



UNIVERSIDADE DA BEIRA INTERIOR  
Ciências

# Índice de Cruzamentos: Propriedades e Inferência

**João Renato Caramona Belo Sebastião**

Tese para obtenção do Grau de Doutor em  
**Matemática**  
(3º ciclo de estudos)

Orientador: Prof. Doutora Ana Paula André Martins Fernandes  
Co-orientador: Prof. Doutora Luísa Maria Jota Pereira Amaral

**Covilhã, abril de 2013**



À Manuela, à Márcia e ao João.



# Agradecimentos

As minhas primeiras palavras de profundo agradecimento são dirigidas à Professora Doutora Ana Paula Martins pela excelente orientação deste trabalho, pela confiança que sempre em mim depositou, pela disponibilidade e generosidade reveladas ao longo destes anos, assim como pelas críticas, correções e sábias sugestões. O meu muito obrigado.

Agradeço também especialmente à Professora Doutora Luísa Pereira, minha co-orientadora, pelas suas sugestões, confiança e incentivo na realização deste trabalho.

À Professora Doutora Helena Ferreira pelas suas sugestões e comentários que permitiram elevar a qualidade deste trabalho.

Aos meus familiares, em especial aos meus filhos Márcia e João e esposa Manuela, pela infinita compreensão e inspiração, imprescindíveis nesta etapa tão importante da minha vida.

À FCT pelo apoio concedido através da bolsa de doutoramento: FCT/SFRH/BD/41439/2007.

À Escola Superior de Gestão do Instituto Politécnico de Castelo Branco pela facilidade na organização das atividades letivas.



# Resumo

Para sucessões estacionárias que verifiquem certas condições de dependência, se a sucessão de processos pontuais de cruzamentos converge em distribuição, então o processo pontual limite é necessariamente um processo de Poisson composto, cuja intensidade se relaciona com um parâmetro denominado índice de cruzamentos  $\eta$ ,  $0 \leq \eta \leq 1$ . Este coeficiente extremal pode ser visto como uma medida do agrupamento de cruzamentos de níveis elevados pelas variáveis de uma sucessão estacionária e fornece informação diferente e complementar à dada por um dos parâmetros mais importantes na Teoria de Valores Extremos, o índice extremal  $\theta$ ,  $0 \leq \theta \leq 1$ .

Nesta tese avaliamos o efeito que a subamostragem exerce sobre o valor do índice de cruzamentos  $\eta$  e, com base em diferentes caracterizações assintóticas deste parâmetro e na sua relação com o índice extremal, propomos diversos métodos para o estimar. Demonstramos várias propriedades dos estimadores propostos e aplicamo-los em amostras de dados simulados e de dados reais.

## Palavras-chave

Índice de cruzamentos, condições de dependência local, subamostragem, estimação, consistência e normalidade assintótica



# Abstract

For stationary sequences, under general dependence restrictions, if the sequence of point processes of upcrossings converges in distribution, then the limiting point process is necessarily a compound Poisson process, with intensity linked to a parameter called upcrossings index  $\eta$ ,  $0 \leq \eta \leq 1$ . This extremal coefficient can be viewed as a measure of the clustering of upcrossings of high levels by the variables of a stationary sequence and it provides different and complementary information to that provided by the key parameter in Extreme Value Theory, the extremal index  $\theta$ ,  $0 \leq \theta \leq 1$ .

In this thesis we evaluate the effect that subsampling has on the value of the upcrossings index  $\eta$  and exploring the asymptotic characterizations of this parameter, as well as its relation with the extremal index, we propose different estimating methods. Several properties of the proposed estimators are proved and applications to simulated and real data given.

## Keywords

Upcrossings index, local dependence conditions, subsampling, estimation, consistency and asymptotic normality.



# Conteúdo

Introdução	1
<b>1 Conceitos Fundamentais</b>	<b>5</b>
1.1 Introdução	5
1.2 Sucessões de variáveis aleatórias independentes e identicamente distribuídas	5
1.3 Sucessões estacionárias	9
1.3.1 O índice extremal: caracterização e estimação	14
1.3.1.1 Fórmulas de cálculo do índice extremal	16
1.3.1.2 Estimação do índice extremal	17
1.3.1.2.1 Método dos blocos	17
1.3.1.2.2 Método dos <i>runs</i>	20
1.3.1.2.3 Método do tempo entre excedências	21
1.3.2 O índice de cruzamentos: caracterização	22
1.3.2.1 Fórmulas de cálculo do índice de cruzamentos	23
<b>2 Índice de cruzamentos após subamostragem sistemática</b>	<b>27</b>
2.1 Introdução	27
2.2 Processo pontual de $T$ -cruzamentos	28
2.3 Cálculo do índice de $T$ -cruzamentos sob novas condições de dependência local	33
2.4 Cálculo do índice de cruzamentos após subamostragem sistemática	40
<b>3 O estimador de <i>runs</i> do índice de cruzamentos</b>	<b>51</b>
3.1 Introdução	51
3.2 O estimador de <i>runs</i> do índice de cruzamentos	52
3.3 Consistência fraca	55
3.4 Normalidade assintótica	62
3.5 A escolha dos níveis	68
3.6 Estudo de simulação: alguns exemplos	70
3.6.1 O processo ARMAX	71
3.6.2 O processo AR(1)	74

3.7	Método <i>Bootstrap</i> . . . . .	77
3.8	Aplicações a dados reais . . . . .	78
3.8.1	Dados meteorológicos: Ozono . . . . .	78
3.8.2	Dados financeiros: DAX30 . . . . .	80
3.8.3	Dados hídricos: Caudal efluente numa barragem . . . . .	82
<b>4</b>	<b>Estimadores de blocos do índice de cruzamentos</b>	<b>85</b>
4.1	Introdução . . . . .	85
4.2	O estimador de blocos $\hat{\eta}_n^B$ . . . . .	86
4.2.1	Consistência fraca . . . . .	87
4.2.2	Normalidade assintótica . . . . .	88
4.3	Os estimadores de blocos $\hat{\eta}_n^{dj}$ e $\hat{\eta}_n^{sl}$ . . . . .	89
4.3.1	Consistência e normalidade assintótica . . . . .	93
4.4	Estudo de simulação . . . . .	95
4.4.1	O número de blocos . . . . .	95
4.4.2	Decomposição do erro quadrático médio . . . . .	98
4.4.3	Comparações com o estimador de <i>runs</i> . . . . .	103
4.5	Aplicações a dados reais . . . . .	109
	<b>Conclusão</b>	<b>111</b>
	<b>Bibliografia</b>	<b>112</b>

# Lista de Figuras

1.1	Distribuições de valores extremos (Gumbel, Fréchet e Weibull) com $\alpha = 1$ . . . . .	7
2.1	Amostras simuladas de um modelo ARMAX com distribuição geradora de Fréchet unitária, para $k = 0.1$ (à esquerda), $k = 0.5$ (ao centro) e $k = 0.9$ (à direita). . . . .	38
2.2	Trajetórias do processo $AR(1)$ com $r = 2$ e $n = 100$ e do processo obtido por subamostragem simples com $T = 3$ (a negrito). . . . .	46
3.1	Trajetória do processo ARMAX definido em (3.35). . . . .	71
3.2	Estimativas de $\eta$ em função de $k$ , para uma amostra de dimensão $n = 5000$ do processo estacionário ARMAX em (3.35), numa escala linear (à esquerda) e numa escala logarítmica (à direita). . . . .	72
3.3	Valores médios estimados (à esquerda), erros quadráticos médios, variâncias e quadrado do viés estimados (à direita), para amostras de dimensão $n = 500$ do processo ARMAX em (3.35), numa escala logarítmica. ( $\eta = 0.5$ ). . . . .	73
3.4	Valores estimados do desvio padrão de $\hat{\eta}_n(k)$ obtidos pela simulação multi-amostra (linha a cheio) e valores de $\widehat{\sigma}_n(\hat{\eta}_n(k))$ (linha tracejada) para amostras de dimensão $n = 100, 500$ e $1000$ do processo ARMAX dado por (3.35). . . . .	74
3.5	Valores médios estimados de $\hat{\eta}_n(k)$ (linha a cheio) e intervalos de confiança a 95% obtidos a partir de (3.28) (linha tracejada) e as aproximações Monte Carlo dos intervalos de confiança a 95% (quantis 2.5% e 97.5%) (sombreado), para amostras de dimensão $n = 100, 500$ e $1000$ do processo ARMAX em (3.35). . . . .	74
3.6	Trajetória do processo estacionário AR(1) definido por (2.13), com $r = 2$ (à esquerda), $r = 3$ (ao centro) e $r = 5$ (à direita). . . . .	75
3.7	Estimativas de $\eta$ , em função de $k$ , para uma amostra de dimensão $n = 5000$ do processo estacionário AR(1) em (2.13), com $r = 2$ (à esquerda, $\eta = 0.75$ ), $r = 3$ (ao centro, $\eta = 0.89$ ) e $r = 5$ (à direita, $\eta = 0.96$ ), em escala logarítmica. . . . .	75
3.8	Valores médios estimados de $\hat{\eta}_n(k)$ (linha a cheio) e intervalos de confiança a 95% obtidos a partir de (3.28) (linha tracejada) e as aproximações Monte Carlo dos intervalos de confiança a 95% (quantis 2.5% e 97.5%) (sombreado), para amostras de dimensão $n = 500$ do processo AR(1) em (2.13), para $r = 2$ (à esquerda), $r = 3$ (ao centro) e $r = 5$ (à direita). . . . .	76
3.9	Erros quadráticos médios, variâncias e quadrado do viés estimados (em escala logarítmica) de $\hat{\eta}_n(k)$ , a partir de amostras de dimensão $n = 500$ do processo AR(1) em (2.13), para $r = 2$ (à esquerda), $r = 3$ (ao centro) e $r = 5$ (à direita). . . . .	77

3.10	Valores (em partes por milhão) de máximos semanais da concentração média, por hora, de ozono verificada na área da baía de São Francisco nos EUA. . . . .	79
3.11	Estimativas de $\eta$ para máximos semanais da concentração média, por hora, de ozono (em partes por milhão) na área da baía de São Francisco nos EUA, numa escala linear (à esquerda) e numa escala logarítmica (à direita). . . . .	79
3.12	Estimativas bootstrap ( $B = 5000$ ) de $\hat{\eta}$ (a cheio) e respetivos intervalos de confiança a 95% (a tracejado) em função de $k$ dos máximos semanais da concentração média, por hora, de ozono (em partes por milhão) na área da baía de São Francisco nos EUA, numa escala linear (à esquerda) e numa escala logarítmica (à direita). . . . .	80
3.13	Dados diários de fecho do índice DAX30 no período em análise (à esquerda) e respetivos <i>Log-returns</i> (à direita). . . . .	81
3.14	Estimativas de $\eta$ para os <i>log-returns</i> dos valores diários de fecho do índice DAX30 no período em análise, em função de $k$ , numa escala linear (à esquerda) e numa escala logarítmica (à direita). . . . .	82
3.15	Estimativas bootstrap ( $B = 5000$ ) de $\eta$ (a cheio) e respetivos intervalos de confiança a 95% (a tracejado) para os <i>log-returns</i> dos valores diários de fecho do índice DAX30 no período em análise, em função de $k$ , numa escala linear (à esquerda) e numa escala logarítmica (à direita). . . . .	82
3.16	Valores do caudal efluente médio diário (em $m^3/s$ ), na barragem do Fratel nos meses de novembro a março, entre 01-01-1984 e 30-11-2008. . . . .	83
3.17	Estimativas de $\eta$ para o caudal efluente médio diário (em $m^3/s$ ), na barragem do Fratel no período em análise, numa escala linear (à esquerda) e numa escala logarítmica (à direita). . . . .	84
3.18	Estimativas bootstrap ( $B = 5000$ ) de $\eta$ (a cheio) e respetivos intervalos de confiança a 95% (a tracejado) do caudal efluente médio diário (em $m^3/s$ ), na barragem do Fratel no período em análise, numa escala linear (à esquerda) e numa escala logarítmica (à direita). . . . .	84
4.1	Estimativas de $\eta$ obtidas com $\hat{\eta}_n^B(k)$ (à esquerda), $\hat{\eta}_n^{dj}(k)$ (ao centro) e $\hat{\eta}_n^{sl}(k)$ (à direita) para o processo ARMAX, para vários valores de $r_n$ e para $n = 50000$ , numa escala logarítmica ( $\eta = 0.5$ ) . . . . .	95
4.2	Estimativas de $\eta$ obtidas com $\hat{\eta}_n^B(k)$ (à esquerda), $\hat{\eta}_n^{dj}(k)$ (ao centro) e $\hat{\eta}_n^{sl}(k)$ (à direita) para o processo AR(1), para vários valores de $r_n$ e com $n = 50000$ e $r = 2$ , numa escala logarítmica ( $\eta = 0.75$ ) . . . . .	96
4.3	Estimativas de $\eta$ obtidas com $\hat{\eta}_n^B(k)$ (à esquerda), $\hat{\eta}_n^{dj}(k)$ (ao centro) e $\hat{\eta}_n^{sl}(k)$ (à direita) para o processo AR(1), para vários valores de $r_n$ e com $n = 50000$ e $r = 3$ , numa escala logarítmica ( $\eta = 0.89$ ) . . . . .	96

4.4	Estimativas de $\eta$ obtidas com $\hat{\eta}_n^B(k)$ (à esquerda), $\hat{\eta}_n^{dj}(k)$ (ao centro) e $\hat{\eta}_n^{sl}(k)$ (à direita) para o processo AR(1), para vários valores de $r_n$ e com $n = 50000$ e $r = 5$ , numa escala logarítmica ( $\eta = 0.96$ ) . . . . .	96
4.5	Erro quadrático médio estimado de $\hat{\eta}_n^B(k)$ (à esquerda), $\hat{\eta}_n^{dj}(k)$ (ao centro) e $\hat{\eta}_n^{sl}(k)$ (à direita) para o processo ARMAX, para vários valores de $r_n$ e com $n = 500$ , numa escala logarítmica ( $\eta = 0.5$ ) . . . . .	97
4.6	Erro quadrático médio estimado de $\hat{\eta}_n^B(k)$ (à esquerda), $\hat{\eta}_n^{dj}(k)$ (ao centro) e $\hat{\eta}_n^{sl}(k)$ (à direita) para o processo AR(1), para vários valores de $r_n$ e com $n = 500$ e $r = 2$ , numa escala logarítmica ( $\eta = 0.75$ ) . . . . .	97
4.7	Erro quadrático médios estimado de $\hat{\eta}_n^B(k)$ (à esquerda), $\hat{\eta}_n^{dj}(k)$ (ao centro) e $\hat{\eta}_n^{sl}(k)$ (à direita) para o processo AR(1), para vários valores de $r_n$ e com $n = 500$ e $r = 3$ , numa escala logarítmica ( $\eta = 0.89$ ) . . . . .	97
4.8	Erro quadrático médio estimado de $\hat{\eta}_n^B(k)$ (à esquerda), $\hat{\eta}_n^{dj}(k)$ (ao centro) e $\hat{\eta}_n^{sl}(k)$ (à direita) para o processo AR(1), para vários valores de $r_n$ e com $n = 500$ e $r = 5$ , numa escala logarítmica ( $\eta = 0.96$ ) . . . . .	98
4.9	Erro quadrático médio estimado, variância estimada e quadrado do viés estimado de $\hat{\eta}_n^B(k)$ (cima à esquerda), $\hat{\eta}_n^{dj}(k)$ (cima à direita), $\hat{\eta}_n^{sl}(k)$ (baixo à esquerda) e $\hat{\eta}_n^{FS}(k)$ (baixo à direita) para o processo ARMAX, com $n = 500$ , numa escala logarítmica ( $\eta = 0.5$ ) . . . . .	99
4.10	Erro quadrático médio estimado, variância estimada e quadrado do viés estimado de $\hat{\eta}_n^B(k)$ (cima à esquerda), $\hat{\eta}_n^{dj}(k)$ (cima à direita), $\hat{\eta}_n^{sl}(k)$ (baixo à esquerda) e $\hat{\eta}_n^{FS}(k)$ (baixo à direita) para o processo AR(1), com $n = 500$ e $r = 2$ , numa escala logarítmica ( $\eta = 0.75$ ) . . . . .	100
4.11	Erro quadrático médio estimado, variância estimada e quadrado do viés estimado de $\hat{\eta}_n^B(k)$ (cima à esquerda), $\hat{\eta}_n^{dj}(k)$ (cima à direita), $\hat{\eta}_n^{sl}(k)$ (baixo à esquerda) e $\hat{\eta}_n^{FS}(k)$ (baixo à direita) para o processo AR(1), com $n = 500$ e $r = 3$ , numa escala logarítmica ( $\eta = 0.89$ ) . . . . .	101
4.12	Erro quadrático médio estimado, variância estimada e quadrado do viés estimado de $\hat{\eta}_n^B(k)$ (cima à esquerda), $\hat{\eta}_n^{dj}(k)$ (cima à direita), $\hat{\eta}_n^{sl}(k)$ (baixo à esquerda) e $\hat{\eta}_n^{FS}(k)$ (baixo à direita) para o processo AR(1), com $n = 500$ e $r = 5$ , numa escala logarítmica ( $\eta = 0.96$ ) . . . . .	102
4.13	Valores médios estimados dos estimadores em estudo para o processo ARMAX, com $n = 500$ , numa escala logarítmica. ( $\eta = 0.5$ ) . . . . .	104
4.14	Comparação dos erros quadráticos médios estimados (à esquerda) e das variâncias estimadas (à direita) dos estimadores em estudo para o processo ARMAX com $n = 500$ , numa escala logarítmica. . . . .	104
4.15	Valores médios estimados dos estimadores em estudo para o processo AR(1), com $n = 500$ e $r = 2$ , numa escala logarítmica. ( $\eta = 0.75$ ) . . . . .	105

4.16	Comparação dos erros quadráticos médios estimados (à esquerda) e das variâncias estimadas (à direita) dos estimadores em estudo para o processo AR(1), com $n = 500$ e $r = 2$ , numa escala logarítmica. . . . .	105
4.17	Valores médios estimados dos estimadores em estudo para o processo AR(1), com $n = 500$ e $r = 3$ , numa escala logarítmica. ( $\eta = 0.89$ ) . . . . .	105
4.18	Comparação dos erros quadráticos médios estimados (à esquerda) e das variâncias estimadas (à direita) dos estimadores em estudo para o processo AR(1), com $n = 500$ e $r = 3$ , numa escala logarítmica. . . . .	106
4.19	Valores médios estimados dos estimadores em estudo para o processo AR(1), com $n = 500$ e $r = 5$ , numa escala logarítmica. ( $\eta = 0.96$ ) . . . . .	106
4.20	Comparação dos erros quadráticos médios estimados (à esquerda) e das variâncias estimadas (à direita) dos estimadores em estudo para o processo AR(1), com $n = 500$ e $r = 5$ , numa escala logarítmica. . . . .	106
4.21	Estimativas de $\eta$ obtidas com $\hat{\eta}_n^B(k)$ , $\hat{\eta}_n^{dj}(k)$ , $\hat{\eta}_n^{sl}(k)$ , $\hat{\eta}_n^{FS}(k)$ e $\hat{\eta}_n^R(k)$ para os <i>log-returns</i> dos valores diários de fecho do índice DAX30 no período compreendido entre 03-01-2000 e 29-06-2012. . . . .	109
4.22	Estimativas de $\eta$ obtidas com $\hat{\eta}_n^B(k)$ , $\hat{\eta}_n^{dj}(k)$ , $\hat{\eta}_n^{sl}(k)$ , $\hat{\eta}_n^{FS}(k)$ e $\hat{\eta}_n^R(k)$ do caudal efluente médio diário (em $m^3/s$ ), na barragem do Fratel nos meses de novembro a março, entre 01-01-1984 e 30-11-2008. . . . .	110

# Lista de Tabelas

2.1	Valores de $\eta^{(T)}$ do processo ARMAX com distribuição geradora de Fréchet unitária para alguns valores de $k$ e $T$ . . . . .	38
3.1	Fração de amostragem ótima, valores médios, erros quadráticos médios e desvio padrão nos seus níveis ótimos, do processo ARMAX, com $\eta = 0.5$ . . . . .	72
3.2	Fração ótima de amostragem, valores médios, erros quadráticos médios e desvios padrão do estimador nos seus níveis ótimos, para o processo AR(1) em (2.13), com $\eta = 0.75$ , $\eta = 0.89$ e $\eta = 0.96$ . . . . .	76
4.1	Medidas de eficiência relativa dos estimadores em análise do processo ARMAX dado por (3.35), com $n = 500$ . . . . .	107
4.2	$REFF_0^{I/J}$ dos estimadores em estudo para o processo AR(1) dado por (2.13), com $n = 500$ . . . . .	107
4.3	Indicador de redução do viés dos estimadores em análise do processo ARMAX dado por (3.35), com $n = 500$ . . . . .	108
4.4	$BRJ_0^{I/J}$ dos estimadores em estudo para o processo AR(1) dado por (2.13), com $n = 500$ . . . . .	108



# Introdução

A Teoria de Valores Extremos emergiu da necessidade de caracterizar o comportamento de observações extremas, em particular caracterizar a distribuição do máximo das  $n$  primeiras variáveis de uma sucessão  $\mathbf{X} = \{X_n\}_{n \geq 1}$ , de variáveis aleatórias independentes e identicamente distribuídas (iid). Dado que o conhecido Teorema Limite Central de Lindberg-Lévy não respondia a esta necessidade, Fréchet (1927), Fisher e Tippett (1928) e Gnedenko (1943) provam que, independentemente da distribuição das variáveis que compõem a sucessão  $\mathbf{X}$ , a classe de possíveis leis limite para o máximo, linearmente normalizado, das  $n$  primeiras variáveis de  $\mathbf{X}$ , caso exista, se restringe a três tipos distintos: Fréchet, Gumbel ou Weibull.

O estudo de observações extremas em áreas como a hidrologia, as finanças ou mesmo a engenharia cedo fizeram perceber que a hipótese de independência entre as variáveis da sucessão seria demasiado restritiva na modelação de dados reais. A investigação nesta área procurou assim estender de alguma forma os resultados já conhecidos, para o caso iid, a sucessões que apresentam dependência entre as variáveis, ou até mesmo, a sucessões em que as variáveis não apresentam a mesma distribuição.

A introdução de condições convenientes de dependência local e assintótica foi fundamental para a aplicação da Teoria de Valores Extremos desenvolvida no contexto iid, como o Teorema dos Tipos Extremos, a sucessões dependentes, em particular a sucessões estacionárias. As condições utilizadas com maior frequência são as condições de independência assintótica  $\Delta(\mathbf{u})$  de Hsing *et al.* (1988) e  $D(\mathbf{u})$  de Leadbetter (1974) e de dependência local  $D'(\mathbf{u})$  de Leadbetter (1974),  $D''(\mathbf{u})$  de Leadbetter e Nandagopalan (1989) e  $D^{(k)}(\mathbf{u})$ ,  $k \geq 1$ , de Chernick *et al.* (1991).

Nesta passagem do contexto iid para o contexto estacionário encontramos um dos parâmetros mais relevantes nesta teoria, o índice extremal,  $\theta \in [0, 1]$ . Este parâmetro definido por Leadbetter (1983), permite relacionar a distribuição limite do máximo das  $n$  primeiras variáveis de uma sucessão estacionária, com a distribuição limite que obteríamos se a sucessão fosse iid, o que justifica a sua extrema importância.

As condições de dependência local referidas acima permitem obter fórmulas simples para o cálculo do índice extremal a partir de um número finito de variáveis consecutivas da sucessão. Estas fórmulas estão na base de muitos dos métodos propostos para estimar este coeficiente extremal, sendo os principais o método dos blocos (e.g. Hsing (1991), Weissman e Novak (1998), Robert (2009), Robert *et al.* (2009), Drees (2011)), o método dos *runs* (e.g. Nandagopalan (1990), Hsing (1993), Weissman e Novak (1998), Gomes *et al.* (2008)) e o método do tempo entre excedências (e.g. Ferro e Segers (2003)), Süveges (2007), Süveges e Davison (2010)).

Excedências de níveis elevados são sem dúvida os acontecimentos mais estudados na literatura de extremos, sendo o índice extremal uma medida do agrupamento desses acontecimentos. Podem, no entanto, ocorrer situações adversas com outros tipos de acontecimentos extremos, como cruzamentos de níveis elevados. H. Ferreira (2006) define um novo coeficiente extremal, o índice de cruzamentos,  $\eta \in [0, 1]$ , que é uma medida do agrupamento de cruzamentos ascen-

dentos de um nível elevado pelas variáveis de uma sucessão estacionária  $\mathbf{X}$ . Este parâmetro permite relacionar o número médio limite de excedências,  $\tau$ , com o número médio limite de cruzamentos ascendentes,  $\nu$ , de um mesmo nível, através da relação  $\eta = \frac{\tau}{\nu} \theta$ , onde  $\theta$  é o índice extremal de  $\mathbf{X}$ . Sob a validade de novas condições de dependência local H. Ferreira (2006, 2007) obtém fórmulas simples de cálculo deste parâmetro.

Robinson e Tawn (2000) concluíram que a frequência amostral influencia o valor do índice extremal ao provarem que o índice extremal de uma sucessão subamostrada a partir de uma sucessão estacionária inicial  $\mathbf{X}$ , em intervalos fixos de  $T$ -unidades, *i.e.*, de uma subsucessão  $\mathbf{Y} = \{Y_n = X_{1+(n-1)T}\}_{n \geq 1}$ ,  $T \geq 1$ , é sempre superior ou igual ao da sucessão inicial. Será assim de esperar que conclusões semelhantes possam ser retiradas para o índice de cruzamentos  $\eta$ , até porque estes dois coeficientes extremos se relacionam entre si. A confirmação deste facto é dada no segundo capítulo onde, seguindo de perto a estrutura de Sebastião *et al.* (2010), analisamos o efeito que a subamostragem sistemática de passo constante  $T \geq 1$ , considerada em Robinson e Tawn (2000), tem no índice de cruzamentos.

Provamos que a intensidade do processo pontual limite do processo pontual de  $T$ -cruzamentos é caracterizada por um parâmetro  $\eta^{(T)} \in [0, 1]$ , a que chamamos índice de  $T$ -cruzamentos e introduzimos condições de dependência local que permitem obter fórmulas simples de cálculo deste parâmetro. Mostramos de que forma os índices de  $T$ -cruzamentos se relacionam entre si e uma vez que  $\eta := \eta^{(1)}$  generalizamos naturalmente a relação existente entre o índice extremal e o índice de cruzamentos. Assumindo que a sucessão estacionária  $\mathbf{X}$  verifica determinadas condições de dependência, provamos que a sucessão subamostrada  $\mathbf{Y}$  verifica condições análogas e calculamos o seu índice de cruzamentos  $\eta^{(1, \mathbf{Y})}$  a partir da relação com o índice de  $T$ -cruzamentos de  $\mathbf{X}$ ,  $\eta^{(T, \mathbf{X})}$ .

A relação entre o índice de cruzamentos e o índice extremal permite-nos estimar o índice de cruzamentos a partir dos conhecidos estimadores do índice extremal e de estimadores de  $\tau$ , número médio de excedências e  $\nu$  número médio de cruzamentos. Contudo, conseguimos propor, a partir das caracterizações assintóticas do índice de cruzamentos, estimadores com melhores propriedades assintóticas. Debruçamos-nos sobre este assunto no terceiro capítulo, onde propomos um estimador não paramétrico relacionado com *runs* de cruzamentos e por conseguinte será denominado estimador de *runs* do índice de cruzamentos. Este estimador surge naturalmente a partir dos resultados do capítulo anterior, uma vez que, sob a validade de determinadas condições de dependência, a distribuição condicional do tamanho dos *runs* de cruzamentos, do processo autoregressivo considerado, converge para uma distribuição geométrica de valor esperado coincidente com o inverso do índice de cruzamentos. Torna-se assim natural estimar o índice de cruzamentos através do quociente entre o número de *runs* de cruzamentos e o número total de cruzamentos. Provamos a consistência e normalidade assintótica deste estimador, inicialmente sugerido por H. Ferreira (2007). Analisamos o seu comportamento em diversas amostras de dimensão finita, simuladas a partir de processos para os quais o índice de cruzamentos é bem conhecido, bem como em conjuntos de dados reais. A maioria dos resultados deste capítulo são apresentados em Sebastião *et al.* (2012).

No último capítulo propomos um método de estimação do índice de cruzamentos baseado no facto de, mediante certas condições, o índice de cruzamentos ser aproximadamente o inverso da dimensão média dos grupos de cruzamentos. À definição de grupo de cruzamentos está associada uma partição da amostra  $X_1, \dots, X_n$  em blocos com uma certa dimensão  $r_n$  conveniente, e o número de pontos em cada grupo de cruzamentos será o número de cruzamentos de um nível elevado em cada um desses blocos. Define-se assim um estimador de blocos do índice de cruzamentos. Estudamos as propriedades assintóticas deste estimador. Neste capítulo ainda propomos outros estimadores baseados na metodologia dos blocos e procedemos à comparação, via técnicas de Monte Carlo, do conjunto de estimadores propostos.

Terminamos com algumas considerações finais onde fazemos uma síntese da investigação realizada e destacamos alguns pontos inacabados que poderão ser objeto de trabalho futuro.



# Capítulo 1

## Conceitos Fundamentais

### 1.1 Introdução

Com o objetivo de tornar mais clara a leitura dos capítulos subsequentes, apresentamos a seguir algumas definições e resultados que serão relevantes ao longo deste trabalho e que referiremos sempre que julgarmos oportuno. Aproveitamos para introduzir notações que serão usadas ao longo do mesmo.

Da Teoria Clássica de Valores Extremos salientamos o Teorema dos Tipos Extremos ou do Limite Extremal que caracteriza a classe das possíveis leis limite do máximo, linearmente normalizado, de  $n$  variáveis aleatórias independentes e identicamente distribuídas.

A maioria dos resultados que apresentamos referem-se à situação de não independência, mais concretamente ao estudo de extremos de sucessões estacionárias onde se insere este trabalho. Uma grande diferença que surge no estudo deste tipo de sucessões, decorre do facto de ser natural encontrarmos grupos de excedências ou de cruzamentos de níveis elevados, em vez de valores isolados como no caso iid. Neste contexto, o tamanho médio dos grupos de excedências de um nível elevado é assintoticamente caracterizado por um parâmetro denominado índice extremal, que por sua vez, se relaciona, entre outros, com o índice de cruzamentos (ascendentes) que caracteriza o tamanho médio dos grupos destes cruzamentos.

Fazemos a revisão de algumas condições de dependência que permitem, por um lado obter a distribuição limite do máximo de sucessões estacionárias, e por outro, calcular o valor do índice extremal e do índice de cruzamentos. Apresentamos ainda alguns dos estimadores conhecidos para o índice extremal.

Omitimos as demonstrações dos resultados referenciados neste capítulo por constarem nas referências bibliográficas que apresentamos. Os assuntos aqui abordados poderão ser aprofundados nas seguintes obras de referência da Teoria de Valores Extremos: Gumbel (1958), Leadbetter *et al.* (1983), Tiago de Oliveira (1984), Embrechts *et al.* (1997), Coles (2001), Beirlant *et al.* (2004), Haan e Ferreira (2006) e Reiss e Thomas (2007).

### 1.2 Sucessões de variáveis aleatórias independentes e identicamente distribuídas

Seja  $\mathbf{X} = \{X_n\}_{n \geq 1}$  uma sucessão de variáveis aleatórias reais iid com função de distribuição (f.d.) comum  $F$  e denotemos por  $M_n = \max\{X_1, X_2, \dots, X_n\}$  o máximo das  $n$  primeiras variá-

veis desta sucessão.

A Teoria Clássica de Valores Extremos começou por querer dar resposta à questão de, se para  $M_n$  sujeito a uma normalização linear se teria como limite uma f.d. não degenerada, mais concretamente, se

$$P(a_n^{-1}(M_n - b_n) \leq x) = F^n(a_n x + b_n) \xrightarrow{n \rightarrow +\infty} G(x) \quad (1.1)$$

para todo o ponto de continuidade,  $x$ , da função de distribuição  $G$ , com  $G$  alguma f.d. não degenerada e  $\{a_n > 0\}_{n \geq 1}$  e  $\{b_n\}_{n \geq 1}$  sucessões de constantes reais. Esta questão fazia sentido uma vez que para somas existia tal convergência, garantida pelo Teorema do Limite Central.

O problema poderia colocar-se igualmente para o mínimo, mas tendo em conta que  $\min\{X_1, X_2, \dots, X_n\} = -\max\{-X_1, -X_2, \dots, -X_n\}$  os resultados para o mínimo obtêm-se facilmente a partir dos do máximo,  $M_n$ .

Quando (1.1) se verifica dizemos que  $F$  pertence ao domínio de atração, neste caso para máximos, de  $G$ , simbolicamente  $F \in D(G)$ , e designamos as constantes  $a_n > 0$  e  $b_n, n \geq 1$ , por constantes de atração de  $F$  para  $G$ , ou constantes de normalização.

A classe das distribuições a que  $G$  pertence, ou equivalentemente, o domínio de atração a que pertence a f.d.  $F$  comum às variáveis da sucessão, é então estabelecida no clássico Teorema dos Tipos Extremos devido a Fréchet (1927), Fisher e Tippett (1928) e formalmente demonstrado por Gnedenko (1943), que a seguir recordamos.

**Teorema 1.2.1** *Se a sucessão  $X$  de variáveis aleatórias iid verifica (1.1) para algumas constantes reais  $a_n > 0$  e  $b_n, n \geq 1$ , então  $G$  é do tipo de uma das três seguintes distribuições:*

$$\text{Tipo I (Gumbel)} \quad H(x) \equiv \Lambda(x) = \exp(-e^{-x}), \quad x \in \mathbb{R};$$

$$\text{Tipo II (Fréchet)} \quad H(x) \equiv \Phi_\alpha(x) = \exp(-x^{-\alpha}), \quad x > 0, \quad \text{para algum } \alpha > 0;$$

$$\text{Tipo III (Weibull)} \quad H(x) \equiv \Psi_\alpha(x) = \exp(-(-x)^\alpha), \quad x \leq 0, \quad \text{para algum } \alpha > 0.$$

Reciprocamente, qualquer f.d. da forma anterior pode aparecer como limite em (1.1), o que de facto acontece quando  $G$  é, ela própria, a f.d. comum às variáveis da sucessão  $X$ .

Dizer que a f.d. não degenerada  $G$ , é do tipo de uma das três distribuições de valores extremos (Gumbel, Fréchet ou Weibull) significa dizer que existem constantes reais  $a > 0$  e  $b$  tais que  $G(x) = H(ax + b)$ .

As três distribuições de valores extremos apresentam comportamentos na cauda muito diferentes, como podemos observar na Figura 1.1.

A classe de leis limite para o máximo de sucessões iid apresentada no Teorema 1.2.1 coincide com a classe das distribuições max-estáveis, isto é, funções de distribuição  $G$ , tais que

$$G^n(A_n x + B_n) = G(x)$$

para algumas constantes reais  $A_n > 0$  e  $B_n, n \geq 1$ . Esta importante caracterização das leis

limite do máximo de sucessões iid é consequência de (1.1) e de um lema relativo à convergência de funções de distribuição, conhecido como Lema dos Tipos de Khintchine.

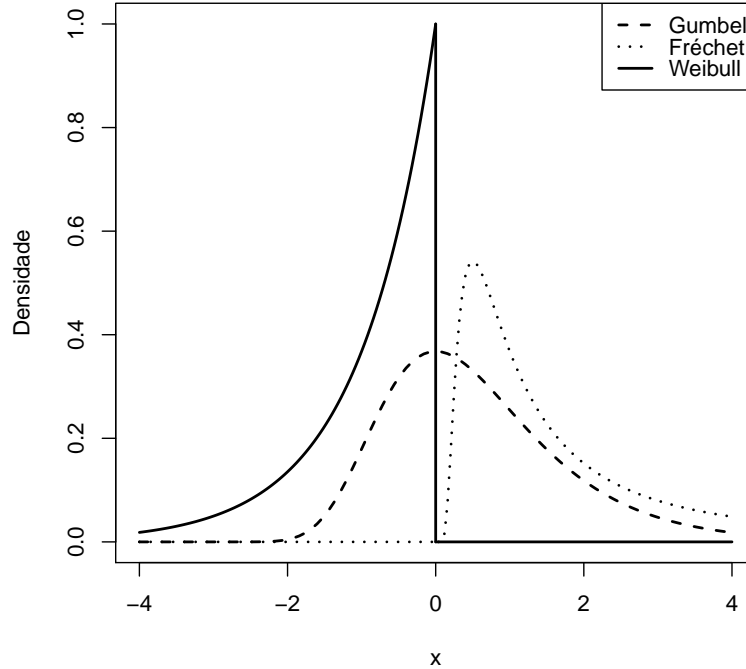


Figura 1.1: Distribuições de valores extremos (Gumbel, Fréchet e Weibull) com  $\alpha = 1$ .

**Lema 1.2.1** *Sejam  $\{F_n\}_{n \geq 1}$  uma sucessão de funções de distribuição,  $G$  e  $G^*$  funções de distribuição não degeneradas. Se existirem constantes  $a_n > 0$ ,  $b_n$ ,  $\alpha_n > 0$ ,  $\beta_n$ ,  $n \geq 1$ , para as quais*

$$F_n(a_n x + b_n) \xrightarrow[n \rightarrow +\infty]{w} G(x) \quad \text{e} \quad F_n(\alpha_n x + \beta_n) \xrightarrow[n \rightarrow +\infty]{w} G^*(x)$$

então

$$\frac{\alpha_n}{a_n} \xrightarrow[n \rightarrow +\infty]{} a > 0, \quad \frac{\beta_n - b_n}{a_n} \xrightarrow[n \rightarrow +\infty]{} b \quad \text{e} \quad G(ax + b) = G^*(x).$$

Este lema tem também um papel preponderante na extensão do Teorema dos Tipos Extremos, Teorema 1.2.1, às sucessões dependentes.

Conhecendo o domínio de atração da distribuição  $F$ , comum às variáveis da sucessão  $\mathbf{X}$ , podemos usar a distribuição de valores extremos adequada para modelar a distribuição do máximo  $M_n$ , tendo em conta o Teorema 1.2.1. Contudo, o conhecimento do domínio de atração de  $F$  não é obrigatório, uma vez que as famílias Gumbel, Fréchet e Weibull podem ser unificadas numa única família paramétrica de distribuições, denotada pela sigla GEV (do inglês Generalized Extreme Value), generalizada de valores extremos (von Mises (1936)). A família GEV tem

funções de distribuição da forma

$$GEV_{\xi,\mu,\sigma}(x) = \exp \left\{ - \left[ 1 + \xi \left( \frac{x - \mu}{\sigma} \right) \right]_+^{-\frac{1}{\xi}} \right\}, \quad (1.2)$$

definidas no conjunto  $\{x : 1 + \xi(x - \mu)/\sigma > 0\}$ , onde  $\mu \in \mathbb{R}$  é o parâmetro de localização,  $\sigma > 0$  é o parâmetro de escala,  $\xi \in \mathbb{R}$  é o parâmetro de forma e  $x_+ = \max\{0, x\}$ .

O parâmetro  $\xi$ , designado índice de cauda, determina o comportamento da cauda da distribuição (ver Figura 1.1): se  $\xi$  é positivo então a cauda é pesada (Fréchet), se  $\xi = 0$  (tomando  $\lim_{\xi \rightarrow 0}$  em (1.2)) então a cauda é leve (Gumbel) e se  $\xi < 0$  então a cauda é limitada (Weibull).

Conhecida a forma da distribuição limite do máximo  $M_n$ , é importante saber em que condições há convergência para esse limite. O resultado apresentado a seguir dá-nos essas condições e a sua extensão às sucessões estacionárias assumirá grande importância como veremos na próxima secção.

**Teorema 1.2.2** *Seja  $\mathbf{X}$  uma sucessão de variáveis aleatórias iid com função de distribuição comum  $F$  e  $\mathbf{u} = \{u_n\}_{n \geq 1}$  uma sucessão de níveis reais. Então, para  $0 \leq \tau \leq +\infty$ , tem-se*

$$P(M_n \leq u_n) \xrightarrow{n \rightarrow +\infty} e^{-\tau} \quad (1.3)$$

se e só se

$$n(1 - F(u_n)) \xrightarrow{n \rightarrow +\infty} \tau. \quad (1.4)$$

Notemos que (1.1) é um caso particular de (1.3) quando consideramos  $\tau = \tau(x) = -\log G(x)$  e  $u_n = u_n(x) = a_n x + b_n$ . Consequentemente, uma condição necessária e suficiente para  $F \in D(G)$  é,

$$n[1 - F(u_n(x))] \xrightarrow{n \rightarrow +\infty} \tau = \tau(x) = -\log G(x)$$

para todo o  $x$  e algumas constantes reais  $a_n > 0$  e  $b_n, n \geq 1$ .

Se considerarmos que a sucessão  $\mathbf{X}$  tem uma excedência do nível  $u_n$  no instante  $i$  quando o acontecimento  $\{X_i > u_n\}$  ocorre, temos que a probabilidade de ocorrência de uma excedência é  $1 - F(u_n)$  e, por conseguinte, o número médio de excedências do nível  $u_n$  pelas  $n$  primeiras variáveis da sucessão  $\mathbf{X}$  é dado por  $n[1 - F(u_n)]$ . O nível  $u_n$  diz-se normalizado para excedências se o número médio de excedências for assintoticamente constante, isto é, quando (1.4) se verificar para algum  $\tau \geq 0$ . Neste caso denotaremos  $u_n$  por  $u_n^{(\tau)}$ .

Existindo uma sucessão de níveis normalizados para  $\mathbf{X}$ ,  $\mathbf{u}^{(\tau)} = \{u_n^{(\tau)}\}_{n \geq 1}$  para algum  $\tau > 0$ , então existe para todo  $\tau > 0$ . Mais ainda, toda a f.d. pertencente a algum dos três domínios de atração clássicos verifica (1.4), contudo o recíproco é falso. Salientamos ainda que, quando  $n \rightarrow +\infty$ ,  $u_n^{(\tau)}$  tende para  $x_F = \sup\{x : F(x) < 1\}$ , o limite superior do suporte de  $F$ .

Nas últimas décadas, o estudo da distribuição limite do processo pontual das excedências

de níveis reais  $\mathbf{u} = \{u_n\}_{n \geq 1}$  pelas variáveis de uma sucessão  $\mathbf{X}$ ,

$$N_n(u_n)(B) = \sum_{i=1}^n \mathbb{I}_{\{X_i > u_n\}} \delta_{\frac{i}{n}}(B), \quad B \subseteq \mathcal{B}([0, 1]) \quad (1.5)$$

onde  $\delta_a(B)$  é a medida de Dirac do conjunto  $B$  induzida pelo ponto  $a$  e  $\mathcal{B}([0, 1])$  representa a tribo de Borel sobre  $[0, 1]$ , adquiriu um estatuto especial na Teoria de Valores Extremos.

Esta importância é justificada pela relação

$$\left\{ \frac{M_n^{(k)} - b_n}{a_n} \leq x \right\} = \{N_n(u_n(x))([0, 1]) \leq k - 1\}$$

onde  $u_n(x) = a_n x + b_n$  e  $M_n^{(k)}$  denota a  $k$ -ésima estatística ordinal (e.o.) superior associada a  $X_1, X_2, \dots, X_n$ . De facto, se conhecermos a distribuição assintótica de  $N_n(u_n(x))$  dispomos da lei limite da  $k$ -ésima e.o. superior linearmente normalizada.

Para níveis normalizados para excedências,  $u_n^{(\tau)}$ , tem-se que a sucessão dos processos pontuais das excedências desses níveis pelas variáveis de uma sucessão  $\mathbf{X}$  de variáveis aleatórias iid converge em distribuição para um processo de Poisson homogéneo sobre  $[0, 1]$  de intensidade  $\tau$ .

A utilização de sucessões iid na modelação de dados reais é, como já referimos, inadequada em inúmeras situações. Basta pensarmos que, por exemplo, os dados climatéricos são geralmente séries temporais para as quais existe dependência temporal. Tendo em conta esta realidade, é natural que a investigação na área de valores extremos tenha procurado abranger situações de não independência e/ou não igualdade da distribuição das variáveis da sucessão. A maioria dos resultados que se encontram na literatura referem-se à situação de não independência, mais concretamente ao estudo dos extremos de sucessões estacionárias.

### 1.3 Sucessões estacionárias

Dizemos que a sucessão de variáveis aleatórias reais  $\mathbf{X} = \{X_n\}_{n \geq 1}$  é (estritamente) estacionária se para cada escolha de índices,  $1 \leq i_1 < \dots < i_n$  e todo o inteiro  $m \geq 1$ , os vetores aleatórios  $(X_{i_1}, \dots, X_{i_n})$  e  $(X_{i_1+m}, \dots, X_{i_n+m})$  têm a mesma distribuição.

No estudo dos valores extremos de sucessões estacionárias é usual considerar certo tipo de condições de dependência que garantem que a distribuição limite do máximo é ainda, neste caso, do tipo max-estável. De uma maneira geral admite-se que à medida que  $|i - j|$  cresce, a dependência entre  $X_i$  e  $X_j$  tende a desaparecer. O exemplo mais simples deste tipo de restrições é a  $m$ -dependência que garante que, desde que  $|i - j| > m$ ,  $X_i$  e  $X_j$  são variáveis aleatórias independentes.

Outras condições de dependência fraca, também designadas por condições de independência assintótica, têm sido consideradas por forma a estender os resultados conhecidos para sucessões iid às sucessões estacionárias. Destacamos aqui as condições  $\Delta(\mathbf{u})$  de Hsing *et al.* (1988) e  $D(\mathbf{u})$  de Leadbetter (1974).

**Definição 1.3.1** A sucessão  $\mathbf{X}$  verifica a condição  $\Delta(\mathbf{u})$ , para uma dada sucessão de reais  $\mathbf{u} = \{u_n\}_{n \geq 1}$ , se

$$\alpha_{n,l} = \sup \{ |P(A \cap B) - P(A)P(B)| : A \in \mathcal{B}_1^k(u_n), B \in \mathcal{B}_{k+l}^n(u_n), 1 \leq k \leq n-l \}, \quad (1.6)$$

onde  $B_i^j(u_n)$  representa a tribo gerada pelos acontecimentos  $\{X_s \leq u_n\}$ ,  $i \leq s \leq j$ , é tal que  $\alpha_{n,l_n} \xrightarrow{n \rightarrow +\infty} 0$ , para alguma sucessão  $l_n = o(n)$ .

Hsing *et al.* (1988), recorrendo ao lema que a seguir apresentamos, provam que sob a validade de  $\Delta(\mathbf{u})$  se a sucessão de processos pontuais de excedências  $\{N_n(u_n)\}_{n \geq 1}$ , gerada por uma sucessão estacionária  $\mathbf{X}$ , converge então o processo pontual limite é um processo de Poisson composto sobre  $[0, 1]$ .

**Lema 1.3.1** Seja  $\mathbf{X}$  uma sucessão estacionária que verifica a condição  $\Delta(\mathbf{u})$  para sucessões  $\mathbf{k} = \{k_n\}_{n \geq 1}$ ,  $\mathbf{l} = \{l_n\}_{n \geq 1}$  e  $\mathbf{u} = \{u_n\}_{n \geq 1}$  tais que

$$k_n \xrightarrow{n \rightarrow +\infty} +\infty, \quad \frac{k_n l_n}{n} \xrightarrow{n \rightarrow +\infty} 0, \quad k_n \alpha_{n,l_n} \xrightarrow{n \rightarrow +\infty} 0. \quad (1.7)$$

Para cada  $n$ , sejam  $J_{n,i}$ ,  $1 \leq i \leq k_n$ , subconjuntos disjuntos de inteiros em  $\{1, 2, \dots, n\}$ , tais que  $\# \left( \bigcup_{i=1}^{k_n} J_{n,i} \right) \sim n$  e  $\chi_{n,j} = \mathbb{I}_{\{X_j > u_n\}}$ .

Então, para qualquer sucessão de constantes não negativas  $\{a_n\}_{n \geq 1}$  e qualquer função,  $f$  definida sobre  $[0, 1]$ , não negativa, contínua ou em escada  $f$  em  $[0, 1]$ , tem-se que,

$$E \left[ \exp \left\{ -a_n \sum_{m=1}^n f(m/n) \chi_{n,m} \right\} \right] - \prod_{j=1}^{k_n} E \left[ \exp \left\{ -a_n \sum_{m \in J_{n,i}} f(m/n) \chi_{n,m} \right\} \right] \xrightarrow{n \rightarrow +\infty} 0.$$

**Teorema 1.3.1** Suponhamos que a condição  $\Delta(\mathbf{u})$  se verifica para  $\mathbf{X}$  e  $\{N_n(u_n)\}_{n \geq 1}$  converge em distribuição para algum processo pontual  $N$ . Então  $N$  é necessariamente um processo de Poisson composto com transformada de Laplace

$$L_N(f) = \exp \left( -\beta \int_0^1 \left( 1 - \sum_{j=1}^{+\infty} \pi(j) e^{-f(x)j} \right) dx \right), \quad (1.8)$$

para cada função  $f$  em  $[0, 1]$ , não negativa e mensurável, onde

$$\beta = -\log \lim_{n \rightarrow +\infty} P(N_n(u_n)([0, 1]) = 0) \quad \text{e} \quad \pi(j) = \lim_{n \rightarrow +\infty} \pi_n(j) \quad (1.9)$$

com

$$\pi_n(j) = P \left( \sum_{i=1}^{r_n} \mathbb{I}_{\{X_i > u_n\}} = j \mid \sum_{i=1}^{r_n} \mathbb{I}_{\{X_i > u_n\}} > 0 \right), \quad j = 1, 2, \dots, \quad (1.10)$$

para alguma sucessão  $r_n = [n/k_n]$  com  $\{k_n\}$  nas condições de (1.7). Além do mais, sob  $\Delta(\mathbf{u})$ , se (1.9) se verificar para algum  $\beta > 0$ , uma distribuição de probabilidades  $\pi$  e uma sucessão  $\{k_n\}_{n \geq 1}$  nas condições (1.7), então  $\{N_n(u_n)\}_{n \geq 1}$  converge em distribuição para o processo de Poisson composto anterior.

As excedências do nível  $u_n$  por  $X_i$  com  $i \in J_{n,j} = \{(j-1)r_n + 1, \dots, jr_n\}$ ,  $j = 1, \dots, k_n$ ,  $r_n = \lfloor n/k_n \rfloor$  com  $k_n$  nas condições de (1.7), ou seja, as excedências ocorrendo no mesmo bloco, são consideradas como pertencentes a um mesmo grupo de excedências. A distribuição  $\pi_n(j)$  em (1.9) é designada por distribuição condicional do tamanho de um grupo de excedências.

Se em (1.6) considerarmos apenas acontecimentos

$$A = \bigcap_{s=1}^p \{X_{i_s} \leq u_n\} \quad \text{e} \quad B = \bigcap_{s=l}^q \{X_{j_s} \leq u_n\}$$

tais que  $1 \leq i_1 < \dots < i_p < j_1 < \dots < j_q \leq n$ ,  $j_1 - i_p > l$  encontramos a condição de dependência mais fraca  $D(\mathbf{u})$ .

Esta condição, que garante a independência assintótica de máximos de blocos de variáveis suficientemente afastadas, permitiu a Leadbetter e Nandagopalan (1989) obterem a independência assintótica de máximos sobre intervalos disjuntos de sucessões estacionárias. Consequentemente, a validade da condição  $D(\mathbf{u})$ , permitiu provar que o Teorema dos Tipos Extremos continua a ser válido para sucessões estacionárias.

**Teorema 1.3.2** *Seja  $\mathbf{X}$  uma sucessão estacionária e  $a_n > 0$  e  $b_n$  constantes reais tais que  $P(a_n^{-1}(M_n - b_n) \leq x)$  converge para alguma função de distribuição não degenerada  $G$ . Se  $D(\mathbf{u})$  se verifica com  $\mathbf{u} = \{u_n = a_n x + b_n\}_{n \geq 1}$ , para cada  $x \in \mathbb{R}$ , então  $G$  é de um dos três tipos referidos no Teorema 1.2.1.*

O'Brien (1987) sob a validade de uma condição de independência assintótica mais fraca que a condição  $D(\mathbf{u})$  obtém a distribuição assintótica de  $M_n$  a partir de um número finito de variáveis de  $\mathbf{X}$ .

**Definição 1.3.2** *Dizemos que a sucessão estacionária  $\mathbf{X}$  tem independência assintótica de máximos relativamente à sucessão de números reais  $\{c_n\}_{n \geq 1}$ , e escreveremos  $\mathbf{X}$  tem  $AIM(c_n)$ , se existe uma sucessão de inteiros positivos  $\{q_n\}_{n \geq 1}$  com  $q_n = o(n)$  tal que*

$$\alpha_n := \max |P(M_{1,i}, M_{i+q+1, i+q+j} \leq c_n) - P(M_{1,i} \leq c_n)P(M_{1,j} \leq c_n)| \xrightarrow{n \rightarrow +\infty} 0,$$

para todo  $i \geq q$  e  $j \geq q$  tais que  $i + q + j \leq n$ , onde  $M_{i,j} = \max\{X_i, X_{i+1}, \dots, X_j\}$ .

**Teorema 1.3.3** *Suponhamos que  $\mathbf{X}$  tem  $AIM(c_n)$ , onde  $\{c_n\}_{n \geq 1}$  é uma sucessão de números reais e seja  $\{p_n\}_{n \geq 1}$  uma sucessão de inteiros positivos tal que  $p = o(n)$ ,  $n\alpha = o(p)$  e  $q = o(p)$ .*

Se

$$\liminf_{n \rightarrow +\infty} F^n(c_n) > 0 \quad \text{ou} \quad \liminf_{n \rightarrow +\infty} P(M_{1,p} \leq c_n \mid X_1 > c_n) > 0$$

então

$$P(M_n \leq c_n) - (F(c_n))^{nP(M_{1,p} \leq c_n \mid X_1 > c_n)} \xrightarrow{n \rightarrow +\infty} 0$$

e

$$P(M_n \leq c_n) - \exp\{-nP(X_1 > c_n, M_{1,p} \leq c_n)\} \xrightarrow{n \rightarrow +\infty} 0. \quad (1.11)$$

O processo pontual das excedências comporta-se assintoticamente como se fosse gerado por uma sucessão iid, ou seja, sem pontos múltiplos, se a sucessão estacionária  $X$  verificar uma condição de dependência local introduzida por Leadbetter (1974) e denotada por  $D'(\mathbf{u})$ .

**Definição 1.3.3** A sucessão  $X$  verifica a condição  $D'(\mathbf{u})$  para uma dada sucessão de números reais  $\mathbf{u} = \{u_n\}_{n \geq 1}$ , se

$$\limsup_{n \rightarrow +\infty} n \sum_{j=2}^{\lfloor \frac{n}{k} \rfloor} P(X_1 > u_n, X_j > u_n) \xrightarrow{k \rightarrow +\infty} 0,$$

onde  $\lfloor a \rfloor$  representa a parte inteira do real  $a$ .

A condição  $D'(\mathbf{u})$  limita a probabilidade de ocorrência de duas ou mais excedências do nível  $u_n$  em cada bloco de dimensão  $\lfloor n/k \rfloor$ . Este facto permitiu a Leadbetter (1983) estender o Teorema 1.2.1 às sucessões estacionárias como a seguir enunciamos.

**Teorema 1.3.4** Seja  $X$  uma sucessão estacionária de variáveis aleatórias com função de distribuição comum  $F$  e  $\mathbf{u} = \{u_n\}_{n \geq 1}$  uma sucessão de níveis reais e suponhamos que se verificam as condições  $D(\mathbf{u})$  e  $D'(\mathbf{u})$ . Então para  $0 \leq \tau < +\infty$ , tem-se

$$P(M_n \leq u_n) \xrightarrow{n \rightarrow +\infty} e^{-\tau}$$

se e só se

$$n(1 - F(u_n)) \xrightarrow{n \rightarrow +\infty} \tau. \quad (1.12)$$

O resultado anterior garante que a lei limite do máximo, linearmente normalizado, de uma sucessão estacionária fracamente dependente no sentido da condição  $D(\mathbf{u})$  e verificando  $D'(\mathbf{u})$ , coincide com a lei limite do máximo de uma sucessão iid.

Enquanto a condição  $D'(\mathbf{u})$  “controla” a ocorrência de excedências de  $u_n$ , a condição de dependência local  $D''(\mathbf{u})$  de Leadbetter e Nandagopalan (1989) “controla” a ocorrência de cruzamentos ascendentes de  $u_n$ . Entendemos por ocorrência de um cruzamento ascendente, ou simplesmente cruzamento (uma vez que não iremos considerar outro tipo de cruzamentos), do nível  $u_n$  pela sucessão  $X$  no instante  $i$ , a ocorrência do acontecimento  $\{X_i \leq u_n < X_{i+1}\}$ .

**Definição 1.3.4** A sucessão  $X$  verifica a condição  $D''(\mathbf{u})$  para uma dada sucessão de números reais  $\mathbf{u} = \{u_n\}_{n \geq 1}$  se existe uma sucessão de inteiros  $\mathbf{k} = \{k_n\}_{n \geq 1}$  nas condições (1.7),

$$k_n P(X_1 > u_n) \xrightarrow{n \rightarrow +\infty} 0 \text{ e}$$

$$\lim_{n \rightarrow +\infty} n \sum_{j=2}^{r_n-1} P(X_1 > u_n, X_j \leq u_n < X_{j+1}) = 0.$$

Sob a validade desta condição, mais fraca que a condição  $D'(\mathbf{u})$ , e supondo assintoticamente constante o número médio de cruzamentos de  $u_n$  entre as  $n$  primeiras variáveis de  $\mathbf{X}$ , Leadbetter e Nandagopalan (1989) obtêm a distribuição limite de  $M_n$ .

**Teorema 1.3.5** *Se  $\mathbf{X}$  verifica as condições  $D(\mathbf{u})$  e  $D''(\mathbf{u})$ , então*

$$nP(X_1 \leq u_n < X_2) \xrightarrow{n \rightarrow +\infty} \nu \quad \text{se e só se} \quad P(M_n \leq u_n) \xrightarrow{n \rightarrow +\infty} e^{-\nu}, \quad 0 \leq \nu < +\infty.$$

A condição de dependência local  $D''(\mathbf{u})$ , que restringe localmente as oscilações rápidas junto de níveis elevados tem, como seria de esperar, efeito sobre o processo pontual dos cruzamentos de  $u_n$ , gerado pela sucessão  $\mathbf{X}$ , definido por

$$\tilde{N}_n(u_n)(B) = \sum_{i=1}^{n-1} \mathbb{I}_{\{X_i \leq u_n < X_{i+1}\}} \delta_{\frac{i}{n}}(B), \quad B \subseteq \mathcal{B}([0, 1]). \quad (1.13)$$

Esta condição tem sobre  $\tilde{N}_n(u_n)$  um efeito análogo ao que  $D'(\mathbf{u})$  tem sobre  $N_n(u_n)$ , uma vez que evita o agrupamento de cruzamentos. Assim, sob a validade de  $D''(\mathbf{u})$  e supondo assintoticamente constante e igual a  $\nu$  o número médio de cruzamentos de  $u_n$  pelas  $n$  primeiras variáveis de  $\mathbf{X}$ ,  $\{\tilde{N}_n(u_n)\}_{n \geq 1}$  converge em distribuição para um processo de Poisson homogêneo sobre  $[0, 1]$  de intensidade  $\nu$ .

Em relação ao processo pontual das excedências (1.5), sob a validade de  $D''(\mathbf{u})$  e supondo assintoticamente constante o número médio de cruzamentos de  $u_n$  entre as  $n$  primeiras variáveis de  $\mathbf{X}$ , as excedências agrupadas ocorrem consecutivamente e por conseguinte, a distribuição condicional do tamanho de um grupo de excedências é assintoticamente igual à distribuição condicional do comprimento de um *run* de excedências (sequência de excedências sucessivas) que se encontra definida no resultado seguinte de Leadbetter e Nandagopalan (1989).

**Teorema 1.3.6** *Se a sucessão estacionária  $\mathbf{X}$  verifica as condições  $D(\mathbf{u})$  e  $D''(\mathbf{u})$ , com  $\mathbf{u} = \{u_n\}_{n \geq 1}$  uma sucessão tal que*

$$nP(X_1 \leq u_n < X_2) \xrightarrow{n \rightarrow +\infty} \nu > 0, \quad (1.14)$$

*então  $\pi_n(j) - \pi_n^*(j) \xrightarrow{n \rightarrow +\infty} 0$  para todo o inteiro  $j = 1, 2, \dots$ , onde*

$$\pi_n^*(j) = P(X_2 > u_n, X_3 > u_n, \dots, X_{j+1} > u_n, X_{j+2} \leq u_n \mid X_1 \leq u_n < X_2). \quad (1.15)$$

Se um nível  $u_n$  verificar (1.14) diremos que é normalizado para cruzamentos e representá-lo-emos por  $\tilde{u}_n^{(\nu)}$ .

O Teorema 1.3.1 permanece válido se considerarmos o processo pontual de cruzamentos  $\tilde{N}_n(u_n)$  e substituirmos excedências por cruzamentos como podemos constatar no próximo resultado de H. Ferreira (2006).

**Teorema 1.3.7** *Suponhamos que a condição  $\Delta(\mathbf{u})$  se verifica para  $\mathbf{X}$  e que  $\{\tilde{N}_n(u_n)\}_{n \geq 1}$  converge em distribuição para algum processo de Poisson  $\tilde{N}$ . Então  $\tilde{N}$  é necessariamente um processo de Poisson composto com transformada de Laplace*

$$L_{\tilde{N}}(f) = \exp \left( -\tilde{\beta}(0, +\infty) \int_0^1 \left( 1 - \sum_{j=1}^{+\infty} \tilde{\pi}(j) e^{-f(x)j} \right) dx \right), \quad (1.16)$$

para cada função  $f$  definida em  $[0, 1]$ , não negativa e mensurável, onde  $\tilde{\beta}(\cdot)$  é uma medida positiva concentrada nos inteiros positivos,  $\mathbb{N}$ ,

$$\tilde{\beta} \equiv \tilde{\beta}(0, +\infty) = -\log \lim_{n \rightarrow +\infty} P(\tilde{N}_n(u_n)([0, 1]) = 0), \quad (1.17)$$

$$\tilde{\pi}(\cdot) = \tilde{\beta}(\cdot) / \tilde{\beta}(\mathbb{N}) = \lim_{n \rightarrow +\infty} \tilde{\pi}_n(\cdot) \quad (1.18)$$

onde

$$\tilde{\pi}_n(j) = P \left( \sum_{i=1}^{r_n} \mathbb{I}_{\{X_i \leq u_n < X_{i+1}\}} = j \mid \sum_{i=1}^{r_n} \mathbb{I}_{\{X_i \leq u_n < X_{i+1}\}} > 0 \right), \quad j = 1, 2, \dots, \quad (1.19)$$

para alguma sucessão  $r_n = \lfloor n/k_n \rfloor$  com  $\{k_n\}$  nas condições de (1.7). Além do mais, sob  $\Delta(\mathbf{u})$ , se as convergências em (1.17) e (1.18) se verificarem para algum  $\tilde{\beta} > 0$ , uma distribuição de probabilidades  $\tilde{\pi}$  e uma sucessão  $\{k_n\}_{n \geq 1}$  nas condições (1.7), então  $\{\tilde{N}_n(u_n)\}_{n \geq 1}$  converge em distribuição para o processo de Poisson composto anterior.

Encontramos processos pontuais limite para as sucessões de processos pontuais de excedências  $\{N_n(u_n)\}_{n \geq 1}$  e de cruzamentos  $\{\tilde{N}_n(u_n)\}_{n \geq 1}$  interessantes e úteis, quando consideramos, respetivamente, níveis normalizados para excedências e níveis normalizados para cruzamentos. Nestes casos a intensidade dos processos pontuais limite depende de parâmetros importantes da Teoria de Valores Extremos, o índice extremal (Leadbetter (1983)) e o índice de cruzamentos (H. Ferreira (2006)). Falaremos mais detalhadamente sobre estes parâmetros no que se segue.

### 1.3.1 O índice extremal: caracterização e estimação

O índice extremal é um parâmetro de grande importância na Teoria de Valores Extremos uma vez que permite relacionar a distribuição limite do máximo da sucessão estacionária  $\mathbf{X}$  com a da sucessão de variáveis aleatórias independentes associada à sucessão  $\mathbf{X}$ , isto é, a sucessão iid  $\hat{\mathbf{X}} = \{\hat{X}_n\}_{n \geq 1}$  tal que  $\hat{X}_n$  e  $X_n$  têm a mesma distribuição, para todo  $n \geq 1$ . Como poderemos constatar mais à frente, este parâmetro também tem um papel fundamental na determinação do tamanho médio dos grupos de excedências.

**Definição 1.3.5** A sucessão estacionária  $\mathbf{X} = \{X_n\}_{n \geq 1}$  tem índice extremal  $\theta$ , com  $0 \leq \theta \leq 1$ , se para todo  $\tau > 0$  existir uma sucessão de níveis normalizados  $\mathbf{u}^{(\tau)} = \{u_n^{(\tau)}\}_{n \geq 1}$  tal que,

$$P(M_n \leq u_n^{(\tau)}) \xrightarrow[n \rightarrow +\infty]{} e^{-\theta\tau},$$

com  $\theta$  independente de  $\tau$ .

Se uma sucessão estacionária tem índice extremal  $\theta > 0$  então  $M_n$  tem distribuição limite não degenerada se e só se o mesmo se tem para  $\widehat{M}_n = \max\{\widehat{X}_1, \dots, \widehat{X}_n\}$ . As distribuições limite são do mesmo tipo podendo ser usadas as mesmas constantes de atração ou alteradas de modo a obter-se exatamente a mesma distribuição. Assim, se

$$P(\widehat{M}_n \leq u_n(x)) \xrightarrow[n \rightarrow +\infty]{} \widehat{G}(x)$$

e  $\mathbf{X}$  tem índice extremal  $\theta > 0$ , então

$$G(x) = \lim_{n \rightarrow +\infty} P(M_n \leq u_n(x)) = \widehat{G}^\theta(x). \quad (1.20)$$

Se (1.20) for válida e  $\mathbf{X}$  verificar a condição  $D(u_n(x))$ ,  $x \in \mathbb{R}$ , então, para  $x$  tal que  $H(x) > 0$ , temos

$$u_n(x) = u_n^{(\tau)} \quad \text{com} \quad \tau = \tau(x) = -\log G^{1/\theta}(x),$$

o que justifica em grande parte a importância dada ao estudo do processo pontual das excedências do nível  $u_n^{(\tau)}$ .

Recordemos o resultado de Hsing *et al.* (1988).

**Teorema 1.3.8** Suponhamos que, para cada  $\tau > 0$ , a condição  $\Delta(\mathbf{u}^{(\tau)})$  se verifica para  $\mathbf{X}$ . Se, para algum  $\tau_0 > 0$ ,  $\{N_n(u_n^{(\tau_0)})\}_{n \geq 1}$  converge em distribuição para algum processo pontual  $N^{(\tau_0)}$ , então para todo  $\tau > 0$ ,  $\{N_n(u_n^{(\tau)})\}_{n \geq 1}$  converge em distribuição para um processo de Poisson composto com transformada de Laplace dada por (1.8), com medida intensidade  $\beta = \theta\tau$ ,

$\theta = -\log \lim_{n \rightarrow +\infty} P(N_n^{(1)}(u_n^{(1)})([0, 1]) = 0)$ ,  $0 \leq \theta \leq \left( \sum_{j \geq 1} j \pi(j) \right)^{-1} \leq 1$ , a distribuição de probabilidade  $\pi$  sobre  $\mathbb{N}$  e  $\theta$  independente de  $\tau$ .

Este resultado garante que se  $\mathbf{X}$  verificar  $\Delta(\mathbf{u}^{(\tau)})$  e tiver índice extremal  $\theta > 0$ , então o processo pontual das excedências converge em distribuição para um processo de Poisson composto de intensidade  $\theta\tau$  e com multiplicidade média igual a  $1/\theta$ , desde que seja possível permutar o limite com a soma na seguinte igualdade provada por Nandagopalan (1990)

$$\lim_{n \rightarrow +\infty} \sum_{j=1}^{\tau_n} j \pi_n(j) = \frac{1}{\theta}. \quad (1.21)$$

com  $\pi_n$  definido em (1.10) e  $u_n = u_n^{(\tau)}$ .

A relação (1.21) garante que podemos aproximar o índice extremal pelo inverso aritmético do tamanho médio dos grupos de excedências.

A utilização prática desta relação no cálculo do índice extremal é em geral difícil e por conseguinte a obtenção de fórmulas simples de cálculo deste parâmetro tornou-se uma necessidade. Veremos a seguir que a introdução de condições de dependência local, tais como a condição  $D'(\mathbf{u})$  de Leadbetter (1983) e a condição  $D''(\mathbf{u})$  de Leadbetter e Nandagopalan (1989), tem permitido a obtenção de tais fórmulas.

### 1.3.1.1 Fórmulas de cálculo do índice extremal

Tendo em conta os resultados apresentados até aqui, torna-se claro que para sucessões iid, ou sucessões estacionárias que verifiquem as condições  $D(\mathbf{u}^{(\tau)})$  e  $D'(\mathbf{u}^{(\tau)})$ , para todo  $\tau > 0$ , se tem sempre  $\theta = 1$ . Para sucessões que verifiquem  $D''(\mathbf{u}^{(\tau)})$  todos os valores de  $\theta \in [0, 1]$  são possíveis, como podemos constatar pelo próximo resultado de Leadbetter e Nandagopalan (1989). Salientamos que o caso em que  $\theta = 0$  corresponde a uma distribuição limite para  $M_n$  degenerada.

**Teorema 1.3.9** *Se a sucessão estacionária  $\mathbf{X}$  verifica as condições  $D(\mathbf{u})$  e  $D''(\mathbf{u})$  para sucessões  $\mathbf{u}, k$  e  $l$  nas condições (1.7), e tiver índice extremal  $\theta > 0$ , então*

$$nP(X_1 \leq u_n < X_2) \xrightarrow[n \rightarrow +\infty]{} \nu \quad \text{se e somente se} \quad nP(X_1 > u_n) \xrightarrow[n \rightarrow +\infty]{} \frac{\nu}{\theta}.$$

Este resultado relaciona assim a existência de índice extremal com o comportamento assintótico do número médio de cruzamentos de  $u_n$  pelas  $n$  primeiras variáveis de  $\mathbf{X}$ , mais concretamente, sob a validade das condições do teorema anterior,

$$\theta = \lim_{n \rightarrow +\infty} \frac{P(X_1 \leq u_n < X_2)}{P(X_1 > u_n)}.$$

Chernick *et al.* (1991) definiram uma classe de condições de dependência local mais fracas que  $D''(\mathbf{u})$  e que possibilitam a determinação do índice extremal, a partir da distribuição conjunta de  $k$  variáveis consecutivas da sucessão.

**Definição 1.3.6** *Seja  $\mathbf{X}$  uma sucessão estacionária que verifica a condição  $D(\mathbf{u})$ . A condição  $D^{(k)}(\mathbf{u})$  é verificada para um inteiro  $k \geq 1$ , se existe uma sucessão  $\mathbf{k} = \{k_n\}_{n \geq 1}$  nas condições de (1.7) e*

$$\lim_{n \rightarrow +\infty} nP\left(X_1 > u_n \geq M_{2,k}, M_{k+1, \lfloor \frac{n}{k_n} \rfloor} > u_n\right) = 0$$

onde  $M_{i,j} = \max\{X_i, X_{i+1}, \dots, X_j\}$  para  $i \leq j$  e  $M_{i,j} = -\infty$  para  $i > j$ .

Em particular, se  $k = 1$ , a condição  $D^{(1)}(\mathbf{u})$  é equivalente à condição  $D'(\mathbf{u})$ , e se  $k = 2$ , a condição  $D^{(2)}(\mathbf{u})$  é implicada pela condição  $D''(\mathbf{u})$ .

Estas condições limitam a distância entre excedências não consecutivas no mesmo bloco. Concretamente, no mesmo bloco as excedências terão que estar separadas, no máximo por  $k - 2$  não excedências.

Sob a validade de  $D^{(k)}(\mathbf{u}^{(\tau)})$  para algum  $k \geq 1$  e para todo  $\tau > 0$ , o índice extremal de  $\mathbf{X}$  existe e é igual a  $\theta$  se e somente se

$$\frac{P(X_1 > u_n^{(\tau)} \geq M_{2,k})}{P(X_1 > u_n^{(\tau)})} \xrightarrow{n \rightarrow +\infty} \theta, \text{ para todo } \tau > 0.$$

### 1.3.1.2 Estimação do índice extremal

As fórmulas apresentadas anteriormente para o cálculo do índice extremal estão na base de muitos dos métodos propostos para estimar este coeficiente extremal, existindo uma vasta literatura sobre a estimação deste parâmetro. Os principais métodos propostos na literatura são o método dos blocos, o método dos *runs* e o método do tempo entre excedências.

No que se segue iremos apresentar alguns dos estimadores mais utilizados dentro de cada metodologia. Outros estimadores poderão ser encontrados, por exemplo, em Canto e Castro (1992), Gomes (1993), Smith e Weissman (1994) e Ancona-Navarrete e Tawn (2000).

#### 1.3.1.2.1 Método dos blocos

O índice extremal surge, da relação (1.21), como o inverso do limite médio de pontos em cada grupo de excedências. Por outro lado, à definição de grupo de excedências está associada uma partição da amostra  $X_1, X_2, \dots, X_n$  em blocos de dimensão  $r_n$  e o número de pontos em cada grupo de excedências será o número de excedências de um nível elevado  $u_n$  em cada um desses blocos. Mais concretamente, será representado pela variável aleatória  $N(r_n, u_n) = \sum_{i=1}^{r_n} \mathbb{I}_{\{X_i > u_n\}}$ , que não é mais do que  $N_n(u_n)([0, r_n/n])$ .

A relação (1.21) é assim equivalente a

$$\lim_{n \rightarrow +\infty} E[N(r_n, u_n^{(\tau)}) \mid N(r_n, u_n^{(\tau)}) > 0] = \frac{1}{\theta},$$

o que sugere que uma forma de estimar  $1/\theta$  é contar o número total de excedências de um nível normalizado  $u_n^{(\tau)}$  na amostra  $X_1, \dots, X_n$  e dividir pelo número de blocos em que ocorre pelo menos uma excedência. Esta foi a motivação para o estimador de blocos do índice extremal apresentada por Hsing (1991).

Esta forma de estimar o parâmetro  $\theta$  levanta alguns problemas práticos, nomeadamente a escolha do nível normalizado e do número de blocos a usar. Estas escolhas influenciam inclusivamente a consistência do estimador, como demonstrado em Hsing (1991), sendo mesmo necessário considerar um nível mais baixo para garantir a consistência do estimador. No próximo resultado, de Hsing (1991), encontramos a definição destes níveis não normalizados.

**Teorema 1.3.10** *Seja  $X$  uma sucessão estacionária com índice extremal  $\theta > 0$  e sejam  $\{r_n\}_{n \geq 1}$ ,  $\{k_n\}_{n \geq 1}$  e  $\{c_n\}_{n \geq 1}$  sucessões de números reais, tais que,*

$$c_n, k_n, r_n \rightarrow +\infty, \quad c_n/k_n \rightarrow 0, \quad k_n r_n \leq n, \quad k_n r_n \sim n \quad \text{quando } n \rightarrow +\infty.$$

*Seja  $\{v_n\}_{n \geq 1}$  uma sucessão de níveis não normalizados satisfazendo*

$$nP(X_1 > v_n) \sim c_n \tau \quad \text{quando } n \rightarrow +\infty \quad (1.22)$$

*e suponhamos que existe uma sucessão  $l_n = o(r_n)$  para a qual  $\frac{k_n}{c_n} \alpha_{n, l_n} \xrightarrow[n \rightarrow +\infty]{} 0$ , onde  $\alpha_{n, l_n}$ ,  $n \geq 1$ , representam os coeficientes da condição  $\Delta(v_n)$ . Então*

$$\lim_{n \rightarrow +\infty} \frac{k_n}{c_n} P(M_{r_n} > v_n) = \theta \tau$$

e

$$\lim_{n \rightarrow +\infty} \sum_{j \geq 1} j \pi_n(j) = \frac{1}{\theta}.$$

Notemos que a sucessão de níveis  $\{v_n\}_{n \geq 1}$  é uma sucessão de níveis normalizados relativamente à amostra  $X_1, X_2, \dots, X_{[n/c_n]}$ , *i.e.*,  $v_n = u_{[n/c_n]}^{(\tau)}$ .

Hsing (1991) propõe assim o seguinte estimador de blocos para  $\theta$

$$\hat{\theta}_n^H = \frac{\sum_{i=1}^{k_n} \mathbb{I}_{\{N^{(i)}(r_n, v_n)\}}}{N(n, v_n) > 0}$$

onde  $N^{(i)}(r_n, v_n)$  representa o número de excedências do nível  $v_n$  no  $i$ -ésimo bloco.

Impondo condições sobre o comportamento assintótico dos primeiro e segundo momentos da variável aleatória  $N(r_n, v_n)$ , Hsing (1991) demonstrou a consistência fraca deste estimador  $\hat{\theta}_n^H$ . Mais ainda, sob as condições do Teorema 1.3.10 e admitindo a convergência das séries

$$\sum_{n=1}^{+\infty} r_n k_n \alpha_{n, r_n} \quad \text{e} \quad \sum_{n=1}^{+\infty} \exp(-ac_n/r_n), \quad a > 0$$

para algum  $\tau > 0$ , prova que se tem a normalidade assintótica do estimador  $\hat{\theta}_n^H$ , depois de convenientemente normalizado, *i.e.*,

$$\sqrt{c_n}(\hat{\theta}_n^H - \theta_n) \xrightarrow[n \rightarrow +\infty]{d} N\left(0, \sqrt{\theta(\theta^2 \sigma^2 - 1)}/\tau\right)$$

onde

$$\theta_n = \frac{P(N(r_n, v_n) > 0)}{E[N(r_n, v_n)]} \quad \text{e} \quad \sigma^2 = \lim_{n \rightarrow +\infty} E[N^2(r_n, v_n)].$$

Em termos práticos, a escolha do nível  $v_n$  está condicionada ao conhecimento da função

de distribuição da variável aleatória  $X_1$ . Sendo esta informação desconhecida na maioria dos problemas, Hsing (1991) substituiu o nível  $v_n$  pela estatística ordinal superior  $X_{n-[c_n\tau]:n}$ , pois por (1.22), o número médio de excedências de  $v_n$  é aproximadamente  $c_n\tau$ . Demonstrou que este novo estimador é consistente, sendo, no entanto, ainda desconhecida a respetiva distribuição assintótica.

Robert *et al.* (2009) propuseram outros dois estimadores de blocos do índice extremal, motivados pela definição deste parâmetro. Deste modo, definindo, para um inteiro positivo  $r$ ,

$$F_r(u) = P(M_r \leq u), \quad \tau_r(u) = rP(X_1 > u), \quad \text{e} \quad \theta_r(u) = -\frac{\log F_r(u)}{\tau_r(u)},$$

tem-se, atendendo à definição do índice extremal, que  $\theta = \lim_{r \rightarrow +\infty} \theta_r(u_r)$ , onde  $u_r$  é um nível que verifica  $\lim_{r \rightarrow +\infty} rP(X_1 > u_r) = \tau$ . Os estimadores propostos distinguem-se pela forma como a função de distribuição  $F_r$  é estimada.

Primeiro, consideram a amostra  $X_1, \dots, X_n$  dividida em  $k = [n/r]$  blocos disjuntos e estimam  $F_r$  por

$$\widehat{F}_{n,r}^{dj}(u) = \frac{1}{k} \sum_{i=1}^k \mathbb{I}_{\{M_{(i-1)r+1,ir} \leq u\}},$$

onde  $M_{s,r} = \max\{X_s, \dots, X_r\}$ ,  $0 \leq s < r$ , e  $M_{0,r} = \max\{X_1, \dots, X_r\}$ .

Segundo, consideram que os  $k$  blocos de dimensão  $r$  são deslizantes e portanto estimam  $F_r$ , por

$$\widehat{F}_{n,r}^{sl}(u) = \frac{1}{n-k+1} \sum_{i=0}^{n-r} \mathbb{I}_{\{M_{i+1,i+r} \leq u\}}.$$

Os estimadores de blocos disjuntos e de blocos deslizantes de  $\theta$  são assim definidos, respetivamente, por

$$\widehat{\theta}_{n,r}^{dj}(u) = -\frac{\log \widehat{F}_{n,r}^{dj}(u)}{\widehat{\tau}_{n,r}(u)} \quad \text{e} \quad \widehat{\theta}_{n,r}^{sl}(u) = -\frac{\log \widehat{F}_{n,r}^{sl}(u)}{\widehat{\tau}_{n,r}(u)},$$

onde

$$\widehat{\tau}_{n,r}(u) = \frac{1}{k} \sum_{i=1}^{kr} \mathbb{I}_{\{X_i > u\}}.$$

Estes estimadores são consistentes e tem-se

$$\sqrt{k_n} \begin{pmatrix} \widehat{\theta}_{n,r_n}^{dj}(u_{r_n}) - \theta \\ \widehat{\theta}_{n,r_n}^{sl}(u_{r_n}) - \theta \end{pmatrix} \xrightarrow[n \rightarrow +\infty]{d} N(0, \mathbf{V})$$

onde  $\mathbf{V} = [v_{i,j}]_{2 \times 2}$  é uma matriz simétrica com

$$v_{11} = \frac{\theta^2}{\alpha^2}(e^\alpha - 1 - \alpha) + \frac{\theta^2 c^2}{\alpha},$$

$$v_{22} = v_{12} = 2\frac{\theta^2}{\alpha^3}(e^\alpha - 1 - \alpha - \frac{\alpha^2}{2}) + \frac{\theta^2 c^2}{\alpha},$$

em que  $\alpha = \theta\tau$  e  $c^2 = \frac{\sum_{j \geq 1} j^2 \pi(j) - \theta^{-2}}{\theta^{-2}}$  representa o quadrado do coeficiente de variação da distribuição  $\pi$  definida no Teorema 1.3.1.

### 1.3.1.2.2 Método dos runs

A motivação para o estimador de *runs* do índice extremal, proposto por Nandagopalan (1990), surgiu do Teorema 1.3.6 que permite contornar a dificuldade da identificação dos blocos na utilização do estimador  $\hat{\theta}_n^H$ . Sob a validade de  $D''(\mathbf{u})$  não há oscilações bruscas sempre que há um cruzamento o que significa que há um grupo de excedências consecutivas ou *run* de excedências na vizinhança de uma excedência.

A dimensão esperada de um *run* de excedências é caracterizada no próximo resultado de Nandagopalan (1990).

**Teorema 1.3.11** *Seja X uma sucessão estacionária com índice extremal  $\theta > 0$  verificando as condições  $D(\mathbf{u}^{(\tau)})$  e  $D''(\mathbf{u}^{(\tau)})$ , para cada  $\tau > 0$ . Se  $P(X_1 > u_n^{(\tau)}, \dots, X_j > u_n^{(\tau)}) \xrightarrow{j \rightarrow +\infty} 0$ , então*

$$\sum_{j \geq 1} j \pi_n^*(j) = \frac{P(X_1 > u_n^{(\tau)})}{P(X_1 \leq u_n^{(\tau)} < X_2)} = \frac{E[N(n, u_n^{(\tau)})]}{E[\tilde{N}(n, u_n^{(\tau)})]} \xrightarrow{n \rightarrow +\infty} \frac{1}{\theta},$$

com  $\pi_n^*$  definida em (1.15),  $N(n, u_n^{(\tau)}) = N_n(u_n^{(\tau)})([0, 1])$  e  $\tilde{N}(n, u_n^{(\tau)}) = \tilde{N}(u_n^{(\tau)})([0, 1])$ .

Com base neste resultado Nandagopalan (1990), assumindo que a amostra  $X_1, \dots, X_n$  tem pelo menos um cruzamento, propôs o seguinte estimador de *runs* para o índice extremal

$$\hat{\theta}_n^N = \frac{\tilde{N}(n, v_n)}{N(n, v_n)}$$

onde  $v_n$  são níveis não normalizados definidos de acordo com (1.22).

A utilização destes níveis não normalizados é essencial para se ter a consistência de  $\hat{\theta}_n^N$ . Sob certas condições Nandagopalan (1990) estabelece também a distribuição assintótica do estimador

$$\sqrt{c_n}(\hat{\theta}_n^N - \theta_n) \xrightarrow[n \rightarrow +\infty]{d} N\left(0, \sqrt{\theta(\theta^2 \sigma^2 - 1)/\tau}\right)$$

onde,

$$\theta_n = \frac{E[\tilde{N}(n, v_n)]}{E[N(n, v_n)]} \quad \text{e} \quad \sigma^2 = \lim_{n \rightarrow +\infty} \sum_{j \geq 1} j^2 \pi_n^*(j).$$

Nandagopalan (1990) também propôs uma variante ao seu estimador de *runs*, que como anteriormente resulta da substituição do nível  $v_n$  pela estatística ordinal superior  $V_n = X_{n-[c_n \tau]:n}$

$$\tilde{\theta}_n^N = \frac{\tilde{N}(n, V_n)}{[c_n \tau]}.$$

A consistência deste estimador está assegurada, no entanto, permanece desconhecida a sua distribuição assintótica.

### 1.3.1.2.3 Método do tempo entre excedências

A distribuição assintótica dos tempos, devidamente normalizados, entre excedências consecutivas de um nível elevado  $u_n$ , pertence a uma família de distribuições indexada por um parâmetro que é o índice extremal como provado por Ferro e Segers (2003). Mais concretamente, se a variável aleatória  $T(u)$  representa o tempo que separa duas excedências consecutivas do nível elevado  $u$ , pelas variáveis de  $\mathbf{X}$  com distribuição comum  $F$ , então

$$T(u) \stackrel{d}{=} \min\{n \geq 1 : X_{n+1} > u\}, \quad \text{dado que } X_1 > u,$$

isto é,

$$P(T(u) = n) = P(M_{2,n} \leq u, X_{n+1} > u \mid X_1 > u), \quad n \geq 1$$

ou, alternativamente,

$$P(T(u) > n) = P(M_{2,n+1} \leq u \mid X_1 > u), \quad n \geq 1,$$

com  $M_{i,j} = \max\{X_i, \dots, X_j\}$  para  $0 \leq i < j$ .

Sob a validade de uma condição de independência assintótica mais fraca que  $D(\mathbf{u})$  e para níveis  $u_n$  satisfazendo certas condições Ferro e Segers (2003) provam, recorrendo ao Teorema 1.3.3, que

$$P(\bar{F}(u_n)T(u_n) \leq x) \xrightarrow{n \rightarrow +\infty} 1 - \theta \exp\{-\theta x\}, \quad x > 0,$$

com  $\bar{F}(u) = 1 - F(u)$ . Tal significa que

$$\bar{F}(u)T(u) \xrightarrow{d} T_\theta, \quad \text{quando } u \uparrow x_F, \quad (1.23)$$

com  $T_\theta$  uma variável aleatória com distribuição dada pela seguinte mistura de distribuições,

$$(1 - \theta)\varepsilon_0 + \theta\mu_\theta,$$

onde  $\varepsilon_0$  representa uma distribuição de probabilidade degenerada em 0 e  $\mu_\theta$  representa a distribuição exponencial com valor médio  $1/\theta$ .

O índice extremal pode assim ser interpretado como a proporção de tempos não nulos entre excedências e como o inverso da média dos tempos não nulos entre excedências. Deste modo, tendo em conta (1.23) e o facto de  $E(T_\theta) = 1$ ,  $E(T_\theta^2) = 2/\theta$  e  $P(T(u_n) > n) = \theta F(u_n)^{n\theta} + o(1)$ , Ferro e Segers (2003) propõem o seguinte estimador para o índice extremal

$$\hat{\theta}_n^{FS}(u) = \begin{cases} \min\{1, \hat{\theta}_1^{FS}(u)\} & \text{se } \max\{T_i : 1 \leq i \leq N-1\} \leq 2 \\ \min\{1, \hat{\theta}_2^{FS}(u)\} & \text{se } \max\{T_i : 1 \leq i \leq N-1\} > 2 \end{cases}$$

onde

$$\hat{\theta}_1^{FS}(u) = \frac{2 \left( \sum_{i=1}^{N-1} T_i \right)^2}{(N-1) \sum_{i=1}^{N-1} T_i^2} \quad \text{e} \quad \hat{\theta}_2^{FS}(u) = \frac{2 \left[ \sum_{i=1}^{N-1} (T_i - 1) \right]^2}{(N-1) \sum_{i=1}^{N-1} (T_i - 1)(T_i - 2)},$$

com  $T_1, \dots, T_{N-1}$  uma amostra de tempos entre excedências.

Este estimador é consistente para processos  $m$ -dependentes.

### 1.3.2 O índice de cruzamentos: caracterização

No Teorema 1.3.7 vimos que se a sucessão de processos pontuais de cruzamentos  $\{\tilde{N}_n(u_n)\}_{n \geq 1}$  converge em distribuição, então o processo pontual limite é necessariamente um processo de Poisson composto. Obtém-se o mesmo processo pontual limite quando consideramos um nível  $\tilde{u}_n^{(\nu)}$  nas condições (1.14), contudo agora a intensidade do processo de Poisson composto está relacionada com um parâmetro  $\eta$ , denominado índice de cruzamentos (H. Ferreira (2006)).

**Teorema 1.3.12** *Suponhamos que para cada  $\nu > 0$ ,  $\mathbf{X}$  verifica a condição  $\Delta(\tilde{\mathbf{u}}^{(\nu)})$ . Se  $\{\tilde{N}_n(\tilde{u}_n^{(1)})\}_{n \geq 1}$  converge em distribuição para algum processo pontual  $\tilde{N}^{(1)}$ , então para todo  $\nu > 0$ ,  $\{\tilde{N}_n(\tilde{u}_{[n/\nu]}^{(1)})\}_{n \geq 1}$  converge em distribuição para um processo de Poisson composto com transformada de Laplace dada por (1.16)-(1.18), com intensidade  $\tilde{\beta} = \eta\nu$ ,  $\nu = -\log \lim_{n \rightarrow +\infty} P(\tilde{N}_n(\tilde{u}_n^{(1)})([0, 1]) = 0)$ ,*

$$0 \leq \eta \leq \left( \sum_{j \geq 1} j \tilde{\pi}(j) \right)^{-1} \leq 1, \quad (1.24)$$

medida de probabilidade  $\tilde{\pi}$  sobre  $\mathbb{N}$  e  $\eta$  independente de  $\nu$ . Para cada  $\tilde{v}_n^{(\nu)}$  tal que

$$nP(\min\{\tilde{v}_n^{(\nu)}, \tilde{u}_{[n/\nu]}^{(1)}\} < X_1 \leq \max\{\tilde{v}_n^{(\nu)}, \tilde{u}_{[n/\nu]}^{(1)}\}) \xrightarrow{n \rightarrow +\infty} 0, \quad (1.25)$$

o mesmo processo de Poisson surge como limite de  $\{\tilde{N}_n(\tilde{v}_n^{(\nu)})\}_{n \geq 1}$ .

Por (1.24) o índice de cruzamentos pode ser interpretado como uma medida do tamanho dos agrupamentos de cruzamentos de um nível  $u_n$  pelas variáveis de  $\mathbf{X}$ .

**Definição 1.3.7** Se para cada  $\nu > 0$ , existe uma sucessão de níveis  $\{\tilde{u}_n^{(\nu)}\}_{n \geq 1}$  e  $\lim_{n \rightarrow +\infty} P(\tilde{N}_n(\tilde{u}_n^{(\nu)}) = 0) = e^{-\eta\nu}$ , para alguma constante  $0 \leq \eta \leq 1$ , então dizemos que a sucessão  $\mathbf{X}$  tem índice de cruzamentos  $\eta$ .

H. Ferreira (2006), mostrou que se existem níveis  $\tilde{u}_n^{(\nu)}$  e  $\tilde{u}_n^{(\nu)} = u_n^{(\tau)}$  para algum  $\tau > 0$ , então  $P(\tilde{N}_n(\tilde{u}_n^{(\nu)})([0, 1]) = 0) - P(N_n(u_n^{(\tau)})([0, 1]) = 0) \xrightarrow{n \rightarrow +\infty} 0$ , com  $N_n(u_n^{(\tau)})$  dado em (1.5), e o índice de cruzamentos  $\eta$  existe se e só se existe o índice extremal  $\theta$ . Neste caso,

$$\theta = \frac{\nu}{\tau} \eta. \quad (1.26)$$

Quando  $\eta = 1$ , o que acontece sobre a validade da condição  $D''(\mathbf{u})$  encontramos a fórmula para  $\theta$  obtida por Leadbetter e Nandagopalan (1989) e apresentada no Teorema 1.3.9.

### 1.3.2.1 Fórmulas de cálculo do índice de cruzamentos

Em H. Ferreira (2006) são introduzidas condições de dependência local que permitem calcular  $\lim_{n \rightarrow +\infty} P(\tilde{N}_n(\tilde{u}_n^{(\nu)}) = 0)$  a partir de  $k$  variáveis consecutivas de  $\mathbf{X}$ .

**Definição 1.3.8** Seja  $\mathbf{X}$  uma sucessão estacionária que verifica a condição  $\Delta(\mathbf{u})$ . Para todo  $k \geq 2$ ,  $\mathbf{X}$  verifica a condição  $\tilde{D}^{(k)}(\mathbf{u})$  se

$$\lim_{n \rightarrow +\infty} nP(X_1 \leq u_n < X_2, \tilde{N}_{3,k} = 0, \tilde{N}_{k+1,r_n} > 0) = 0$$

para alguma sucessão de inteiros  $\{r_n = [n/k_n]\}_{n \geq 1}$  com  $\mathbf{k} = \{k_n\}_{n \geq 1}$  nas condições de (1.7) e  $\tilde{N}_{i,j} = \tilde{N}_n(u_n)([i/n, j/n])$ ,  $\tilde{N}_{i,j} = 0$  se  $j < i$ .

Estas condições de dependência local  $\tilde{D}^{(k)}(\mathbf{u})$  são ligeiramente mais fortes que as condições  $D^{(k)}(\mathbf{u})$  de Chernick *et al.* (1991) e quando  $k = 2$  obtemos uma condição ligeiramente mais forte que  $D''(\mathbf{u})$ . Sob a validade de  $\tilde{D}^{(k)}(\mathbf{u})$ , é assintoticamente improvável que, num mesmo bloco, os cruzamentos ascendentes estejam separados por mais de  $k - 3$  não cruzamentos.

**Teorema 1.3.13** Suponhamos que a sucessão estacionária  $\mathbf{X}$  verifica a condição  $\Delta(\mathbf{u})$  e que  $\liminf_{n \rightarrow +\infty} P(\tilde{N}_n(u_n)([0, 1]) = 0) > 0$ . Então para cada inteiro  $k > 0$ ,

$$P(\tilde{N}_n(u_n)([0, 1]) = 0) - \exp(-nP(X_1 \leq u_n < X_2, \tilde{N}_{3,k} = 0)) \xrightarrow{n \rightarrow +\infty} 0$$

se e só se  $\mathbf{X}$  verifica a condição  $\tilde{D}^{(k)}(\mathbf{u})$ .

Os resultados anteriores permitem obter, sob a validade de alguma condição  $\tilde{D}^{(k)}(\tilde{\mathbf{u}}^{(\nu)})$ ,  $k \geq 2$ , a seguinte fórmula de cálculo do índice de cruzamentos  $\eta$ .

**Corolário 1.3.1** *Suponhamos que a sucessão  $\mathbf{X}$  verifica as condições  $\Delta(\mathbf{u})$  e, para algum  $k \geq 2$ ,  $\tilde{D}^{(k)}(\tilde{\mathbf{u}}^{(\nu)})$  para cada  $\nu > 0$ . Então o índice de cruzamentos de  $\mathbf{X}$  existe e é igual a  $\eta$ , se e só se,*

$$P(\tilde{N}_{3,k}(\tilde{u}_n^{(\nu)}) = 0 \mid X_1 \leq \tilde{u}_n^{(\nu)} < X_2) \xrightarrow{n \rightarrow +\infty} \eta, \quad (1.27)$$

para cada  $\nu > 0$ .

O próximo resultado, de H. Ferreira (2006), garante que as condições  $\tilde{D}^{(k)}(\tilde{\mathbf{u}}^{(\nu)})$ ,  $k \geq 2$ , são necessárias e suficientes para garantir que se tenha a relação (1.26).

**Corolário 1.3.2** *Seja  $\mathbf{X}$  uma sucessão estacionária que verifica as condições  $\Delta(\mathbf{u})$  e, para algum  $k \geq 2$ ,  $\tilde{D}^{(k)}(\tilde{\mathbf{u}}^{(\nu)})$  para cada  $\nu > 0$ . Se  $\tilde{u}_n^{(\nu)} = u_n^{(\tau)}$  para algum  $\tau > 0$ , então o índice extremal de  $\mathbf{X}$  existe e é igual a  $\theta = \frac{\nu}{\tau} \eta$  se e só se*

$$P(\tilde{N}_{3,k}(\tilde{u}_n^{(\nu)}) = 0 \mid X_1 \leq \tilde{u}_n^{(\nu)} < X_2) \xrightarrow{n \rightarrow +\infty} \eta,$$

para cada  $\nu > 0$ .

Considerando agora  $\tilde{\pi}_n^*$  a distribuição condicional do tamanho de um *run* de cruzamentos definida por

$$\tilde{\pi}_n^*(j) = P(X_3 \leq u_n < X_4, \dots, X_{2j+1} \leq u_n < X_{2j+2}, \tilde{N}_{2j+3,2j+3} = 0 \mid \tilde{N}_{1,1} = 0, X_3 \leq u_n < X_4), \quad j \geq 1, \quad (1.28)$$

encontramos no resultado seguinte de H. Ferreira (2006) uma forma de calcularmos um valor aproximado para o tamanho médio destes *runs* de cruzamentos.

**Teorema 1.3.14** *Se para cada  $\nu > 0$ , a sucessão estacionária  $\mathbf{X}$  verifica a condição  $\tilde{D}^{(3)}(\tilde{\mathbf{u}}^{(\nu)})$  e  $P(X_3 \leq \tilde{u}_n^{(\nu)} < X_4, \dots, X_{2j-1} \leq \tilde{u}_n^{(\nu)} < X_{2j}) \rightarrow 0$ , com  $j \rightarrow +\infty$ , então  $\mathbf{X}$  tem índice de cruzamentos  $\eta$  se e só se*

$$\sum_{j \geq 1} j \tilde{\pi}_n^*(j) = \frac{P(X_1 \leq \tilde{u}_n^{(\nu)} < X_2)}{P(X_1 \leq \tilde{u}_n^{(\nu)} < X_2, \tilde{N}_{3,3} = 0)} \xrightarrow{n \rightarrow +\infty} \frac{1}{\eta}, \quad (1.29)$$

para cada  $\nu > 0$ .

O índice de cruzamentos para além de se relacionar com o índice extremal também se relaciona com outros coeficientes de dependência, nomeadamente, sob a validade da condição  $\tilde{D}^{(3)}(\mathbf{u})$ , relaciona-se com o coeficiente de dependência na cauda para cruzamentos

$$\mu_1 = \lim_{x \rightarrow x_F} P(X_3 \leq \tilde{u}_n^{(\nu)} < X_4 \mid X_1 \leq \tilde{u}_n^{(\nu)} < X_2)$$

introduzido em M. Ferreira e Ferreira (2012), através da relação  $\eta = 1 - \mu_1$ .

Facilmente se obtêm expressões alternativas para o índice de cruzamentos  $\eta$  considerando apenas a estacionaridade, como se constata no próximo resultado.

**Lema 1.3.2** *Se  $nP(X_1 \leq u_n < X_2, \tilde{N}_{3,3} = 0) \xrightarrow[n \rightarrow +\infty]{} \xi > 0$  então as seguintes convergências são equivalentes*

i)  $P(\tilde{N}_{3,3} = 0 \mid X_1 \leq u_n < X_2) \xrightarrow[n \rightarrow +\infty]{} \eta;$

ii)  $nP(X_1 \leq u_n < X_2) \xrightarrow[n \rightarrow +\infty]{} \frac{\xi}{\eta};$

iii)  $n(1 - P(\tilde{N}_{1,1} = 0, \tilde{N}_{3,3} = 0)) \xrightarrow[n \rightarrow +\infty]{} \xi + \frac{\xi}{\eta}.$



# Capítulo 2

## Índice de cruzamentos após subamostragem sistemática

### 2.1 Introdução

Na última década tem-se assistido a um aumento do interesse no comportamento extremal de processos em frequências distintas da observada, em geral uma frequência unitária. Este interesse é muitas vezes motivado pela análise de diferentes esquemas de monitorização em fenómenos médicos, financeiros, ambientais, entre outros.

Robinson e Tawn (2000) salientaram a importância que a frequência amostral tem nas propriedades extremas. Neste trabalho consideraram uma sucessão subamostrada a partir de uma sucessão estacionária inicial  $\mathbf{X} = \{X_n\}_{n \geq 1}$ , em intervalos fixos de  $T$ -unidades, *i.e.*, a sub-sucessão  $\mathbf{Y} = \{Y_n = X_{1+(n-1)T}\}_{n \geq 1}$ ,  $T \geq 1$ . Neste caso de subamostragem sistemática de passo constante  $T$ , a estacionaridade é preservada e se  $\mathbf{X}$  e  $\mathbf{Y}$  têm, respetivamente, índices extremos  $\theta_{\mathbf{X}}$  e  $\theta_{\mathbf{Y}}$ , então

$$\theta_{\mathbf{X}} \leq \theta_{\mathbf{Y}} \leq T\theta_{\mathbf{X}} \left( 1 - \sum_{j=1}^{T-1} \left( 1 - \frac{j}{T} \pi(j) \right) \right), \quad (2.1)$$

onde  $\pi$  é a distribuição de probabilidade limite do tamanho dos grupos de excedências de níveis elevados por  $\mathbf{X}$ , definida em (1.9). O limite superior é atingido quando os grupos de excedências consistem em observações consecutivas, o que se verifica sob a validade da condição  $D''(u)$  de Leadbetter e Nandagopalan (1989).

Os resultados deste trabalho pioneiro foram aplicados ao estudo das alturas das ondas do mar na costa sul de Inglaterra. Em particular, perante uma sucessão das alturas das ondas observadas por hora, Robinson e Tawn (2000) relacionaram o máximo destas com o máximo das alturas das ondas coincidentes com as marés altas, que ocorrem sensivelmente de 12 em 12 horas. Estes resultados foram generalizados em Scotto *et al.* (2003) e Scotto e Turkman (2002) assumindo estruturas lineares e não lineares do processo original. Scotto *et al.* (2003) analisaram também o efeito que a subamostragem sistemática exerce sobre as propriedades extremas de processos de médias móveis.

No contexto de sucessões estacionárias de valores inteiros, Hall e Scotto (2003) analisaram o impacto deste tipo de subamostragem nas propriedades extremas de sucessões de médias móveis de valores inteiros não negativos, frisando a importância desta metodologia na análise de séries temporais.

Martins e Ferreira (2004) considerando uma outra função determinista de subamostragem obtêm padrões de subamostragem mais complexos que o anterior e que incluem este como caso particular. Analisam o efeito que esta subamostragem por blocos tem nas condições de dependência  $D(\mathbf{u})$  e  $D''(\mathbf{u})$ , que assumem válidas para a sucessão inicial, e obtêm uma fórmula para o cálculo do índice extremal da sucessão subamostrada. Esta fórmula permite-lhes relacionar o índice extremal da sucessão inicial com o da sucessão subamostrada. Contrariamente à majoração (2.1), obtida por Robinson e Tawn (2000), para o índice extremal da sucessão subamostrada, conseguem obter uma expressão exata para este parâmetro.

Os trabalhos anteriores permitiram concluir que a frequência amostral influencia o valor do índice extremal, no sentido em que uma redução da frequência amostral reduz a dependência entre valores extremos. Será assim de esperar que conclusões semelhantes possam ser retiradas para o índice de cruzamentos  $\eta$  de H. Ferreira (2006), até porque estes dois coeficientes extremos se relacionam entre si.

Seguindo de perto a estrutura de Sebastião *et al.* (2010), analisamos o efeito que a subamostragem sistemática de passo constante  $T \geq 1$  tem no índice de cruzamentos. Para tal, começamos por provar, na Secção 2.2, que a intensidade do processo pontual limite do processo pontual de  $T$ -cruzamentos é caracterizada por um parâmetro  $\eta^{(T)} \in [0, 1]$ , a que chamamos índice de  $T$ -cruzamentos e que pode ser interpretado como uma medida do agrupamento de  $T$ -cruzamentos de um nível elevado pelas variáveis de  $\mathbf{X}$ .

Na Secção 2.3, introduzimos condições de dependência local que permitem obter o índice de  $T$ -cruzamentos a partir da distribuição conjunta de um número finito de variáveis de  $\mathbf{X}$ . Mostramos de que forma os índices de  $T$ -cruzamentos se relacionam entre si e uma vez que  $\eta := \eta^{(1)}$  generalizamos naturalmente a relação (1.26).

Apresentamos, na Secção 2.4, uma análise detalhada do efeito que a subamostragem sistemática tem no valor do índice de cruzamentos. Assumindo que a sucessão estacionária  $\mathbf{X}$  verifica determinadas condições de dependência, provamos que a sucessão subamostrada, a partir de  $\mathbf{X}$ ,  $\mathbf{Y} = \{Y_n = X_{1+(n-1)T}\}_{n \geq 1}$ , verifica condições análogas e calculamos o seu índice de cruzamentos  $\eta^{(1, \mathbf{Y})}$  a partir da relação com o índice de  $T$ -cruzamentos de  $\mathbf{X}$ ,  $\eta^{(T, \mathbf{X})}$ .

Os resultados das duas secções anteriores são ilustrados com o processo autoregressivo de máximos de Alpuim (1989) e com o processo autoregressivo de primeira ordem, negativamente correlacionado, considerado em Chernick *et al.* (1991).

## 2.2 Processo pontual de $T$ -cruzamentos

Seja  $\mathbf{X} = \{X_n\}_{n \geq 1}$  uma sucessão estacionária e  $\mathbf{u} = \{u_n\}_{n \geq 1}$  uma sucessão de níveis reais. Para todo o inteiro  $T \geq 1$ , dizemos que  $\mathbf{X}$  tem um  $T$ -cruzamento de  $u_n$  em  $i \geq 1$  se o acontecimento  $\{X_i \leq u_n < X_{i+T}\}$  se realiza.

Definimos o processo pontual de  $T$ -cruzamentos por

$$N_n^{(T)}(B) = N_n^{(T)}(u_n)(B) = \sum_{i=1}^n \mathbb{I}_{\{X_i \leq u_n < X_{i+T}\}} \delta_{\frac{i}{n}}(B), \quad B \subset [0, 1], \quad n \geq 1. \quad (2.2)$$

Quando  $T = 1$ , obtemos o processo pontual de cruzamentos definido em (1.13).

Considerando que para cada  $j = 1, 2, \dots, k_n$ , os  $T$ -cruzamentos de  $u_n$  em  $i$  com  $i \in J_{n,j} = \{(j-1)r_n + 1, \dots, jr_n\}$  pertencem ao mesmo grupo de  $T$ -cruzamentos, demonstramos de seguida a independência assintótica dos  $T$ -cruzamentos nestes blocos disjuntos  $J_{n,j}$ . Este resultado permitirá provar, com argumentos análogos aos apresentados por Hsing *et al.* (1988) para o processo pontual de excedências e aos de H. Ferreira (2006) para o processo pontual de cruzamentos, que sob a validade da condição  $\Delta(\mathbf{u})$ , o processo pontual limite  $N^{(T)}$  de  $\{N_n^{(T)}\}_{n \geq 1}$ , quando  $n \rightarrow +\infty$ , é um processo pontual de Poisson composto.

**Lema 2.2.1** *Suponhamos que  $\mathbf{X}$  verifica a condição  $\Delta(\mathbf{u})$  e que  $\mathbf{k} = \{k_n\}_{n \geq 1}$  satisfaz (1.7). Então*

$$P\left(N_n^{(T)}([0, 1]) = 0\right) - P^{k_n}\left(N_n^{(T)}\left(\left[0, \frac{r_n}{n}\right]\right) = 0\right) \xrightarrow{n \rightarrow +\infty} 0.$$

**Demonstração:** Considerando no Lema 1.3.1  $\chi_{n,i}$  igual à função indicatriz do acontecimento  $\{X_i \leq u_n < X_{i+T}\}$ ,  $f = 1$ ,  $a_n \xrightarrow{n \rightarrow +\infty} +\infty$  e  $k_n e^{-a_n} \xrightarrow{n \rightarrow +\infty} 0$ , obtemos

$$E\left[e^{-a_n N_n^{(T)}([0,1])}\right] - \prod_{j=1}^{k_n} E\left[e^{-a_n N_n^{(T)}\left(\frac{J_{n,j}}{n}\right)}\right] \xrightarrow{n \rightarrow +\infty} 0.$$

O resultado fica demonstrado atendendo a que

$$\begin{aligned} E\left[e^{-a_n N_n^{(T)}([0,1])}\right] - P(N_n^{(T)}([0, 1]) = 0) &= \sum_{r=1}^{+\infty} e^{-a_n r} P(N_n^{(T)}([0, 1]) = r) \\ &\leq \frac{e^{-a_n}}{1 - e^{-a_n}} = o(1) \end{aligned}$$

e

$$\begin{aligned} &\left| \prod_{j=1}^{k_n} E\left[e^{-a_n N_n^{(T)}\left(\frac{J_{n,j}}{n}\right)}\right] - \prod_{j=1}^{k_n} P\left(N_n^{(T)}\left(\frac{J_{n,j}}{n}\right) = 0\right) \right| \\ &\leq \sum_{j=1}^{k_n} \left| E\left[e^{-a_n N_n^{(T)}\left(\frac{J_{n,j}}{n}\right)}\right] - P\left(N_n^{(T)}\left(\frac{J_{n,j}}{n}\right) = 0\right) \right| \\ &\leq \sum_{j=1}^{k_n} \sum_{r=1}^{+\infty} e^{-a_n r} P\left(N_n^{(T)}\left(\frac{J_{n,j}}{n}\right) = r\right) \\ &\leq k_n \frac{e^{-a_n}}{1 - e^{-a_n}} = o(1). \end{aligned}$$

□

Sabemos que um processo de Poisson composto fica completamente especificado se conhecermos a intensidade do processo de Poisson simples que lhe está associado e a distribuição de probabilidade das suas multiplicidades. Assim, o teorema que se segue indica-nos que essas características para o processo de Poisson composto  $N^{(T)}$ , limite de  $N_n^{(T)}$  quando  $n \rightarrow +\infty$ , estão relacionadas com a probabilidade de ocorrência de um  $T$ -cruzamento e com a distribuição do número de pontos em cada grupo de  $T$ -cruzamentos.

**Teorema 2.2.1** *Suponhamos que  $\mathbf{X}$  verifica a condição  $\Delta(\mathbf{u})$  e que a sucessão  $\{N_n^{(T)}\}_{n \geq 1}$  converge em distribuição para um processo pontual  $N^{(T)}$ . Então  $N^{(T)}$  é necessariamente um processo de Poisson composto com transformada de Laplace*

$$L_{N^{(T)}}(f) = \exp \left( -\beta(0, +\infty) \int_0^1 \left( 1 - \sum_{j \geq 1} \pi(j) e^{-f(x)j} \right) dx \right), \quad (2.3)$$

para cada função  $f$ , definida em  $[0, 1]$ , não negativa e mensurável, onde

$$\beta \equiv \beta(0, +\infty) = -\log \lim_{n \rightarrow +\infty} P(N_n^{(T)}([0, 1]) = 0) \quad (2.4)$$

representa uma medida finita mensurável concentrada em  $\mathbb{N}$  e

$$\pi(\cdot) = \frac{\beta(\cdot)}{\beta(\mathbb{N})} = \lim_{n \rightarrow +\infty} \pi_n(\cdot), \quad (2.5)$$

com

$$\pi_n(j) = P \left( \sum_{i=1}^{r_n} \mathbb{I}_{\{X_i \leq u_n < X_{i+T}\}} = j \mid \sum_{i=1}^{r_n} \mathbb{I}_{\{X_i \leq u_n < X_{i+T}\}} > 0 \right), j = 1, 2, \dots,$$

para alguma sucessão  $\{r_n = [n/k_n]\}_{n \geq 1}$  com  $\mathbf{k} = \{k_n\}_{n \geq 1}$  satisfazendo (1.7). Ainda, sob a validade de  $\Delta(\mathbf{u})$ , se os limites em (2.4) e (2.5) se verificam para algum  $\beta > 0$ , alguma distribuição de probabilidade  $\pi$  e uma sucessão  $\mathbf{k}$  satisfazendo (1.7), então  $\{N_n^{(T)}\}_{n \geq 1}$  converge em distribuição para o processo de Poisson composto definido anteriormente.

**Demonstração:** Os argumentos utilizados para provar (1.8) no Teorema 1.3.1 permanecem válidos para o processo pontual de  $T$ -cruzamentos,  $N_n^{(T)}$ , e consequentemente tem-se a validade de (2.3) e (2.4).

Para a demonstração da convergência (2.5), comecemos por notar que pelo Lema 1.3.1

$$\lim_{n \rightarrow +\infty} E[\exp(-tN_n^{(T)}([0, 1]))] = \lim_{n \rightarrow +\infty} E^{k_n} \left[ \exp \left( -t \sum_{i=1}^{r_n} \mathbb{I}_{\{X_i \leq u_n < X_{i+T}\}} \right) \right], t \in \mathbb{R}.$$

Agora, tendo em conta o Lema 2.2.1 e (2.4), obtemos

$$\begin{aligned}
& E \left[ \exp \left( -t \sum_{i=1}^{r_n} \mathbb{I}_{\{X_i \leq u_n < X_{i+T}\}} \right) \right] \\
&= 1 - P \left( \sum_{i=1}^{r_n} \mathbb{I}_{\{X_i \leq u_n < X_{i+T}\}} > 0 \right) \left( 1 - \sum_{j \geq 1} e^{-tj} \pi_n(j) \right) \\
&= 1 - \frac{\beta}{k_n} \left( 1 - \sum_{j \geq 1} e^{-tj} \pi_n(j) \right) (1 + o(1))
\end{aligned}$$

e como

$$E^{k_n} \left[ \exp \left( -t \sum_{i=1}^{r_n} \mathbb{I}_{\{X_i \leq u_n < X_{i+T}\}} \right) \right] \approx E[\exp(-tN_n^{(T)}([0, 1]))]$$

converge, então também  $\alpha_n = \beta(1 - \sum_{j \geq 1} e^{-tj} \pi_n(j))$  converge, ou seja, existe uma medida  $\pi'$  concentrada sobre  $\mathbb{N}$  tal que  $\pi'(j) = \lim_{n \rightarrow +\infty} \pi_n(j)$ ,  $j = 1, 2, \dots$ . Assim, temos que

$$\begin{aligned}
\lim_{n \rightarrow +\infty} E[\exp(-tN_n^{(T)}([0, 1]))] &= \lim_{n \rightarrow +\infty} E^{k_n} \left[ \exp \left( -t \sum_{i=1}^{r_n} \mathbb{I}_{\{X_i \leq u_n < X_{i+T}\}} \right) \right] \\
&= \lim_{n \rightarrow +\infty} \left( 1 - \frac{\alpha_n}{k_n} \right)^{k_n} \\
&= \exp \left( -\beta \left( 1 - \sum_{j \geq 1} e^{-tj} \pi'(j) \right) \right).
\end{aligned}$$

No entanto,

$$\begin{aligned}
\lim_{n \rightarrow +\infty} E[\exp(-tN_n^{(T)}([0, 1]))] &= \lim_{n \rightarrow +\infty} E[\exp(-tN^{(T)}([0, 1]))] \\
&= \exp \left( -\beta \left( 1 - \sum_{j \geq 1} e^{-tj} \pi(j) \right) \right)
\end{aligned}$$

o que significa que  $\pi \equiv \pi'$ , donde resulta (2.5).

A última parte do resultado obtém-se com argumentos análogos aos utilizados por Hsing *et al.* (1988) na demonstração da segunda parte do Teorema 1.3.1.  $\square$

Notemos que a distribuição de multiplicidade  $\pi$  do processo pontual limite  $N^{(T)}$  de  $N_n^{(T)}$ , quando  $n \rightarrow +\infty$ , pode ser interpretada como a distribuição limite do tamanho dos grupos de  $T$ -cruzamentos de  $u_n$  por  $\mathbf{X}$ . Mais ainda, o Teorema 2.2.1 estende o Teorema 1.3.7 a valores de

$T$  superiores a 1.

Consideremos agora a sucessão de níveis normalizados para  $T$ -cruzamentos, isto é, a sucessão  $\tilde{\mathbf{u}}^{(\nu^{(T)})} = \{\tilde{u}_n^{(\nu^{(T)})}\}_{n \geq 1}$  tal que

$$nP(X_1 \leq \tilde{u}_n^{(\nu^{(T)})} < X_{1+T}) \xrightarrow{n \rightarrow +\infty} \nu^{(T)} > 0. \quad (2.6)$$

Salientamos que para dois níveis normalizados diferentes,  $u_{n,1} \equiv \tilde{u}_{n,1}^{(\nu^{(T)})}$  e  $u_{n,2} \equiv \tilde{u}_{n,2}^{(\nu^{(T)})}$  satisfazendo (2.6), ou seja, no limite com o mesmo número médio de  $T$ -cruzamentos, a convergência

$$P(N_n^{(T)}(u_{n,1})([0, 1]) \neq N_n^{(T)}(u_{n,2})([0, 1])) \xrightarrow{n \rightarrow +\infty} 0$$

só se tem, se

$$nP(\min\{u_{n,1}, u_{n,2}\} < X_1 \leq \max\{u_{n,1}, u_{n,2}\}) \xrightarrow{n \rightarrow +\infty} 0. \quad (2.7)$$

Esta convergência (2.7) tem-se, por exemplo, quando os dois níveis também são normalizados para o mesmo número de excedências, ou seja, quando

$$\lim_{n \rightarrow +\infty} nP(X_1 > u_{n,1}) = \tau = \lim_{n \rightarrow +\infty} nP(X_1 > u_{n,2}),$$

para algum  $\tau > 0$ , uma vez que, nestas condições,

$$\begin{aligned} & P\left(N_n^{(T)}(u_{n,1})([0, 1]) \neq N_n^{(T)}(u_{n,2})([0, 1])\right) \\ & \leq nP(\min\{u_{n,1}, u_{n,2}\} < X_1 \leq \max\{u_{n,1}, u_{n,2}\}) \\ & = n(P(X_1 > \min\{u_{n,1}, u_{n,2}\}) - P(X_1 > \max\{u_{n,1}, u_{n,2}\})) \xrightarrow{n \rightarrow +\infty} 0. \end{aligned}$$

Tal significa que para níveis para os quais se tem a convergência (2.7), a sucessão de processos pontuais  $\{N_n^{(T)}\}_{n \geq 1}$  converge em distribuição para o mesmo processo pontual limite. Deste modo, para dois níveis  $u_n^{(1)}$  e  $\tilde{u}_{[n/\nu^{(T)}}^{(1)}$  podemos aplicar argumentos semelhantes aos usados na demonstração do Teorema 2.2.1 e obter o resultado seguinte.

**Teorema 2.2.2** *Suponhamos que, para cada  $\nu^{(T)} > 0$ , se verifica a condição  $\Delta(\tilde{\mathbf{u}}^{(\nu^{(T)})})$  para  $\mathbf{X}$ . Se a sucessão de processos pontuais de  $T$ -cruzamentos do nível  $\tilde{u}_n^{(1)}$  converge em distribuição para algum processo pontual  $N^{(T),(1)}$ , então, para qualquer  $\nu^{(T)} > 0$ , a sucessão de processos pontuais de  $T$ -cruzamentos de  $\tilde{u}_{[n/\nu^{(T)}}^{(1)}$  converge em distribuição para um processo de Poisson composto com transformada de Laplace dada por (2.3)-(2.5), com medida de intensidade  $\beta = \eta^{(T)} \nu^{(T)}$ , onde  $\eta^{(T)} \in [0, 1]$  representa a intensidade de  $N^{(T),(1)}$ ,*

$$0 \leq \eta^{(T)} \leq \left( \sum_{j \geq 1} j \pi(j) \right)^{-1} \leq 1,$$

a distribuição de multiplicidade  $\pi$  sobre  $\mathbb{N}$  e  $\eta^{(T)}$  independente de  $\nu^{(T)}$ . Se  $\tilde{u}_n^{(\nu^{(T)})}$  e  $\tilde{u}_{[n/\nu^{(T)}]}^{(1)}$  verificam

$$nP \left( \min\{\tilde{u}_n^{(\nu^{(T)})}, \tilde{u}_{[n/\nu^{(T)}]}^{(1)}\} < X_1 \leq \max\{\tilde{u}_n^{(\nu^{(T)})}, \tilde{u}_{[n/\nu^{(T)}]}^{(1)}\} \right) \xrightarrow{n \rightarrow +\infty} 0,$$

então surge o mesmo processo de Poisson composto como limite da sucessão de processos pontuais de  $T$ -cruzamentos de  $\tilde{u}_n^{(\nu^{(T)})}$ .

## 2.3 Cálculo do índice de $T$ -cruzamentos sob novas condições de dependência local

O parâmetro  $\eta^{(T)}$  que surge na caracterização da medida intensidade do processo de Poisson composto limite da sucessão de processos pontuais de  $T$ -cruzamentos de  $\tilde{u}_n^{(\nu^{(T)})}$ , apresentado no Teorema 2.2.2, será designado por índice de  $T$ -cruzamentos. A sua definição formal é dada a seguir.

**Definição 2.3.1** Se para cada  $\nu^{(T)} > 0$ , existir uma sucessão  $\tilde{\mathbf{u}}^{(\nu^{(T)})} = \{\tilde{u}_n^{(\nu^{(T)})}\}_{n \geq 1}$  nas condições (2.6) e tal que

$$\lim_{n \rightarrow +\infty} P(N_n^{(T)}([0, 1]) = 0) = e^{-\eta^{(T)}\nu^{(T)}}$$

para alguma constante  $\eta^{(T)}$ ,  $0 \leq \eta^{(T)} \leq 1$ , então diz-se que a sucessão  $\mathbf{X}$  tem índice de  $T$ -cruzamentos  $\eta^{(T)}$ .

Se a sucessão estacionária  $\mathbf{X}$  possuir índice de  $T$ -cruzamentos  $\eta^{(T)} > 0$ , então a distribuição condicional do tamanho de um grupo de  $T$ -cruzamentos  $\pi_n$  definida no Teorema 2.2.1 relaciona-se com  $\eta^{(T)} > 0$  de uma forma interessante.

**Teorema 2.3.1** Suponhamos que  $\mathbf{X}$  verifica  $\Delta(\tilde{\mathbf{u}}^{(\nu^{(T)})})$ ,  $\nu^{(T)} > 0$ , e seja  $\pi_n$  a distribuição definida no Teorema 2.2.1 com  $u_n = \tilde{u}_n^{(\nu^{(T)})}$ . Então

$$\lim_{n \rightarrow +\infty} \sum_{j \geq 1} j \pi_n(j) = \frac{1}{\eta^{(T)}}. \quad (2.8)$$

**Demonstração:** Começemos por notar que

$$\begin{aligned} \sum_{j \geq 1} j \pi_n(j) &= E[N_n^{(T)}(\tilde{u}_n^{(\nu^{(T)})})([0, r_n/n]) \mid N_n^{(T)}(\tilde{u}_n^{(\nu^{(T)})})([0, r_n/n]) > 0] \\ &= \frac{r_n P(X_1 \leq \tilde{u}_n^{(\nu^{(T)})} < X_{1+T})}{P(N_n^{(T)}(\tilde{u}_n^{(\nu^{(T)})})([0, r_n/n]) \neq 0)}. \end{aligned}$$

Agora, pelo Lema 2.2.1 se  $X$  tem índice de  $T$ -cruzamentos  $\eta^{(T)} > 0$ , então

$$\begin{aligned} \lim_{n \rightarrow +\infty} P^{k_n}(N_n^{(T)}(\tilde{u}_n^{(\nu^{(T)})})([0, r_n/n]) = 0) &= e^{-\eta^{(T)}\nu^{(T)}} \\ \Leftrightarrow \lim_{n \rightarrow +\infty} \left( 1 + \frac{-k_n P(N_n^{(T)}(\tilde{u}_n^{(\nu^{(T)})})([0, r_n/n]) \neq 0)}{k_n} \right)^{k_n} &= e^{-\eta^{(T)}\nu^{(T)}} \\ \Leftrightarrow \lim_{n \rightarrow +\infty} k_n P(N_n^{(T)}(\tilde{u}_n^{(\nu^{(T)})})([0, r_n/n]) \neq 0) &= \eta^{(T)}\nu^{(T)}. \end{aligned}$$

Logo,  $\lim_{n \rightarrow +\infty} \sum_{j \geq 1} j \pi_n(j) = \eta^{(T)}/(\eta^{(T)}\nu^{(T)})$  uma vez que  $\lim_{n \rightarrow +\infty} k_n r_n/n = 1$ , o que prova o pretendido.  $\square$

Este resultado diz-nos que podemos aproximar o tamanho médio dos agrupamentos de  $T$ -cruzamentos pelo inverso aritmético do índice de  $T$ -cruzamentos. Além do mais, desde que em (2.8) seja possível permutar o limite com a soma, o processo de Poisson composto limite de  $\{N_n^{(T)}(\tilde{u}_n^{(\nu^{(T)})})\}_{n \geq 1}$  terá uma multiplicidade média igual a  $1/\eta^{(T)}$ .

Convencionando que  $\eta^{(0)} \equiv \theta$  e  $\nu^{(0)} \equiv \tau$ , encontramos no próximo resultado a generalização de (1.26) onde demonstramos como se relacionam os índices de  $T$ -cruzamentos,  $\eta^{(T)}$ , para diferentes valores de  $T \geq 0$ .

**Teorema 2.3.2** *Seja  $X$  uma sucessão estacionária para a qual existe  $\mathbf{u} \equiv \tilde{\mathbf{u}}^{(\nu^{(T)})}$ , para todo  $\nu^{(T)} > 0$ . Se  $P(X_1 > \tilde{u}_n^{(\nu^{(T)})}) \xrightarrow{n \rightarrow +\infty} 0$  e  $\tilde{\mathbf{u}}^{(\nu^{(T)})} \equiv \tilde{\mathbf{u}}^{(\nu^{(T')})}$  para algum  $T' \neq T$  então o índice de  $T$ -cruzamentos existe se e só se existe o índice de  $T'$ -cruzamentos e tem-se*

$$\eta^{(T')} = \frac{\nu^{(T)}}{\nu^{(T')}} \eta^{(T)}. \quad (2.9)$$

**Demonstração:** Representemos por  $N_n^{(T)}(u_n)$  e  $N_n^{(T')}(u_n)$  os processos pontuais de cruzamentos de passo  $T$  e  $T'$ , respetivamente. Tem-se

$$\begin{aligned} \limsup_{n \rightarrow +\infty} P(N_n^{(T')}([0, 1]) > 0) &= \limsup_{n \rightarrow +\infty} P(N_n^{(T')}([0, 1]) > 0, X_1 \leq u_n, \dots, X_T \leq u_n) + \\ &+ \limsup_{n \rightarrow +\infty} P \left( N_n^{(T')}([0, 1]) > 0, \bigcup_{j=1}^T \{X_j > u_n\} \right) \\ &= \limsup_{n \rightarrow +\infty} P(N_n^{(T')}([0, 1]) > 0, X_1 \leq u_n, \dots, X_T \leq u_n) + o(1) \\ &\leq \limsup_{n \rightarrow +\infty} P(N_n^{(T)}([0, 1]) > 0) \end{aligned}$$

e

$$\begin{aligned}
& \limsup_{n \rightarrow +\infty} P(N_n^{(T)}([0, 1]) > 0) \\
&= \limsup_{n \rightarrow +\infty} P(N_n^{(T)}([0, 1]) > 0, X_1 \leq u_n, \dots, X_{T'} \leq u_n) + o(1) \\
&\leq \limsup_{n \rightarrow +\infty} P(N_n^{(T')}([0, 1]) > 0).
\end{aligned}$$

O mesmo raciocínio poderá ser feito para os limites inferiores. Portanto, existindo os índices, temos

$$\lim_{n \rightarrow +\infty} P(N_n^{(T')}([0, 1]) = 0) = \lim_{n \rightarrow +\infty} P(N_n^{(T)}([0, 1]) = 0),$$

ou equivalentemente,  $\nu^{(T')}\eta^{(T')} = \nu^{(T)}\eta^{(T)}$ . □

Este resultado dá-nos uma fórmula simples para o cálculo do índice de  $T$ -cruzamentos  $\eta^{(T)}$ , caso exista, a partir de, por exemplo, o índice de cruzamentos  $\eta^{(1)}$ .

Ilustramos esta relação entre os índices de  $T$ -cruzamentos com a sucessão extremal Markoviana ou processo autoregressivo de máximos (ARMAX), estudada em Alpuim (1989).

**Exemplo 2.3.1 (Processo ARMAX Fréchet)** *Seja  $X_0$  uma variável aleatória, com função de distribuição  $H_0$ , independente da sucessão de variáveis aleatórias i.i.d.  $\{Y_n\}_{n \geq 1}$  com f.d.  $G$ . Consideremos a sucessão Markoviana  $\mathbf{X} = \{X_n\}_{n \geq 1}$ , de Alpuim (1989), definida por*

$$X_n = k \max\{X_{n-1}, Y_n\}, \quad n \geq 1, \quad 0 < k < 1,$$

ou equivalentemente, por

$$X_n = \max \left\{ \max_{1 \leq i \leq n-1} k^i Y_{n-i+1}, k^{n-1} X_1 \right\}.$$

*Iremos supor que  $\mathbf{X}$  é uma sucessão estacionária, ou seja, como provou Alpuim (1989), que existe  $x > 0$  tal que*

$$G\left(\frac{x}{k}\right) > 0 \quad \text{e} \quad 0 < \sum_{j=1}^{+\infty} \left(1 - G\left(\frac{x}{k^j}\right)\right) < +\infty.$$

*Então a distribuição não degenerada  $H$  de  $X_n$ ,  $n \geq 1$ , satisfaz a equação*

$$G(x) = \frac{H(kx)}{H(x)}, \quad x \geq \alpha(H)/k.$$

onde  $\alpha(H) = \inf\{x : H(x) > 0\} \geq 0$ .

*A sucessão  $\mathbf{X}$  verifica a condição  $\Delta(\mathbf{u})$ , qualquer que seja a sucessão  $\mathbf{u} = \{u_n\}_{n \geq 1}$ , uma vez*

que Alpuim (1989) provou que  $\mathbf{X}$  verifica a condição de mistura forte, i.e.

$$m(k) = \sup\{|P(A \cap B) - P(A)P(B)| : A \in \mathcal{B}(X_1, \dots, X_p), B \in \mathcal{B}(X_{p+k+1}, X_{p+k+2}, \dots)\}$$

é tal que  $\lim_{k \rightarrow +\infty} m(k) = 0$ , onde  $\mathcal{B}(\dots, X_i, \dots)$  representa a tribo gerada pelas variáveis entre parêntesis.

Para  $u_n > 0$  verifica-se facilmente que

$$P(X_j \leq u_n < X_{j+1}) = P\left(X_1 \leq \frac{u_n}{k^{j-1}}, Y_2 \leq \frac{u_n}{k^{j-1}}, Y_3 \leq \frac{u_n}{k^{j-2}}, \dots, Y_j \leq \frac{u_n}{k}, Y_{j+1} > \frac{u_n}{k}\right).$$

Admitindo que  $H$  pertence ao domínio de atração de uma distribuição de Fréchet  $\Phi_\alpha$ ,  $\alpha \geq 1$ , temos que  $u_n^{(\tau)} = a_n \tau^{-1/\alpha}$  com  $a_n = H^{-1}(1 - 1/n)$  e  $\mathbf{X}$  verifica a condição  $D''(\mathbf{u}^{(\tau)})$  uma vez que

$$\begin{aligned} n \sum_{j=2}^{r_n-1} P(X_1 > u_n^{(\tau)}, X_j \leq u_n^{(\tau)} < X_{j+1}) \\ \leq nr_n(1 - H(u_n^{(\tau)}))^2 + nr_n(1 - H(u_n^{(\tau)}))(1 - H_0(u_n^{(\tau)})) = o(1). \end{aligned}$$

(ver H. Ferreira (1994)). Neste caso,  $\mathbf{X}$  tem índice extremal (Alpuim (1989) e H. Ferreira (1994))

$$\theta = 1 - k^\alpha, \quad 0 < k < 1, \quad \alpha \geq 1.$$

Para os outros domínios de atração  $\theta = 1$ .

Tendo agora em conta que,

$$\begin{aligned} \nu^{(1)} &= \lim_{n \rightarrow +\infty} nP(X_1 \leq u_n^{(\tau)} < X_2) \\ &= \lim_{n \rightarrow +\infty} nP(X_1 \leq u_n^{(\tau)}, Y_2 > u_n^{(\tau)}/k) \\ &= \lim_{n \rightarrow +\infty} nH(u_n^{(\tau)}) \left[1 - G(u_n^{(\tau)}/k)\right] \\ &= \lim_{n \rightarrow +\infty} n \left(1 - G(u_n^{(\tau)}/k)\right) \\ &= \lim_{n \rightarrow +\infty} n \left(1 - \frac{H(u_n^{(\tau)})}{H(u_n^{(\tau)}/k)}\right) \\ &= \lim_{n \rightarrow +\infty} n \left(H(u_n^{(\tau)}/k) - H(u_n^{(\tau)})\right) \\ &= \lim_{n \rightarrow +\infty} n \left[\left(1 - H(u_n^{(\tau)})\right) - \left(1 - H(u_n^{(\tau)}/k)\right)\right] \end{aligned}$$

$$\begin{aligned}
&= \lim_{n \rightarrow +\infty} n \left[ \left( 1 - H(a_n \tau^{-1/\alpha}) \right) - \left( 1 - H(a_n (\tau k^\alpha)^{-1/\alpha}) \right) \right] \\
&= \tau(1 - k^\alpha)
\end{aligned}$$

obtemos o índice de cruzamentos da sucessão  $\mathbf{X}$ ,

$$\eta^{(1)} = \frac{\tau\theta}{\tau(1 - k^\alpha)} = 1$$

como seria de esperar, uma vez que  $D''(\mathbf{u}^{(\tau)})$  se verifica para esta sucessão.

Generalizando, para cada  $T \geq 1$ , obtemos

$$\begin{aligned}
\nu^{(T)} &= \lim_{n \rightarrow +\infty} nP(X_1 \leq u_n^{(\tau)} < X_{1+T}) \\
&= \lim_{n \rightarrow +\infty} nP\left(X_1 \leq u_n^{(\tau)}, \max\{k^T X_1, k^T Y_2, \dots, k^T Y_{T+1}\} > u_n^{(\tau)}\right) \\
&= \lim_{n \rightarrow +\infty} nP\left(X_1 \leq u_n^{(\tau)}\right) - nP\left(X_1 \leq u_n^{(\tau)}, Y_1 \leq \frac{u_n^{(\tau)}}{k^T}, \dots, Y_{T+1} \leq \frac{u_n^{(\tau)}}{k}\right) \\
&= \lim_{n \rightarrow +\infty} nH\left(u_n^{(\tau)}\right) \left[ 1 - \frac{H(u_n^{(\tau)})}{H(u_n^{(\tau)}/k^T)} \right] \\
&= \lim_{n \rightarrow +\infty} n \left[ 1 - \frac{H(u_n^{(\tau)})}{H(u_n^{(\tau)}/k^T)} \right] \\
&= \lim_{n \rightarrow +\infty} n \left( H(u_n^{(\tau)}/k^T) - H(u_n^{(\tau)}) \right) \\
&= \lim_{n \rightarrow +\infty} n \left[ \left( 1 - H(u_n^{(\tau)}) \right) - \left( 1 - H(u_n^{(\tau)}/k^T) \right) \right] \\
&= \lim_{n \rightarrow +\infty} n \left[ \left( 1 - H(a_n \tau^{-1/\alpha}) \right) - \left( 1 - H(a_n (\tau k^{T\alpha})^{-1/\alpha}) \right) \right] \\
&= \tau(1 - k^{T\alpha})
\end{aligned}$$

e consequentemente por (2.9) vem que o índice de  $T$ -cruzamentos de  $\mathbf{X}$  é dado por

$$\eta^{(T)} = \frac{1 - k^\alpha}{1 - k^{T\alpha}}, \quad 0 < k < 1, \quad \alpha \geq 1. \quad (2.10)$$

Se  $H$  pertence ao domínio de atração de uma distribuição de Fréchet unitária, então  $\alpha = 1$  e os possíveis valores de  $\eta^{(T)}$  em função de diferentes valores de  $k$  e  $T$ , com  $0 < k < 1$  e  $T \geq 1$ , são apresentados na Tabela 2.1.

Tabela 2.1: Valores de  $\eta^{(T)}$  do processo ARMAX com distribuição geradora de Fréchet unitária para alguns valores de  $k$  e  $T$ .

		$T$						
		1	2	3	5	8	10	30
$k$	0.1	1.0000	0.9091	0.9009	0.9000	0.9000	0.9000	0.9000
	0.5	1.0000	0.6667	0.5714	0.5161	0.5020	0.5005	0.5000
	0.9	1.0000	0.5263	0.3690	0.2442	0.1756	0.1535	0.1000

Como podemos constatar pela Tabela 2.1, os valores de  $\eta^{(T)}$  diminuem à medida que  $k$  e  $T$  aumentam. Mais ainda, para  $k$  fixo e valores grandes de  $T$ , constatamos que  $\eta^{(T)} = \theta$ , o que seria de esperar tendo em conta (2.10). Tal significa que o número médio de cruzamentos a passo  $T$ , suficientemente grande, é assintoticamente igual ao número médio de excedências.

Para melhor ilustrarmos a riqueza destas sucessões estacionárias autoregressivas de máximos no que diz respeito ao agrupamento de  $T$ -cruzamentos, apresentamos na Figura 2.1 exemplos de amostras simuladas de modelos ARMAX com distribuição geradora de Fréchet unitária e  $k \in \{0.1, 0.5, 0.9\}$ .

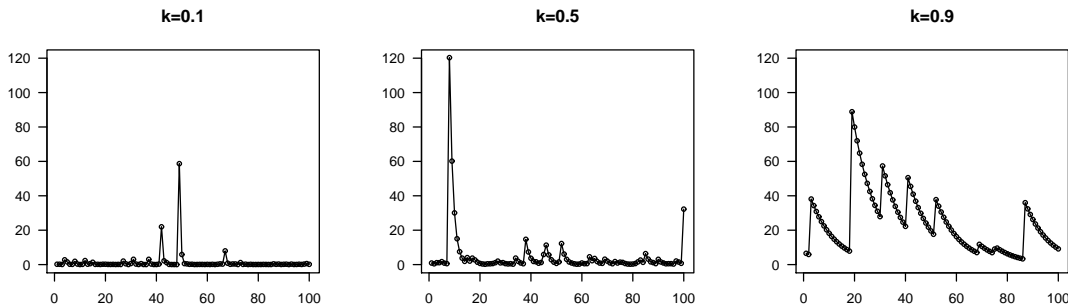


Figura 2.1: Amostras simuladas de um modelo ARMAX com distribuição geradora de Fréchet unitária, para  $k = 0.1$  (à esquerda),  $k = 0.5$  (ao centro) e  $k = 0.9$  (à direita).

Notemos que à medida que o valor de  $k$  aumenta encontramos mais frequentemente grupos de  $T$ -cruzamentos de níveis elevados.

Consideramos a seguir uma condição de dependência local que permitirá calcular o índice de  $T$ -cruzamentos a partir da distribuição conjunta de um número finito de variáveis da sucessão  $\mathbf{X}$ . Iremos denotar  $N_n^{(T)}(u_n) \left( \left[ \frac{i}{n}, \frac{j}{n} \right] \right)$  simplesmente por  $N_{i,j}^{(T)}$  e convencionamos que  $N_{i,j}^{(T)} = 0$  se  $j < i$ . No caso de ser necessário identificar a sucessão  $\mathbf{X}$  envolvida no processo pontual, escreveremos  $N_{i,j}^{(T,\mathbf{X})}$ .

**Definição 2.3.2** Seja  $\mathbf{X}$  uma sucessão estacionária verificando a condição  $D(\mathbf{u})$ . Para qualquer  $k \geq 2$ ,  $\mathbf{X}$  satisfaz a condição  $\tilde{D}_T^{(k)}(\mathbf{u})$  se

$$\lim_{n \rightarrow +\infty} nP(X_1 \leq u_n < X_{1+T}, N_{2,k}^{(T)} = 0, N_{k+1,r_n}^{(T)} > 0) = 0,$$

para alguma sucessão  $\{r_n = [n/k_n]\}_{n \geq 1}$  com  $k$  nas condições de (1.7).

Notemos que para  $T = 1$ , obtemos condições equivalentes às condições  $\tilde{D}^{(k)}(\mathbf{u})$ ,  $k \geq 1$ , de H. Ferreira (2006), uma vez que os acontecimentos  $\{N_{1,1}^{(1)} = 1\}$  e  $\{N_{2,2}^{(1)} = 1\}$  são incompatíveis, logo os resultados seguintes generalizam os obtidos em H. Ferreira (2006).

Toda a sucessão que verifique a condição  $D''(\mathbf{u})$ , verifica também a condição  $\tilde{D}_T^{(k)}(\mathbf{u})$  para qualquer escolha de  $T \geq 1$  e  $k > 2$ , uma vez que

$$\begin{aligned} & \lim_{n \rightarrow +\infty} nP(X_1 \leq u_n < X_{1+T}, N_{2,k}^{(T)} = 0, N_{k+1,r_n}^{(T)} > 0) \\ & \leq \lim_{n \rightarrow +\infty} n \sum_{i=1}^T P(X_i \leq u_n < X_{i+1}, N_{i+1,r_n}^{(1)} > 0). \end{aligned}$$

Como exemplo destacamos a sucessão ARMAX do Exemplo 2.3.1.

O resultado que apresentamos a seguir mostra que a condição de dependência local  $\tilde{D}_T^{(k)}(\mathbf{u})$  é necessária e suficiente para obter  $\lim_{n \rightarrow +\infty} P(N_n^{(T)}(u_n)([0, 1]) = 0)$ , a partir da distribuição conjunta de um número finito de variáveis da sucessão.

**Teorema 2.3.3** *Suponhamos que  $\mathbf{X}$  verifica a condição  $D(\mathbf{u})$  e que  $\liminf_{n \rightarrow +\infty} P(N_n^{(T)}(u_n)([0, 1]) = 0) > 0$ . Então, para qualquer inteiro  $k \geq 2$*

$$P(N_n^{(T)}(u_n)([0, 1]) = 0) - \exp(-nP(X_1 \leq u_n < X_{1+T}, N_{2,k}^{(T)} = 0)) \xrightarrow{n \rightarrow +\infty} 0 \quad (2.11)$$

se e só se  $\tilde{D}_T^{(k)}(\mathbf{u})$  se verifica para  $\mathbf{X}$ .

**Demonstração:** Substituindo excedências por  $T$ -cruzamentos nos argumentos usados por O'Brien (1987) para demonstrar (1.11), obtemos a seguinte convergência para a probabilidade de não haver nenhum  $T$ -cruzamento,

$$P(N_n^{(T)}(u_n)([0, 1]) = 0) - \exp(-nP(X_1 \leq u_n < X_{1+T}, N_{2,r_n}^{(T)} = 0)) \xrightarrow{n \rightarrow +\infty} 0.$$

Concluindo-se o pretendido do facto de

$$\begin{aligned} & nP(X_1 \leq u_n < X_{1+T}, N_{2,r_n}^{(T)} = 0) \\ & = nP(X_1 \leq u_n < X_{1+T}, N_{2,k}^{(T)} = 0, N_{k+1,r_n}^{(T)} = 0) \\ & = nP(X_1 \leq u_n < X_{1+T}, N_{2,k}^{(T)} = 0) + o(1) \end{aligned}$$

se e só se  $\mathbf{X}$  verifica a condição  $\tilde{D}_T^{(k)}(\mathbf{u})$ . □

O limite de  $P(N_n^{(T)}(u_n)([0, 1]) = 0)$  pode assim ser obtido a partir do limite do número médio de  $T$ -cruzamentos, entre as  $n$  primeiras variáveis de  $\mathbf{X}$ , seguidos de  $k - 1$  não  $T$ -cruzamentos.

Como consequência imediata deste resultado obtemos a seguinte fórmula simples para o cálculo do índice de  $T$ -cruzamentos de uma sucessão estacionária  $\mathbf{X}$  que verifica a condição  $\tilde{D}_T^{(k)}(\tilde{\mathbf{u}}^{(\nu^{(T)})})$ ,  $k \geq 2$ .

**Corolário 2.3.1** *Se  $\mathbf{X}$  verifica a condição  $D(\mathbf{u})$  e, para algum  $k \geq 2$ , também verifica a condição  $\tilde{D}_T^{(k)}(\tilde{\mathbf{u}}^{(\nu^{(T)})})$  para cada  $\nu^{(T)} > 0$ , então o índice de  $T$ -cruzamentos de  $\mathbf{X}$  existe e é igual a  $\eta^{(T)}$  se e só se*

$$P(N_{2,k}^{(T)}(\tilde{u}_n^{(\nu^{(T)})}) = 0 \mid X_1 \leq \tilde{u}_n^{(\nu^{(T)})} < X_{1+T}) \xrightarrow{n \rightarrow +\infty} \eta^{(T)}, \quad (2.12)$$

para cada  $\nu^{(T)} > 0$ .

**Demonstração:** Se existe  $\eta^{(T)} > 0$  então, por (2.11), verifica-se a convergência (2.12). Se (2.12) se verifica, então necessariamente teremos que  $\liminf_{n \rightarrow \infty} P(N_n^{(T)}(\tilde{\mathbf{u}}^{(\nu^{(T)})}) = 0) > 0$ , pois caso contrário  $P(N_n^{(T)}(\tilde{\mathbf{u}}^{(\nu^{(T)})}) = 0) \xrightarrow{n \rightarrow +\infty} 0$  ao longo de alguma subsucessão de  $\mathbb{N}$  e consequentemente

$$nP(X_1 \leq \tilde{u}_n^{(\nu^{(T)})} < X_{1+T}, N_{2,k}^{(T)}(\tilde{u}_n^{(\nu^{(T)})}) = 0) \xrightarrow{n \rightarrow +\infty} +\infty \neq \eta^{(T)} \nu^{(T)}$$

ao longo dessa subsucessão. Assim, estamos em condições de aplicar o Teorema 2.3.3 e concluir que

$$\lim_{n \rightarrow +\infty} P(N_n^{(T)}(\tilde{u}_n^{(\nu^{(T)})}) = 0) = e^{-\eta^{(T)} \nu^{(T)}},$$

onde resulta o pretendido. □

A facilidade de aplicação deste resultado ficará patente quando, na próxima secção, analisarmos detalhadamente um processo autoregressivo de primeira ordem que verifica a condição  $\tilde{D}_3^{(7)}(\mathbf{u})$ .

## 2.4 Cálculo do índice de cruzamentos após subamostragem sistemática

De modo a analisarmos o efeito que a subamostragem sistemática exerce sobre o valor do índice de cruzamentos, considere-se a sucessão  $\mathbf{Y} = \{Y_n = X_{1+(n-1)T}\}_{n \geq 1}$ , subamostrada a partir de  $\mathbf{X}$  em intervalos fixos de  $T \geq 1$  unidades. Notemos que quando  $T = 1$  todos os valores da sucessão  $\mathbf{X}$  serão subamostrados, coincidindo neste caso ambas as sucessões.

Uma vez que agora estão envolvidas várias sucessões, teremos que adaptar a notação em função da sucessão que se pretenda utilizar. Por exemplo,  $N_n^{(T, \mathbf{X})}(u_n)$  e  $\nu^{(T, \mathbf{X})}$  representam, respetivamente,  $N_n^{(T)}(u_n)$  e  $\nu^{(T)}$  para a sucessão  $\mathbf{X}$ .

Suponhamos que  $\mathbf{X}$  verifica a condição  $\Delta(\mathbf{u})$  e sejam  $A$  e  $B$  acontecimentos pertencentes, respetivamente, às tribos geradas por  $\{Y_s \leq u_{nT}\}$ ,  $s = 1, \dots, k$  e  $\{Y_s \leq u_{nT}\}$ ,

$s = k + l, \dots, n$ . Então  $A$  e  $B$  pertencem, respetivamente, às tribos geradas por  $\{X_s \leq u_{nT}\}$ ,  $s = 1, \dots, 1 + (k - 1)T$  e  $\{X_s \leq u_{nT}\}$ ,  $s = 1 + (k + l - 1)T, \dots, 1 + (n - 1)T$  e portanto

$$\alpha_{n, l_{nT}}^{(\mathbf{Y})} \leq \alpha_{nT, l_{nT}T}^{(\mathbf{X})} \leq \alpha_{nT, l_{nT}}^{(\mathbf{X})} = o(1),$$

uma vez que na definição da condição  $\Delta(\mathbf{u})$  os coeficientes  $\alpha_{n, l}$ ,  $n \geq 1$ , podem ser tomados como sendo não crescentes em  $l$ . Além disso,  $l_{nT} = o(n)$  se  $l_n = o(n)$  e conseqüentemente  $\mathbf{Y}$  verifica a condição  $\Delta(\mathbf{v})$  com  $\mathbf{v} = \{v_n = u_{nT}\}_{n \geq 1}$  e  $l_n^{(\mathbf{Y})} = l_{nT}^{(\mathbf{X})}$ .

Uma comparação análoga poderá ser feita para os coeficientes de dependência  $\alpha_{n, l}$  associados à condição mais fraca  $D(\mathbf{u})$ .

Introduzimos de seguida uma condição de dependência local que implica  $\tilde{D}_T^{(1+2T)}(\mathbf{u})$  para  $\mathbf{X}$  e  $\tilde{D}_1^{(3)}(\mathbf{v})$  para  $\mathbf{Y}$ , com  $\mathbf{v}$  definida anteriormente. Esta nova condição vai permitir-nos aplicar os resultados da secção anterior e calcular os respetivos índices de  $T$ -cruzamentos de  $\mathbf{X}$  e  $\mathbf{Y}$ ,  $\eta^{(T, \mathbf{X})}$  e  $\eta^{(T, \mathbf{Y})}$ , a partir da distribuição conjunta de um número finito de variáveis e relacionar os valores destes dois índices.

**Definição 2.4.1** *Seja  $\mathbf{X}$  uma sucessão estacionária verificando  $D(\mathbf{u})$ . Para  $k \geq 2$ ,  $\mathbf{X}$  satisfaz a condição  $\overline{D}_T^{(k)}(\mathbf{u})$  se*

$$\lim_{n \rightarrow +\infty} nP(X_1 \leq u_n < X_{1+T}, N_{k, k}^{(T)} = 0, N_{k+1, r_n}^{(T)} > 0) = 0$$

para alguma sucessão  $\{r_n = [n/k_n]\}_{n \geq 1}$  com  $\mathbf{k} = \{k_n\}_{n \geq 1}$  nas condições de (1.7).

Notemos que toda a sucessão que verifica a condição  $\overline{D}_T^{(k)}(\mathbf{u})$ , para algum  $k \geq 2$ , também verifica a condição  $\tilde{D}_T^{(k)}(\mathbf{u})$ . Mais ainda, o próximo resultado garante-nos que se  $\mathbf{X}$  verifica  $\overline{D}_T^{(1+2T)}(\mathbf{u})$  então  $\mathbf{Y}$  verifica necessariamente  $\tilde{D}_1^{(3)}(\mathbf{v})$ , com  $\mathbf{v} = \{v_n = u_{nT}\}_{n \geq 1}$ .

**Teorema 2.4.1** *Se  $\mathbf{X}$  verifica  $\overline{D}_T^{(1+2T)}(\mathbf{u})$  para alguma sucessão  $\mathbf{k} = \{k_n\}_{n \geq 1}$  tal que*

$$k_{[n/T]} \xrightarrow{n \rightarrow +\infty} +\infty, \quad \frac{k_{[n/T]} l_{nT}^{(\mathbf{X})}}{n} \xrightarrow{n \rightarrow +\infty} 0 \quad \text{e} \quad k_{[n/T]} \alpha_{n, l_{nT}}^{(\mathbf{X})} \xrightarrow{n \rightarrow +\infty} 0,$$

então a sucessão  $\mathbf{Y}$  verifica  $\tilde{D}_1^{(3)}(\mathbf{v})$  com  $\mathbf{v} = \{v_n = u_{nT}\}_{n \geq 1}$  e  $\mathbf{k}' = \{k'_n = k_{[n/T]}\}_{n \geq 1}$ .

**Demonstração:** Atendendo a que  $\overline{D}_T^{(1+2T)}(\mathbf{u})$  se verifica para  $\mathbf{X}$ , obtemos

$$\begin{aligned} & nP(Y_1 \leq v_n < Y_2, N_{3,3}^{(1, \mathbf{Y})} = 0, N_{4, r_n}^{(1, \mathbf{Y})} > 0) \\ &= nP(X_1 \leq u_{nT} < X_{1+T}, N_{1+2T, 1+2T}^{(T, \mathbf{X})} = 0, N_{1+3T, 1+(r_n-1)T}^{(T, \mathbf{X})} > 0) \\ &\leq nTP(X_1 \leq u_{nT} < X_{1+T}, N_{1+2T, 1+2T}^{(T, \mathbf{X})} = 0, N_{1+3T, [nT/k_n]}^{(T, \mathbf{X})} > 0) \\ &= o(1). \end{aligned}$$

Além do mais, a sucessão  $k'$  está nas condições da definição da condição  $\tilde{D}_1^{(3)}(\mathbf{v})$ , uma vez que

$$k'_n = k_{[n/T]} \xrightarrow{n \rightarrow +\infty} +\infty, \quad \frac{k'_n \alpha_{n, l_n}^{(\mathbf{Y})}}{n} \leq \frac{k_{[n/T]} \alpha_{n, l_{nT}}^{(\mathbf{X})}}{n} \xrightarrow{n \rightarrow +\infty} 0$$

e

$$k'_n \alpha_{n, l_n}^{(\mathbf{Y})} \leq k_{[n/T]} \alpha_{n, l_{nT}}^{(\mathbf{X})} \xrightarrow{n \rightarrow +\infty} 0.$$

□

Estão finalmente reunidas as condições que nos permitem relacionar o índice de cruzamentos da sucessão subamostrada  $\mathbf{Y}$ ,  $\eta^{(1, \mathbf{Y})}$ , com o índice de  $T$ -cruzamentos da sucessão inicial  $\mathbf{X}$ ,  $\eta^{(T, \mathbf{X})}$ .

**Teorema 2.4.2** *Sob a validade das condições estabelecidas no Teorema 2.4.1 e se existir  $\eta^{(T, \mathbf{X})}$ , então existe  $\eta^{(1, \mathbf{Y})}$  se e só se existe*

$$\gamma_T = \lim_{n \rightarrow +\infty} \frac{nP(X_1 \leq u_{nT} < X_{1+T}, N_{2, 2T}^{(T, \mathbf{X})} > 0, N_{1+2T, 1+2T}^{(T, \mathbf{X})} = 0)}{\nu^{(T, \mathbf{X})}/T}$$

e, neste caso,  $\eta^{(1, \mathbf{Y})} = \eta^{(T, \mathbf{X})} + \gamma_T$ .

**Demonstração:** Como já referimos, se a condição  $\overline{D}_T^{(1+2T)}(\mathbf{u})$  se verifica para  $\mathbf{X}$ , então também se verifica a condição  $\tilde{D}_T^{(1+2T)}(\mathbf{u})$  para esta sucessão e conseqüentemente

$$\eta^{(T, \mathbf{X})} = \lim_{n \rightarrow +\infty} \frac{P(X_1 \leq u_n < X_{1+T}, N_{2, 1+2T}^{(T, \mathbf{X})} = 0)}{P(X_1 \leq u_n < X_{1+T})}.$$

Mais ainda,  $\tilde{D}_1^{(3)}(\mathbf{v})$  verifica-se para  $\mathbf{Y}$  e portanto

$$\begin{aligned} \eta^{(1, \mathbf{Y})} &= \lim_{n \rightarrow +\infty} \frac{P(Y_1 \leq u_{nT} < Y_2, N_{3, 3}^{(1, \mathbf{Y})} = 0)}{P(Y_1 \leq u_{nT} < Y_2)} \\ &= \lim_{n \rightarrow +\infty} \frac{P(X_1 \leq u_{nT} < X_{1+T}, N_{1+2T, 1+2T}^{(T, \mathbf{X})} = 0)}{P(X_1 \leq u_{nT} < X_{1+T})} \\ &= \lim_{n \rightarrow +\infty} \frac{P(X_1 \leq u_{nT} < X_{1+T}, N_{2, 1+2T}^{(T, \mathbf{X})} = 0)}{P(X_1 \leq u_{nT} < X_{1+T})} + \\ &\quad + \lim_{n \rightarrow +\infty} \frac{nTP(X_1 \leq u_{nT} < X_{1+T}, N_{2, 2T}^{(T, \mathbf{X})} > 0, N_{1+2T, 1+2T}^{(T, \mathbf{X})} = 0)}{nTP(X_1 \leq u_{nT} < X_{1+T})} \\ &= \eta^{(T, \mathbf{X})} + \gamma_T. \end{aligned}$$

□

Se considerarmos  $T = 1$  no resultado anterior vem  $\gamma_T = 0$ , como seria de esperar pois neste

caso as duas sucessões  $\mathbf{X}$  e  $\mathbf{Y}$  coincidam.

Ilustramos a seguir os resultados desta secção com o processo autoregressivo de primeira ordem,  $\text{AR}(1)$ , negativamente correlacionado, considerado em Chernick *et al.* (1991). Em Chernick e Davis (1982) encontramos o estudo do comportamento extremal deste processo.

**Exemplo 2.4.1 (Processo  $\text{AR}(1)$ )** Consideremos a sucessão estacionária autoregressiva  $\mathbf{X} = \{X_n\}_{n \geq 1}$ , definida em Chernick *et al.* (1991) por

$$X_n = -\frac{1}{r} X_{n-1} + \varepsilon_n, \quad n \geq 1, \quad (2.13)$$

onde  $\{\varepsilon_n\}_{n \geq 1}$  é uma sucessão de variáveis i.i.d. tais que, para um inteiro fixo  $r \geq 2$ ,  $\varepsilon_n \sim U\left\{\frac{1}{r}, \dots, \frac{r-1}{r}, 1\right\}$  e  $X_0 \sim U\{0, 1\}$  é independente de  $\varepsilon_n$ .

Chernick *et al.* (1991) provaram que este processo autoregressivo de primeira ordem, negativamente correlacionado, satisfaz as condições  $D(\mathbf{u})$  e  $D^{(3)}(\mathbf{u})$  com  $\mathbf{u} = \{u_n = 1 - x/n\}_{n \geq 1}$ ,  $x > 0$ ,  $k_n = [n^{3/4}]$ ,  $l_n = [n^{1/8}]$ ,  $\alpha_{n, l_n} = \left(\frac{1}{r}\right)^{l_n} \left(1 - \frac{1}{r}\right)$  e tem índice extremal  $\theta^{(\mathbf{X})} = 1 - 1/r^2$ . Outra característica deste processo é que  $nP(X_i > u_n, X_{i+1} > u_n) \xrightarrow[n \rightarrow +\infty]{} 0$ , uma vez que para  $n$  suficientemente grande,

$$\begin{aligned} P(X_1 \leq u_n < X_2) &= P\left(\varepsilon_1 \leq 1 - \frac{x}{n} + \frac{X_0}{r}, -\frac{1}{r}\varepsilon_1 + \varepsilon_2 > 1 - \frac{x}{n} - \frac{X_0}{r^2}\right) \\ &= \int_0^1 P\left(\varepsilon_1 \leq 1 - \frac{x}{n} + \frac{u}{r}, -\frac{1}{r}\varepsilon_1 + \varepsilon_2 > 1 - \frac{x}{n} - \frac{u}{r^2}\right) du \\ &= \int_{1 - \frac{r^2 x}{n}}^1 P\left(\varepsilon_2 - \frac{1}{r}\varepsilon_1 > 1 - \frac{x}{n} - \frac{u}{r^2}\right) du \\ &= r^2 \frac{x}{n} P\left(\varepsilon_2 = 1, \varepsilon_1 = \frac{1}{r}\right) = \frac{x}{n}. \end{aligned} \quad (2.14)$$

No próximo resultado mostramos que  $\mathbf{X}$  também satisfaz a condição  $\tilde{D}_1^{(3)}(\mathbf{u})$ , obtemos o índice de cruzamentos deste processo e provamos que a distribuição condicional do tamanho dos *runs* de 1-cruzamentos, ou simplesmente cruzamentos, converge para uma distribuição geométrica.

**Teorema 2.4.3** Seja  $\mathbf{X} = \{X_n\}_{n \geq 1}$  uma sucessão definida por (2.13) e  $\mathbf{u} = \{u_n = 1 - \tau/n\}_{n \geq 1}$  com  $\tau > 0$ . Então

- i)  $\mathbf{u} = \tilde{\mathbf{u}}^{(\nu^{(1)})}$  com  $\nu^{(1)} = \tau$ ;
- ii)  $\mathbf{X}$  verifica a condição  $\tilde{D}_1^{(3)}(\mathbf{u})$ ;
- iii)  $\eta^{(1, \mathbf{X})} = 1 - 1/r^2$ ;

$$\begin{aligned}
\text{iv)} \quad \lim_{n \rightarrow +\infty} P\left(\bigcap_{k=1}^j \{X_{2k+1} \leq u_n < X_{2k+2}\}, N_{2j+3, 2j+3}^{(1, \mathbf{X})} = 0 \mid N_{1,1}^{(1, \mathbf{X})} = 0, X_3 \leq u_n < X_4\right) = \\
= \left(\frac{1}{r^2}\right)^{j-1} \left(1 - \frac{1}{r^2}\right), \quad j \geq 1.
\end{aligned}$$

**Demonstração:** i) É consequência imediata de (2.14).

ii) Em vez de usarmos a definição de  $\tilde{D}_1^{(3)}(\mathbf{u})$  vamos aplicar o Teorema 2.3.3 uma vez que, neste caso, já conhecemos o valor do índice extremal de  $\mathbf{X}$ ,  $\theta^{(\mathbf{X})} = 1 - 1/r^2$  e por i) sabemos que  $\nu^{(1)} = \tau$ . Assim,

$$\begin{aligned}
\lim_{n \rightarrow +\infty} P(N_n^{(1, \mathbf{X})}(u_n)([0, 1]) = 0) &= \lim_{n \rightarrow +\infty} P(X_1 \leq u_n, \dots, X_n \leq u_n) \\
&= \exp\left(-\left(1 - \frac{1}{r^2}\right)\tau\right)
\end{aligned}$$

e

$$\begin{aligned}
&\lim_{n \rightarrow +\infty} nP(X_1 \leq u_n < X_2, N_{3,3}^{(1, \mathbf{X})} = 0) \\
&= \lim_{n \rightarrow +\infty} nP(N_{1,1}^{(1, \mathbf{X})} = 0, X_3 \leq u_n < X_4) \\
&= \lim_{n \rightarrow +\infty} nP(X_1 \leq u_n, X_2 \leq u_n, X_3 \leq u_n < X_4) + \\
&\quad + \lim_{n \rightarrow +\infty} nP(X_1 > u_n, X_2 \leq u_n, X_3 \leq u_n < X_4) + \\
&\quad + \lim_{n \rightarrow +\infty} nP(X_1 > u_n, X_2 > u_n, X_3 \leq u_n < X_4) \\
&= \lim_{n \rightarrow +\infty} nP(X_1 \leq u_n, X_2 \leq u_n, X_3 \leq u_n < X_4) \\
&= \lim_{n \rightarrow +\infty} nP(X_2 \leq u_n, X_3 \leq u_n < X_4) - \\
&\quad - \lim_{n \rightarrow +\infty} nP(X_1 > u_n, X_2 \leq u_n, X_3 \leq u_n < X_4) \\
&= \lim_{n \rightarrow +\infty} nP(X_1 \leq u_n, X_2 \leq u_n < X_3) \\
&= \lim_{n \rightarrow +\infty} nP(X_3 > u_n)P(X_1 \leq u_n, X_2 \leq u_n \mid X_3 > u_n) \\
&= \left(1 - \frac{1}{r^2}\right)\tau,
\end{aligned}$$

pois  $\mathbf{X}$  verifica  $D^{(3)}(\mathbf{u})$  e  $nP(X_i > u_n, X_{i+1} > u_n) \xrightarrow{n \rightarrow +\infty} 0$ .

Logo

$$P(N_n^{(1, \mathbf{X})}(u_n)([0, 1]) = 0) - \exp(nP(X_1 \leq u_n < X_2, N_{3,3}^{(1, \mathbf{X})} = 0)) \xrightarrow{n \rightarrow +\infty} 0$$

e pelo Teorema 2.3.3 concluímos que  $\mathbf{X}$  verifica a condição  $\tilde{D}_1^{(3)}(\mathbf{u})$ .

iii) Resulta imediatamente de i) e do limite obtido em ii).

iv) Tendo em conta o obtido em ii), é imediato que para  $j \geq 1$

$$\begin{aligned} & \lim_{n \rightarrow +\infty} P\left(\bigcap_{k=1}^j \{X_{2k+1} \leq u_n < X_{2k+2}\}, N_{2j+3, 2j+3}^{(1, \mathbf{X})} = 0 \mid N_{1,1}^{(1, \mathbf{X})} = 0, X_3 \leq u_n < X_4\right) \\ &= \lim_{n \rightarrow +\infty} \frac{n}{\tau} \frac{r^2}{r^2 - 1} P(X_1 \leq u_n, X_2 \leq u_n < X_3, \dots, X_{2j} \leq u_n < X_{2j+1}, X_{2j+3} \leq u_n). \end{aligned}$$

Para obter o limite do numerador, consideremos a seguinte representação para  $\mathbf{X}$ . Para  $j \geq 2$ ,

$$X_{2j-1} = \sum_{\substack{k=3 \\ k \text{ ímpar}}}^{2j-1} -\frac{1}{r^{2j-k}} \left( \varepsilon_{k-1} - \frac{1}{r} \varepsilon_{k-2} \right) + \varepsilon_{2j-1} - \frac{1}{r^{2j-1}} X_0$$

e

$$X_{2j} = \sum_{\substack{k=2 \\ k \text{ par}}}^{2j} \frac{1}{r^{2j-k}} \left( \varepsilon_k - \frac{1}{r} \varepsilon_{k-1} \right) + \frac{1}{r^{2j}} X_0.$$

Aplicando esta representação obtemos, para  $n$  suficientemente grande,

$$\begin{aligned} & P(X_1 \leq u_n, X_2 \leq u_n < X_3, \dots, X_{2j} \leq u_n < X_{2j+1}, X_{2j+3} \leq u_n) \\ &= \int_{1 - \frac{r^2 \tau}{n}}^{1 - \frac{\tau}{n}} P\left(\varepsilon_1 \leq 1 - \frac{\tau}{n} + \frac{x}{r}, -\frac{1}{r} \varepsilon_1 + \varepsilon_2 > 1 - \frac{\tau}{n} - \frac{x}{r^2}, \dots, \right. \\ & \quad \left. \dots, \varepsilon_{2j-1} \leq 1 - \frac{\tau}{n} + \frac{x}{r^{2j-1}} + \sum_{\substack{k=3 \\ k \text{ ímpar}}}^{2j-1} -\frac{1}{r^{2j-k}} \left( \varepsilon_{k-1} - \frac{1}{r} \varepsilon_{k-2} \right), \right. \\ & \quad \left. \varepsilon_{2j} - \frac{1}{r} \varepsilon_{2j-1} > 1 - \frac{\tau}{n} + \frac{x}{r^{2j}} - \sum_{\substack{k=2 \\ k \text{ par}}}^{2j} \frac{1}{r^{2j-k}} \left( \varepsilon_k - \frac{1}{r} \varepsilon_{k-1} \right), \right. \\ & \quad \left. \varepsilon_{2j+2} - \frac{1}{r} \varepsilon_{2j+1} \leq 1 - \frac{\tau}{n} + \frac{x}{r^{2j+2}} - \sum_{k=2}^{2j+2} \frac{1}{r^{2j-k}} \left( \varepsilon_k - \frac{1}{r} \varepsilon_{k-1} \right) \right) dx \end{aligned}$$

$$\begin{aligned}
&= \int_{1-\frac{\tau}{n}}^{1-\frac{\tau}{n}} P\left(\varepsilon_1 = \frac{1}{r}, \varepsilon_2 = 1, \varepsilon_3 = \frac{1}{r}, \varepsilon_4 = 1, \dots, \varepsilon_{2j-1} = \frac{1}{r}, \varepsilon_{2j} = 1\right) \times \\
&\quad \times \left(1 - P\left(\varepsilon_{2j+2} = \frac{1}{r}, \varepsilon_{2j+1} = 1\right)\right) dx \\
&= \frac{\tau}{n}(r^2 - 1) \left(\frac{1}{r^2}\right)^j \left(1 - \frac{1}{r^2}\right),
\end{aligned}$$

o que permite obter o limite expresso em iv). □

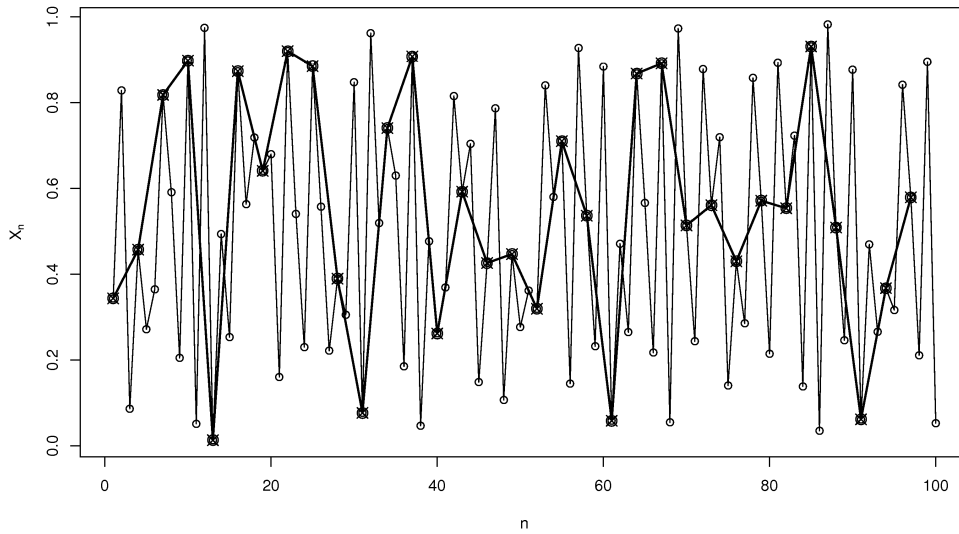


Figura 2.2: Trajetórias do processo  $AR(1)$  com  $r = 2$  e  $n = 100$  e do processo obtido por subamostragem simples com  $T = 3$  (a negroito).

Salientamos que iii) e iv) permitem concluir que a distribuição condicional do tamanho dos *runs* de cruzamentos converge para uma distribuição geométrica de valor médio igual a  $1/\eta^{(1, \mathbf{X})} = 1/(1 - 1/r^2)$ .

Para ilustrarmos os resultados da Secção 2.4 efetuamos agora uma subamostragem sistemática deste processo com passo  $T = 3$ . A sucessão subamostrada é agora definida por  $\mathbf{Y} = \{Y_n = X_{1+(n-1)3}\}_{n \geq 1}$  e encontra-se representada na Figura 2.2 a negroito.

**Teorema 2.4.4** *Seja  $\mathbf{X}$  a sucessão definida em (2.13),  $\mathbf{u} = \{u_n = 1 - \tau/n\}_{n \geq 1}$  para algum  $\tau > 0$ ,  $\mathbf{Y} = \{X_{1+(n-1)3}\}_{n \geq 1}$  e  $\mathbf{v} = \{u_{3n}\}_{n \geq 1}$ .*

Então

i)  $\mathbf{X}$  verifica  $\overline{D}_3^{(\tau)}(\mathbf{u})$ ;

ii)  $\mathbf{Y}$  verifica  $\tilde{D}_1^{(3)}(\mathbf{v})$ ;

$$\text{iii)} \eta^{(3, \mathbf{X})} = 1 - 1/r^2;$$

$$\text{iv)} \eta^{(1, \mathbf{Y})} = \eta^{(3, \mathbf{X})} + \left(1 - \frac{1}{r^2}\right) \left(\frac{1}{r^2} + \left(\frac{1}{r^2}\right)^2\right) = \left(1 - \frac{1}{r^2}\right) \left(1 + \frac{1}{r^2} + \left(\frac{1}{r^2}\right)^2\right).$$

**Demonstração:** i) Como  $\mathbf{X}$  verifica a condição  $\tilde{D}_1^{(3)}(\mathbf{u})$  e  $nP(X_i > u_n, X_{i+1} > u_n) \xrightarrow{n \rightarrow +\infty} 0$  temos que

$$\begin{aligned} & nP(X_1 \leq u_n < X_4, N_{7,7}^{(3, \mathbf{X})} = 0, N_{8,r_n}^{(3, \mathbf{X})} > 0) \\ &= nP(X_1 \leq u_n < X_4, X_5 \leq u_n, X_6 \leq u_n, N_{7,7}^{(3, \mathbf{X})} = 0, N_{8,r_n}^{(3, \mathbf{X})} > 0) + \\ & \quad + nP(X_1 \leq u_n < X_4, X_5 \leq u_n < X_6, N_{7,7}^{(3, \mathbf{X})} = 0, N_{8,r_n}^{(3, \mathbf{X})} > 0) + o(1) \\ &\leq nP(X_3 \leq u_n < X_4, X_5 \leq u_n, X_6 \leq u_n, N_{7,r_n}^{(3, \mathbf{X})} > 0) + \\ & \quad + nP(X_3 \leq u_n < X_4, X_5 \leq u_n < X_6, X_7 \leq u_n, X_{10} \leq u_n, N_{8,r_n}^{(3, \mathbf{X})} > 0) + o(1) = o(1), \end{aligned}$$

donde se conclui que  $\mathbf{X}$  também verifica a condição  $\bar{D}_3^{(7)}(\mathbf{u})$ .

iii) e iv) Atendendo ao Teorema 2.4.2 podemos calcular  $\eta^{(1, \mathbf{Y})}$  como  $\eta^{(1, \mathbf{Y})} = \eta^{(3, \mathbf{X})} + \gamma_3$ , com

$$\eta^{(3, \mathbf{X})} = \lim_{n \rightarrow +\infty} \frac{P(X_1 \leq u_n < X_4, N_{2,7}^{(3, \mathbf{X})} = 0)}{P(X_1 \leq u_n < X_4)}.$$

Mas por iv) do Teorema 2.4.3

$$\begin{aligned} & nP(X_1 \leq u_n < X_4, N_{2,7}^{(3, \mathbf{X})} = 0) \\ &= nP(X_1 \leq u_n < X_2, X_3 \leq u_n < X_4, X_5 \leq u_n, X_6 \leq u_n, X_7 \leq u_n, \\ & \quad X_8 \leq u_n, X_9 \leq u_n, X_{10} \leq u_n) + \\ & \quad + nP(X_1 \leq u_n, X_2 \leq u_n, X_3 \leq u_n < X_4, X_5 \leq u_n, X_6 \leq u_n, X_7 \leq u_n, \\ & \quad X_8 \leq u_n, X_9 \leq u_n, X_{10} \leq u_n) + o(1) \\ &= nP(X_1 \leq u_n < X_2, X_3 \leq u_n < X_4, X_6 \leq u_n) + \\ & \quad + nP(X_1 \leq u_n, X_2 \leq u_n, X_3 \leq u_n < X_4, X_6 \leq u_n) + o(1) \\ &\xrightarrow{n \rightarrow +\infty} \tau \frac{1}{r^2} \left(1 - \frac{1}{r^2}\right) + \tau \left(\frac{1}{r^2}\right)^0 \left(1 - \frac{1}{r^2}\right)^2 = \tau \left(1 - \frac{1}{r^2}\right) \end{aligned}$$

e

$$\begin{aligned}
& nP(X_1 \leq u_n < X_4) \\
&= nP(X_1 \leq u_n, X_2 \leq u_n, X_3 \leq u_n < X_4) + nP(X_1 \leq u_n < X_2, X_3 \leq u_n < X_4) + o(1) \\
&= \sum_{j \geq 2} nP(X_1 \leq u_n, X_2 \leq u_n, X_3 \leq u_n < X_4, \dots, X_{2j-1} \leq u_n < X_{2j}, X_{2j+1} \leq u_n) + \\
&\quad + \sum_{j \geq 2} nP(X_1 \leq u_n < X_2, X_3 \leq u_n < X_4, \dots, X_{2j-1} \leq u_n < X_{2j}, X_{2j+1} \leq u_n) + o(1) \\
&\xrightarrow{n \rightarrow +\infty} \sum_{j \geq 1} \tau \left( \frac{1}{r^2} \right)^{j-1} \left( 1 - \frac{1}{r^2} \right)^2 + \sum_{j \geq 2} \tau \left( \frac{1}{r^2} \right)^{j-1} \left( 1 - \frac{1}{r^2} \right) \\
&= \tau \left( 1 - \frac{1}{r^2} \right) \sum_{j \geq 2} \left( \frac{1}{r^2} \right)^{j-1} \left( 1 - \frac{1}{r^2} \right) + \tau \left( \frac{1}{r^2} \right) \sum_{j \geq 2} \left( \frac{1}{r^2} \right)^{j-2} \left( 1 - \frac{1}{r^2} \right) \\
&= \tau \left( 1 - \frac{1}{r^2} \right) + \tau \left( \frac{1}{r^2} \right) \\
&= \tau = \nu^{(3, \mathbf{X})},
\end{aligned}$$

$$\text{logo } \eta^{(3, \mathbf{X})} = 1 - \frac{1}{r^2} = \eta^{(1, \mathbf{X})} = \theta^{(\mathbf{X})}.$$

Em relação a  $\gamma_3$  temos

$$\begin{aligned}
& \lim_{n \rightarrow +\infty} 3nP(X_1 \leq u_{n3} < X_4, N_{2,6}^{(3, \mathbf{X})} > 0, N_{7,7}^{(3, \mathbf{X})} = 0) \\
&= \lim_{n \rightarrow +\infty} 3n \left( P(X_1 \leq u_{n3} < X_4, N_{7,7}^{(3, \mathbf{X})} = 0) - P(X_1 \leq u_{n3} < X_4, N_{2,7}^{(3, \mathbf{X})} = 0) \right) \\
&= \tau \left( 1 - \frac{1}{r^2} \right) \left[ 1 + \frac{1}{r^2} + \left( \frac{1}{r^2} \right)^2 \right] - \tau \left( 1 - \frac{1}{r^2} \right) \\
&= \left( 1 - \frac{1}{r^2} \right) \left( \frac{1}{r^2} + \left( \frac{1}{r^2} \right)^2 \right) \tau.
\end{aligned}$$

□

Notemos que este processo AR(1) confirma a ideia inicial, e um tanto intuitiva, de que a frequência amostral influenciaria o valor do índice de cruzamentos da sucessão, no sentido em que  $\eta^{(1, \mathbf{Y})} > \eta^{(1, \mathbf{X})}$ .

Para finalizar destacamos a importância da propriedade **iv)** do Teorema 2.4.3 na estimação do índice de cruzamentos do processo AR(1) considerado. Esta propriedade garante-nos que a distribuição condicional do tamanho dos *runs* de cruzamentos converge para uma distribuição geométrica de valor esperado coincidente com  $1/\eta^{(1,\mathbf{X})} = 1/(1 - 1/r^2)$ . Torna-se assim natural estimar o índice de cruzamentos através do quociente entre o número de *runs* de cruzamentos e o número total de cruzamentos, isto é, considerando o estimador

$$\hat{\eta}^{(1,\mathbf{X})} = \frac{\sum_{i=1}^n \mathbb{I}_{B_{i,n}}}{\sum_{i=1}^n \mathbb{I}_{\{X_i \leq u < X_{i+1}\}}}, \quad (2.15)$$

com

$$B_{i,n} = \bigcup_{s \geq 1} \left\{ N_{i,i}^{(1,\mathbf{X})} = 0, X_{i+2} \leq u < X_{i+3}, \dots, X_{i+2s} \leq u < X_{i+2s+1}, N_{i+2s+2, i+2s+2}^{(1,\mathbf{X})} = 0 \right\}$$

onde  $u$  é um nível adequado.

No próximo capítulo estudaremos as propriedades teóricas deste tipo de estimadores.



# Capítulo 3

## O estimador de *runs* do índice de cruzamentos

### 3.1 Introdução

O índice de cruzamentos,  $\eta$ , introduzido em H. Ferreira (2006), enquanto medida do agrupamento de cruzamentos de níveis elevados em sucessões estacionárias, revela-se um parâmetro importante no estudo de acontecimentos extremos. A sua relação com o índice extremal,  $\theta$ , permite-nos estimá-lo a partir dos conhecidos estimadores para  $\theta$ . Relembramos que nas últimas décadas muitos estimadores consistentes foram propostos para o índice extremal, no entanto nem todos podem ser usados na estimação do índice de cruzamentos. Por exemplo, o estimador de *runs* do índice extremal é definido sob a validade da condição  $D''(\mathbf{u})$  que implica  $\eta = 1$ . Parece-nos assim interessante, e analogamente ao que foi feito na estimação do índice extremal, sugerir estimadores para o índice de cruzamentos, a partir das caracterizações assintóticas conhecidas para este parâmetro, e estudar as suas propriedades assintóticas.

No capítulo anterior vimos que sob a validade da condição  $\tilde{D}^{(3)}(\mathbf{u})$  a distribuição condicional do tamanho dos runs de cruzamentos, do processo autoregressivo considerado, convergia para uma distribuição geométrica de valor esperado coincidente com o inverso do índice de cruzamentos, o que levou a sugerir que  $\eta$  fosse estimado através do quociente entre o número de runs de cruzamentos e o número total de cruzamentos. O Teorema 1.3.14 de H. Ferreira (2007) garante que cada run de cruzamentos pode ser identificado a partir de cada ocorrência de um não cruzamento seguido de um cruzamento, uma vez os cruzamentos ocorrem sucessivamente. Deste modo, definimos na Secção 3.2, o estimador, não paramétrico, de *runs* do índice de cruzamentos, dado pelo quociente entre o número total de não-cruzamentos seguidos de um cruzamento e o número total de cruzamentos, como sugerido em H. Ferreira (2007).

Nas Secções 3.3 e 3.4 obtemos, respetivamente, a consistência fraca e a normalidade assintótica do estimador de *runs* do índice de cruzamentos. Os resultados são obtidos com argumentos análogos aos usados por Nandagopalan (1990) para o seu estimador de *runs* do índice extremal.

A escolha dos níveis a usar na estimação é uma das dificuldades encontradas na estimação de muitos dos coeficientes extremos e por conseguinte debatemos este assunto na Secção 3.5. Terminamos o capítulo com a análise do comportamento do estimador de *runs* do índice de cruzamentos em diversas amostras de dimensão finita, simuladas a partir de processos para os quais o índice de cruzamentos é bem conhecido, bem como em conjuntos de dados reais. Na aplicação a dados reais avaliamos a incerteza da estimação com um método *bootstrap* que propomos na Secção 3.7.

A maioria dos resultados deste capítulo são apresentados em Sebastião *et al.* (2012).

### 3.2 O estimador de *runs* do índice de cruzamentos

Ao longo deste capítulo iremos assumir que a sucessão estacionária  $\mathbf{X}$  verifica a condição  $\tilde{D}^{(3)}(\mathbf{u})$ , de H. Ferreira (2006, 2007) (ver Definição 1.3.8). Notemos que esta condição é implicada pela condição

$$\lim_{n \rightarrow +\infty} n \sum_{j=4}^{r_n-1} P(X_1 \leq u_n < X_2, \tilde{N}_{3,3} = 0, X_j \leq u_n < X_{j+1}) = 0, \quad (3.1)$$

que restringe a possibilidade de se ter mais de um não-cruzamento seguido de um cruzamento num dado bloco, portanto permite a identificação de grupos de cruzamentos e de *runs* de cruzamentos. Aliás, o próximo resultado diz-nos que a distribuição condicional do tamanho de um grupo de cruzamentos  $\tilde{\pi}_n$  em (1.19) pode ser substituída pela distribuição condicional do tamanho de um *run* de cruzamentos  $\tilde{\pi}_n^*$  em (1.28).

**Teorema 3.2.1** *Suponhamos que as condições  $\Delta(\mathbf{u})$  e (3.1) se verificam para  $\mathbf{X}$  com  $\mathbf{u} = \{u_n\}_{n \geq 1}$  tal que  $nP(\tilde{N}_{1,1} = 0, X_3 \leq u_n < X_4) \xrightarrow{n \rightarrow +\infty} \gamma$  para algum  $\gamma > 0$ . Então  $\tilde{\pi}_n(j) - \tilde{\pi}_n^*(j) \xrightarrow{n \rightarrow +\infty} 0$ , para cada  $j \geq 1$ .*

**Demonstração:** Por ser equivalente ao pretendido, iremos demonstrar que  $\tilde{Q}_n(j) - \tilde{Q}_n^*(j) \xrightarrow{n \rightarrow +\infty} 0$ , com  $\tilde{Q}_n(j) = \sum_{s=j}^{+\infty} \tilde{\pi}_n(s)$  e  $\tilde{Q}_n^*(j) = \sum_{s=j}^{+\infty} \tilde{\pi}_n^*(s)$ ,  $j \geq 1$ .

Com  $J = [0, r_n/n]$ , temos, pelo Lema 2.2.1 com  $T = 1$  e o Teorema 1.3.13, para  $j \geq 1$

$$\begin{aligned} \tilde{Q}_n(j) &= P(\tilde{N}_n(J) \geq j \mid \tilde{N}_n(J) > 0) \\ &= \frac{P(\tilde{N}_n(J) \geq j)}{P(\tilde{N}_n(J) > 0)} \\ &= \frac{k_n}{\gamma} P(\tilde{N}_n(J) \geq j)(1 + o(1)) \\ &= \frac{k_n}{\gamma} \left[ P\left(X_1 \leq u_n < X_2, \tilde{N}_n\left(\left[\frac{3}{n}, \frac{r_n}{n}\right]\right) \geq j - 1\right) + \right. \\ &\quad \left. + \sum_{i=3}^{r_n-(2j-4)} P\left(\tilde{N}_{1,i-2} = 0, X_i \leq u_n < X_{i+1}, \tilde{N}_n\left(\left[\frac{i+2}{n}, \frac{r_n}{n}\right]\right) \geq j - 1\right) \right] (1 + o(1)). \end{aligned}$$

Mas, pela definição de  $u_n$  e tendo em conta o Teorema 1.3.13 e o Lema 1.3.2, vem

$$\frac{k_n}{\gamma} P\left(X_1 \leq u_n < X_2, \tilde{N}_n\left(\left[\frac{3}{n}, \frac{r_n}{n}\right]\right)\right) \leq \frac{k_n}{\gamma} P(X_1 \leq u_n < X_2) = o(1)$$

e, por (3.1)

$$\begin{aligned}
0 &\leq P\left(\tilde{N}_{1,i-2} = 0, X_i \leq u_n < X_{i+1}, \tilde{N}_n\left(\left[\frac{i+2}{n}, \frac{r_n}{n}\right]\right) \geq j-1\right) - \\
&- P\left(\tilde{N}_{1,i-2} = 0, X_i \leq u_n < X_{i+1}, X_{i+2} \leq u_n < X_{i+3}, \dots, X_{i+2j-2} \leq u_n < X_{i+2j-1}\right) \\
&\leq \sum_{j=4}^{r_n-1} P(X_1 \leq u_n < X_2, \tilde{N}_{3,3} = 0, X_j \leq u_n < X_{j+1}) = o(1/n),
\end{aligned}$$

logo

$$\begin{aligned}
\tilde{Q}_n(j) &= \frac{k_n}{\gamma} \left[ P(\tilde{N}_{1,i-2} = 0, X_i \leq u_n < X_{i+1}, \dots, X_{i+2j-2} \leq u_n < X_{i+2j-1}) \right] \times \\
&\times (1 + o(1)) + o(1).
\end{aligned}$$

Por outro lado, também se tem

$$\begin{aligned}
0 &\leq P(\tilde{N}_{i-2,i-2} = 0, X_i \leq u_n < X_{i+1}, \dots, X_{i+2j-2} \leq u_n < X_{i+2j-1}) \\
&\leq \sum_{j=4}^{r_n-1} P(X_1 \leq u_n < X_2, \tilde{N}_{3,3} = 0, X_j \leq u_n < X_{j+1}) - \\
&- P\left(\tilde{N}_{1,i-2} = 0, X_i \leq u_n < X_{i+1}, \dots, X_{i+2j-2} \leq u_n < X_{i+2j-1}\right) \\
&\leq \sum_{j=4}^{r_n-1} P(X_1 \leq u_n < X_2, \tilde{N}_{3,3} = 0, X_j \leq u_n < X_{j+1}) = o(1/n)
\end{aligned}$$

e portanto

$$\begin{aligned}
\tilde{Q}_n(j) &= \frac{k_n}{\gamma} \left[ \sum_{i=3}^{r_n-(2j-4)} P\left(\tilde{N}_{i-2,i-2} = 0, X_i \leq u_n < X_{i+1}, \dots, X_{i+2j-2} \leq u_n < X_{i+2j-1}\right) \right] \times \\
&\times (1 + o(1)) + o(1) \\
&\sim \frac{k_n}{\gamma} (r_n - j + 2) P(X_3 \leq u_n < X_4, \dots, X_{2j+1} \leq u_n < X_{2j+2} \mid \tilde{N}_{1,1} = 0, X_3 \leq u_n < X_4) \times \\
&\times \frac{\gamma}{n} (1 + o(1)) + o(1) \\
&= \tilde{Q}_n^*(j) (1 + o(1)) + o(1) = \tilde{Q}_n^*(j) + o(1),
\end{aligned}$$

como pretendido. □

Sob a validade de (3.1), parece assim natural, tendo em conta o resultado anterior, estimar  $\eta$  usando *runs* de cruzamentos em vez de grupos de cruzamentos. Tal pode, de facto, ser feito tendo em conta o Teorema 1.3.14, como veremos no que se segue.

Para uma dada sucessão de níveis  $\mathbf{u} = \{u_n\}_{n \geq 1}$  consideremos o processo pontual de não-cruzamentos seguidos por um cruzamento de  $u_n$ , pelas primeiras  $n$  variáveis da sucessão  $\mathbf{X}$ , *i.e.*,

$$\widehat{N}_n(B) = \widehat{N}_n(u_n)(B) = \sum_{i=1}^{n-3} \mathbb{I}_{\{\widetilde{N}_{i,i}=0, X_{i+2} \leq u_n < X_{i+3}\}} \delta_{\frac{i}{n}}(B), \quad B \subseteq \mathcal{B}([0, 1]). \quad (3.2)$$

Definam-se as variáveis aleatórias  $Y_i = Y_i(u_n)$ ,  $i = 1, \dots, n$ , representativas do número de cruzamentos consecutivos de  $u_n$  a partir do instante  $i + 2$  ou do comprimento dos *runs* de cruzamentos de  $u_n$  a partir do instante  $i + 2$ , ou seja, tais que,

$$Y_i = \begin{cases} 0 & \text{se } \widetilde{N}_{i+2, i+2} = 0 \\ k & \text{se } X_{i+2} \leq u_n < X_{i+3}, \dots, X_{i+2k} \leq u_n < X_{i+2k+1}, \widetilde{N}_{i+2k+2, i+2k+2} = 0, \quad k \geq 1. \end{cases}$$

Represente-se por  $Z_i(u_n)$ ,  $i = 1, \dots, n$ , o comprimento de cada uma dessas sequências, condicionado à ocorrência de um não-cruzamento seguido de um cruzamento de  $u_n$  no instante  $i + 2$ . As variáveis aleatórias  $Z_i(u_n)$  têm assim distribuição de probabilidade  $\widetilde{\pi}_n^*$  dada por (1.28), *i.e.*,

$$\widetilde{\pi}_n^*(k) = P(Z_i(u_n) = k) = P(Y_i = k \mid \widetilde{N}_{i,i} = 0, X_{i+2} \leq u_n < X_{i+3}), \quad k \geq 1, \quad (3.3)$$

que é independente de  $i$  devido à estacionaridade de  $\mathbf{X}$ .

Com esta notação, se a sucessão  $\mathbf{X}$  tem índice de cruzamentos  $\eta > 0$ ,  $\mathbf{u} = \{u_n = u_n^{(\nu)}\}_{n \geq 1}$  for uma sucessão de níveis normalizados de acordo com (1.14) e  $\mathbf{X}$  verificar a condição  $\widetilde{D}^{(3)}(\mathbf{u})$ , então pelo Teorema 1.3.14 vem

$$\begin{aligned} \frac{1}{\eta} &= \lim_{n \rightarrow +\infty} E[Z_1(u_n)] = \lim_{n \rightarrow +\infty} E[Y_1 \mid \widetilde{N}_{1,1} = 0, X_3 \leq u_n < X_4] \\ &= \lim_{n \rightarrow +\infty} \frac{P(X_1 \leq u_n < X_2)}{P(\widetilde{N}_{1,1} = 0, X_3 \leq u_n < X_4)} = \lim_{n \rightarrow +\infty} \frac{E[\widehat{N}_n([0, 1])]}{E[\widetilde{N}_n([0, 1])]}. \end{aligned} \quad (3.4)$$

Salientamos que pela estacionaridade de  $\mathbf{X}$  se tem

$$P(\widetilde{N}_{1,1} = 0, X_3 \leq u_n < X_4) = P(X_1 \leq u_n < X_2, \widetilde{N}_{3,3} = 0)$$

e conseqüentemente um *run* de cruzamentos tanto pode ser identificado no seu início como no seu fim.

O índice de cruzamentos surge em (3.4) como o inverso do limite do comprimento médio de cada *run* de cruzamentos e, sob a validade da condição  $\widetilde{D}^{(3)}(\mathbf{u})$ , cada *run* de cruzamentos pode ser identificado a partir de cada ocorrência de um não cruzamento seguido de um cruzamento,

uma vez que os cruzamentos ocorrem sucessivamente. A partir deste resultado é natural definir o estimador não paramétrico para  $\eta$ , através do quociente entre o número total de não-cruzamentos seguidos de um cruzamento e o número total de cruzamentos, como foi sugerido em H. Ferreira (2007),

$$\hat{\eta}_n = \hat{\eta}_n(u) := \frac{\hat{N}_n([0, 1])}{\tilde{N}_n([0, 1])} = \frac{\sum_{i=1}^{n-3} \mathbb{I}_{\{\tilde{N}_{i,i}=0, X_{i+2} \leq u_n < X_{i+3}\}}}{\sum_{i=1}^{n-1} \mathbb{I}_{\{X_i \leq u_n < X_{i+1}\}}}. \quad (3.5)$$

Denominaremos este estimador por “Estimador de *runs*” do índice de cruzamentos por este depender do número de *runs* de cruzamentos de um nível  $u_n$  pelas  $n$  primeiras variáveis de  $\mathbf{X}$ . A motivação apresentada para a definição deste estimador é semelhante à considerada por Nandagopalan (1990) na sua definição do estimador de *runs* do índice extremal de sucessões estacionárias que verificam a condição  $D''(\mathbf{u})$ .

Nas aplicações práticas usaremos sempre o estimador (3.5), contudo as propriedades teóricas apresentadas nas próximas secções serão demonstradas para um estimador assintoticamente equivalente a  $\hat{\eta}_n$  que definimos no que se segue.

Seja  $\hat{N}_n^{(T)}$  o processo pontual marcado em  $[0, 1]$  definido por

$$\hat{N}_n^{(T)}(B) = \sum_{i=1}^{n-3} T(Y_i) \mathbb{I}_{\{\tilde{N}_{i,i}=0, X_{i+2} \leq u_n < X_{i+3}\}} \delta_{\frac{i}{n}}(B), \quad B \subseteq \mathcal{B}([0, 1]), \quad (3.6)$$

onde  $T$  representa uma aplicação de  $\mathbb{N}$  em  $\mathbb{R}$ . Assim,  $\hat{N}_n^{(T)}$  tem massa de probabilidade igual a  $T(Y_i)$  no ponto  $i/n$  quando  $\mathbf{X}$  tiver um não-cruzamento em  $i$  seguido de um cruzamento.

Se em (3.6) considerarmos  $T(y) \equiv 1$  obtemos o processo pontual  $\hat{N}_n$  em (3.2), enquanto que, se considerarmos  $T(y) = y \geq 1$  obtemos um novo processo pontual que denotaremos por  $\bar{N}_n$ . Este processo pontual distingue-se do processo pontual de cruzamentos  $\tilde{N}_n$  em (1.13) num acontecimento, com probabilidade majorada por  $P(X_1 \leq u_n < X_2)$  que converge para zero quando  $n \rightarrow +\infty$  e  $u_n = u_n^{(\nu)}$ . Deste modo, definiremos um novo estimador  $\hat{\eta}_n^* = \hat{N}_n([0, 1])/\bar{N}_n([0, 1])$  que é assintoticamente equivalente a  $\hat{\eta}_n$ , visto que para os níveis  $u_n = u_n^{(\nu)}$  se tem  $\hat{\eta}_n = \hat{\eta}_n^* + o_p(1)$ . Assim, as propriedades demonstradas a seguir para  $\hat{\eta}_n^*$  também serão válidas para  $\hat{\eta}_n$  que usaremos nos estudos de simulação.

### 3.3 Consistência fraca

Iremos assumir de agora em diante que  $\mathbf{X}$  é uma sucessão estacionária que verifica a condição  $\tilde{D}^{(3)}(\mathbf{u})$  e tem índice de cruzamentos  $\eta > 0$ . Salientamos que os resultados obtidos permanecem válidos sob a condição (3.1).

Para provarmos que o estimador  $\hat{\eta}_n^*$  é um estimador consistente de  $\eta$ , e por conseguinte o es-

timador  $\widehat{\eta}_n$  também o será, precisamos do resultado auxiliar que apresentamos de seguida. Este lema obtém-se do Lemma 5.2.1 de Nandagopalan (1990) (pags. 77-79), substituindo o processo pontual de excedências e o processo pontual marcado de cruzamentos, respetivamente pelo processo pontual de cruzamentos e o processo pontual marcado de não cruzamentos seguido de um cruzamento.

**Lema 3.3.1** *Seja  $\{k_n\}_{n \geq 1}$  uma sucessão de inteiros que verifica  $k_n \xrightarrow{n \rightarrow +\infty} +\infty$  e suponhamos que existe uma sucessão  $\{l_n\}_{n \geq 1}$  para a qual*

$$k_n[\alpha_{n,l_n-2} + P(\widetilde{N}_n([0, l_n/n]) > 0)] \xrightarrow{n \rightarrow +\infty} 0, \quad (3.7)$$

onde  $\alpha_{n,l_n-2}$ ,  $n \geq 1$ , são os coeficientes associados à condição  $\Delta(\mathbf{u})$ .

Então

$$E[e^{ia_n \widehat{N}_n^{(T)}(J_n)}] - \prod_{j=1}^{k_n} E[e^{ia_n \widehat{N}_n^{(T)}(J_{nj})}] \xrightarrow{n \rightarrow +\infty} 0,$$

para qualquer sucessão de reais  $\{a_n\}_{n \geq 1}$ , onde  $J_n \subset [0, 1]$ ,  $n \geq 1$ , é uma sucessão de intervalos, tais que para cada  $n$ ,  $J_n \supset \bigcup_{j=1}^{k_n} J_{nj}$ , com  $J_{nj}$ ,  $j = 1, \dots, k_n$  subintervalos disjuntos verificando  $m(J_n) - m(\bigcup_{j=1}^{k_n} J_{nj}) \leq k_n/n$  ( $m(\cdot)$  denota a medida de Lebesgue).

A condição (3.7) verifica-se para qualquer sucessão de níveis  $\mathbf{u}^{(\nu)}$ , definida de acordo com (1.14), se  $\mathbf{X}$  verificar a condição  $\Delta(\mathbf{u}^{(\nu)})$ , uma vez que (3.7) implica que  $k_n \alpha_{n,l_n-2} \xrightarrow{n \rightarrow +\infty} 0$  e, para  $\mathbf{u}^{(\nu)}$ , temos que  $k_n P(\widetilde{N}_n([0, l_n/n]) > 0) \xrightarrow{n \rightarrow +\infty} 0$ .

A utilização de níveis normalizados para cruzamentos definidos de acordo com (1.14) não nos permite obter a consistência do estimador  $\widehat{\eta}_n^*$ , porque quando  $n$  aumenta, os valores de  $\widehat{\eta}_n^*$  não convergem adequadamente para o valor de  $\eta$ . É assim necessário considerar níveis menos elevados que  $u_n^{(\nu)}$ , i.e., níveis não normalizados  $v_n = \widetilde{u}_{[n/c_n]}^{(\nu)}$ , para algum  $\nu > 0$  fixo, que satisfaçam

$$nP(X_1 \leq v_n < X_2) - c_n \nu \xrightarrow{n \rightarrow +\infty} 0 \quad (3.8)$$

onde  $\{c_n\}_{n \geq 1}$  e  $\{k_n\}_{n \geq 1}$  são sucessões de números reais que verificam  $c_n, k_n \xrightarrow{n \rightarrow +\infty} +\infty$ ,  $n/c_n \xrightarrow{n \rightarrow +\infty} +\infty$  e  $c_n/k_n \xrightarrow{n \rightarrow +\infty} 0$ . Notemos que para esta sucessão de níveis também se verifica a convergência

$$\lim_{n \rightarrow +\infty} E[Z_1(v_n)] = 1/\eta.$$

De agora em diante usaremos a seguinte notação

$$\overline{N}_n = \overline{N}_n([0, 1]) = \sum_{i=1}^{n-3} Y_i \mathbb{I}_{\{\widetilde{N}_{i,i}=0, X_{i+2} \leq u_n < X_{i+3}\}}$$

e

$$\widehat{N}_n = \widehat{N}_n([0, 1]) = \sum_{i=1}^{n-3} \mathbb{I}_{\{\tilde{N}_{i,i}=0, X_{i+2} \leq u_n < X_{i+3}\}},$$

para que  $\overline{N}_n$  e  $\widehat{N}_n$  representem variáveis aleatórias em vez de processos pontuais. Será conveniente representar  $\overline{N}_n$  por  $\overline{N}_n = \sum_{i=1}^{k_n} \overline{N}_{ni}$  com  $\overline{N}_{ni} = \sum_{j=(i-1)r_n+1}^{ir_n-3} Y_i \mathbb{I}_{\{\tilde{N}_{j,j}=0, X_{j+2} \leq u_n < X_{j+3}\}},$

$$1 \leq i \leq k_n, \text{ e } \widehat{N}_n \text{ por } \widehat{N}_n = \sum_{i=1}^{k_n} \widehat{N}_{ni} \text{ com } \widehat{N}_{ni} = \sum_{j=(i-1)r_n+1}^{ir_n-3} \mathbb{I}_{\{\tilde{N}_{j,j}=0, X_{j+2} \leq u_n < X_{j+3}\}}, 1 \leq i \leq k_n.$$

Notemos que no caso de (3.7) se verificar e tendo em conta que  $n - k_n r_n \leq k_n$ , o Lema 3.3.1 permanece válido para  $\overline{N}_n$  e  $\widehat{N}_n$ , e consequentemente pela estacionaridade de  $\mathbf{X}$  podemos concluir que

$$E[\exp(ia_n \overline{N}_n)] = \left( E[\exp(ia_n \overline{N}_{r_n})] \right)^{k_n} + o(1) \quad (3.9)$$

e

$$E[\exp(ia_n \widehat{N}_n)] = \left( E[\exp(ia_n \widehat{N}_{r_n})] \right)^{k_n} + o(1)$$

onde  $\overline{N}_{r_n} = \overline{N}_{n1}$  e  $\widehat{N}_{r_n} = \widehat{N}_{n1}$ . Usaremos estas relações nos resultados que se seguem.

Consideremos agora a seguinte variável aleatória de Poisson composta

$$\overline{N}_n^* = \sum_{j=1}^{\widehat{N}_n^*} Z_j^*(v_n) \quad (3.10)$$

em que  $\widehat{N}_n^*$  é uma variável aleatória de Poisson com valor médio  $c_n \eta \nu$  e  $Z_j^*(v_n)$  são variáveis aleatórias independentes e identicamente distribuídas com  $Z_1(v_n)$ , ou seja, têm distribuição  $\tilde{\pi}_n^*$  dada por (1.28).

Demonstramos no próximo resultado, recorrendo a argumentos semelhantes aos usados em Nandagopalan (1990) que, sob certas condições, as distribuições limite de  $\overline{N}_n$  e  $\widehat{N}_n$  são idênticas às distribuições de  $\overline{N}_n^*$  e  $\widehat{N}_n^*$ , respetivamente.

**Teorema 3.3.1** *Seja  $\{v_n\}_{n \geq 1}$  uma sucessão de níveis que verifique (3.8) e suponhamos que existe uma sucessão  $\{l_n\}_{n \geq 1}$  para a qual se verifique (3.7) para esses níveis.*

Se

$$E \left[ Z_j(v_n) \mathbb{I}_{\{\widehat{N}_{r_n} > 1\}} \right] = o(1) \quad (3.11)$$

uniformemente para  $j = 1, \dots, r_n$ , ( $r_n = [n/k_n]$ ), então

$$E \left[ \exp(itc_n^{-1} \overline{N}_n) \right] - E \left[ \exp(itc_n^{-1} \overline{N}_n^*) \right] \xrightarrow{n \rightarrow +\infty} 0 \quad (3.12)$$

e

$$E \left[ \exp(itc_n^{-1} \widehat{N}_n) \right] - E \left[ \exp(itc_n^{-1} \widehat{N}_n^*) \right] \xrightarrow{n \rightarrow +\infty} 0, \quad (3.13)$$

para cada  $t \in \mathbb{R}$ .

**Demonstração:** Sejam  $Z_j = Z_j(v_n)$ ,  $\Psi_{Z_1}(t) = E[\exp(itZ_1)]$  a função característica da variável aleatória  $Z_1$  com distribuição  $\tilde{\pi}_n^*$  dada por (1.28) e  $R_j = \{\tilde{N}_{j,j} = 0, X_{j+2} \leq v_n < X_{j+4}\}$ . Como  $\overline{N}_n^*$  é uma variável aleatória de Poisson composta, temos que

$$E \left[ \exp(itc_n^{-1} \overline{N}_n^*) \right] = \exp(-c_n \nu \eta (1 - \Psi_{Z_1}(tc_n^{-1})))$$

e por conseguinte é suficiente provar que

$$E \left[ \exp(itc_n^{-1} \overline{N}_n) \right] = \exp(-c_n \nu \eta (1 - \Psi_{Z_1}(tc_n^{-1}))) + o(1), \quad (3.14)$$

ou equivalentemente, por (3.9) que

$$(E \left[ \exp(itc_n^{-1} \overline{N}_{r_n}) \right])^{k_n} = \exp(-c_n \nu \eta (1 - \Psi_{Z_1}(tc_n^{-1}))) + o(1).$$

Note-se que

$$\begin{aligned} E \left[ \exp(itc_n^{-1} \overline{N}_{r_n}) \right] &= 1 - E \left[ (1 - \exp(itc_n^{-1} \overline{N}_{r_n})) \{ \mathbb{I}_{\{\widehat{N}_{r_n}=1\}} + \mathbb{I}_{\{\widehat{N}_{r_n}>1\}} \} \right] \\ &= 1 - E \left[ (1 - \exp(itc_n^{-1} \overline{N}_{r_n})) \mathbb{I}_{\{\widehat{N}_{r_n}=1\}} \right] - \rho_n^{(1)} \\ &= 1 - E \left[ (1 - \exp(itc_n^{-1} \overline{N}_{r_n})) \mathbb{I}_{\{\cup_{j=1}^{r_n} (R_j \cap (\widehat{N}_{r_n}=1))\}} \right] - \rho_n^{(1)} \\ &= 1 - E \left[ \sum_{j=1}^{r_n} (1 - \exp(itc_n^{-1} Z_j)) \mathbb{I}_{\{R_j \cap (\widehat{N}_{r_n}=1)\}} \right] - \rho_n^{(1)} \\ &= 1 - E \left[ \sum_{j=1}^{r_n} (1 - \exp(itc_n^{-1} Z_j)) \{ \mathbb{I}_{\{R_j\}} - \mathbb{I}_{\{R_j \cap (\widehat{N}_{r_n}>1)\}} \} \right] - \rho_n^{(1)} \\ &= 1 - E \left[ \sum_{j=1}^{r_n} (1 - \exp(itc_n^{-1} Z_j)) \mathbb{I}_{\{R_j\}} \right] + \rho_n^{(2)} - \rho_n^{(1)} \\ &= 1 - r_n E \left[ (1 - \exp(itc_n^{-1} Z_1)) \mathbb{I}_{\{R_1\}} \right] + \rho_n^{(2)} - \rho_n^{(1)} \\ &= 1 - r_n E \left[ (1 - \exp(itc_n^{-1} Z_1)) \right] P(R_1) + \rho_n^{(2)} - \rho_n^{(1)} \\ &= 1 - \frac{c_n}{k_n} (\eta \nu + o(1)) \times (1 - \Psi_{Z_1}(tc_n^{-1})) + o(k_n^{-1}) \end{aligned}$$

com

$$\begin{aligned}
|\rho_n^{(1)}| &= \left| E \left[ (1 - \exp(itc_n^{-1}\bar{N}_{r_n})) \mathbb{I}_{\{\hat{N}_{r_n} > 1\}} \right] \right| \\
&\leq |tc_n^{-1}| E \left[ \bar{N}_{r_n} \mathbb{I}_{\{\hat{N}_{r_n} > 1\}} \right] \\
&\leq |tc_n^{-1}| E \left[ \sum_{j=1}^{r_n} Z_j \mathbb{I}_{\{R_j \cap \{\hat{N}_{r_n} > 1\}\}} \right]
\end{aligned}$$

e

$$\begin{aligned}
|\rho_n^{(2)}| &= \left| E \left[ \sum_{j=1}^{r_n} (1 - \exp(itc_n^{-1}Z_j)) \mathbb{I}_{\{R_j \cap \{\hat{N}_{r_n} > 1\}\}} \right] \right| \\
&\leq |tc_n^{-1}| E \left[ \sum_{j=1}^{r_n} Z_j \mathbb{I}_{\{R_j \cap \{\hat{N}_{r_n} > 1\}\}} \right],
\end{aligned}$$

pois para  $n \geq 1$  e  $t \in \mathbb{R}$  se tem  $|e^{it} - 1 - it/1! - \dots - (it)^{n-1}/(n-1)!| \leq |t|^n/n!$ . Logo, pela desigualdade triangular vem

$$\begin{aligned}
|\rho_n^{(2)} - \rho_n^{(1)}| &\leq 2|tc_n^{-1}| \sum_{j=1}^{r_n} E \left[ Z_j \mathbb{I}_{\{\hat{N}_{r_n} > 1\}} \right] P(R_j) \\
&= 2|tc_n^{-1}| P(R_1) \sum_{j=1}^{r_n} E \left[ Z_j \mathbb{I}_{\{\hat{N}_{r_n} > 1\}} \right] \\
&= 2|tc_n^{-1}| r_n P(R_1) o(1) \\
&= o(k_n^{-1}),
\end{aligned}$$

uma vez que se verifica (3.11), e sob a validade de  $\tilde{D}^{(3)}(v_n)$ , temos que  $\frac{c_n}{k_n} P(R_1) = \eta\nu + o(1)$ . Além do mais, como  $|1 - \Psi_{Z_1}(tc_n^{-1})| \leq |tc_n^{-1}| E[Z_1(v_n)]$  e  $E[Z_1(v_n)] \xrightarrow[n \rightarrow +\infty]{c_n} \eta^{-1}$  vem que

$$E \left[ \exp(itc_n^{-1}\bar{N}_n) \right] = \left( 1 - \frac{c_n}{k_n} \eta\nu (1 - \Psi_{Z_1}(tc_n^{-1})) + o(k_n^{-1}) \right)^{k_n} + o(1),$$

de onde resulta (3.14) e consequentemente prova (3.12).

A convergência (3.13) poderá ser obtida com argumentos similares e tendo em consideração que

$$\begin{aligned}
E \left[ \exp(itc_n^{-1}\hat{N}_{r_n}) \right] &= 1 - r_n E \left[ (1 - \exp(itc_n^{-1})) \mathbb{I}_{\{R_1\}} \right] + \rho_n^{(2)} - \rho_n^{(1)} \\
&= 1 - r_n (1 - \exp(itc_n^{-1})) P(R_1) + o(k_n^{-1})
\end{aligned}$$

$$= 1 - \frac{c_n}{k_n}(\eta\nu + o(1))(1 - \exp(itc_n^{-1})) + o(k_n^{-1}),$$

com

$$\begin{aligned} |\rho_n'^{(2)} - \rho_n'^{(1)}| &\leq 2tc_n^{-1} E \left[ \sum_{j=1}^{r_n} \mathbb{I}_{\{R_j \cap (\hat{N}_{r_n} > 1)\}} \right] \\ &= 2|tc_n^{-1}| r_n P(R_1) P(\hat{N}_{r_n} > 1) \\ &= o(k_n^{-1}). \end{aligned}$$

□

Se no resultado anterior fossem considerados níveis normalizados  $u_n^{(\nu)}$ , as convergências (3.12) e (3.13) verificar-se-iam sem haver necessidade da constante normalizadora  $c_n$ , sendo neste caso  $\hat{N}_n^*$  em (3.10) uma variável aleatória de Poisson com valor médio  $\eta\nu$ .

Encontramos no próximo resultado a justificação para a utilização dos níveis  $v_n = \tilde{u}_{[n/c_n]}^{(\nu)}$  para garantir a consistência fraca do estimador.

**Teorema 3.3.2** *Suponhamos que  $\{v_n\}_{n \geq 1}$  é uma sucessão de níveis satisfazendo (3.8),*

$$E [Z_1^*(v_n) \mathbb{I}_{\{Z_1^*(v_n) > c_n\}}] \xrightarrow{n \rightarrow +\infty} 0 \quad (3.15)$$

e

$$E [(Z_1^*(v_n))^2 \mathbb{I}_{\{Z_1^*(v_n) \leq c_n\}}] \xrightarrow{n \rightarrow +\infty} 0, \quad (3.16)$$

então

$$c_n^{-1} \bar{N}_n^* \xrightarrow[n \rightarrow +\infty]{P} \nu.$$

**Demonstração:** Começemos por notar que a convergência (3.15) conjuntamente com o facto de  $E[Z_1(v_n)] \xrightarrow{n \rightarrow +\infty} \eta^{-1}$  implicam que  $\eta\nu E[Z_1^*(v_n) \mathbb{I}_{\{Z_1^*(v_n) \leq c_n\}}] \xrightarrow{n \rightarrow +\infty} \nu$ . Assim, será suficiente demonstrar que

$$c_n^{-1} \bar{N}_n^* - \eta\nu E[Z_1^*(v_n) \mathbb{I}_{\{Z_1^*(v_n) \leq c_n\}}] \xrightarrow[n \rightarrow +\infty]{P} 0.$$

Considerando agora  $\gamma_n = E[Z_1^*(v_n) \mathbb{I}_{\{Z_1^*(v_n) \leq c_n\}}]$  obtemos, para qualquer  $\varepsilon > 0$  e  $\delta < \varepsilon / \max\{\gamma_n\}$ ,

$$P(|c_n^{-1} \bar{N}_n^* - \eta\nu\gamma_n| > 2\varepsilon) = \sum_{k \geq 0} P\left(|c_n^{-1} \bar{N}_n^* - \eta\nu\gamma_n| > 2\varepsilon \mid \hat{N}_n^* = k\right) P(\hat{N}_n^* = k)$$

$$\begin{aligned}
&= \sum_{k: |k - c_n \eta \nu| \leq c_n \delta} P \left( \left| c_n^{-1} \sum_{j=1}^k Z_j^*(v_n) - \eta \nu \gamma_n \right| > 2\varepsilon \right) P(\widehat{N}_n^* = k) + \\
&\quad + \sum_{k: k > c_n \eta \nu + c_n \delta} P \left( \left| c_n^{-1} \sum_{j=1}^k Z_j^*(v_n) - \eta \nu \gamma_n \right| > 2\varepsilon \right) P(\widehat{N}_n^* = k) \\
&\leq \underbrace{(\delta + \eta \nu) E[Z_1^*(v_n) \mathbb{I}_{\{Z_1^*(v_n) > c_n\}}] + \frac{E[(Z_1^*(v_n) \mathbb{I}_{\{Z_1^*(v_n) \leq c_n\}})^2]}{c_n(\eta \nu - \delta) [\varepsilon / (\delta + \eta \nu)]^2}}_{(A)} + \\
&\quad + \underbrace{P(|\widehat{N}_n^* - c_n \eta \nu| > \delta c_n)}_{(B)}.
\end{aligned}$$

A expressão (A) resulta do facto de  $\gamma_n \leq E[Z_1^*(v_n)] \xrightarrow{n \rightarrow +\infty} \eta^{-1}$  e portanto para  $\delta < \varepsilon / \max\{\gamma_n\}$  se ter  $|k - c_n \eta \nu| \leq \delta c_n$  e

$$\begin{aligned}
&P \left( \left| c_n^{-1} \sum_{j=1}^k Z_j^*(v_n) - \eta \nu \gamma_n \right| > 2\varepsilon \right) \\
&= P \left( \left| \frac{k}{c_n} (k^{-1} \sum_{j=1}^k Z_j^*(v_n) - \gamma_n) + \frac{k}{c_n} \gamma_n - \eta \nu \gamma_n \right| > 2\varepsilon \right) \\
&\leq P \left( \left| \frac{k}{c_n} \left( \frac{1}{k} \sum_{j=1}^k Z_j^*(v_n) - \gamma_n \right) \right| > \varepsilon \right), \text{ porque } \varepsilon - \gamma_n \left| \frac{k}{c_n} - \eta \nu \right| \geq 0 \\
&\leq P \left( \left| \frac{1}{k} \sum_{j=1}^k Z_j^*(v_n) - \gamma_n \right| > \frac{\varepsilon}{\eta \nu + \delta} \right), \text{ porque } \frac{k}{c_n} \leq \eta \nu + \delta \\
&\leq P \left( Z_j^*(v_n) \neq Z_j^*(v_n) \mathbb{I}_{\{Z_j^*(v_n) \leq c_n\}} \text{ para algum } j \leq k \right) + \\
&\quad + P \left( \left| \frac{1}{k} \sum_{j=1}^k Z_j^*(v_n) \mathbb{I}_{\{Z_j^*(v_n) \leq c_n\}} - \gamma_n \right| > \frac{\varepsilon}{\eta \nu + \delta} \right) \\
&\leq k P(Z_1^*(v_n) > c_n) + \frac{E[(Z_1^*(v_n) \mathbb{I}_{\{Z_1^*(v_n) \leq c_n\}})^2]}{k [\varepsilon / (\delta + \eta \nu)]^2} \\
&\leq (\delta + \eta \nu) E[Z_1^*(v_n) \mathbb{I}_{\{Z_1^*(v_n) > c_n\}}] + \frac{E[(Z_1^*(v_n) \mathbb{I}_{\{Z_1^*(v_n) \leq c_n\}})^2]}{c_n(\eta \nu - \delta) [\varepsilon / (\delta + \eta \nu)]^2}.
\end{aligned}$$

A convergência de (A) para zero resulta agora de (3.15) e (3.16), enquanto que (B) converge

para zero uma vez que  $\widehat{N}_n^*$  é uma variável aleatória de Poisson de média  $c_n \eta \nu$ . Consequentemente, como  $\varepsilon > 0$  é arbitrário, temos que

$$c_n^{-1} \overline{N}_n^* - \eta \nu \xrightarrow[n \rightarrow +\infty]{P} 0$$

como pretendido.  $\square$

Demonstrámos no Teorema 3.3.1 que  $c_n^{-1} \overline{N}_n$  e  $c_n^{-1} \widehat{N}_n$  têm assintoticamente a mesma distribuição que, respetivamente,  $c_n^{-1} \overline{N}_n^*$  e  $c_n^{-1} \widehat{N}_n^*$  e no Teorema 3.3.2 que  $c_n^{-1} \overline{N}_n^* \xrightarrow[n \rightarrow +\infty]{P} \nu$ . Mais ainda, como  $\widehat{N}_n^*$  é uma variável aleatória de Poisson com valor médio  $c_n \eta \nu$ , segue-se que  $c_n^{-1} \widehat{N}_n^* \xrightarrow[n \rightarrow +\infty]{P} \eta \nu$ . Destas duas convergências resulta a consistência fraca de  $\widehat{\eta}_n^*$ , enunciada no próximo resultado.

**Corolário 3.3.1** *Nas condições dos Teoremas 3.3.1 e 3.3.2 tem-se*

$$\widehat{\eta}_n^* \xrightarrow[n \rightarrow +\infty]{P} \eta.$$

### 3.4 Normalidade assintótica

A normalidade assintótica do estimador  $\widehat{\eta}_n^*$  é obtida impondo algumas condições adicionais aos primeiro e segundo momentos da variável aleatória  $Z_1(v_n)$ . Para tal, o próximo resultado revela-se essencial.

**Teorema 3.4.1** *Seja  $\{v_n\}_{n \geq 1}$  uma sucessão de níveis satisfazendo (3.8) e suponhamos que existe uma sucessão  $\{l_n\}_{n \geq 1}$  para a qual se verifica (3.7) para esses níveis.*

Se

$$E[\overline{N}_{r_n}^2 \mathbb{I}_{\{\widehat{N}_{r_n} > 1\}}] = o(c_n/k_n), \quad (3.17)$$

$$E[(Z_j(v_n))^2 \mathbb{I}_{\{\widehat{N}_{r_n} > 1\}}] = o(1) \quad (3.18)$$

uniformemente em  $j = 1, \dots, r_n$ , ( $r_n = [n/k_n]$ ) e

$$\sigma_n^2 = E[(Z_1(v_n))^2]$$

é uma sucessão limitada, então  $c_n^{-1/2} \begin{bmatrix} \overline{N}_n - E[\overline{N}_n] \\ \widehat{N}_n - E[\widehat{N}_n] \end{bmatrix}$  converge em distribuição, se e só se

$c_n^{-1/2} \begin{bmatrix} \overline{N}_n^* - E[\overline{N}_n^*] \\ \widehat{N}_n^* - E[\widehat{N}_n^*] \end{bmatrix}$  converge, e nesse caso, os limites coincidem.

**Demonstração:** Fixemos  $t, s \in \mathbb{R}$  e denotemos as funções características dos vetores aleatórios

anteriores, no ponto  $(t, s)$  por, respetivamente,

$$\Psi_1(t, s) = E \left[ \exp\{itc_n^{-1/2}(\bar{N}_n - E[\bar{N}_n]) + isc_n^{-1/2}(\hat{N}_n - E[\hat{N}_n])\} \right]$$

e

$$\Psi_2(t, s) = E \left[ \exp\{itc_n^{-1/2}(\bar{N}_n^* - E[\bar{N}_n^*]) + isc_n^{-1/2}(\hat{N}_n^* - E[\hat{N}_n^*])\} \right].$$

Teremos que provar que para cada  $(t, s)$ ,  $\Psi_2(t, s) - \Psi_1(t, s) \xrightarrow[n \rightarrow +\infty]{} 0$ .

Denotando por  $\Psi_{Z_1}(t) = E[\exp(itZ_1(v_n))]$  a função característica da variável aleatória  $Z_1(v_n)$  com distribuição  $\tilde{\pi}_n^*$  dada por (1.28) e  $R_j = \{\tilde{N}_{j,j} = 0, X_{j+2} \leq v_n < X_{j+4}\}$  e tendo em conta as definições de  $\bar{N}_n^*$  e  $\hat{N}_n^*$  podemos concluir que

$$\begin{aligned} \Psi_2(t, s) &= \exp \left\{ -c_n \nu \eta \left[ 1 - \Psi_{Z_1}(tc_n^{-1/2}) e^{isc_n^{-1/2}} + ic_n^{-1/2} (tE[Z_1(v_n)] + s) \right] \right\} \\ &= \exp \left\{ -c_n \nu \eta \left[ \rho_n^{(1)} - isc_n^{-1/2} \rho_n^{(2)} + \rho_n^{(3)} \right] \right\} + o(1), \end{aligned} \quad (3.19)$$

uma vez que podemos escrever

$$\begin{aligned} 1 - \Psi_{Z_1}(tc_n^{-1/2}) e^{isc_n^{-1/2}} &= 1 - \Psi_{Z_1}(tc_n^{-1/2}) + \Psi_{Z_1}(tc_n^{-1/2}) (1 - e^{isc_n^{-1/2}}) \\ &= (-itc_n^{-1/2} E[Z_1(v_n)] + \rho_n^{(1)}) + (1 + \rho_n^{(2)}) (-isc_n^{-1/2} + \rho_n^{(3)}), \end{aligned}$$

com  $|\rho_n^{(1)}| \leq \frac{t}{2} c_n^{-1} \sigma_n^2$ ,  $|\rho_n^{(2)}| \leq tc_n^{-1/2} E[Z_1(v_n)]$ ,  $|\rho_n^{(3)}| \leq \frac{s^2}{2} c_n^{-1}$  e  $\rho_n^{(2)} \times \rho_n^{(3)} = o(c_n^{-1})$ .

Por outro lado, tendo em conta o Lema 3.3.1 temos que

$$\Psi_1(t, s) = \left( E \left[ \exp\{itc_n^{-1/2}(\bar{N}_{r_n} - E[\bar{N}_{r_n}]) + isc_n^{-1/2}(\hat{N}_{r_n} - E[\hat{N}_{r_n}])\} \right] \right)^{k_n} + o(1).$$

Procedendo agora como na demonstração do Teorema 3.3.1 vem

$$\begin{aligned} E \left[ \exp(itc_n^{-1/2} \bar{N}_{r_n} + isc_n^{-1/2} \hat{N}_{r_n}) \right] &= \\ &= 1 - E \left[ \left( 1 - \exp(itc_n^{-1/2} \bar{N}_{r_n} + isc_n^{-1/2} \hat{N}_{r_n}) \right) \{ \mathbb{I}_{\{\hat{N}_{r_n}=1\}} + \mathbb{I}_{\{\hat{N}_{r_n}>1\}} \} \right] \\ &= 1 - r_n E \left[ 1 - \exp(itc_n^{-1/2} Z_1(v_n) + isc_n^{-1/2}) \right] \mathbb{I}_{\{R_1\}} - \rho_n^{(4)} + \rho_n^{(5)} \\ &= 1 - r_n P(R_1) \left[ 1 - \Psi_{Z_1}(tc_n^{-1/2}) \exp(isc_n^{-1/2}) \right] + o(k_n^{-1}), \end{aligned}$$

pois por (3.17) e (3.18) prova-se que  $|\rho_n^{(4)} + \rho_n^{(5)}| = o(k_n^{-1})$ .

Como também

$$\exp(-itc_n^{-1/2} E[\bar{N}_{r_n}]) = 1 - itc_n^{-1/2} r_n P(R_1) E[Z_1(v_n)] + o(k_n^{-1}),$$

$$\exp(-isc_n^{-1/2}E[\widehat{N}_{r_n}]) = 1 - isc_n^{-1/2}r_nP(R_1) + o(k_n^{-1})$$

e tendo em consideração que  $r_nP(R_1) = \frac{c_n}{k_n}(\nu\eta + o(1))$ , vem

$$\begin{aligned} & E \left[ \exp\{itc_n^{-1/2}(\overline{N}_{r_n} - E[\overline{N}_{r_n}]) + isc_n^{-1/2}(\widehat{N}_{r_n} - E[\widehat{N}_{r_n}])\} \right] \\ &= \left\{ 1 - r_nP(R_1)[1 - \Psi_{Z_1}(tc_n^{-1/2}) \exp(isc_n^{-1/2})] + o(k_n^{-1}) \right\} \times \\ & \quad \times \left\{ 1 - itc_n^{-1/2}r_nP(R_1)E[Z_1(v_n)] + o(k_n^{-1}) \right\} \left\{ 1 - isc_n^{-1/2}r_nP(R_1) + o(k_n^{-1}) \right\} \\ &= 1 - \frac{c_n}{k_n}\nu\eta \left[ \rho_n^{(1)} - isc_n^{-1/2}\rho_n^{(2)} + \rho_n^{(3)} \right] + o(k_n^{-1}). \end{aligned}$$

Donde

$$\begin{aligned} \Psi_1(t, s) &= \left\{ 1 - \frac{c_n}{k_n}\nu\eta \left[ \rho_n^{(1)} - isc_n^{-1/2}\rho_n^{(2)} + \rho_n^{(3)} \right] + o(k_n^{-1}) \right\}^{k_n} + o(1) \\ &= \exp \left\{ -c_n\nu\eta \left[ \rho_n^{(1)} - isc_n^{-1/2}\rho_n^{(2)} + \rho_n^{(3)} \right] \right\} + o(1). \end{aligned} \quad (3.20)$$

De (3.19) e (3.20) resulta que, para qualquer par  $(t, s)$ ,  $\Psi_2(t, s) - \Psi_1(t, s) \xrightarrow[n \rightarrow +\infty]{} 0$ .  $\square$

Tendo em conta este resultado, provamos a seguir que as sucessões de variáveis aleatórias  $\{\overline{N}_n^*\}_{n \geq 1}$  e  $\{\widehat{N}_n^*\}_{n \geq 1}$  convergem em distribuição para variáveis aleatórias normalmente distribuídas.

**Teorema 3.4.2** *Seja  $\{v_n\}_{n \geq 1}$  uma sucessão de níveis satisfazendo (3.8) e suponhamos que existe uma sucessão  $\{l_n\}_{n \geq 1}$  para a qual se verifique (3.7) para esses níveis.*

Se

$$\sigma_n^2 = E[(Z_1(v_n))^2] \xrightarrow[n \rightarrow +\infty]{} \sigma^2 < +\infty \quad (3.21)$$

e para cada  $\varepsilon > 0$ ,

$$E[(Z_1^*(v_n))^2 \mathbb{I}_{\{(Z_1^*(v_n))^2 > \varepsilon c_n\}}] \xrightarrow[n \rightarrow +\infty]{} 0. \quad (3.22)$$

Então

$$c_n^{-1/2} \begin{bmatrix} \overline{N}_n^* - E[\overline{N}_n^*] \\ \widehat{N}_n^* - E[\widehat{N}_n^*] \end{bmatrix} \xrightarrow[n \rightarrow +\infty]{d} \mathcal{N} \left( \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \nu\eta\sigma^2 & \nu \\ \nu & \nu\eta \end{bmatrix} \right).$$

**Demonstração:** Atendendo a que  $\overline{N}_n^*$  é uma variável aleatória de Poisson composta tem-se  $E[\overline{N}_n^*] = E[\widehat{N}_n^*]E[Z_1(v_n)]$  e então podemos escrever

$$\overline{N}_n^* - E[\overline{N}_n^*] = (\overline{N}_n^* - \widehat{N}_n^*E[Z_1(v_n)]) + E[Z_1(v_n)](\widehat{N}_n^* - E[\widehat{N}_n^*]). \quad (3.23)$$

A condição de Lindeberg (3.22) e a definição de  $\widehat{N}_n^*$  garantem a validade do Teorema do Limite Central, ou seja,

$$c_n^{-1/2}(\overline{N}_n^* - \widehat{N}_n^* E[Z_1(v_n)]) \xrightarrow[n \rightarrow +\infty]{d} \mathcal{N}\left(0, \eta\nu\left(\sigma^2 - \frac{1}{\eta^2}\right)\right).$$

Donde

$$\begin{aligned} E[\exp(itc_n^{-1/2}(\overline{N}_n^* - \widehat{N}_n^* E[Z_1(v_n)]))] &= \exp(-c_n\eta\nu[1 - \Psi_{Z_1}(tc_n^{-1/2})e^{-itc_n^{-1/2}E[Z_1(v_n)]}]) \\ &\xrightarrow[n \rightarrow +\infty]{} \exp\left(-\frac{\eta\nu}{2}\left(\sigma^2 - \frac{1}{\eta^2}\right)t^2\right). \end{aligned}$$

e conseqüentemente

$$c_n\eta\nu[1 - \Psi_{Z_1}(tc_n^{-1/2})e^{-itc_n^{-1/2}E[Z_1(v_n)]}] \xrightarrow[n \rightarrow +\infty]{} \frac{\eta\nu}{2}\left(\sigma^2 - \frac{1}{\eta^2}\right)t^2. \quad (3.24)$$

Como também  $c_n^{-1/2}(\widehat{N}_n^* - E[\widehat{N}_n^*]) \xrightarrow[n \rightarrow +\infty]{d} \mathcal{N}(0, \eta\nu)$ , temos que

$$\begin{aligned} E[\exp\{isc_n^{-1/2}(\widehat{N}_n^* - E[\widehat{N}_n^*])\}] &= \exp(-c_n\eta\nu(1 - e^{isc_n^{-1/2}} + isc_n^{-1/2})) \\ &\xrightarrow[n \rightarrow +\infty]{} \exp\left(-\frac{\eta\nu s^2}{2}\right) \end{aligned}$$

e conseqüentemente

$$c_n\eta\nu(1 - e^{isc_n^{-1/2}} + isc_n^{-1/2}) \xrightarrow[n \rightarrow +\infty]{} \frac{\eta\nu s^2}{2}. \quad (3.25)$$

Agora,

$$\begin{aligned} &E[\exp\{itc_n^{-1/2}(\overline{N}_n^* - \widehat{N}_n^* E[Z_1(v_n)]) + isc_n^{-1/2}(\widehat{N}_n^* - E[\widehat{N}_n^*])\}] \\ &= \exp(-c_n\eta\nu(1 - \Psi_{Z_1}(tc_n^{-1/2})e^{-itc_n^{-1/2}E[Z_1(v_n)]}e^{isc_n^{-1/2}} + isc_n^{-1/2})), \end{aligned} \quad (3.26)$$

com  $\Psi_{Z_1}(tc_n^{-1/2}) = 1 + itc_n^{-1/2}E[Z_1(v_n)] + \rho_n^{(1)}$  e  $e^{-itc_n^{-1/2}E[Z_1(v_n)]} = 1 - itc_n^{-1/2}E[Z_1(v_n)] + \rho_n^{(2)}$  em que  $\rho_n^{(1)}$  e  $\rho_n^{(2)}$  são ambos majorados por  $t^2 c_n^{-1} \sigma_n^2 / 2$ .

Portanto, de (3.21), (3.24) e (3.25) obtemos

$$\begin{aligned} &\eta\nu c_n [1 - \Psi_{Z_1}(tc_n^{-1/2})e^{-itc_n^{-1/2}E[Z_1(v_n)]}e^{isc_n^{-1/2}} + isc_n^{-1/2}] \\ &= \eta\nu c_n [(1 - \Psi_{Z_1}(tc_n^{-1/2})e^{-itc_n^{-1/2}E[Z_1(v_n)]}) + (1 - e^{isc_n^{-1/2}} + isc_n^{-1/2}) + o(c_n^{-1})] \\ &\xrightarrow[n \rightarrow +\infty]{} \frac{\eta\nu(\sigma^2 - 1/\eta^2)t^2}{2} + \frac{\eta\nu s^2}{2}, \end{aligned}$$

que conjuntamente com (3.26) implica que

$$c_n^{-1/2} \begin{bmatrix} \bar{N}_n^* - \hat{N}_n^* E[Z_1(v_n)] \\ \hat{N}_n^* - E[\hat{N}_n^*] \end{bmatrix} \xrightarrow[n \rightarrow +\infty]{d} \mathcal{N} \left( \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \eta\nu \left( \sigma^2 - \frac{1}{\eta^2} \right) & 0 \\ 0 & \nu\eta \end{bmatrix} \right).$$

Para concluir basta agora ter em conta (3.23) e o facto de  $E[Z_1(v_n)] \xrightarrow[n \rightarrow +\infty]{} 1/\eta$ .  $\square$

A normalidade assintótica de  $\hat{\eta}_n^*$  é agora uma consequência imediata dos resultados anteriores.

**Corolário 3.4.1** *Nas condições dos Teoremas 3.4.1 e 3.4.2, tem-se*

$$\sqrt{c_n}(\hat{\eta}_n^* - \eta_n) \xrightarrow[n \rightarrow +\infty]{d} \mathcal{N} \left( 0, \sqrt{\frac{\eta}{\nu}(\eta^2\sigma^2 - 1)} \right),$$

em que  $\eta_n = E[\hat{N}_n]/E[\bar{N}_n]$ .

**Demonstração:** Pelos Teoremas 3.4.1 e 3.4.2 temos que

$$c_n^{-1/2} \begin{bmatrix} \bar{N}_n - E[\bar{N}_n] \\ \hat{N}_n - E[\hat{N}_n] \end{bmatrix} \xrightarrow[n \rightarrow +\infty]{d} \mathcal{N} \left( \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \nu\eta\sigma^2 & \nu \\ \nu & \nu\eta \end{bmatrix} \right).$$

As condições (3.18), (3.21) e (3.22) implicam as condições dos Teoremas 3.3.2 e 3.3.1, por conseguinte  $c_n^{-1}\bar{N}_n \xrightarrow[n \rightarrow +\infty]{P} \nu$ . Por outro lado, a condição  $\tilde{D}^{(3)}(v_n)$  implica que  $\eta_n \xrightarrow[n \rightarrow +\infty]{} \eta$ . Logo, para concluir basta notar que

$$\sqrt{c_n}(\hat{\eta}_n - \eta_n) = \frac{1}{c_n^{-1}\bar{N}_n} \{c_n^{-1/2}(\hat{N}_n - E[\hat{N}_n]) - \eta_n c_n^{-1/2}(\bar{N}_n - E[\bar{N}_n])\}$$

e

$$Var \left[ c_n^{-1/2}(\hat{N}_n - E[\hat{N}_n]) - \eta_n c_n^{-1/2}(\bar{N}_n - E[\bar{N}_n]) \right] = \eta\nu + \eta_n^2 \eta\nu\sigma^2 - 2\eta_n\nu.$$

$\square$

Notemos que a variância de  $\hat{\eta}_n^*$  é da ordem de  $1/\nu$  e portanto, se  $\nu$  se mantivesse fixo como em (1.14), não tínhamos garantida a consistência do estimador. Assim, teremos que assumir  $\nu = \nu_n \rightarrow +\infty$  quando  $n \rightarrow +\infty$ , que é conseguido recorrendo aos níveis não-normalizados  $v_n$  definidos em (3.8).

Um aspecto importante a salientar dos resultados anteriores é o facto de para  $\eta = 1$  a distribuição assintótica não estar bem definida, uma vez que neste caso a variância vem igual a zero, devido a não haver agrupamentos de cruzamentos. Esta situação ocorre por exemplo em processos que verifiquem a condição  $D''(\mathbf{u})$  e consequentemente também verificam a condição  $\tilde{D}^{(3)}(\mathbf{u})$ .

Apresentamos a seguir um resultado que do ponto de vista prático tem mais utilidade que o Corolário 3.4.1, uma vez que permite obter intervalos de confiança assintóticos ou testes de hipóteses para o parâmetro  $\eta$ , e deste modo caracterizar o agrupamento de cruzamentos de níveis elevados nos dados observados.

**Corolário 3.4.2** *Suponhamos que se verificam as condições do Corolário 3.4.1,*

$$\sqrt{c_n}(\eta_n - \eta) \xrightarrow[n \rightarrow +\infty]{} 0$$

e

$$c_n^{-1} E[(Z_1^*(v_n))^4 \mathbb{I}_{\{(Z_1^*(v_n))^2 \leq c_n\}}] \xrightarrow[n \rightarrow +\infty]{} 0. \quad (3.27)$$

Então

$$\sqrt{\frac{\bar{N}_n}{\hat{\eta}_n^* ((\hat{\eta}_n^* \hat{\sigma}_n)^2 - 1)}} (\hat{\eta}_n^* - \eta) \xrightarrow[n \rightarrow +\infty]{d} \mathcal{N}(0, 1),$$

$$\text{onde } \hat{\sigma}_n^2 = \frac{\sum_{i=1}^{n-3} Y_i^2 \mathbb{I}_{\{\tilde{N}_{i,i}=0, X_{i+2} \leq v_n < X_{i+3}\}}}{\hat{N}_n}.$$

**Demonstração:** Pelos resultados anteriores e (3.27) temos que

$$\frac{\bar{N}_n}{c_n \nu} \xrightarrow[n \rightarrow +\infty]{P} 1 \quad \text{e} \quad \left( \frac{c_n \nu}{\eta((\eta\sigma)^2 - 1)} \right)^{1/2} (\hat{\eta}_n^* - \eta) \xrightarrow[n \rightarrow +\infty]{d} \mathcal{N}(0, 1).$$

Aplicando agora os Teoremas 3.3.1 e 3.3.2 às variáveis  $Z_1^2(v_n)$  e  $(Z_1^*)^2(v_n)$  vem

$$c_n^{-1} \sum_{i=1}^{n-3} Y_i^2 \mathbb{I}_{\{\tilde{N}_{i,i}=0, X_{i+2} \leq v_n < X_{i+3}\}} \xrightarrow[n \rightarrow +\infty]{P} \eta \nu \sigma^2.$$

Como  $\hat{N}_n/c_n \xrightarrow[n \rightarrow +\infty]{P} \eta \nu$  segue-se que  $\hat{\sigma}_n^2 \xrightarrow[n \rightarrow +\infty]{P} \sigma^2$  e conseqüentemente

$$\left( \frac{c_n \nu}{\eta((\eta\sigma)^2 - 1)} \right)^{1/2} / \left( \frac{\bar{N}_n}{\hat{\eta}_n^* ((\hat{\eta}_n^* \hat{\sigma}_n)^2 - 1)} \right)^{1/2} \xrightarrow[n \rightarrow +\infty]{P} 1,$$

o que demonstra o resultado. □

Dada uma amostra aleatória  $X_1, \dots, X_n$  e um nível  $u$ , o Corolário 3.4.2 permite-nos construir o seguinte intervalo de confiança assintótico a  $100(1 - \alpha)\%$  para  $\eta$

$$\left( \hat{\eta}_n^* - z_{1-\alpha/2} \sqrt{\frac{\hat{\eta}_n^* ((\hat{\eta}_n^* \hat{\sigma}_n)^2 - 1)}{\bar{N}_n}}, \hat{\eta}_n^* + z_{1-\alpha/2} \sqrt{\frac{\hat{\eta}_n^* ((\hat{\eta}_n^* \hat{\sigma}_n)^2 - 1)}{\bar{N}_n}} \right) \quad (3.28)$$

onde  $z_{1-\alpha/2}$  denota o quantil de probabilidade  $1 - \alpha/2$  da distribuição normal padrão,

$$\hat{\eta}_n^* = \frac{\sum_{i=1}^{n-3} \mathbb{I}_{\{\tilde{N}_{i,i}=0, X_{i+2} \leq u < X_{i+3}\}}}{\sum_{i=1}^{n-3} Y_i \mathbb{I}_{\{\tilde{N}_{i,i}=0, X_{i+2} \leq u < X_{i+3}\}}} \quad \text{e} \quad \hat{\sigma}_n^2 = \frac{\sum_{i=1}^{n-3} Y_i^2 \mathbb{I}_{\{\tilde{N}_{i,i}=0, X_{i+2} \leq u < X_{i+3}\}}}{\sum_{i=1}^{n-3} \mathbb{I}_{\{\tilde{N}_{i,i}=0, X_{i+2} \leq u < X_{i+3}\}}}.$$

Os intervalos de confiança em (3.28) são aproximações dos verdadeiros intervalos de confiança para amostras finitas. A região de confiança exata para uma amostra de dimensão  $n$  poderá ser melhor aproximada recorrendo a técnicas de reamostragem *bootstrap*. O bootstrap pode ser utilizado para quantificar a incerteza associada às estimativas. A ideia principal (Efron (1979)) é tratar o conjunto de dados, ou seja a amostra, como se fosse a população e retirar amostras com reposição desse conjunto de dados. Este procedimento é repetido um número elevado de vezes,  $B$ , e para cada amostra calculam-se as quantidades de interesse. Os  $B$  valores calculados são então utilizados para estimar a distribuição desconhecida dessas quantidades.

Singh (1981) demonstrou que esta teoria clássica *bootstrap* é inadequada para dados dependentes. Neste contexto, é usual agruparem-se os dados em blocos e proceder à reamostragem destes blocos de modo a manter a estrutura de dependência local dos modelos em estudo. Deste modo, se a dimensão dos blocos tender para infinito assim como a dimensão da amostra, é possível encontrar uma teoria assintótica que garanta os resultados obtidos por este processo de reamostragem. Encontramos na literatura vários autores que se dedicaram ao estudo deste tipo de reamostragem *bootstrap*, dos quais destacamos Carlstein (1986), Künsch (1989), Liu e Singh (1992), Politis e Romano (1992, 1994) e Lahiri (2003).

Os métodos de reamostragem por blocos têm a desvantagem de ser necessário otimizar a dimensão dos blocos, sendo que a determinação da dimensão ótima não parece ser uma tarefa simples de generalizar para qualquer tipo de modelo. Em Ferro e Segers (2003) encontramos um procedimento *bootstrap* que evita a identificação dos blocos e conseqüentemente a determinação da sua dimensão. Este método foi aplicado no contexto da estimação do índice extremal e será por nós adaptada mais adiante na estimação do índice de cruzamentos.

Recordemos que as propriedades que demonstrámos para o estimador  $\hat{\eta}_n^*$  se mantêm válidas para  $\hat{\eta}_n$  que iremos usar a partir de agora.

### 3.5 A escolha dos níveis

Na Secção 3.3 demonstrámos que a consistência do estimador de *runs* do índice de cruzamentos  $\hat{\eta}_n$  se tem para níveis  $v_n$  mais baixos (que os normalizados para cruzamentos  $u_n^{(\nu)}$ ) satisfazendo (3.8). Assim, a escolha precisa dos níveis  $v_n$  depende da distribuição conjunta do par aleatório  $(X_1, X_2)$ , geralmente desconhecida. Na prática, estes níveis determinísticos terão de ser substituídos por níveis aleatórios sugeridos pela relação

$$\frac{n}{c_n} P(X_1 \leq v_n < X_2) \sim \nu. \quad (3.29)$$

Esta relação indica-nos que o número esperado de cruzamentos do nível  $v_n$  é aproximadamente igual a  $c_n \nu$ . Contrariamente ao que usualmente é usado na estimação do índice extremal, estes níveis aleatórios não poderão ser representados por estatísticas ordinais de topo. Todavia, para níveis  $v_n$  tais que  $\frac{n}{c_n} P(X_1 > v_n) \sim \tau$ , o número esperado de cruzamentos é necessariamente inferior ou igual a  $c_n \tau$ . Parece assim natural substituir também os níveis  $v_n$  pela estatística ordinal

apropriada usada na estimação do índice extremal (ver Nandagopalan (1990)), nomeadamente  $\hat{v}_n = X_{n-[c_n\tau]:n}$ . Tendo em consideração estes argumentos, definimos o seguinte estimador para o índice de cruzamentos,

$$\tilde{\eta}_n = \frac{\hat{N}_n(\hat{v}_n)}{\tilde{N}_n(\hat{v}_n)},$$

em que  $\hat{N}_n(\hat{v}_n) = \hat{N}_n([0, 1])$  e  $\tilde{N}_n(\hat{v}_n) = \tilde{N}_n([0, 1])$  com  $u_n = \hat{v}_n$ .

A consistência fraca deste estimador é obtida demonstrando que o valor das suas estimativas é aproximadamente igual ao valor das estimativas obtidas a partir do mesmo estimador baseado em níveis determinísticos. Para tal, comecemos por verificar que para dois níveis  $v_1$  e  $v_2$  tais que  $v_1 \neq v_2$  e  $n$  fixo, se tem sempre que

$$|\tilde{N}_n(v_1) - \tilde{N}_n(v_2)| \leq |N_n(v_1) - N_n(v_2)| \quad (3.30)$$

e

$$|\hat{N}_n(v_1) - \hat{N}_n(v_2)| \leq |N_n(v_1) - N_n(v_2)| \quad (3.31)$$

onde  $N_n(v_n) = \sum_{i=1}^n \mathbb{I}_{\{X_i > v_n\}}$ .

**Teorema 3.5.1** *Suponhamos que para cada  $\nu > 0$  existem  $\tilde{u}_{[n/c_n]}^{(\nu)} = u_{[n/c_n]}^{(\tau)}$  para algum  $\tau > 0$ , que se verificam as condições das Proposições 5.3.1 e 5.3.2 de Nandagopalan (1990) (pags. 81 e 83) para cada  $\tau'$  numa vizinhança de  $\tau$  e que as condições (3.11), (3.15) e (3.16) se verificam para cada  $\nu'$  numa vizinhança de  $\nu$ . Então  $\tilde{\eta}_n \xrightarrow[n \rightarrow +\infty]{P} \eta$ .*

**Demonstração:** Para  $\varepsilon > 0$ ,

$$P(|c_n^{-1}\tilde{N}_n(\hat{v}_n) - \nu| > 6\varepsilon) \leq P(|c_n^{-1}[c_n\tau] - N_n(v_n)| > 3\varepsilon) + P(|c_n^{-1}\tilde{N}_n(v_n) - \nu| > 3\varepsilon) \quad (3.32)$$

por (3.30), com  $v_1$  e  $v_2$  substituídos por  $\hat{v}_n$  e  $v_n = u_{[n/c_n]}^{(\tau+\varepsilon)}$ , respetivamente. Se  $\varepsilon$  é suficientemente pequeno, os resultados em Nandagopalan (1990) e obtidos na Secção 3.3 permitem concluir que  $c_n^{-1}N_n(v_n) \xrightarrow[n \rightarrow +\infty]{P} \tau + \varepsilon$  e  $c_n^{-1}\tilde{N}_n(v_n) \xrightarrow[n \rightarrow +\infty]{P} \nu + \varepsilon$ . Assim, como  $c_n^{-1}[c_n\tau] \xrightarrow[n \rightarrow +\infty]{P} \tau$ , para  $n$  suficientemente elevado a expressão de (3.32) é dominada por

$$P(|c_n^{-1}N_n(v_n) - (\tau + \varepsilon)| > \varepsilon) + P(|c_n^{-1}\tilde{N}_n(v_n) - (\nu + \varepsilon)| > \varepsilon) \xrightarrow[n \rightarrow +\infty]{P} 0,$$

o que prova que  $c_n^{-1}\tilde{N}_n(\hat{v}_n) \xrightarrow[n \rightarrow +\infty]{P} \nu$ .

Por outro lado, para  $\varepsilon > 0$ , a partir de (3.31) com  $v_1$  e  $v_2$  substituídos, respetivamente por  $\hat{v}_n$  e  $v_n = u_{[n/c_n]}(\tau + \varepsilon)$ , temos que

$$P(|c_n^{-1}\hat{N}_n(\hat{v}_n) - \eta\nu| > 6\varepsilon) \leq P(|c_n^{-1}[c_n\tau] - N_n(v_n)| > 3\varepsilon) + P(|c_n^{-1}\hat{N}_n(v_n) - \nu| > 3\varepsilon). \quad (3.33)$$

Consequentemente, se  $\varepsilon$  é suficientemente pequeno, recorrendo aos resultados da Secção 3.3,

temos que  $c_n^{-1}\widehat{N}_n(v_n) \xrightarrow[n \rightarrow +\infty]{P} \eta(\nu + \varepsilon)$  e poderemos concluir que (3.33) é dominada por

$$P(|c_n^{-1}N_n(v_n) - (\tau + \varepsilon)| > \varepsilon) + P(|c_n^{-1}\widehat{N}_n(v_n) - \eta(\nu + \varepsilon)| > \varepsilon) \xrightarrow[n \rightarrow +\infty]{} 0,$$

o que demonstra que  $c_n^{-1}\widehat{N}_n(\widehat{v}_n) \xrightarrow[n \rightarrow +\infty]{P} \eta\nu$ .

O resultado é agora uma consequência imediata das duas convergências estabelecidas.  $\square$

A demonstração da normalidade assintótica de  $\widetilde{\eta}_n$  continua um problema em aberto.

### 3.6 Estudo de simulação: alguns exemplos

Estudamos nesta secção alguns exemplos de sucessões que apresentam agrupamentos de cruzamentos de níveis elevados. Usando simulações multi-amostra, ilustraremos o desempenho do estimador de *runs* do índice de cruzamentos no processo autoregressivo de máximos (ARMAX) de H. Ferreira (2006) e no processo autoregressivo de primeira ordem negativamente correlacionado (AR(1)) de Chernick *et al.* (1991), para os quais o valor do índice de cruzamentos  $\eta$  é bem conhecido.

Segundo Gomes e Oliveira (2001) a simulação multi-amostra é uma prática comum quando é necessário estimar medidas de dispersão relativas a estatísticas de interesse, tais como por exemplo a variância ou o erro quadrático médio (MSE). Num processo de simulação multi-amostra de tamanho  $m \times r$  ( $m$  corridas e  $r$  réplicas), são obtidas  $m$  estimativas da estatística de interesse a partir de  $m$  amostras independentes de dimensão  $n$  do modelo, de seguida consideram-se  $r$  réplicas independentes desta experiência. Temos assim  $r$  conjuntos de  $m$  estimativas. O valor final simulado da estimativa é obtido pela média das  $r$  estimativas calculadas em cada réplica. Por exemplo, a variância de  $\widehat{\eta}_n$  será estimada por

$$\widehat{Var}[\widehat{\eta}_n] = \frac{1}{r} \sum_{j=1}^r s_n^{2(j)},$$

com

$$s_n^{2(j)} = \frac{1}{m-1} \sum_{i=1}^m (\widehat{\eta}_{ni}^{(j)} - \bar{\eta}_n^{(j)})^2 \quad \text{e} \quad \bar{\eta}_n^{(j)} = \frac{1}{m} \sum_{i=1}^m \widehat{\eta}_{ni}^{(j)}.$$

onde  $\widehat{\eta}_{ni}^{(j)}$  representa a  $i$ -ésima estimativa de  $\eta$  na  $j$ -ésima réplica da experiência.

Atendendo a que o estimador global do parâmetro de interesse é uma média, sabemos que, em condições bastante gerais, podemos aproximar a distribuição do estimador pela distribuição  $t$ -student com  $r-1$  graus de liberdade. Podemos também, a partir das  $r$  réplicas, estimar o erro padrão do estimador global e calcular os respetivos intervalos de confiança. Deste modo, por exemplo, um intervalo de confiança a  $100 \times (1 - \alpha)\%$  para  $E[\widehat{\eta}_n]$  é dado por,

$$\widehat{E}[\widehat{\eta}_n] \pm t_{r-1}(1 - \alpha/2) \frac{SD_{\widehat{E}[\widehat{\eta}_n]}}{\sqrt{r}}, \quad (3.34)$$

onde,

$$\widehat{E}[\widehat{\eta}_n] = \frac{1}{r} \sum_{j=1}^r \widehat{\eta}_n^{(j)}$$

e

$$SD_{\widehat{E}[\widehat{\eta}_n]} = \sqrt{\frac{1}{r-1} \sum_{j=1}^r \left( \widehat{\eta}_n^{(j)} - \widehat{\eta}_n \right)^2},$$

com

$$\widehat{\eta}_n^{(j)} = \frac{1}{m} \sum_{i=1}^m \widehat{\eta}_{ni}^{(j)} \quad \text{e} \quad \widehat{\eta}_n = \frac{1}{r} \sum_{j=1}^r \widehat{\eta}_n^{(j)}.$$

### 3.6.1 O processo ARMAX

Consideremos o processo ARMAX de H. Ferreira (2006) definido por,

$$X_n = \max\{Y_n, Y_{n-2}, Y_{n-3}\}, \quad n \geq 1, \quad (3.35)$$

onde  $\{Y_n\}_{n \geq -2}$  é uma sucessão independente e uniformemente distribuída em  $[0, 1]$ . Apresentamos na Figura 3.1 uma trajetória deste processo.

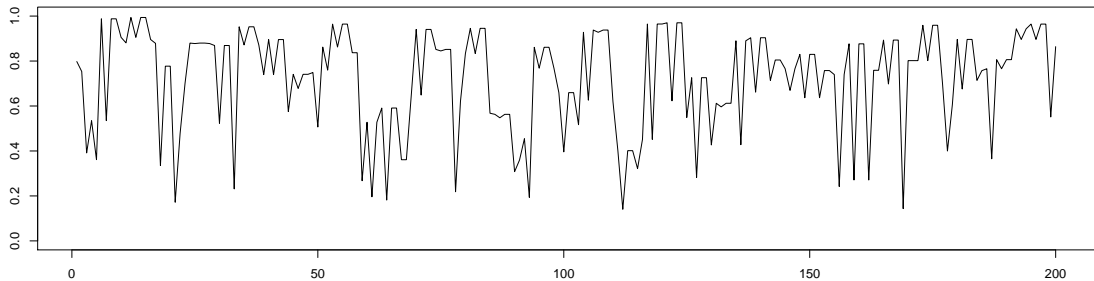


Figura 3.1: Trajetória do processo ARMAX definido em (3.35).

H. Ferreira (2006) provou que esta sucessão estacionária verifica a condição  $\widetilde{D}^{(3)}(\mathbf{u})$  com  $\mathbf{u} = \{u_n = 1 - \tau'/n\}_{n \geq 1}$ ,  $\tau > 0$ , tem índice extremal  $\theta = 1/3$  e índice de cruzamentos  $\eta = 1/2$ . Demonstrou ainda que se tem  $\lim_{n \rightarrow +\infty} nP(X_1 > u_n) = 3\tau'$  e  $\lim_{n \rightarrow +\infty} nP(X_1 \leq u_n < X_2) = 2\tau'$ , ou seja, os níveis  $u_n$  anteriores são simultaneamente normalizados para excedências e para cruzamentos. Notemos que as convergências anteriores implicam que para uma amostra  $(X_1, \dots, X_n)$  da sucessão  $\mathbf{X}$ , com  $n$  suficientemente elevado, o número de cruzamentos do nível elevado  $u_n$  é aproximadamente igual a  $2/3$  do número de excedências desse mesmo nível.

Nas simulações que se seguem consideramos níveis  $u = X_{n-k:n}$ ,  $1 \leq k \leq n-1$ , em (3.5), correspondentes às estatísticas ordinais de topo associadas à amostra aleatória  $(X_1, \dots, X_n)$ ,

comumente utilizados na estimação do índice extremal (ver por exemplo Nandagopalan (1990) e Gomes *et al.* (2008)). Deste modo, o estimador do índice de cruzamento é agora uma função de  $k$ , que representaremos por  $\hat{\eta}_n(k)$ . De facto,  $k$  está a substituir  $c_n\tau$ ,  $c_n \xrightarrow{n \rightarrow +\infty} +\infty$ , e por conseguinte a consistência será alcançada se  $k = k_n \xrightarrow{n \rightarrow +\infty} +\infty$  e  $k = o(n)$  quando  $n \rightarrow +\infty$ .

Na Figura 3.2 apresentamos, na escala linear e na escala logarítmica, uma trajetória das estimativas de  $\eta$ , obtidas com o estimador  $\hat{\eta}_n$ , definido por (3.5), em função de  $k \geq 1$ , recorrendo a uma amostra de dimensão  $n = 5000$  de (3.35). Como podemos verificar, a escala logarítmica põe em evidência o desempenho do estimador para valores pequenos de  $k$ , dando uma melhor percepção da sua região de estabilidade em torno de  $\eta = 0.5$  para estes valores. A região de estabilidade é sem dúvida pequena e as estimativas obtidas aumentam substancialmente à medida que  $k$  aumenta afastando-se assim do verdadeiro valor de  $\eta = 0.5$ .

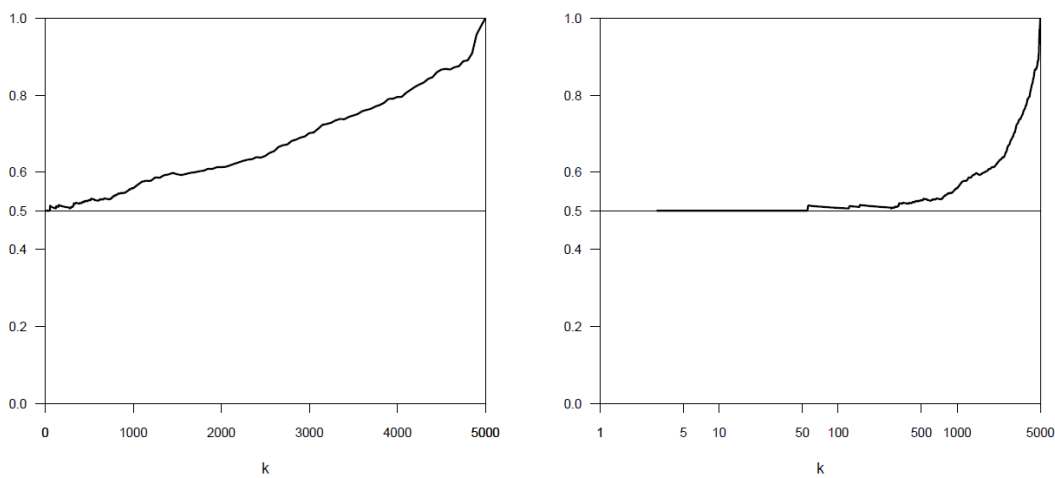


Figura 3.2: Estimativas de  $\eta$  em função de  $k$ , para uma amostra de dimensão  $n = 5000$  do processo estacionário ARMAX em (3.35), numa escala linear (à esquerda) e numa escala logarítmica (à direita).

Para amostras de dimensão  $n = 100, 200, 500, 1000, 2000$  e  $5000$ , do processo ARMAX de (3.35), realizámos uma simulação multi-amostra com 5000 corridas e 10 réplicas. Simulámos para o estimador  $\hat{\eta}_n(k)$ , o valor médio ( $E[\bullet]$ ), o erro quadrático médio ( $MSE[\bullet]$ ), o desvio padrão ( $SD[\bullet]$ ) e a fração ótima de amostragem, dada por  $k_0$  com  $k_0 := \arg \min_k MSE[\hat{\eta}_n(k)]$ .

Na Tabela 3.1 encontram-se as principais propriedades distribucionais do estimador em estudo, bem como os respetivos intervalos de confiança a 95%, calculados a partir de (3.34). Como seria de esperar os resultados melhoram à medida que  $n$  aumenta.

Tabela 3.1: Fração de amostragem ótima, valores médios, erros quadráticos médios e desvio padrão nos seus níveis ótimos, do processo ARMAX, com  $\eta = 0.5$ .

$n$	$k_0$	$k_0/n$	$E[\bullet]$	$MSE[\bullet]$	$SD[\bullet]$
100	12	0.120	$0.54265 \pm 0.000423$	$0.00546 \pm 0.000112$	$0.06034 \pm 0.000659$
200	14	0.070	$0.52440 \pm 0.000263$	$0.00224 \pm 0.000032$	$0.04059 \pm 0.000335$
500	19	0.038	$0.51200 \pm 0.000123$	$0.00072 \pm 0.000015$	$0.02408 \pm 0.000252$
1000	25	0.025	$0.50733 \pm 0.000164$	$0.00031 \pm 0.000008$	$0.01595 \pm 0.000188$
2000	27	0.014	$0.50403 \pm 0.000072$	$0.00014 \pm 0.000003$	$0.01112 \pm 0.000101$
5000	25	0.005	$0.50143 \pm 0.000049$	$0.00005 \pm 0.000001$	$0.00703 \pm 0.000093$

A sensibilidade do valor médio  $E[\bullet]$  a alterações de  $k$  evidenciam a necessidade de estudar as propriedades do viés deste estimador. Na Figura 3.3 observamos que para o processo ARMAX o comportamento do erro quadrático médio é descrito por uma função convexa, o que nos permite identificar o valor de  $k$  que minimiza esta função. Confirmamos ainda que, a partir de um determinado valor de  $k$ , o MSE é completamente dominado pelo viés, apresentando a variância valores bastantes reduzidos. Conclusões semelhantes foram retiradas para as amostras de outras dimensões.

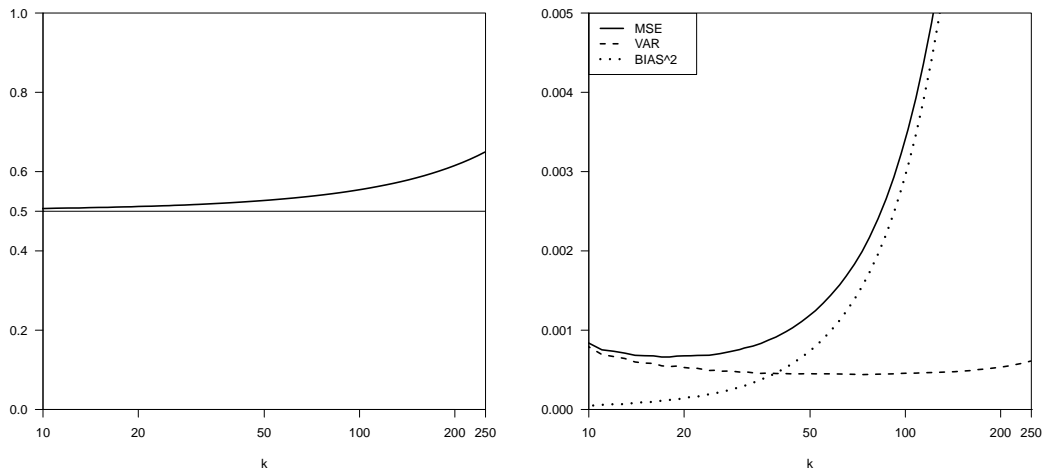


Figura 3.3: Valores médios estimados (à esquerda), erros quadráticos médios, variâncias e quadrado do viés estimados (à direita), para amostras de dimensão  $n = 500$  do processo ARMAX em (3.35), numa escala logarítmica. ( $\eta = 0.5$ ).

No Corolário 3.4.2 encontramos um estimador para o desvio padrão de  $\hat{\eta}_n(k)$ . Para percebermos se podemos utilizar este estimador para estimar o desvio padrão de  $\hat{\eta}_n(k)$  em amostras de dimensão finita, comparámos os valores obtidos com este estimador com as estimativas obtidas pela simulação multi-amostra. Para tal, calculámos para cada  $k$  e para cada amostra simulada do processo ARMAX estimativas do desvio padrão  $\hat{\sigma}_n(\hat{\eta}_n(k))$  usando o estimador do Corolário 3.4.2 e por fim considerámos para cada  $k$  a média dos valores obtidos  $\overline{\hat{\sigma}_n(\hat{\eta}_n(k))}$ . Na Figura 3.4, podemos visualizar os valores calculados e como podemos constatar pela aproximação entre as duas linhas,  $\hat{\sigma}_n(\hat{\eta}_n(k))$  parece ser um estimador viável para o desvio padrão de  $\hat{\eta}_n(k)$ , mesmo quando consideramos amostras de dimensão moderada.

Ilustramos na Figura 3.5 os intervalos de confiança gaussianos assintóticos a 95% para o estimador de *runs* do índice de cruzamentos apresentados em (3.34) e comparamo-los com os obtidos com as simulações de Monte Carlo. Como podemos verificar os intervalos de confiança assintóticos (3.34) (as linhas representam médias do limites dos intervalos de confiança) fornecem aproximações fidedignas dos intervalos de confiança exatos mesmo para amostras de dimensão moderada. Estes intervalos podem assim ser usados para avaliar a precisão das estimativas de  $\eta$  nas aplicações a dados reais.

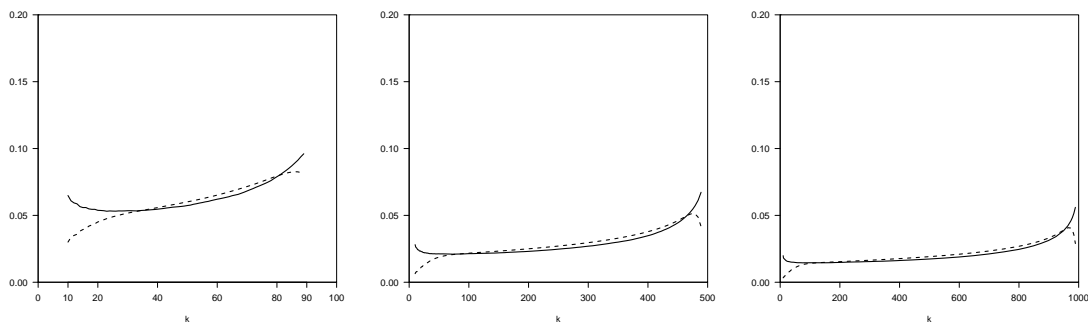


Figura 3.4: Valores estimados do desvio padrão de  $\hat{\eta}_n(k)$  obtidos pela simulação multi-amostra (linha a cheio) e valores de  $\widehat{\sigma}_n(\hat{\eta}_n(k))$  (linha tracejada) para amostras de dimensão  $n = 100, 500$  e  $1000$  do processo ARMAX dado por (3.35).

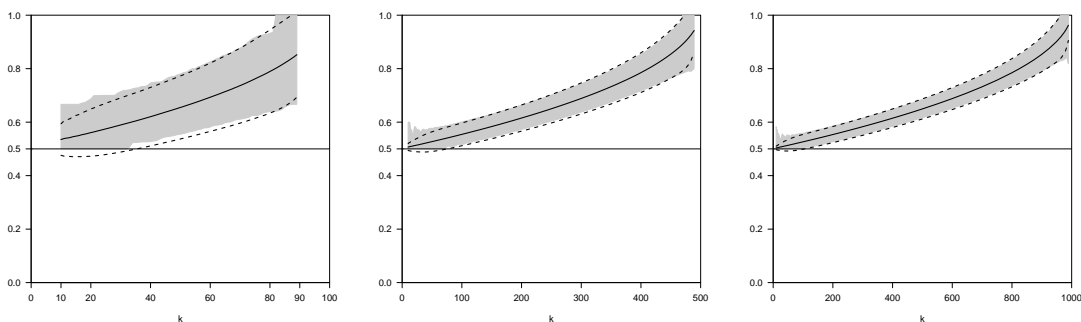


Figura 3.5: Valores médios estimados de  $\hat{\eta}_n(k)$  (linha a cheio) e intervalos de confiança a 95% obtidos a partir de (3.28) (linha tracejada) e as aproximações Monte Carlo dos intervalos de confiança a 95% (quantis 2.5% e 97.5%) (sombreado), para amostras de dimensão  $n = 100, 500$  e  $1000$  do processo ARMAX em (3.35).

### 3.6.2 O processo AR(1)

Consideremos agora o processo AR(1) de Chernick *et al.* (1991) definido no Exemplo 2.4.1. Como vimos na Secção 2.4 este processo verifica a condição  $\tilde{D}^{(3)}(\mathbf{u})$  com  $\mathbf{u} = \{u_n = 1 - \tau/n\}_{n \geq 1}$ ,  $\tau > 0$ , tem-se  $\lim_{n \rightarrow +\infty} nP(X_1 > u_n) = \tau = \nu = \lim_{n \rightarrow +\infty} P(X_1 \leq u_n < X_2)$  e  $\eta = \theta = 1 - 1/r^2$ .

A condição  $D''(\mathbf{u})$  não é geralmente verificada por este tipo de sucessões que tendem a oscilar rapidamente próximo de valores extremos. Para ilustrarmos esta característica, apresentamos na Figura 3.6 uma trajetória deste processo AR(1) com  $r = 2$ ,  $r = 3$  e  $r = 5$  ( $\eta = \theta = 0.75$ ,  $\eta = \theta = 0.89$  e  $\eta = \theta = 0.96$ , respetivamente).

Na Figura 3.7 podemos observar as trajetórias das estimativas de  $\eta$  obtidas a partir do estimador de *runs* do índice de cruzamentos, em função de  $k$ , para uma amostra de dimensão  $n = 5000$  do modelo AR(1), para diferentes valores de  $r$ .

Procedendo como no exemplo anterior apresentamos na Figura 3.8 os valores médios estimados para uma amostra de dimensão  $n = 500$  com os respetivos intervalos de confiança. Parece existir uma estabilização inicial das estimativas em torno dos verdadeiros valores de  $\eta$

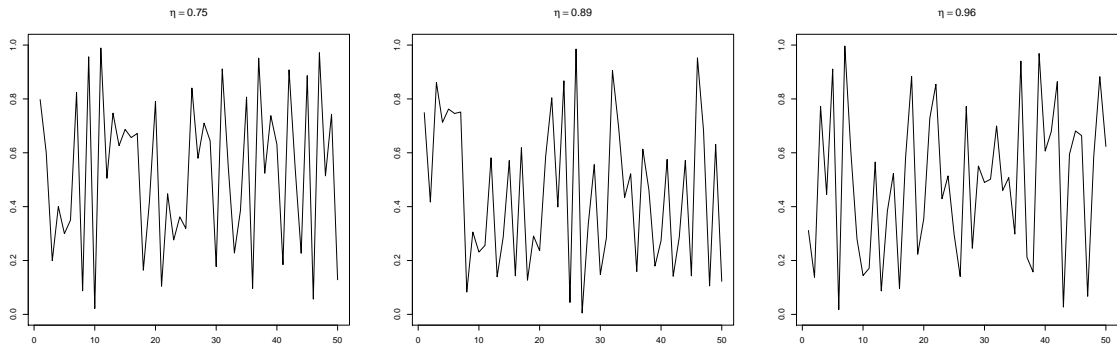


Figura 3.6: Trajetória do processo estacionário AR(1) definido por (2.13), com  $r = 2$  (à esquerda),  $r = 3$  (ao centro) e  $r = 5$  (à direita).

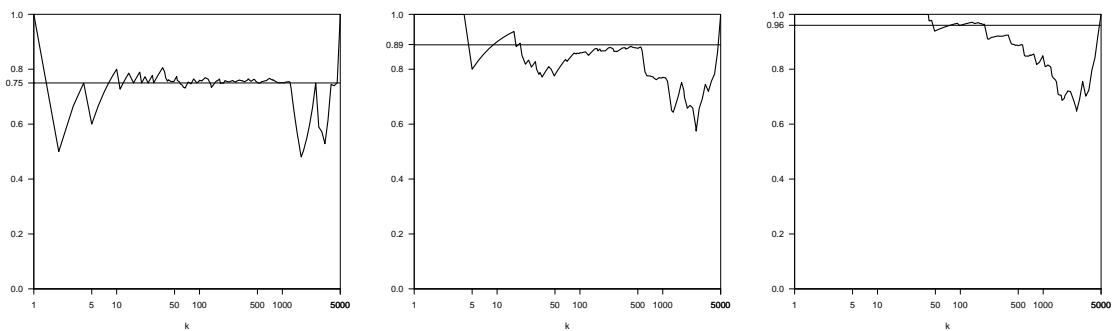


Figura 3.7: Estimativas de  $\eta$ , em função de  $k$ , para uma amostra de dimensão  $n = 5000$  do processo estacionário AR(1) em (2.13), com  $r = 2$  (à esquerda,  $\eta = 0.75$ ),  $r = 3$  (ao centro,  $\eta = 0.89$ ) e  $r = 5$  (à direita,  $\eta = 0.96$ ), em escala logarítmica.

assinalados por uma linha horizontal. A região de estabilidade vai diminuindo à medida que o valor de  $r$  aumenta, ou seja, para valores de  $\eta$  superiores. Os intervalos de confiança assintóticos (3.34) também aqui parecem fornecer aproximações fidedignas dos intervalos de confiança exatos mesmo para amostras de dimensão moderada, contudo para valores de  $\eta$  próximos de um, tal deixa de ser verdade como seria de esperar, pois nestes casos não podemos aplicar o Corolário 3.4.2.

Uma análise mais detalhada do comportamento do estimador de  $r_{uns}$  do índice de cruzamentos em processos AR(1) definidos em (2.13) é apresentada na Tabela 3.2.

Observando a Figura 3.9 conseguimos perceber que neste processo AR(1), para valores iniciais de  $k$  (mais pequenos) o erro quadrático médio é fortemente dominado pela variância, sendo o viés bastante reduzido. No entanto, a partir de determinados valores  $k$ , o viés tem um crescimento muito acentuado, passando a dominar claramente o erro quadrático médio.

Os dois exemplos apresentados permitem concluir que as estimativas do índice de cruzamentos obtidas, em função de  $k$ , apresentam padrões muito diferentes. Enquanto que no processo ARMAX as estimativas tendem a aumentar com o valor de  $k$ , para o processo AR(1) o gráfico das estimativas em função de  $k$  decresce até ao verdadeiro valor de  $\eta$  e a partir daí as estimativas

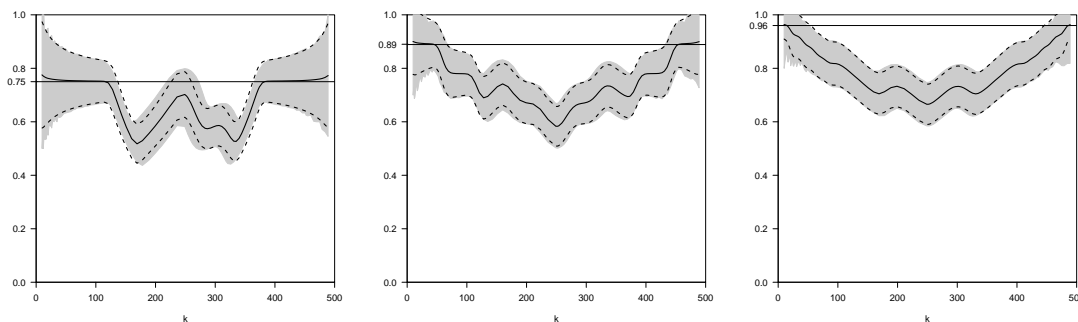


Figura 3.8: Valores médios estimados de  $\hat{\eta}_n(k)$  (linha a cheio) e intervalos de confiança a 95% obtidos a partir de (3.28) (linha tracejada) e as aproximações Monte Carlo dos intervalos de confiança a 95% (quantis 2.5% e 97.5%) (sombreado), para amostras de dimensão  $n = 500$  do processo AR(1) em (2.13), para  $r = 2$  (à esquerda),  $r = 3$  (ao centro) e  $r = 5$  (à direita).

Tabela 3.2: Fração ótima de amostragem, valores médios, erros quadráticos médios e desvios padrão do estimador nos seus níveis ótimos, para o processo AR(1) em (2.13), com  $\eta = 0.75$ ,  $\eta = 0.89$  e  $\eta = 0.96$ .

$n$	$k_0$	$k_0/n$	$E[\bullet]$	$MSE[\bullet]$	$SD[\bullet]$
<b><math>\eta = 0.75</math></b>					
100	25	0.250	$0.73998 \pm 0.000668$	$0.00663 \pm 0.000109$	$0.08081 \pm 0.000641$
200	50	0.250	$0.73659 \pm 0.000453$	$0.00330 \pm 0.000031$	$0.05582 \pm 0.000210$
500	123	0.246	$0.74389 \pm 0.000197$	$0.00131 \pm 0.000017$	$0.03567 \pm 0.000229$
1000	247	0.247	$0.74502 \pm 0.000203$	$0.00064 \pm 0.000010$	$0.02487 \pm 0.000200$
2000	495	0.248	$0.74647 \pm 0.000127$	$0.00033 \pm 0.000005$	$0.01772 \pm 0.000116$
5000	1175	0.235	$0.75023 \pm 0.000083$	$0.00016 \pm 0.000001$	$0.01261 \pm 0.000057$
<b><math>\eta = 0.89</math></b>					
100	11	0.110	$0.87571 \pm 0.001000$	$0.01045 \pm 0.000123$	$0.10138 \pm 0.000557$
200	20	0.100	$0.88229 \pm 0.000375$	$0.00541 \pm 0.000076$	$0.07323 \pm 0.000510$
500	48	0.096	$0.88814 \pm 0.000338$	$0.00211 \pm 0.000029$	$0.04595 \pm 0.000314$
1000	98	0.098	$0.88860 \pm 0.000199$	$0.00103 \pm 0.000016$	$0.03212 \pm 0.000254$
2000	202	0.101	$0.88869 \pm 0.000148$	$0.00050 \pm 0.000005$	$0.02232 \pm 0.000116$
5000	501	0.100	$0.88911 \pm 0.000103$	$0.00020 \pm 0.000002$	$0.01403 \pm 0.000058$
<b><math>\eta = 0.96</math></b>					
100	2	0.020	$0.97854 \pm 0.000841$	$0.01061 \pm 0.000353$	$0.10129 \pm 0.001911$
200	2	0.010	$0.98044 \pm 0.000540$	$0.00982 \pm 0.000227$	$0.09692 \pm 0.001315$
500	2	0.004	$0.97961 \pm 0.000896$	$0.01016 \pm 0.000376$	$0.09883 \pm 0.002117$
1000	34	0.034	$0.95664 \pm 0.000195$	$0.00125 \pm 0.000011$	$0.03527 \pm 0.000150$
2000	66	0.033	$0.95943 \pm 0.000188$	$0.00060 \pm 0.000007$	$0.02441 \pm 0.000143$
5000	171	0.034	$0.96002 \pm 0.000131$	$0.00023 \pm 0.000003$	$0.01512 \pm 0.000111$

tornam-se um pouco instáveis devido essencialmente ao elevado valor do viés.

Nestes processos, as estimativas apresentam alguma estabilização em torno dos verdadeiros valores de  $\eta$  para valores pequenos de  $k$ . Para estes valores, os gráficos dos erros quadráticos médios apresentam um padrão semelhante ao que usualmente se denomina por “banheira”, ou em forma de U, que nos permite identificar claramente os valores em que se atinge o menor erro quadrático médio. No caso ARMAX esta forma é mais evidente do que no processo AR(1).

Para valores pequenos de  $k$ , ou equivalentemente para níveis elevados, obtivemos boas

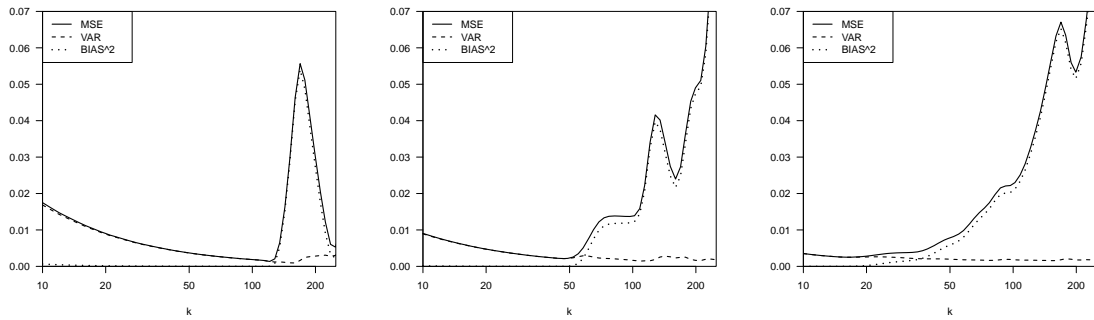


Figura 3.9: Erros quadráticos médios, variâncias e quadrado do viés estimados (em escala logarítmica) de  $\hat{\eta}_n(k)$ , a partir de amostras de dimensão  $n = 500$  do processo AR(1) em (2.13), para  $r = 2$  (à esquerda),  $r = 3$  (ao centro) e  $r = 5$  (à direita).

estimativas de  $\eta$ . Relembramos que o problema prático da escolha da estatística ordinal de topo a usar é um problema complexo da estatística de extremos.

A sensibilidade do valor médio relativamente às alterações no valor de  $k$ , parece evidenciar a necessidade de serem estudadas as propriedades do viés do estimador. Contudo, para os valores de  $k$  considerados, coincidentes com as zonas de decrescimento do erro quadrático médio nos gráficos das Figuras 3.3 e 3.9 (valores iniciais de  $k$ ), é sempre a variância que maior influência tem no erro quadrático médio. Será assim importante tentar encontrar alterações ao estimador de *runs* que consigam melhorar o seu desempenho em termos de redução da variância.

### 3.7 Método *Bootstrap*

A estimação de parâmetros, num contexto de dados reais, é muitas vezes dificultada pela impossibilidade de, na prática, compararmos os resultados obtidos com resultados teóricos calculados a partir de modelos probabilísticos. Neste sentido, propomos um método *bootstrap* para avaliar a incerteza da estimação, uma vez que permitirá calcular intervalos de confiança para as estimativas de  $\eta$  obtidas.

O método proposto usa as técnicas de *bootstrap* adaptadas à reamostragem por blocos de modo a preservar a dependência local das sucessões, e é uma adaptação do método proposto por Ferro e Segers (2003), na estimação do índice extremal.

Atendendo a que, no caso de existir, o processo pontual limite da sucessão de processos pontuais de cruzamentos  $\{\tilde{N}_n\}_{n \geq 1}$  é um processo de Poisson composto que corresponde à soma aleatória de variáveis independentes que representam o tamanho dos agrupamentos de cruzamentos, poderemos decompor a sucessão observada em duas partições distintas: por um lado temos grupos assintoticamente independentes de cruzamentos e por outro lado temos grupos assintoticamente independentes de não-cruzamentos. Se definirmos a forma como se deter-

minam os grupos de cruzamentos que consideramos independentes, teremos a possibilidade de usar o bootstrap, reamostrando com reposição elementos de um grupo e intercalando com elementos selecionados aleatoriamente com reposição do outro grupo, formando assim uma réplica bootstrap. Na prática procedemos da seguinte forma:

- a) Considerar que num grupo de cruzamentos, os cruzamentos estão separados no máximo por  $t$  não-cruzamentos;
- b) Identificar os  $c$  grupos de cruzamentos;
- c) Identificar os  $(c - 1)$  grupos de não-cruzamentos (desprezamos as observações iniciais e finais que não pertencem ao grupo de cruzamentos);
- d) Selecionar aleatoriamente e com reposição  $c$  grupos de cruzamentos a partir dos grupos identificados em b) e intercalar com  $(c - 1)$  grupos de não-cruzamentos selecionados aleatoriamente com reposição a partir dos grupos identificados em c);
- e) Repetir o ponto d)  $B$  vezes e calcular  $\hat{\eta}_n(k)$  em cada uma das  $B$  repetições;
- f) A partir das estimativas  $(\hat{\eta}_n(k)_{(1)}, \hat{\eta}_n(k)_{(2)}, \dots, \hat{\eta}_n(k)_{(B)})$  que constituem a distribuição empírica de  $\hat{\eta}_n(k)$  calculam-se os intervalos de confiança recorrendo aos quantis  $(\alpha)$  e  $(1 - \alpha)$  desta distribuição empírica. As estimativas  $\hat{\eta}_n(k)$  são calculadas por  $\frac{1}{B} \sum_{i=1}^B \hat{\eta}_n(k)_{(i)}$ .

Notemos que assumindo a validade da condição  $\tilde{D}^{(3)}(\mathbf{u})$  os grupos de cruzamentos são assintoticamente *runs* de cruzamentos, ou seja, admitindo  $t = 0$  no algoritmo anterior, temos a justificação assintótica do método bootstrap neste caso.

## 3.8 Aplicações a dados reais

Apresentamos nesta secção alguns resultados da estimação do índice de cruzamentos em três conjuntos de dados reais.

### 3.8.1 Dados meteorológicos: Ozono

Consideramos a seguir um conjunto de dados que denominámos por “Ozono”, correspondente ao registo de  $n = 120$  máximos semanais da concentração média, por hora, de ozono (em partes por milhão) na área da baía de São Francisco nos Estados Unidos da América (EUA). Os dados estão disponíveis no *package Xtremes* (Reiss e Thomas (2007)) e foram utilizados por Gomes *et al.* (2008) como exemplo de aplicação de várias metodologias de estimação do índice extremal. Na Figura 3.10 encontramos a representação destes dados.

Como na prática não é possível validarmos a condição  $\tilde{D}^{(3)}$ , recorreremos aos argumentos de Gomes *et al.* (2008), onde afirmam que a maioria dos modelos paramétricos ajustados a estes tipos de dados meteorológicos satisfazem a condição  $D''$ , consequentemente também verificarão a condição  $\tilde{D}^{(3)}$ .

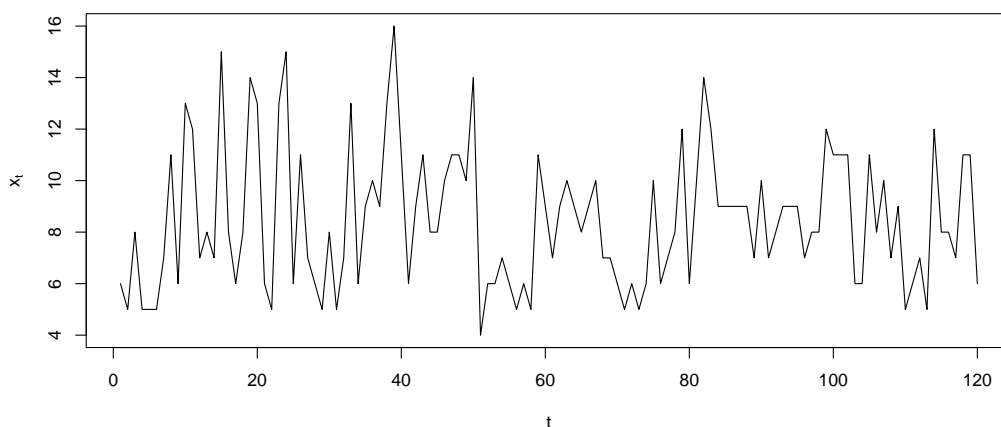


Figura 3.10: Valores (em partes por milhão) de máximos semanais da concentração média, por hora, de ozono verificada na área da baía de São Francisco nos EUA.

Na Figura 3.11 apresentamos as estimativas de  $\eta$  e na Figura 3.12 apresentamos as estimativas bootstrap com  $B = 5000$  réplicas e os respectivos intervalos de confiança a 95%.

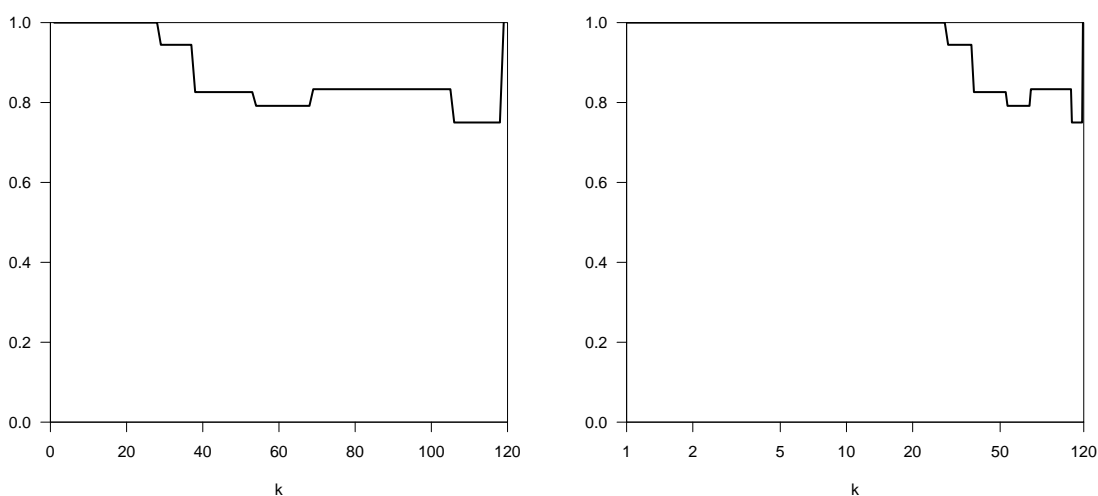


Figura 3.11: Estimativas de  $\eta$  para máximos semanais da concentração média, por hora, de ozono (em partes por milhão) na área da baía de São Francisco nos EUA, numa escala linear (à esquerda) e numa escala logarítmica (à direita).

A estabilidade das estimativas para valores pequenos de  $k$  em torno de um, vai ao encontro das conclusões retiradas em Gomes *et al.* (2008) para este conjunto de dados, onde assumiram a validade de  $D''$  e portanto  $\eta = 1$ . No entanto, a dimensão reduzida da amostra disponível não nos permite dar grande credibilidade às estimativas pontuais obtidas, uma vez que neste caso o número de cruzamentos e o número de cruzamentos seguidos de não cruzamentos do nível envolvido, a partir dos quais obtemos as estimativas é muito reduzido.

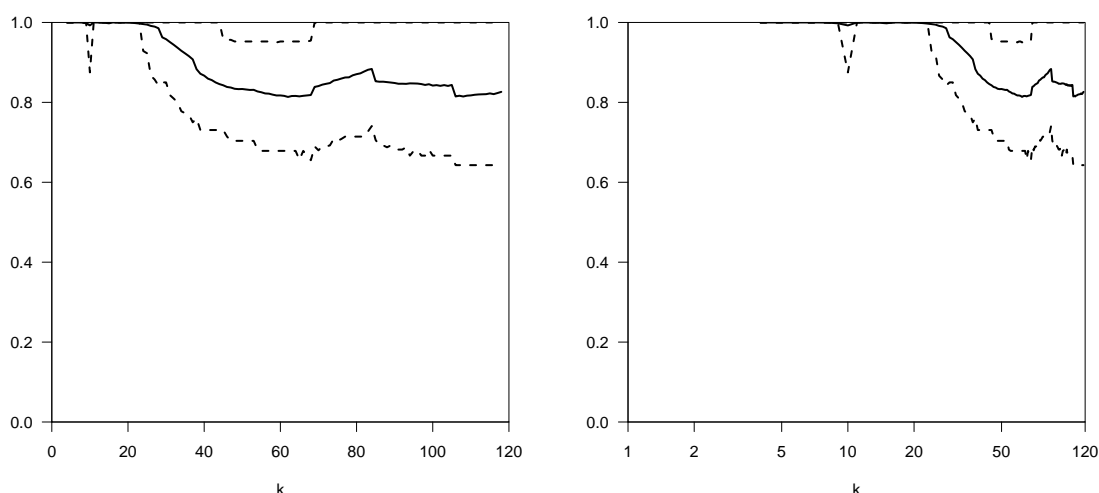


Figura 3.12: Estimativas bootstrap ( $B = 5000$ ) de  $\hat{\eta}$  (a cheio) e respetivos intervalos de confiança a 95% (a tracejado) em função de  $k$  dos máximos semanais da concentração média, por hora, de ozono (em partes por milhão) na área da baía de São Francisco nos EUA, numa escala linear (à esquerda) e numa escala logarítmica (à direita).

### 3.8.2 Dados financeiros: DAX30

Os dados provenientes de séries temporais financeiras apresentam geralmente elevada volatilidade, o que torna muito pouco verosímil a satisfação da condição  $D''$  pela respetiva sucessão geradora, tornando-os aliciantes neste contexto de estimação de  $\eta$  sob a validade da condição  $\tilde{D}^{(3)}$ .

Analisamos assim nesta secção dados do índice bolsista DAX30, calculados a partir das cotações das 30 maiores empresas da bolsa de Frankfurt. Em particular, iremos estudar os  $n = 3185$  valores diários de fecho do DAX30 no período entre 03-01-2000 e 29-06-2012, obtidos no endereço eletrónico <http://www.markt-daten.de/daten/2006dax.txt>. A estes dados foram retiradas 7 observações porque os valores de cotação do índice no dia eram idênticas às cotações do dia anterior. Os valores do índice coincidem quando, por algum motivo, o mercado bolsista não abriu num determinado dia.

Este tipo de dados financeiros são geralmente bem modelados por processos GARCH(1,1). Estes processos definem-se por

$$X_t = \sigma_t Z_t, \quad t \in \mathbb{Z}$$

onde  $\{Z_t\}_{t \in \mathbb{Z}}$  é uma sucessão de variáveis aleatórias iid simétricas com  $E[Z_1^2] = 1$ . A sucessão  $\{\sigma_t\}_{t \in \mathbb{Z}}$ , geralmente denominada por volatilidade estocástica de  $X_t$ , é constituída por variáveis aleatórias não negativas, tais que  $Z_t$  e  $\sigma_t$  são independentes para qualquer valor fixo de  $t$ , e satisfaz

$$\sigma_t^2 = \alpha_0 + \beta_1 \sigma_{t-1}^2 + \alpha_1 X_{t-1}^2 = \alpha_0 + \sigma_{t-1}^2 (\beta_1 + \alpha_1 Z_{t-1}^2), \quad t \in \mathbb{Z}, \quad (3.36)$$

com  $\alpha_0$ ,  $\alpha_1$  e  $\beta_1$  parâmetros não negativos.

Por Mikosch e Starica (2000), temos que se

$$\alpha_0 > 0 \quad \text{e} \quad E[\ln(\alpha_1 Z_1^2 + \beta_1)] < 0 \quad (3.37)$$

$\{X_t\}_{t \in \mathbb{Z}}$  é uma sucessão estacionária. Além do mais, se  $Z_1$  tem densidade positiva em  $\mathbb{R}$ ,  $\{X_t\}_{t \in \mathbb{Z}}$  satisfaz a condição de mistura, usualmente denominada por  $\alpha$ -mixing, com decaimento geométrico então também satisfaz as condições  $\Delta(\cdot)$  e  $D(\cdot)$  (Mikosch e Starica (2004)).

Na Figura 3.13 apresentamos os valores diários de fecho do índice DAX30 no período em consideração,  $x_t$ , e os respectivos *log-returns* definidos por  $100 \times (\ln x_t - \ln x_{t-1})$ , que correspondem aos dados a analisar.

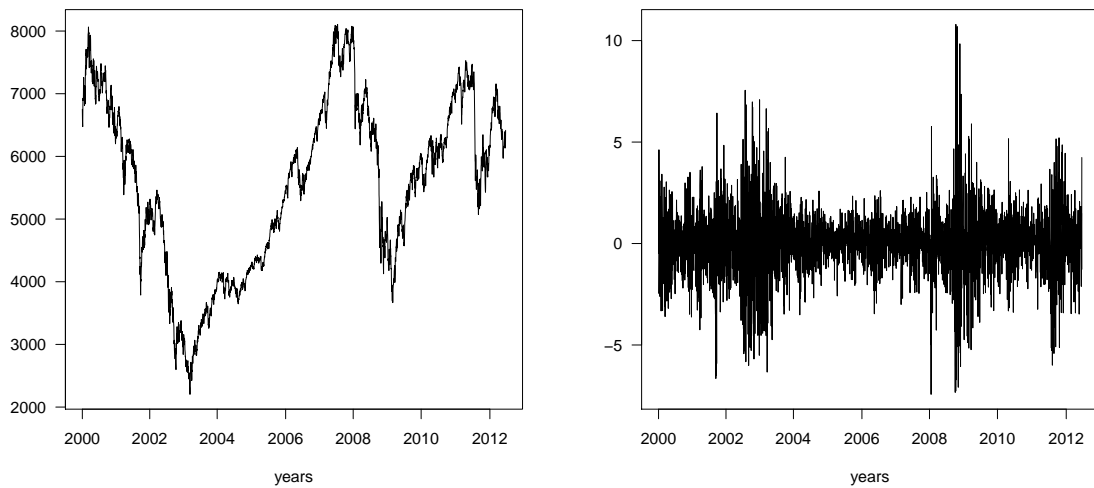


Figura 3.13: Dados diários de fecho do índice DAX30 no período em análise (à esquerda) e respectivos *Log-returns* (à direita).

Uma descrição estatística completa de parte destes dados poderá ser encontrada em Klar *et al.* (2012). Os autores analisaram três séries estatísticas do DAX30 em períodos diferentes: entre 1991 e 1998; entre 2005 e 2007; e entre 2007 e 2009. Nestas três séries não rejeitaram a hipótese de os dados serem provenientes de um processo GARCH(1,1), admitindo que  $Z_1$  tem distribuição *t-student* e os parâmetros estimados em (3.36) verificam sempre  $\hat{\alpha}_0 > 0$  e  $\hat{\alpha}_1 + \hat{\beta}_1 < 1$ , o que empiricamente não contradiz (3.37).

As Figuras 3.14 e 3.15 parecem indicar um valor de  $\eta$  diferente de um, o que está de acordo com o facto de ser pouco provável que se verifique a condição  $D''$ . As estimativas apresentam alguma estabilidade em torno do valor  $\eta = 0.85$ .

Mikosch e Starica (2000) determinaram o valor do índice extremal para o quadrado dos processos GARCH(1,1), enquanto que Laurini e Tawn (2006) propõem um algoritmo para a determinação do índice extremal em processos GARCH(1,1), com inovações seguindo uma distribuição

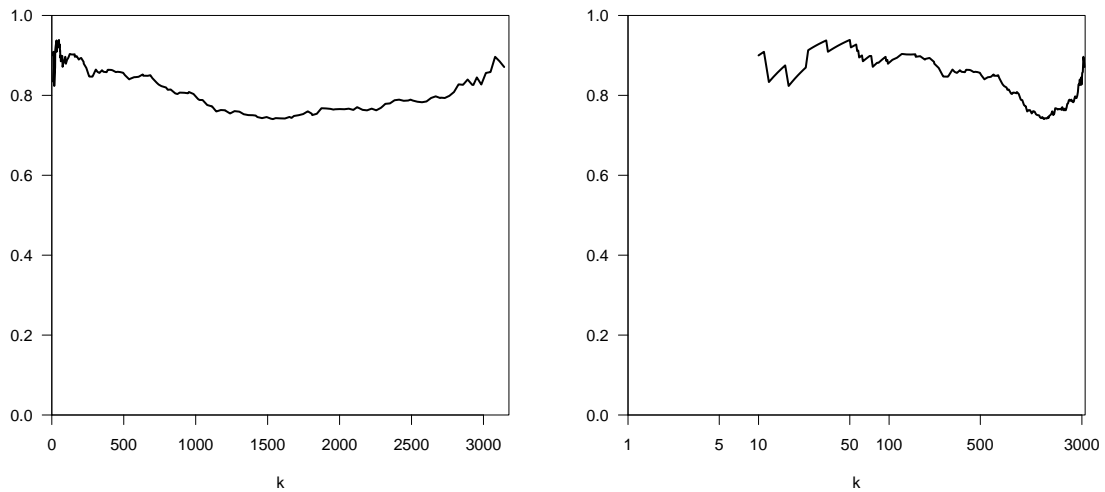


Figura 3.14: Estimativas de  $\eta$  para os *log-returns* dos valores diários de fecho do índice DAX30 no período em análise, em função de  $k$ , numa escala linear (à esquerda) e numa escala logarítmica (à direita).

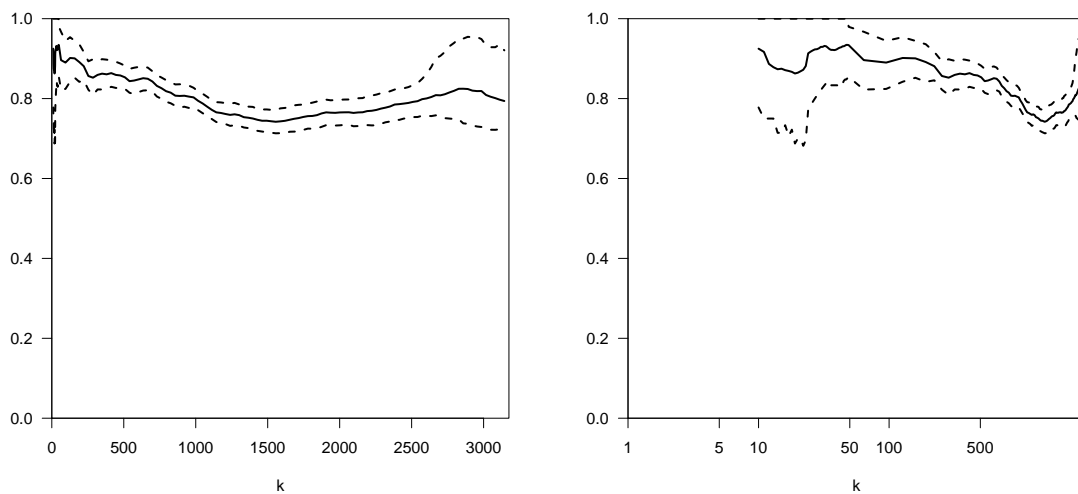


Figura 3.15: Estimativas bootstrap ( $B = 5000$ ) de  $\eta$  (a cheio) e respetivos intervalos de confiança a 95% (a tracejado) para os *log-returns* dos valores diários de fecho do índice DAX30 no período em análise, em função de  $k$ , numa escala linear (à esquerda) e numa escala logarítmica (à direita).

*t-student*. Seria interessante, num trabalho futuro, obter resultados similares para o índice de cruzamentos de processos GARCH(1,1).

### 3.8.3 Dados hídricos: Caudal efluente numa barragem

Consideramos agora os dados disponíveis no endereço <http://www.snirh.pt> referentes ao caudal efluente <sup>1</sup> médio diário (em  $m^3/s$ ), verificado na barragem do Fratel nos meses de

<sup>1</sup>Caudal que sai da barragem.

novembro a março, entre 01-01-1984 e 30-11-2008. Este conjunto de dados que apresentamos na Figura 3.16, depois de filtrado para exclusão dos valores nulos que significam que a barragem não libertou caudal, contém 3513 observações. Por questões de estacionaridade considerámos apenas os meses de novembro a março.

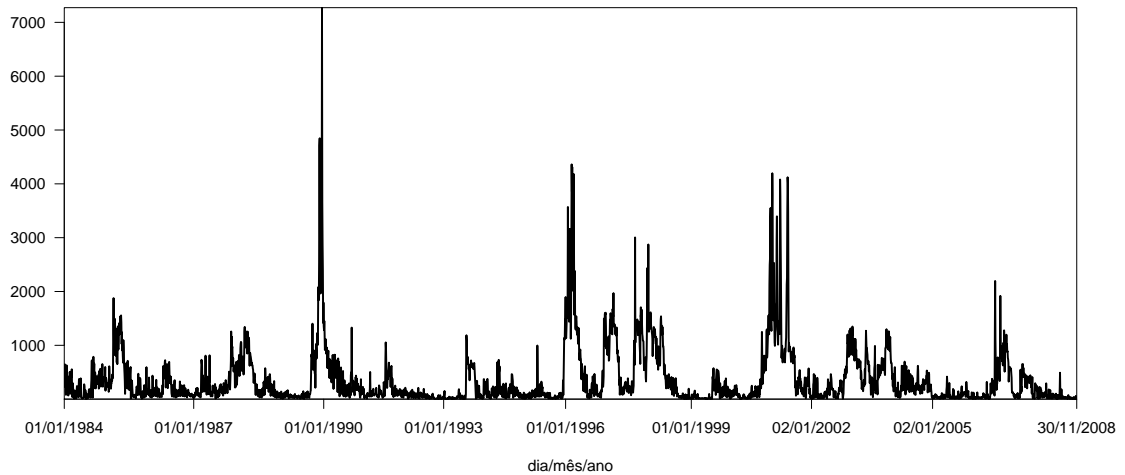


Figura 3.16: Valores do caudal efluente médio diário (em  $m^3/s$ ), na barragem do Fratel nos meses de novembro a março, entre 01-01-1984 e 30-11-2008.

Os gráficos das Figuras 3.17 e 3.18 revelam que para estes dados teremos agrupamentos de cruzamentos, pois as estimativas obtidas de  $\eta$  não nos parecem indicar que o valor do índice de cruzamentos seja igual a um, contudo as estimativas apresentam alguma estabilidade para valores de  $\eta$  não muito distantes de um. Notemos que para  $k$  pequeno, o valor um pertence a muitos intervalos de confiança, como se poderá observar na Figura 3.18.

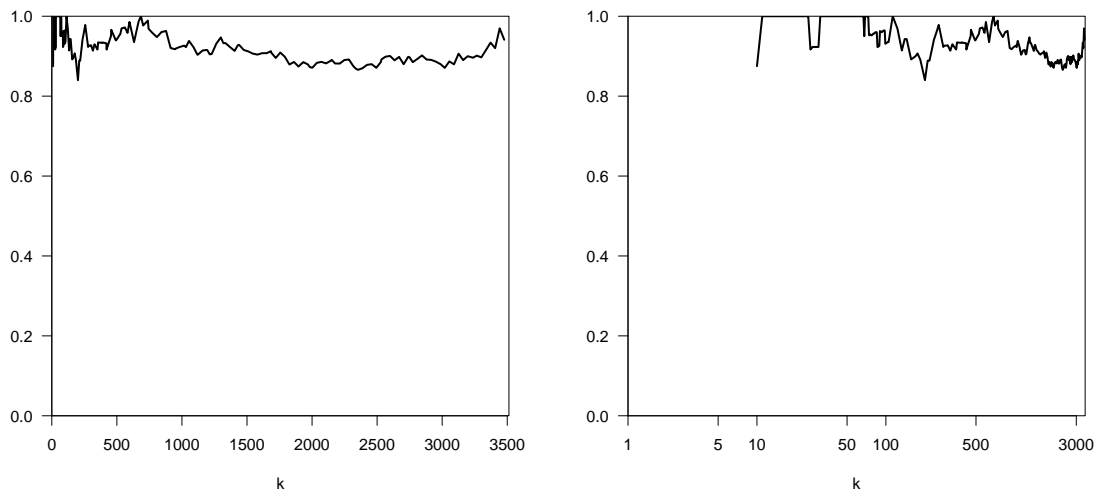


Figura 3.17: Estimativas de  $\eta$  para o caudal efluente médio diário (em  $m^3/s$ ), na barragem do Fratel no período em análise, numa escala linear (à esquerda) e numa escala logarítmica (à direita).

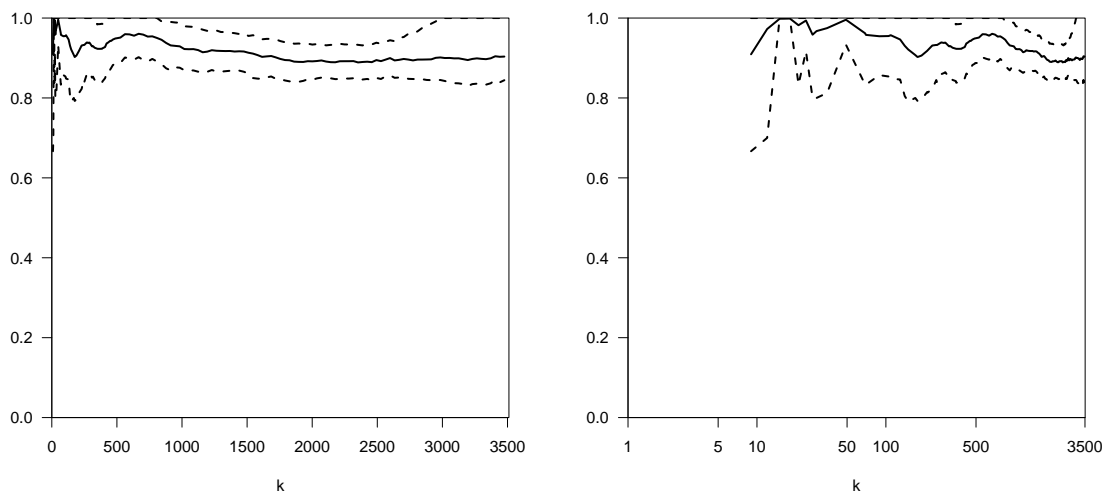


Figura 3.18: Estimativas bootstrap ( $B = 5000$ ) de  $\eta$  (a cheio) e respetivos intervalos de confiança a 95% (a tracejado) do caudal efluente médio diário (em  $m^3/s$ ), na barragem do Fratel no período em análise, numa escala linear (à esquerda) e numa escala logarítmica (à direita).

# Capítulo 4

## Estimadores de blocos do índice de cruzamentos

### 4.1 Introdução

A validade da condição  $\tilde{D}^{(3)}(\mathbf{u})$ , exigida no capítulo anterior, para a definição do estimador de *runs* é bastante restritiva, uma vez que impõe que num grupo de cruzamentos estes deverão ocorrer de forma sucessiva. O número de processos em que o estimador de *runs* poderá ser aplicado é assim bastante limitado e conseqüentemente poderá questionar-se a sua aplicabilidade em termos práticos. Fará então todo o sentido encontrar estimadores que possam ser usados em condições mais abrangentes, *i.e.*, estimadores que não exijam a validade de nenhuma condição de dependência local. Este será o objetivo deste capítulo.

Os estimadores do índice extremal definidos segundo a metodologia dos blocos (e.g. Hsing (1991), Weissman e Novak (1998), Robert (2009), Robert *et al.* (2009), Drees (2011)) não exigem que a sucessão estacionária  $X$  verifique alguma condição de dependência local. A gênese deste método reside na partição da amostra  $X_1, \dots, X_n$  em blocos com uma certa dimensão  $r_n$  conveniente, a partir dos quais se definem estatísticas capazes de estimar consistentemente o parâmetro em causa. Nos trabalhos de Hsing (1991), encontramos o estimador de blocos do índice extremal, obtido a partir do facto de o índice extremal ser assintoticamente igual ao inverso do tamanho médio dos grupos de excedências. Weissman e Novak (1998) demonstraram a consistência e a normalidade assintótica deste estimador impondo condições mais fracas do que as impostas por Hsing (1991). Desde então, outros estimadores têm surgido dentro desta abordagem, nomeadamente Robert *et al.* (2009), com base na definição do índice extremal, propuseram dois estimadores: um baseado na utilização de blocos disjuntos, como no estimador de blocos de Hsing (1991) e outro baseado em blocos deslizantes. O estimador de blocos deslizantes tem uma menor variância assintótica e por conseguinte é preferível ao estimador de blocos disjuntos.

Tendo por base os estimadores de blocos propostos para o índice extremal propomos neste capítulo três novos estimadores para o índice de cruzamentos. Assim, na Secção 4.2 apresentamos o estimador de blocos do índice de cruzamentos, que denotamos por  $\hat{\eta}_n^B$ , e demonstramos algumas das suas propriedades. A Secção 4.3, está reservada aos estimadores de blocos disjuntos e de blocos deslizantes do índice de cruzamentos, denotados respetivamente por  $\hat{\eta}_n^{dj}$  e  $\hat{\eta}_n^{sl}$ . Para cada um destes estimadores provamos a consistência e a normalidade assintótica.

Para os modelos e os dados reais do capítulo anterior, efetuamos, no término deste capítulo, um estudo comparativo da eficiência dos estimadores propostos, do estimador de *runs* do capítulo anterior e do o estimador de  $\eta$  sugerido pela relação entre o índice de cruzamentos e

o índice extremal (1.26), *i.e.*, quando estimamos  $\eta$  a partir de um estimador consistente de  $\theta$  e estimadores consistentes do número médio de excedências,  $\tau$ , e de cruzamentos,  $\nu$ , de um nível elevado.

## 4.2 O estimador de blocos $\hat{\eta}_n^B$

A relação (2.8), para  $T = 1$ , obtida na Secção 2.3 do capítulo 2, indica-nos que, sob a validade da condição  $\Delta(\tilde{u}^{(\nu)})$ , se a sucessão estacionária  $X$  tem índice de cruzamentos,  $\eta > 0$ , então este surge como o inverso do limite do número médio de pontos em cada grupo de cruzamentos, estando a definição de grupo de cruzamentos associada a uma partição da amostra  $X_1, \dots, X_n$  em blocos de uma certa dimensão  $r_n$  conveniente.

Notemos que a relação (2.8), para  $T = 1$ , é equivalente a

$$\begin{aligned} \frac{1}{\eta} &= \lim_{n \rightarrow +\infty} E[\tilde{N}_{r_n}(u_n) \mid \tilde{N}_{r_n}(u_n) > 0] \\ &= \lim_{n \rightarrow +\infty} \frac{r_n P(X_1 \leq u_n < X_2)}{P(\tilde{N}_{r_n}(u_n) > 0)} \end{aligned} \quad (4.1)$$

onde  $\tilde{N}_{r_n}(u_n) = \tilde{N}_n(u_n)([0, r_n/n])$ , desde que  $u_n = u_n^{(\nu)}$ ,  $n \geq 1$ , seja uma sucessão de níveis normalizados e  $r_n = [n/k_n]$  com  $k = \{k_n\}_{n \geq 1}$  satisfazendo (1.7). Esta relação sugere uma forma de estimar  $1/\eta$ : contar o número total de cruzamentos do nível  $u_n$  na amostra  $X_1, \dots, X_n$  e dividir pelo número de blocos em que ocorre pelo menos um cruzamento. Com base nesta ideia, propomos o seguinte estimador de blocos para o índice de cruzamentos,  $\eta$ ,

$$\hat{\eta}_n^B(u_n) := \frac{\sum_{i=1}^{k_n} \mathbb{I}_{\{\tilde{N}_{r_n}^{(i)}(u_n) > 0\}}}{\sum_{i=1}^{n-1} \mathbb{I}_{\{X_i \leq u_n < X_{i+1}\}}}, \quad (4.2)$$

onde  $\tilde{N}_{r_n}^{(i)}(u_n) = \tilde{N}_n(u_n)([(i-1)r_n+1/n, ir_n/n])$  representa o número de cruzamentos do nível  $u_n$  no  $i$ -ésimo bloco.

Para provarmos a consistência deste estimador é necessário considerarmos, como no capítulo anterior para o estimador de *runs* do índice de cruzamentos, níveis  $v_n = \tilde{u}_{[n/c_n]}^{(\nu)}$ ,  $n \geq 1$ , verificando (3.8). No próximo resultado provamos que para estes níveis ainda se tem a validade de (4.1).

**Teorema 4.2.1** *Suponhamos que a sucessão estacionária  $X$  tem índice de cruzamentos  $\eta > 0$  e sejam  $\{r_n\}_{n \geq 1}$ ,  $\{k_n\}_{n \geq 1}$  e  $\{c_n\}_{n \geq 1}$  sucessões de números reais tais que*

$$c_n \rightarrow +\infty, \quad k_n \rightarrow +\infty, \quad r_n \rightarrow +\infty, \quad c_n/k_n \rightarrow 0, \quad k_n r_n \sim n, \quad \text{quando } n \rightarrow +\infty.$$

*Seja  $\mathbf{v} = \{v_n\}_{n \geq 1}$  uma sucessão de níveis definidos por (3.8) ( $nP(X_1 \leq v_n < X_2) - c_n \nu \xrightarrow[n \rightarrow +\infty]{} 0$ ) e suponhamos que existe  $l_n = o(r_n)$  tal que  $\frac{k_n}{c_n} \alpha_{n, l_n} \xrightarrow[n \rightarrow +\infty]{} 0$ , onde  $\alpha_{n, l_n}$ ,  $n \geq 1$ , são os coeficientes associados à condição  $\Delta(\mathbf{v})$ .*

Então

$$\lim_{n \rightarrow +\infty} \frac{k_n}{c_n} P(\tilde{N}_{r_n}(v_n) > 0) = \eta\nu, \quad (4.3)$$

onde  $\tilde{N}_{r_n}(v_n) = \tilde{N}_{r_n}^{(1)}(v_n)$ , e (4.1) é válida para  $v_n$ .

**Demonstração:** Como a sucessão de níveis  $\mathbf{v} = \{v_n\}_{n \geq 1}$  é normalizada em relação à amostra  $X_1, \dots, X_{[n/c_n]}$ , se considerarmos esta amostra dividida em  $k_n/c_n$  blocos disjuntos podemos concluir que

$$P(\tilde{N}_{[n/c_n]}(v_n) = 0) \sim P^{k_n/c_n}(\tilde{N}_{r_n}(v_n) = 0), \quad (4.4)$$

tendo em conta o Lema 2.1 de H. Ferreira (2006), que garante a independência assintótica de cruzamentos em intervalos disjuntos. Mas, pela existência de índice de cruzamentos  $\eta > 0$  temos que  $P(\tilde{N}_{[n/c_n]}(v_n) = 0) \xrightarrow{n \rightarrow +\infty} e^{-\eta\nu}$  que conjuntamente com (4.4) prova (4.3).

Agora, tendo em conta (4.3) e a definição dos níveis  $v_n$  obtemos

$$E[\tilde{N}_{r_n}(u_n) \mid \tilde{N}_{r_n}(u_n) > 0] \sim \frac{r_n c_n \nu / n}{c_n \eta \nu / k_n},$$

o que completa a demonstração, pois por hipótese  $k_n r_n \sim n$ . □

No que se segue iremos estudar as propriedades do estimador  $\hat{\eta}_n^B = \hat{\eta}_n^B(v_n)$ , com  $\mathbf{v} = \{v_n\}_{n \geq 1}$  a sucessão de níveis definidos por (3.8).

#### 4.2.1 Consistência fraca

Impondo condições análogas às impostas por Weissman e Novak (1998) para obter a consistência do estimador dos blocos do índice extremal, obtemos no próximo resultado a consistência do estimador dos blocos do índice de cruzamentos  $\hat{\eta}_n^B = \hat{\eta}_n^B(v_n)$ .

**Teorema 4.2.2** *Suponhamos válidas as condições do Teorema 4.2.1,*

$$\eta_n^B := \frac{P(\tilde{N}_{r_n}(v_n) > 0)}{r_n P(X_1 \leq v_n < X_2)} \xrightarrow{n \rightarrow +\infty} \eta \quad (4.5)$$

e

$$\gamma_n := \sum_{i=1}^{k_n} \frac{k_n - i}{k_n} \left( P(\tilde{N}_{r_n}^{(i+1)}(v_n) > 0 \mid \tilde{N}_{r_n}(v_n) > 0) - P(\tilde{N}_{r_n}(v_n) > 0) \right) = o(nP(X_1 \leq v_n < X_2)). \quad (4.6)$$

Então

$$\hat{\eta}_n^B(v_n) \xrightarrow{n \rightarrow +\infty} \eta.$$

**Demonstração:** Atendendo à definição dos níveis  $v_n$ ,  $n \geq 1$ , e ao facto de  $k_n r_n \sim n$  temos que

$$E \left[ \sum_{i=1}^{k_n} \mathbb{I}_{\{\tilde{N}_{r_n}^{(i)}(v_n) > 0\}} \right] = k_n \eta_n^B r_n P(X_1 \leq v_n < X_2) \sim \eta_n^B n P(X_1 \leq v_n < X_2) \xrightarrow{n \rightarrow +\infty} +\infty.$$

Por outro lado, de (4.6) resulta

$$\begin{aligned} & Var \left[ \sum_{i=1}^{k_n} \mathbb{I}_{\{\tilde{N}_{r_n}^{(i)}(v_n) > 0\}} \right] \\ &= \sum_{i=1}^{k_n} Var \left[ \mathbb{I}_{\{\tilde{N}_{r_n}^{(i)}(v_n) > 0\}} \right] + 2 \sum_{i=1}^{k_n-1} \sum_{j=i+1}^{k_n} Cov \left( \mathbb{I}_{\{\tilde{N}_{r_n}^{(i)}(v_n) > 0\}}, \mathbb{I}_{\{\tilde{N}_{r_n}^{(j)}(v_n) > 0\}} \right) \\ &= k_n P(\tilde{N}_{r_n}(v_n) > 0) [1 - P(\tilde{N}_{r_n}(v_n) > 0) + 2\gamma_n] \\ &= o((nP(X_1 \leq v_n < X_2))^2). \end{aligned}$$

Consequentemente, pela Desigualdade de Chebychev vem, para todo  $\varepsilon > 0$ ,

$$P \left( \left| \frac{\sum_{i=1}^{k_n} \mathbb{I}_{\{\tilde{N}_{r_n}^{(i)}(u_n) > 0\}}}{k_n P(\tilde{N}_{r_n}(v_n) > 0)} - 1 \right| \geq \varepsilon \right) \leq \frac{k_n P(\tilde{N}_{r_n}(v_n) > 0) [1 - P(\tilde{N}_{r_n}(v_n) > 0) + 2\gamma_n]}{(k_n P(\tilde{N}_{r_n}(v_n) > 0))^2} = o(1),$$

donde

$$\frac{\sum_{i=1}^{k_n} \mathbb{I}_{\{\tilde{N}_{r_n}^{(i)}(u_n) > 0\}}}{k_n P(\tilde{N}_{r_n}(v_n) > 0)} \xrightarrow[n \rightarrow +\infty]{P} 1. \quad (4.7)$$

Quando  $r_n = 1$ ,  $n \geq 1$ , (4.7) implica que  $\frac{\sum_{i=1}^{n-1} \mathbb{I}_{\{X_i \leq v_n < X_{i+1}\}}}{nP(X_1 \leq v_n < X_2)} \xrightarrow[n \rightarrow +\infty]{P} 1$ , que conjuntamente com (4.7) prova o pretendido.  $\square$

## 4.2.2 Normalidade assintótica

A normalidade assintótica de  $\hat{\eta}_n^B(v_n)$  é obtida recorrendo a um resultado de Utev (1990) sobre arranjos triangulares de variáveis aleatórias que passamos a enunciar.

**Lema 4.2.1** *Seja  $Y_{1,n}, \dots, Y_{s_n,n}$ ,  $n \geq 1$ , um arranjo triangular de variáveis aleatórias de média zero e variância finita,*

$$S_n = \sum_{i=1}^{s_n} Y_{i,n}, \quad \sigma_n^2 = E[S_n^2]$$

e

$$\phi_n(k) = \max_{1 \leq m \leq s_n} \sup_{\substack{A \in \mathcal{B}_1^m, P(A) > 0 \\ B \in \mathcal{B}_{m+k}^{+\infty}}} |P(B|A) - P(B)|, \quad k \geq 1,$$

onde  $\mathcal{B}_a^b$  denota a  $\sigma$ -álgebra gerada pelas variáveis aleatórias  $Y_{i,n}$ ,  $a \leq i \leq b$ .

Suponhamos que existe uma sucessão de números naturais  $\{j_n\}_{n \geq 1}$  tais que

$$\sup_n \phi_n(kj_n) \xrightarrow[k \rightarrow +\infty]{} 0$$

e, para todo  $\varepsilon > 0$ ,

$$j_n \sigma_n^{-2} \sum_{i=1}^{s_n} E \left[ Y_{i,n}^2 \mathbb{I}_{\{|Y_{i,n}| \geq \varepsilon \sigma_n / j_n\}} \right] \xrightarrow[n \rightarrow +\infty]{} 0. \quad (4.8)$$

Então

$$S_n / \sigma_n \xrightarrow[n \rightarrow +\infty]{d} \mathcal{N}(0, 1).$$

**Teorema 4.2.3** Se as condições do Teorema 4.2.2 são válidas,  $\sup_n \phi(k_n r_n) \xrightarrow[n \rightarrow +\infty]{} 0$ ,  $r_n^4 = o(nP(X_1 \leq v_n < X_2))$  e

$$\frac{\text{Var} \left[ \sum_{i=1}^{k_n} \mathbb{I}_{\{\tilde{N}_{r_n}^{(i)}(v_n) > 0\}} - \eta_n^B(v_n) \sum_{j=1+(i-1)r_n}^{ir_n} \mathbb{I}_{\{X_j \leq v_n < X_{j+1}\}} \right]}{nP(X_1 \leq v_n < X_2)} \xrightarrow[n \rightarrow +\infty]{} \sigma_B^2,$$

então

$$\sqrt{nP(X_1 \leq v_n < X_2)} (\hat{\eta}_n^B(v_n) - \eta_n^B(v_n)) \xrightarrow[n \rightarrow +\infty]{d} \mathcal{N}(0, \sigma_B^2).$$

**Demonstração:** Consideremos  $Z_i = \mathbb{I}_{\{\tilde{N}_{r_n}^{(i)}(v_n) > 0\}} - \eta_n^B(v_n) \sum_{j=1+(i-1)r_n}^{ir_n} \mathbb{I}_{\{X_j \leq v_n < X_{j+1}\}}$ ,  $i \geq 1$ . Notemos que  $E[Z_i] = 0$  e  $\text{Var}[\sum_{i=1}^{k_n} Z_i] \sim \sigma_B^2 nP(X_1 \leq v_n < X_2)$ .

A condição  $r_n^4 = o(nP(X_1 \leq v_n < X_2))$  implica (4.8).

Logo pelo Lema 4.2.1, obtemos

$$\frac{\sum_{i=1}^{k_n} Z_i}{\sqrt{nP(X_1 \leq v_n < X_2)}} \xrightarrow[n \rightarrow +\infty]{d} N(0, \sigma_B^2).$$

Tendo em conta que

$$\frac{\sqrt{nP(X_1 \leq v_n < X_2)} (\hat{\eta}_n^B(v_n) - \eta_n^B(v_n))}{nP(X_1 \leq v_n < X_2) / \sum_{i=1}^n \mathbb{I}_{\{X_i \leq v_n < X_{i+1}\}}} \sim \frac{\sum_{i=1}^{k_n} Z_i}{\sqrt{nP(X_1 \leq v_n < X_2)}}$$

e que pelo Teorema 4.2.2  $\frac{\sum_{i=1}^n \mathbb{I}_{\{X_i \leq v_n < X_{i+1}\}}}{nP(X_1 \leq v_n < X_2)} \xrightarrow[n \rightarrow +\infty]{P} 1$ , resulta o pretendido.  $\square$

### 4.3 Os estimadores de blocos $\hat{\eta}_n^{dj}$ e $\hat{\eta}_n^{sl}$

Baseando-nos no trabalho de Robert *et al.* (2009) definimos, nesta secção, dois novos estimadores de blocos do índice de cruzamentos,  $\hat{\eta}_n^{dj}$  e  $\hat{\eta}_n^{sl}$ , correspondendo o primeiro a blocos disjuntos e o segundo a blocos deslizantes.

Para uma sucessão de níveis  $\{u_n\}_{n \geq 1}$  e  $r_n > 0$ ,  $n \geq 1$ , consideremos

$$F_{r_n}(u_n) = P(M_{r_n} \leq u_n), \quad \nu_n(u_n) = r_n P(X_1 \leq u_n < X_2) \quad \text{e} \quad \eta_{r_n}(u_n) = -\frac{\log F_{r_n}(u_n)}{\nu_n(u_n)},$$

onde  $M_{0,r_n} = \max\{X_1, \dots, X_{r_n}\}$ .

Com esta notação e da definição de índice de cruzamentos vem que

$$\eta = \lim_{r_n \rightarrow +\infty} \eta_{r_n}(u_n), \quad (4.9)$$

desde que os níveis  $u_n$ ,  $n \geq 1$ , satisfaçam  $\lim_{r_n \rightarrow +\infty} r_n P(X_1 \leq u_n < X_2) = \nu$  e  $\lim_{r_n \rightarrow +\infty} r_n P(X_1 > u_n) = \tau$ , uma vez que nestas condições, existindo  $\eta > 0$  e  $\theta > 0$ , se tem  $\eta\nu = \theta\tau$ .

Tendo em conta a relação (4.9) basta substituímos  $F_{r_n}$  e  $\nu_n$  pelas suas versões empíricas em  $\eta_{r_n}(u_n)$ , para obtemos estimadores para  $\eta$ . A partir dos dois estimadores propostos por Robert *et al.* (2009) para  $F_{r_n}$ , que a seguir apresentamos, iremos obter os nossos dois estimadores para  $\eta$ .

$$\widehat{F}_{n,r_n}^{dj}(u_n) = \frac{1}{k_n} \sum_{i=1}^{k_n} \mathbb{I}_{\{M_{(i-1)r_n, ir_n} \leq u_n\}}$$

e

$$\widehat{F}_{n,r_n}^{sl}(u_n) = \frac{1}{n - k_n + 1} \sum_{i=0}^{n-r_n} \mathbb{I}_{\{M_{i, i+r_n} \leq u_n\}}.$$

O primeiro estimador,  $\widehat{F}_{n,r_n}^{dj}(u_n)$ , corresponde a uma partição da amostra  $X_1, \dots, X_n$  em  $r_n$  blocos disjuntos enquanto que o segundo,  $\widehat{F}_{n,r_n}^{sl}(u_n)$ , corresponde a considerar  $n - r_n + 1$  blocos deslizantes. Quando nos referimos a blocos deslizantes, estamos a supor uma sequência de blocos de tamanho  $r_n$  com início em cada uma das  $n - r_n + 1$  observações iniciais da amostra.

Considerando para  $\nu_n$  o seguinte estimador

$$\widehat{\nu}_{n,r_n}(u_n) = \frac{1}{k_n} \sum_{j=1}^{k_n} \sum_{i=1}^{r_n} \mathbb{I}_{\{X_{r_n(j-1)+i} \leq u_n < X_{r_n(j-1)+i+1}\}} = \frac{1}{k_n} \sum_{i=1}^{k_n r_n} \mathbb{I}_{\{X_i \leq u_n < X_{i+1}\}},$$

obtemos então os seguintes estimadores do índice de cruzamentos,  $\eta$ ,

$$\widehat{\eta}_{n,r_n}^{dj}(u_n) = -\frac{\log \widehat{F}_{n,r_n}^{dj}(u_n)}{\widehat{\nu}_{n,r_n}(u_n)} \quad \text{e} \quad \widehat{\eta}_{n,r_n}^{sl}(u_n) = -\frac{\log \widehat{F}_{n,r_n}^{sl}(u_n)}{\widehat{\nu}_{n,r_n}(u_n)}.$$

Com condições análogas às consideradas por Robert *et al.* (2009) obtêm-se a consistência e a normalidade assintótica dos estimadores  $\widehat{\eta}_{n,r_n}^{dj}(u_n)$  e  $\widehat{\eta}_{n,r_n}^{sl}(u_n)$ . Assim sendo, consideremos os

coeficientes de correlação máxima de  $\mathbf{X}$ ,

$$\rho_{n,l}(u_n) = \max_{t=1,\dots,n-l} \sup_{\substack{\xi \in L_2(\mathcal{B}_t^i(u_n)) \\ \psi \in L_2(\mathcal{B}_{t+l}^n(u_n))}} |\text{corr}(\xi, \psi)|,$$

onde  $\mathcal{B}_a^b(u_n)$  denota a  $\sigma$ -álgebra gerada pelos acontecimentos  $\{X_i \leq u_n\}$ ,  $a \leq i \leq b$ ,  $L_2(\mathcal{B})$  é o espaço das variáveis aleatórias de quadrado integrável e as variáveis aleatórias  $\xi$  e  $\psi$  têm variância positiva.

Tem-se  $\rho_{n,l}(u_n) \geq \alpha_{n,l}$ , com  $\alpha_{n,l}(u)$  os coeficientes da condição  $\Delta(\mathbf{u})$ . Consideremos que as sucessões de inteiros positivos  $\{r_n\}_{n \geq 1}$  e  $\{l_n\}_{n \geq 1}$  verificam

$$l_n = o(r_n), \quad r_n = o(n) \quad \text{e} \quad k_n \rho_{n,l_n}(u_n) \xrightarrow{n \rightarrow +\infty} 0. \quad (4.10)$$

Sob a validade destas condições e tendo em conta que para os níveis  $u_n$ ,  $n \geq 1$ ,  $\eta\nu = \theta\tau$  as seguintes convergências de Robert *et al.* (2009) também são válidas,

$$\begin{aligned} k_n \text{Var}[\widehat{F}_{n,r_n}^{dj}(u_n)] &\xrightarrow{n \rightarrow +\infty} e^{-\alpha}(1 - e^{-\alpha}), \\ k_n \text{Var}[\widehat{F}_{n,r_n}^{sl}(u_n)] &\xrightarrow{n \rightarrow +\infty} 2\alpha^{-1}e^{-\alpha}(1 - (1 + \alpha)e^{-\alpha}), \\ k_n \text{Cov}(\widehat{F}_{n,r_n}^{sl}(u_n), \widehat{F}_{n,r_n}^{dj}(u_n)) &\xrightarrow{n \rightarrow +\infty} 2\alpha^{-1}e^{-\alpha}(1 - (1 + \alpha)e^{-\alpha}), \end{aligned}$$

com  $\alpha = \eta\nu$ .

A covariância assintótica entre  $\widehat{\nu}_{n,r_n}(u_n)$  e os estimadores de blocos disjuntos e deslizantes de  $F_{r_n}(u_n)$  é obtida no próximo resultado de modo semelhante ao obtido em Robert *et al.* (2009) uma vez que apenas a definição de  $\widehat{\nu}_{n,r_n}(u_n)$  é diferente da aí considerada.

**Teorema 4.3.1** *Se forem válidas as condições desta secção e*

$$\sum_{l=1}^n \rho_{n,l}(u_n) = o(r_n^{1/2}), \quad n \rightarrow +\infty, \quad (4.11)$$

então

$$\left. \begin{aligned} k_n \text{Cov}(\widehat{F}_{n,r_n}^{dj}(u_n), \widehat{\nu}_{n,r_n}(u_n)) \\ k_n \text{Cov}(\widehat{F}_{n,r_n}^{sl}(u_n), \widehat{\nu}_{n,r_n}(u_n)) \end{aligned} \right\} \xrightarrow{n \rightarrow +\infty} -\nu e^{-\alpha}. \quad (4.12)$$

**Demonstração:** *Tem-se*

$$k_n \text{Cov}(\widehat{F}_{n,r_n}^{dj}(u_n), \widehat{\nu}_{n,r_n}(u_n)) = \frac{1}{k_n} \sum_{i=1}^k \sum_{j=1}^{k_n r_n} \text{cov}(\mathbb{I}_{\{M_{(i-1)r_n, ir_n} \leq u_n\}}, \mathbb{I}_{\{X_j \leq u_n < X_{j+1}\}}).$$

Decompondo a expressão anterior em duas parcelas  $I_n + II_n$ , em que  $I_n$  corresponde aos índices  $i$  e  $j$  tais que  $(i-1)r_n < j \leq r$  e  $II_n$  aos outros índices, obtemos, pela estacionaridade de  $\mathbf{X}$ ,

que a primeira parcela é dada por

$$I_n = - \sum_{j=1}^{r_n} Cov(\mathbb{I}_{\{M_{0,r_n} \leq u_n\}}, \mathbb{I}_{\{\tilde{N}_{i,j}(u_n)=0\}}) = -r_n P(X_1 \leq u_n < X_2) F_{r_n}(u_n). \quad (4.13)$$

e a segunda parcela é tal que

$$\begin{aligned} |II_n| &\leq 2 \left\{ Var[\mathbb{I}_{\{M_{0,r_n} \leq u_n\}}] Var[\mathbb{I}_{\{\tilde{N}_{1,1}(u_n)=0\}}] \right\}^{1/2} \sum_{l=1}^n \rho_{n,l}(u_n) \\ &\leq 2P(X_1 \leq u_n < X_2)^{1/2} \sum_{l=1}^n \rho_{n,l}(u_n). \end{aligned} \quad (4.14)$$

Temos que  $I_n \xrightarrow{n \rightarrow +\infty} -\nu e^{-\alpha}$  e as hipóteses (4.10) e (4.11) implicam que  $II_n \xrightarrow{n \rightarrow +\infty} 0$ .

Quanto a  $k_n Cov(\hat{F}_{n,r_n}^{sl}(u_n), \hat{\nu}_{n,r_n}(u_n))$  podemos escrever

$$k_n Cov(\hat{F}_{n,r_n}^{sl}(u_n), \hat{\nu}_{n,r_n}(u_n)) = \frac{1}{n - r_n + 1} \sum_{i=0}^{n-r_n} \sum_{j=1}^{k_n r_n} Cov(\mathbb{I}_{\{M_{i,i+r_n} \leq u_n\}}, \mathbb{I}_{\{X_j \leq u_n < X_{j+1}\}})$$

e decompor também em duas parcelas de acordo com o facto de  $i < j \leq i + r$ , ou não. A primeira parcela é igual a  $I_n$  em (4.13) e a segunda é majorada pela expressão em (4.14) e portanto é assintoticamente desprezável.  $\square$

Para conseguirmos obter a variância assintótica de  $\hat{\nu}_{n,r_n}(u_n)$ , temos que impor um conjunto de condições que assegurem que, no limite,  $\tilde{N}_{r_n}(u_n)$  tem distribuição de Poisson composta, ou seja,

$$\tilde{N}_{r_n}(u_n) \xrightarrow{n \rightarrow +\infty} \tilde{N} = \sum_{i=1}^N Z_i,$$

com  $N$  uma variável aleatória de Poisson de valor médio  $\eta\nu$  e  $\{Z_i\}_{i \in \mathbb{N}}$  uma sucessão de variáveis aleatórias independentes e identicamente distribuídas, e independentes de  $N$ , que representam o tamanho dos grupos de cruzamentos. Deste modo, suponhamos, que existe uma sucessão de inteiros positivos  $\{s_n\}_{n \geq 1}$  tal que

$$l_n = o(s_n), \quad s_n = o(r_n), \quad \rho_{n,l_n}(u_n) = o(s_n/n) \quad (4.15)$$

e

$$\begin{aligned} P(\tilde{N}_{s_n}(u_n) = j \mid \tilde{N}_{s_n}(u_n) > 0) &\xrightarrow{n \rightarrow +\infty} \tilde{\pi}(j), \quad j = 1, 2, \dots \\ E \left[ \tilde{N}_{s_n}^2(u_n) \mid \tilde{N}_{s_n}(u_n) > 0 \right] &\xrightarrow{n \rightarrow +\infty} \sum_{j=1}^{+\infty} j^2 \tilde{\pi}(j) < +\infty, \end{aligned} \quad (4.16)$$

onde  $\tilde{\pi}(j)$ ,  $j = 1, 2, \dots$ , é a distribuição limite da distribuição condicional do tamanho de um grupo de cruzamentos definida em (1.19). Relembramos que,  $E[Z_1] = \sum_{j \geq 1} j \tilde{\pi}(j) = \eta^{-1}$ ,

$E[\tilde{N}] = \nu$  e  $var[\tilde{N}] = \alpha \sum_{j \geq 1} j^2 \tilde{\pi}(j)$ , com  $\alpha = \eta\nu$ .

**Teorema 4.3.2** *Seja  $X$  uma sucessão estacionária com índice extremal  $\theta > 0$  e índice de cruzamentos  $\eta > 0$ . Se  $\{u_n\}_{n \geq 1}$  é uma sucessão de níveis tais que  $\lim_{r_n \rightarrow +\infty} r_n P(X_1 > u_n) = \tau > 0$  e  $\lim_{r_n \rightarrow +\infty} r_n P(X_1 \leq u_n < X_2) = \nu > 0$ , e se verificam as condições (4.10), (4.15) e (4.16), então*

$$k_n Var[\hat{\nu}_{n,r_n}(u_n)] \xrightarrow{n \rightarrow +\infty} Var[\tilde{N}] = \alpha \sum_{j \geq 1} j^2 \tilde{\pi}(j). \quad (4.17)$$

**Demonstração:** Notemos que

$$k_n Var[\hat{\nu}_{n,r_n}(u_n)] = \frac{1}{k_n} Var[\tilde{N}_{r_n k_n}(u_n)].$$

Fixando  $1 \leq l_n < s_n < n$  e considerando  $m_n = [r_n k_n / s_n]$  temos que

$$\tilde{N}_{r_n k_n}(u_n) = \underbrace{\sum_{i=1}^{m_n} \tilde{N}_{(i-1)s_n, i s_n - l_n}(u_n)}_{A_n} + \underbrace{\sum_{i=1}^{m_n} \tilde{N}_{i s_n - l_n, i s_n}(u_n)}_{B_n} + \underbrace{\tilde{N}_{m_n s_n, r_n k_n}(u_n)}_{C_n}.$$

Com argumentos semelhantes aos usados por Robert *et al.* (2009) prova-se que

$$(1/k_n) Var[A_n] \xrightarrow{n \rightarrow +\infty} \alpha \sum_{j \geq 1} j^2 \tilde{\pi}(j), \quad (1/k_n) Var[B_n] \xrightarrow{n \rightarrow +\infty} 0, \quad (1/k_n) Var[C_n] \xrightarrow{n \rightarrow +\infty} 0,$$

o que conjuntamente com a desigualdade de Cauchy-Schwarz demonstra o resultado.  $\square$

### 4.3.1 Consistência e normalidade assintótica

Iniciamos esta secção com um resultado que prova a consistência fraca dos estimadores de blocos disjuntos e blocos deslizantes.

**Teorema 4.3.3** *Seja  $X$  uma sucessão estacionária com índice extremal  $\theta > 0$ , índice de cruzamentos  $\eta > 0$  e suponhamos que existe uma sucessão de níveis  $\{u_n\}_{n \geq 1}$  tais que,  $\lim_{r_n \rightarrow +\infty} r_n P(X_1 > u_n) = \tau > 0$  e  $\lim_{r_n \rightarrow +\infty} r_n P(X_1 \leq u_n < X_2) = \nu > 0$ . Se as condições (4.10), (4.15) e (4.16) se verificarem, então*

$$\hat{\eta}_{n,r_n}^{dj}(u_n) \xrightarrow[n \rightarrow +\infty]{P} \eta \quad \text{e} \quad \hat{\eta}_{n,r_n}^{sl}(u_n) \xrightarrow[n \rightarrow +\infty]{P} \eta.$$

**Demonstração:** Dos resultados anteriores e da desigualdade de Chebychev, temos que

$$|\hat{F}_{n,r_n}^{dj}(u_n) - F_{n,r_n}(u_n)| + |\hat{F}_{n,r_n}^{sl}(u_n) - F_{n,r_n}(u_n)| + |\hat{\nu}_{n,r_n}(u_n) - \nu_{n,r_n}(u_n)| \xrightarrow[n \rightarrow +\infty]{P} 0.$$

Pela definição de índice extremal  $F_{n,r_n}(u_n) \xrightarrow[n \rightarrow +\infty]{} e^{-\theta\tau}$ ,  $\theta\tau = \eta\nu$ , e  $\nu_{n,r_n}(u_n) \xrightarrow[n \rightarrow +\infty]{} \nu$ . O resultado é agora consequência da continuidade da função  $(x, y) \mapsto -\log(x)/y$  em  $(0, +\infty) \times (0, +\infty)$ .  $\square$

A normalidade assintótica dos estimadores  $\hat{F}_{n,r_n}^{dj}(u_n)$ ,  $\hat{F}_{n,r_n}^{sl}(u_n)$  e  $\hat{\nu}_{n,r_n}(u_n)$  é obtida no próximo resultado.

**Teorema 4.3.4** *Seja  $X$  uma sucessão estacionária com índice extremal  $\theta > 0$ , índice de cruzamentos  $\eta > 0$  e suponhamos que existe uma sucessão de níveis  $\{u_n\}_{n \geq 1}$  tais que,  $\lim_{r_n \rightarrow +\infty} r_n P(X_1 > u_n) = \tau > 0$  e  $\lim_{r_n \rightarrow +\infty} r_n P(X_1 \leq u_n < X_2) = \nu > 0$ . Se as condições (4.10), (4.11), (4.15), (4.16) se verificarem e se existir uma constante  $p > 1$  tal que  $E[\tilde{N}_{r_n}^{2p}(u_n)] = O(1)$ , quando  $n \rightarrow +\infty$ , então*

$$\sqrt{k_n} \begin{pmatrix} \hat{F}_{n,r_n}^{dj}(u_n) - F_{n,r_n}(u_n) \\ \hat{F}_{n,r_n}^{sl}(u_n) - F_{n,r_n}(u_n) \\ \hat{\nu}_{n,r_n}(u_n) - \nu_{n,r_n}(u_n) \end{pmatrix} \xrightarrow[n \rightarrow +\infty]{d} N(0, \Sigma)$$

onde  $\Sigma = [\sigma_{i,j}]_{3 \times 3}$  é uma matriz simétrica com

$$\begin{aligned} \sigma_{11} &= e^{-\alpha}(1 - e^{-\alpha}), \\ \sigma_{22} = \sigma_{12} &= 2\alpha^{-1}e^{-\alpha}(1 - (1 + \alpha)e^{-\alpha}), \\ \sigma_{31} = \sigma_{32} &= -\nu e^{-\alpha}, \\ \sigma_{33} &= \alpha \sum_{j \geq 1} j^2 \tilde{\pi}(j), \end{aligned}$$

e  $\alpha = \eta\nu$ .

**Demonstração:** Basta considerar  $Z_{n,3} = \sqrt{k_n}(\hat{\nu}_{n,r_n}(u_n) - \nu_{n,r_n}(u_n))$  e substituir  $\mathbb{I}_{\{X_t > u_n\}}$  por  $\mathbb{I}_{\{X_t \leq u_n < X_{t+1}\}}$  na demonstração Teorema 4.2 de Robert *et al.* (2009).  $\square$

Deste resultado e da aplicação do método delta obtemos agora a distribuição assintótica dos estimadores  $\hat{\eta}_{n,r_n}^{dj}(u_n)$  e  $\hat{\eta}_{n,r_n}^{sl}(u_n)$ .

**Corolário 4.3.1** *Sob as condições do Teorema 4.3.4 e se*

$$\sqrt{k_n}(\eta_{r_n} - \eta) \xrightarrow[n \rightarrow +\infty]{} 0,$$

então

$$\sqrt{k_n} \begin{pmatrix} \hat{\eta}_{n,r_n}^{dj}(u_n) - \eta \\ \hat{\eta}_{n,r_n}^{sl}(u_n) - \eta \end{pmatrix} \xrightarrow[n \rightarrow +\infty]{d} N(0, \mathbf{V})$$

onde  $\mathbf{V} = [v_{i,j}]_{2 \times 2}$  é uma matriz simétrica com

$$\begin{aligned} v_{11} &= \frac{\eta^2}{\alpha^2}(e^\alpha - 1 - \alpha) + \frac{\eta^2 c^2}{\alpha}, \\ v_{22} = v_{12} &= 2\frac{\eta^2}{\alpha^3}(e^\alpha - 1 - \alpha - \frac{\alpha^2}{2}) + \frac{\eta^2 c^2}{\alpha}, \end{aligned}$$

em que  $\alpha = \eta\nu$  e  $c^2 = \frac{\sum_{j \geq 1} j^2 \tilde{\pi}(j) - \eta^{-2}}{\eta^{-2}}$  representa o quadrado do coeficiente de variação da distribuição condicional do tamanho de um grupo de cruzamentos  $\tilde{\pi}(j)$ .

## 4.4 Estudo de simulação

Comparamos o desempenho dos estimadores do índice de cruzamentos, propostos anteriormente, nos modelos ARMAX e AR(1), já apresentados no capítulo anterior. Começamos por analisar o efeito que o número de blocos considerados,  $k_n$ , ou o tamanho desses blocos,  $r_n$  tem sobre as trajetórias das estimativas de  $\eta$  obtidas com estes estimadores. Seguidamente analisamos o comportamento do erro quadrático médio, da variância e do viés dos estimadores e por último comparamos as estimativas obtidas com estes estimadores com as obtidas com o estimador de *runs* definido no capítulo anterior.

Nas simulações iremos considerar, como anteriormente, níveis determinísticos  $u_n = X_{n-k:n}$ ,  $1 \leq k \leq n-1$ , e portanto os estimadores serão agora funções de  $k$ , *i.e.*,  $\hat{\eta}_n^B(k)$ ,  $\hat{\eta}_n^{dj}(k)$ ,  $\hat{\eta}_n^{sl}(k)$  e  $\hat{\eta}_n^R(k)$ , onde  $\hat{\eta}_n^R(k)$  denota o estimador dos *runs*.

### 4.4.1 O número de blocos

Para tentarmos perceber o efeito que o valor de  $k_n$  (ou  $r_n$ ) exerce sobre as trajetórias das estimativas de  $\eta$  obtidas com os estimadores de blocos propostos anteriormente, simulámos amostras dos processos ARMAX e AR(1) definidos nos capítulos anteriores. Atendendo a que a validade dos resultados teóricos é obtida recorrendo argumentos de natureza assintótica, considerámos amostras de dimensão  $n = 50000$ . Não sendo, do ponto de vista prático, exequível a apresentação de gráficos para todos os valores possíveis de  $k_n$ , a estratégia seguida passou por considerar primeiro valores pequenos de  $r_n$  (valores elevados de  $k_n$ ) e posteriormente ir considerando valores moderados até atingirmos valores elevados de  $r_n$  (valores pequenos de  $k_n$ ). Nas Figuras 4.1-4.4 encontramos o resultado destas simulações, onde relembremos que as linhas horizontais assinaladas correspondem aos verdadeiros valores do índice de cruzamentos.

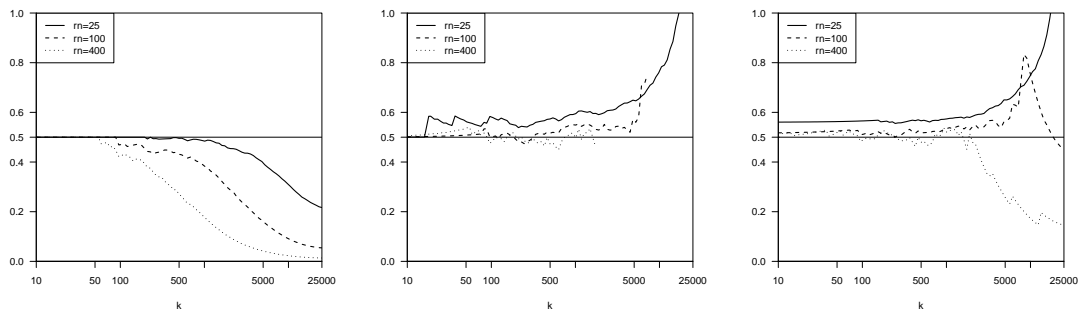


Figura 4.1: Estimativas de  $\eta$  obtidas com  $\hat{\eta}_n^B(k)$  (à esquerda),  $\hat{\eta}_n^{dj}(k)$  (ao centro) e  $\hat{\eta}_n^{sl}(k)$  (à direita) para o processo ARMAX, para vários valores de  $r_n$  e para  $n = 50000$ , numa escala logarítmica ( $\eta = 0.5$ )

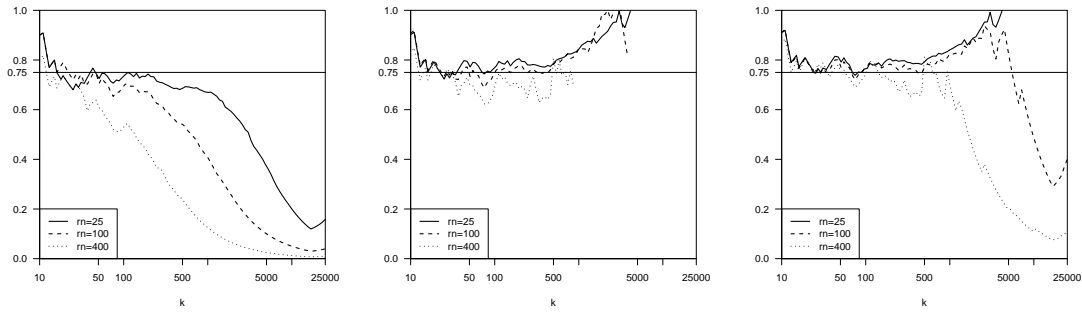


Figura 4.2: Estimativas de  $\eta$  obtidas com  $\hat{\eta}_m^B(k)$  (à esquerda),  $\hat{\eta}_n^{dj}(k)$  (ao centro) e  $\hat{\eta}_n^{sl}(k)$  (à direita) para o processo AR(1), para vários valores de  $r_n$  e com  $n = 50000$  e  $r = 2$ , numa escala logarítmica ( $\eta = 0.75$ )

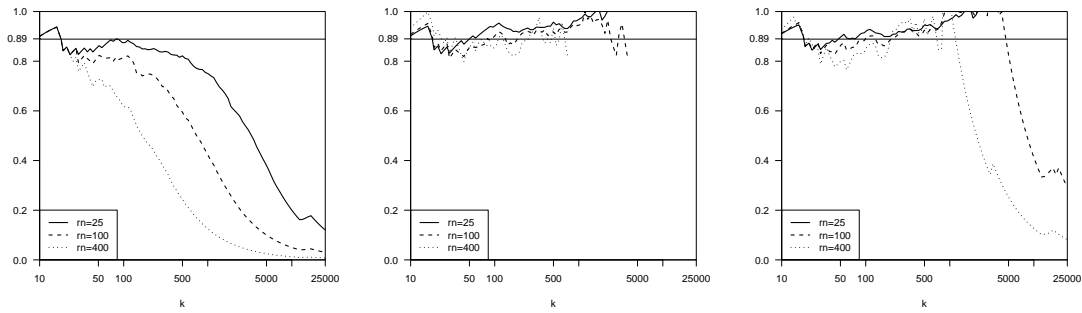


Figura 4.3: Estimativas de  $\eta$  obtidas com  $\hat{\eta}_m^B(k)$  (à esquerda),  $\hat{\eta}_n^{dj}(k)$  (ao centro) e  $\hat{\eta}_n^{sl}(k)$  (à direita) para o processo AR(1), para vários valores de  $r_n$  e com  $n = 50000$  e  $r = 3$ , numa escala logarítmica ( $\eta = 0.89$ )

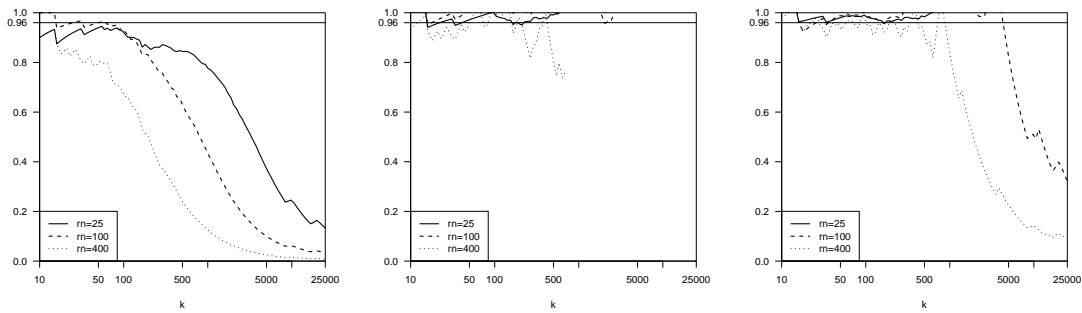


Figura 4.4: Estimativas de  $\eta$  obtidas com  $\hat{\eta}_m^B(k)$  (à esquerda),  $\hat{\eta}_n^{dj}(k)$  (ao centro) e  $\hat{\eta}_n^{sl}(k)$  (à direita) para o processo AR(1), para vários valores de  $r_n$  e com  $n = 50000$  e  $r = 5$ , numa escala logarítmica ( $\eta = 0.96$ )

A análise das Figuras 4.1-4.4 vem confirmar algumas das considerações teóricas feitas anteriormente. Assim o número de blocos deverá ser grande,  $k_n$  deverá tender para infinito, mas não demasiado grande pois é necessário que  $n/k_n$  tenda para infinito. Para os três estimadores identificamos zonas de estabilidade das estimativas junto do verdadeiro valor de  $\eta$ . Estas zonas parecem ser mais evidentes para o estimador  $\hat{\eta}_n^{sl}(k)$  de blocos deslizantes.

Quando consideramos amostras de menor dimensão, por exemplo  $n = 500$ , é natural que

as zonas de estabilidade das estimativas sejam mais difíceis de identificar. Nestes casos, para tentarmos perceber o efeito da escolha da dimensão dos blocos  $r_n$ , realizamos uma simulação multi-amostra com 5000 corridas e 10 réplicas, para  $n = 500$ , e comparamos o comportamento do erro quadrático médio estimado dos estimadores. Nas Figuras 4.5-4.8 apresentamos os resultados obtidos para os dois modelos em análise.

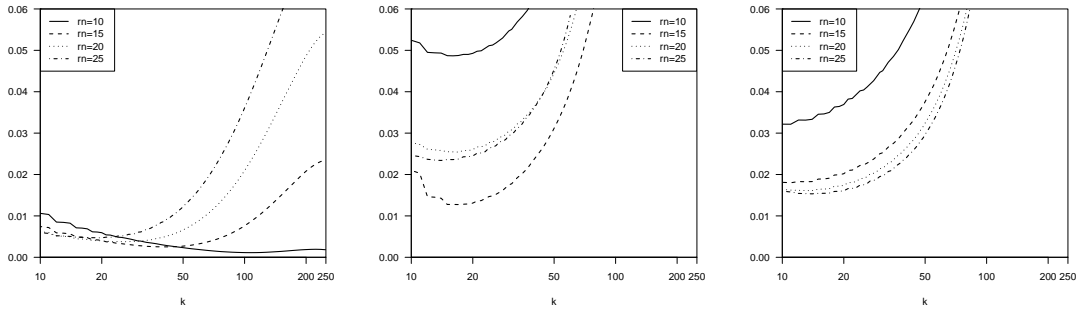


Figura 4.5: Erro quadrático médio estimado de  $\hat{\eta}_n^B(k)$  (à esquerda),  $\hat{\eta}_n^{dj}(k)$  (ao centro) e  $\hat{\eta}_n^{sl}(k)$  (à direita) para o processo ARMAX, para vários valores de  $r_n$  e com  $n = 500$ , numa escala logarítmica ( $\eta = 0.5$ )

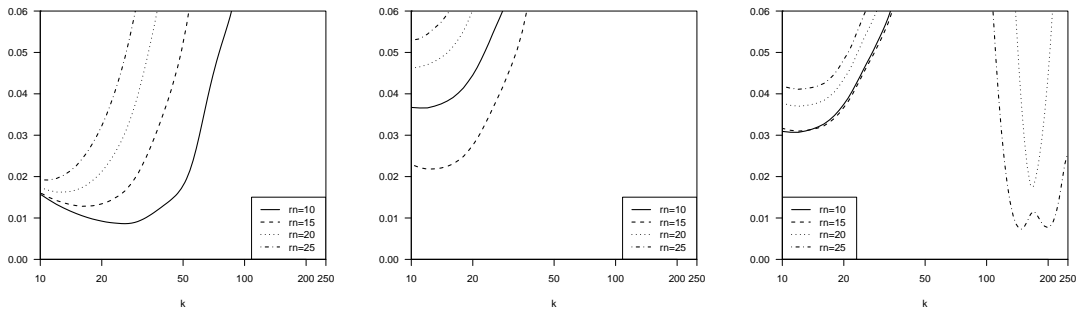


Figura 4.6: Erro quadrático médio estimado de  $\hat{\eta}_n^B(k)$  (à esquerda),  $\hat{\eta}_n^{dj}(k)$  (ao centro) e  $\hat{\eta}_n^{sl}(k)$  (à direita) para o processo AR(1), para vários valores de  $r_n$  e com  $n = 500$  e  $r = 2$ , numa escala logarítmica ( $\eta = 0.75$ )

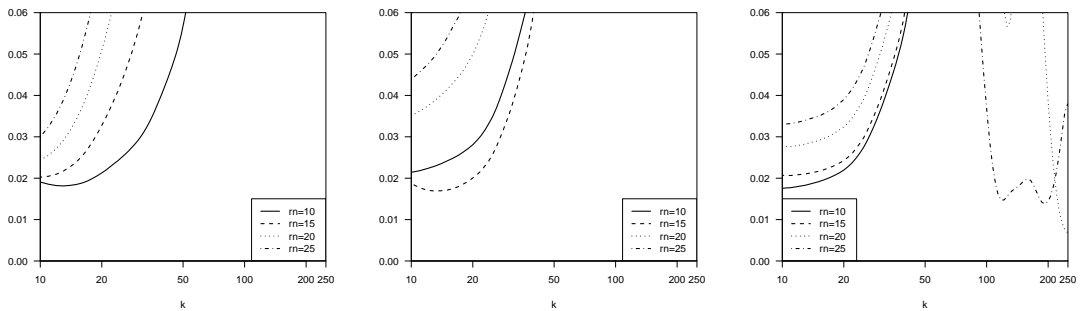


Figura 4.7: Erro quadrático médios estimado de  $\hat{\eta}_n^B(k)$  (à esquerda),  $\hat{\eta}_n^{dj}(k)$  (ao centro) e  $\hat{\eta}_n^{sl}(k)$  (à direita) para o processo AR(1), para vários valores de  $r_n$  e com  $n = 500$  e  $r = 3$ , numa escala logarítmica ( $\eta = 0.89$ )

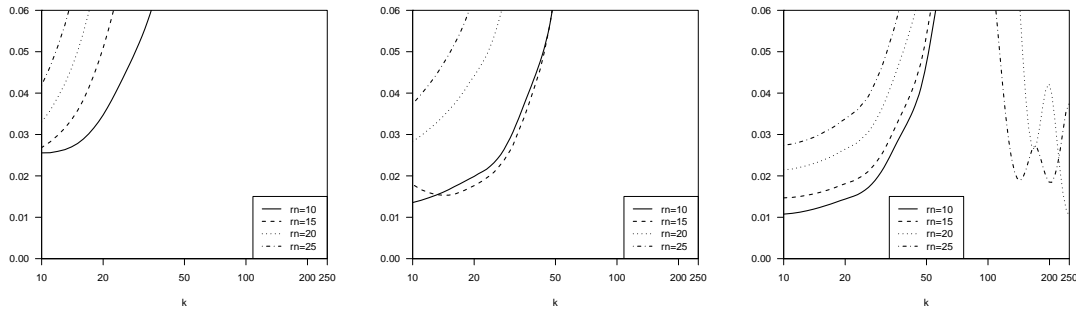


Figura 4.8: Erro quadrático médio estimado de  $\hat{\eta}_n^B(k)$  (à esquerda),  $\hat{\eta}_n^{dj}(k)$  (ao centro) e  $\hat{\eta}_n^{sl}(k)$  (à direita) para o processo AR(1), para vários valores de  $r_n$  e com  $n = 500$  e  $r = 5$ , numa escala logarítmica ( $\eta = 0.96$ )

As Figuras 4.5-4.8 mostram que para  $r_n = 15$  os resultados são consideravelmente bons, no sentido em que este valor ou produz os menores erros quadráticos médios para valores de  $k$  pequenos, ou se aproxima bastante dos melhores que conseguimos obter com outros valores de  $r_n$ . Assim, no estudo comparativo dos estimadores do índice de cruzamentos que apresentaremos a seguir, será usado como referência o valor de  $r_n = 15$ .

#### 4.4.2 Decomposição do erro quadrático médio

Tendo em consideração a relação entre o erro quadrático médio do estimador, a respetiva variância e o viés, podemos perceber a forma como estas duas últimas quantidades dominam o erro quadrático médio dos modelos em estudo. Nas estimativas calculadas a partir dos estimadores dos blocos considerámos  $r_n = 15$ , tendo em conta as justificações apresentadas anteriormente.

Consideramos ainda o seguinte estimador de  $\eta$

$$\hat{\eta}_n^{FS}(u_n) = \hat{\theta}_n^{FS}(u_n) \frac{\sum_{i=1}^{n-1} \mathbb{I}_{\{X_i \leq u_n < X_{i+1}\}}}{\sum_{i=1}^n \mathbb{I}_{\{X_i > u_n\}}},$$

sugerido pela relação entre o índice extremal e o índice de cruzamentos (1.26), onde  $\hat{\theta}_n^{FS}(u_n)$  é o estimador do índice extremal de Ferro e Segers (2003), apresentado na Secção 1.3.1.2.3 do Capítulo 1.

Atendendo a que pretendemos estimadores de  $\eta$  para  $0 < \eta \leq 1$ , não fará sentido usar estimadores do índice extremal definidos sob a validade da condição  $D''(\mathbf{u})$ , uma vez que nestes casos se tem  $\eta = 1$ .

Para processos  $m$ -dependentes o estimador  $\hat{\eta}_n^{FS}(u_n)$  é consistente, pois Ferro e Segers (2003) provaram que para estes processos  $\hat{\theta}_n^{FS}(u_n)$  é consistente e as outras estatísticas que aparecem em  $\hat{\eta}_n^{FS}(u_n)$  são estimadores consistentes de  $\tau$  e de  $\nu$ .

A Figura 4.9 contém os resultados para o modelo ARMAX, com  $n = 500$ .

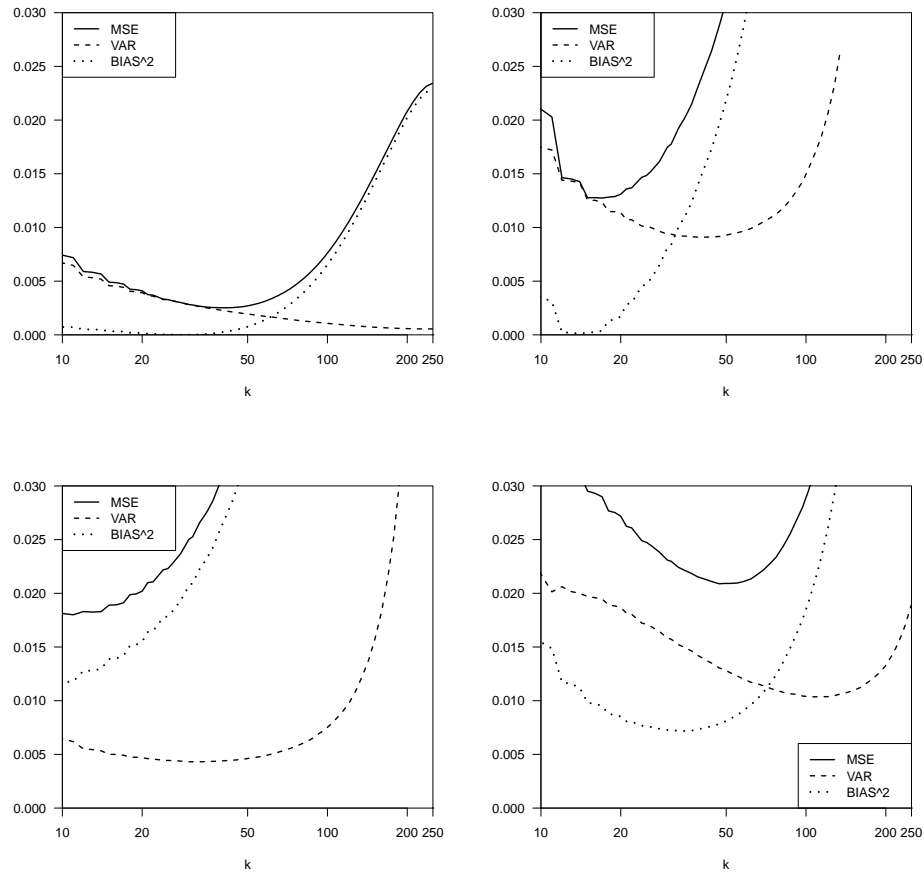


Figura 4.9: Erro quadrático médio estimado, variância estimada e quadrado do viés estimado de  $\hat{\eta}_n^B(k)$  (cima à esquerda),  $\hat{\eta}_n^{dj}(k)$  (cima à direita),  $\hat{\eta}_n^{sl}(k)$  (baixo à esquerda) e  $\hat{\eta}_n^{FS}(k)$  (baixo à direita) para o processo ARMAX, com  $n = 500$ , numa escala logarítmica ( $\eta = 0.5$ )

Para o modelo AR(1) apresentamos os resultados obtidos nas Figuras 4.10-4.12.

A partir das Figuras 4.9-4.12 destacamos o facto do estimador  $\hat{\eta}_n^B(k)$  apresentar, em geral, valores para a variâncias inferiores aos dos outros estimadores analisados. No entanto, à medida que o verdadeiro valor de  $\eta$  se aproxima de 1, como no caso do modelo AR(1) com  $r = 3$  e  $r = 5$ , este estimador tende a apresentar um viés mais acentuado, fazendo com que o seu erro quadrático médio seja em alguns casos superior aos erros quadráticos médios dos restantes estimadores. Em relação aos estimadores  $\hat{\eta}_n^{dj}(k)$  e  $\hat{\eta}_n^{sl}(k)$ , concluímos, como Robert *et al.* (2009), que  $\hat{\eta}_n^{sl}(k)$  apresenta, para os modelos analisados, uma variância inferior à de  $\hat{\eta}_n^{dj}(k)$ . Podemos ainda salientar que o estimador  $\hat{\eta}_n^{FS}(k)$  tem um comportamento relativamente bom no caso AR(1) com  $r = 5$ , em que a respetiva variância parece ser semelhante à variância obtida a partir de  $\hat{\eta}_n^B(k)$ . Neste caso os erros quadráticos médios distanciam-se pelo facto de, como já foi referido, o estimador  $\hat{\eta}_n^B(k)$  apresentar um viés considerável. Na generalidade, verificamos que os resultados obtidos por  $\hat{\eta}_n^{FS}(k)$  são melhores para os modelos com  $\eta$  mais próximo de 1. O mau comportamento deste estimador nas restantes situações poderá justificar-se pelo facto do estimador  $\hat{\eta}_n^{FS}(u_n)$  não ser consistente para modelos que não são  $m$ - dependentes, como é

o caso do processo AR(1).

Para valores pequenos de  $k$ , o erro quadrático médio é fortemente dominado pela variância, alterando-se o cenário a partir de certos valores de  $k$ .

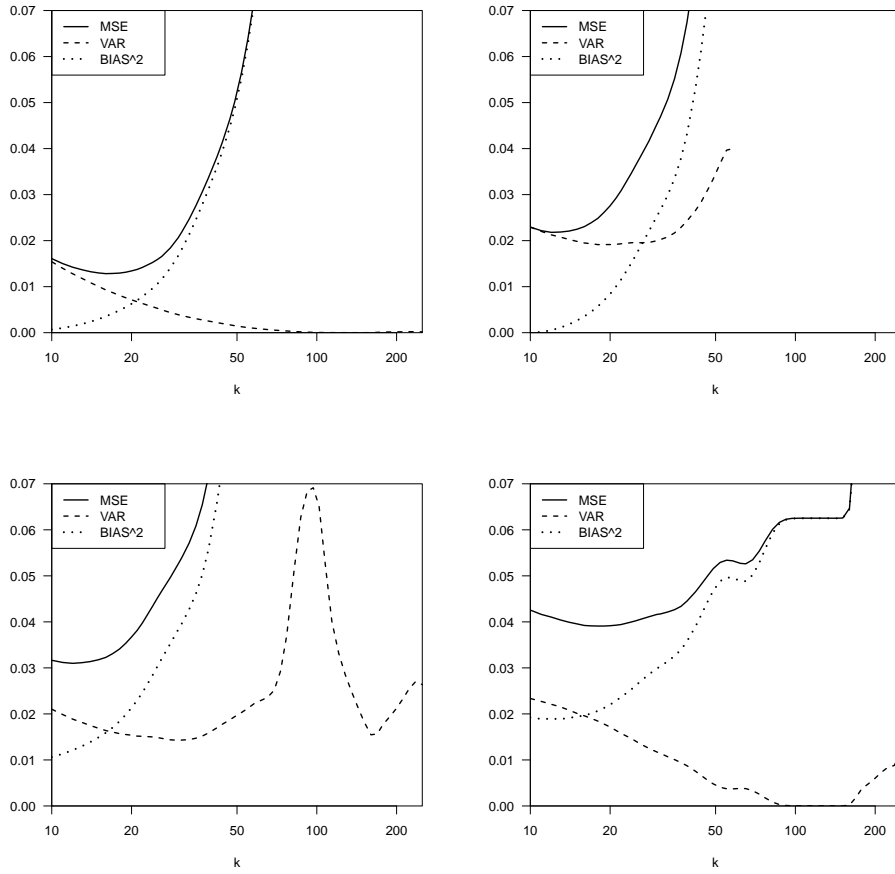


Figura 4.10: Erro quadrático médio estimado, variância estimada e quadrado do viés estimado de  $\hat{\eta}_n^B(k)$  (cima à esquerda),  $\hat{\eta}_n^{dj}(k)$  (cima à direita),  $\hat{\eta}_n^{sl}(k)$  (baixo à esquerda) e  $\hat{\eta}_n^{FS}(k)$  (baixo à direita) para o processo AR(1), com  $n = 500$  e  $r = 2$ , numa escala logarítmica ( $\eta = 0.75$ )

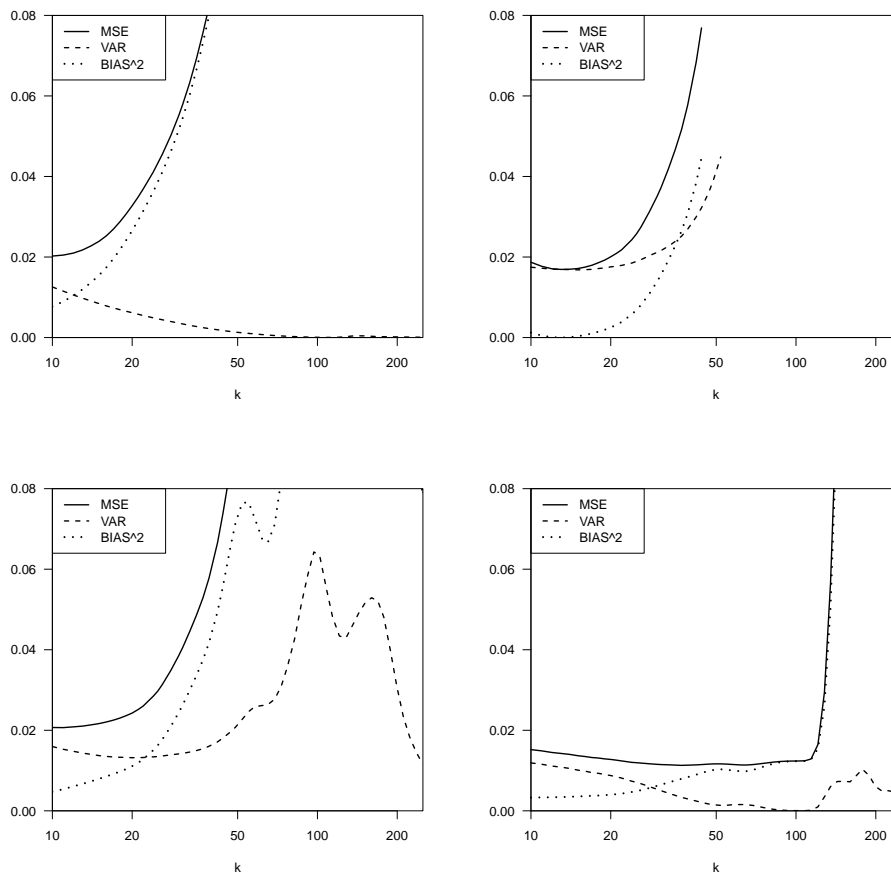


Figura 4.11: Erro quadrático médio estimado, variância estimada e quadrado do viés estimado de  $\hat{\eta}_n^B(k)$  (cima à esquerda),  $\hat{\eta}_n^{dj}(k)$  (cima à direita),  $\hat{\eta}_n^{sl}(k)$  (baixo à esquerda) e  $\hat{\eta}_n^{FS}(k)$  (baixo à direita) para o processo AR(1), com  $n = 500$  e  $r = 3$ , numa escala logarítmica ( $\eta = 0.89$ )

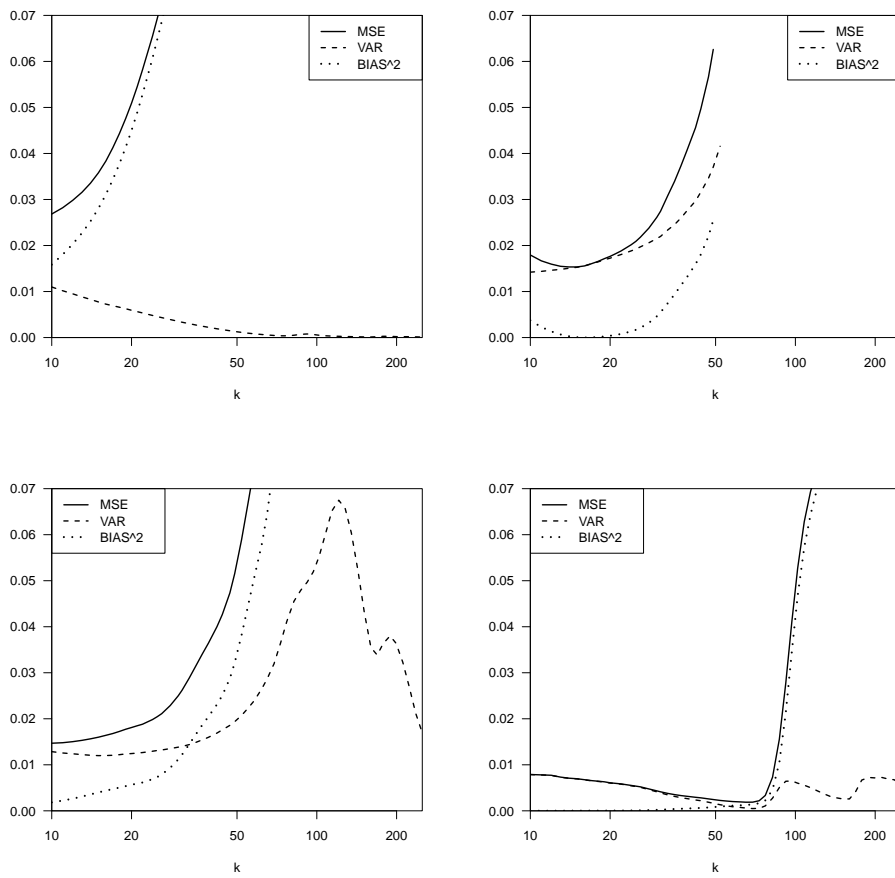


Figura 4.12: Erro quadrático médio estimado, variância estimada e quadrado do viés estimado de  $\hat{\eta}_n^B(k)$  (cima à esquerda),  $\hat{\eta}_n^{dj}(k)$  (cima à direita),  $\hat{\eta}_n^{sl}(k)$  (baixo à esquerda) e  $\hat{\eta}_n^{FS}(k)$  (baixo à direita) para o processo AR(1), com  $n = 500$  e  $r = 5$ , numa escala logarítmica ( $\eta = 0.96$ )

### 4.4.3 Comparações com o estimador de *runs*

Procedemos agora à comparação do comportamento dos estimadores propostos neste capítulo e o estimador de *runs* do Capítulo 3.

Podemos verificar nos gráficos das Figuras 4.13 a 4.20 que nas situações analisadas, as estimativas do valor esperado de  $\hat{\eta}_n^R(k)$  apresentam valores próximos do verdadeiro valor de  $\eta$ , ou seja, facilmente identificamos zonas de estabilidade com viés reduzido. Para os outros estimadores em estudo, à exceção de  $\hat{\eta}_n^{FS}(k)$  para o AR(1) com  $r = 5$ , também encontramos zonas que apresentam alguma estabilidade, no entanto, o viés parece-nos muito elevado para que sejam consideradas credíveis.

Salientamos a existência de estimadores que não estão definidos para todos os valores de  $k$ . Por exemplo, o estimador  $\hat{\eta}_n^{dj}(k)$ , pelo facto de ser calculado a partir do número de blocos em que não se verificam excedências, quando o valor de  $k$  cresce (níveis mais baixos), atingem-se situações em que todos os blocos apresentam excedências, o que equivale a calcular o logaritmo de zero, fazendo com que nestas situações não se possa determinar o seu valor.

Também nos parece importante mencionar que a variância estimada de  $\hat{\eta}_n^B(k)$  e  $\hat{\eta}_n^R(k)$  é em geral inferior à variância estimada dos outros estimadores. Voltamos a referir que no caso de  $\hat{\eta}_n^{FS}(k)$  para o processo AR(1) com  $r = 5$ , os resultados obtidos por este estimador em termos de variância são comparáveis aos obtidos por  $\hat{\eta}_n^B(k)$ . No entanto é a partir do estimador de *runs* que se obtêm, quase sempre, os melhores resultados em termos de menor variância comparativamente aos outros estimadores, sobretudo para valores pequenos de  $k$ . Notemos ainda que o estimador  $\hat{\eta}_n^{dj}(k)$  apresenta sempre variância superior à de  $\hat{\eta}_n^{sl}(k)$ .

Analisarmos de seguida o comportamento dos estimadores nos seus níveis ótimos ( $k_0$ ), ou seja, nos níveis que minimizam o erro quadrático médio, através da medida de eficiência relativa apresentada em Gomes *et al.* (2008) e definida por,

$$REFF_0^{I/J} = \sqrt{\frac{MSE[\hat{\eta}_n^I(k_0)]}{MSE[\hat{\eta}_n^J(k_0)]}} \quad (4.18)$$

onde os índices  $I$  e  $J$  poderão ser iguais a  $R$ ,  $B$ ,  $dj$ ,  $sl$  ou  $FS$  para indicarem, respetivamente, os estimadores  $\hat{\eta}_n^R(k)$ ,  $\hat{\eta}_n^B(k)$ ,  $\hat{\eta}_n^{dj}(k)$ ,  $\hat{\eta}_n^{sl}(k)$  e  $\hat{\eta}_n^{FS}(k)$ . Notemos que no caso de os estimadores terem a mesma eficiência, deveremos obter ( $REFF_0^{I/J} = 1$ ). Valores de ( $REFF_0^{I/J} < 1$ ), significam que o estimador assinalado pelo índice  $I$  é mais eficiente do que o estimador assinalado pelo índice  $J$ . Se pelo contrário, tivermos ( $REFF_0^{I/J} > 1$ ), será o estimador assinalado por  $J$  mais eficiente que o assinalado por  $I$ .

Nas Tabelas 4.1 e 4.2 encontramos os resultados referentes à eficiência relativa dos estimadores em análise. Apenas para o processo AR(1), com  $r = 5$ , o estimador  $\hat{\eta}_n^{FS}(k)$  é mais eficiente que os restantes, nos restantes casos o estimador de *runs* é sempre o mais eficiente. Destacamos ainda a eficiência superior de  $\hat{\eta}_n^{FS}(k)$  comparativamente a  $\hat{\eta}_n^B(k)$ ,  $\hat{\eta}_n^{dj}(k)$  e  $\hat{\eta}_n^{sl}(k)$  no processo AR(1) com  $r = 3$  e  $r = 5$ .

Nas Tabelas 4.3 e 4.4 apresentamos os valores estimados de um indicador de redução do viés definido por (ver Gomes *et al.* (2008)),

$$BRI_0^{I/J} = \left| \frac{BIAS[\hat{\eta}_n^I(k_0)]}{BIAS[\hat{\eta}_n^J(k_0)]} \right|,$$

onde os índices  $I$  e  $J$  têm o mesmo significado que em (4.18). Se  $BRI_0^{I/J} > 1$ , o estimador assinalado com o índice  $I$  é mais enviesado que o estimador assinalado com  $J$ , sendo que o contrário se verifica se  $BRI_0^{I/J} < 1$ . Nos casos em que  $BRI_0^{I/J} = 1$  os dois estimadores terão o mesmo viés.

Mais uma vez os melhores resultados, são obtidos para o estimador de *runs*. Verificamos ainda que  $\hat{\eta}_n^{sl}(k)$  apresenta sempre um viés superior a  $\hat{\eta}_n^{dj}(k)$ . Notemos também que  $\hat{\eta}_n^{FS}(k)$  apresenta os piores resultados na maioria das situações analisadas.

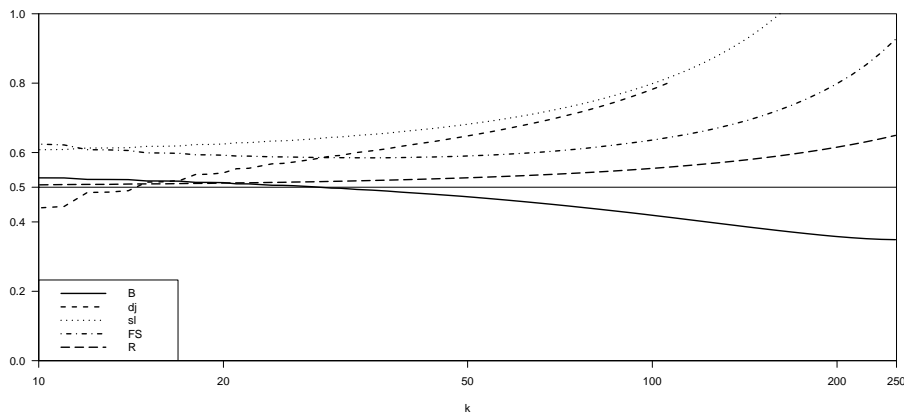


Figura 4.13: Valores médios estimados dos estimadores em estudo para o processo ARMAX, com  $n = 500$ , numa escala logarítmica. ( $\eta = 0.5$ )

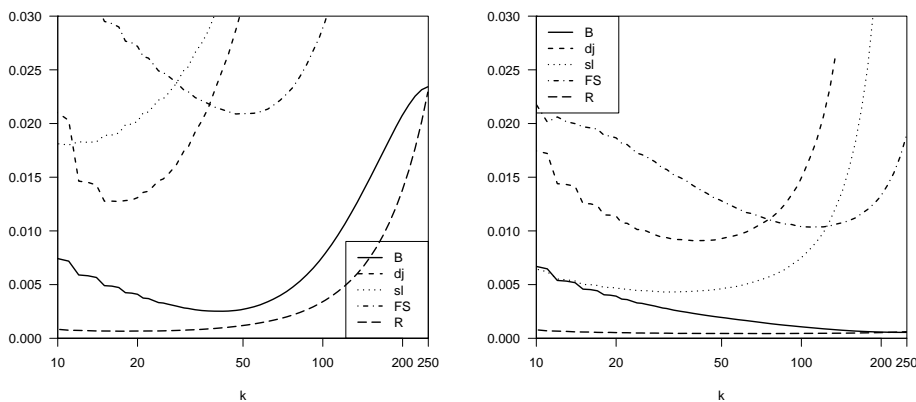


Figura 4.14: Comparação dos erros quadráticos médios estimados (à esquerda) e das variâncias estimadas (à direita) dos estimadores em estudo para o processo ARMAX com  $n = 500$ , numa escala logarítmica.

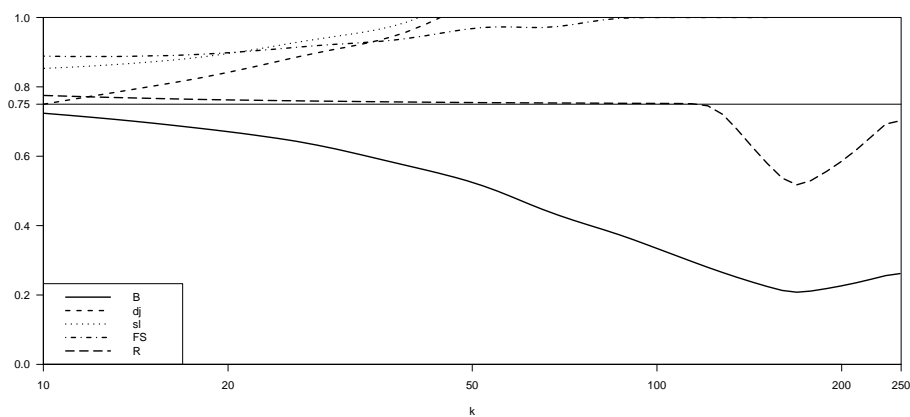


Figura 4.15: Valores médios estimados dos estimadores em estudo para o processo AR(1), com  $n = 500$  e  $r = 2$ , numa escala logarítmica. ( $\eta = 0.75$ )

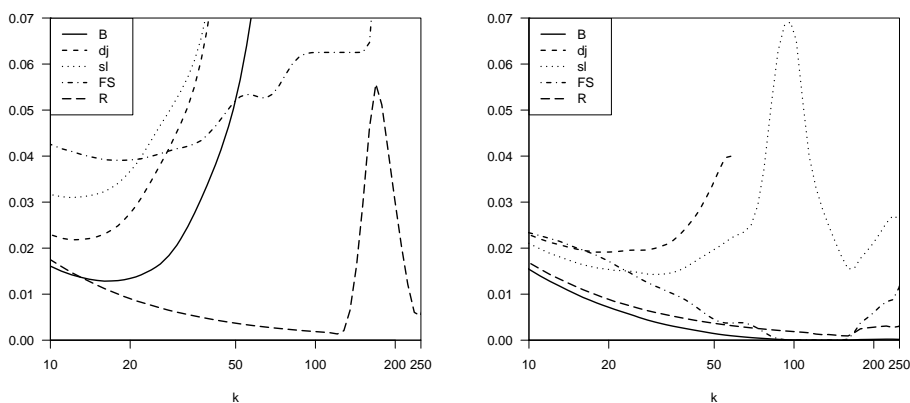


Figura 4.16: Comparação dos erros quadráticos médios estimados (à esquerda) e das variâncias estimadas (à direita) dos estimadores em estudo para o processo AR(1), com  $n = 500$  e  $r = 2$ , numa escala logarítmica.

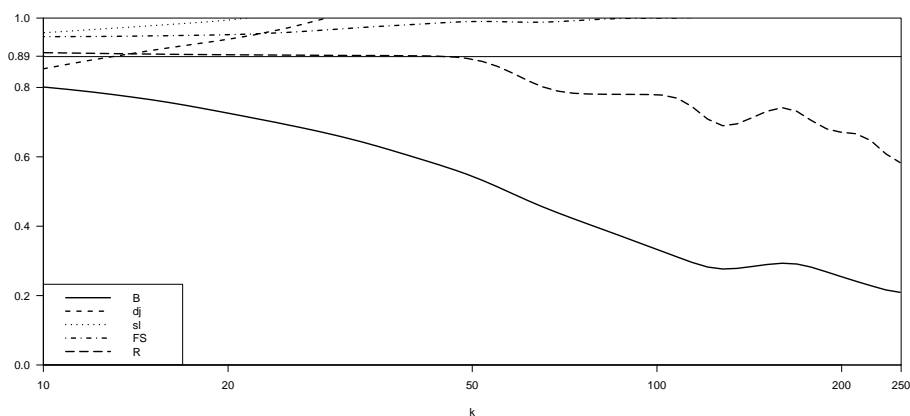


Figura 4.17: Valores médios estimados dos estimadores em estudo para o processo AR(1), com  $n = 500$  e  $r = 3$ , numa escala logarítmica. ( $\eta = 0.89$ )

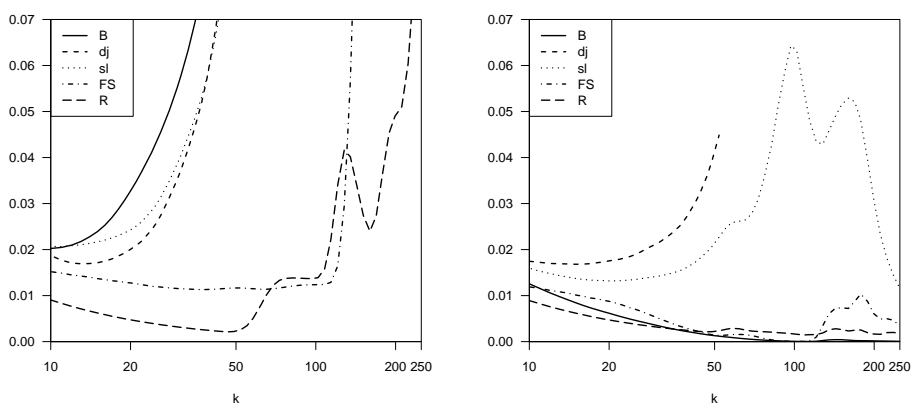


Figura 4.18: Comparação dos erros quadráticos médios estimados (à esquerda) e das variâncias estimadas (à direita) dos estimadores em estudo para o processo AR(1), com  $n = 500$  e  $r = 3$ , numa escala logarítmica.

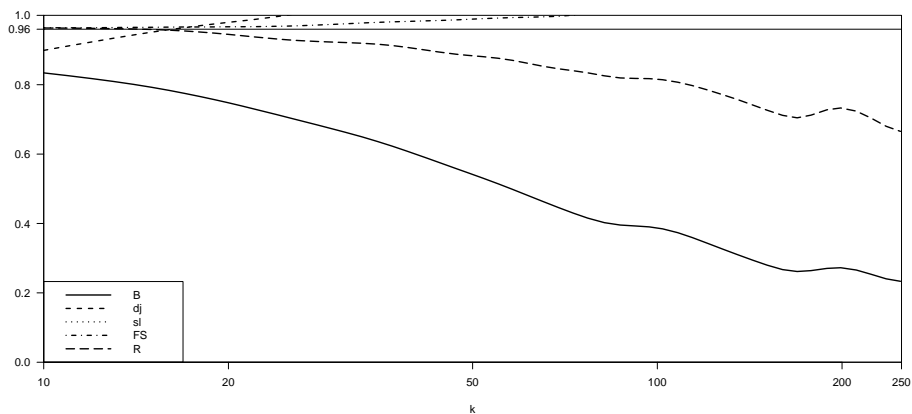


Figura 4.19: Valores médios estimados dos estimadores em estudo para o processo AR(1), com  $n = 500$  e  $r = 5$ , numa escala logarítmica. ( $\eta = 0.96$ )

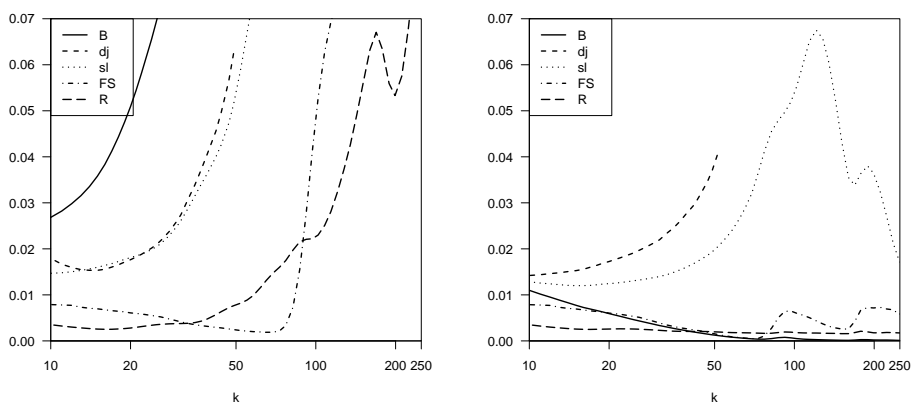


Figura 4.20: Comparação dos erros quadráticos médios estimados (à esquerda) e das variâncias estimadas (à direita) dos estimadores em estudo para o processo AR(1), com  $n = 500$  e  $r = 5$ , numa escala logarítmica.

Tabela 4.1: Medidas de eficiência relativa dos estimadores em análise do processo ARMAX dado por (3.35), com  $n = 500$ .

Denominador ( $J$ )	Numerador ( $I$ )			
	R	B	dj	sl
B	0.5134	-	-	-
dj	0.2280	0.444	-	-
sl	0.1918	0.3735	0.8412	-
FS	0.1780	0.3467	0.7809	0.9283

Tabela 4.2:  $REFF_0^{I/J}$  dos estimadores em estudo para o processo AR(1) dado por (2.13), com  $n = 500$ .

$(r = 2)$				
Denominador ( $J$ )	Numerador ( $I$ )			
	R	B	dj	sl
B	0.3245	-	-	-
dj	0.2490	0.7671	-	-
sl	0.2088	0.6434	0.8387	-
FS	0.1857	0.5728	0.7468	0.8903
$(r = 3)$				
Denominador ( $J$ )	Numerador ( $I$ )			
	R	B	dj	sl
B	0.3251	-	-	-
dj	0.3555	1.0937	-	-
sl	0.3217	0.9897	0.9049	-
FS	0.4353	1.3390	1.2243	1.3529
$(r = 5)$				
Denominador ( $J$ )	Numerador ( $I$ )			
	R	B	dj	sl
B	0.3063	-	-	-
dj	0.4052	1.3229	-	-
sl	0.4140	1.3517	1.0218	-
FS	1.1593	3.785	2.8612	2.8001

Tabela 4.3: Indicador de redução do viés dos estimadores em análise do processo ARMAX dado por (3.35), com  $n = 500$ .

Denominador ( $J$ )	Numerador ( $I$ )			
	R	B	dj	sl
B	0.5618	-	-	-
dj	0.5369	0.9557	-	-
sl	0.0970	0.1726	0.1806	-
FS	0.1191	0.212	0.2218	1.228

Tabela 4.4:  $BRI_0^{I/J}$  dos estimadores em estudo para o processo AR(1) dado por (2.13), com  $n = 500$ .

$(r = 2)$				
Denominador ( $J$ )	Numerador ( $I$ )			
	R	B	dj	sl
B	0.0775	-	-	-
dj	0.1983	2.5572	-	-
sl	0.0419	0.5401	0.2112	-
FS	0.0321	0.4138	0.1618	0.7661

$(r = 3)$				
Denominador ( $J$ )	Numerador ( $I$ )			
	R	B	dj	sl
B	0.0173	-	-	-
dj	4.2395	245.3512	-	-
sl	0.0206	1.1940	0.0049	-
FS	0.017	0.9826	0.004	0.823

$(r = 5)$				
Denominador ( $J$ )	Numerador ( $I$ )			
	R	B	dj	sl
B	0.0213	-	-	-
dj	0.3193	14.9889	-	-
sl	0.0627	2.9416	0.1963	-
FS	0.0753	3.5352	0.2359	1.2018

## 4.5 Aplicações a dados reais

Os dados financeiros e hídricos da Secção 3.8 também nos serviram como exemplos de aplicação dos resultados deste capítulo. Deste modo, estimámos o índice de cruzamentos para estes dois conjuntos de dados “DAX30” e “Caudal efluente numa barragem”, e apresentamos os resultados nas Figuras 4.21 e 4.22.

Como já tínhamos constatado no capítulo anterior, as estimativas obtidas com o estimador de *runs* para os dados “DAX30”, parecem indicar-nos que o valor de  $\eta$  é aproximadamente igual a 0.9. No entanto, as estimativas obtidas com os outros estimadores sugerem um valor mais próximo de 0.8. Relembremos que para este conjunto de dados, muitas vezes modelados por processos GARCH(1,1), não nos foi possível verificar a validade de  $\tilde{D}^{(3)}(\mathbf{u})$ . Se, de facto, esta condição não é válida, poderemos estar a sobreavaliar o número de *runs* de cruzamentos, conduzindo a estimativas mais elevadas do índice de cruzamentos. Quanto ao estimador  $\hat{\eta}_n^{FS}(k)$ , as estimativas encontram-se em zonas muito inferiores às obtidas pelos outros estimadores.

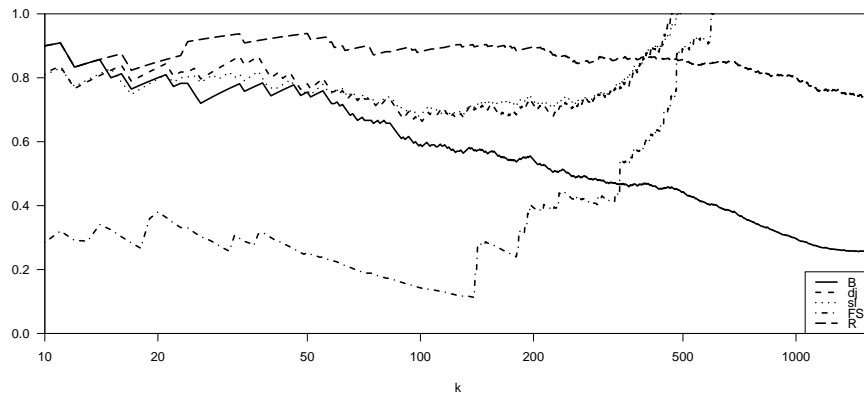


Figura 4.21: Estimativas de  $\eta$  obtidas com  $\hat{\eta}_n^B(k)$ ,  $\hat{\eta}_n^{dj}(k)$ ,  $\hat{\eta}_n^{sl}(k)$ ,  $\hat{\eta}_n^{FS}(k)$  e  $\hat{\eta}_n^R(k)$  para os *log-returns* dos valores diários de fecho do índice DAX30 no período compreendido entre 03-01-2000 e 29-06-2012.

Também para os dados “Caudal efluente numa barragem” não temos compatibilidade de resultados entre o estimador de *runs* e os outros estimadores, sendo que neste caso as diferenças ainda são mais acentuadas do que no exemplo DAX30. O estimador de *runs* aponta para valores de  $\eta$  próximos de um, mas pelos outros métodos, esta conjetura parece não se confirmar. Este é um exemplo interessante de dados reais em que a partilha de conhecimento, entre as áreas da teoria de valores extremos e as áreas específicas de hidrologia, poderiam ajudar a tomar uma decisão sobre a eventualidade de podermos, ou não, assumir a condição  $\tilde{D}^{(3)}(\mathbf{u})$  e usar o estimador de *runs* como referência para estimarmos  $\eta$  para estes dados.

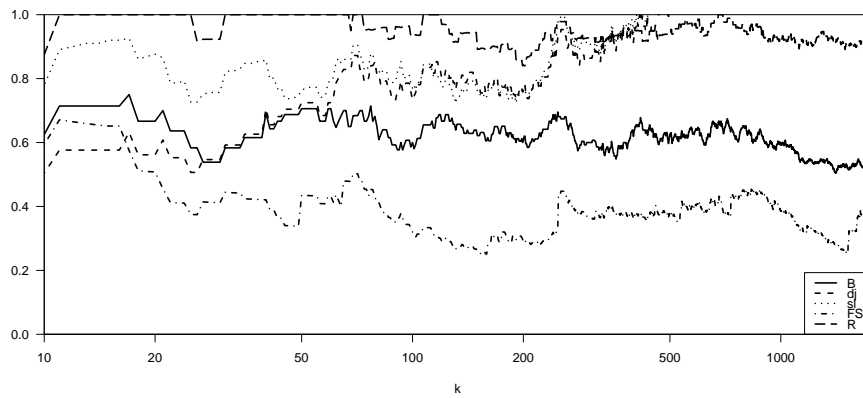


Figura 4.22: Estimativas de  $\eta$  obtidas com  $\hat{\eta}_n^B(k)$ ,  $\hat{\eta}_n^{dj}(k)$ ,  $\hat{\eta}_n^{sl}(k)$ ,  $\hat{\eta}_n^{FS}(k)$  e  $\hat{\eta}_n^R(k)$  do caudal efluente médio diário (em  $m^3/s$ ), na barragem do Fratel nos meses de novembro a março, entre 01-01-1984 e 30-11-2008.

## Conclusão

Os temas abordados ao longo deste trabalho poderão dividir-se em duas partes, a primeira dedicada ao estudo de propriedades do índice de cruzamentos e a segunda dedicada à sua estimação. Na primeira parte, formada pelo Capítulo 2, provámos que a intensidade do processo pontual limite do processo pontual de  $T$ -cruzamentos,  $T \geq 1$ , é caracterizada por um parâmetro  $\eta^{(T)} \in [0, 1]$ , a que chamámos índice de  $T$ -cruzamentos e introduzimos condições de dependência local que permitiram obter fórmulas simples de cálculo deste parâmetro. Mostrámos de que forma os índices de  $T$ -cruzamentos se relacionavam entre si e com o índice de cruzamentos. Calculámos o índice de cruzamentos após subamostragem sistemática de passo constante  $T$  a partir de uma relação com o índice de  $T$ -cruzamentos da sucessão inicial. Os resultados desta primeira parte foram publicados em Sebastião *et al.* (2010).

A segunda parte, é sem dúvida a maior e engloba os Capítulos 3 e 4 deste trabalho. Nesta parte apresentámos cinco estimadores para o índice de cruzamentos e provámos as suas propriedades assintóticas. Estes estimadores foram obtidos a partir de diferentes caracterizações assintóticas do índice de cruzamentos. Aplicámos estes estimadores em modelos existentes na literatura, para os quais o valor do índice de cruzamentos é conhecido, e em conjuntos de dados obtidos a partir de fenómenos reais. Comparámos, via técnicas de Monte Carlo, o conjunto de estimadores propostos.

No futuro seria interessante analisar os resultados do segundo capítulo com outras funções deterministas de subamostragem e utilizar metodologias *Jackknife* para construir estimadores com menor viés e erro quadrático médio do que os aqui propostos.

## Bibliografia

- Alpuim, M. (1989). An extremal markovian sequence. *J. Appl. Prob.*, 26, 219-232.
- Ancona-Navarrete, M. A. e Tawn, J. A. (2000). A comparison of methods for estimating the extremal index. *Extremes*, 3(1), 5-38.
- Beirlant, J., Goegebeur, Y., Teugels, J., e Segers, J. (2004). *Statistics of extremes*. Chichester: John Wiley & Sons Ltd. (Theory and applications, With contributions from Daniel De Waal and Chris Ferro)
- Canto e Castro, L. (1992). *Sobre a teoria assintótica de extremos*. Tese de Doutorado, Faculdade de Ciências da Universidade de Lisboa.
- Carlstein, E. (1986). The use of subseries values for estimating the variance of a general statistic from a stationary sequence. *Ann. Statist.*, 14(3), 1171-1179.
- Chernick, M. e Davis, R. A. (1982). Extremes in autoregressive processes with uniform marginal distributions. *Statist. Prob. Lett.*, 1, 85-88.
- Chernick, M., Hsing, T., e McCormick, W. (1991). Calculating the extremal index for a class of stationary sequences. *Adv. Appl. Prob.*, 23, 835-850.
- Coles, S. (2001). *An introduction to statistical modeling of extreme values*. London: Springer-Verlag London Ltd.
- Drees, H. (2011). Bias correction for estimators of the extremal index. *ArXiv e-prints*.
- Efron, B. (1979). Bootstrap methods: another look at the jackknife. *Ann. Statist.*, 7(1), 1-26.
- Embrechts, P., Klüppelberg, C., e Mikosch, T. (1997). *Modelling extremal events* (Vol. 33). Berlin: Springer-Verlag. (For insurance and finance)
- Ferreira, H. (1994). Multivariate extreme values in t-periodic random sequences under mild oscillation restrictions. *Stochastic Process. Appl.*, 49, 111-125.
- Ferreira, H. (2006). The upcrossings index and the extremal index. *J. Appl. Prob.*, 43, 927-937.
- Ferreira, H. (2007). Runs of high values and the upcrossings index for a stationary sequence. In *Proceedings of the 56th session of the ISI*.
- Ferreira, M. e Ferreira, H. (2012). On extremal dependence: some contributions. *TEST*, 21, 566-583.
- Ferro, C. A. T. e Segers, J. (2003). Inference for clusters of extreme values. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 65(2), 545-556.
- Fisher, R. e Tippett, L. (1928). Limiting forms of the frequency distribution of the largest or smallest member of a sample. *Mathematical Proceedings of the Cambridge Philosophical Society*, 24, 180-190.
- Fréchet, M. (1927). Sur la loi de probabilité de l'écart maximum. *Ann. de la Soc. Polonaise de Math.*, 6(93).
- Gnedenko, B. (1943). Sur la distribution limite du terme maximum d'une série aléatoire. *Ann. Math.*, 44, 423-453.
- Gomes, M. (1993). On the estimation of parameters of rare events in environmental time series. In V. Barnett e K. Turkman (Eds.), (chap. Statistics for the environment 2: Water related issues). John Wiley & Sons.

- Gomes, M., Hall, A., e Miranda, M. (2008). Subsampling techniques and the jackknife methodology in the estimation of the extremal index. *Comput. Statist. Data Anal.*, 52(4), 2022-2041.
- Gomes, M. e Oliveira, O. (2001). The bootstrap methodology in statistics of extremes—choice of the optimal sample fraction. *Extremes*, 4(4), 331-358.
- Gumbel, E. J. (1958). *Statistics of extremes*. New York: Columbia University Press.
- Haan, L. de e Ferreira, A. (2006). *Extreme value theory*. New York: Springer. (An introduction)
- Hall, A. e Scotto, M. (2003). Extremes of sub-sampled integer-valued moving average models with heavy-tailed innovations. *Statistics & Probability Letters*, 63, 97-105.
- Hsing, T. (1991). Estimating the parameters of rare events. *Stochastic Process. Appl.*, 37, 117-139.
- Hsing, T. (1993). Extremal index estimation for a weakly dependent stationary sequence. *Ann. Statist.*, 21(4), 2043-2071.
- Hsing, T., Hüsler, J., e Leadbetter, M. (1988). On the exceedance point process for a stationary sequences. *Prob. Th. Rel. Fields*, 78, 97-112.
- Klar, B., Lindner, F., e Meintanis, S. (2012). Specification tests for the error distribution in garch models. *Computational Statistics & Data Analysis*, 56(11), 3587 - 3598.
- Künsch, H. (1989). The jackknife and the bootstrap for general stationary observations. *Ann. Math.*, 17, 1217-1241.
- Lahiri, S. N. (2003). *Resampling methods for dependent data*. New York: Springer-Verlag.
- Laurini, F. e Tawn, J. A. (2006). *The extremal index for garch(1,1) processes with t-distributed innovations* (Tech. Rep.).
- Leadbetter, M. (1974). On extreme values in stationary sequences. *Z. Warsch. Gebiete*, 28, 289-303.
- Leadbetter, M. (1983). Extremes and local dependence in stationary sequences. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 65, 291-306.
- Leadbetter, M., Lindgren, G., e Rootzén, H. (1983). *Extremes and related properties of random sequences and processes*. New York: Springer-Verlag.
- Leadbetter, M. e Nandagopalan, S. (1989). On exceedance point process for stationay sequences under mild oscillation restrictions. In J. Hüsler e D. Reiss (Eds.), *Extreme value theory* (p. 69-80). New York: Springer-Verlag.
- Leadbetter, M. e Rootzén, H. (1988). Extremal theory for stochastic processes. *The Annals of Probability*, 16(2), 431-478.
- Liu, R. Y. e Singh, K. (1992). Moving blocks jackknife and bootstrap capture weak dependence. In *Exploring the limits of bootstrap (East Lansing, MI, 1990)* (pp. 225-248). New York: Wiley.
- Martins, A. (2004). *Coeficientes extremais*. Tese de Doutorado, Universidade da Beira Interior. Covilhã.
- Martins, A. e Ferreira, H. (2004). The extremal index of a sub-sampled process. *J. Statist. Plann. Inf.*, 124, 145-152.
- Mikosch, T. e Starica, C. (2000). Limit theory for the sample autocorrelations and extremes of a GARCH (1,1) process. *Ann. Statist.*, 28(5), 1427-1451.

- Mikosch, T. e Starica, C. (2004). Changes of structure in financial time series and the Garch model. *REVSTAT*, 2(1), 41-73.
- Nandagopalan, S. (1990). *Multivariate extremes and estimation of the extremal index*. Tese de Doutorado, University of North Carolina. Chapel Hill.
- O'Brien, G. (1987). Extreme values for stationary and markov sequences. *The Annals of Probability*, 15(1).
- Politis, D. N. e Romano, J. P. (1992). A circular block-resampling procedure for stationary data. In *Exploring the limits of bootstrap (East Lansing, MI, 1990)* (pp. 263-270). New York: Wiley.
- Politis, D. N. e Romano, J. P. (1994). The stationary bootstrap. *J. Amer. Statist. Assoc.*, 89(428), 1303-1313.
- Reiss, R.-D. e Thomas, M. (2007). *Statistical analysis of extreme values with applications to insurance, finance, hydrology and other fields* (Third ed.). Basel, Switzerland: Birkhäuser Verlag.
- Robert, C. Y. (2009). Asymptotic distributions for the intervals estimators of the extremal index and the cluster-size probabilities. *J. Statist. Plann. Inference*, 139(9), 3288-3309.
- Robert, C. Y., Segers, J., e Ferro, C. A. T. (2009). A sliding blocks estimator for the extremal index. *Electron. J. Stat.*, 3, 993-1020.
- Robinson, M. e Tawn, J. (2000). Extremal analysis of processes sampled at different frequencies. *Journal of the Royal Statistical Society*, 62, 117-135.
- Scotto, M. e Turkman, K. (2002). On the extremal behavior of sub-sampled solutions of stochastic difference equations. *Portugal Mathematics*, 59, 267-282.
- Scotto, M., Turkman, K., e Anderson, C. (2003). Extremes of some sub-sampled time series. *Journal of time series analysis*, 24, 579-590.
- Sebastião, J. R., Martins, A. P., Ferreira, H., e Pereira, L. (2012). Estimating the Upcrossings Index. *Em processo de revisão na Test*, <http://arxiv.org/abs/1204.1905>.
- Sebastião, J., Martins, A., Pereira, L., e Ferreira, H. (2010). Clustering of upcrossings of high values. *J. Statist. Plann. Inference*, 140(4), 1003-1012.
- Singh, K. (1981). On the asymptotic accuracy of Efron's bootstrap. *Ann. Statist.*, 9(6), 1187-1195.
- Smith, R. e Weissman, I. (1994). Estimating the extremal index. *Journal of the Royal Statistical Society*, 53(1).
- Süveges, M. (2007). Likelihood estimation of the extremal index. *Extremes*, 10(1-2), 41-55.
- Süveges, M. e Davison, A. C. (2010). Model misspecification in peaks over threshold analysis. *Ann. Appl. Stat.*, 4(1), 203-221.
- Tiago de Oliveira, J. (Ed.). (1984). *Statistical extremes and applications* (Vol. 131). Dordrecht: D. Reidel Publishing Co.
- Utev, S. A. (1990). Central limit theorem for dependent random variables. In *Probability theory and mathematical statistics, Vol. II (Vilnius, 1989)* (pp. 519-528). Vilnius: "Mokslas".
- von Mises, R. (1936). La distribution de la plus grande de n valeurs. In *Selected papers ii* (pp. 271-294). Providence, R.I., 1954: Amer. Math. Soc.
- Weissman, I. e Novak, S. Y. (1998). On blocks and runs estimators of the extremal index. *J.*

*Statist. Plann. Inference*, 66(2), 281-288.