} \sum_{k=1}^n \log(1 - 2\lambda \lambda_k^n), & \text{se } \lambda \in \Delta_n \\ +\infty, & \text{caso contrário,} \end{cases}$$

onde

$$\Delta_n = \left\{ \lambda \in \mathbb{R} : \max(\lambda \lambda_1^n, \lambda \lambda_n^n) < \frac{1}{2} \right\},$$

ou equivalentemente,

$$L_n(\lambda, Z_n(c)) = \begin{cases} -\frac{1}{2n} \log \det(I_n - 2\lambda M_n T_n(g)), & \text{se } I_n - 2\lambda M_n T_n(g) \text{ é positiva definida,} \\ +\infty, & \text{caso contrário.} \end{cases} \quad (3.73)$$



**Observação 3.27:** Aqui estamos convencendo que  $\log(z) = -\infty$ , se  $z \leq 0$ , o que explica o fato de exigirmos que  $I_n - 2\lambda M_n T_n(g)$  seja positiva definida na equação (3.73), pois assim, os autovalores desta matriz, nomeadamente  $1 - 2\lambda\lambda_k^n$ , serão todos positivos, de tal modo que a equação (3.73) faça sentido.

Se definirmos

$$\mu_n = \frac{1}{n} \sum_{k=1}^n \delta_{\lambda_k^n}, \quad (3.74)$$

concluimos que, para qualquer  $\lambda \in \Delta_n$ ,

$$L_n(\lambda, Z_n(c)) = -\frac{1}{2} \int_{\mathbb{R}} \log(1 - 2\lambda x) \mu_n(dx). \quad (3.75)$$

Assim, o seguinte lema é útil para estabelecer conexão com a Proposição 3.26.

**Lema 3.28.** *Seja  $(\mu_n)_{n \in \mathbb{N}}$  uma sequência de medidas de probabilidade em  $\mathbb{R}$  cujos suportes estão contidos em um conjunto compacto fixado. Para uma medida de probabilidade  $\mu$  com suporte compacto, assumimos que existe  $v > 0$  tal que*

$$\lim_{n \rightarrow \infty} \int_{\mathbb{R}} \log(1 - 2\lambda x) \mu_n(dx) = \int_{\mathbb{R}} \log(1 - 2\lambda x) \mu(dx), \quad \text{para } |\lambda| \leq v. \quad (3.76)$$

Então,  $(\mu_n)_{n \in \mathbb{N}}$  converge fracamente para  $\mu$ .

*Demonstração.* Assim como referenciam Bercu et al. (1997), a demonstração deste lema segue as mesmas linhas da prova apresentada por Grenander e Szegö (1958, p. 63).  $\square$

Com uma simples multiplicação entre matrizes, podemos verificar que a inversa de  $T_n(g)$  é dada pela matriz tri-diagonal

$$T_n^{-1}(g) = \begin{pmatrix} 1 & -\theta & 0 & \cdots & 0 \\ -\theta & 1 + \theta^2 & -\theta & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & \cdots & -\theta & 1 + \theta^2 & -\theta \\ 0 & \cdots & 0 & -\theta & 1 \end{pmatrix}, \quad \text{se } n \geq 3, \quad (3.77)$$

$$T_2^{-1}(g) = \begin{pmatrix} 1 & -\theta \\ -\theta & 1 \end{pmatrix}, \quad \text{para } n = 2 \quad \text{e} \quad T_1^{-1}(g) = 1 - \theta^2, \quad \text{para } n = 1.$$

Defina  $D_n = (I_n - 2\lambda M_n T_n(g)) T_n^{-1}(g) = T_n^{-1}(g) - 2\lambda M_n$ , para  $n \geq 2$ . A forma explícita para  $D_n$  é dada por

$$D_n = \begin{pmatrix} 1 + 2c\lambda & -\theta - \lambda & 0 & \cdots \\ -\theta - \lambda & 1 + \theta^2 + 2c\lambda & -\theta - \lambda & \cdots \\ \cdots & \cdots & \cdots & \cdots \\ \cdots & -\theta - \lambda & 1 + \theta^2 + 2c\lambda & -\theta - \lambda \\ \cdots & 0 & -\theta - \lambda & 1 \end{pmatrix} = \begin{pmatrix} r & q & 0 & \cdots \\ q & p & q & \cdots \\ \cdots & \cdots & \cdots & \cdots \\ \cdots & q & p & q \\ \cdots & 0 & q & 1 \end{pmatrix},$$

onde  $p = 1 + \theta^2 + 2c\lambda$ ,  $q = -\theta - \lambda$  e  $r = p - \theta^2$ .

Observe que esta matriz é tri-diagonal, o que se comparada com a matriz  $I_n - 2\lambda M_n T_n(g)$ , apresenta maior simplicidade computacional no instante em que investigamos suas propriedades. Conforme mostramos no Lema B.10, a matriz  $I_n - 2\lambda M_n T_n(g)$  é positiva definida no caso em que  $D_n$  também o for. Assim, surge o Lema 3.29.

**Lema 3.29.** *Para  $n$  suficientemente grande, a matriz tri-diagonal  $D_n$  é positiva definida apenas no domínio  $\mathcal{D} = \mathcal{D}_1 \cup \mathcal{D}_2 \cup \mathcal{D}_3$ , onde  $\mathcal{D}_1 = \{\theta^2 < p \leq 2\theta^2 \text{ e } q^2 < \theta^2(p - \theta^2)\}$ ,  $\mathcal{D}_2 = \{2\theta^2 < p < 2 \text{ e } p > 2|q|\}$  e  $\mathcal{D}_3 = \{p \geq 2 \text{ e } q^2 < p - 1\}$ .*

*Demonstração.* A matriz  $D_n$  é tri-diagonal, logo, é simples de ser escalonada. Assim, podemos usar o fato de que uma matriz é positiva definida se, e somente se, seus pivôs são todos positivos (vide seção 17.G de Lima, 2014).

Para cada  $k = 1, \dots, n$ , considere o *menor principal* de ordem  $k$  da matriz  $D_n$  como sendo o determinante da submatriz principal  $D_n^k = (d_{l,j})_{1 \leq l, j \leq k}$ . O *critério de Sylvester* (teorema 19.10 de Lima, 2014) permite concluir que a matriz  $D_n$  é positiva definida se, e somente se, os seus menores principais são todos positivos.

Dado que  $d_{1,1} = r$ , podemos inferir do critério de Sylvester que  $D_n$  é positiva definida se, e somente se,  $r = p - \theta^2$  é uma constante positiva. Desta forma, podemos nos concentrar apenas nos valores de  $\lambda$  para os quais  $p > \theta^2$ .

Escalonando a matriz  $D_n$ , obtemos a matriz triangular superior  $n \times n$ ,

$$E_n = \begin{pmatrix} r & q & 0 & 0 & \cdots \\ 0 & F(r) & q & 0 & \cdots \\ 0 & 0 & F_2(r) & q & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & 0 & 0 & F_{n-2}(r) & q \\ \cdots & 0 & 0 & 0 & F_{n-1}(r) \end{pmatrix}, \quad (3.78)$$

onde  $F : (0, \infty) \rightarrow \mathbb{R}$  é a transformada definida por

$$F(r) = p - \frac{q^2}{r},$$

para  $r > 0$ ,  $F_k(\cdot)$  é a  $k$ -ésima iterada de  $F(\cdot)$ , para  $2 \leq k \leq n - 1$ , e cujos elementos na diagonal de  $E_n$  são os pivôs da matriz  $D_n$ . Desta forma, basta averiguarmos para quais valores de  $p$ ,  $q$  e  $r$ , a transformada  $F_k(r)$  é positiva, para todo  $2 \leq k \leq n - 1$ . Este resultado está demonstrado no lema 9.4.1 de Jensen (1995) e dele podemos inferir que a matriz  $D_n$  é definida positiva apenas em  $\mathcal{D} = \mathcal{D}_1 \cup \mathcal{D}_2 \cup \mathcal{D}_3$ , vide Jensen (1995, p. 274).  $\square$

Como consequência deste lema, podemos estabelecer o seguinte lema, cujo resultado fornece uma forma mais prática para a função geradora de cumulantes normalizada da estatística  $Z_n(c)$ .

**Lema 3.30.** *Considere  $\mathcal{D} = \mathcal{D}_1 \cup \mathcal{D}_2 \cup \mathcal{D}_3$  o domínio determinado pelo Lema 3.29. Para  $p > 2|q|$ , defina*

$$\rho = \frac{p + \sqrt{p^2 - 4q^2}}{2} \quad e \quad \tau = \frac{p - \sqrt{p^2 - 4q^2}}{2}. \quad (3.79)$$

Então, para  $n$  suficientemente grande, obtemos

$$L_n(\lambda, Z_n(c)) = \frac{1}{2n} \log[(1 - \theta^2)(\rho - \tau)] - \frac{1}{2n} \log[(r - \tau)(1 - \tau)\rho^{n-1} - (r - \rho)(1 - \rho)\tau^{n-1}], \quad (3.80)$$

se  $\lambda \in \mathcal{D}$  e  $L_n(\lambda, Z_n(c)) = +\infty$ , caso contrário.

*Demonstração.* Começamos observando que, para  $\lambda \in \mathcal{D}$ , temos sempre que  $p > 2|q|$ , o que indica que  $\rho$  e  $\tau$  são números reais ( $p > 2|q| \Rightarrow p^2 - 4q^2 > 0$ ). De fato, se  $\lambda \in \mathcal{D}$ , então pode pertencer à qualquer um dos conjuntos  $\mathcal{D}_1$ ,  $\mathcal{D}_2$  ou  $\mathcal{D}_3$ . Analisamos a seguir cada um dos casos.

**Caso 1:** Se  $\lambda \in \mathcal{D}_1$ , então por definição do conjunto  $\mathcal{D}_1$ , temos que  $\theta^2(p - \theta^2) > q^2$ . Dessa inequação, da monotonia da função  $h(x) = \sqrt{x}$  e de

$$\begin{aligned} \left(\frac{p}{2} - \theta^2\right)^2 \geq 0 &\Rightarrow \frac{p^2}{4} - p\theta^2 + \theta^4 \geq 0 \\ &\Rightarrow \frac{p^2}{4} \geq p\theta^2 - \theta^4 \Rightarrow \frac{p^2}{4} \geq \theta^2(p - \theta^2) > q^2, \end{aligned}$$

segue que

$$q^2 < \frac{p^2}{4} \Rightarrow |q| < \frac{p}{2} \Rightarrow 2|q| < p.$$

**Caso 2:** Se  $\lambda \in \mathcal{D}_2$ , o resultado parte da própria definição do conjunto  $\mathcal{D}_2$ .

**Caso 3:** Se  $\lambda \in \mathcal{D}_3$ , temos pela definição do conjunto  $\mathcal{D}_3$  que  $q^2 < p - 1$ . Basta então fixarmos  $\theta = 1$ , para que a prova siga as mesmas linhas do caso  $\lambda \in \mathcal{D}_1$ .

Denote por  $H_n$  o determinante da matriz quadrada de ordem  $n$

$$\begin{pmatrix} p & q & 0 & \cdots \\ q & p & q & \cdots \\ \cdots & \cdots & \cdots & \cdots \\ \cdots & q & p & q \\ \cdots & 0 & q & 1 \end{pmatrix}_{n \times n}. \quad (3.81)$$

Provamos no Lema B.11 que  $H_n$  tem a forma fechada

$$H_n = \frac{(1 - \tau)\rho^n + (\rho - 1)\tau^n}{\rho - \tau},$$

onde  $\rho$  e  $\tau$  são as constantes da expressão (3.79). Assim, o cálculo do determinante da matriz  $D_n$  se torna relativamente simples, visto que

$$\begin{aligned}
\det(D_n) &= rH_{n-1} - q^2H_{n-2} \\
&= r \frac{(1-\tau)\rho^{n-1} + (\rho-1)\tau^{n-1}}{\rho-\tau} - q^2 \frac{(1-\tau)\rho^{n-2} + (\rho-1)\tau^{n-2}}{\rho-\tau} \\
&= \frac{(\rho r - q^2)(1-\tau)\rho^{n-2} - (\tau r - q^2)(1-\rho)\tau^{n-2}}{\rho-\tau} \\
&= \frac{(\rho r - \rho\tau)(1-\tau)\rho^{n-2} - (\tau r - \rho\tau)(1-\rho)\tau^{n-2}}{\rho-\tau} \\
&= \frac{(r-\tau)(1-\tau)\rho^{n-1} - (r-\rho)(1-\rho)\tau^{n-1}}{\rho-\tau} \tag{3.82}
\end{aligned}$$

A partir do que mostramos no Lema B.9, sabemos que o determinante da matriz  $T_n^{-1}(g)$  é igual à  $1 - \theta^2$ , de onde concluímos que o determinante de  $T_n(g)$  é o inverso deste valor, isto é,  $\det[T_n(g)] = (1 - \theta^2)^{-1}$ . Assim, das expressões (3.73) e (3.82), obtemos para  $\lambda \in \mathcal{D}$  que

$$\begin{aligned}
L_n(\lambda, Z_n(c)) &= -\frac{1}{2n} \log\{\det[I_n - 2\lambda M_n T_n(g)]\} = -\frac{1}{2n} \log\{\det[D_n \times T_n(g)]\} \\
&= -\frac{1}{2n} \log\{\det(D_n) \times \det[T_n(g)]\} \\
&= -\frac{1}{2n} \log\left[ \frac{(r-\tau)(1-\tau)\rho^{n-1} - (r-\rho)(1-\rho)\tau^{n-1}}{\rho-\tau} \times \frac{1}{1-\theta^2} \right] \\
&= \frac{1}{2n} \log[(1-\theta^2)(\rho-\tau)] - \frac{1}{2n} \log[(r-\tau)(1-\tau)\rho^{n-1} - (r-\rho)(1-\rho)\tau^{n-1}].
\end{aligned}$$

Quando  $\lambda \notin \mathcal{D}$ , pelos Lemas 3.29 e B.10, a matriz  $I_n - 2\lambda M_n T_n(g)$  conterà em seu espectro pelo menos um autovalor não-positivo, e desta forma, a partir da expressão  $L_n(\lambda, Z_n(c)) = -1/2n \log \det(I_n - 2\lambda M_n T_n(g))$ , obtemos que  $L_n(\lambda, Z_n(c)) = +\infty$ .  $\square$

**Observação 3.31:** As constantes  $\rho$  e  $\tau$  que aparecem no enunciado do Lema 3.30, são raízes do polinômio  $\phi(x) = x^2 - px + q^2$ . É então imediato que

$$\rho + \tau = p \quad \text{e} \quad \rho\tau = q^2.$$

### 3.3.2 Demonstração do Teorema 3.5

A partir da expressão obtida para  $L_n(\lambda, Z_n(c))$  no Lema 3.30, é simples mostrar que, quando  $n \rightarrow \infty$ , temos

$$\lim_{n \rightarrow \infty} L_n(\lambda, Z_n(c)) = \begin{cases} -\frac{1}{2} \log(\rho) = -\frac{1}{2} \log\left(\frac{p + \sqrt{p^2 - 4q^2}}{2}\right), & \text{se } \lambda \in \mathcal{D}, \\ +\infty, & \text{caso contrário.} \end{cases} \tag{3.83}$$

De fato, observe inicialmente que

$$\begin{aligned} \lim_{n \rightarrow \infty} -\frac{1}{2n} \log[(r - \tau)(1 - \tau)\rho^{n-1}] &= \lim_{n \rightarrow \infty} \left\{ -\frac{1}{2n} [\log[(r - \tau)(1 - \tau)] + \log(\rho^{n-1})] \right\} \\ &= \lim_{n \rightarrow \infty} -\left(\frac{n-1}{2n}\right) \log(\rho) = -\frac{1}{2} \log(\rho). \end{aligned} \quad (3.84)$$

Além disso, pela hipótese de que  $\lambda \in \mathcal{D}$ , é também verdade que  $0 < \tau/\rho < 1$ , o que implica que

$$\lim_{n \rightarrow \infty} \left(\frac{\tau}{\rho}\right)^{n-1} = 0 \quad \Rightarrow \quad \lim_{n \rightarrow \infty} \left\{ \log \left[ 1 - \frac{(r - \rho)(1 - \rho)}{(r - \tau)(1 - \tau)} \left(\frac{\tau}{\rho}\right)^{n-1} \right] \right\} = \log(1) = 0.$$

Quando tomamos o limite em  $L_n(\lambda, Z_n(c))$  na expressão (3.80), para  $n \rightarrow \infty$ , o termo  $1/2n \log[(1 - \theta^2)(\rho - \tau)]$  desaparece, de onde somando e subtraindo a constante  $1/2 \log(\rho)$  e lembrando que esta constante pode ser expressada através do limite dado em (3.84), obtemos que

$$\begin{aligned} \lim_{n \rightarrow \infty} L_n(\lambda, Z_n(c)) &= \\ &= \lim_{n \rightarrow \infty} \left\{ -\frac{1}{2n} \log[(r - \tau)(1 - \tau)\rho^{n-1} - (r - \rho)(1 - \rho)\tau^{n-1}] \right\} \\ &= \lim_{n \rightarrow \infty} \left\{ -\frac{1}{2n} \log \left[ \frac{(r - \tau)(1 - \tau)\rho^{n-1} - (r - \rho)(1 - \rho)\tau^{n-1}}{(r - \tau)(1 - \tau)\rho^{n-1}} \right] \right\} - \frac{1}{2} \log(\rho) \\ &= \lim_{n \rightarrow \infty} \left\{ -\frac{1}{2n} \log \left[ 1 - \frac{(r - \rho)(1 - \rho)}{(r - \tau)(1 - \tau)} \left(\frac{\tau}{\rho}\right)^{n-1} \right] \right\} - \frac{1}{2} \log(\rho) \\ &= \lim_{n \rightarrow \infty} \left\{ -\frac{1}{2n} \right\} \lim_{n \rightarrow \infty} \left\{ \log \left[ 1 - \frac{(r - \rho)(1 - \rho)}{(r - \tau)(1 - \tau)} \left(\frac{\tau}{\rho}\right)^{n-1} \right] \right\} - \frac{1}{2} \log(\rho) = -\frac{1}{2} \log(\rho). \end{aligned}$$

Observamos agora que a expressão do limite de  $L_n(\lambda, Z_n(c))$  em (3.83), para  $\lambda \in \mathcal{D}$ , é exatamente igual à função  $L(fg, \lambda)$ . Para provar isso, vamos usar o fato de que (ver equação (4.224)(9) de Gradshteyn e Ryzhik, 1965)

$$\int_0^\pi \log(a + b \cos(t)) dt = \pi \log \left( \frac{a + \sqrt{a^2 - b^2}}{2} \right), \quad \text{se } a \geq |b| > 0. \quad (3.85)$$

Pela equação (3.32) temos

$$\begin{aligned} L(fg, \lambda) &= -\frac{1}{4\pi} \int_{\mathbb{T}} \log \left[ 1 - 2\lambda(\cos(t) - c) \left( \frac{1}{1 + \theta^2 - 2\theta \cos(t)} \right) \right] dt \\ &= -\frac{1}{4\pi} \int_{\mathbb{T}} \log \left[ \frac{1 + \theta^2 - 2\theta \cos(t) - 2\lambda \cos(t) + 2\lambda c}{1 + \theta^2 - 2\theta \cos(t)} \right] dt \\ &= -\frac{1}{4\pi} \int_{\mathbb{T}} \log \left[ \frac{p + 2q \cos(t)}{1 + \theta^2 - 2\theta \cos(t)} \right] dt \\ &= -\frac{1}{4\pi} \int_{\mathbb{T}} [\log[p + 2q \cos(t)] - \log[1 + \theta^2 - 2\theta \cos(t)]] dt \\ &= -\frac{1}{4\pi} \int_{\mathbb{T}} \log[p + 2q \cos(t)] dt + \frac{1}{4\pi} \int_{\mathbb{T}} \log[1 + \theta^2 - 2\theta \cos(t)] dt. \end{aligned}$$

Como já observamos na demonstração do Lema 3.30, temos sempre  $p > 2|q| = |2q|$ , desde que  $\lambda \in \mathcal{D}$ . Além disso, também vale que  $1 + \theta^2 \geq |-2\theta|$ , pois

$$\begin{cases} 1 + 2\theta + \theta^2 = (1 + \theta)^2 \geq 0, \\ 1 - 2\theta + \theta^2 = (1 - \theta)^2 \geq 0, \end{cases} \Rightarrow \begin{cases} 1 + \theta^2 \geq -2\theta, \\ 1 + \theta^2 \geq +2\theta, \end{cases} \Rightarrow 1 + \theta^2 \geq \max\{2\theta, -2\theta\}. \quad (3.86)$$

Assim, de (3.85) e da paridade da função  $\log[p + 2q \cos(t)]$ , obtemos que

$$\begin{aligned} -\frac{1}{4\pi} \int_{\mathbb{T}} \log[p + 2q \cos(t)] dt &= -\frac{1}{2\pi} \int_0^\pi \log[p + 2q \cos(t)] dt \\ &= -\frac{1}{2\pi} \left[ \pi \log \left( \frac{p + \sqrt{p^2 - 4q^2}}{2} \right) \right] = -\frac{1}{2} \log(\rho). \end{aligned}$$

Analogamente,

$$\begin{aligned} \frac{1}{4\pi} \int_{\mathbb{T}} \log[1 + \theta^2 - 2\theta \cos(t)] dt &= \frac{1}{2\pi} \int_0^\pi \log[1 + \theta^2 - 2\theta \cos(t)] dt \\ &= \frac{1}{2\pi} \left[ \pi \log \left( \frac{1 + \theta^2 + \sqrt{(1 + \theta^2)^2 - 4\theta^2}}{2} \right) \right] \\ &= \frac{1}{2} \log \left( \frac{1 + \theta^2 + \sqrt{1 + 2\theta^2 + \theta^4 - 4\theta^2}}{2} \right) \\ &= \frac{1}{2} \log \left( \frac{1 + \theta^2 + \sqrt{(1 - \theta^2)^2}}{2} \right) \\ &= \frac{1}{2} \log \left( \frac{1 + \theta^2 + 1 - \theta^2}{2} \right) = \frac{1}{2} \log \left( \frac{2}{2} \right) = 0, \end{aligned}$$

Combinando o resultado destas duas últimas expressões, mostramos que

$$L(fg, \lambda) = \lim_{n \rightarrow \infty} L_n(\lambda, Z_n(c)) = -1/2 \log(\rho), \quad \text{se } \lambda \in \mathcal{D}, \quad (3.87)$$

onde  $\rho = \frac{p + \sqrt{p^2 - 4q^2}}{2}$ .

Assim, como  $\lim_{n \rightarrow \infty} L_n(\lambda, Z_n(c)) = L(fg, \lambda)$ , para  $\lambda \in \mathcal{D}$ , notamos que,

$$\begin{aligned} \lim_{n \rightarrow \infty} -\frac{1}{2} \int_{\mathbb{R}} \log(1 - 2\lambda x) \mu_n(dx) &= \lim_{n \rightarrow \infty} L_n(\lambda, Z_n(c)) = -\frac{1}{4\pi} \int_{\mathbb{T}} \log(1 - 2\lambda(fg)(t)) dt \\ &= -\frac{1}{2} \int_{\mathbb{R}} \log(1 - 2\lambda x) P_{fg}(dx), \end{aligned}$$

o que pelo Lema 3.28, é suficiente para concluirmos que

$$\mu_n \Rightarrow P_{fg}.$$

Usando a notação introduzida no início da Subseção 3.2.3, podemos inferir que é unitário o conjunto  $\mathcal{A}(f, g)$ , dos pontos de acumulação de seqüências da forma  $(\underline{a}_n, \bar{a}_n)$ , onde  $\underline{a}_n$  e  $\bar{a}_n$  são respectivamente, o menor e o maior dos  $n$  autovalores da matriz  $M_n T_n(g)$ . Como consequência da Proposição 3.26, concluimos que a estatística  $Z_n(c)$  dada por

$$Z_n(c) = \frac{1}{n} \mathbf{X}_n^T M_n \mathbf{X}_n = \frac{1}{n} \left( \sum_{k=2}^n X_k X_{k-1} - c \sum_{k=1}^{n-1} X_k^2 \right),$$

satisfaz um princípio de grandes desvios com função taxa

$$J(x) = \begin{cases} K_{fg}(x), & \text{se } x \in (x_1, x_2), \\ K_{fg}(x_1) + \frac{1}{2\underline{a}}(x - x_1), & \text{se } x \in (-\infty, x_1], \\ \frac{1}{2\bar{a}}(x - x_2) + K_{fg}(x_2), & \text{se } x \in [x_2, +\infty) \end{cases} \quad (3.88)$$

onde  $x_1$  e  $x_2$  são dados respectivamente por

$$x_1 = \begin{cases} L' \left( fg, \frac{1}{2\underline{a}} \right), & \text{se } \underline{a} < 0 \text{ e } \underline{a} < m_{fg}, \\ -\infty, & \text{caso contrário;} \end{cases}$$

$$x_2 = \begin{cases} L' \left( fg, \frac{1}{2\bar{a}} \right), & \text{se } \bar{a} > 0 \text{ e } \bar{a} > M_{fg}, \\ -\infty, & \text{caso contrário.} \end{cases}$$

Para mostrar que coincidem as funções taxa  $R(\cdot)$  e  $J(\cdot)$ , dadas respectivamente pelas expressões (3.7) e (3.88), devemos analisar o comportamento dos autovalores de  $M_n T_n(g)$ . Para isso, é necessário um estudo minucioso do conjunto  $\mathcal{D}$ , dado no Lema 3.29, com relação à variável  $\lambda$ . Quando  $\lambda = 0$ , a função  $L_n(\lambda, Z_n(c))$  é identicamente nula, sendo um caso de menor interesse e podemos descartá-lo. Considere  $\lambda \neq 0$ . Analisamos a seguir os casos em que  $\lambda$  está em  $\mathcal{D}_1$ ,  $\mathcal{D}_2$  ou  $\mathcal{D}_3$ .

**Caso 1:** Se  $\lambda \in \mathcal{D}_1 = \{\theta^2 < p \leq 2\theta^2 \text{ e } q^2 < \theta^2(p - \theta^2)\}$ , então, lembrando que  $p = 1 + \theta^2 + 2c\lambda$  e  $q = -\theta - \lambda$ , obtemos

$$\begin{aligned} q^2 < \theta^2(p - \theta^2) &\Rightarrow (\theta + \lambda)^2 < \theta^2(1 + \theta^2 + 2c\lambda - \theta^2) \\ &\Rightarrow \theta^2 + 2\theta\lambda + \lambda^2 < \theta^2 + 2c\lambda\theta^2 \Rightarrow \lambda(\lambda + 2\theta - 2c\theta^2) < 0 \\ &\Rightarrow \lambda < -2\theta(1 - \theta c) \text{ ou } \lambda > -2\theta(1 - \theta c) \quad (\text{conforme o sinal de } \lambda). \end{aligned} \quad (3.89)$$

**Caso 2:** Se  $\lambda \in \mathcal{D}_2 = \{2\theta^2 < p < 2 \text{ e } p > 2|q|\}$ , então temos dois casos para averiguar. Note que

$$\begin{aligned} p > 2|q| &\Rightarrow -p < 2q < p \Rightarrow -(1 + \theta^2 + 2c\lambda) < 2(-\theta - \lambda) < 1 + \theta^2 + 2c\lambda \\ &\Rightarrow -1 - \theta^2 - 2c\lambda < -2\theta - 2\lambda < 1 + \theta^2 + 2c\lambda. \end{aligned} \quad (3.90)$$

Da desigualdade à esquerda em (3.90) e lembrando que  $c \in (-1, 1)$ , segue que

$$2\theta + 2\lambda < 1 + \theta^2 + 2c\lambda \Rightarrow -(1 - 2\theta + \theta^2) < 2(c - 1)\lambda \Rightarrow \frac{(1 - \theta)^2}{2(1 - c)} < \lambda. \quad (3.91)$$

Já da desigualdade à direita em (3.90), segue que

$$-2\theta - 2\lambda < 1 + \theta^2 + 2c\lambda \Rightarrow -(1 + \theta)^2 < 2(c + 1)\lambda \Rightarrow -\frac{(1 + \theta)^2}{2(1 + c)} < \lambda. \quad (3.92)$$

**Caso 3:** Se  $\lambda \in \mathcal{D}_3 = \{p \geq 2 \text{ e } q^2 < p - 1\}$ , então analogamente ao Caso 1 obtemos

$$\begin{aligned} q^2 < p - 1 &\Rightarrow (\theta + \lambda)^2 < 1 + \theta^2 + 2c\lambda - 1 \\ &\Rightarrow \theta^2 + 2\theta\lambda + \lambda^2 < \theta^2 + 2c\lambda \Rightarrow 2\theta + \lambda < 2c \Rightarrow \lambda < 2(c - \theta). \end{aligned} \quad (3.93)$$

A partir desta análise de casos, vamos denotar por  $\lambda_1, \lambda_2, \lambda_3$  e  $\lambda_4$  os limites para  $\lambda$ , encontrados nas expressões (3.89), (3.91), (3.92) e (3.93), onde

$$\lambda_1 = -2\theta(1 - \theta c) \quad \text{e} \quad \lambda_2 = 2(c - \theta), \quad (3.94)$$

$$\lambda_3 = -\frac{(1 + \theta)^2}{2(1 + c)} \quad \text{e} \quad \lambda_4 = \frac{(1 - \theta)^2}{2(1 - c)}. \quad (3.95)$$

Assim, o conjunto  $\mathcal{A}(f, g)$  é unitário da forma  $(\underline{a}, \bar{a})$ , onde

$$\underline{a}, \bar{a} \in \left\{ \frac{1}{2\lambda_1}, \frac{1}{2\lambda_2}, \frac{1}{2\lambda_3}, \frac{1}{2\lambda_4} \right\}. \quad (3.96)$$

Estes valores devem ser comparados com os de

$$m_{fg}, M_{fg} \in \left\{ \frac{1}{2\lambda_3}, \frac{1}{2\lambda_4} \right\}. \quad (3.97)$$

A classificação acima depende da localização do ponto  $c$ . A função produto  $fg(\cdot)$  é dada por

$$fg(x) = \frac{\cos(x) - c}{1 + \theta^2 - 2\theta \cos(x)}, \quad \text{se } x \in \mathbb{T} \text{ e } c \in (-1, 1). \quad (3.98)$$

Derivando com relação à  $x$ , obtemos que  $fg(\cdot)$  é constante, se, e somente se, para todo  $x \in \mathbb{T}$ ,

$$\begin{aligned} 0 &= \frac{d}{dx} [fg(x)] = \frac{[-\sin(x)][1 + \theta^2 - 2\theta \cos(x)] - [\cos(x) - c][2\theta \sin(x)]}{[1 + \theta^2 - 2\theta \cos(x)]^2} \\ \Leftrightarrow 0 &= \sin(x)[1 + \theta^2 - 2\theta \cos(x) + 2\theta \cos(x) - 2\theta c] \Leftrightarrow 0 = 1 + \theta^2 - 2\theta c \Leftrightarrow c = \frac{1 + \theta^2}{2\theta}. \end{aligned}$$

Portanto, se  $c_0 = (1 + \theta^2)/(2\theta)$ , então temos

$$m_{fg} = \frac{1}{2\lambda_3} = -\frac{1 + c}{(1 + \theta)^2} \quad \text{e} \quad M_{fg} = \frac{1}{2\lambda_4} = \frac{1 - c}{(1 - \theta)^2}, \quad \text{se } c > c_0 \quad (3.99)$$

e

$$m_{fg} = \frac{1}{2\lambda_4} = \frac{1 - c}{(1 - \theta)^2} \quad \text{e} \quad M_{fg} = \frac{1}{2\lambda_3} = -\frac{1 + c}{(1 + \theta)^2}, \quad \text{se } c < c_0. \quad (3.100)$$

De acordo com Bercu et al. (2000, p. 18), a função  $K_{fg}(\cdot)$  pode ser calculada através de

$$K_{fg}(c) = -L(fg, \lambda_c) = \frac{1}{2} \log(\rho), \quad (3.101)$$



onde  $\rho$  é a constante definida em 3.79 e  $\lambda_c$  é definida por

$$\lambda_c = \frac{\theta(1+c^2) - c(1+\theta^2)}{c^2 - 1}. \quad (3.102)$$

Considere  $p = 1 + \theta^2 + 2c\lambda$  e  $q = -\theta - \lambda$  como funções de  $\lambda$ . Vamos usar a notação  $p(\lambda)$  e  $q(\lambda)$  para denotar tais funções. Assim,  $\rho$  também é função de  $\lambda$  com expressão

$$\rho(\lambda) = \frac{p(\lambda) + \sqrt{p^2(\lambda) - 4q^2(\lambda)}}{2}. \quad (3.103)$$

Calculando  $p(\lambda)$  em  $\lambda_c$ , obtemos que

$$\begin{aligned} p(\lambda_c) &= 1 + \theta^2 + 2c\lambda_c = \frac{(1 + \theta^2)(c^2 - 1) + 2c[\theta(1 + c^2) - c(1 + \theta^2)]}{c^2 - 1} \\ &= \frac{(1 + \theta^2)(c^2 - 1) + 2c\theta(1 + c^2) - 2c^2(1 + \theta^2)}{c^2 - 1} \\ &= \frac{(1 + \theta^2)(-c^2 - 1) + 2c\theta(1 + c^2)}{c^2 - 1} = \frac{2c\theta(1 + c^2) - (1 + \theta^2)(c^2 + 1)}{c^2 - 1} \\ &= \frac{(1 + c^2)(2c\theta - 1 - \theta^2)}{c^2 - 1}. \end{aligned} \quad (3.104)$$

Para  $q^2(\lambda_c)$ , analogamente concluímos que

$$\begin{aligned} q^2(\lambda_c) &= \left( \theta + \frac{\theta(1+c^2) - c(1+\theta^2)}{c^2 - 1} \right)^2 = \left( \frac{\theta(c^2 - 1) + \theta(1 + c^2) - c(1 + \theta^2)}{c^2 - 1} \right)^2 \\ &= \left( \frac{2c^2\theta - c(1 + \theta^2)}{c^2 - 1} \right)^2 = \left( \frac{c(2c\theta - 1 - \theta^2)}{c^2 - 1} \right)^2 = \frac{c^2(2c\theta - 1 - \theta^2)^2}{(c^2 - 1)^2}. \end{aligned} \quad (3.105)$$

Logo,

$$\begin{aligned} p^2(\lambda_c) - 4q^2(\lambda_c) &= \frac{(1 + c^2)^2(2c\theta - 1 - \theta^2)^2 - 4c^2(2c\theta - 1 - \theta^2)^2}{(c^2 - 1)^2} \\ &= \frac{(1 + 2c^2 + c^4 - 4c^2)(2c\theta - 1 - \theta^2)^2}{(c^2 - 1)^2} = \frac{(1 - c^2)^2}{(c^2 - 1)^2} (2c\theta - 1 - \theta^2)^2 \\ &= (2c\theta - 1 - \theta^2)^2. \end{aligned} \quad (3.106)$$

Tirando a raiz, vemos que

$$\sqrt{p^2(\lambda_c) - 4q^2(\lambda_c)} = |2c\theta - 1 - \theta^2|. \quad (3.107)$$

Chegamos então na expressão

$$\begin{aligned} \rho(\lambda_c) &= \frac{p(\lambda_c) + \sqrt{p^2(\lambda_c) - 4q^2(\lambda_c)}}{2} = \frac{(1 + c^2)(2c\theta - 1 - \theta^2) + |2c\theta - 1 - \theta^2|(c^2 - 1)}{2(c^2 - 1)} \\ &= \begin{cases} \frac{c^2(1 + \theta^2 - 2\theta c)}{1 - c^2}, & \text{se } c > \frac{1 + \theta^2}{2\theta} \\ \frac{1 + \theta^2 - 2\theta c}{1 - c^2}, & \text{se } c < \frac{1 + \theta^2}{2\theta}. \end{cases} \end{aligned} \quad (3.108)$$

Juntando todos estes argumentos que acabamos de exibir com mais alguns cálculos extensivos, é possível mostrar que a expressão (3.88) gera a função taxa  $R(\cdot)$ , definida na expressão (3.7), terminando assim, a prova do Teorema 3.5.

### 3.4 Comparação Entre os Estimadores

No início da Seção 3.2, mostramos que os estimadores de Yule-Walker e de mínimos quadrados do processo AR(1) Gaussiano, definido pela equação (3.1), são ambos limitados por 1. Outra propriedade importante que não devemos esquecer de mencionar é que, dada uma amostra aleatória  $X_1, \dots, X_n$  do processo  $(X_n)_{n \in \mathbb{N}}$ , o estimador de Yule-Walker do parâmetro  $\theta$  é sempre majorado pelo estimador de mínimos quadrados deste mesmo parâmetro, isto é,

$$|\tilde{\theta}_n| = \left| \frac{\sum_{k=2}^n X_k X_{k-1}}{\sum_{k=1}^n X_k^2} \right| < \left| \frac{\sum_{k=2}^n X_k X_{k-1}}{\sum_{k=1}^{n-1} X_k^2} \right| = |\hat{\theta}_n|$$

para todo  $n \geq 2$ .

Com relação às propriedades de grandes desvios dos dois estimadores, quando  $\theta = 0$ , as funções  $R(\cdot)$  e  $S(\cdot)$ , apresentadas respectivamente nos Teoremas 3.5 e 3.4, assumem as seguintes formas

$$R(c) = \begin{cases} \frac{1}{2} \log \left( \frac{1}{1-c^2} \right), & \text{se } c \in \left[ -\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2} \right], \\ \log |2c|, & \text{caso contrário.} \end{cases}$$

$$S(c) = \begin{cases} \frac{1}{2} \log \left( \frac{1}{1-c^2} \right), & \text{se } c \in (-1, 1), \\ +\infty, & \text{caso contrário.} \end{cases}$$

Quando  $\theta = 0$ , as funções taxa  $R(\cdot)$  e  $S(\cdot)$  são ambas simétricas em torno de  $c = 0$ . O gráfico destas duas funções é apresentado de forma simultânea na Figura 3.1 e a partir desta mesma figura, fica evidente que o comportamento do estimador de Yule-Walker é melhor do que o do estimador de mínimos quadrados, levando em consideração as propriedades de grandes desvios destes estimadores e a estimação do parâmetro  $\theta$ .

Para demais valores de  $\theta$  compreendidos entre  $-1$  e  $1$ , não possuímos simetria com relação ao parâmetro  $c$ . Todavia, outra propriedade interessante pode ser observada no gráfico destas funções. A Figura 3.2 mostra o gráfico das funções taxa  $R(\cdot)$  e  $S(\cdot)$ , definidas respectivamente pelas equações (3.6) e (3.7), para valores de  $\theta$  compreendidos no conjunto  $\{-0.7, -0.5, -0.3, 0.3, 0.5, 0.7\}$ . Assim como na Figura 3.1, as funções  $R(\cdot)$  e  $S(\cdot)$  são apresentadas em vermelho e azul, respectivamente, e onde estas coincidem, encontram-se delineadas na cor preta. As duas linhas verticais interiores representam os pontos  $a$  e  $b$  dados pela expressão (3.8), ambos dependendo de  $\theta$ .

Podemos observar um comportamento “espelhado” de tais funções com relação aos parâmetros  $\theta$  e  $c$ . Este comportamento é decorrente da paridade das funções  $R(\cdot)$  e  $S(\cdot)$  com relação ao parâmetro  $\theta$ . Com esta paridade queremos dizer que, dado  $\theta_0 \in (-1, 1)$ ,

a função  $S(\cdot)$  com relação ao parâmetro  $\theta_0$ , tem sinal contrário ao da função  $S(\cdot)$  com relação ao parâmetro simétrico  $-\theta_0$ . Mais explicitamente, sejam

$$S_{\theta_0}(c) = \begin{cases} \frac{1}{2} \log \left( \frac{1 + \theta_0^2 - 2\theta_0 c}{1 - c^2} \right), & \text{se } c \in (-1, 1), \\ +\infty, & \text{caso contrário.} \end{cases}$$

e

$$S_{-\theta_0}(c) = \begin{cases} \frac{1}{2} \log \left( \frac{1 + \theta_0^2 + 2\theta_0 c}{1 - c^2} \right), & \text{se } c \in (-1, 1), \\ +\infty, & \text{caso contrário.} \end{cases}$$

Então, podemos verificar trivialmente que  $S_{\theta_0}(c) = S_{-\theta_0}(-c)$ , para todo  $c \in \mathbb{R}$ . Este fato justifica o comportamento “espelhado” do gráfico de  $S(\cdot)$  que observamos na Figura 3.2. A mesma propriedade pode ser constatada para a função taxa  $R(\cdot)$ .

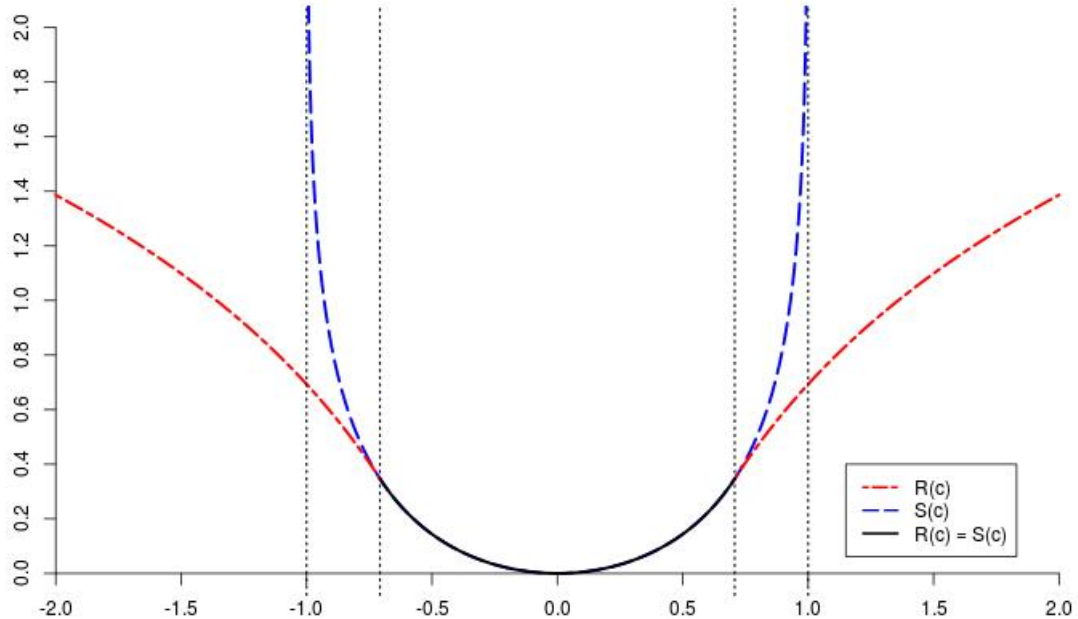


Figura 3.1 – Gráfico das funções taxa  $R(\cdot)$  e  $S(\cdot)$  para  $\theta = 0$ .

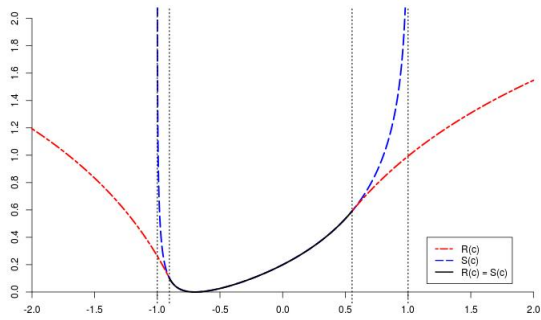
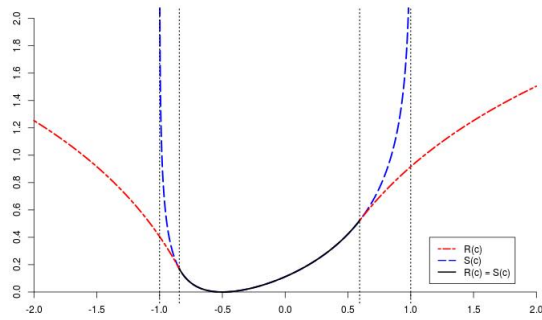
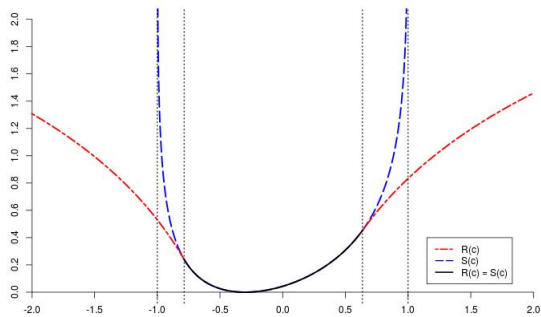
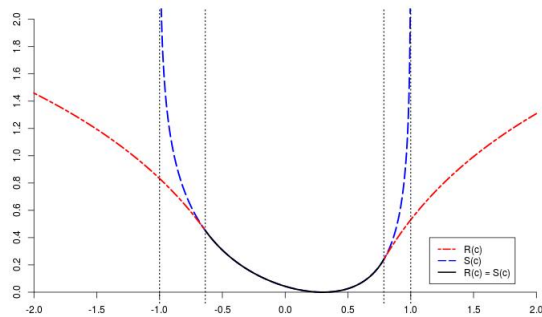
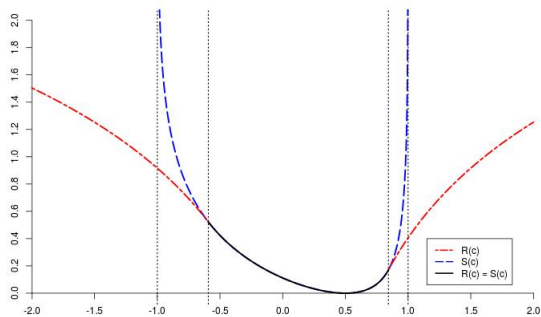
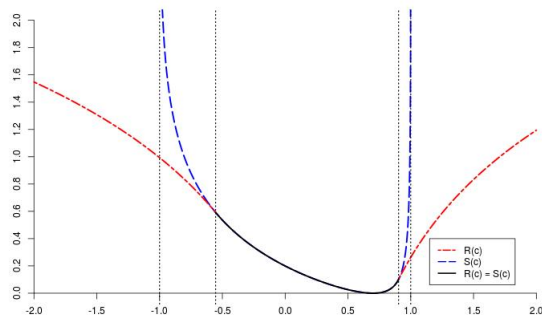
(a)  $\theta = -0.7$ .(b)  $\theta = -0.5$ (c)  $\theta = -0.3$ (d)  $\theta = 0.3$ (e)  $\theta = 0.5$ (f)  $\theta = 0.7$ 

Figura 3.2 – Gráfico das funções taxa  $R(\cdot)$  e  $S(\cdot)$  para valores de  $\theta$  compreendidos no conjunto  $\{-0.7, -0.5, -0.3, 0.3, 0.5, 0.7\}$ .

### 3.5 Princípio Forte de Grandes Desvios e o Processo AR(1) com Inovações Gaussianas

Estabelecendo relação com o assunto dissertado na Subseção 2.2.5, é ainda possível estender o resultado encontrado no Teorema 3.4 para o princípio forte de grandes desvios. Bercu et al. (2000, p. 16-18) provam o seguinte resultado.

**Teorema 3.32.** *Considere o processo  $(X_n)_{n \in \mathbb{N}}$  definido pela equação (3.1) e assuma que  $X_1$  é independente de  $(\varepsilon_n)_{n \geq 2}$ , com  $X_1 \sim \mathcal{N}(0, (1 - \theta^2)^{-1})$ . Considere também  $(\tilde{\theta}_n)_{n \geq 2}$ , a sequência dos estimadores de Yule-Walker do parâmetro  $\theta$ , para  $\tilde{\theta}_n$  definido como em (3.3). Defina para  $|c| < 1$ ,*

$$t_c = \frac{(\theta c - 1)(\theta - c)}{1 - c^2} \quad e \quad \sigma_c^2 = \frac{1 - c^2}{(1 + \theta^2 - 2\theta c)^2}. \quad (3.109)$$

Então existe uma sequência  $(d_{c,k})_{k \in \mathbb{N}}$  tal que, para todo inteiro  $p > 0$  e  $n$  suficientemente grande, temos

$$\mathbb{P}(\tilde{\theta}_n \geq c) = \frac{\exp(-nS(c))}{\sigma_c t_c \sqrt{2\pi n}} J(c) \left[ 1 + \sum_{k=1}^p \frac{d_{c,k}}{n^k} + \mathcal{O}\left(\frac{1}{n^{p+1}}\right) \right], \quad \text{para } 1 > c > \theta, \quad (3.110)$$

enquanto que

$$\mathbb{P}(\tilde{\theta}_n \leq c) = \frac{-\exp(-nS(c))}{\sigma_c t_c \sqrt{2\pi n}} J(c) \left[ 1 + \sum_{k=1}^p \frac{d_{c,k}}{n^k} + \mathcal{O}\left(\frac{1}{n^{p+1}}\right) \right], \quad \text{para } -1 < c < \theta, \quad (3.111)$$

onde  $S(\cdot)$  é a função taxa definida pela expressão (3.6) e

$$J(c) = \left( \frac{(1 - \theta^2)(1 + \theta^2 - 2\theta c)(1 - c^2)^2}{(\theta c - 1)^4} \right)^{1/2}. \quad (3.112)$$

Em suma, o que o Teorema 3.32 mostra é que a sequência de estimadores de Yule-Walker,  $(\tilde{\theta}_n)_{n \geq 2}$ , do parâmetro  $\theta$  dado na equação (3.1), satisfaz um SLDP de ordem  $p \in \mathbb{N}$ . As expressões (3.110) e (3.111) fornecem aproximações extremamente mais precisas para a cauda do estimador  $\tilde{\theta}_n$ , do que as obtidas através do princípio de grandes desvios, fornecidas pelo Teorema 3.4.

Ao leitor interessado na prova do Teorema 3.32 e em mais resultados relacionados ao princípio forte de grandes, aconselhamos a consulta de Bercu et al. (2000).



# 4 Processos AR(1) com Inovações $\alpha$ -Estáveis

Neste capítulo vamos generalizar o processo autorregressivo definido na equação (3.1), para o caso em que as inovações  $(\varepsilon_n)_{n \geq 2}$  pertencem a classe das distribuições  $\alpha$ -estáveis. Estamos interessados em determinar um princípio de grandes desvios para os estimadores de Yule-Walker e de mínimos quadrados do parâmetro  $\theta$  deste processo.

Como já citamos na Introdução deste trabalho, até onde temos conhecimento, nenhum LDP foi providenciado para qualquer um dos dois estimadores, tanto para o estimador de Yule-Walker, quanto para o de mínimos quadrados, no contexto do processo autorregressivo com inovações  $\alpha$ -estáveis. A ideia é tentar adaptar o estudo realizado por Bercu et al. (1997) para esta situação mais geral, visto que o processo Gaussiano resulta em um caso particular do processo  $\alpha$ -estável. Podemos antecipar nossas conclusões e afirmar que a teoria proposta por Bercu et al. (1997) não pode ser adaptada com tanta facilidade, visto que para  $\alpha \in (0, 2)$ , as distribuições  $\alpha$ -estáveis não possuem momentos de segunda ordem finitos (vide Propriedade 4.5 (d) abaixo).

Iniciamos este capítulo com um pequeno estudo sobre as distribuições  $\alpha$ -estáveis, que é uma classe de distribuições bastante geral e não costumam ser estudadas em cursos básicos de Probabilidade ou Estatística. Citamos com frequência Samorodnitsky e Taqqu (2000), que fornecem uma discussão bastante abrangente e atualizada sobre o assunto em pauta. Para o leitor interessado em referências clássicas sobre distribuições  $\alpha$ -estáveis, referenciamos Gnedenko et al. (1954), Feller (1970) e Sato (1999).

## 4.1 Distribuições $\alpha$ -Estáveis

Seja  $X$  uma variável aleatória seguindo uma distribuição normal com média  $\mu$  e variância  $\sigma^2$ , isto é,  $X \sim \mathcal{N}(\mu, \sigma^2)$ . Sejam  $X_1$  e  $X_2$  variáveis i.i.d., com  $X_i \stackrel{d}{=} X, i = 1, 2$ . Uma vez que  $X$  tem distribuição normal, sua função característica é definida por

$$\varphi_X(t) = \exp \left\{ it\mu - \frac{\sigma^2 t^2}{2} \right\}.$$

A partir das propriedades da função característica e da independência das variáveis  $X_1$  e  $X_2$ , temos para  $c = \sqrt{a^2 + b^2}$  e  $d = \mu(a + b - \sqrt{a^2 + b^2})$ ,

$$\begin{aligned}
\varphi_{cX+d}(t) &= e^{itd} \varphi_X(ct) = e^{itd} \exp \left\{ ict\mu - \frac{\sigma^2 c^2 t^2}{2} \right\} \\
&= e^{i(a+b-\sqrt{a^2+b^2})t\mu} \exp \left\{ i \left( \sqrt{a^2+b^2} \right) t\mu - \frac{\sigma^2 (a^2+b^2) t^2}{2} \right\} \\
&= \exp \left\{ i \left( a+b-\sqrt{a^2+b^2} + \sqrt{a^2+b^2} \right) t\mu - \frac{\sigma^2 (a^2+b^2) t^2}{2} \right\} \\
&= \exp \left\{ i(a+b)t\mu - \frac{\sigma^2 (a^2+b^2) t^2}{2} \right\} \\
&= \exp \left\{ iat\mu - \frac{\sigma^2 a^2 t^2}{2} \right\} \exp \left\{ ibt\mu - \frac{\sigma^2 b^2 t^2}{2} \right\} \\
&= \varphi_X(at) \varphi_X(bt) = \varphi_{aX_1}(t) \varphi_{bX_2}(t) = \varphi_{aX_1+bX_2}(t). \tag{4.1}
\end{aligned}$$

Pela unicidade da função característica de uma variável aleatória, podemos inferir da equação (4.1) que

$$aX_1 + bX_2 \stackrel{d}{=} cX + d. \tag{4.2}$$

Baseado na propriedade da distribuição normal acima demonstrada, a extensão às demais variáveis aleatórias pode ser realizada através da seguinte definição, adotada por Samorodnitsky e Taqqu (2000, p. 2-3).

**Definição 4.1.** Uma variável aleatória  $X$  é *estável* se, para  $X_1$  e  $X_2$  cópias independentes de  $X$  e para quaisquer  $a$  e  $b$  reais positivos, (4.2) é satisfeita para determinados  $c$  e  $d$ , onde  $c$  é real positivo e  $d$  é um número real qualquer. Além disso, uma variável aleatória  $X$  é *estritamente estável* se cumpre (4.2) para  $d = 0$ . Finalmente, a variável aleatória  $X$  é dita *simetricamente estável* se for estável e, ao mesmo tempo, simetricamente distribuída em torno de 0, o que denotamos por  $X \stackrel{d}{=} -X$ .

Podemos provar que, para qualquer variável aleatória estável  $X$ , existe um único número  $\alpha \in (0, 2]$ , tal que o terno  $a$ ,  $b$  e  $c$  em (4.2) satisfaz

$$c^\alpha = a^\alpha + b^\alpha.$$

A prova deste resultado pode ser encontrada em Feller (1970, p. 170-171). De acordo com Samorodnitsky e Taqqu (2000, p. 3), o número  $\alpha$  é chamado de *índice de estabilidade* ou *exponente característico* e uma variável aleatória  $X$  com tal índice é chamada de  $\alpha$ -estável.

#### 4.1.1 Definições Equivalentes de Estabilidade

A Definição 4.1 não é a única maneira de caracterizar a classe das distribuições  $\alpha$ -estáveis. Apresentamos em seguida duas definições equivalentes que descrevem bem estas distribuições, o que dependendo do contexto estudado, costumam ser mais versáteis do que aquela apresentada na Definição 4.1.



**Definição 4.2.** Uma variável aleatória  $X$  não-degenerada (veja a observação a seguir) é *estável* se, para todo  $n > 1$ , existem constantes  $c_n > 0$  e  $d_n \in \mathbb{R}$  de modo que

$$X_1 + \cdots + X_n \stackrel{d}{=} c_n X + d_n, \quad (4.3)$$

onde  $X_1, \dots, X_n$  são variáveis i.i.d. com  $X_i \stackrel{d}{=} X, i = 1, \dots, n$ . De maneira similar ao caso  $n = 2$ , dizemos que  $X$  é *estritamente estável* se  $d_n = 0$ , para todo  $n > 1$ .

**Observação 4.3:** De acordo com Rohatgi e Saleh (2000, p. 49), dizemos que uma variável aleatória  $X$  é *degenerada* em  $c \in \mathbb{R}$  se  $\mathbb{P}(X = c) = 1$ . Um exemplo clássico de tal variável é aquela que indica a face observada num lançamento de uma moeda com faces duplicadas, isto é, uma moeda com duas caras ou duas coroas. Em ambos os casos, a probabilidade de sair respectivamente, cara ou coroa, é sempre igual a 1/2. Uma variável aleatória deste tipo, apesar de ser sempre estável, não agrega em muito os nossos estudos, sendo que por este motivo vamos desconsiderá-las de nosso escopo.

É também possível definir variáveis  $\alpha$ -estáveis através de sua função característica. Denotando por  $\varphi_X(t) = \mathbb{E}(\exp\{itX\})$ , a seguinte definição revela uma fórmula fechada para a função característica de uma variável aleatória  $X$ , cuja distribuição é  $\alpha$ -estável.

**Definição 4.4.** Dizemos que uma variável aleatória  $X$  é  $\alpha$ -estável se existirem parâmetros  $0 < \alpha \leq 2$ ,  $\sigma > 0$ ,  $-1 \leq \beta \leq 1$  e  $\mu \in \mathbb{R}$ , tais que sua função característica, denotada por  $\varphi_X(\cdot)$ , é da forma

$$\varphi_X(t) = \mathbb{E}(\exp\{itX\}) = \begin{cases} \exp\{-\sigma^\alpha |t|^\alpha [1 - i\beta(\text{sign}(t)) \tan \frac{\pi\alpha}{2}] + i\mu t\}, & \alpha \neq 1, \\ \exp\{-\sigma |t| [1 + i\beta \frac{2}{\pi}(\text{sign}(t)) \log |t|] + i\mu t\}, & \alpha = 1, \end{cases} \quad (4.4)$$

onde  $\text{sign}(\cdot)$  representa a função sinal, definida por

$$\text{sign}(t) = \begin{cases} -1, & t < 0, \\ 0, & t = 0, \\ 1, & t > 0. \end{cases} \quad (4.5)$$

Com exceção do parâmetro  $\sigma$ , temos que  $\alpha$ ,  $\beta$  e  $\mu$  na definição acima, determinam de forma única a distribuição da variável aleatória  $X$  (observe que  $\sigma$  é uma exceção à regra porque, exemplificando para o caso  $\alpha = 2$ , se definirmos como  $\sigma_1 = 1$  e  $\sigma_2 = -1$ , ambos parâmetros determinam a mesma distribuição, dado que  $\sigma_1^2 = \sigma_2^2$ ). Tais parâmetros tem nomes específicos,  $\alpha$  será chamado o *índice de estabilidade*,  $\sigma$  o *parâmetro de escala*,  $\beta$  é o *parâmetro de simetria* e por fim,  $\mu$  é chamado de *parâmetro de localização*. Empregaremos a notação  $X \sim S_\alpha(\sigma, \beta, \mu)$  para designar uma variável  $\alpha$ -estável que é caracterizada por estes quatro parâmetros. Também encontramos com frequência na literatura (por exemplo em Samorodnitsky e Taqqu, 2000), a notação  $S\alpha S$ , utilizada para denotar distribuições

$\alpha$ -estáveis simétricas, o que de acordo com a Propriedade 4.5(c), ocorre se, e somente se,  $\beta = \mu = 0$ .

Vamos agora apresentar algumas das propriedades mais relevantes das distribuições  $\alpha$ -estáveis. Estas e outras propriedades mais, em conjunto com suas respectivas demonstrações, podem ser encontradas, por exemplo, em Samorodnitsky e Taqqu (2000).

**Propriedade 4.5. (Propriedades das  $\alpha$ -estáveis).** Sejam  $X_1$  e  $X_2$  duas variáveis aleatórias independentes  $\alpha$ -estáveis, seguindo respectivamente, as distribuições  $S_\alpha(\sigma_1, \beta_1, \mu_1)$  e  $S_\alpha(\sigma_2, \beta_2, \mu_2)$ . Então temos as seguintes propriedades:

(a) A variável aleatória  $Y = X_1 + X_2$  é  $\alpha$ -estável  $S_\alpha(\sigma, \beta, \mu)$  com

$$\sigma = (\sigma_1^\alpha + \sigma_2^\alpha)^{1/\alpha}, \quad \beta = \frac{\beta_1\sigma_1^\alpha + \beta_2\sigma_2^\alpha}{\sigma_1^\alpha + \sigma_2^\alpha} \quad e \quad \mu = \mu_1 + \mu_2; \quad (4.6)$$

(b) Para quaisquer constantes reais não-nulas  $a$  e  $b$ , a variável aleatória  $aX_1 + b$  é também  $\alpha$ -estável com distribuição

$$aX_1 + b \sim \begin{cases} S_\alpha(|a|\sigma_1, \text{sign}(a)\beta, a\mu_1 + b), & \text{se } \alpha \neq 1, \\ S_\alpha(|a|\sigma_1, \text{sign}(a)\beta, a\mu_1 - \frac{2}{\pi}a(\ln|a|)\sigma_1\beta_1 + b), & \text{se } \alpha = 1; \end{cases} \quad (4.7)$$

(c)  $X_1$  é simétrica se, e somente se,  $\beta_1 = \mu_1 = 0$ ;

(d) Se  $0 < \alpha < 2$ , então

$$\begin{cases} \mathbb{E}(|X_1|^p) < \infty, & \text{para todo } 0 < p < \alpha, \\ \mathbb{E}(|X_1|^p) = \infty, & \text{para todo } p \geq \alpha. \end{cases} \quad (4.8)$$

Explicitamente, a equação (4.8) afirma que uma variável aleatória que pertence à classe das distribuições  $\alpha$ -estáveis, para  $0 < \alpha < 2$ , não possui segundo momento finito. Sendo assim, perdemos alguns dos conceitos mais importantes utilizados no estudo de variáveis aleatórias, como por exemplo, a função de autocovariância e a função densidade espectral. Não obstante, contornamos esta adversidade estendendo a função de autocovariância para uma função com propriedades similares e que não necessita da hipótese de momentos finitos. A esta função chamaremos de *codiferença* e apresentamos sua definição e algumas propriedades na Subseção 4.2.1. Já na Subseção 4.3, apresentamos a definição de função densidade espectral para a classe das distribuições  $\alpha$ -estáveis.

### 4.1.2 Funções de Densidade das Distribuições $\alpha$ -Estáveis e Grandes Desvios para a Média Amostral

Segundo Samorodnitsky e Taqqu (2000, p. 9-10), ainda que as funções de densidade de variáveis aleatórias  $\alpha$ -estáveis existam e sejam contínuas, são apenas três os casos em

que existe uma forma fechada para tais densidades. Vamos apresentar logo em seguida as três distribuições estáveis com densidade conhecida, juntamente com as propriedades de grandes desvios para a média amostral  $\bar{X}_n = n^{-1} \sum_{k=1}^n X_k$ , de uma amostra aleatória  $X_1, \dots, X_n$ , cuja distribuição é a  $\alpha$ -estável em discussão.

**Exemplo 4.6. (Distribuição Gaussiana).** Considere uma variável aleatória  $X$  cuja distribuição é  $S_2(\sigma_X, 0, \mu)$ . Vamos mostrar que  $X \sim \mathcal{N}(\mu, 2\sigma_X^2)$ .

De fato, sabemos que a função característica  $\varphi_{\mathcal{N}(\mu, 2\sigma_X^2)}(\cdot)$  de uma distribuição normal de média  $\mu$  e variância  $2\sigma_X^2$  é dada pela expressão

$$\varphi_{\mathcal{N}(\mu, 2\sigma_X^2)}(t) = e^{it\mu - 2\sigma_X^2 t^2}.$$

Pela expressão (4.4), se fixarmos  $\alpha = 2$ ,  $\sigma = \sigma_X \sqrt{2}$  e  $\beta = 0$ , obtemos

$$\begin{aligned} \varphi_X(t) &= \exp\{-\sigma^\alpha |t|^\alpha [1 - i\beta(\text{sign}(t)) \tan \frac{\pi\alpha}{2}] + i\mu t\} \\ &= \exp\{-(\sigma_X \sqrt{2})^2 t^2 + i\mu t\} = e^{i\mu t - 2\sigma_X^2 t^2}. \end{aligned}$$

Dado que as duas funções características são iguais, concluímos que

$$\mathcal{N}(\mu, 2\sigma_X^2) \stackrel{d}{=} S_2(\sigma_X, 0, \mu). \quad (4.9)$$

Considere uma amostra aleatória  $X_1, \dots, X_n$ , seguindo uma distribuição  $\mathcal{N}(\mu, \sigma^2)$ . Müller (2016, p. 23-24) mostra através do Teorema de Cramér-Chernoff que a função taxa para a sequência  $\bar{X}_n = n^{-1} \sum_{k=1}^n X_k$  é dada pela expressão

$$I(x) = \frac{(x - \mu)^2}{2\sigma^2}, \quad x \in \mathbb{R}.$$

□

**Exemplo 4.7. (Distribuição de Cauchy).** Dizemos que  $X$  tem *distribuição de Cauchy* com parâmetro de locação  $\mu$  e parâmetro de escala  $\sigma$ , isto é,  $X \sim \mathcal{C}(\mu, \sigma)$ , se possui função densidade dada por

$$f_X(x) = \frac{\sigma}{\pi[\sigma^2 + (x - \mu)^2]}. \quad (4.10)$$

Além disso, dizemos que  $X$  tem *distribuição Cauchy padrão* quando  $\mu = 0$  e  $\sigma = 1$ .

Não é difícil mostrar que a função  $f_X(\cdot)$  definida em (4.10) é uma função de densidade (ver Apêndice C). Além disso, a distribuição de Cauchy é também conhecida por ser um caso particular das distribuições  $\alpha$ -estáveis quando  $\alpha = 1$  e  $\beta = 0$ , em outras palavras, as distribuições  $\mathcal{C}(\mu, \sigma)$  e  $S_1(\sigma, 0, \mu)$  são equivalentes (ver Proposição C.2).

Sejam  $X_1, \dots, X_n$ , variáveis aleatórias i.i.d. com distribuição Cauchy padrão. Vamos determinar a função taxa  $I(\cdot)$  para a sequência das médias amostrais  $(\bar{X}_n)_{n \in \mathbb{N}}$  definida por  $\bar{X}_n = n^{-1} \sum_{k=1}^n X_k$ .

A função geradora de momentos da variável aleatória  $X_1$  é dada por

$$M_{X_1}(t) = \mathbb{E} \left( e^{tX_1} \right) = \int_{\mathbb{R}} \frac{e^{tx}}{\pi(1+x^2)} dx, \quad t \in \mathbb{R}.$$

Se  $t = 0$ , então é óbvio que  $M_{X_1}(0) = 1$ . Se  $t > 0$ , visto que a função  $e^{tx}$  é convexa, podemos concluir pela equação da reta tangente no ponto  $x = 0$ , que  $e^{tx} \geq tx + 1$ , para todo  $x$  real. Além disso, como  $e^{tx}/[\pi(1+x^2)]$  é uma função positiva em todo ponto da reta, segue que

$$\int_{-\infty}^0 \frac{e^{tx}}{\pi(1+x^2)} dx \geq 0.$$

Destes dois últimos argumentos podemos inferir que, se  $t > 0$ ,

$$\begin{aligned} M_{X_1}(t) &= \int_{-\infty}^0 \frac{e^{tx}}{\pi(1+x^2)} dx + \int_0^{\infty} \frac{e^{tx}}{\pi(1+x^2)} dx \geq \int_0^{\infty} \frac{e^{tx}}{\pi(1+x^2)} dx \\ &\geq \int_0^{\infty} \frac{tx+1}{\pi(1+x^2)} dx = \frac{t}{\pi} \int_0^{\infty} \frac{x}{1+x^2} dx + \frac{1}{\pi} \int_0^{\infty} \frac{1}{1+x^2} dx = \\ &= \frac{t}{2\pi} \left[ \lim_{b \rightarrow \infty} \int_0^b \frac{2x}{1+x^2} dx \right] + \frac{1}{\pi} \left[ \lim_{b \rightarrow \infty} \arctan(b) - \arctan(0) \right] \\ &= \frac{t}{2\pi} \left[ \lim_{b \rightarrow \infty} \log(1+b^2) - \log(1) \right] + \frac{1}{\pi} \times \frac{\pi}{2} = \infty. \end{aligned}$$

Se  $t < 0$ , usamos argumentos análogos aos do caso  $t > 0$ . Com efeito, se  $t < 0$ ,

$$\begin{aligned} M_{X_1}(t) &= \int_{-\infty}^0 \frac{e^{tx}}{\pi(1+x^2)} dx + \int_0^{\infty} \frac{e^{tx}}{\pi(1+x^2)} dx \geq \int_{-\infty}^0 \frac{e^{tx}}{\pi(1+x^2)} dx \\ &\geq \int_{-\infty}^0 \frac{tx+1}{\pi(1+x^2)} dx = \frac{t}{2\pi} \lim_{b \rightarrow \infty} \int_{-b}^0 \frac{2x}{1+x^2} dx + \frac{1}{\pi} \lim_{b \rightarrow \infty} \int_{-b}^0 \frac{1}{1+x^2} dx \\ &= -\frac{t}{2\pi} \lim_{b \rightarrow \infty} \int_0^{-b} \frac{2x}{1+x^2} dx + \frac{1}{\pi} \lim_{b \rightarrow \infty} \int_0^{-b} \frac{1}{1+x^2} dx \\ &= -\frac{t}{2\pi} \left[ \lim_{b \rightarrow \infty} \log(1+(-b)^2) - \log(1) \right] + \frac{1}{2} = \infty, \end{aligned}$$

onde acima

$$\int_{-b}^0 \frac{1}{1+x^2} dx = \int_0^b \frac{1}{1+x^2} dx$$

é consequência da paridade da função  $1/(1+x^2)$  e a igualdade

$$\int_{-b}^0 \frac{2x}{1+x^2} dx = - \int_0^{-b} \frac{2x}{1+x^2} dx,$$

foi obtida realizando a troca dos limites de integração.

Em conclusão, obtemos que a função geradora de momentos da variável  $X_1$  é dada por

$$M_{X_1}(t) = \begin{cases} 1, & \text{se } t = 0, \\ +\infty, & \text{caso contrário.} \end{cases} \quad (4.11)$$

O Teorema de Cramér-Chernoff aplicado à este caso afirma que a função taxa é dada por

$$I(x) = \sup_{t \in \mathbb{R}} \{tx - \mathcal{L}_{X_1}(t)\}, \quad x \in \mathbb{R},$$

onde  $\mathcal{L}_{X_1}(t) = \log M_{X_1}(t)$  é a função de log-Laplace de  $X_1$ . Convencionando que  $\log(\infty) = \infty$ , concluímos que

$$-\mathcal{L}_{X_1}(t) = -\log M_{X_1}(t) = \begin{cases} 0, & \text{se } t = 0, \\ -\infty, & \text{se } t \neq 0, \end{cases}$$

e portanto, o supremo só precisa ser considerado para  $t = 0$ , o que simplifica a função taxa para

$$I(x) = \sup_{t=0} \{tx - \mathcal{L}_{X_1}(t)\} = 0, \quad \text{para todo } x \in \mathbb{R}.$$

□

**Exemplo 4.8. (Distribuição de Lévy).** Se  $X$  segue uma distribuição  $\alpha$ -estável com parâmetros  $\alpha = 1/2$  e  $\beta = 1$ , então diremos que  $X$  segue uma *distribuição de Lévy*, isto é,  $X \sim S_{1/2}(\sigma, 1, \mu)$ . A função densidade de tal variável aleatória é dada por

$$f_X(x) = \left(\frac{\sigma}{2\pi}\right)^{1/2} \frac{1}{(x - \mu)^{3/2}} \exp\left[-\frac{\sigma}{2(x - \mu)}\right], \quad x \in (\mu, +\infty). \quad (4.12)$$

Além disso, dizemos que  $X$  tem *distribuição Lévy padrão* quando  $\mu = 0$  e  $\sigma = 1$ . No Apêndice C, mostramos que a função  $f_X(\cdot)$  definida em (4.12) é de fato, uma função de densidade.

Considere uma amostra  $X_1, \dots, X_n$ , de variáveis aleatórias i.i.d. com distribuição Lévy padrão. Da mesma forma que procedemos com a distribuição de Cauchy, vamos determinar a função taxa  $I(\cdot)$  da sequência de médias amostrais  $(\bar{X}_n)_{n \in \mathbb{N}}$ , mas agora para o caso Lévy.

A função geradora de momentos da variável aleatória  $X_1$  é dada por

$$M_{X_1}(t) = \mathbb{E}\left(e^{tX_1}\right) = \int_0^\infty e^{tx} \left(\frac{1}{2\pi}\right)^{1/2} x^{-3/2} e^{-1/(2x)} dx, \quad t \in \mathbb{R}.$$

Se  $t = 0$ , então claramente  $M_{X_1}(0) = 1$ . Se  $t > 0$ , note que a função  $-1/(2x)$  é crescente no intervalo  $(0, \infty)$ . Deste fato e da monotonia da função exponencial obtemos

$$-\frac{1}{2} \leq -\frac{1}{2x}, \quad \forall x \geq 1 \quad \Rightarrow \quad e^{-1/2} \leq e^{-1/(2x)}, \quad \forall x \geq 1. \quad (4.13)$$

Pela desigualdade de Jensen (ver p. 127 de James, 2015) sabemos que

$$M_{X_1}(t) = \mathbb{E}\left(e^{tX_1}\right) \geq e^{t\mathbb{E}(X_1)}. \quad (4.14)$$

Calculando a esperança de  $X_1$ , verificamos que

$$\begin{aligned}\mathbb{E}(X_1) &= \int_0^\infty x \left(\frac{1}{2\pi}\right)^{1/2} x^{-3/2} e^{-1/(2x)} dx = \int_0^\infty \left(\frac{1}{2\pi}\right)^{1/2} x^{-1/2} e^{-1/(2x)} dx \\ &\geq \int_1^\infty \left(\frac{1}{2\pi}\right)^{1/2} x^{-1/2} e^{-1/(2x)} dx \geq \int_1^\infty \left(\frac{1}{2\pi}\right)^{1/2} x^{-1/2} e^{-1/2} dx \\ &= \left(\frac{1}{2\pi e}\right)^{1/2} \int_1^\infty \frac{1}{x^{1/2}} dx = \infty,\end{aligned}\tag{4.15}$$

onde na última igualdade usamos o resultado do Cálculo Integral que afirma que a integral  $\int_1^\infty \frac{1}{x^p} dx$  diverge para qualquer  $p \leq 1$  (ver teorema 8.8.2 de Anton et al., 2007). Portanto, segue de (4.14) que  $M_{X_1}(t) = \infty$ , para  $t > 0$ .

Para  $t < 0$ , os cálculos já não são tão simples. Observe que podemos reescrever  $M_{X_1}(\cdot)$  como

$$\begin{aligned}M_{X_1}(t) &= \frac{1}{\sqrt{2\pi}} \int_0^\infty x^{-\frac{1}{2}-1} \exp\left[-\frac{1/2}{x} - (-t)x\right] dx = \frac{2}{\sqrt{2\pi}} \left(\frac{\beta}{\gamma}\right)^{\nu/2} K_\nu\left(2\sqrt{\beta\gamma}\right) \\ &= \sqrt{\frac{2}{\pi}} \left(\frac{1/2}{-t}\right)^{\frac{-1/2}{2}} K_{-1/2}\left(2\sqrt{-\frac{t}{2}}\right) = \sqrt{\frac{2}{\pi}} (-2t)^{1/4} K_{-1/2}\left(\sqrt{-2t}\right),\end{aligned}\tag{4.16}$$

onde  $K_\nu(\cdot)$  representa a função Bessel com argumento imaginário, para  $\nu = -1/2$ ,  $\beta = 1/2$  e  $\gamma = -t > 0$  (ver equação 3.471 (9) de Gradshteyn e Ryzhik, 1965). Como estamos no caso particular com  $\nu = -1/2$ , a expressão 8.469 (3) de Gradshteyn e Ryzhik (1965) nos diz que

$$K_{-1/2}(z) = \sqrt{\frac{\pi}{2z}} e^{-z}.$$

Aplicando este resultado à equação (4.16), concluímos que

$$M_{X_1}(t) = \sqrt{\frac{2}{\pi}} (-2t)^{1/4} K_{-1/2}\left(\sqrt{-2t}\right) = \sqrt{\frac{2}{\pi}} (-2t)^{1/4} \sqrt{\frac{\pi}{2\sqrt{-2t}}} e^{-\sqrt{-2t}} = e^{-\sqrt{-2t}}, \quad t < 0.$$

A expressão encontrada para a função geradora de momentos da variável  $X_1$ ,  $M_{X_1}(t) = e^{-\sqrt{-2t}}$ , se  $t < 0$ , coincide com aquela apresentada na proposição 1.2.12 de Samorodnitsky e Taqqu (2000). De fato, esta proposição afirma que se  $X \sim S_\alpha(\sigma, 1, 0)$ , para  $0 < \alpha \leq 2$  e  $\sigma > 0$ , então a Transformada de Laplace  $\mathbb{E}(e^{-\gamma X})$ , para  $\gamma \geq 0$ , é dada por

$$\mathbb{E}(e^{-\gamma X}) = \exp\left\{-\frac{\sigma^\alpha}{\cos\left(\frac{\pi\alpha}{2}\right)} \gamma^\alpha\right\}, \quad \text{se } \alpha \neq 1.\tag{4.17}$$

Definindo  $\alpha = 1/2$  e  $\sigma = 1$ , a expressão (4.17) assume a forma

$$\mathbb{E}(e^{-\gamma X}) = \exp\left\{-\frac{1}{\cos\left(\frac{\pi}{4}\right)} \sqrt{\gamma}\right\} = \exp\left\{-\frac{2}{\sqrt{2}} \sqrt{\gamma}\right\} = e^{-\sqrt{2\gamma}},$$

que para  $\gamma = -t \geq 0$ , é justamente igual à  $M_{X_1}(t) = e^{-\sqrt{-2t}}$ .

Usando novamente o Teorema de Cramér-Chernoff para encontrar a função taxa para a sequência  $(X_n)_{n \in \mathbb{N}}$ , com  $X_1, \dots, X_n$ , variáveis aleatórias i.i.d. seguindo uma distribuição Lévy padrão, temos que

$$M_{X_1}(t) = \begin{cases} e^{-\sqrt{-t}}, & \text{se } t \leq 0, \\ \infty, & \text{se } t > 0. \end{cases}$$

É então suficiente considerar na expressão (2.12) que o supremo seja tomado apenas para  $t < 0$ . Assim, segue que

$$\begin{aligned} I(x) &= \sup_{t < 0} \{tx - \mathcal{L}_{X_1}(t)\} = \sup_{t < 0} \{tx - \log[M_{X_1}(t)]\} = \sup_{t < 0} \{tx - \log(e^{-\sqrt{-2t}})\} \\ &= \sup_{t < 0} \{tx + \sqrt{-2t}\}. \end{aligned}$$

Se  $x = 0$ ,  $I(0) = \sup_{t < 0} \{\sqrt{-2t}\} = \infty$ . Se  $x < 0$ , considere a função  $\phi(t) = tx + \sqrt{-2t}$ , para  $t < 0$ . Tal função é decrescente com relação à  $t$ , e portanto,

$$I(x) = \sup_{t < 0} \{tx + \sqrt{-2t}\} = \lim_{t \rightarrow -\infty} \phi(t) = \infty.$$

Se  $x > 0$ , consideremos novamente a função  $\phi(t) = tx + \sqrt{-2t}$ , para  $t < 0$ . Derivando esta função em relação à  $t$ , obtemos

$$\phi'(t) = x - \frac{1}{\sqrt{-2t}} = 0 \Rightarrow x = \frac{1}{\sqrt{-2t}} \Rightarrow x^2 = \frac{1}{-2t} \Rightarrow t = -\frac{1}{2x^2}, \quad t < 0,$$

enquanto que a segunda derivada de  $\phi(\cdot)$  é dada por

$$\phi''(t) = -\frac{1}{(-2t)^{3/2}} < 0, \quad t < 0.$$

Da análise das duas primeiras derivadas de  $\phi(\cdot)$ , concluímos que esta função atinge o máximo em  $t_0 = -1/(2x^2)$ . Assim,

$$I(x) = \sup_{t < 0} \{tx + \sqrt{-2t}\} = -\frac{1}{2x^2}x + \sqrt{-2 \left(-\frac{1}{2x^2}\right)} = -\frac{1}{2x} + \sqrt{\frac{1}{x^2}} = -\frac{1}{2x} + \frac{1}{x} = \frac{1}{2x}.$$

Em conclusão, a sequência  $(\bar{X}_n)_{n \in \mathbb{N}} = (n^{-1} \sum_{k=1}^n X_k)_{n \in \mathbb{N}}$ , para  $X_1, \dots, X_n$ , i.i.d. Lévy padrão, satisfaz um LDP com função taxa dada por

$$I(x) = \begin{cases} \frac{1}{2x}, & \text{se } x > 0, \\ +\infty, & \text{se } x \leq 0. \end{cases} \quad (4.18)$$

Na Figura 4.1 apresentamos o gráfico da função taxa encontrada. Repare que para  $x > 0$ , esta função é convexa, o que já era esperado.  $\square$

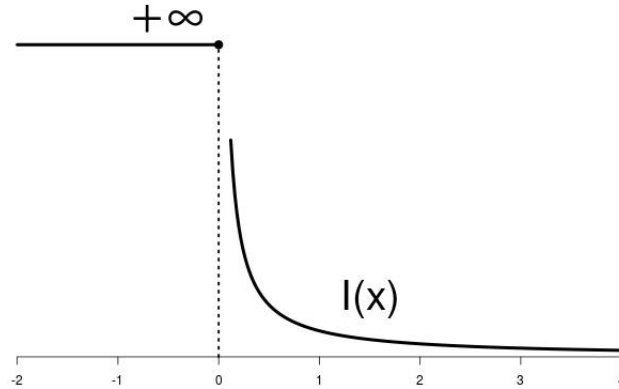


Figura 4.1 – Gráfico da função taxa  $I(\cdot)$  apresentada na expressão (4.18).

## 4.2 Processo AR(1) com Inovações $\alpha$ -Estáveis Simétricas

Para o que segue, consideramos apenas distribuições  $\alpha$ -estáveis simétricas não-Gaussianas, que como já referimos na seção anterior,  $X$  é variável  $S\alpha S$  ( $\alpha$ -estável simétrica) não-Gaussiana se, e somente se,  $X \sim S_\alpha(\sigma, 0, 0)$ , para  $0 < \alpha < 2$  e  $\sigma > 0$  (conforme Propriedade 4.5 (c)). Ainda observamos que, de acordo com a equação (4.4), se  $X \sim S\alpha S$ , então sua função característica  $\varphi_X(t) = \mathbb{E}(e^{itX})$  é dada por

$$\varphi_X(t) = e^{-\sigma^\alpha |t|^\alpha}, \quad \text{para todo } t \in \mathbb{R}. \quad (4.19)$$

**Definição 4.9.** Considere o processo autorregressivo

$$X_{n+1} = \theta X_n + \varepsilon_{n+1}, \quad (4.20)$$

para  $|\theta| < 1$ ,  $n \in \mathbb{N}$  e  $(\varepsilon_n)_{n \geq 2}$  uma sequência de variáveis aleatórias i.i.d. com distribuição  $S_\alpha(\sigma, 0, 0)$ , onde  $0 < \alpha < 2$  e  $\sigma > 0$ . Assim como no caso Gaussiano, assumimos que a variável  $X_1$  é independente da sequência  $(\varepsilon_n)_{n \geq 2}$  e além disso,  $X_1$  tem distribuição  $S_\alpha\left(\frac{\sigma}{(1-|\theta|^\alpha)^{1/\alpha}}, 0, 0\right)$ .

**Observação 4.10:** Quando  $\alpha = 2$  e  $\sigma = 1/\sqrt{2}$ , recaímos no caso Gaussiano estudado no Capítulo 3, pois  $S_2\left(\frac{1}{\sqrt{2}(1-|\theta|^2)^{1/2}}, 0, 0\right) \stackrel{d}{=} \mathcal{N}(0, (1-\theta^2)^{-1})$  (ver Exemplo 4.6).

Com estas hipóteses, podemos afirmar que o vetor  $\mathbf{X}_n = (X_1, \dots, X_n)$ , solução da equação (4.20), possui componentes identicamente distribuídas e que cada uma dessas componentes tem distribuição  $S_\alpha\left(\frac{\sigma}{(1-|\theta|^\alpha)^{1/\alpha}}, 0, 0\right)$ . Podemos observar também que, o parâmetro de escala  $\frac{\sigma}{(1-|\theta|^\alpha)^{1/\alpha}}$ , é o único que torna o processo  $(X_n)_{n \in \mathbb{N}}$  identicamente distribuído.



**Proposição 4.11.** *Considere o processo autorregressivo  $(X_n)_{n \in \mathbb{N}}$  definido na equação (4.20). Assuma que  $X_1$  é independente da sequência  $(\varepsilon_n)_{n \geq 2}$  e além disso, que  $X_1$  tem distribuição  $S_\alpha\left(\frac{\sigma}{(1-|\theta|^\alpha)^{1/\alpha}}, 0, 0\right)$ . Então, a sequência de variáveis aleatórias  $(X_n)_{n \in \mathbb{N}}$ , que é solução da equação (4.20), satisfaz*

$$X_n \sim S_\alpha\left(\frac{\sigma}{(1-|\theta|^\alpha)^{1/\alpha}}, 0, 0\right), \quad (4.21)$$

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

*Demonstração.* Vamos mostrar por indução que vale a expressão (4.21), para qualquer  $n$  natural. Se  $\theta = 0$ , temos que  $X_n = \varepsilon_n$ , para todo  $n \geq 2$ , sendo o resultado imediato a partir da definição das variáveis  $(\varepsilon_n)_{n \geq 2}$ . Vamos supor então que  $0 < |\theta| < 1$ .

Se  $n = 1$ , o resultado segue da convenção de que  $X_1 \sim S_\alpha\left(\frac{\sigma}{(1-|\theta|^\alpha)^{1/\alpha}}, 0, 0\right)$ . Se  $n = 2$ , da maneira que o processo foi definido sabemos que

$$X_2 = \theta X_1 + \varepsilon_2.$$

Da Propriedade 4.5 (b), obtemos que  $\theta X_1 \sim S_\alpha\left(\frac{|\theta|\sigma}{(1-|\theta|^\alpha)^{1/\alpha}}, 0, 0\right)$ . Como por hipótese  $X_1$  é independente de  $(\varepsilon_n)_{n \geq 2}$ , segue que  $\theta X_1$  e  $\varepsilon_2$  são independentes e portanto, a Propriedade 4.5 (a) se aplica, mostrando que

$$X_2 = \theta X_1 + \varepsilon_2 \sim S_\alpha(\tilde{\sigma}, 0, 0),$$

onde  $\tilde{\sigma}$  é determinado pela expressão

$$\tilde{\sigma}^\alpha = \left(\frac{|\theta|\sigma}{(1-|\theta|^\alpha)^{1/\alpha}}\right)^\alpha + \sigma^\alpha = \frac{|\theta|^\alpha \sigma^\alpha + \sigma^\alpha - |\theta|^\alpha \sigma^\alpha}{1-|\theta|^\alpha} = \frac{\sigma^\alpha}{1-|\theta|^\alpha}.$$

Por indução, vamos assumir que (4.21) vale para qualquer  $n \in \mathbb{N}$  e mostremos que é também válida para  $n + 1$ . De fato, note inicialmente que  $X_{n+1}$  pode ser escrito como

$$X_{n+1} = \theta X_n + \varepsilon_{n+1} = \theta^n X_1 + \sum_{k=0}^{n-1} \theta^k \varepsilon_{(n+1)-k}.$$

Sendo a sequência  $(\varepsilon_n)_{n \geq 2}$  i.i.d. com distribuição  $S_\alpha(\sigma, 0, 0)$ , então a combinação linear  $Y_n = \sum_{k=0}^{n-1} \theta^k \varepsilon_{(n+1)-k}$ , para  $n \geq 1$ , segue uma distribuição  $S_\alpha(\sigma_n, 0, 0)$ , onde  $\sigma_n^\alpha = \sigma^\alpha \left(\sum_{k=0}^{n-1} |\theta|^{k\alpha}\right)$ .

De fato, quando  $n = 1$ , temos que  $Y_1 = \varepsilon_2$ , e portanto,  $\sigma_1 = \sigma$ . Se  $n = 2$ , temos que  $Y_2 = \varepsilon_3 + \theta \varepsilon_2$ . Logo,  $\theta \varepsilon_2 \sim S_\alpha(|\theta|\sigma, 0, 0)$ , de onde segue que

$$\sigma_2^\alpha = \sigma^\alpha + |\theta|^\alpha \sigma^\alpha = \sigma^\alpha (1 + |\theta|^\alpha).$$

Por indução, temos que  $Y_{n+1} = \theta Y_n + \varepsilon_{n+2}$ , e conseqüentemente

$$\sigma_{n+1}^\alpha = |\theta|^\alpha (\sigma_n)^\alpha + \sigma^\alpha = |\theta|^\alpha \sigma^\alpha \left(\sum_{k=0}^{n-1} |\theta|^{k\alpha}\right) + \sigma^\alpha = \sigma^\alpha \left(\sum_{k=1}^n |\theta|^{k\alpha} + 1\right) = \sigma^\alpha \left(\sum_{k=0}^n |\theta|^{k\alpha}\right).$$

Demonstrado este fato, observe que  $X_{n+1} = \theta^n X_1 + Y_n$ , onde  $Y_n = \sum_{k=0}^{n-1} \theta^k \varepsilon_{(n+1)-k}$  é a soma definida no penúltimo parágrafo. Note que  $X_1$  é independente da sequência  $(\varepsilon_n)_{n \geq 2}$  e portanto, é também independente da soma  $Y_n$ . Dado que  $\theta^n X_1$  tem distribuição  $S_\alpha \left( |\theta|^n \frac{\sigma}{(1+|\theta|^\alpha)^{1/\alpha}}, 0, 0 \right)$  e  $Y_n$  tem distribuição  $S_\alpha(\sigma_n, 0, 0)$ , onde  $\sigma_n^\alpha = \sigma^\alpha \left( \sum_{k=0}^{n-1} |\theta|^{k\alpha} \right)$ , pela Propriedade 4.5 (a), podemos concluir que  $X_{n+1}$  é  $S\alpha S$  com parâmetro de escala  $\sigma$  satisfazendo a relação

$$\begin{aligned} \sigma^\alpha &= \left( \frac{|\theta|^n \sigma}{(1+|\theta|^\alpha)^{1/\alpha}} \right)^\alpha + \sigma_n^\alpha = \left[ \frac{|\theta|^n \sigma}{(1-|\theta|^\alpha)^{1/\alpha}} \right]^\alpha + \left[ \sigma(1+|\theta|^\alpha + \dots + |\theta|^{(n-1)\alpha})^{1/\alpha} \right]^\alpha \\ &= \frac{|\theta|^{n\alpha} \sigma^\alpha + (1-|\theta|^\alpha) \sigma^\alpha (1+|\theta|^\alpha + \dots + |\theta|^{(n-1)\alpha})}{1-|\theta|^\alpha} \\ &= \frac{\sigma^\alpha}{1-|\theta|^\alpha} \left( |\theta|^{n\alpha} + 1 + |\theta|^\alpha + \dots + |\theta|^{(n-1)\alpha} - |\theta|^\alpha - |\theta|^{2\alpha} - \dots - |\theta|^{n\alpha} \right) \\ &= \frac{\sigma^\alpha}{1-|\theta|^\alpha}, \end{aligned}$$

ou seja,  $X_{n+1}$  tem distribuição  $S_\alpha \left( \frac{\sigma}{(1-|\theta|^\alpha)^{1/\alpha}}, 0, 0 \right)$ , como queríamos demonstrar.  $\square$

Mostramos na proposição a seguir que o processo  $(X_n)_{n \in \mathbb{N}}$ , definido na equação (4.20), é estacionário.

**Proposição 4.12.** *O processo autorregressivo  $(X_n)_{n \in \mathbb{N}}$ , definido pela equação (4.20), é estacionário.*

Para provar esta proposição, faremos uso do seguinte lema.

**Lema 4.13.** *Considere as sequências de variáveis aleatórias  $(X_n)_{n \in \mathbb{N}}$  e  $(\varepsilon_n)_{n \geq 2}$ , apresentadas na Definição 4.9. Então, para todo par de naturais  $m$  e  $n$ , tais que  $m > n$ ,  $X_n$  e  $\varepsilon_m$  são variáveis aleatórias independentes.*

*Demonstração.* Sejam  $m$  e  $n$  dois números naturais pré-fixados, tais que  $m > n$ . Observe que o processo  $X_n$ , dado na expressão (4.20), pode ser reescrito na forma

$$X_n = \theta^{n-1} X_1 + \sum_{k=2}^n \theta^{n-k} \varepsilon_k.$$

Assim, como  $X_n$  é uma combinação linear de  $X_1, \varepsilon_2, \dots, \varepsilon_n$ , e estas variáveis são por hipótese, independentes de  $\varepsilon_m$ , para todo  $m > n$ , o resultado segue.  $\square$

**Demonstração da Proposição 4.12.** Para mostrar que o processo  $(X_n)_{n \in \mathbb{N}}$  é estacionário, devemos provar que as distribuições conjuntas de

$$\tilde{\mathbf{X}}_0 = (X_{n_1}, X_{n_2}, \dots, X_{n_j}) \quad \text{e de} \quad \tilde{\mathbf{X}}_h = (X_{n_1+h}, X_{n_2+h}, \dots, X_{n_j+h}) \quad (4.22)$$

são iguais, para quaisquer naturais  $n_1, n_2, \dots, n_j, j$  e  $h$ . Provamos apenas o caso em que  $n_1 < n_2 < \dots < n_j$ . Os demais seguem deste.

A função característica do vetor aleatório  $\tilde{\mathbf{X}}_0$  é dada por

$$\varphi_{\tilde{\mathbf{X}}_0}(t_1, t_2, \dots, t_j) = \mathbb{E}\{\exp[i(t_1 X_{n_1} + t_2 X_{n_2} + \dots + t_j X_{n_j})]\}.$$

Escrevendo as variáveis aleatórias  $X_{n_2}, \dots, X_{n_j}$  em função de  $X_{n_1}$  e da sequência  $(\varepsilon_n)_{n \geq 2}$ , tal como na equação (3.23), obtemos

$$\begin{aligned} X_{n_2} &= \theta^{n_2 - n_1} X_{n_1} + \sum_{k'=1}^{n_2 - n_1} \theta^{n_2 - n_1 - k'} \varepsilon_{n_1 + k'} = \theta^{n_2 - n_1} X_{n_1} + \sum_{k=n_1+1}^{n_2} \theta^{n_2 - k} \varepsilon_k, \\ X_{n_3} &= \theta^{n_3 - n_1} X_{n_1} + \sum_{k'=1}^{n_3 - n_1} \theta^{n_3 - n_1 - k'} \varepsilon_{n_1 + k'} = \theta^{n_3 - n_1} X_{n_1} + \sum_{l=2}^3 \left( \sum_{k=n_{l-1}+1}^{n_l} \theta^{n_3 - k} \varepsilon_k \right), \\ &\vdots \\ X_{n_j} &= \theta^{n_j - n_1} X_{n_1} + \sum_{k'=1}^{n_j - n_1} \theta^{n_j - n_1 - k'} \varepsilon_{n_1 + k'} = \theta^{n_j - n_1} X_{n_1} + \sum_{l=2}^j \left( \sum_{k=n_{l-1}+1}^{n_l} \theta^{n_j - k} \varepsilon_k \right). \end{aligned} \quad (4.23)$$

Segue da expressão (4.23) que a função característica de  $\tilde{\mathbf{X}}_0$  é igual à

$$\begin{aligned} \varphi_{\tilde{\mathbf{X}}_0}(t_1, t_2, \dots, t_j) &= \mathbb{E}\{\exp[i(t_1 X_{n_1} + t_2 X_{n_2} + \dots + t_j X_{n_j})]\} \\ &= \mathbb{E}\left\{\exp\left[i\left(\sum_{k=1}^j t_k \theta^{n_k - n_1} X_{n_1} + t_2 \sum_{k=n_1+1}^{n_2} \theta^{n_2 - k} \varepsilon_k + \dots + t_j \sum_{l=2}^j \left(\sum_{k=n_{l-1}+1}^{n_l} \theta^{n_j - k} \varepsilon_k\right)\right)\right]\right\} \\ &= \mathbb{E}\left\{\exp\left[i\left(\sum_{k=1}^j t_k \theta^{n_k - n_1} X_{n_1} + \sum_{m=2}^j \sum_{l=m}^j t_l \sum_{k=n_{m-1}+1}^{n_m} \theta^{n_l - k} \varepsilon_k\right)\right]\right\}. \end{aligned} \quad (4.24)$$

Observe agora que a parcela do somatório que contém apenas as variáveis  $\varepsilon_n$ 's na expressão (4.24), começa em  $n_1 + 1$ . Pelo Lema 4.13, segue que  $X_{n_1}$  e  $(\varepsilon_n)_{n > n_1}$  são independentes, de onde podemos concluir que  $\sum_{k=1}^j t_k \theta^{n_k - n_1} X_{n_1}$  e  $\sum_{m=2}^j \sum_{l=m}^j t_l \sum_{k=n_{m-1}+1}^{n_m} \theta^{n_l - k} \varepsilon_k$  também são independentes. Das propriedades da função característica, obtemos que

$$\begin{aligned} \varphi_{\tilde{\mathbf{X}}_0}(t_1, t_2, \dots, t_j) &= \\ &= \mathbb{E}\left\{\exp\left[i\left(\sum_{k=1}^j t_k \theta^{n_k - n_1} X_{n_1}\right)\right]\right\} \times \mathbb{E}\left\{\exp\left[i\left(\sum_{m=2}^j \sum_{l=m}^j t_l \sum_{k=n_{m-1}+1}^{n_m} \theta^{n_l - k} \varepsilon_k\right)\right]\right\} \\ &= \varphi_{X_{n_1}}\left(\sum_{k=1}^j t_k \theta^{n_k - n_1}\right) \times \mathbb{E}\left\{\exp\left[i\left(\sum_{m=2}^j \sum_{l=m}^j t_l \sum_{k=n_{m-1}+1}^{n_m} \theta^{n_l - k} \varepsilon_k\right)\right]\right\}. \end{aligned} \quad (4.25)$$

Raciocinando da mesma maneira para o vetor  $\tilde{\mathbf{X}}_h = (X_{n_1+h}, X_{n_2+h}, \dots, X_{n_j+h})$ , observamos que, para qualquer  $j \geq 2$ ,

$$\begin{aligned} X_{n_j+h} &= \theta^{n_j - n_1} X_{n_1+h} + \sum_{k'=1}^{n_j - n_1} \theta^{n_j - n_1 - k'} \varepsilon_{n_1 + k' + h} \\ &= \theta^{n_j - n_1} X_{n_1+h} + \sum_{l=2}^j \left( \sum_{k=n_{l-1}+1}^{n_l} \theta^{n_j - k} \varepsilon_{k+h} \right), \end{aligned}$$

de onde obtemos a função característica

$$\begin{aligned}
\varphi_{\tilde{\mathbf{X}}_h}(t_1, t_2, \dots, t_j) &= \mathbb{E}\{\exp[i(t_1 X_{n_1+h} + t_2 X_{n_2+h} + \dots + t_j X_{n_j+h})]\} \\
&= \mathbb{E}\left\{\exp\left[i\left(\sum_{k=1}^j t_k \theta^{n_k - n_1} X_{n_1+h} + \sum_{m=2}^j \sum_{l=m}^j t_l \sum_{k=n_{m-1}+1}^{n_m} \theta^{n_l - k} \varepsilon_{k+h}\right)\right]\right\} \\
&= \mathbb{E}\left\{\exp\left[i\left(\sum_{k=1}^j t_k \theta^{n_k - n_1} X_{n_1+h}\right)\right]\right\} \times \mathbb{E}\left\{\exp\left[i\left(\sum_{m=2}^j \sum_{l=m}^j t_l \sum_{k=n_{m-1}+1}^{n_m} \theta^{n_l - k} \varepsilon_{k+h}\right)\right]\right\} \\
&= \varphi_{X_{n_1+h}}\left(\sum_{k=1}^j t_k \theta^{n_k - n_1}\right) \times \mathbb{E}\left\{\exp\left[i\left(\sum_{m=2}^j \sum_{l=m}^j t_l \sum_{k=n_{m-1}+1}^{n_m} \theta^{n_l - k} \varepsilon_{k+h}\right)\right]\right\}. \quad (4.26)
\end{aligned}$$

Uma vez que  $X_{n_1+h}$  e  $X_{n_1}$  são variáveis aleatórias identicamente distribuídas, concluímos que as suas funções características são iguais. Além disso, da hipótese de que  $(\varepsilon_n)_{n \geq 2}$  é uma sequência i.i.d., sabemos que é também um processo estacionário (ver exemplo A da p. 445 de Karlin e Taylor, 1975). Destes dois fatos, segue a igualdade das funções características  $\varphi_{\tilde{\mathbf{X}}_0}(t_1, t_2, \dots, t_j)$  e  $\varphi_{\tilde{\mathbf{X}}_h}(t_1, t_2, \dots, t_j)$ , dadas respectivamente pelas expressões (4.25) e (4.26), mostrando que  $\tilde{\mathbf{X}}_0 = (X_{n_1}, \dots, X_{n_j})$  e  $\tilde{\mathbf{X}}_h = (X_{n_1+h}, \dots, X_{n_j+h})$  são identicamente distribuídas, concluindo assim, a prova da Proposição 4.12.  $\square$

**Observação 4.14:** Note que na demonstração da Proposição 4.12, em nenhum momento usamos a forma particular da função característica de uma distribuição  $\alpha$ -estável. Na verdade, o que provamos, é um resultado muito mais forte do que o estabelecido pela proposição. Dado que a função característica de uma variável aleatória sempre existe, vamos enunciar a seguir este resultado e sua demonstração é idêntica à da Proposição 4.12.

**Proposição 4.15.** *Considere o processo autorregressivo*

$$X_{n+1} = \theta X_n + \varepsilon_{n+1},$$

para  $|\theta| < 1$ ,  $n \in \mathbb{N}$  e  $(\varepsilon_n)_{n \geq 2}$  uma sequência de variáveis aleatórias i.i.d. com distribuição qualquer. Assuma também que  $X_1$  é independente da sequência  $(\varepsilon_n)_{n \geq 2}$  e que  $(X_n)_{n \in \mathbb{N}}$  é uma sequência identicamente distribuída. Então o processo  $(X_n)_{n \in \mathbb{N}}$  é estacionário.

Note que a Proposição 4.12, que aborda o caso  $\alpha$ -estável, é desta forma apenas um caso particular da Proposição 4.15. Ressaltamos também que, o resultado da Proposição 4.15 não foi encontrado em nenhuma das referências pesquisadas, representando desta forma, um de nossos achados durante este trabalho.

### 4.2.1 Codiferença

Queremos estender os resultados de LDP's que encontramos no Capítulo 3, para o parâmetro  $\theta$  dado na equação 4.20. Para tanto, observamos que, para  $0 < \alpha < 2$ , não

podemos definir a matriz de autocovariância do processo, devido ao fato de que os momentos de segunda ordem são infinitos (ver Propriedade 4.5 (d)). Assim, precisamos generalizar esta matriz para tais processos, de tal forma que, no caso  $\alpha = 2$ , sua generalização coincida com a matriz de autocovariância.

**Definição 4.16.** Seja  $\mathbf{X}_n = (X_1, \dots, X_n)$  um vetor aleatório com componentes  $S\alpha S$ , para  $0 < \alpha \leq 2$ . A matriz de codiferença do vetor  $\mathbf{X}_n$  é dada por

$$\Upsilon_{\mathbf{X}_n} = \begin{pmatrix} \tau_{X_1, X_1} & \tau_{X_1, X_2} & \cdots & \tau_{X_1, X_n} \\ \tau_{X_2, X_1} & \tau_{X_2, X_2} & \cdots & \tau_{X_2, X_n} \\ \vdots & \vdots & \ddots & \vdots \\ \tau_{X_n, X_1} & \tau_{X_n, X_2} & \cdots & \tau_{X_n, X_n} \end{pmatrix}, \quad (4.27)$$

onde  $\tau_{X,Y}$  é a função codiferença das variáveis  $X$  e  $Y$ , definida por

$$\tau_{X,Y} = \|X\|_\alpha^\alpha + \|Y\|_\alpha^\alpha - \|X - Y\|_\alpha^\alpha, \quad (4.28)$$

desde que  $\|X\|_\alpha$ ,  $\|Y\|_\alpha$  e  $\|X - Y\|_\alpha$  denotem, respectivamente, os parâmetros de escala de  $X$ ,  $Y$  e  $X - Y$ .

**Observação 4.17:** Quando estivermos perante um processo estacionário  $(X_n)_{n \in \mathbb{N}}$ , com índice de estabilidade  $\alpha \in (0, 2]$ , vamos usar a notação  $\tau_X(k; \alpha)$  para representar a função codiferença no lag  $k$ , isto é,

$$\tau_X(k; \alpha) = \tau_{X_l, X_j},$$

sempre que  $k = l - j \in \mathbb{Z}$ . A notação  $\tau(k; \alpha)$ , com o ponto e vírgula (;) separando as variáveis  $k$  e  $\alpha$ , serve para dizer que  $k$  é uma variável muda, variando nos inteiros, e que  $\alpha$  é uma variável livre, variando no intervalo  $(0, 2]$ . O que queremos transmitir com esta notação é que, apesar da função  $\tau(k; \alpha)$  ser dependente de duas variáveis, apenas a primeira delas, nomeadamente  $k$ , é uma variável que está variando de fato. Já a segunda variável, nomeadamente  $\alpha$ , é uma variável que iremos fixar previamente de acordo com o contexto, sendo portanto,  $\{\tau(\cdot; \alpha) \mid 0 < \alpha \leq 2\}$  uma classe de funções reais de variáveis inteiras.

A matriz de codiferença possui propriedades similares às da matriz de autocovariância, herdadas da função codiferença. Em concordância com a seção 2.10 de Samorodnitsky e Taqqu (2000), enunciamos a seguir uma proposição que reúne as propriedades mais pertinentes ao nosso estudo.

**Proposição 4.18.** *Sejam  $X$  e  $Y$  variáveis aleatórias  $S\alpha S$ , para  $0 < \alpha \leq 2$ . Com respeito à função codiferença, apresentada na Definição 4.16, são válidas as seguintes afirmações:*

(a)  $\tau_{X,Y} = \tau_{Y,X}$ ;

(b) Se  $\alpha = 2$ , então  $\tau_{X,Y} = \text{Cov}(X, Y)$ .

Da Proposição 4.18 acima e da propriedade 2.10.7, apresentada em Samorodnitsky e Taqqu (2000, p. 105-106), podemos deduzir as seguintes propriedades da matriz de codiferença.

**Propriedade 4.19. (Propriedades da matriz de codiferença).** Seja  $\mathbf{X}_n$  um vetor aleatório cujas componentes são  $S\alpha S$ , para  $0 < \alpha \leq 2$  e  $\Upsilon_{\mathbf{X}_n}$  a sua respectiva matriz de codiferença, dada pela expressão (4.27). Então:

- (a)  $\Upsilon_{\mathbf{X}_n}$  é simétrica;
- (b)  $\Upsilon_{\mathbf{X}_n}$  é uma matriz não-negativa definida;
- (c) Se  $\alpha = 2$ , a matriz  $\Upsilon_{\mathbf{X}_n}$  coincide com a matriz de autocovariância do vetor  $\mathbf{X}_n$ .

#### 4.2.2 Matriz de Codiferença de um Processo AR(1)

Vamos calcular a matriz de codiferença para o vetor  $\mathbf{X}_n = (X_1, \dots, X_n)$ , cujas componentes são as respectivas variáveis do processo  $(X_n)_{n \in \mathbb{N}}$ , apresentado na equação 4.20.

**Proposição 4.20.** Considere o processo  $(X_n)_{n \in \mathbb{N}}$  definido pela equação (4.20), onde  $(\varepsilon_n)_{n \geq 2}$  é uma seqüência de variáveis aleatórias i.i.d. tal que  $\varepsilon_n \sim S_\alpha(\sigma, 0, 0)$ , para todo  $n \geq 2$ , com  $0 < \alpha \leq 2$  e  $\sigma > 0$ . A matriz de codiferença do vetor aleatório  $\mathbf{X}_n = (X_1, X_2, \dots, X_n)$ , cujas componentes são as respectivas variáveis do processo  $(X_n)_{n \in \mathbb{N}}$ , é dada por

$$\Upsilon_{\mathbf{X}_n} = \frac{\sigma^\alpha}{1 - |\theta|^\alpha} \left( 1 + |\theta|^{\alpha|l-j|} - (1 - \theta^{l-j})^\alpha \right)_{1 \leq l, j \leq n}. \quad (4.29)$$

*Demonstração.* Para qualquer  $X_k$ ,  $1 \leq k \leq n$ , temos que  $X_k \sim S_\alpha \left( \frac{\sigma}{(1 - |\theta|^\alpha)^{1/\alpha}}, 0, 0 \right)$ . Assim, da expressão (4.28) segue que

$$\tau_{X_k, X_k} = 2 \|X_k\|_\alpha^\alpha = \frac{2\sigma^\alpha}{1 - |\theta|^\alpha}, \quad \text{para todo } 1 \leq k \leq n.$$

Assuma que as variáveis  $X_1$  e  $X_2$  sejam identicamente distribuídas, podendo ser dependentes. Isso impede a aplicação da Propriedade 4.5 (a) de uma maneira direta. No entanto,  $X_2$  é da forma  $X_2 = \theta X_1 + \varepsilon_2$ , de onde segue que

$$X_1 - X_2 = (1 - \theta)X_1 - \varepsilon_2. \quad (4.30)$$

Assim, da hipótese de independência entre  $X_1$  e  $(\varepsilon_n)_{n \geq 2}$  e da Propriedade 4.5 (b),  $(1 - \theta)X_1$  e  $-\varepsilon_2$  são independentes, com distribuições  $S_\alpha \left( \frac{(1 - \theta)\sigma}{(1 - |\theta|^\alpha)^{1/\alpha}}, 0, 0 \right)$  e  $S_\alpha(\sigma, 0, 0)$ , respectivamente.

Aplicando a Propriedade 4.5 (a), concluímos da equação (4.30) que  $X_1 - X_2$  segue uma distribuição  $S_\alpha(\sigma_{1,2}, 0, 0)$ , onde  $\sigma_{1,2}$  é dado por

$$\sigma_{1,2}^\alpha = \left[ \frac{(1-\theta)\sigma}{(1-|\theta|^\alpha)^{1/\alpha}} \right]^\alpha + \sigma^\alpha = \frac{(1-\theta)^\alpha \sigma^\alpha + \sigma^\alpha - |\theta|^\alpha \sigma^\alpha}{1-|\theta|^\alpha} = \sigma^\alpha \left( \frac{1 + (1-\theta)^\alpha - |\theta|^\alpha}{1-|\theta|^\alpha} \right).$$

Por fim, a codiferença entre  $X_1$  e  $X_2$  pode ser expressa como

$$\begin{aligned} \tau_{X_1, X_2} &= \|X_1\|_\alpha^\alpha + \|X_2\|_\alpha^\alpha - \|X_1 - X_2\|_\alpha^\alpha \\ &= 2 \frac{\sigma^\alpha}{1-|\theta|^\alpha} - \sigma^\alpha \left( \frac{1 + (1-\theta)^\alpha - |\theta|^\alpha}{1-|\theta|^\alpha} \right) \\ &= \sigma^\alpha \left( \frac{1 + |\theta|^\alpha - (1-\theta)^\alpha}{1-|\theta|^\alpha} \right). \end{aligned}$$

Uma vez que a função codiferença é simétrica, concluímos também que

$$\tau_{X_2, X_1} = \tau_{X_1, X_2} = \sigma^\alpha \left( \frac{1 + |\theta|^\alpha - (1-\theta)^\alpha}{1-|\theta|^\alpha} \right).$$

Além disso, sendo o processo  $(X_n)_{n \in \mathbb{N}}$  estacionário, segue que no lag 1 a função codiferença é igual à  $\tau_{X_1, X_2}$ , isto é,

$$\tau_X(1; \alpha) = \tau_{X_1, X_2} = \sigma^\alpha \left( \frac{1 + |\theta|^\alpha - (1-\theta)^\alpha}{1-|\theta|^\alpha} \right).$$

Vamos agora considerar naturais  $l$  e  $j$  tais que  $1 \leq l, j \leq n$ . Pretendemos determinar o valor de  $\tau_{X_l, X_j}$ . Sem perda de generalidade, assumimos que  $l < j$  e  $j - l = h \geq 2$  (no caso de termos  $l > j$ , basta usar a simetria da função codiferença). Assim como no caso em que  $l = 1$  e  $j = 2$ , não temos independência entre as variáveis  $X_l$  e  $X_j$ , mas podemos escrever  $X_j$  em função de  $X_l$  e das demais variáveis  $(\varepsilon_n)_{n>l}$ . De fato, pela equação (3.23), temos que

$$X_j = \theta^{j-l} X_l + \sum_{k=1}^{j-l} \theta^{j-l-k} \varepsilon_{l+k} = \theta^h X_l + \sum_{k=1}^h \theta^{h-k} \varepsilon_{l+k},$$

de onde segue imediatamente que

$$X_l - X_j = (1 - \theta^h) X_l - \sum_{k=1}^h \theta^{h-k} \varepsilon_{l+k}.$$

Com cálculos análogos aos apresentados na demonstração da Proposição 4.11, chegamos à conclusão que

$$\sum_{k=1}^h \theta^{h-k} \varepsilon_{l+k} \sim S_\alpha(\sigma_h, 0, 0),$$

onde  $\sigma_h^\alpha = \sigma^\alpha \left( \sum_{k=0}^{h-1} |\theta|^{k\alpha} \right)$ . Procedendo de maneira análoga aos casos anteriores, obtemos que  $X_l - X_j$  segue uma distribuição  $S_\alpha(\sigma_{l,j}, 0, 0)$ , onde  $\sigma_{l,j}$  é dado por

$$\begin{aligned} \sigma_{l,j}^\alpha &= \left[ \frac{(1-\theta^h)\sigma}{(1-|\theta|^\alpha)^{1/\alpha}} \right]^\alpha + \sigma_h^\alpha = \sigma^\alpha \frac{(1-\theta^h)^\alpha}{1-|\theta|^\alpha} + \sigma^\alpha \sum_{k=0}^{h-1} (|\theta|^\alpha)^k \\ &= \sigma^\alpha \left[ \frac{(1-\theta^h)^\alpha}{1-|\theta|^\alpha} + \frac{1-|\theta|^{\alpha h}}{1-|\theta|^\alpha} \right] = \sigma^\alpha \left( \frac{(1-\theta^h)^\alpha + 1 - |\theta|^{\alpha h}}{1-|\theta|^\alpha} \right). \end{aligned}$$

E por fim, concluímos que  $\tau_X(h; \alpha) = \tau_{X_l, X_j}$  é dado por

$$\begin{aligned} \tau_X(h; \alpha) &= \tau_{X_l, X_j} = \|X_l\|_\alpha^\alpha + \|X_j\|_\alpha^\alpha - \|X_l - X_j\|_\alpha^\alpha \\ &= 2 \frac{\sigma^\alpha}{1 - |\theta|^\alpha} - \sigma^\alpha \left( \frac{(1 - \theta^h)^\alpha + 1 - |\theta|^{\alpha h}}{1 - |\theta|^\alpha} \right) \\ &= \sigma^\alpha \left( \frac{2 - 1 - (1 - \theta^h)^\alpha + |\theta|^{\alpha h}}{1 - |\theta|^\alpha} \right) = \sigma^\alpha \left( \frac{1 + |\theta|^{\alpha h} - (1 - \theta^h)^\alpha}{1 - |\theta|^\alpha} \right). \end{aligned}$$

□

**Observação 4.21:** Nowicka-Zagrajek e Wyłomańska (2008) mostram no exemplo 3.1, o resultado da Proposição 4.20 no caso em que  $\alpha \in (1, 2]$  e o processo  $(X_n)$  tem índice  $n$  inteiro.

Como corolário, obtemos o caso Gaussiano.

**Corolário 4.22.** *Considere o processo autorregressivo  $(X_n)_{n \in \mathbb{N}}$  definido pela equação*

$$X_{n+1} = \theta X_n + \varepsilon_{n+1}, \quad |\theta| < 1,$$

onde  $(\varepsilon_n)_{n \geq 2}$  é uma sequência de variáveis aleatórias i.i.d. com distribuição  $\mathcal{N}(0, 1)$ . Assuma que  $X_1$  é independente da sequência  $(\varepsilon_n)_{n \geq 2}$  e que tenha distribuição  $\mathcal{N}(0, (1 - \theta^2)^{-1})$ . Então, o vetor aleatório  $\mathbf{X}_n = (X_1, \dots, X_n)$  tem matriz de autocovariância

$$\left( \gamma_{X_l, X_j} \right)_{1 \leq l, j \leq n} = \left( \mathbb{E}(X_l X_j) \right)_{1 \leq l, j \leq n} = \left( \frac{\theta^{|l-j|}}{1 - \theta^2} \right)_{1 \leq l, j \leq n}. \quad (4.31)$$

*Demonstração.* De fato, fixando  $\alpha = 2$  e  $\sigma := 1/\sqrt{2}$ , o elemento na linha  $l$  e coluna  $j$  da matriz  $\Upsilon_{\mathbf{X}_n}$  exibida em (4.29) é dado por

$$\begin{aligned} (\Upsilon_{\mathbf{X}_n})_{l,j} &= \frac{\sigma^\alpha \left( 1 + |\theta|^{\alpha|l-j|} - (1 - \theta^{|l-j|})^\alpha \right)}{1 - |\theta|^\alpha} = \frac{1}{2(1 - \theta^2)} \left[ 1 + \theta^{2|l-j|} - (1 - \theta^{|l-j|})^2 \right] \\ &= \frac{1}{2(1 - \theta^2)} \left( 1 + \theta^{2|l-j|} - 1 + 2\theta^{|l-j|} - \theta^{2|l-j|} \right) = \frac{\theta^{|l-j|}}{1 - \theta^2}, \end{aligned}$$

para todo par  $(l, j)$ , com  $1 \leq l, j \leq n$ , o que coincide com a expressão apresentada em (4.31). □

**Observação 4.23:** Lembramos que, pela forma que definimos a matriz de Toeplitz na equação (3.16), a matriz  $T_n(g)$  da função  $g(t) = (1 + \theta^2 - 2\theta \cos(t))^{-1}$ , para  $t \in \mathbb{T} = [-\pi, \pi)$ , coincide, pelo Teorema de Herglotz (ver Teorema 4.37), com a matriz de autocovariância  $\left( \gamma_{X_l, X_j} \right)_{1 \leq l, j \leq n}$  apresentada na equação (4.31).

Não podemos deixar de observar que a matriz de codiferença encontrada na Proposição 4.20, é ainda uma matriz de Toeplitz. De fato, no sentido definido por Gray



(2009), uma matriz quadrada  $A = (a_{l,j})_{1 \leq l, j \leq n}$  de ordem  $n$ , é dita ser uma *matriz de Toeplitz* se  $a_{l,j} = a_{l-j}$ , isto é,  $A$  é uma matriz da forma

$$A = \begin{pmatrix} a_0 & a_{-1} & a_{-2} & \dots & a_{-(n-1)} \\ a_1 & a_0 & a_{-1} & \dots & a_{-(n-2)} \\ a_2 & a_1 & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & a_{-1} \\ a_{n-1} & a_{n-2} & \dots & a_1 & a_0 \end{pmatrix}. \quad (4.32)$$

Particularizando para o caso da matriz de codiferença do vetor  $\mathbf{X}_n = (X_1, \dots, X_n)$ , devemos lembrar que sendo  $(X_n)_{n \in \mathbb{N}}$  um processo estacionário, então podemos definir

$$a_k = \tau_X(k; \alpha) = \tau_{X_l, X_j}, \quad \text{sempre que } k = l - j \in \mathbb{Z},$$

de maneira que  $\Upsilon_{\mathbf{X}_n}$  tem a forma da matriz exibida em (4.32). Note que, pela estacionariedade do processo,  $k$  não depende da representação específica do par  $(l, j) \in \mathbb{N}^2$ .

Com este fato em mente, nos perguntamos: Será possível existir uma função real  $g(\cdot)$ , definida no toro  $\mathbb{T} = [-\pi, \pi)$ , tal que  $\Upsilon_{\mathbf{X}_n} = T_n(g)$ ? Em outras palavras, será que a matriz  $\Upsilon_{\mathbf{X}_n}$  pode ser reescrita como uma matriz de Toeplitz de alguma função  $g(\cdot)$ , no sentido clássico definido por Grenander e Szegö (1958, p. 102), que coincide com nossa definição de matriz de Toeplitz (ver Definição 3.6)?

Na próxima seção, daremos uma resposta concreta às perguntas levantadas no parágrafo anterior e mostramos que de fato, a função  $g(\cdot)$  existe e que podemos escrever a matriz de codiferença do vetor  $\mathbf{X}_n$ , como uma matriz de Toeplitz no sentido clássico, para o caso  $\alpha$ -estável não-Gaussiano. Mas antes disso, analisamos algumas propriedades assintóticas da codiferença e propriedades ergódicas do processo  $S\alpha S (X_n)_{n \in \mathbb{N}}$ .

### 4.2.3 Comportamento Assintótico da Codiferença de um Processo AR(1)

Estabelecida uma fórmula fechada para a função codiferença do processo  $(X_n)_{n \in \mathbb{N}}$ , definido pela equação (4.20), é importante analisar o comportamento assintótico da sua função codiferença.

No caso Gaussiano, é bem conhecida a propriedade de que, a função de autocovariância de um processo AR(1) estacionário, converge à zero quando o lag desta mesma função converge ao infinito, isto é,

$$\lim_{h \rightarrow \infty} \gamma_X(h) = \lim_{h \rightarrow \infty} \mathbb{E} \{ [X_{k+h} - \mathbb{E}(X_{k+h})][X_k - \mathbb{E}(X_k)] \} = 0, \quad (4.33)$$

para todo  $k \in \mathbb{N}$  (veja o exemplo 3.1.2 de Brockwell e Davis, 1991). Vamos mostrar que, a partir da expressão (4.29), conseguimos uma propriedade análoga no caso  $\alpha$ -estável.

**Corolário 4.24.** Considere o processo  $(X_n)_{n \in \mathbb{N}}$  definido pela equação (4.20), onde  $(\varepsilon_n)_{n \geq 2}$  é uma sequência de variáveis aleatórias i.i.d. tal que  $\varepsilon_n \sim S_\alpha(\sigma, 0, 0)$ , para todo  $n \geq 2$ , com  $0 < \alpha < 2$  e  $\sigma > 0$ . Seja  $\tau_X(k; \alpha)$  a função codiferença associada ao processo  $(X_n)_{n \in \mathbb{N}}$  no lag  $k$ , cuja expressão é dada por  $\tau_X(k; \alpha) = \tau_{X_{k+1}, X_1}$ , para todo  $k \geq 0$ . Então

$$\lim_{k \rightarrow \infty} \tau_X(k; \alpha) = 0. \quad (4.34)$$

*Demonstração.* Pela expressão (4.29), sabemos que

$$\tau_X(k; \alpha) = \tau_{X_{k+1}, X_1} = \frac{\sigma^\alpha}{1 - |\theta|^\alpha} \left( 1 + |\theta|^{\alpha k} - (1 - \theta^k)^\alpha \right).$$

Passando ao limite a expressão acima e lembrando que  $|\theta| < 1$ , notamos que  $|\theta|^{\alpha k}$  e  $\theta^k$  convergem à zero, quando  $k \rightarrow \infty$ . Isto é suficiente para concluirmos à partir da linearidade e continuidade do limite que

$$\begin{aligned} \lim_{k \rightarrow \infty} \tau_X(k; \alpha) &= \lim_{k \rightarrow \infty} \frac{\sigma^\alpha}{1 - |\theta|^\alpha} \left( 1 + |\theta|^{\alpha k} - |1 - \theta^k|^\alpha \right) \\ &= \frac{\sigma^\alpha}{1 - |\theta|^\alpha} \left( 1 + \lim_{k \rightarrow \infty} |\theta|^{\alpha k} - |1 - \lim_{k \rightarrow \infty} \theta^k|^\alpha \right) = \frac{\sigma^\alpha}{1 - |\theta|^\alpha} (1 - |1|) = 0. \end{aligned}$$

□

#### 4.2.4 Processos Mixing e Ergódicos

Nesta subseção, vamos recordar o que são processos mixing e ergódicos (ver seção 5.4 de Lopes e Lopes, 2016). Em seguida, iremos mostrar que o processo  $(X_n)_{n \in \mathbb{N}}$ , apresentado na equação 4.20, é um processo mixing e ergódico.

**Definição 4.25.** Seja  $\mathbf{x} = (x_1, x_2, \dots)$  uma sequência real e denote por  $\vec{T}(\cdot)$  o operador *shift*, que à  $\mathbf{x}$ , associa a sequência  $\vec{T}(\mathbf{x}) = (x_2, x_3, \dots)$ . Então:

- (a) Dizemos que um conjunto  $A \subseteq \mathbb{R}^\infty$  de sequências reais é *invariante pelo shift* se  $\vec{T}^{-1}(A) = A$ . Em outras palavras,  $A$  é invariante para o *shift* quando  $\mathbf{x} \in A$  se, e somente se,  $\vec{T}(\mathbf{x}) \in A$ .
- (b) Um processo estacionário  $(Y_n)_{n \in \mathbb{N}}$  é dito *ergódico* se  $\mathbb{P}\{(Y_1, Y_2, \dots) \in A\}$  é igual à zero ou um, sempre que  $A$  for um conjunto invariante. Isto equivale à dizer que os únicos conjuntos invariantes são triviais ( $A$  é *trivial* se  $\mathbb{P}(A) \in \{0, 1\}$ ).
- (c) Dizemos que  $(Y_n)_{n \in \mathbb{N}}$  é um processo *mixing* se, para quaisquer conjuntos  $A$  e  $B$  de sequências reais,

$$\begin{aligned} &\lim_{n \rightarrow \infty} \mathbb{P}\{(Y_1, Y_2, \dots) \in A, (Y_{n+1}, Y_{n+2}, \dots) \in B\} = \\ &= \mathbb{P}\{(Y_1, Y_2, \dots) \in A\} \times \mathbb{P}\{(Y_1, Y_2, \dots) \in B\}. \end{aligned} \quad (4.35)$$

Estabelecidas as definições de processos *mixing* e *ergódicos*, obtemos a seguinte proposição, cuja demonstração é simples e pode ser encontrada em Lopes e Lopes (2016, p. 477).

**Proposição 4.26.** *Se a medida de probabilidade  $\mathbb{P}$  é mixing para o shift  $\vec{T}(\cdot)$ , então  $\mathbb{P}$  é ergódica para  $\vec{T}(\cdot)$ .*

A Proposição 4.26 exhibe uma condição suficiente para que um processo seja ergódico. Esporadicamente, a prova de que um processo tem a propriedade de ser *mixing*, pode parecer mais simples do que a demonstração de que o processo é ergódico. Vamos apresentar a seguir, um critério extremamente prático para mostrar quando um processo infinitamente divisível é mixing. Este critério contorna a dificuldade de recorrermos imediatamente à Definição 4.25, pois às vezes pode ser difícil calcular o limite que figura na equação (4.35), para certos processos estacionários. Mas antes, vamos definir o que são processos infinitamente divisíveis (para uma explanação mais detalhada do assunto, veja o capítulo 3 de Gnedenko et al., 1954, para o caso real, ou o capítulo 2 de Sato, 1999, para o caso em  $\mathbb{R}^k$ , para  $k \in \mathbb{N}$ ). Definimos vetores infinitamente divisíveis através de sua função característica.

**Definição 4.27.** Dizemos que o vetor aleatório  $\mathbf{Z}$  definido em  $\mathbb{R}^k$ , para  $k \in \mathbb{N}$ , é *infinitamente divisível* se, para cada número natural  $n$ , dada a função característica  $\varphi_{\mathbf{Z}}(\cdot)$  de  $\mathbf{Z}$ , existe uma variável aleatória  $\mathbf{Z}_n$  definida em  $\mathbb{R}^k$  e com função característica  $\varphi_n(\cdot)$  tal que

$$\varphi_{\mathbf{Z}}(\mathbf{z}) = [\varphi_n(\mathbf{z})]^n, \quad \text{para todo } \mathbf{z} \in \mathbb{R}^k. \quad (4.36)$$

As funções de distribuição de vetores aleatórios infinitamente divisíveis, são chamadas de *funções de distribuição infinitamente divisíveis*.

Tendo em vista a definição de vetores infinitamente divisíveis, podemos definir também processos estocásticos infinitamente divisíveis (ver Maruyama, 1970).

**Definição 4.28.** Dizemos que  $(Y_n)_{n \in \mathbb{N}}$  é um *processo estocástico infinitamente divisível* se, para todo  $k \in \mathbb{N}$  e para toda  $k$ -upla de subíndices  $(n_1, \dots, n_k) \in \mathbb{N}^k$ , o vetor  $(Y_{n_1}, \dots, Y_{n_k})$  for um vetor infinitamente divisível em  $\mathbb{R}^k$ , conforme a Definição 4.27.

Sato (1999) apresenta na equação (8.1) uma representação para a função característica de qualquer distribuição infinitamente divisível sobre  $\mathbb{R}^k$ . Esta representação é comumente chamada na literatura de *representação de Lévy-Khintchine*, em homenagem aos dois matemáticos que introduziram e demonstraram esta fórmula. Apresentamos a seguir, tal representação na forma de um teorema (ver teorema 8.1 de Sato, 1999).

**Teorema 4.29.** *Seja  $\mathbf{D}_k = \{\mathbf{x} \in \mathbb{R}^k : \|\mathbf{x}\| \leq 1\}$  o disco unitário fechado em  $\mathbb{R}^k$ , onde  $\|\cdot\|$  representa a norma Euclidiana neste espaço. Se  $\mathbf{Z}$  é um vetor aleatório infinitamente*

divisível em  $\mathbb{R}^k$ , então sua função característica é dada por

$$\varphi_{\mathbf{Z}}(\mathbf{z}) = \exp \left[ -\frac{1}{2} \langle \mathbf{z}, A\mathbf{z} \rangle + i \langle \zeta, \mathbf{z} \rangle + \int_{\mathbb{R}^d} \left( e^{i \langle \mathbf{z}, \mathbf{x} \rangle} - 1 - i \langle \mathbf{z}, \mathbf{x} \rangle \mathbf{1}_{D_k}(\mathbf{x}) \right) \nu(d\mathbf{x}) \right], \quad (4.37)$$

para todo  $\mathbf{z} \in \mathbb{R}^k$ , onde  $A$  é uma matriz simétrica não-negativa definida  $k \times k$ ,  $\zeta \in \mathbb{R}^k$  e  $\nu(\cdot)$  é uma medida em  $\mathcal{B}(\mathbb{R}^k)$  satisfazendo

$$\nu(\{\mathbf{0}\}) = 0 \quad e \quad \int_{\mathbb{R}^k} (\|\mathbf{x}\|^2 \wedge 1) \nu(d\mathbf{x}) < \infty, \quad (4.38)$$

onde  $a \wedge b \equiv \min\{a, b\}$ .

**Observação 4.30:** A representação de  $\varphi_{\mathbf{Z}(\cdot)}$  por  $A, \nu$  e  $\zeta$  em (4.37) é única e  $(A, \nu, \zeta)$  é chamada de *tripla geradora* da distribuição do vetor aleatório  $\mathbf{Z}$ . Além disso, Sato (1999) mostra que, se existem  $A$  uma matriz simétrica não-negativa definida  $d \times d$ ,  $\nu(\cdot)$  uma medida em  $\mathcal{B}(\mathbb{R}^k)$ , satisfazendo as duas condições em (4.38), e  $\zeta \in \mathbb{R}^k$ , então existe uma distribuição infinitamente divisível cuja função característica associada é dada pela expressão à direita da igualdade em (4.37).

Apresentamos a seguir um exemplo simples de distribuições infinitamente divisíveis.

**Exemplo 4.31. (Distribuições  $\alpha$ -estáveis são infinitamente divisíveis).** Seja  $Z$  uma variável aleatória tal que  $Z \sim S_\alpha(\sigma, \beta, \mu)$ , então  $Z$  é infinitamente divisível.

De fato, vamos assumir que  $\alpha \neq 1$  (para o caso  $\alpha = 1$  é análogo). A função característica da variável aleatória  $Z$  é dada por

$$\varphi_Z(t) = \exp \left\{ -\sigma^\alpha |t|^\alpha \left( 1 - i\beta(\text{sign}(t)) \tan \left( \frac{\pi\alpha}{2} \right) \right) + i\mu t \right\}. \quad (4.39)$$

Considere  $Z_n \sim S_\alpha(\sigma/n^{1/\alpha}, \beta, \mu/n)$ . Então sabemos que

$$\begin{aligned} [\varphi_{Z_n}(t)]^n &= \left\{ \exp \left[ -\left( \frac{\sigma}{n^{1/\alpha}} \right)^\alpha |t|^\alpha \left( 1 - i\beta(\text{sign}(t)) \tan \left( \frac{\pi\alpha}{2} \right) \right) + i \left( \frac{\mu}{n} \right) t \right] \right\}^n \\ &= \exp \left[ -n \left( \frac{\sigma^\alpha}{n} \right) |t|^\alpha \left( 1 - i\beta(\text{sign}(t)) \tan \left( \frac{\pi\alpha}{2} \right) \right) + i \left( \frac{n\mu}{n} \right) t \right] = \varphi_Z(t), \end{aligned}$$

e portanto, como  $n \in \mathbb{N}$  foi escolhido arbitrariamente, mostramos que toda variável  $\alpha$ -estável é infinitamente divisível.  $\square$

O resultado do Exemplo 4.31 pode ser estendido para vetores  $\alpha$ -estáveis.

**Definição 4.32.** Dizemos que um vetor aleatório  $\mathbf{X}$  em  $\mathbb{R}^k$ , para  $k \in \mathbb{N}$ , é  $\alpha$ -estável se, para todo  $n \geq 2$ , existir  $\alpha \in (0, 2]$  e um vetor  $\mathbf{d}_n \in \mathbb{R}^k$  tais que

$$\mathbf{X}^{(1)} + \dots + \mathbf{X}^{(n)} \stackrel{d}{=} n^{1/\alpha} \mathbf{X} + \mathbf{d}_n, \quad (4.40)$$

para  $\mathbf{X}^{(1)}, \dots, \mathbf{X}^{(n)}$  cópias independentes de  $\mathbf{X}$ . Além disso,  $\mathbf{X}$  é um *vetor estritamente estável* se  $\mathbf{d}_n$  é o vetor nulo em (4.40).

**Exemplo 4.33. (Extensão do Exemplo 4.31).** Seja  $\mathbf{Z}$  um vetor  $\alpha$ -estável em  $\mathbb{R}^k$ . Kuelbs (1973) mostra uma correspondência entre as distribuições  $\alpha$ -estáveis num espaço de Hilbert  $H$  e as distribuições infinitamente divisíveis. Particularizando para o caso em que  $H = \mathbb{R}^k$ , podemos usar o lema 2.2 de Kuelbs (1973) para mostrar que se  $\mathbf{Z}$  tem distribuição  $\alpha$ -estável, então existem um vetor  $\zeta \in \mathbb{R}^k$  e uma medida finita  $\tilde{\nu}(\cdot)$  em  $\mathcal{B}(S_k) = \mathcal{B}(\{\mathbf{x} \in \mathbb{R}^k : \|\mathbf{x}\| = 1\})$  tal que, a função característica do vetor aleatório  $\mathbf{Z}$  pode ser expressa através de uma representação de Lévy-Khintchine, dada por

$$\varphi_{\mathbf{Z}}(\mathbf{z}) = \exp \left\{ i\langle \mathbf{z}, \zeta \rangle + \int_{S_k} \int_0^\infty \left[ e^{i\langle \mathbf{z}, \mathbf{s} \rangle r} - 1 - \frac{i\langle \mathbf{z}, \mathbf{s} \rangle r}{1 + r^2} \right] \frac{dr}{r^{1+\alpha}} \tilde{\nu}(d\mathbf{s}) \right\} \quad (4.41)$$

para todo  $\mathbf{z} \in \mathbb{R}^k$  e  $0 < \alpha < 2$ . Em suma, os vetores aleatórios  $\alpha$ -estáveis tem distribuições infinitamente divisíveis.  $\square$

A expressão (4.41) ainda pode ser usada para deduzir a expressão clássica das funções características de vetores  $\alpha$ -estáveis, para  $0 < \alpha < 2$  (ver teorema 2.1 de Kuelbs, 1973 e teorema 2.3.1 de Samorodnitsky e Taqqu, 2000). O fato de que vetores  $\alpha$ -estáveis tem distribuições infinitamente divisíveis é extremamente relevante para o nosso estudo. Enunciaremos a seguir, uma condição necessária e suficiente que permite dizer quando um processo estacionário infinitamente divisível é mixing ou não. Assim como Gross (1994, p. 280) referencia, apesar de Maruyama (1970) não enunciar de maneira explícita a Proposição 4.34 abaixo, ele utiliza este resultado para demonstrar o teorema 6 de seu artigo.

**Proposição 4.34.** *Um processo estocástico à tempo contínuo, estacionário e infinitamente divisível  $(X_t)_{t \in \mathbb{R}}$  é mixing se, e somente se, para quaisquer  $z_1$  e  $z_2$  reais, temos*

$$\mathbb{E}(\exp[iz_1 X_1 + iz_2 X_t]) \xrightarrow{t \rightarrow \infty} \mathbb{E}(\exp[iz_1 X_1]) \mathbb{E}(\exp[iz_2 X_1]).$$

*Em particular, se considerarmos um processo estacionário à tempo discreto e infinitamente divisível  $(X_n)_{n \in \mathbb{N}}$ , é natural que o processo  $(X_n)_{n \in \mathbb{N}}$  seja mixing se, e somente se, para qualquer  $z_1$  e  $z_2$  reais, tivermos*

$$\mathbb{E}(\exp[iz_1 X_1 + iz_2 X_n]) \xrightarrow{n \rightarrow \infty} \mathbb{E}(\exp[iz_1 X_1]) \mathbb{E}(\exp[iz_2 X_1]). \quad (4.42)$$

**Observação 4.35:** Note que a condição (4.42) acima, se resume à mostrar que, quando  $n \rightarrow \infty$ , a função característica do vetor aleatório  $(X_1, X_n)$ , calculada em  $(z_1, z_2)$ , converge para o produto das funções características de  $X_1$ , calculadas nos pontos  $z_1$  e  $z_2$ , isto é,

$$\lim_{n \rightarrow \infty} \varphi_{X_1, X_n}(z_1, z_2) = \varphi_{X_1}(z_1) \varphi_{X_1}(z_2). \quad (4.43)$$

O resultado da Proposição 4.34 pode ser usado para mostrar que o processo  $(X_n)_{n \in \mathbb{N}}$ , definido pela equação (4.20), é mixing.

**Teorema 4.36.** *O processo  $(X_n)_{n \in \mathbb{N}}$  apresentado na equação 4.20 é mixing e ergódico.*

*Demonstração.* Para demonstrar este teorema, devemos primeiro notar que as distribuições finito-dimensionais do processo  $(X_n)_{n \in \mathbb{N}}$  são  $\alpha$ -estáveis, isto é, para todo subconjunto de índices  $(n_1, \dots, n_d) \in \mathbb{N}^d$  e  $d \in \mathbb{N}$ ,  $(X_{n_1}, \dots, X_{n_d})$  é um vetor  $\alpha$ -estável.

De fato, para todo  $X_{n_j}$ , tal que  $1 \leq j \leq d$ , temos que

$$X_{n_j} = \theta^{n_j-1} X_1 + \sum_{k=2}^{n_j} \theta^{n_j-k} \varepsilon_k. \quad (4.44)$$

Por convenção, suponha que  $n_1 < n_2 < \dots < n_d$ . Considere a matriz  $A$  real  $n_d \times n_d$  tal que, na linha  $j$  os primeiros  $n_j$  elementos são os coeficientes da combinação linear em (4.44) e os restantes  $n_d - n_j$  elementos desta linha são preenchidos com zeros, isto é,  $A$  é uma matriz triangular inferior, da forma

$$A = \begin{pmatrix} \theta^{n_1-1} & \theta^{n_1-2} & \dots & 1 & 0 & \dots & 0 & 0 & \dots & 0 \\ \theta^{n_2-1} & \theta^{n_2-2} & \dots & \theta^{n_2-n_1} & \theta^{n_2-n_1-1} & \dots & 1 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ \theta^{n_d-1} & \theta^{n_d-2} & \dots & \theta^{n_d-n_1} & \theta^{n_d-n_1-1} & \dots & \theta^{n_d-n_2} & \theta^{n_d-n_2-1} & \dots & 1 \end{pmatrix}.$$

Defina o vetor aleatório  $\mathbf{Y}_{n_d} = (X_1, \varepsilon_2, \dots, \varepsilon_{n_d})$  em  $\mathbb{R}^{n_d}$  e note que

$$(X_{n_1}, X_{n_2}, \dots, X_{n_d}) = A\mathbf{Y}_{n_d}.$$

Uma vez que as variáveis aleatórias  $X_1, \varepsilon_2, \dots, \varepsilon_{n_d}$  são por hipótese, independentes e  $\alpha$ -estáveis, segue do exemplo 2.3.6 de Samorodnitsky e Taquq (2000), que o vetor aleatório  $(X_{n_1}, X_{n_2}, \dots, X_{n_d})$  é  $\alpha$ -estável. Podemos então concluir do Exemplo 4.33 que o vetor aleatório  $(X_{n_1}, \dots, X_{n_d})$  tem distribuição infinitamente divisível.

Como  $d$  e  $(n_1, \dots, n_d)$  foram escolhidos de modo arbitrário, segue que as distribuições finito-dimensionais do processo  $(X_n)_{n \in \mathbb{N}}$  são infinitamente divisíveis e pela Definição 4.28, isto significa que o processo  $(X_n)_{n \in \mathbb{N}}$  é infinitamente divisível.

Vamos então usar a condição apresentada em (4.43) para mostrar que  $(X_n)_{n \in \mathbb{N}}$  é um processo mixing. Sejam  $z_1$  e  $z_2$  reais quaisquer. Se  $z_1$  ou  $z_2$  for nulo, (4.43) é satisfeita trivialmente. Suponhamos que  $z_1$  e  $z_2$  sejam ambos não nulos. Temos que

$$z_1 X_1 + z_2 X_n = z_1 X_1 + z_2 \left( \theta^{n-1} X_1 + \sum_{k=2}^n \theta^{n-k} \varepsilon_k \right) = (z_1 + z_2 \theta^{n-1}) X_1 + z_2 \sum_{k=2}^n \theta^{n-k} \varepsilon_k.$$

Com isso, a função característica do vetor  $(X_1, X_n)$  é dada por

$$\begin{aligned}
\varphi_{X_1, X_n}(z_1, z_2) &= \mathbb{E}(\exp[i(z_1 X_1 + z_2 X_n)]) \\
&= \mathbb{E}\left(\exp\left\{i[(z_1 + z_2 \theta^{n-1})X_1] + i\left[z_2 \sum_{k=2}^n \theta^{n-k} \varepsilon_k\right]\right\}\right) \\
&= \mathbb{E}\left(\exp[i(z_1 + z_2 \theta^{n-1})X_1]\right) \times \prod_{k=2}^n \mathbb{E}\left(\exp\left[i(z_2 \theta^{n-k}) \varepsilon_k\right]\right) \\
&= \exp\left[-\left(\frac{\sigma^\alpha}{1 - |\theta|^\alpha}\right) |z_1 + z_2 \theta^{n-1}|^\alpha\right] \times \prod_{k=2}^n \exp\left(-\sigma^\alpha |z_2 \theta^{n-k}|^\alpha\right) \\
&= \exp\left[-\left(\frac{\sigma^\alpha}{1 - |\theta|^\alpha}\right) |z_1 + z_2 \theta^{n-1}|^\alpha\right] \exp\left[-\sigma^\alpha |z_2|^\alpha \sum_{k=0}^{n-2} (|\theta|^\alpha)^k\right] \\
&= \exp\left[-\left(\frac{\sigma^\alpha}{1 - |\theta|^\alpha}\right) |z_1 + z_2 \theta^{n-1}|^\alpha\right] \exp\left[-\sigma^\alpha |z_2|^\alpha \left(\frac{1 - |\theta|^{\alpha(n-1)}}{1 - |\theta|^\alpha}\right)\right].
\end{aligned} \tag{4.45}$$

Passando ao limite a expressão (4.45) acima e lembrando que  $|\theta| < 1$ , obtemos que

$$\begin{aligned}
&\lim_{n \rightarrow \infty} \varphi_{X_1, X_n}(z_1, z_2) \\
&= \lim_{n \rightarrow \infty} \left\{ \exp\left[-\left(\frac{\sigma^\alpha}{1 - |\theta|^\alpha}\right) |z_1 + z_2 \theta^{n-1}|^\alpha\right] \exp\left[-\sigma^\alpha |z_2|^\alpha \left(\frac{1 - |\theta|^{\alpha(n-1)}}{1 - |\theta|^\alpha}\right)\right] \right\} \\
&= \exp\left\{-\left(\frac{\sigma^\alpha}{1 - |\theta|^\alpha}\right) |z_1 + z_2 \lim_{n \rightarrow \infty} [\theta^{n-1}]|^\alpha\right\} \exp\left\{-\sigma^\alpha |z_2|^\alpha \left(\frac{1 - \lim_{n \rightarrow \infty} [|\theta|^{\alpha(n-1)}]}{1 - |\theta|^\alpha}\right)\right\} \\
&= \exp\left[-\left(\frac{\sigma^\alpha}{1 - |\theta|^\alpha}\right) |z_1|^\alpha\right] \exp\left[-\left(\frac{\sigma^\alpha}{1 - |\theta|^\alpha}\right) |z_2|^\alpha\right] = \varphi_{X_1}(z_1) \varphi_{X_1}(z_2).
\end{aligned}$$

Desta forma, como a condição (4.43) é satisfeita, podemos concluir que o processo  $(X_n)_{n \in \mathbb{N}}$  é mixing. O fato de que  $(X_n)_{n \in \mathbb{N}}$  é um processo ergódico é consequência imediata da Proposição 4.26.  $\square$

### 4.3 Função Densidade Espectral para o Processo AR(1) com Inovações $\alpha$ -Estáveis Simétricas

Nesta seção damos uma resposta para as perguntas realizadas no final da Subseção 4.2.2. Observamos inicialmente que o Teorema de Herglotz para o caso de processos  $(Y_n)_{n \in \mathbb{Z}}$  com variância finita, estabelece uma dualidade entre as funções de distribuição espectral e de autocovariância, denotada por  $\gamma_Y(\cdot)$ . Enunciamos abaixo o Teorema de Herglotz (junto com o corolário 4.3.1 de Brockwell e Davis, 1991), sem a sua demonstração, que pode ser consultada em Brockwell e Davis (1991, p. 118-119).

**Teorema 4.37 (Teorema de Herglotz).** *Uma função complexa  $\gamma(\cdot)$ , definida nos inteiros, é não-negativa definida se, e somente se,*

$$\gamma(h) = \int_{\mathbb{T}} e^{ihv} dF(v), \quad \text{para todo } h \in \mathbb{Z}, \tag{4.46}$$

onde  $F(\cdot)$  é uma função contínua à direita, não-decrescente, limitada em  $[-\pi, \pi]$  e  $F(-\pi) = 0$ . A função  $F(\cdot)$  é chamada de **função de distribuição espectral** da função  $\gamma(\cdot)$  e se

$$F(\omega) = \int_{-\pi}^{\omega} f(v)dv, \quad \text{para } -\pi \leq \omega \leq \pi,$$

então  $f(\cdot)$  é chamada a **função densidade espectral** de  $\gamma(\cdot)$ . Além disso, a função  $\gamma(\cdot)$  é a função de autocovariância de algum processo estacionário  $(Y_n)_{n \in \mathbb{Z}}$  se, e somente se,

- (i) (4.46) é satisfeita para alguma função  $F(\cdot)$  contínua à direita, não-decrescente, limitada em  $[-\pi, \pi]$  e tal que  $F(-\pi) = 0$ , ou
- (ii)  $\sum_{l,j=1}^n a_l \gamma(l-j) \bar{a}_j \geq 0$ , para todo  $n \in \mathbb{N}$  e todo  $\mathbf{a} = (a_1, \dots, a_n) \in \mathbb{C}^n$ .

Portanto, se (i) ou (ii) acima forem satisfeitas ((i) e (ii) são equivalentes), a função de distribuição espectral  $F(\cdot)$  (e a correspondente função densidade espectral, se existir alguma) serão referenciadas como a **função de distribuição espectral** (e **função densidade espectral**), tanto do processo  $(Y_n)_{n \in \mathbb{Z}}$  como também da função de autocovariância  $\gamma(\cdot)$ .

A dificuldade que encontramos em definir a função densidade espectral, no caso em que determinado processo  $(Y_n)_{n \in \mathbb{Z}}$  possui variância infinita, se dá pelo fato de que estamos impedidos de definir a função de autocovariância, e conseqüentemente, perdemos o Teorema de Herglotz. Uma alternativa à função de autocovariância é o uso de diferentes medidas de dependência, cujas definições não dependam da existência ou não dos momentos de primeira e segunda ordem.

Uma das alternativas viáveis para substituir a função de autocovariância no caso  $\alpha$ -estável simétrico, baseia-se no uso da função codiferença, que foi definida na Subseção 4.2.1. Embora a versão análoga do Teorema de Herglotz ainda não tenha sido explicitamente deduzida, podemos citar alguns trabalhos que estão próximos desta linha de pensamento, mas que não tenham preenchido por completo o raciocínio que leva à extensão do Teorema de Herglotz, para variáveis  $\alpha$ -estáveis simétricas, envolvendo a função codiferença:

- para uma representação espectral de um conjunto finito de variáveis  $\alpha$ -estáveis simétricas  $\{X_1, \dots, X_n\}$ , citamos Shilder (1970);
- para a versão envolvendo um conjunto de variáveis  $\alpha$ -estáveis simétricas  $\{X_t : t \in T\}$ , com  $T$  genérico, citamos Kuelbs (1973).

Ambos Shilder (1970) e Kuelbs (1973) definem a norma  $\|X\|$  de uma variável  $\alpha$ -estável simétrica  $X$ , cuja função característica é dada por

$$\varphi_X(t) = \exp[-|b||t|^\alpha], \quad b \in \mathbb{R} \text{ e } \alpha \in (0, 2],$$



como sendo

$$\|X\| = \begin{cases} |b|^{1/\alpha}, & \text{se } 1 \leq \alpha \leq 2, \\ |b|, & \text{se } 0 < \alpha < 1. \end{cases}$$

### 4.3.1 Relação Entre a Codiferença e Função Densidade Espectral

Tendo em mente o Teorema de Herglotz, vamos estabelecer uma representação para a função codiferença do processo  $(X_n)_{n \in \mathbb{N}}$ , definido pela equação (4.20), através de uma função  $g_\alpha(\cdot)$  real, definida no toro  $\mathbb{T} = [-\pi, \pi)$ . Da expressão (4.29), considere a função

$$\tau_X(k; \alpha) = \frac{\sigma^\alpha}{1 - |\theta|^\alpha} \left( 1 + |\theta|^{\alpha|k|} - (1 - \theta^{|k|})^\alpha \right), \quad (4.47)$$

para todo  $k \in \mathbb{Z}$  e  $0 < \alpha < 2$ . Pelo teorema 4.3.2 de Brockwell e Davis (1991), se

$$\sum_{k=-\infty}^{\infty} |\tau_X(k; \alpha)| < \infty,$$

para cada  $0 < \alpha < 2$ , então existe uma função  $g_\alpha(\cdot)$  definida por

$$g_\alpha(\omega) = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} e^{-ik\omega} \tau_X(k; \alpha), \quad (4.48)$$

para todo  $\omega \in \mathbb{T} = [-\pi, \pi)$ , tal que

$$\tau_X(k; \alpha) = \int_{\mathbb{T}} e^{ik\omega} g_\alpha(\omega) d\omega, \quad k \in \mathbb{Z}.$$

**Observação 4.38:** Uma vez que a função codiferença é simétrica, basta provar que

$$\sum_{k=1}^{\infty} |\tau_X(k; \alpha)| < \infty,$$

pois dado que  $\tau_X(-k; \alpha) = \tau_X(k; \alpha)$ , para todo  $k \in \mathbb{Z}$ , segue que

$$\sum_{k=-\infty}^{\infty} |\tau_X(k; \alpha)| = \tau_X(0; \alpha) + 2 \sum_{k=1}^{\infty} |\tau_X(k; \alpha)|.$$

No Apêndice D, apresentamos alguns gráficos da função codiferença, dada pela expressão (4.47), no caso particular em que  $\sigma = 1$ . Neste mesmo apêndice, ilustramos certas propriedades das quais a função  $\tau_X(k; \alpha)$  goza. Vamos obter a seguir, uma função explícita para a função  $g_\alpha(\cdot)$ , definida pela equação (4.48), na situação em que  $\alpha = 1$ .

**Exemplo 4.39. (Voltando ao Exemplo 4.7).** Considere o caso em que  $\alpha = 1$ . Temos

$$\sum_{k=0}^{\infty} |\tau_X(k; 1)| = \sum_{k=0}^{\infty} \left| \frac{\sigma}{1 - |\theta|} \right| \left| 1 + |\theta|^k - (1 - \theta^k) \right| = \frac{\sigma}{1 - |\theta|} \sum_{k=0}^{\infty} (|\theta|^k + \theta^k).$$

Se  $\theta = 0$ , a série  $\sum_{k=0}^{\infty} |\tau_X(k; 1)|$  é trivialmente convergente.

Se  $\theta > 0$ , então  $|\theta| = \theta$  e

$$\sum_{k=0}^{\infty} |\tau_X(k; 1)| = \frac{\sigma}{1-\theta} \sum_{k=0}^{\infty} 2\theta^k = \frac{\sigma}{1-\theta} \frac{2}{1-\theta} = \frac{2\sigma}{(1-\theta)^2} < \infty.$$

Se  $\theta < 0$ , então  $|\theta| = -\theta$  e portanto,

$$|\theta|^k + \theta^k = (-\theta)^k + \theta^k = \begin{cases} 2\theta^{2k'}, & \text{se } k = 2k', \text{ para todo } k' \in \mathbb{N} \cup \{0\}, \\ 0, & \text{se } k = 2k' - 1, \text{ para todo } k' \in \mathbb{N}. \end{cases}$$

Assim, para  $\theta < 0$ , obtemos

$$\sum_{k=0}^{\infty} |\tau_X(k; 1)| = \frac{\sigma}{1+\theta} \sum_{k=0}^{\infty} 2\theta^{2k} = \frac{\sigma}{1+\theta} \frac{2}{1-\theta^2} = \frac{2\sigma}{(1+\theta)^2(1-\theta)} < \infty.$$

Da convergência da série  $\sum_{k=0}^{\infty} |\tau_X(k; 1)|$  segue que, para  $\alpha = 1$ , a função  $g_1(\cdot)$  satisfazendo (4.48), pode ser calculada através de

$$\begin{aligned} g_1(\omega) &= \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} e^{-ik\omega} \tau_X(k; 1) = \frac{\sigma}{2\pi(1-|\theta|)} \sum_{k=-\infty}^{\infty} e^{-ik\omega} (|\theta|^{|k|} + \theta^{|k|}) \\ &= \frac{\sigma}{2\pi(1-|\theta|)} \left[ \sum_{k=-\infty}^{-1} e^{-ik\omega} (|\theta|^{|k|} + \theta^{|k|}) + e^0 (|\theta|^0 + \theta^0) + \sum_{k=1}^{\infty} e^{-ik\omega} (|\theta|^{|k|} + \theta^{|k|}) \right] \\ &= \frac{\sigma}{2\pi(1-|\theta|)} \left[ 2 + \sum_{k=1}^{\infty} e^{ik\omega} (|\theta|^k + \theta^k) + \sum_{k=1}^{\infty} e^{-ik\omega} (|\theta|^k + \theta^k) \right], \quad \omega \in \mathbb{T}. \end{aligned}$$

Separando em casos, se  $\theta = 0$ , então  $g_1(\omega) = \sigma/\pi$ , para todo  $\omega \in \mathbb{T}$ . Da equação 1.447 (3) de Gradshteyn e Ryzhik (1965) sabemos que

$$1 + 2 \sum_{k=1}^{\infty} p^k \cos(kx) = \frac{1-p^2}{1-2p \cos(x) + p^2}, \quad \text{se } |p| < 1, \quad (4.49)$$

logo, usando este resultado com  $p = \theta$  e  $x = \omega$ , segue que para  $\theta > 0$ ,

$$\begin{aligned} g_1(\omega) &= \frac{\sigma}{2\pi(1-\theta)} \left[ 2 + \sum_{k=1}^{\infty} e^{ik\omega} (\theta^k + \theta^k) + \sum_{k=1}^{\infty} e^{-ik\omega} (\theta^k + \theta^k) \right] \\ &= \frac{\sigma}{\pi(1-\theta)} \left[ 1 + \sum_{k=1}^{\infty} \theta^k e^{ik\omega} + \sum_{k=1}^{\infty} \theta^k e^{-ik\omega} \right] = \frac{\sigma}{\pi(1-\theta)} \left[ 1 + \sum_{k=1}^{\infty} \theta^k (e^{ik\omega} + e^{-ik\omega}) \right] \\ &= \frac{\sigma}{\pi(1-\theta)} \left[ 1 + 2 \sum_{k=1}^{\infty} \theta^k \cos(k\omega) \right] = \frac{\sigma}{\pi(1-\theta)} \left[ \frac{1-\theta^2}{1-2\theta \cos(\omega) + \theta^2} \right] \\ &= \frac{\sigma}{\pi(1-\theta)} \frac{(1-\theta)(1+\theta)}{[1-2\theta \cos(\omega) + \theta^2]} = \frac{\sigma(1+\theta)}{\pi[1-2\theta \cos(\omega) + \theta^2]}, \quad \omega \in \mathbb{T}. \end{aligned}$$

No caso  $\theta < 0$ , aplicando o resultado da equação (4.49) com  $p = \theta^2$  e  $x = 2\omega$ , obtemos que

$$\begin{aligned} g_1(\omega) &= \frac{\sigma}{2\pi[1 - (-\theta)]} \left\{ 2 + \sum_{k=1}^{\infty} e^{ik\omega} [(-\theta)^k + \theta^k] + \sum_{k=1}^{\infty} e^{-ik\omega} [(-\theta)^k + \theta^k] \right\} \\ &= \frac{\sigma}{\pi(1 + \theta)} \left( 1 + \sum_{k=1}^{\infty} \theta^{2k} e^{i2k\omega} + \sum_{k=1}^{\infty} \theta^{2k} e^{-i2k\omega} \right) \\ &= \frac{\sigma}{\pi(1 + \theta)} \left[ 1 + \sum_{k=1}^{\infty} \theta^{2k} (e^{i2k\omega} + e^{-i2k\omega}) \right] = \frac{\sigma}{\pi(1 + \theta)} \left\{ 1 + 2 \sum_{k=1}^{\infty} (\theta^2)^k \cos[k(2\omega)] \right\} \\ &= \frac{\sigma}{\pi(1 + \theta)} \left[ \frac{1 - \theta^4}{1 - 2\theta^2 \cos(2\omega) + \theta^4} \right], \quad \omega \in \mathbb{T}. \end{aligned}$$

Resumindo, para  $\alpha = 1$ , a função  $g_1(\cdot)$  definida em  $\mathbb{T} = [-\pi, \pi)$  e satisfazendo  $\tau_X(k; 1) = \int_{\mathbb{T}} e^{ik\omega} g_1(\omega) d\omega$ , para qualquer inteiro  $k$ , é dada por

$$g_1(\omega) = \begin{cases} \frac{\sigma(1 + \theta)}{\pi(1 - 2\theta \cos(\omega) + \theta^2)}, & \text{se } \theta \geq 0, \\ \frac{\sigma(1 - \theta^4)}{\pi(1 + \theta)(1 - 2\theta^2 \cos(2\omega) + \theta^4)}, & \text{se } \theta < 0. \end{cases} \quad (4.50)$$

□

### 4.3.2 Caso $\alpha$ Geral

Vamos estender o Exemplo 4.39 considerando o caso geral de distribuições  $\alpha$ -estáveis, para  $\alpha \in (0, 2)$ . Precisamos mostrar primeiro que

$$\sum_{k=1}^{\infty} |\tau_X(k; \alpha)| = \frac{\sigma^\alpha}{1 - |\theta|^\alpha} \sum_{k=1}^{\infty} |1 + |\theta|^{\alpha k} - (1 - \theta^k)^\alpha| < \infty. \quad (4.51)$$

E para mostrar este fato, usamos o seguinte critério (ver p. 42 de Lima, 2013)

**Proposição 4.40 (Critério de d'Alembert).** *Seja  $(a_n)_{n \in \mathbb{N}}$  uma sequência real tal que  $a_n \neq 0$ , para todo  $n \in \mathbb{N}$ . Se existir uma constante  $c$  tal que*

$$\lim_{n \rightarrow \infty} \left| \frac{a_{n+1}}{a_n} \right| = c,$$

com  $c < 1$ , então a série  $\sum_{n=1}^{\infty} a_n$  converge absolutamente.

**Lema 4.41.** *Se  $0 < \theta < 1$  e  $\alpha \in (0, 2)$ , então a série*

$$\sum_{k=0}^{\infty} |\tau_X(k; \alpha)| = \frac{\sigma^\alpha}{1 - |\theta|^\alpha} \sum_{k=0}^{\infty} |1 + |\theta|^{\alpha k} - (1 - \theta^k)^\alpha|$$

converge.

*Demonstração.* Se  $\theta > 0$ , a função  $\tau_X(k; \alpha)$  é decrescente em  $k$ . De fato, sendo  $\theta > 0$  a codiferença assume a forma  $\tau_X(k; \alpha) = \frac{\sigma^\alpha}{1 - |\theta|^\alpha} (1 + \theta^{\alpha k} - (1 - \theta^k)^\alpha)$ . Além disso,

$$0 < \theta < 1 \Rightarrow \theta^{\alpha(k+1)} < \theta^{\alpha k} \Rightarrow 1 + \theta^{\alpha(k+1)} < 1 + \theta^{\alpha k} \quad (4.52)$$

e

$$0 < \theta < 1 \Rightarrow -\theta^k < -\theta^{k+1} \Rightarrow 1 - \theta^k < 1 - \theta^{k+1} \Rightarrow -(1 - \theta^{k+1})^\alpha < -(1 - \theta^k)^\alpha, \quad (4.53)$$

o que mostra o resultado, pois das inequações (4.52) e (4.53) segue que

$$\begin{aligned} \tau_X(k+1; \alpha) &= \frac{\sigma^\alpha}{1 - |\theta|^\alpha} \left( 1 + \theta^{\alpha(k+1)} - (1 - \theta^{k+1})^\alpha \right) \\ &< \frac{\sigma^\alpha}{1 - |\theta|^\alpha} \left( 1 + \theta^{\alpha k} - (1 - \theta^k)^\alpha \right) = \tau_X(k; \alpha), \quad k \in \mathbb{N}. \end{aligned}$$

**Observação 4.42:** Sendo  $1 + \theta^{\alpha k} - (1 - \theta^k)^\alpha$  uma função decrescente e convergente à zero, necessariamente devemos ter que  $1 + \theta^{\alpha k} - (1 - \theta^k)^\alpha > 0$ , para qualquer  $k \in \mathbb{N}$ , desde que  $\theta > 0$ .

Vamos mostrar que as séries

$$\sum_{k=0}^{\infty} \theta^{\alpha k} \quad \text{e} \quad \sum_{k=0}^{\infty} 1 - (1 - \theta^k)^\alpha \quad (4.54)$$

convergem absolutamente.

É óbvio que a primeira delas é absolutamente convergente, dado que é uma série geométrica com razão  $\theta^\alpha \in (0, 1)$ . Para ver que a segunda também é, vamos usar o critério de d'Alembert. Mas antes, observe que sendo  $0 < \theta < 1$ , temos

$$0 < (1 - \theta^k)^\alpha < 1 \Rightarrow 0 < 1 - (1 - \theta^k)^\alpha < 1,$$

para todo  $k \in \mathbb{Z}$ . Assim, pelo critério de d'Alembert

$$\lim_{k \rightarrow \infty} \left| \frac{1 - (1 - \theta^{(k+1)})^\alpha}{1 - (1 - \theta^k)^\alpha} \right| = \lim_{k \rightarrow \infty} \frac{1 - (1 - \theta^{(k+1)})^\alpha}{1 - (1 - \theta^k)^\alpha} = \frac{0}{0}. \quad (\text{Indeterminado!}) \quad (4.55)$$

Como o limite é indeterminado, vamos usar a regra de l'Hospital para calcular o limite da fração

$$\lim_{x \rightarrow \infty} \frac{\phi_1(x)}{\phi_2(x)} = \lim_{x \rightarrow \infty} \frac{1 - (1 - \theta^{(x+1)})^\alpha}{1 - (1 - \theta^x)^\alpha},$$

onde  $\phi_1(x) = 1 - (1 - \theta^{(x+1)})^\alpha = \phi_2(x - 1)$ . Vamos mostrar que este último limite é igual à  $\theta$ . Logo, poderemos concluir que o limite em (4.55) também existe e que é igual à  $\theta$ .

Ora, temos que

$$\frac{\phi_1'(x)}{\phi_2'(x)} = \frac{-\alpha(1 - \theta^{(x+1)})^{\alpha-1} [-\theta^{(x+1)} \ln(\theta)]}{-\alpha(1 - \theta^x)^{\alpha-1} [-\theta^x \ln(\theta)]} = \theta \frac{(1 - \theta^{(x+1)})^{\alpha-1}}{(1 - \theta^x)^{\alpha-1}}.$$

Tomando o limite para  $x \rightarrow \infty$ , segue que

$$\lim_{x \rightarrow \infty} \frac{\phi_1'(x)}{\phi_2'(x)} = \theta \left[ \lim_{x \rightarrow \infty} \frac{(1 - \theta^{(x+1)})^{\alpha-1}}{(1 - \theta^x)^{\alpha-1}} \right] = \theta. \quad (4.56)$$

Assim, dado que o limite  $\lim_{x \rightarrow \infty} \phi'_1(x)/\phi'_2(x)$  existe e é finito, a Regra de l'Hôpital garante que

$$\lim_{x \rightarrow \infty} \frac{1 - (1 - \theta^{(x+1)})^\alpha}{1 - (1 - \theta^x)^\alpha} = \lim_{x \rightarrow \infty} \frac{\phi_1(x)}{\phi_2(x)} = \lim_{x \rightarrow \infty} \frac{\phi'_1(x)}{\phi'_2(x)} = \theta.$$

Mas observe que o limite em (4.55) é apenas a versão inteira deste último limite, logo chegamos à conclusão de que também

$$\lim_{k \rightarrow \infty} \left| \frac{1 - (1 - \theta^{(k+1)})^\alpha}{1 - (1 - \theta^k)^\alpha} \right| = \lim_{k \rightarrow \infty} \frac{1 - (1 - \theta^{(k+1)})^\alpha}{1 - (1 - \theta^k)^\alpha} = \theta. \quad (4.57)$$

Sendo  $\theta < 1$ , o Critério de d'Alembert se aplica e assegura que a série  $\sum_{k=0}^{\infty} [1 - (1 - \theta^k)^\alpha]$  é absolutamente convergente para  $0 < \theta < 1$  e  $\alpha \in (0, 2)$ .

Finalmente, como as séries em (4.54) são absolutamente convergentes e temos sempre

$$|1 + |\theta|^{\alpha k} - (1 - \theta^k)^\alpha| \leq |1 - (1 - \theta^k)^\alpha| + |\theta|^{\alpha k},$$

o Critério de Comparação para séries se aplica, o que prova a afirmação do lema.  $\square$

Vamos abordar no lema seguinte, o caso em que  $-1 < \theta < 0$ .

**Lema 4.43.** *Se  $-1 < \theta < 0$  e  $\alpha \in (0, 2)$ , então a série*

$$\sum_{k=1}^{\infty} |\tau_X(k; \alpha)| = \frac{\sigma^\alpha}{1 - |\theta|^\alpha} \sum_{k=1}^{\infty} |1 + |\theta|^{\alpha k} - (1 - \theta^k)^\alpha| \quad (4.58)$$

*converge.*

*Demonstração.* Vamos usar o mesmo raciocínio empregado na prova do Lema 4.41, mas como  $-1 < \theta < 0$ , precisamos proceder com cautela e averiguar os casos em que  $k$  é par ou ímpar na sequência  $(\tau_X(k; \alpha))_{k \in \mathbb{N}}$ . Queremos mais uma vez aplicar o Critério de d'Alembert e para isso, vamos definir uma sequência  $(z_k)_{k \in \mathbb{N}}$  cujos termos são dados por

$$z_k = \left| \frac{1 - (1 - \theta^{k+1})^\alpha}{1 - (1 - \theta^k)^\alpha} \right|, \quad k \in \mathbb{N}. \quad (4.59)$$

Considere as subsequências  $(x_k)_{k \in \mathbb{N}}$  e  $(y_k)_{k \in \mathbb{N}}$  de  $(z_k)_{k \in \mathbb{N}}$ , dadas respectivamente por

$$x_k = \frac{|1 - (1 - \theta^{2k})^\alpha|}{|1 - (1 - \theta^{2k-1})^\alpha|} \quad \text{e} \quad y_k = \frac{|1 - (1 - \theta^{2k+1})^\alpha|}{|1 - (1 - \theta^{2k})^\alpha|}, \quad k \in \mathbb{N}. \quad (4.60)$$

Vamos provar que  $\lim_{k \rightarrow \infty} x_k = \lim_{k \rightarrow \infty} y_k = \theta$ . Com efeito, consideremos primeiro a sequência  $(x_k)_{k \in \mathbb{N}}$ . Sabemos que

$$-1 < \theta < 0 \Rightarrow -1 < -\theta^{2k} < 0 \Rightarrow 0 < 1 - \theta^{2k} < 1 \Rightarrow 0 < 1 - (1 - \theta^{2k})^\alpha < 1$$

e que

$$\begin{aligned} -1 < \theta < 0 &\Rightarrow 0 < -\theta^{2k-1} = (-\theta)^{2k-1} < 1 \Rightarrow 1 < 1 + (-\theta)^{2k-1} < 2 \\ &\Rightarrow 1 < [1 + (-\theta)^{2k-1}]^\alpha < 2^\alpha < 4 \Rightarrow -4 < -2^\alpha < -[1 + (-\theta)^{2k-1}]^\alpha < -1 \\ &\Rightarrow -3 < 1 - 2^\alpha < 1 - [1 + (-\theta)^{2k-1}]^\alpha < 0, \end{aligned}$$

para todo  $k \in \mathbb{N}$ . Desta forma, observamos que

$$\lim_{k \rightarrow \infty} x_k = \lim_{k \rightarrow \infty} \frac{|1 - [1 - \theta^{2k}]^\alpha|}{|1 - [1 - \theta^{2k-1}]^\alpha|} = \lim_{k \rightarrow \infty} \frac{1 - [1 - \theta^{2k}]^\alpha}{[1 + (-\theta)^{2k-1}]^\alpha - 1} = \frac{0}{0}. \quad (\text{Indeterminado!})$$

Defina

$$\phi_1(x) = 1 - [1 - \theta^{2x}]^\alpha, \quad \text{para } x \in (0, \infty)$$

e

$$\phi_2(x) = [1 + (-\theta)^{2x-1}]^\alpha - 1, \quad \text{para } x \in (1/2, \infty).$$

Ambas as funções  $\phi_1(\cdot)$  e  $\phi_2(\cdot)$  são deriváveis em seu domínio e são respectivamente, as extensões reais das funções definidas nos naturais

$$\phi_1(k) = 1 - [1 - \theta^{2k}]^\alpha \quad \text{e} \quad \phi_2(k) = [1 + (-\theta)^{2k-1}]^\alpha - 1.$$

Recorrendo à regra de l'Hôpital, vemos por argumentos análogos aos da prova do Lema 4.41, que

$$\begin{aligned} \lim_{k \rightarrow \infty} x_k &= \lim_{k \rightarrow \infty} \frac{1 - [1 - \theta^{2k}]^\alpha}{[1 + (-\theta)^{2k-1}]^\alpha - 1} = \lim_{x \rightarrow \infty} \frac{\phi_1(x)}{\phi_2(x)} = \lim_{x \rightarrow \infty} \frac{\phi_1'(x)}{\phi_2'(x)} \\ &= \lim_{x \rightarrow \infty} \frac{-\alpha[1 - \theta^{2x}]^{\alpha-1}[-\theta^{2x} \ln(\theta^2)]}{\alpha[1 + (-\theta)^{2x-1}]^{\alpha-1}[(-\theta)^{2x-1} 2 \ln(-\theta)]} = \theta \left\{ \lim_{x \rightarrow \infty} \frac{[1 - \theta^{2x}]^{\alpha-1}}{[1 + (-\theta)^{2x-1}]^{\alpha-1}} \right\} \\ &= \theta. \end{aligned}$$

Consideremos agora a sequência  $(y_k)_{k \in \mathbb{N}}$ . Note que também

$$\lim_{k \rightarrow \infty} y_k = \lim_{k \rightarrow \infty} \frac{|1 - [1 - \theta^{2k+1}]^\alpha|}{|1 - [1 - \theta^{2k}]^\alpha|} = \lim_{k \rightarrow \infty} \frac{[1 + (-\theta)^{2k+1}]^\alpha - 1}{1 - [1 - \theta^{2k}]^\alpha} = \frac{0}{0}. \quad (\text{Indeterminado!})$$

Assim, recorrendo mais uma vez à regra de l'Hôpital e à argumentos similares aos usados na prova do Lema 4.41, observamos que

$$\begin{aligned} \lim_{k \rightarrow \infty} y_k &= \lim_{k \rightarrow \infty} \frac{[1 + (-\theta)^{2k+1}]^\alpha - 1}{1 - [1 - \theta^{2k}]^\alpha} = \lim_{k \rightarrow \infty} \frac{\phi_2(x+1)}{\phi_1(x)} = \lim_{k \rightarrow \infty} \frac{\phi_2'(x+1)}{\phi_1'(x)} \\ &= \lim_{x \rightarrow \infty} \frac{\alpha[1 + (-\theta)^{2x+1}]^{\alpha-1}[(-\theta)^{2x+1} 2 \ln(-\theta)]}{-\alpha[1 - \theta^{2x}]^{\alpha-1}[-\theta^{2x} \ln(\theta^2)]} = \theta \left\{ \lim_{x \rightarrow \infty} \frac{[1 + (-\theta)^{2x+1}]^{\alpha-1}}{[1 - \theta^{2x}]^{\alpha-1}} \right\} \\ &= \theta. \end{aligned}$$

Observe agora que as sequências  $(x_k)_{k \in \mathbb{N}}$ ,  $(y_k)_{k \in \mathbb{N}}$  e  $(z_k)_{k \in \mathbb{N}}$  são associadas pelas relações

$$z_{2k-1} = x_k \quad \text{e} \quad z_{2k} = y_k,$$

para qualquer  $k \in \mathbb{N}$ . Vamos mostrar que também  $\lim_{k \rightarrow \infty} z_k = \theta$ .

De fato, mostramos acima que  $\lim_{k \rightarrow \infty} x_k = \lim_{k \rightarrow \infty} y_k = \theta$ . Pela definição formal de limite de uma sequência, dado  $\delta > 0$ , existem  $k_1, k_2 \in \mathbb{N}$  tais que

$$n > k_1 \Rightarrow |x_n - \theta| < \delta \quad \text{e} \quad n > k_2 \Rightarrow |y_n - \theta| < \delta. \quad (4.61)$$

Tome  $k_0 = \max\{2k_1 - 1, 2k_2\}$ . Se  $k = 2n - 1$ , para  $n \in \mathbb{N}$ , então  $z_k = z_{2n-1} = x_n$  e

$$k > k_0 \Rightarrow 2n - 1 = k > k_0 > 2k_1 - 1 \Rightarrow n > k_1 \stackrel{(4.61)}{\Rightarrow} |z_k - \theta| = |x_n - \theta| < \delta.$$

Por outro lado, se  $k = 2n$ , para  $n \in \mathbb{N}$ , então  $z_k = z_{2n} = y_n$  e

$$k > k_0 \Rightarrow 2n = k > k_0 > 2k_2 \Rightarrow n > k_2 \stackrel{(4.61)}{\Rightarrow} |z_k - \theta| = |y_n - \theta| < \delta.$$

Em outras palavras, acabamos de mostrar que

$$\forall \delta > 0, \exists k_0 \in \mathbb{N}; k > k_0 \Rightarrow |z_k - \theta| < \delta,$$

isto é,  $\lim_{k \rightarrow \infty} z_k = \theta$ .

Em conclusão, provamos que

$$\lim_{k \rightarrow \infty} \left| \frac{1 - (1 - \theta^{k+1})^\alpha}{1 - (1 - \theta^k)^\alpha} \right| = \lim_{k \rightarrow \infty} z_k = \theta, \quad \text{para } -1 < \theta < 0,$$

e portanto, o Critério de d'Alembert permite concluir que a série de termos  $1 - (1 - \theta^k)^\alpha$ , para  $k \in \mathbb{N}$ ,  $\alpha \in (0, 2)$  e  $-1 < \theta < 0$  é absolutamente convergente, isto é,

$$\sum_{k=1}^{\infty} |1 - (1 - \theta^k)^\alpha| < \infty.$$

Como a série  $\sum_{k=1}^{\infty} |\theta|^{\alpha k}$  também é convergente e

$$|1 + |\theta|^{\alpha k} - (1 - \theta^k)^\alpha| \leq |1 - (1 - \theta^k)^\alpha| + |\theta|^{\alpha k}, \quad k \in \mathbb{N},$$

podemos inferir que a série em (4.58) converge.  $\square$

### 4.3.3 Cálculo da Função $g_\alpha(\cdot)$

Tendo conhecimento da convergência da série

$$\sum_{k=-\infty}^{\infty} |\tau_X(k; \alpha)| = \frac{\sigma^\alpha}{1 - |\theta|^\alpha} \sum_{k=-\infty}^{\infty} |1 + |\theta|^{\alpha|k|} - (1 - \theta^{|k|})^\alpha|, \quad (4.62)$$

vamos determinar a função  $g_\alpha(\cdot)$  definida na equação (4.48), para  $\alpha \in (0, 2)$ .

Observamos primeiramente que, sendo  $\tau_X(k; \alpha)$  uma função simétrica em  $k$ , da equação (4.48) obtemos

$$\begin{aligned}
g_\alpha(\omega) &= \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} e^{-ik\omega} \tau_X(k; \alpha) \\
&= \frac{1}{2\pi} \left[ \tau_X(0, \alpha) + \sum_{k=-\infty}^{-1} e^{-ik\omega} \tau_X(k; \alpha) + \sum_{k=1}^{\infty} e^{-ik\omega} \tau_X(k; \alpha) \right] \\
&= \frac{1}{2\pi} \left[ \frac{2\sigma^\alpha}{1 - |\theta|^\alpha} + \sum_{k=1}^{\infty} e^{ik\omega} \tau_X(k; \alpha) + \sum_{k=1}^{\infty} e^{-ik\omega} \tau_X(k; \alpha) \right] \\
&= \frac{1}{2\pi} \left[ \frac{2\sigma^\alpha}{1 - |\theta|^\alpha} + \sum_{k=1}^{\infty} \tau_X(k; \alpha) (e^{ik\omega} + e^{-ik\omega}) \right] = \frac{1}{2\pi} \left[ \frac{2\sigma^\alpha}{1 - |\theta|^\alpha} + 2 \sum_{k=1}^{\infty} \tau_X(k; \alpha) \cos(k\omega) \right] \\
&= \frac{\sigma^\alpha}{2\pi(1 - |\theta|^\alpha)} \left\{ 2 + 2 \sum_{k=1}^{\infty} [1 + |\theta|^{\alpha k} - (1 - \theta^k)^\alpha] \cos(k\omega) \right\}. \tag{4.63}
\end{aligned}$$

Observe que, sendo a função cosseno limitada por 1, temos que

$$\begin{aligned}
|[1 + |\theta|^{\alpha k} - (1 - \theta^k)^\alpha] \cos(k\omega)| &\leq |\theta|^{\alpha k} |\cos(k\omega)| + |1 - (1 - \theta^k)^\alpha| |\cos(k\omega)| \\
&\leq |\theta|^{\alpha k} + |1 - (1 - \theta^k)^\alpha|, \quad \text{para todo } k \in \mathbb{N}.
\end{aligned}$$

Assim, pela convergência absoluta, podemos separar as séries em (4.63), dando origem às duas somas da expressão (4.64) abaixo

$$g_\alpha(\omega) = \frac{\sigma^\alpha}{2\pi(1 - |\theta|^\alpha)} \left\{ \left[ 1 + 2 \sum_{k=1}^{\infty} |\theta|^{\alpha k} \cos(k\omega) \right] + \left[ 1 + 2 \sum_{k=1}^{\infty} (1 - (1 - \theta^k)^\alpha) \cos(k\omega) \right] \right\}. \tag{4.64}$$

Vamos avaliar cada uma destas séries separadamente. Para a primeira série em 4.64, note que sendo  $|\theta|^\alpha < 1$ , a equação (4.49), para  $p = |\theta|^\alpha$  e  $x = \omega$ , nos diz que

$$1 + 2 \sum_{k=1}^{\infty} |\theta|^{\alpha k} \cos(k\omega) = 1 + 2 \sum_{k=1}^{\infty} (|\theta|^\alpha)^k \cos(k\omega) = \frac{1 - |\theta|^{2\alpha}}{1 - 2|\theta|^\alpha \cos(\omega) + |\theta|^{2\alpha}}. \tag{4.65}$$

Para a segunda série em 4.64, vamos reescrever  $h(\theta) = (1 - \theta)^\alpha$ , para  $|\theta| < 1$ , na sua série de Taylor em torno do ponto 0. Observe que  $h(\cdot)$  é de classe  $C^\infty$  no intervalo aberto  $(-1, 1)$  e suas  $n$  primeiras derivadas são dadas por

$$\begin{aligned}
h'(\theta) &= -\alpha(1 - \theta)^{\alpha-1} \\
h''(\theta) &= \alpha(\alpha - 1)(1 - \theta)^{\alpha-2} \\
h'''(\theta) &= -\alpha(\alpha - 1)(\alpha - 2)(1 - \theta)^{\alpha-3} \\
&\dots \\
h^{(n)} &= (-1)^n \alpha(\alpha - 1) \dots (\alpha - n + 1)(1 - \theta)^{\alpha-n}.
\end{aligned}$$



O polinômio de Taylor  $h_n(\cdot)$  de ordem  $n$  da função  $h(\cdot)$  em torno do ponto  $\theta = 0$  é dado por

$$\begin{aligned} h_n(\theta) &= h(0) + \frac{h'(0)}{1!}\theta + \frac{h''(0)}{2!}\theta^2 + \frac{h'''(0)}{3!}\theta^3 + \dots + \frac{h^{(n)}(0)}{n!}\theta^n \\ &= 1 - \frac{\alpha}{1!}\theta + \frac{\alpha(\alpha-1)}{2!}\theta^2 + \dots + (-1)^n \frac{\alpha(\alpha-1)\dots(\alpha-n+1)}{n!}\theta^n \\ &= \sum_{j=0}^n (-1)^j \frac{\Gamma(\alpha+1)}{\Gamma(\alpha+1-j)} \frac{\theta^j}{\Gamma(j+1)}. \end{aligned}$$

Assim, a série de Taylor da função  $h(\cdot)$  no ponto 0 é dada por

$$h(\theta) = (1-\theta)^\alpha = \sum_{j=0}^{\infty} (-1)^j \frac{\Gamma(\alpha+1)}{\Gamma(\alpha+1-j)} \frac{\theta^j}{\Gamma(j+1)}, \quad |\theta| < 1. \quad (4.66)$$

Usando a expressão (4.66) para reescrever a função  $(1-\theta^k)^\alpha$ , para qualquer  $k$  natural, temos

$$(1-\theta^k)^\alpha = \sum_{j=0}^{\infty} (-1)^j \frac{\Gamma(\alpha+1)}{\Gamma(\alpha+1-j)} \frac{\theta^{jk}}{\Gamma(j+1)}, \quad |\theta| < 1. \quad (4.67)$$

**Observação 4.44:** A série dada na equação (4.67) pertence à classe das séries hipergeométricas e conforme a equação 2.1.6 de Andrews et al. (1999), temos

$$(1-\theta^k)^\alpha = {}_1F_0 \left( \begin{matrix} -\alpha \\ - \end{matrix} ; \theta^k \right).$$

O primeiro termo da série em (4.67) é igual à 1, para qualquer  $k \in \mathbb{N}$ . Consequentemente,

$$\begin{aligned} 1 - (1-\theta^k)^\alpha &= - \sum_{j=1}^{\infty} (-1)^j \frac{\Gamma(\alpha+1)}{\Gamma(\alpha+1-j)} \frac{\theta^{jk}}{\Gamma(j+1)} \\ &= \sum_{j=1}^{\infty} (-1)^{j+1} \frac{\Gamma(\alpha+1)}{\Gamma(\alpha+1-j)} \frac{\theta^{jk}}{\Gamma(j+1)}, \end{aligned}$$

para  $|\theta| < 1$ . E assim, a segunda série que aparece na equação (4.64), pode ser reescrita como

$$\begin{aligned} 1 + 2 \sum_{k=1}^{\infty} \left( 1 - (1-\theta^k)^\alpha \right) \cos(k\omega) &= \\ &= 1 + 2 \sum_{k=1}^{\infty} \left( \sum_{j=1}^{\infty} (-1)^{j+1} \frac{\Gamma(\alpha+1)}{\Gamma(\alpha+1-j)} \frac{\theta^{jk}}{\Gamma(j+1)} \right) \cos(k\omega). \end{aligned} \quad (4.68)$$

Usando as equações (4.65) e (4.68) para reescrever a função  $g_\alpha(\cdot)$ , obtemos da equação (4.64) que

$$\begin{aligned} g_\alpha(\omega) &= \frac{\sigma^\alpha}{2\pi(1-|\theta|^\alpha)} \left( \frac{1-|\theta|^{2\alpha}}{1-2|\theta|^\alpha \cos(\omega) + |\theta|^{2\alpha}} \right) + \\ &+ \frac{\sigma^\alpha}{2\pi(1-|\theta|^\alpha)} \left( 1 + 2 \sum_{k=1}^{\infty} \left( \sum_{j=1}^{\infty} (-1)^{j+1} \frac{\Gamma(\alpha+1)}{\Gamma(\alpha+1-j)} \frac{\theta^{jk}}{\Gamma(j+1)} \right) \cos(k\omega) \right) \\ &= \frac{\sigma^\alpha(1+|\theta|^\alpha)}{2\pi(1-2|\theta|^\alpha \cos(\omega) + |\theta|^{2\alpha})} + \\ &+ \frac{\sigma^\alpha}{2\pi(1-|\theta|^\alpha)} \left( 1 + 2 \sum_{k=1}^{\infty} \left( \sum_{j=1}^{\infty} (-1)^{j+1} \frac{\Gamma(\alpha+1)}{\Gamma(\alpha+1-j)} \frac{\theta^{jk}}{\Gamma(j+1)} \right) \cos(k\omega) \right). \end{aligned} \quad (4.69)$$

Com isto provamos o seguinte teorema.

**Teorema 4.45.** *Considere o processo  $(X_n)_{n \in \mathbb{N}}$ , definido pela equação (4.20), para  $|\theta| < 1$  e  $(\varepsilon_n)_{n \geq 2}$  uma sequência de variáveis aleatórias i.i.d. com distribuição  $S_\alpha(\sigma, 0, 0)$ , com  $0 < \alpha < 2$  e  $\sigma > 0$ . Assuma também que a variável  $X_1$  é independente da sequência  $(\varepsilon_n)_{n \geq 2}$  e além disso,  $X_1$  tem distribuição  $S_\alpha\left(\frac{\sigma}{(1-|\theta|^\alpha)^{1/\alpha}}, 0, 0\right)$ . Se  $\Upsilon_{\mathbf{X}_n}$  representa a matriz de codiferença do vetor  $\mathbf{X}_n = (X_1, \dots, X_n)$ , cuja expressão é dada por*

$$\Upsilon_{\mathbf{X}_n} = \frac{\sigma^\alpha}{1-|\theta|^\alpha} \left( 1 + |\theta|^{\alpha|l-j|} - (1 - \theta^{|l-j|})^\alpha \right)_{1 \leq l, j \leq n},$$

então temos que

$$T_n(g_\alpha) = [T_n(g_\alpha)_{l,j}]_{1 \leq l, j \leq n} = \left[ \frac{1}{2\pi} \int_{\mathbb{T}} \exp[i(l-j)t] g_\alpha(t) dt \right]_{1 \leq l, j \leq n} = \Upsilon_{\mathbf{X}_n},$$

onde  $g_\alpha : \mathbb{T} \rightarrow \mathbb{R}$  é a função definida pela equação (4.69) e  $T_n(g_\alpha)$  é a matriz de Toeplitz associada à função  $g_\alpha(\cdot)$ .

## 4.4 Grandes Desvios e o Processo AR(1) com Inovações $\alpha$ -Estáveis Não-Gaussianas

O objetivo desta seção é mostrar que, o método utilizado por Bercu et al. (1997) na determinação de um princípio de grandes desvios dos estimadores de Yule-Walker e de mínimos quadrados, para o parâmetro  $\theta$  do processo autorregressivo Gaussiano, não pode ser adaptado passo a passo para o caso em que as inovações são  $\alpha$ -estáveis não-Gaussianas, vide Definição 4.9.

A principal dificuldade se encontra na determinação da função geradora de cumulantes normalizada. Considere  $\mathbf{X}_n = (X_1, \dots, X_n)$ , com  $X_1, \dots, X_n$  soluções da equação (4.20). Analogamente à Proposição 3.12 (b), suponhamos que exista um vetor aleatório

$\mathbf{Y}_n$ , cuja distribuição é  $S\alpha S$  com  $n$  componentes i.i.d., e tal que o vetor aleatório  $\mathbf{X}_n$  pode ser escrito como um produto da forma

$$\mathbf{X}_n = \Upsilon_{\mathbf{X}_n}^{1/2} \mathbf{Y}_n, \quad (4.70)$$

onde  $\Upsilon_{\mathbf{X}_n}^{1/2}$  representa a matriz raiz quadrada da matriz de codiferença do vetor aleatório  $\mathbf{X}_n$ . Note que sendo  $\Upsilon_{\mathbf{X}_n}$  uma matriz simétrica e não-negativa definida (ver Proposição 4.19), a sua matriz raiz quadrada  $\Upsilon_{\mathbf{X}_n}^{1/2}$  existe, é simétrica e não-negativa definida.

Considere novamente a função  $f(x) \equiv \cos(x) - c$ , para  $x \in \mathbb{T}$  e  $c \in (-1, 1)$ . Considere também a matriz de Toeplitz  $T_n(f)$  associada a função  $f(\cdot)$ , dada pela expressão (3.18). A matriz  $\Upsilon_{\mathbf{X}_n}^{1/2} T_n(f) \Upsilon_{\mathbf{X}_n}^{1/2}$  é real simétrica e existe uma matriz  $P_n$  ortogonal  $n \times n$  tal que

$$\Upsilon_{\mathbf{X}_n}^{1/2} T_n(f) \Upsilon_{\mathbf{X}_n}^{1/2} = P_n \Lambda_n P_n^T, \quad (4.71)$$

onde  $\Lambda_n = \text{diag}(\lambda_1^n, \dots, \lambda_n^n)$  e  $\lambda_1^n, \dots, \lambda_n^n$  são os autovalores de  $T_n(f) \Upsilon_{\mathbf{X}_n}$ .

O Teorema 4.45 mostra que existe uma função  $g_\alpha(\cdot)$ , definida pela equação (4.69), tal que  $\Upsilon_{\mathbf{X}_n} = T_n(g_\alpha)$ . Logo, a expressão (4.71) pode ser reescrita como

$$T_n^{1/2}(g_\alpha) T_n(f) T_n^{1/2}(g_\alpha) = P_n \Lambda_n P_n^T,$$

e desta forma,  $\lambda_1^n, \dots, \lambda_n^n$  são os autovalores da matriz  $T_n(f) T_n(g_\alpha)$ . Combinando as equações (4.70) e (4.71) para reescrever a estatística  $\mathcal{W}_n(f)$ , apresentada na Seção 3.2, temos que

$$\begin{aligned} \mathcal{W}_n(f) &= \frac{1}{n} \left( \sum_{k=2}^n X_k X_{k-1} - c \sum_{k=1}^n X_k^2 \right) = \frac{1}{n} \mathbf{X}_n^T T_n(f) \mathbf{X}_n = \frac{1}{n} (\Upsilon_{\mathbf{X}_n}^{1/2} \mathbf{Y}_n)^T T_n(f) (\Upsilon_{\mathbf{X}_n}^{1/2} \mathbf{Y}_n) \\ &= \frac{1}{n} \mathbf{Y}_n^T \Upsilon_{\mathbf{X}_n}^{1/2} T_n(f) \Upsilon_{\mathbf{X}_n}^{1/2} \mathbf{Y}_n = \frac{1}{n} \mathbf{Y}_n^T P_n \Lambda_n P_n^T \mathbf{Y}_n = \frac{1}{n} (P_n^T \mathbf{Y}_n)^T \Lambda_n (P_n^T \mathbf{Y}_n). \end{aligned} \quad (4.72)$$

Como  $\mathbf{Y}_n$  é vetor aleatório  $S\alpha S$  com componentes i.i.d, segue do exemplo 2.3.6 de Samorodnitsky e Taquq (2000) que  $\mathbf{Z}_n = P_n^T \mathbf{Y}_n = (Z_1, \dots, Z_n)$  é vetor  $\alpha$ -estável. Assim, da equação (4.72), obtemos que

$$\mathcal{W}_n(f) = \frac{1}{n} (P_n^T \mathbf{Y}_n)^T \Lambda_n (P_n^T \mathbf{Y}_n) = \frac{1}{n} \mathbf{Z}_n^T \Lambda_n \mathbf{Z}_n = \frac{1}{n} \sum_{k=1}^n \lambda_k^n Z_k^2. \quad (4.73)$$

Mas  $\mathcal{W}_n(f)$  abordada desta forma, não demonstra utilidades práticas, pois a função geradora de cumulantes normalizada da soma  $n^{-1} \sum_{k=1}^n \lambda_k^n Z_k^2$ , não está definida para qualquer vizinhança de zero! De fato, pela desigualdade de Jensen, a função geradora de momentos da variável  $Z_k^2$  satisfaz

$$M_{Z_k^2}(t) = \mathbb{E} \left( \exp\{t Z_k^2\} \right) \geq \exp \left\{ t \mathbb{E} \left( Z_k^2 \right) \right\} = +\infty, \quad \text{se } t > 0,$$

para todo  $1 \leq k \leq n$ , uma vez que  $Z_k^2$  é uma variável aleatória  $\alpha$ -estável. Mas então, independente do valor que  $M_{Z_k^2}(t)$  assume para  $t \leq 0$ , a função geradora de cumulantes normalizada de  $\mathcal{W}_n(f)$  não está definida em uma vizinhança de 0, pois

$$\begin{aligned} L_n(\lambda, \mathcal{W}_n(f)) &= \frac{1}{n} \log \mathbb{E}(\exp\{n\lambda\mathcal{W}_n(f)\}) = \frac{1}{n} \log \mathbb{E}\left(\exp\left\{\sum_{k=1}^n \lambda\lambda_k^n Z_k^2\right\}\right) \\ &= \frac{1}{n} \sum_{k=1}^n \log M_{Z_k^2}(\lambda\lambda_k^n) = +\infty, \quad \text{se } \lambda\lambda_k^n > 0, \text{ para algum } k \in \{1, \dots, n\}. \end{aligned} \tag{4.74}$$

Para conseguir aplicar o teorema de Gärtner-Ellis, devemos analisar com muito cuidado a expressão 4.74 e também o seu limite. Esta análise pode ser exaustiva e muito mais complicada do que no caso Gaussiano. Aliás, ainda não conhecemos a distribuição das variáveis  $Z_k^2$ , para  $k \in \{1, \dots, n\}$ , e muito menos, caso exista, a função geradora de momentos destas variáveis aleatórias. Citamos Rathie et al. (2016), para a distribuição do produto e quociente de duas variáveis aleatórias  $X$  e  $Y$ , onde ambas são  $\alpha$ -estáveis e independentes, situação da qual não podemos tirar muito proveito, mas que ainda pode interessar ao leitor.

#### 4.4.1 Princípio de Grandes Desvios Para o Estimador de Yule-Walker na Versão $\alpha$ -Estável

O Teorema de Gärtner-Ellis não é imediatamente aplicável, quando a função geradora de cumulantes normalizada não está definida em qualquer vizinhança em torno de zero, pois a Hipótese 2.20 é violada. O que perdemos em tal situação, é a existência de uma “boa” função taxa, embora ainda sejamos capazes de determinar uma função taxa, cujos conjuntos de nível não sejam necessariamente compactos. Bercu e Richou (2017) descobriram uma nova técnica para determinar princípios de grandes desvios, para o estimador de máxima verossimilhança do parâmetro de um processo Ornstein-Uhlenbeck. Podemos utilizar esta mesma técnica para resolver o problema de nosso interesse. Vamos descrever a seguir, um pouquinho da metodologia aplicada por Bercu e Richou (2017), já adaptando-a ao estimador de Yule-Walker do parâmetro  $\theta$ , exibido na equação (4.20).

Observamos na Seção 3.2 que as propriedades de grandes desvios da sequência  $(\tilde{\theta}_n)_{n \geq 2}$ , dos estimadores de Yule-Walker do parâmetro  $\theta$ , coincidem com as da sequência  $(\mathcal{W}_n(f))_{n \geq 2}$ , definida por

$$\mathcal{W}_n(f) = \frac{1}{n} \left( \sum_{k=2}^n X_k X_{k-1} - c \sum_{k=1}^n X_k^2 \right), \tag{4.75}$$

onde  $c \in (-1, 1)$ . Defina portanto a sequência  $(V_n)_{n \geq 2} \in (\mathbb{R} \times (0, +\infty))^{\mathbb{N} \setminus \{1\}}$ , cuja expressão

é dada por

$$V_n = \left( \sum_{k=2}^n \frac{X_k X_{k-1}}{n}, \sum_{k=1}^n \frac{X_k^2}{n} \right), \quad \text{para } n \geq 2. \quad (4.76)$$

Ao provarmos um LDP para a sequência  $(V_n)_{n \geq 2}$ , podemos utilizar o princípio da contração, discutido na Subseção 2.2.4, e deduzir um LDP para a sequência de variáveis aleatórias  $(F(V_n))_{n \geq 2}$ , induzida por alguma função contínua  $F : \mathbb{R} \times (0, +\infty) \rightarrow \mathbb{R}$ .

No caso particular em que a função  $F(\cdot, \cdot)$  for definida pela lei

$$F(x, y) = x - cy, \quad \text{para } c \in (-1, 1), \quad (4.77)$$

observamos que

$$F(V_n) = F \left( \sum_{k=2}^n \frac{X_k X_{k-1}}{n}, \sum_{k=1}^n \frac{X_k^2}{n} \right) = \frac{1}{n} \left( \sum_{k=2}^n X_k X_{k-1} - c \sum_{k=1}^n X_k^2 \right) = \mathcal{W}_n(f). \quad (4.78)$$

Assim, basta provarmos um princípio de grandes desvios para a sequência  $(V_n)_{n \geq 2}$ , à fim de que o LDP para o estimador de Yule-Walker possa ser obtido pelo princípio da contração.

Com o mesmo raciocínio, podemos obter um princípio de grandes desvios para a soma  $n^{-1} \sum_{k=1}^n X_k^2$ . Para isto, basta utilizarmos a função contínua  $F(x, y) = y$ , conhecida por ser a função projeção na segunda variável. Esta ideia estende para a classe das  $\alpha$ -estáveis, o resultado provado por Bryc e Smolenski (1993), conforme mencionado na Observação 3.25.

O ponto principal dentro de todo este esquema, é tentar aplicar novamente o Teorema de Gärtner-Ellis. Defina  $L_n(\cdot, \cdot)$ , como sendo a função geradora de cumulantes normalizada do vetor aleatório  $V_n$ , isto é,

$$L_n(t_1, t_2) = \frac{1}{n} \log \{ \mathbb{E} (n \langle (t_1, t_2), V_n \rangle) \}, \quad (4.79)$$

onde  $\langle \cdot, \cdot \rangle$  denota o produto interno usual em  $\mathbb{R}^2$ . Se a função  $L_n(\cdot, \cdot)$  estiver bem definida, tomando o limite quando  $n \rightarrow \infty$ , obtemos uma função  $L(\cdot, \cdot)$ , definida por

$$L(t_1, t_2) = \lim_{n \rightarrow \infty} L_n(t_1, t_2),$$

e que indica, a partir do Teorema de Gärtner-Ellis, quando ou não, a sequência  $(V_n)_{n \in \mathbb{N}}$  satisfaz um princípio de grandes desvios.

Para poder aplicar o Teorema de Gärtner-Ellis, precisamos mais uma vez estudar o domínio da função  $L(\cdot, \cdot)$ , definido pelo conjunto

$$\mathcal{D}_L = \{(t_1, t_2) \in \mathbb{R}^2 : L(t_1, t_2) < \infty\}.$$

Esta análise pode ser extremamente complicada, visto que estamos à trabalhar com uma classe ampla de distribuições.

Para mais detalhes técnicos, indicamos a consulta do artigo de Bercu e Richou (2017).



## 5 Conclusão

Neste trabalho estudamos uma maneira bastante intrigante de aplicar a Teoria de Grandes Desvios, na estimação do parâmetro de um processo autorregressivo de primeira ordem, definido pela equação (3.1), cujas inovações são Gaussianas. No Capítulo 3 estudamos o método apresentado por Bercu et al. (1997), a fim de encontrar princípios de grandes desvios para os estimadores de Yule-Walker e de mínimos quadrados do parâmetro deste processo. Realizamos uma análise criteriosa dos autovalores da sequência de matrizes  $(T_n(f)T_n(g))_{n \geq 2}$ , cujas funções  $f(\cdot)$  e  $g(\cdot)$  foram definidas respectivamente, nas equações (3.11) e (3.2).

Ao tentarmos aplicar os mesmos métodos usados por Bercu et al. (1997), na busca de um LDP para o processo autorregressivo  $\alpha$ -estável não-Gaussiano, definido pela equação (4.20), nos deparamos com a dificuldade em encontrar uma expressão fechada para a função geradora de cumulantes normalizada. Esta dificuldade decorre do fato que a função geradora de momentos de uma distribuição  $\alpha$ -estável não-Gaussiana, não está completamente definida nas vizinhanças de zero. O próximo passo para dar continuidade a este trabalho, é tentar contornar esta dificuldade através de uma análise minuciosa do problema como um todo. Talvez a técnica utilizada por Bercu et al. (1997), exibida no Capítulo 3, não seja a mais adequada para resolvermos o problema no caso  $\alpha$ -estável geral.

Embora ainda não tenhamos conseguido estabelecer um princípio de grandes desvios para os estimadores de Yule-Walker e de mínimos quadrados, no caso  $\alpha$ -estável não-Gaussiano, obtivemos excelentes resultados paralelos ao que desejávamos. Conseguimos estender o processo  $(X_n)_{n \in \mathbb{N}}$ , apresentado na equação (3.1), para o processo mais geral apresentado na equação (4.20), onde as inovações deste último são variáveis aleatórias  $\alpha$ -estáveis simétricas. Além disso, fomos capazes de provar que, este processo mais geral  $(X_n)_{n \in \mathbb{N}}$  é estacionário, mixing e ergódico. Estas três últimas propriedades foram provadas com ajuda da expressão (3.23), que é consequência imediata da definição do processo  $(X_n)_{n \in \mathbb{N}}$ .

Dentre outros resultados que provamos, não podemos deixar de citar a expressão fechada para a função codiferença do processo  $(X_n)_{n \in \mathbb{N}}$   $\alpha$ -estável não-Gaussiano, apresentada na equação (4.29). Além disso, mostramos que esta mesma função codiferença, converge à zero quando o seu lag converge à infinito, propriedade que compartilha com a função de autocovariância do processo  $(X_n)_{n \in \mathbb{N}}$ , definido pela equação (3.1).

Na Seção 4.3, provamos que a matriz de codiferença do vetor aleatório  $\mathbf{X}_n = (X_1, \dots, X_n)$ , cujas componentes são as variáveis aleatórias que são solução da equação (4.20), pode ser reescrita como uma matriz de Toeplitz associada a função  $g_\alpha(\cdot)$ , onde

$g_\alpha(\cdot)$  é exibida na equação (4.69). Mesmo que a matriz de codiferença do vetor  $\mathbf{X}_n$  possua propriedades semelhantes às da matriz de autocovariância do processo Gaussiano  $(X_n)_{n \in \mathbb{N}}$ , definido na equação (3.1), notamos na Seção 4.4 que tal matriz de codiferença pode não servir da maneira que esperávamos. Ainda não temos certeza se podemos usá-la de fato, para encontrar as propriedades de grandes desvios dos estimadores de Yule-Walker e de mínimos quadrados do parâmetro  $\theta$ , apresentado na equação (4.20).

Apresentamos no final da Seção 4.4, uma alternativa ao problema de grandes desvios para o processo  $\alpha$ -estável não-Gaussiano. Esta ideia, como já mencionado na mesma seção, surgiu a partir da leitura do artigo de Bercu e Richou (2017). A metodologia utilizada por estes autores consiste, resumidamente, em passar o problema para o espaço das sequências em  $\mathbb{R}^2$ , aplicar o Teorema de Gärtner-Ellis (quando possível) e usar o Princípio da Contração para retornar ao problema original. Temos esperanças de que esta técnica seja mais eficaz na resolução de nosso problema e pretendemos explorá-la em estudos futuros.



# Referências

- Andrews, B. e Davis, R. A. (2013). “Model Identification for Infinite Variance Autoregressive Processes”. *Journal of Econometrics*, Vol. **172** (2), 222–234. Citado na página 13.
- Andrews, G. E.; Askey, R. e Roy, R. (1999). *Special Functions*. New York: Cambridge University Press. Citado na página 103.
- Anton, H.; Bivens, I. e Davis, S. (2007). *Cálculo*. Porto Alegre: Bookman, Vol. **1**. Citado na página 76.
- Astrauskas, A. (1983). “Limit Theorems for Sums of Linearly Generated Random Variables”. *Lithuanian Mathematical Journal*, Vol. **23** (2), 127–134. Citado na página 14.
- Avram, F. (1988). “On Bilinear Forms in Gaussian Random Variables and Toeplitz Matrices”. *Probability Theory and Related Fields*, Vol. **79** (1), 37–45. Citado 4 vezes nas páginas 13, 40, 43 e 122.
- Bahadur, R. R. e Rao, R. R. (1960). “On Deviations of the Sample Mean”. *The Annals of Mathematical Statistics*, Vol. **31** (4), 1015–1027. Citado 2 vezes nas páginas 13 e 29.
- Bartle, R. G. (1995). *The Elements of Integration and Lebesgue Measure*. New York: John Wiley & Sons. Citado 2 vezes nas páginas 35 e 42.
- Bercu, B. (2001). “On Large Deviations in the Gaussian Autoregressive Process: Stable, Unstable and Explosive Case”. *Bernoulli*, Vol. **7** (2), 299–316. Citado 2 vezes nas páginas 31 e 32.
- Bercu, B.; Delyon, B. e Rio, E. (2015). *Concentration Inequalities for Sums and Martingales*. New York: Springer. Citado 4 vezes nas páginas 15, 19, 20 e 29.
- Bercu, B.; Gamboa, F. e Lavielle, M. (2000). “Sharp Large Deviations for Gaussian Quadratic Forms with Applications”. *ESAIM: Probability and Statistics*, Vol. **4** (1), 1–24. Citado 5 vezes nas páginas 29, 31, 32, 62 e 67.
- Bercu, B.; Gamboa, F. e Rouault, A. (1997). “Large Deviations for Quadratic Forms of Stationary Gaussian Processes”. *Stochastic Processes and their Applications*, Vol. **71** (1), 75–90. Citado 12 vezes nas páginas 14, 17, 31, 32, 33, 40, 47, 53, 55, 69, 104 e 109.
- Bercu, B. e Richou, A. (2017). “Large Deviations for the Ornstein-Uhlenbeck Process Without Tears”. *Statistics & Probability Letters*, Vol. **123**, 45–55. Citado 3 vezes nas páginas 106, 107 e 110.
- Billingsley, P. (1995). *Probability and Measure*. New York: John Wiley & Sons. Citado 2 vezes nas páginas 43 e 44.
- Bingham, N. H.; Goldie, C. M. e Teugels, J. L. (1987). *Regular Variation*. Cambridge: Cambridge University Press. Citado na página 17.

- Bogachev, V. (2007). *Measure Theory*. Berlin: Springer-Verlag, Vol. 1. Citado 2 vezes nas páginas 41 e 43.
- Brockwell, P. J. e Davis, R. A. (1991). *Time Series: Theory and Methods*. New York: Springer. Citado 11 vezes nas páginas 13, 31, 32, 37, 38, 39, 87, 93, 95, 117 e 124.
- Bryc, W. (1993). “A Remark on the Connection Between the Large Deviation Principle and the Central Limit Theorem”. *Statistics & Probability Letters*, Vol. 18 (4), 253–256. Citado 2 vezes nas páginas 18 e 19.
- Bryc, W. e Dembo, A. (1997). “Large Deviations for Quadratic Functionals of Gaussian Processes”. *Journal of Theoretical Probability*, Vol. 10 (2), 307–332. Citado na página 14.
- Bryc, W. e Smolenski, W. (1993). “On the Large Deviation Principle for a Quadratic Functional of the Autoregressive Process”. *Statistics & Probability Letters*, Vol. 17 (4), 281–285. Citado 2 vezes nas páginas 52 e 107.
- Bucklew, J. A. (1990). *Large Deviation Techniques in Decision, Simulation, and Estimation*. New York: John Wiley & Sons. Citado 6 vezes nas páginas 13, 15, 16, 17, 19 e 26.
- Dembo, A. e Zeitouni, O. (1993). *Large Deviations Techniques and Applications*. Boston: Jones and Bartlett Publishers. Citado 7 vezes nas páginas 13, 15, 16, 17, 18, 26 e 28.
- Durrett, R. (2013). *Probability: Theory and Examples*. New York: Cambridge University Press. Citado na página 45.
- Ellis, R. S. (1985). *Entropy, Large Deviations, and Statistical Mechanics*. New York: Springer-Verlag. Citado 5 vezes nas páginas 13, 15, 16, 48 e 51.
- Feller, W. (1970). *An Introduction to Probability Theory and Its Applications*. New York: John Wiley & Sons. Citado 3 vezes nas páginas 13, 69 e 70.
- Gallagher, C. M. (2001). “A Method for Fitting Stable Autoregressive Models Using the Autocovariation Function”. *Statistics & Probability Letters*, Vol. 53 (4), 381–390. Citado na página 13.
- Gnedenko, B. V.; Kolmogorov, A. N.; Chung, K. L. e Doob, J. L. (1954). *Limit Distributions for Sums of Independent Random Variables*. Cambridge: Addison-Wesley. Citado 3 vezes nas páginas 13, 69 e 89.
- Gradshteyn, I. S. e Ryzhik, I. M. (1965). *Table of Integrals, Series, and Products*. San Diego: Academic Press. Citado 4 vezes nas páginas 36, 59, 76 e 96.
- Gray, R. M. (2009). *Toeplitz and Circulant Matrices: A Review*. Now Publishers. Disponível em: <<http://ee.stanford.edu/~gray/toeplitz.pdf>>. Acesso em: 24 fev. 2017. Citado na página 87.
- Grenander, U. e Szegő, G. (1958). *Toeplitz Forms and Their Applications*. Cambridge: Cambridge University Press. Citado 3 vezes nas páginas 14, 55 e 87.
- Gross, A. (1994). “Some Mixing Conditions for Stationary Symetric Stable Stochastic Processes”. *Stochastic Processes and their Applications*, Vol. 51 (2), 277–295. Citado na página 91.

- Heyde, C. C. (1967). “A Contribution to the Theory of Large Deviations for Sums of Independent Random Variables”. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, Vol. **7** (5), 303–308. Citado na página 14.
- James, B. (2015). *Probabilidade: Um Curso em Nível Intermediário*. Rio de Janeiro: IMPA. Citado 4 vezes nas páginas 15, 33, 46 e 75.
- Jensen, J. L. (1995). *Saddlepoint Approximations*. New York: Oxford University Press. Citado 2 vezes nas páginas 56 e 127.
- Karlin, S. e Taylor, H. (1975). *A First Course in Stochastic Processes*. San Diego: Academic Press. Citado na página 82.
- Kuelbs, J. (1973). “A Representation Theorem for Symetric Stable Processes and Stable Measures on  $H$ ”. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, Vol. **26** (4), 259–271. Citado 2 vezes nas páginas 91 e 94.
- Lai, T. L. e Wei, C. Z. (1983). “Asymptotic Properties of General Autoregressive Models and Strong Consistency of Least-Squares Estimates of Their Parameters”. *Journal of Multivariate Analysis*, Vol. **13** (1), 1–23. Citado na página 33.
- Lima, E. L. (2013). *Análise Real Volume 1: Funções de Uma Variável*. Rio de Janeiro: IMPA. Citado 3 vezes nas páginas 44, 50 e 97.
- Lima, E. L. (2014). *Álgebra Linear*. Rio de Janeiro: IMPA. Citado 5 vezes nas páginas 56, 121, 123, 126 e 127.
- Ling, S. (2005). “Self-weighted Least Absolute Deviation Estimation for Infinite Variance Autoregressive Models”. *Journal of the Royal Statistical Society. Series B, Statistical Methodology*, Vol. **67** (3), 381–393. Citado na página 13.
- Lopes, A. O. e Lopes, S. R. C. (2016). *Introdução aos Processos Estocásticos para Estudantes de Matemática*. Porto Alegre: [s.n.]. Disponível em: <<http://mat.ufrgs.br/~alopes/pub3/PrincipalNovo.pdf>>. Acesso em: 26 mai. 2017. Citado 2 vezes nas páginas 88 e 89.
- Mann, H. B. e Wald, A. (1943). “On the Statistical Treatment of Linear Stochastic Difference Equations”. *Econometrica*, Vol. **11** (3), 173–200. Citado na página 33.
- Maruyama, G. (1970). “Infinitely Divisible Processes”. *Theory of Probability and its Applications*, Vol. **15** (1), 3–23. Citado 2 vezes nas páginas 89 e 91.
- Mikosch, T. (2013). “Precise Large Deviations for Dependent Regularly Varying Sequences”. *Probability Theory and Related Fields*, Vol. **156** (3), 851–887. Citado na página 14.
- Mörters, P. (2008). “Large Deviation Theory and Applications”. *Graduate course notes* Citado na página 17.
- Müller, G. H. (2016). *Uma Introdução aos Grandes Desvios*. Dissertação (Mestrado) — Universidade Federal do Rio Grande do Sul, Porto Alegre. Citado 3 vezes nas páginas 20, 26 e 73.

- Nowicka-Zagrajek, J. e Wyłomańska, A. (2008). “Measures of Dependence for Stable AR(1) Models with Time-Varying Coefficients”. *Stochastic Models*, Vol. **24** (1), 58–70. Citado na página 86.
- Rathie, P. N.; Ozelim, L. C. de S. M. e Otiniano, C. E. G. (2016). “Exact Distribution of the Product and the Quotient of Two Stable Lévy Random Variables”. *Communications in Nonlinear Science and Numerical Simulation*, Vol. **36**, 204–218. Citado na página 106.
- Resnick, S. I. (1997). “Heavy Tail Modeling and Teletraffic Data”. *The Annals of Statistics*, Vol. **25** (5), 1805–1849. Citado na página 13.
- Rockafellar, R. T. (1970). *Convex Analysis*. New Jersey: Princeton University Press. Citado 4 vezes nas páginas 19, 47, 50 e 51.
- Rohatgi, V. K. e Saleh, A. K. M. E. (2000). *An Introduction to Probability and Statistics*. New York: John Wiley & Sons. Citado 6 vezes nas páginas 23, 25, 39, 46, 71 e 131.
- Rozovskii, L. V. (1999). “Large deviations of sums of independent random variables from the domain of attraction of a stable law”. *Journal of Mathematical Sciences*, Vol. **93** (3), 421–435. Citado na página 14.
- Samorodnitsky, G. e Taqqu, M. S. (2000). *Stable Non-Gaussian Random Processes: Stochastic Models with Infinite Variance*. Boca Raton: Chapman & Hall. Citado 14 vezes nas páginas 13, 14, 69, 70, 71, 72, 76, 83, 84, 91, 92, 105, 118 e 136.
- Santana, A. P. e Queiró, J. F. (2010). *Introdução à Álgebra Linear*. Lisboa: Gradiva. Citado 3 vezes nas páginas 121, 122 e 123.
- Sato, K.-I. (1999). *Lévy Processes and Infinitely Divisible Distributions*. Cambridge: Cambridge University Press. Citado 3 vezes nas páginas 69, 89 e 90.
- Shilder, M. (1970). “Some Structure Theorems for the Symmetric Stable Laws”. *The Annals of Mathematical Statistics*, Vol. **41** (2), 412–421. Citado na página 94.
- Zaigraev, A. (1999). “Multivariate Large Deviations With Stable Limit Laws”. *Probability and Mathematical Statistics*, Vol. **19** (2), 323–335. Citado na página 14.

# Apêndices



# Apêndice A – Processos Autorregressivos

Neste apêndice vamos definir algumas das noções básicas referentes à processos autorregressivos, assim como os estimadores de Yule-Walker e de mínimos quadrados.

**Definição A.1.** Dizemos que o processo  $(X_n)_{n \in \mathbb{Z}}$  é (*estritamente*) *estacionário* se as distribuições conjuntas dos vetores aleatórios  $(X_{n_1}, X_{n_2}, \dots, X_{n_j})$  e  $(X_{n_1+h}, X_{n_2+h}, \dots, X_{n_j+h})$  são idênticas, para todo  $j \in \mathbb{N}$  e quaisquer inteiros  $n_1, n_2, \dots, n_j$  e  $h$ .

Na definição acima, a palavra *estritamente* se encontra entre parênteses porque alguns autores, como por exemplo Brockwell e Davis (1991), apresentam dois conceitos distintos de estacionariedade, porém relacionados. O conceito usual de estacionariedade, apresentado na definição 1.3.2 de Brockwell e Davis (1991), exige segundos momentos finitos, o que no caso das distribuições  $\alpha$ -estáveis, para  $0 < \alpha < 2$ , não tem sentido. Já a definição 1.3.3 de Brockwell e Davis (1991), que estabelece a estacionariedade estrita, depende apenas da distribuição conjunta das variáveis do processo. Embora a estacionariedade estrita seja mais forte do que a usual, sabemos que no caso de o processo  $(X_n)_{n \in \mathbb{Z}}$  ser Gaussiano, estas duas concepções são equivalentes (ver p. 13 de Brockwell e Davis, 1991).

**Definição A.2.** Um processo estocástico  $(X_n)_{n \in \mathbb{Z}}$  é dito ser *autorregressivo* de ordem  $p$ , denotado por AR( $p$ ), se, para todo  $n \in \mathbb{Z}$ , satisfizer a equação

$$X_n - \theta_1 X_{n-1} - \theta_2 X_{n-2} - \dots - \theta_p X_{n-p} = \varepsilon_n, \quad (\text{A.1})$$

onde as *inovações*  $(\varepsilon_n)_{n \in \mathbb{Z}}$  são variáveis aleatórias i.i.d. e  $(\theta_1, \dots, \theta_p) \in \mathbb{R}^p$ .

A definição usual de processos autorregressivos não é a que apresentamos na Definição A.2. Em Brockwell e Davis (1991, p. 78), as inovações  $\varepsilon_n$  são definidas como sendo processos *ruídos brancos* média 0 e variância  $\sigma^2$  ( $\text{WN}(0, \sigma^2)$ ), isto é, processos  $(\varepsilon_n)_{n \in \mathbb{N}}$  i.i.d., que como acusa o nome, possuem média 0 e função de autocovariância dada por

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

Novamente observamos que, tal como no momento em que definimos processos estacionários, para o nosso estudo, a definição usual de processo autorregressivo perde o seu significado, visto que inovações  $\alpha$ -estáveis, com  $0 < \alpha < 2$ , não possuem variância finita. A Definição

A.2 é a mesma apresentada por Samorodnitsky e Taqqu (2000, p. 376) e de modo similar as que fazem estes autores, consideramos ao longo de nosso trabalho que as inovações  $(\varepsilon_n)_{n \in \mathbb{N}}$  são i.i.d. com distribuição  $\alpha$ -estável simétrica, isto é,  $\varepsilon_n \sim S_\alpha(\sigma, 0, 0)$ , para todo  $n \in \mathbb{N}$ , com  $0 < \alpha \leq 2$ .

## Estimadores de Yule-Walker

Considere a seguinte definição de autocovariância amostral.

**Definição A.3.** Seja  $X_1, \dots, X_n$  uma amostra aleatória qualquer. A função de *autocovariância amostral*  $\hat{\gamma}_X(\cdot)$  desta amostra é definida por

$$\hat{\gamma}_X(h) = n^{-1} \sum_{k=1}^{n-h} (X_{k+h} - \bar{X}_n)(X_k - \bar{X}_n), \quad 0 \leq h < n,$$

onde  $\bar{X}_n = n^{-1} \sum_{k=1}^n X_k$  denota a média amostral de  $X_1, \dots, X_n$ .

Através da autocovariância amostral, podemos definir um estimador para os parâmetros do processo autorregressivo apresentado na Definição A.2.

**Definição A.4.** Seja  $(X_n)_{n \in \mathbb{Z}}$  um processo AR(p). Os *estimadores de Yule-Walker*  $\hat{\boldsymbol{\theta}}$  e  $\hat{\sigma}_\varepsilon^2$ , respectivamente, de  $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_p)$  e  $\sigma_\varepsilon^2$ , onde  $\sigma_\varepsilon^2 = \text{Var}(\varepsilon_1)$ , podem ser determinados a partir das equações

$$\hat{\Gamma}_p \hat{\boldsymbol{\theta}} = \hat{\boldsymbol{\gamma}}_p \tag{A.2}$$

e

$$\hat{\sigma}_\varepsilon^2 = \hat{\gamma}_X(0) - \hat{\boldsymbol{\theta}}^T \hat{\boldsymbol{\gamma}}_p, \tag{A.3}$$

onde  $\hat{\Gamma}_p = [\hat{\gamma}_X(i-j)]_{i,j=1}^p$  e  $\hat{\boldsymbol{\gamma}}_p = (\hat{\gamma}_X(1), \hat{\gamma}_X(2), \dots, \hat{\gamma}_X(p))^T$ .

Para  $p = 1$  na Definição A.4 e  $n \geq 2$ , a equação (A.2) recai no caso unidimensional

$$\begin{aligned} \hat{\Gamma}_1 \tilde{\theta}_n = \hat{\boldsymbol{\gamma}}_1 &\Rightarrow \tilde{\theta}_n = \frac{\hat{\gamma}_X(1)}{\hat{\gamma}_X(0)} \\ &\Rightarrow \tilde{\theta}_n = \frac{n^{-1} \sum_{k=1}^{n-1} (X_{k+1} - \bar{X}_n)(X_k - \bar{X}_n)}{n^{-1} \sum_{k=1}^n (X_k - \bar{X}_n)(X_k - \bar{X}_n)} \\ &\Rightarrow \tilde{\theta}_n = \frac{\sum_{k=2}^n X_k X_{k-1}}{\sum_{k=1}^n X_k^2}. \end{aligned} \tag{A.4}$$

**Observação A.5:** Para obter a expressão (A.4) a partir da igualdade anterior, levamos em consideração o fato de que o processo  $(X_n)_{n \in \mathbb{N}}$  possui média 0. Assim, para  $n$  suficientemente grande,  $\bar{X}_n \approx 0$ .



## Método dos Mínimos Quadrados

Uma outra alternativa para estimar o parâmetro  $\theta$  da expressão (3.1) é através do método de mínimos quadrados. No caso de um AR(1), o método consiste em minimizar a soma dos quadrados dos ruídos com respeito a  $\theta$ , isto é, queremos minimizar o valor de  $\sum_{k=2}^n \varepsilon_k^2 = \sum_{k=2}^n (X_k - \theta X_{k-1})^2$ . Derivando em relação a  $\theta$  e posteriormente igualando o resultado à zero, obtemos

$$\begin{aligned}
 0 &= \frac{\partial}{\partial \theta} \sum_{k=2}^n (X_k - \theta X_{k-1})^2 \Rightarrow 0 = - \sum_{k=2}^n 2(X_k - \theta X_{k-1})X_{k-1} \\
 &\Rightarrow \sum_{k=2}^n X_k X_{k-1} = \hat{\theta}_n \sum_{k=1}^n X_k^2 \\
 &\Rightarrow \hat{\theta}_n = \frac{\sum_{k=2}^n X_k X_{k-1}}{\sum_{k=1}^{n-1} X_k^2}. \tag{A.5}
 \end{aligned}$$

Assim, a expressão (A.5) fornece um estimador alternativo para o parâmetro  $\theta$  com relação à amostra  $X_1, \dots, X_n$ . Tal estimador é denominado de *estimador de mínimos quadrados* para  $\theta$  e vamos denotá-lo por  $\hat{\theta}_n$ .



# Apêndice B – Álgebra Linear

Neste apêndice reunimos alguns dos resultados de Álgebra Linear que são a base para a teoria desenvolvida neste trabalho. Apesar de básicos, servem como ferramentas poderosas para atacar os problemas abordados nos Capítulos 3 e 4.

**Definição B.1.** Uma matriz simétrica é *não-negativa definida* (resp. *positiva definida*) se, e somente se, seus autovalores são não-negativos (resp. positivos).

**Lema B.2** (Lema 3.16 do Capítulo 3). *Sejam  $f, g \in L^\infty(\mathbb{T})$ . Defina para cada  $n \in \mathbb{N}$ ,  $\underline{a}_n(f, g)$  e  $\bar{a}_n(f, g)$  como sendo o menor e o maior dos  $n$  autovalores da matriz  $T_n(f)T_n(g)$ , respectivamente. Então as sequências  $(\underline{a}_n(f, g))_{n \in \mathbb{N}}$  e  $(\bar{a}_n(f, g))_{n \in \mathbb{N}}$  são ambas limitadas por  $\|f\|_\infty \|g\|_\infty$ .*

Para demonstrar este lema, vamos antes definir uma norma para o espaço vetorial das matrizes reais quadradas  $M_{n \times n}(\mathbb{R})$  de ordem  $n$ .

**Definição B.3.** Considere uma matriz  $A \in M_{n \times n}(\mathbb{R})$ , com valores singulares  $\{s_1, \dots, s_n\}$ , isto é, as raízes quadradas dos autovalores da matriz  $A^*A$ , onde  $A^*$  denota a matriz adjunta de  $A$ . A *norma espectral* da matriz  $A$  será definida por

$$\|A\|_\infty = \max_{1 \leq k \leq n} \{s_k\}. \quad (\text{B.1})$$

**Observação B.4:** Não é difícil mostrar que  $\|\cdot\|_\infty$  definida em (B.1) é de fato uma norma (ver teorema 7.65 de Santana e Queiró, 2010). Além disso, esta norma pode ser estendida para qualquer transformação linear  $A : E \rightarrow F$ , entre espaços vetoriais de dimensão finita munidos de produto interno e podemos mostrar que

$$\|A\|_\infty = s_{\max} = \max\{\|Au\|_F; u \in E, \|u\|_E = 1\}, \quad (\text{B.2})$$

onde  $s_{\max}^2$  é o maior autovalor do operador  $A^*A : E \rightarrow E$  e  $\|\cdot\|_E$  e  $\|\cdot\|_F$  são normas em  $E$  e  $F$ , respectivamente. A prova da segunda igualdade em (B.2) segue da aplicação do Teorema dos Valores Singulares e das propriedades do máximo (veja o teorema 13.10 e o exercício 13.28 de Lima, 2014). Note que  $s_{\max} \geq 0$  pela definição de valor singular.

**Demonstração do Lema B.2.** Observe inicialmente que se  $A$  é uma matriz hermitiana, isto é,  $A^* = A$ , então todos os seus autovalores são também valores singulares. De fato, se  $v$  é autovetor de  $A$  associado ao autovalor  $\lambda$ , então

$$A^*Av = A^*\lambda v = \lambda A^*v = \lambda Av = \lambda^2 v.$$

Como as matrizes  $T_n(f)$  e  $T_n(g)$  são ambas hermitianas, o produto também é uma matriz hermitiana. É então suficiente mostrar que os valores singulares de  $T_n(f)T_n(g)$  são limitados por  $\|f\|_\infty\|g\|_\infty$ , pois neste caso teríamos que, sendo  $\sqrt{\lambda^2}$  valor singular de  $T_n(f)T_n(g)$ , então

$$|\lambda| = \sqrt{\lambda^2} \leq \|f\|_\infty\|g\|_\infty \quad \Leftrightarrow \quad -\|f\|_\infty\|g\|_\infty \leq \lambda \leq \|f\|_\infty\|g\|_\infty,$$

para todo  $\lambda$  autovalor de  $T_n(f)T_n(g)$ .

Ora, note que  $\|T_n(f)T_n(g)\|_\infty \leq \|T_n(f)\|_\infty\|T_n(g)\|_\infty$ . Com efeito, seja  $\|T_n(f)\|_\infty = a$  e  $\|T_n(g)\|_\infty = b$ . Considere  $u \in \mathbb{R}^n$  unitário. Se  $T_n(g)u = 0$ , então  $T_n(f)T_n(g)u = 0 \leq ab$ , pois  $a$  e  $b$  são ambos não-negativos.

Se  $T_n(g)u \neq 0$ , observe que  $\frac{T_n(g)u}{\|T_n(g)u\|}$  é um vetor unitário, e portanto, segue da expressão (B.2) que

$$\left\| T_n(f) \frac{T_n(g)u}{\|T_n(g)u\|} \right\| \leq a \quad \text{e} \quad \|T_n(g)u\| \leq b.$$

Aplicando as propriedades de uma norma, concluímos que

$$\|T_n(f)T_n(g)u\| = \left\| T_n(f)T_n(g) \frac{\|T_n(g)u\|}{\|T_n(g)u\|} u \right\| = \left\| T_n(f) \frac{T_n(g)u}{\|T_n(g)u\|} \right\| \|T_n(g)u\| \leq ab.$$

Por fim, como  $u$  foi escolhido arbitrariamente do conjunto  $\{u \in \mathbb{R}^n : \|u\| = 1\}$ , podemos concluir que  $\|T_n(f)T_n(g)u\| \leq ab$ , dentro deste conjunto. Tomando o máximo, segue que

$$\|T_n(f)T_n(g)\|_\infty = \max\{\|T_n(f)T_n(g)u\| : u \in \mathbb{R}^n, \|u\| = 1\} \leq ab = \|T_n(f)\|_\infty\|T_n(g)\|_\infty. \quad (\text{B.3})$$

Para concluir, a equação (2.6) em Avram (1988) mostra que

$$\|T_n(f)\|_\infty \leq \|f\|_\infty.$$

Combinando este resultado com o que foi obtido na expressão (B.3), obtemos que

$$\|T_n(f)T_n(g)\|_\infty \leq \|T_n(f)\|_\infty\|T_n(g)\|_\infty \leq \|f\|_\infty\|g\|_\infty. \quad (\text{B.4})$$

Assim, como (B.4) é válida para todo  $n \in \mathbb{N}$  e  $\|T_n(f)T_n(g)\|_\infty$  é por definição o máximo dos valores singulares de  $T_n(f)T_n(g)$ , o lema fica demonstrado.  $\square$

**Proposição B.5.** *Dada uma matriz  $\Sigma_{n \times n}$  simétrica, existe uma matriz  $P$  ortogonal  $n \times n$  ( $P^T = P^{-1}$ ) e uma matriz diagonal  $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_n)$ , onde  $\lambda_1, \dots, \lambda_n$  são os autovalores de  $\Sigma_{n \times n}$ , tal que*

$$\Sigma_{n \times n} = P\Lambda P^T. \quad (\text{B.5})$$

*Demonstração.* É simples demonstrar que são reais todos os autovalores próprios de uma matriz simétrica real (Teorema 7.42 de Santana e Queiró (2010, p. 234)). Desta forma, o

Teorema de Schur (ver Santana e Queiró (2010, p. 234-235)) garante que se os autovalores de uma matriz real quadrada são todos reais, então existe  $Q$  ortogonal tal que

$$Q^T A Q = M$$

onde  $M$  é uma matriz triangular superior  $M$ , com os autovalores de  $A$  figurando na diagonal principal de  $M$ . Se  $A$  for simétrica, aplicando a transposição em  $Q^T A Q = M$ , obtém-se

$$Q^T A Q = M^T,$$

pois  $A^T = A$ . Mas então  $M^T = M$ , o que mostra que  $M$  precisa ser uma matriz diagonal.  $\square$

**Definição B.6.** Dizemos que  $X$  é uma *raiz quadrada* da matriz  $A$ , sempre que  $X^2 = A$ . Quando a matriz  $X$  for única, usamos a notação  $A^{1/2}$ .

De acordo com o Teorema 13.8 em Lima (2014), toda matriz real não-negativa definida  $A$ , possui uma única raiz quadrada não-negativa definida, a qual é positiva definida se, e somente se,  $A$  é positiva definida. Desta forma, é plausível o seguinte lema.

**Lema B.7.** *Dadas duas matrizes  $A$  e  $B$ , ambas de ordem  $n \times n$ , com  $B$  não-negativa definida e cujo produto seja comutativo, então:*

- (a)  $\rho$  é autovalor de  $AB$  se, e somente se, é autovalor de  $B^{1/2}AB^{1/2}$ ;
- (b)  $\rho$  e  $v$  são, respectivamente, autovalor e autovetor da matriz  $A$  se, e somente se,  $1 - 2\lambda\rho$  e  $v$  são, respectivamente, autovalor e autovetor da matriz  $I - 2\lambda A$ .

*Demonstração.* a) Seja  $v$  autovetor de  $AB$  associado ao autovalor  $\rho$ . Então como  $B$  é por hipótese não-negativa definida, a matriz  $B^{1/2}$  existe e  $\omega = B^{1/2}v$  é autovetor de  $B^{1/2}AB^{1/2}$  associado ao autovalor  $\rho$ .

De fato,

$$\begin{aligned} ABv = \rho v &\Rightarrow AB^{1/2}B^{1/2}v = \rho v \\ &\Rightarrow B^{1/2}AB^{1/2}\underbrace{B^{1/2}v}_{\omega} = B^{1/2}\rho v = \rho\underbrace{B^{1/2}v}_{\omega} \\ &\Rightarrow B^{1/2}AB^{1/2}\omega = \rho\omega. \end{aligned}$$

Reciprocamente, seja  $v$  um autovetor de  $B^{1/2}AB^{1/2}$  associado ao autovalor  $\rho$ . Segue então que

$$\begin{aligned} B^{1/2}AB^{1/2}v = \rho v &\Rightarrow BA\underbrace{B^{1/2}v}_{\omega} = \rho\underbrace{B^{1/2}v}_{\omega} \\ &\Rightarrow BA\omega = \rho\omega \end{aligned}$$

(o produto entre  $A$  e  $B$  é comutativo)

$$\Rightarrow AB\omega = \rho\omega$$

b) Se  $\rho$  e  $v$  são autovalor e autovetor de  $A$ , então

$$(I - 2\lambda A)v = v - 2\lambda Av = v - 2\lambda\rho v = (1 - 2\lambda\rho)v.$$

Reciprocamente, se  $\lambda = 0$ , o resultado é óbvio. Do contrário, sejam  $(1 - 2\lambda\rho)$  e  $v$  autovalor e autovetor associados à matriz  $I - 2\lambda A$ , segue disto que

$$\begin{aligned} (I - 2\lambda A)v = (1 - 2\lambda\rho)v &\Rightarrow v - 2\lambda Av = v - 2\lambda\rho v \\ &\Rightarrow 2\lambda Av = 2\lambda\rho v \end{aligned}$$

$\lambda \neq 0$

$$\Rightarrow Av = \rho v.$$

□

Vamos agora demonstrar um resultado que combina Álgebra Linear com a Teoria das Probabilidades. Considere o vetor aleatório  $\mathbf{X}_n = (X_1, \dots, X_n)$  cujas coordenadas são as variáveis aleatórias que surgem do processo AR(1) apresentado no Capítulo 3, e que satisfazem a equação (3.1).

Pretendemos mostrar que o vetor aleatório  $\mathbf{X}_n$  pode ser escrito como um produto entre uma matriz  $B$  e um vetor aleatório  $\mathbf{N}_n$  com distribuição normal padrão  $n$ -variada. Para isto, defina  $B = T_n^{1/2}(g)$ , onde  $g(\cdot)$  é a função densidade espectral do processo  $(X_n)_{n \in \mathbb{N}}$  definida por  $g(\omega) = (1 + \theta^2 + 2\theta \cos(\omega))^{-1}$ , para  $-\pi \leq \omega < \pi$ . Observe que  $B$  está bem definida, pois a matriz  $T_n(g)$  é não-negativa definida, e possui portanto, uma raiz quadrada. Definindo

$$\mathbf{Y}_n = B\mathbf{N}_n, \tag{B.6}$$

concluimos que  $\mathbf{Y}_n$  segue uma distribuição normal  $n$ -variada de média  $\mathbb{E}(\mathbf{Y}_n) = \mathbf{0}$  e matriz de variância (ver proposição 1.6.1 de Brockwell e Davis, 1991)

$$\Sigma_{\mathbf{Y}_n \mathbf{Y}_n} = B I_n B^T = T_n^{1/2}(g) T_n^{1/2}(g) = T_n(g).$$

Uma vez que a distribuição normal multivariada é univocamente determinada a partir de sua média e matriz de variância (ver proposição 1.6.4 de Brockwell e Davis, 1991), segue que  $\mathbf{Y}_n$  e  $\mathbf{X}_n$  são identicamente distribuídas.

**Observação B.8:** Dado que  $T_n(g)$  é simétrica, pela Proposição B.5 existe uma matriz ortogonal  $P$  tal que

$$T_n(g) = P \Lambda P^T,$$

com  $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_n)$ , onde  $\lambda_1, \dots, \lambda_n$  são os autovalores de  $T_n(g)$ . Além disso, é não-negativa definida e única a matriz raiz quadrada de  $T_n(g)$ . Defina  $C = P\Lambda^{1/2}P^T$ , onde  $\Lambda^{1/2} = \text{diag}(\sqrt{\lambda_1}, \dots, \sqrt{\lambda_n})$ . A matriz  $C$  é simétrica, pois

$$C^T = (P\Lambda^{1/2}P^T)^T = (P^T)^T\Lambda^{1/2}P^T = C.$$

Uma vez que  $\Lambda^{1/2}\Lambda^{1/2} = \Lambda$  e pela ortogonalidade de  $P$ , vale que

$$CC = P\Lambda^{1/2}P^T P\Lambda^{1/2}P^T = P\Lambda P^T = T_n(g),$$

mostrando por unicidade que  $P\Lambda^{1/2}P^T$  é a raiz quadrada de  $T_n(g)$ . Assim, conhecendo os autovalores da matriz  $T_n(g)$  e a matriz ortogonal  $P$ , é simples de encontrar uma expressão para a sua raiz.

**Lema B.9.** *O determinante da matriz  $T_n^{-1}(g)$  é igual à  $1 - \theta^2$ .*

*Demonstração.* Se  $\theta = 0$ , o resultado é trivial. Assumimos então que  $0 < |\theta| < 1$ . Denote por  $G_n$  o determinante da matriz

$$\begin{pmatrix} 1 + \theta^2 & -\theta & 0 & \dots \\ -\theta & 1 + \theta^2 & -\theta & \dots \\ \dots & \dots & \dots & \dots \\ \dots & -\theta & 1 + \theta^2 & -\theta \\ \dots & 0 & -\theta & 1 + \theta^2 \end{pmatrix}.$$

O teorema de Laplace afirma que o determinante de uma matriz quadrada é igual à soma dos produtos dos elementos de uma qualquer linha pelos respectivos complementos algébricos, isto é, sendo  $A = [a_{l,j}]_{n \times n}$ , temos

$$\det(A) = \sum_{j=1}^n a_{l,j}(-1)^{l+j} \det(A_{l,j}),$$

onde a matriz  $A_{l,j}$  é a submatriz obtida de  $A$  com a linha  $l$  e a coluna  $j$  suprimidas. Assim, é simples ver que

$$G_{n+2} = (1 + \theta^2)G_{n+1} - \theta^2 G_n, \quad \forall n \in \mathbb{N}, \quad (\text{B.7})$$

desde que  $G_1 = 1 + \theta^2$  e  $G_2 = 1 + \theta^2 + \theta^4$ .

Resolvendo a equação de recorrência, observamos que a equação (B.7) tem  $r^2 - (1 - \theta^2)r + \theta^2 = 0$  como equação característica. As raízes desta equação característica são 1 e  $\theta^2$ . Assim, todas as sequências da forma  $a_n = C_1 + C_2\theta^{2n}$ , para  $C_1$  e  $C_2 \in \mathbb{R}$ , são soluções da recorrência em (B.7). Usando as condições iniciais  $G_1 = 1 + \theta^2$  e  $G_2 = 1 + \theta^2 + \theta^4$ , somos capazes de deduzir que  $C_1 = 1/(1 - \theta^2)$  e  $C_2 = -\theta^2/(1 - \theta^2)$ . E portanto,

$$G_n = \frac{1}{1 - \theta^2} - \frac{\theta^2}{1 - \theta^2} \theta^{2n} = \frac{1 - \theta^{2(n+1)}}{1 - \theta^2} \quad (\text{B.8})$$

é uma solução particular da recorrência em (B.7).

Provamos agora, com auxílio da expressão (B.8), o resultado do lema. Para  $n \in \{1, 2, 3, 4\}$ , o resultado é computacionalmente trivial. Entretanto, se  $n \geq 5$ , usando o Teorema de Laplace vemos pela expressão da matriz inversa de  $T_n(g)$ , dada em (3.77), que

$$\begin{aligned} \det(T_n^{-1}(g)) &= G_{n-2} - \theta^2 G_{n-3} - \theta^2(G_{n-3} - \theta^2 G_{n-4}) \\ &= \frac{1 - \theta^{2(n-1)}}{1 - \theta^2} - 2\theta^2 \left( \frac{1 - \theta^{2(n-2)}}{1 - \theta^2} \right) + \theta^4 \left( \frac{1 - \theta^{2(n-3)}}{1 - \theta^2} \right) \\ &= \frac{1 - \theta^{2(n-1)} - 2\theta^2 + 2\theta^{2(n-1)} + \theta^4 - \theta^{2(n-1)}}{1 - \theta^2} = \frac{(1 - \theta^2)^2}{1 - \theta^2} = 1 - \theta^2. \end{aligned}$$

□

O seguinte lema mostra que para conhecer o domínio em que uma matriz da forma  $I_n - 2\lambda M_n T_n(g)$  é positiva definida, para  $M_n$  simétrica real, basta analisar onde a matriz  $D_n = T_n^{-1}(g) - 2\lambda M_n$  o é. Assim, se  $D_n$  for tri-diagonal (caso que nos interessa), será sempre mais simples de ser estudada do que uma matriz genérica  $n \times n$ , tal como  $I_n - 2\lambda M_n T_n(g)$ .

**Lema B.10.** *Se  $D_n = T_n^{-1}(g) - 2\lambda M_n$  é uma matriz positiva definida, então  $I_n - 2\lambda M_n T_n(g)$  também é uma matriz positiva definida.*

*Demonstração.* Uma forma equivalente de definir uma matriz positiva definida é através de operadores auto-adjuntos e seu respectivo produto interno. Para isso, dada a matriz quadrada real  $\mathbf{a} = [a_{i,j}]_{n \times n}$ , defina  $A : \mathbb{R}^n \mapsto \mathbb{R}^n$  como sendo o operador linear cuja matriz na base canônica é  $\mathbf{a}$ . De acordo com Lima (2014, p. 168-169), uma matriz quadrada  $\mathbf{a}$  diz-se *positiva definida* quando o operador  $A$ , que a ela corresponde, é positivo. Em adição, dizemos que o operador  $A$  é *positivo* quando as três seguintes condições são satisfeitas:

- i)  $A$  é auto-adjunto ( $A^* = A$ , onde  $A^*$  é a adjunta de  $A$ );
- ii)  $\langle Av, v \rangle \geq 0$  para todo  $v \neq 0$  e;
- iii)  $A$  é invertível.

Com base nesta definição, provaremos de uma maneira mais geral que, se  $A : \mathbb{R}^n \mapsto \mathbb{R}^n$  é um operador positivo definido, então, para toda transformação linear  $B : \mathbb{R}^n \mapsto \mathbb{R}^n$  invertível, o produto  $B^*AB$  é positivo definido.

De fato,

- i)  $B^*AB$  é auto-adjunto pois

$$(B^*AB)^* = B^*A^*(B^*)^* = B^*AB.$$



ii) Além disso, se  $v \in \mathbb{R}^n$  e  $w = Bv$ , então temos que

$$\langle B^*ABv, v \rangle = \langle ABv, Bv \rangle = \langle Aw, w \rangle.$$

Como  $A$  é positivo definido, segue que  $\langle B^*ABv, v \rangle = \langle Aw, w \rangle \geq 0$ .

iii) Finalmente, suponha que  $B^*ABv = 0$ , então temos que (ver corolário 1 da p. 169 em Lima, 2014)

$$0 = \langle B^*ABv, v \rangle = \langle Aw, w \rangle \Rightarrow w = 0,$$

isto é,  $Bv = 0$ . Como  $B$  é uma transformação invertível, isso só pode ocorrer quando  $v$  for o vetor nulo (ver teorema 6.2 de Lima, 2014), o que mostra que  $B^*AB$  é injetivo, e conseqüentemente (ver corolário da p. 69 em Lima, 2014), sobrejetivo, ou seja,  $B^*AB$  é invertível.

Em conclusão, uma vez que a matriz  $T_n^{1/2}(g)$  é simétrica e invertível, o que na linguagem de operadores se traduziria como um operador auto-adjunto e invertível, segue do que provamos acima, com  $A$  e  $B$  os respectivos operadores lineares associados às matrizes  $D_n$  e  $T_n^{1/2}(g)$  que

$$I_n - 2\lambda M_n T_n(g) = T_n^{1/2}(g) D_n T_n^{1/2}(g)$$

é positiva definida. □

**Lema B.11.** Denote por  $H_n$  o determinante da matriz quadrada de ordem  $n$

$$\begin{pmatrix} p & q & 0 & \cdots \\ q & p & q & \cdots \\ \cdots & \cdots & \cdots & \cdots \\ \cdots & q & p & q \\ \cdots & 0 & q & 1 \end{pmatrix}_{n \times n}.$$

Então, de acordo com Jensen (1995, p. 269), uma forma fechada para  $H_n$  é dada por

$$H_n = \frac{(1 - \tau)\rho^n + (\rho - 1)\tau^n}{\rho - \tau}, \quad (\text{B.9})$$

onde as constantes  $\rho$  e  $\tau$  são dadas respectivamente por

$$\rho = \frac{p + \sqrt{p^2 - 4q^2}}{2} \quad e \quad \tau = \frac{p - \sqrt{p^2 - 4q^2}}{2}. \quad (\text{B.10})$$

*Demonstração.* É fácil ver que, através das operações usuais com determinantes,  $(H_n)_{n \in \mathbb{N}}$  satisfaz a recursão

$$H_{n+2} = p H_{n+1} - q^2 H_n, \quad \forall n \in \mathbb{N}, \quad (\text{B.11})$$

onde por convenção, adotamos que  $H_1 = 1$  e  $H_2 = p - q^2$  é o determinante da matriz

$$\begin{pmatrix} p & q \\ q & 1 \end{pmatrix}.$$

Assim, a equação (B.11) tem  $r^2 - pr + q^2 = 0$  como equação característica. As raízes desta equação característica são as constantes  $\rho$  e  $\tau$ , definidas em (B.10). Assim, todas as sequências da forma  $a_n = C_1\rho^n + C_2\tau^n$ , para  $C_1$  e  $C_2 \in \mathbb{R}$ , são soluções da recorrência em (B.11). Usando as condições iniciais  $H_1 = 1$  e  $H_2 = p - q^2$ , somos capazes de deduzir através da resolução de um sistema linear  $2 \times 2$  que

$$C_1 = \frac{1 - \tau}{\rho - \tau} \quad \text{e} \quad C_2 = \frac{\rho - 1}{\rho - \tau}.$$

Portanto,

$$H_n = \frac{1 - \tau}{\rho - \tau}\rho^n + \frac{\rho - 1}{\rho - \tau}\tau^n$$

é uma solução particular da recorrência em (B.11). □

# Apêndice C – Distribuições $\alpha$ -Estáveis Especiais e Suas Propriedades

Neste apêndice apresentamos algumas distribuições especiais e listamos algumas de suas propriedades, tal como sua função característica, função densidade, relação com as variáveis  $\alpha$ -estáveis, etc.

## Exemplo C.1. (Distribuição de Cauchy).

Dizemos que  $X$  tem *distribuição de Cauchy* com parâmetro de deslocamento  $\mu$  e parâmetro de escala  $\sigma > 0$ , com notação  $X \sim \mathcal{C}(\mu, \sigma)$ , se possui função de densidade

$$f_X(x) = \frac{\sigma}{\pi[\sigma^2 + (x - \mu)^2]}. \quad (\text{C.1})$$

Além disso, dizemos que  $X$  tem *distribuição Cauchy padrão* quando  $\mu = 0$  e  $\sigma = 1$ .

Vamos mostrar que a função  $f_X(\cdot)$  definida em (C.1) é uma função de densidade. De fato, integrando com relação à  $x$  e realizando a substituição de variáveis  $y = (x - \mu)/\sigma$ , obtemos

$$\begin{aligned} \int_{\mathbb{R}} f_X(x) dx &= \int_{\mathbb{R}} \frac{\sigma}{\pi[\sigma^2 + (x - \mu)^2]} dx = \int_{\mathbb{R}} \frac{\sigma}{\pi\sigma^2 \left[1 + \left(\frac{x-\mu}{\sigma}\right)^2\right]} dx = \frac{1}{\pi} \int_{\mathbb{R}} \frac{1}{1 + y^2} dy \\ &= \frac{2}{\pi} \int_0^{+\infty} \frac{1}{1 + y^2} dy = \frac{2}{\pi} \times \frac{\pi}{2} = 1. \end{aligned}$$

A distribuição de Cauchy é também conhecida por ser um caso particular das distribuições  $\alpha$ -estáveis quando  $\alpha = 1$  e  $\beta = 0$ . Com efeito, observamos primeiramente que se  $Y$  é uma variável aleatória tendo função densidade dada por

$$f_Y(y) = \frac{1}{2} e^{-|y|}, \quad y \in \mathbb{R}, \quad (\text{C.2})$$

então a função geradora de momentos de  $Y$  é igual à  $M_Y(\omega) = (1 - \omega^2)^{-1}$ ,  $-1 < \omega < 1$ . Para provar esta afirmação, vamos separar os intervalos de integração em torno de  $y = 0$ .

$$\begin{aligned} M_Y(\omega) &= \int_{\mathbb{R}} e^{\omega y} \frac{e^{-|y|}}{2} dy = \int_{-\infty}^0 \frac{e^{y(\omega+1)}}{2} dy + \int_0^{+\infty} \frac{e^{y(\omega-1)}}{2} dy \\ &= \frac{e^0 - \lim_{y \rightarrow -\infty} e^{y(\omega+1)}}{2(\omega+1)} + \frac{\lim_{y \rightarrow \infty} e^{y(\omega-1)} - e^0}{2(\omega-1)} \end{aligned}$$

(note que  $-1 < \omega < 1 \Rightarrow \omega - 1 < 0 < \omega + 1$ )

$$= \frac{1}{2(\omega + 1)} - \frac{1}{2(\omega - 1)} = \frac{\omega - 1 - (\omega + 1)}{2(\omega^2 - 1)} = \frac{-2}{2(\omega^2 - 1)} = (1 - \omega^2)^{-1}.$$

Da função geradora de momentos de  $Y$ , podemos extrair sua função característica avaliando a função geradora de momentos no argumento  $i\omega$ . Com efeito,

$$\varphi_Y(\omega) = M_Y(i\omega) = (1 - (i\omega)^2)^{-1} = (1 + \omega^2)^{-1}.$$

Isso quer dizer que

$$(1 + \omega^2)^{-1} = \varphi_Y(\omega) = \int_{\mathbb{R}} e^{i\omega y} \frac{e^{-|y|}}{2} dy,$$

equação que multiplicada por  $1/\pi$ , é equivalente à

$$\frac{1}{\pi(1 + \omega^2)} = \frac{1}{2\pi} \int_{\mathbb{R}} e^{i\omega y} e^{-|y|} dy. \quad (\text{C.3})$$

Mas observe que a expressão da direita em (C.3) é explicitamente a inversa da Transformada de Fourier da função  $e^{-|y|}$ , o que nos permite concluir que

$$e^{-|y|} = \int_{\mathbb{R}} e^{-i\omega y} \frac{1}{\pi(1 + \omega^2)} d\omega. \quad (\text{C.4})$$

Note que  $1/[\pi(1 + x^2)]$  é a função de densidade de uma distribuição Cauchy padrão. Sendo assim, se  $X \sim \mathcal{C}(0, 1)$ , sua função característica é dada por

$$\varphi_X(t) = \int_{\mathbb{R}} e^{ixt} \frac{1}{\pi(1 + x^2)} dx = \int_{\mathbb{R}} e^{-ix(-t)} \frac{1}{\pi(1 + x^2)} dx = e^{-|-t|} = e^{-|t|},$$

onde na penúltima igualdade, usamos o resultado da equação (C.4) com  $-t$  e  $x$  nos lugares de  $y$  e  $\omega$ , respectivamente.

No caso em que  $Y \sim \mathcal{C}(\mu, \sigma)$ , basta notar que  $Y = \sigma X + \mu$ , onde  $X$  segue uma distribuição Cauchy padrão. Para obter a função característica da variável  $Y$ , procedemos da seguinte forma

$$\varphi_Y(t) = \mathbb{E}(e^{itY}) = \mathbb{E}(e^{it(\sigma X + \mu)}) = \mathbb{E}(e^{it\sigma X} e^{it\mu}) = e^{it\mu} \varphi_X(t\sigma) = e^{it\mu} e^{-|t\sigma|} = e^{it\mu} e^{-\sigma|t|}, \quad (\text{C.5})$$

lembrado que  $\sigma$  é por hipótese, uma constante positiva.

□

Vamos agora mostrar que a distribuição de Cauchy pertence a classe das distribuições  $\alpha$ -estáveis.

**Proposição C.2.**  $S_1(\sigma, 0, \mu) \stackrel{d}{=} \mathcal{C}(\mu, \sigma)$ .

*Demonstração.* O argumento da prova se baseia na unicidade da função característica. De acordo com a equação (4.4), fixando  $\alpha = 1$  e  $\beta = 0$ , se  $X \sim S_1(\sigma, 0, \mu)$ , sua função característica é dada por

$$\begin{aligned}\varphi_X(t) &= \exp\{-\sigma|t|[1 + i\beta\frac{2}{\pi}(\text{sign}(t))\log|t|] + i\mu t\} \\ &= \exp\{-\sigma|t| + i\mu t\},\end{aligned}$$

o que coincide com a função característica da Equação (C.5), mostrando que também  $X \sim \mathcal{C}(\mu, \sigma)$ .  $\square$

### Exemplo C.3. (Distribuição de Lévy).

No caso em que  $X$  segue uma distribuição  $\alpha$ -estável com parâmetros  $\alpha = 1/2$  e  $\beta = 1$ , diremos que  $X$  segue uma *distribuição de Lévy*, isto é,  $X \sim S_{1/2}(\sigma, 1, \mu)$ . A função densidade de tal variável aleatória é dada por

$$f_X(x) = \left(\frac{\sigma}{2\pi}\right)^{1/2} \frac{1}{(x - \mu)^{3/2}} \exp\left[-\frac{\sigma}{2(x - \mu)}\right], \quad x \in (\mu, +\infty). \quad (\text{C.6})$$

Além disso, dizemos que  $X$  tem *distribuição Lévy padrão* quando  $\mu = 0$  e  $\sigma = 1$ .

Vamos mostrar que a função  $f_X(\cdot)$  definida em (C.6) é uma função de densidade. Com efeito, realizando a substituição  $y = (x - \mu)/\sigma$ , obtemos  $dy = dx/\sigma$ . Assim,

$$\begin{aligned}\int_{\mu}^{+\infty} f_X(x)dx &= \int_{\mu}^{+\infty} \left(\frac{1}{2\pi}\right)^{1/2} \left[\frac{\sigma}{(x - \mu)}\right]^{3/2} \exp\left[-\frac{\sigma}{2(x - \mu)}\right] \frac{dx}{\sigma} \\ &= \int_0^{+\infty} \left(\frac{1}{2\pi}\right)^{1/2} y^{-3/2} e^{-1/(2y)} dy\end{aligned} \quad (\text{C.7})$$

Fazendo uma segunda substituição de  $y$  para  $z$  através de  $z = 1/(2y)$ , obtemos as relações

$$\sqrt{z} = \frac{1}{\sqrt{2y}} \Rightarrow z\sqrt{z} = \frac{1}{2y\sqrt{2y}} \Rightarrow y^{-3/2} = 2\sqrt{2}z^{3/2} \quad (\text{C.8})$$

e

$$dz = -\frac{1}{2y^2}dy \Rightarrow dy = -2y^2dz = -2\left(\frac{1}{2z}\right)^2 dz = -\frac{1}{2z^2}dz \Rightarrow dy = -\frac{1}{2z^2}dz \quad (\text{C.9})$$

Aplicando as relações (C.8) e (C.9) na equação (C.7) e realizando uma mudança adequada dos intervalos de integração de  $(0, +\infty)$  para  $(+\infty, 0)$ , obtemos

$$\int_{\mu}^{+\infty} f_X(x)dx = \int_{+\infty}^0 \left(\frac{1}{2\pi}\right)^{1/2} (2\sqrt{2}z^{3/2})e^{-z} \left(-\frac{1}{2z^2}\right) dz = \frac{1}{\sqrt{\pi}} \int_0^{+\infty} z^{1/2-1}e^{-z}dz. \quad (\text{C.10})$$

A integral no lado direito da equação (C.10) é igual à função Gama calculada em 1/2, isto é,  $\Gamma(1/2)$ , cujo valor é conhecido como sendo  $\Gamma(1/2) = \sqrt{\pi}$  (ver p. 211 de Rohatgi e Saleh, 2000). Portanto, concluímos que

$$\int_{\mu}^{+\infty} f_X(x)dx = \frac{1}{\sqrt{\pi}} \Gamma\left(\frac{1}{2}\right) = \frac{1}{\sqrt{\pi}} \sqrt{\pi} = 1.$$

$\square$



# Apêndice D – Gráficos da Função Codiferença

Neste apêndice apresentamos alguns gráficos da função codiferença do processo autorregressivo com inovações  $\alpha$ -estáveis,  $(X_n)_{n \in \mathbb{N}}$ , definido pela expressão 4.20. Na Subseção 4.2.2 calculamos a matriz de codiferença deste processo e provamos que, se  $\tau_X(k; \alpha)$  representa a função de codiferença do processo  $(X_n)_{n \in \mathbb{N}}$  no lag  $k \in \mathbb{Z}$ , então sua expressão é dada por

$$\tau_X(k; \alpha) = \frac{\sigma^\alpha}{1 - |\theta|^\alpha} \left( 1 + |\theta|^{\alpha|k|} - (1 - \theta^{|k|})^\alpha \right),$$

para  $\alpha \in (0, 2]$  e  $|\theta| < 1$ .

Na Figura D.1 exibimos o gráfico de  $\tau_X(k; \alpha)$  para  $\sigma = 1$ ,  $k = 2$ ,  $\theta \in [-0.9, 0.9]$  e  $\alpha \in [0.1, 2]$ , onde o eixo  $z$  representa a função codiferença. Esta figura apresenta dois gráficos idênticos, cujos eixos foram rotacionados a fim de obtermos uma melhor percepção do comportamento desta função. Além disso, a legenda à direita da figura, indica o valor que a codiferença assume ao longo do eixo  $z$ . Observe pelo gráfico que a função codiferença assume valores muito grandes para  $|\theta|$  e  $\alpha$  convergindo respectivamente, para 1 e 0.

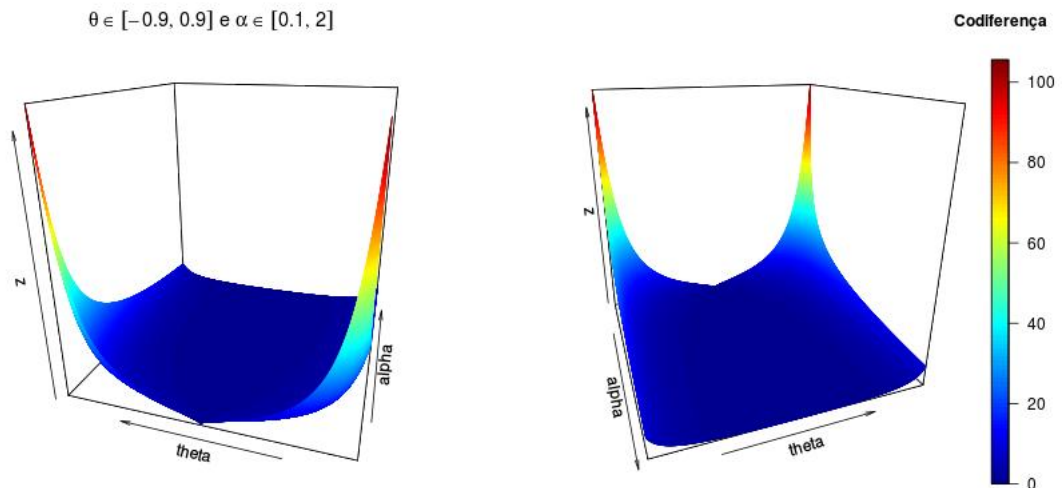


Figura D.1 – Função Codiferença para  $\sigma = 1$ ,  $k = 2$ ,  $\theta \in [-0.9, 0.9]$  e  $\alpha \in [0.1, 2]$ .

Se restringirmos  $\alpha$  e  $\theta$  aos intervalos  $[0.1, 1]$  e  $[-0.4, 0.4]$ , respectivamente, obtemos o gráfico exibido pela Figura D.2. Esta ilustração nos dá a entender que, quando  $\theta$  e  $\alpha$  estão ambos muito próximos de 0, a função codiferença não é diferenciável nestes pontos.

Muito pelo contrário, é simples de mostrar que, para  $\alpha$  e  $k$  fixos, a função

$$\phi(\theta) = \frac{\sigma^\alpha}{1 - |\theta|^\alpha} \left( 1 + |\theta|^{\alpha|k|} - (1 - \theta^{|k|})^\alpha \right)$$

possui derivadas laterais nulas, isto é,

$$\lim_{\theta \rightarrow 0^+} \frac{\phi(\theta) - \phi(0)}{\theta} = \lim_{\theta \rightarrow 0^-} \frac{\phi(\theta) - \phi(0)}{\theta} = 0.$$

E portanto,  $\phi(\theta)$  é diferenciável em 0, mostrando que  $\tau_X(k; \alpha)$  também é. Este argumento contínuo válido para qualquer  $\alpha \in (0, 2]$  e  $k \in \mathbb{Z}$ .

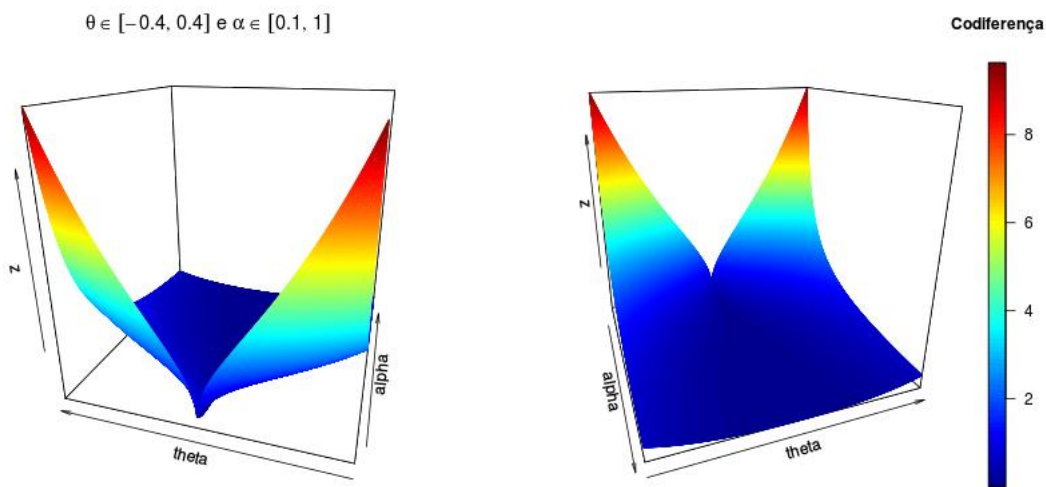


Figura D.2 – Função Codiferença para  $\sigma = 1$ ,  $k = 2$ ,  $\theta \in [-0.4, 0.4]$  e  $\alpha \in [0.1, 1]$ .

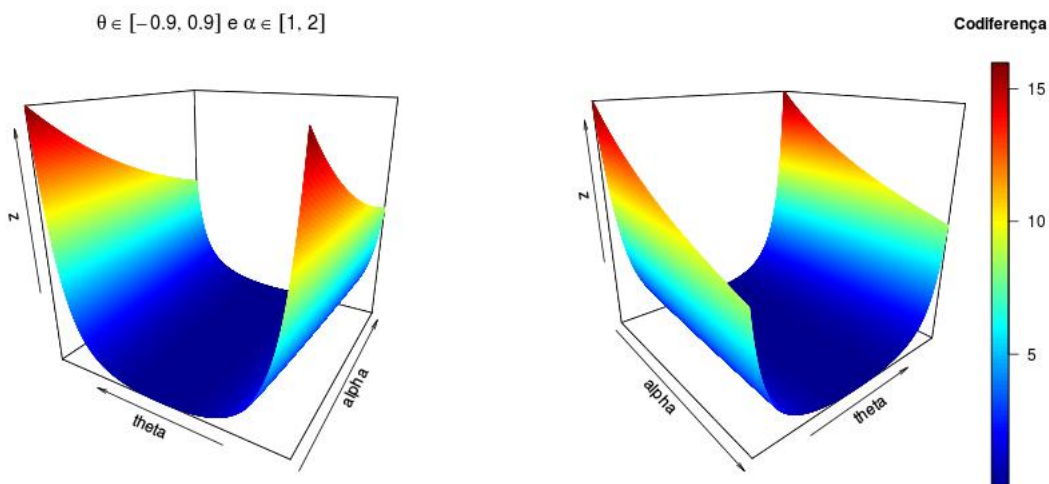


Figura D.3 – Função Codiferença para  $\sigma = 1$ ,  $k = 2$ ,  $\theta \in [-0.9, 0.9]$  e  $\alpha \in [1, 2]$ .



Por outro lado, se restringirmos  $\alpha$  e  $\theta$  aos intervalos  $[1, 2]$  e  $[-0.9, 0.9]$ , respectivamente, o gráfico que obtemos é o dado pela Figura D.3. Observamos que função codiferença é explicitamente côncava nesta situação.

No caso em que  $k = 3$ , o comportamento da função codiferença é ligeiramente diferente. A Figura D.4 mostra um gráfico da função codiferença para  $\sigma = 1$ ,  $k = 3$ ,  $\theta \in [-0.9, 0.9]$  e  $\alpha \in [0.1, 2]$ . Se restringirmos o intervalo de  $\alpha$  para  $[1, 2]$ , o gráfico que obtemos é o apresentado na Figura D.5. Note pela ilustração que, quando  $k = 3$ , a função  $\tau_X(k; \alpha)$  já não é mais côncava.

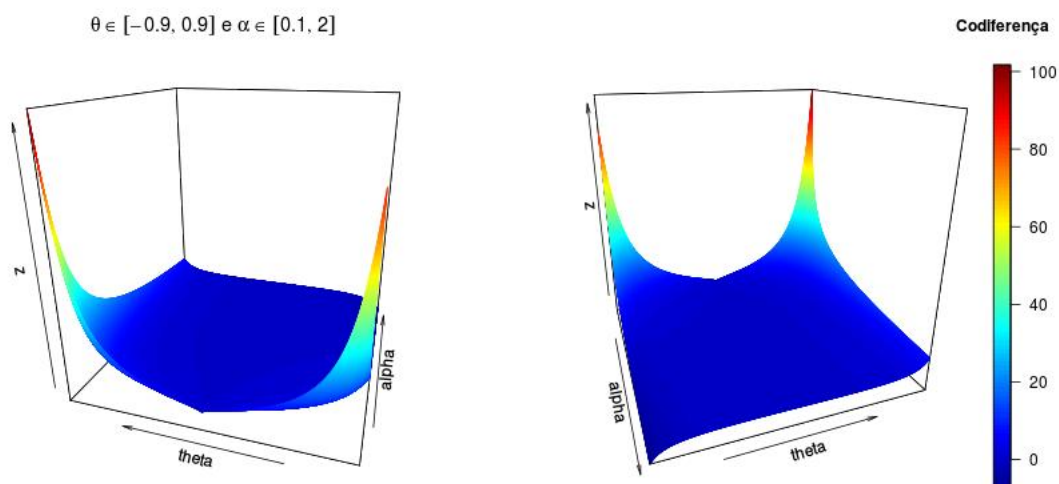


Figura D.4 – Função Codiferença para  $\sigma = 1$ ,  $k = 3$ ,  $\theta \in [-0.9, 0.9]$  e  $\alpha \in [0.1, 2]$ .

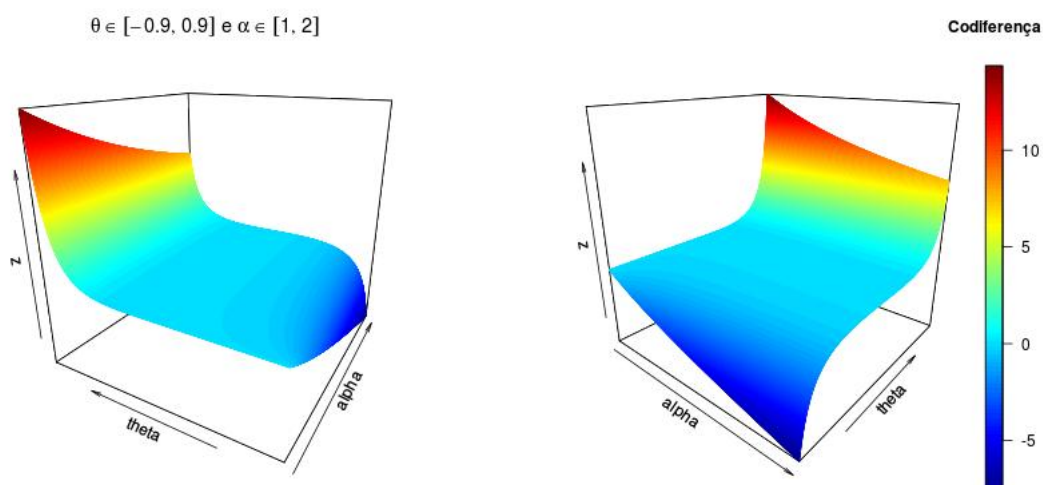


Figura D.5 – Função Codiferença para  $\sigma = 1$ ,  $k = 3$ ,  $\theta \in [-0.9, 0.9]$  e  $\alpha \in [1, 2]$ .

Na Observação 4.42, mostramos que a função  $1 + \theta^{\alpha k} - (1 - \theta^k)^\alpha$  é sempre positiva, para todo  $k \in \mathbb{N}$ ,  $\alpha \in (0, 2)$  e  $\theta > 0$ . Mas se  $k$  é par, então, esta mesma função é positiva para qualquer  $|\theta| < 1$ . Este argumento é suficiente para mostrar que a função codiferença  $\tau_X(k; \alpha)$  é positiva, para todo  $k$  par. O mesmo já não pode ser afirmado se  $k$  for um número inteiro ímpar. De fato, basta olhar para o gráfico da Figura D.5 para ver que  $\tau_X(k; \alpha)$  pode assumir valores negativos.

Para  $\alpha \in (0, 1]$ , existe sempre a garantia de que a codiferença entre duas variáveis aleatórias  $S\alpha S$  é positiva. Este é o resultado da propriedade 2.10.5 dada em Samorodnitsky e Taqqu (2000). De fato, o que esta propriedade diz é que, se  $X$  e  $Y$  forem duas variáveis  $S\alpha S$ , então

$$\tau_{X,Y} \geq 0, \quad \text{se } 0 < \alpha \leq 1. \quad (\text{D.1})$$

O mesmo já não pode ser constatado se  $1 < \alpha \leq 2$ , mas a mesma propriedade 2.10.5 de Samorodnitsky e Taqqu (2000), nos dá uma cota inferior para a codiferença entre  $X$  e  $Y$ , dada por

$$\tau_{X,Y} \geq (1 - 2^{\alpha-1})(\|X\|_\alpha^\alpha + \|Y\|_\alpha^\alpha), \quad \text{se } 1 < \alpha \leq 2. \quad (\text{D.2})$$

A propriedade (D.1) é perceptível nos gráficos que exibimos neste apêndice.

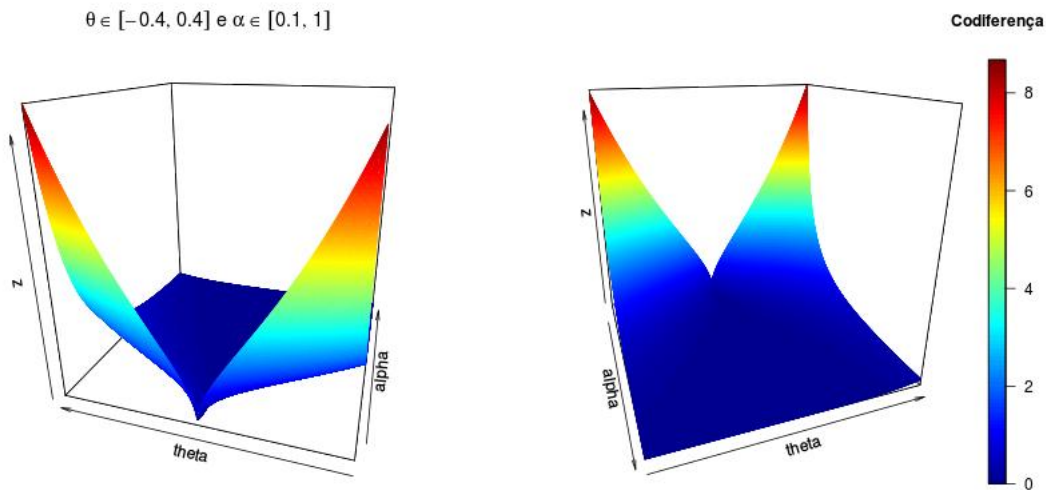


Figura D.6 – Função Codiferença para  $\sigma = 1$ ,  $k = 3$ ,  $\theta \in [-0.4, 0.4]$  e  $\alpha \in [0.1, 1]$ .