



UNIVERSIDADE D
COIMBRA

Ricardo André Oliveira Marques

ESTATÍSTICA DE EVENTOS EXTREMOS
ATÉ ONDE PODEREMOS SALTAR?

VOLUME 1

Dissertação no âmbito do Mestrado em Matemática, Ramo Estatística, Otimização e Matemática Financeira orientada pela Professora Doutora Maria da Graça Santos Temido Neves Mendes e apresentada ao Departamento de Matemática da Faculdade de Ciências e Tecnologia.

Junho de 2022

Estatística de eventos extremos

Ricardo André Oliveira Marques



UNIVERSIDADE D
COIMBRA

Mestrado em Matemática

Junho 2022

Agradecimentos

Uma dissertação de mestrado é um percurso longo, exigente, com muitos percalços e desafios que sem o apoio da pessoas mais importantes para mim e da instituição UC não me seria possível concluir com sucesso.

Gostava de agradecer especialmente à minha orientadora, Professora Maria da Graça Temido, por me ter dado a oportunidade de desenvolver este tema, pelo rigor e competência, pela paciência e dedicação que teve para responder às minhas dúvidas e me guiar da melhor forma durante este percurso e por isto o meu mais sincero obrigado.

Não me posso esquecer de agradecer a todos os professores com quem me cruzei até agora, especialmente na licenciatura e mestrado, sem eles também não seria possível terminar este percurso. Todos de alguma forma contribuíram para a formação académica, pessoal e para a concretização deste objetivo.

À minha professora de matemática de secundário, Fernanda Pacheco, que me incutiu e estimulou o meu gosto pela matemática. Foi e é uma inspiração para mim, sem ela não teria começado a trilhar este caminho gratificante.

Aos meus pais e irmã por me permitirem fazer este percurso, por me apoiarem e motivarem sempre, não deixando nunca haver dúvidas a pairarem na minha cabeça sobre as minhas capacidades. As minhas conquistas serão sempre vossas também.

Gostaria também de fazer referência a alguns dos meus amigos: Nuno Ferreira, João Marcelino, Nuno Freitas, Sofia Ribeiro, Francisco Frota, Diogo Soares, Diana Pimentel e Francisca Martins. Todos eles através do seu apoio e motivação foram determinantes neste percurso.

Por fim, gostava de dedicar esta dissertação aos meus pais. São os melhores do mundo!

Resumo

Na origem da teoria estatística de extremos está o Teorema do limite extremal ou Teorema de Gnedenko. Este resultado fundamental estabelece que, em condições bastante gerais, a classe dos possíveis limites em distribuição do máximo de n variáveis aleatórias, reais independentes e identicamente distribuídas, coincide com a classe das distribuições de valor extremo (ou max-estáveis). Esta classe divide-se nas três sub-classes das funções de distribuição de Fréchet, Weibull e Gumbel.

A motivação para este trabalho foi o conteúdo do artigo "How far can man go?" ([6]). A partir de uma coleção de dados com as maiores marcas de saltos em comprimento (masculino), os autores justificam o facto de o record (8.95m) obtido pelo atleta Mike Powel em 1991 ainda se manter, uma vez que estes dados estimam um salto máximo de aproximadamente 9m.

Este trabalho contém uma introdução à teoria estatística de valores extremos, onde o Teorema do Limite Extremal é motivado e demonstrado. A caracterização dos domínios de atração das três distribuições de valor extremo é baseada e complementada com a introdução de condições adequadas sobre a função quantil. Para o índice de cauda, parâmetro que permite decidir qual das três distribuições de valor extremo está subjacente aos dados, são apresentados vários estimadores e provada a sua consistência. Ao papel fundamental do índice de cauda juntam-se os testes estatísticos de escolha de domínios de atração. Esta análise permite estimar quantis elevados e o extremo superior do suporte da variável em estudo, quando este é finito. O trabalho termina com duas aplicações a dados extremos, nomeadamente, aos dados do triplo salto feminino e aos tempos mínimos de resolução do cubo de Rubik.

Conteúdo

Lista de Figuras	ix
Lista de Tabelas	xi
1 Introdução	1
2 Fundamentos Teóricos	5
2.1 Teorema de Gnedenko	5
2.2 Domínios de atração	14
3 Estimação do índice de cauda e escolha de domínios de atração	23
3.1 Estimadores do índice de cauda	23
3.1.1 Estimador de Hill	23
3.1.2 Estimador de Pickands	24
3.1.3 Estimador dos Momentos	26
3.1.4 Estimador da mistura de momentos	27
3.2 Finitude e estimação de x_F	28
3.3 Escolha de domínios de atração	33
3.4 Estimação dos parâmetros e escolha da função de distribuição G_γ	36
3.4.1 Método dos momentos ponderados	36
3.4.2 Método da Máxima Verosimilhança	38
3.4.3 Teste de Gumbel	38
4 Aplicações	39
4.1 Aplicações ao triplo salto feminino	39
4.2 Aplicação aos tempos do cubo de Rubik	46
Bibliografia	49
Anexo A	51
Anexo B Códigos R	55
B.1 Código Triplo Salto feminino	55
B.2 Código Cubo de Rubik	61

Lista de Figuras

2.1	Esboço do gráfico da densidade da lei de Gumbel com $\lambda = 1$ e $\delta = 0$	11
2.2	Esboço do gráfico da densidade da lei de Fréchet ($\lambda = 0, \delta = 2, \alpha = 2$).	11
2.3	Esboço do gráfico da densidade da lei de Weibull ($\lambda = 0, \delta = 1, \alpha = 2$).	11
4.1	Valores de $\hat{\gamma}_P$ versus k	39
4.2	Valores de $\hat{\gamma}_M$ versus k	40
4.3	Valores de $\hat{\gamma}_-$ versus k	40
4.4	Valores de $\hat{\gamma}_{MM}$ versus k	40
4.5	Valores de $G_{n,k}$ versus k	41
4.6	Valores de $HW_{n,k}$ versus k	42
4.7	Valores de $G_{n,k}^*$ versus k	42
4.8	Valores de $E_{n,k}$ versus k	42
4.9	Valores de \hat{x}_F versus k , considerando o estimador $\hat{\gamma}_-$ (expressão 3.10).	43
4.10	Valores de \hat{x}_F versus k , considerando o estimador $\hat{\gamma}_M$ (expressão 3.10).	44
4.11	Valores de \hat{x}_F versus k , considerando o estimador $\hat{\gamma}_{MM}$ (expressão 3.10).	44
4.12	Valores de \hat{x}_F versus k (expressão 3.16).	44
4.13	Ajustamento de uma densidade e QQ-Plot (G_γ)	47
4.14	Histograma	47

Lista de Tabelas

4.1	Estimativas de γ e de x_F – Triplo salto feminino	45
4.2	Estimativas de γ e de x_F – Triplo salto masculino	45
4.3	Estimativas de γ e de x_F – Salto em comprimento feminino	46
4.4	Estimativas dos parâmetros da f.d. G_γ	48
4.5	Quantis de ordem p estimados e empíricos	48

Capítulo 1

Introdução

"How far can Man go?" é uma pergunta de carácter genérico, mas de grande pertinência quando se pensa em determinados desportos como salto em comprimento, salto em altura, triplo salto, etc.

No campeonato mundial de atletismo realizado em Tokyo, no dia 30 de Agosto de 1991 o atleta norte-americano Mike Powell atingiu a marca record de 8.95 m em salto em comprimento. Este valor ultrapassou o record do norte-americano Bob Beamon (8.90 m), conquistado três décadas antes e que se pensava inultrapassável.

Estes acontecimentos extremos e a sua grande distância no tempo, levaram os autores de *"How far can Man go?"* (Fraga Alves [6]) a aplicar a Teoria Estatística de Extremos aos dados disponíveis. Estes dados permitiram concluir que a variável aleatória que representa o salto em comprimento tem extremo superior do suporte aproximadamente 9 m.

Assim, à pergunta dos autores do referido trabalho,

"What is Largest Jump that Man can achieve, given today's state of art?"

a resposta foi

"it seems possible to attain an ultimate record of 9 m for the Long Jump Men outdoors".

Este artigo foi publicado em 2013 e o record ainda hoje (Junho de 2022) se mantém. A leitura cuidada deste artigo (Fraga Alves [6]) foi a motivação para esta dissertação.

A Teoria de Extremos é um ramo da Teoria da Probabilidade e da Estatística que tem por objetivo a modelação de fenómenos aleatórios de ocorrência excepcional, mas que podem ter um impacto significativo ou um interesse muito especial. De facto, os extremos de um conjunto ordenado de observações expressam o que de melhor ou de pior pode ocorrer na natureza ou no quotidiano das sociedades. Como exemplo, temos a catástrofe ocorrida nos Países Baixos em Fevereiro de 1953, devido à ocorrência de um maremoto e que gerou a necessidade imediata de construir o chamado dique holandês. Isto deu origem ao problema da determinação da altura do dique que pudesse resistir com segurança a futuras catástrofes. Como segundo exemplo temos o contexto do desporto já referido acima.

A primeira forma considerada na obtenção dos dados foi aquela em que estes podem ser entendidos como uma amostra de variáveis aleatórias reais independentes e identicamente distribuídas, situação considerada ao longo desta dissertação.

Na génese da Teoria Estatística de Extremos está o Teorema de Gnedenko, no qual se mostra que a função de distribuição (f.d.) limite do máximo de n variáveis aleatórias (v.a.'s) independentes e identicamente distribuídas (i.i.d.) coincide com a classe das f.d.'s max-estáveis (ou de valor extremo). Concretamente, denotando por F a f.d. destas variáveis, prova-se que, se existir uma normalização linear adequada, tal que a sucessão $F^n(a_n x + b_n)$ converge, então a f.d. limite, G , verifica

$$\forall t > 0, \exists a_t > 0, b_t \in \mathbb{R} : G^t(a_t x + b_t) = G(x), x \in \mathbb{R}, \quad (1.1)$$

o que define a classe das f.d.'s max-estáveis. Esta classe divide-se nas três famílias das f.d.'s Fréchet, Weibull e Gumbel. Sempre que tal limite ocorre dizemos que F pertence ao domínio de atração de G e escreve-se $F \in \mathcal{D}(G)$. Estas três f.d.'s de valor extremo podem ser apresentadas de forma unificada como

$$G_\gamma\left(\frac{x-b}{a}\right) = \begin{cases} \exp(-e^{-\frac{x-b}{a}}) & x \in \mathbb{R}, \quad \gamma = 0 \\ \exp(-(1 + \gamma\frac{x-b}{a})^{-1/\gamma}) & 1 + \gamma\frac{x-b}{a} > 0, \quad \gamma \neq 0 \\ 0 & 1 + \gamma\frac{x-b}{a} \leq 0, \quad \gamma > 0 \\ 1 & 1 + \gamma\frac{x-b}{a} \leq 0, \quad \gamma < 0. \end{cases} \quad (1.2)$$

O parâmetro de forma $\gamma \in \mathbb{R}$ é designado por índice de cauda e define aquelas três famílias Fréchet, Weibull e Gumbel, uma vez que a estas correspondem $\gamma > 0$, $\gamma < 0$ e $\gamma = 0$, respetivamente.

Esta dissertação está dividida em quatro capítulos.

No segundo capítulo, depois de uma exposição mais detalhada destes conceitos, enunciamos e demonstramos o Teorema de Gnedenko. Segue-se a caracterização dos três domínios de atração Fréchet, Weibull e Gumbel, feita quase sempre baseada nas propriedades da função quantil F^\leftarrow . Observamos que na literatura de extremos é usual chamar também função quantil a $U(t) = F^\leftarrow(1 - 1/t)$, $t > 1$). A equivalência entre $F \in \mathcal{D}(G_\gamma)$ e o limite

$$\lim_{t \rightarrow \infty} \frac{U(tx) - U(t)}{a(t)} = G_\gamma^{-1}(e^{-1/x}), x > 0, \quad (1.3)$$

provada neste segundo capítulo, tem uma importância fundamental nos resultados do terceiro capítulo.

O terceiro capítulo começa com a apresentação de quatro estimadores do índice de cauda, sendo provada a convergência em probabilidade dos três primeiros. Quando a estimativa do índice de cauda não é decisiva (por não estar suficientemente afastada de zero), ou mesmo quando o é, torna-se necessário introduzir um teste de escolha de domínios de atração. A este respeito, apresentamos os testes de Greenwood, de Hasofer & Wang e de Fraga Alves [8]. Para este último apresentamos a demonstração da convergência em distribuição da estatística de teste para uma f.d. de Gumbel.

Seguindo também [8], apresentamos um estimador do extremo superior do suporte de F , $x_F = \sup\{x : F(x) < 1\}$, quando este é finito, e provamos a sua convergência quase certa. Incluímos também no terceiro capítulo a apresentação de um teste estatístico à finitude de x_F , isto é, de hipóteses $H_0 : F \in D(G_0)$, $x_F = +\infty$ e $H_1 : F \in D(G_\gamma)$, $x_F < +\infty$, o qual foi introduzido em [12]

Na análise estatística dos dados extremos podemos seguir duas abordagens: a semi-paramétrica e a paramétrica.

Na abordagem semi-paramétrica não se assume nenhum tipo de f.d. concreta para a v.a. em estudo concentrando-se o estudo numa coleção dos k maiores valores observados. Assumindo que F

pertence a algum domínio de atração, estima-se o índice de cauda e realizam-se os testes descritos acima. Estimam-se quantis elevados e/ou x_F e ainda probabilidades de interesse. A determinação do melhor valor de k é um tema já bastante desenvolvido que não foi alvo do nosso estudo. Neste trabalho fizemos uma escolha informal do valor de k sugerida em [8].

A perspectiva paramétrica só pode ser seguida se os dados puderem ser considerados como máximos de n v.a.'s i.i.d.. Nesse caso, e de acordo com o Teorema de Gnedenko, procuramos o ajustamento com uma f.d. G_γ , restando apenas a estimação dos seus parâmetros, a qual pode ser complementada com um teste ao sinal de γ . A secção 3.5 é dedicada a esta abordagem paramétrica.

Por fim, no Capítulo 4 incluímos duas aplicações com dados de marcas de triplo salto feminino (outdoor) e com dados de tempos mínimos de resolução do cubo de Rubik. Os dados do triplo salto foram analisados com a abordagem semi-paramétrica e nos dados do cubo de Rubik foi considerada a perspectiva paramétrica. Uma vez que se trata de valores mínimos usámos a igualdade

$$\min(x_1, \dots, x_n) = -\max(-x_1, \dots, -x_n)$$

tendo tratado os máximos dos simétricos.

No tratamento dos dados foi utilizado o software R studio.

Para além de resultados ou afirmações com referências, o conteúdo deste trabalho encontra-se em [3].

O Anexo A contém informação relevante sobre convergências em distribuição e em probabilidade, bem como notações que lhe estão associadas, resultado da consulta de [1].

O Anexo B contém os códigos R utilizados em ambas as aplicações, tendo a sua elaboração beneficiado da consulta de [15].

Capítulo 2

Fundamentos Teóricos

2.1 Teorema de Gnedenko

Seja $\{X_n\}_{n \geq 1}$ uma sucessão de v.a.'s i.i.d. com f.d. F . Denotemos por M_n a v.a.r que representa o máximo das v.a. (X_1, X_2, \dots, X_n) , isto é, $M_n = \max\{X_1, \dots, X_n\}$. A f.d. de M_n é dada por

$$P(M_n \leq x) = P(X_1 \leq x, X_2 \leq x, \dots, X_n \leq x) = F^n(x), x \in \mathbb{R},$$

o que só tem interesse prático quando conhecemos F . Como $F^n(x)$ tem limite 0 ou 1, devemos procurar uma normalização real u_n tal que $F^n(u_n)$ tenha como limite uma f.d. não degenerada. Concretamente, escolhemos uma normalização linear $u_n = a_n x + b_n$, a qual permite estabelecer um resultado assintótico para a f.d. de $a_n^{-1}(M_n - b_n)$, tal como foi estabelecido para a f.d. de $a_n^{-1}(\sum_{i=1}^n X_i - b_n)$, no âmbito do Teorema do Limite Central. Assim, temos que determinar em que condições existem constantes de normalização reais $a_n > 0$ e b_n tais que

$$F^n(a_n x + b_n) = P(a_n^{-1}(M_n - b_n) \leq x) \rightarrow G(x), n \rightarrow +\infty, \quad (2.1)$$

para alguma f.d. G não degenerada, onde o limite ocorre nos pontos de continuidade de G . De seguida, há que caraterizar a f.d. limite de G .

O Teorema 2.1.2 (ou Teorema de Gnedenko) dá resposta a esta questão. Apresentamos seguidamente algumas definições e resultados necessários à sua demonstração e a muitos outros resultados que se apresentam ao longo do trabalho.

Dada uma função real f não decrescente, representamos por f^{\leftarrow} a função inversa generalizada à esquerda de f , isto é,

$$f^{\leftarrow}(y) = \inf\{x : f(x) \geq y\}, y \in \mathbb{R}.$$

Prova-se [13] que

$$(f^{\leftarrow})^{\leftarrow}(x) = f^-(x) = \lim_{y \downarrow x} f(y). \quad (2.2)$$

A função f^- é a versão contínua à esquerda da função f e $f^+(x) = \lim_{y \uparrow x} f(y)$ é a versão contínua à direita de f .

Dada uma f.d. F denotamos por U a função $(\frac{1}{1-F})^{\leftarrow}$, tendo-se

$$U(t) = F^{\leftarrow}(1 - 1/t), t > 1.$$

As funções F^{\leftarrow} e U são contínuas à esquerda.

Lema 2.1.1 *Seja $\{f_n\}$ uma sucessão de funções reais não decrescentes e g uma função não decrescente. Suponhamos, ainda que, para cada $x \in]a, b[$, ponto continuidade de g , se tem $\lim_{n \rightarrow \infty} f_n(x) = g(x)$. Sejam f_n^{\leftarrow} e g^{\leftarrow} as funções inversas contínuas à esquerda de f_n e g , respetivamente. Para cada $x \in [g(a), g(b)]$, x é ponto de continuidade de g^{\leftarrow} , temos que $\lim_{n \rightarrow \infty} f_n^{\leftarrow}(x) = g^{\leftarrow}(x)$.*

Lema 2.1.2 (Resnick [13]) *Seja $\{f_n\}$ uma sucessão de funções monótonas. Se $f_n \rightarrow f_0$ então $f_n^{\pm} \rightarrow f_0^{\pm}$.*

O teorema seguinte introduz a chamada condição de primeira ordem, da qual assenta num limite para a função U que é equivalente a (2.1).

Teorema 2.1.1 *Sejam $\{a_n > 0\}$ e $\{b_n\}$ sucessões de constantes reais e G uma f.d. real não degenerada. Os seguintes resultados são equivalentes:*

1.

$$\lim_{n \rightarrow \infty} F^n(a_n x + b_n) = G(x), \quad (2.3)$$

para cada ponto de continuidade x de G ;

2.

$$\lim_{t \rightarrow \infty} t(1 - F(a(t)x + b(t))) = -\log G(x), \quad (2.4)$$

para cada ponto de continuidade x de G tal que $0 < G(x) < 1$, $a(t) := a_{[t]}$ e $b(t) := b_{[t]}$, com $[t]$ a parte inteira de t ;

3.

$$\lim_{t \rightarrow \infty} \frac{U(tx) - b(t)}{a(t)} = D(x), \quad (2.5)$$

onde x é ponto de continuidade de $D(x) = G^{\leftarrow}(e^{-\frac{1}{x}})$, $a(t) := a_{[t]}$ e $b(t) = b_{[t]}$;

Prova: Começemos por provar a equivalência de 2. e 3.. Pelo Lema 2.1.1, temos

$$\lim_{t \rightarrow \infty} \left(\frac{1}{t(1 - F(a(t) \cdot + b(t)))} \right)^{\leftarrow}(x) = \left(\frac{1}{-\log G(\cdot)} \right)^{\leftarrow}(x).$$

Observamos que

$$\begin{aligned} \left(\frac{1}{-\log G} \right)^{\leftarrow}(x) &= \inf\{y : \frac{1}{-\log G(y)} \geq x\} = \inf\{y : \log G(y) \geq -\frac{1}{x}\} \\ &= \inf\{y : G(y) \geq e^{-\frac{1}{x}}\} = G^{\leftarrow}(e^{-\frac{1}{x}}) := D(x) \end{aligned}$$

Pela definição de inversa generalizada, vem

$$\begin{aligned}
 \left(\frac{1}{t(1-F(a(t) \cdot +b(t)))} \right)^{\leftarrow} (x) &= \inf \left\{ y : \frac{1}{t(1-F(a(t)y+b(t)))} \geq x \right\} = \\
 &= \inf \left\{ y : \frac{1}{(1-F(a(t)y+b(t)))} \geq xt \right\} \stackrel{z=a(t)y+b(t)}{=} \inf \left\{ \frac{z-b(t)}{a(t)} : \frac{1}{1-F(z)} \geq xt \right\} = \\
 &= \frac{1}{a(t)} \left(\inf \left\{ z : \frac{1}{1-F(z)} \geq xt \right\} - b(t) \right) = \frac{U(tx) - b(t)}{a(t)} \xrightarrow{t \rightarrow \infty} D(x),
 \end{aligned}$$

e assim 2. implica 3..

Por outro lado, desta última convergência e de (2.5) obtemos, aplicando o Lema 2.1.1,

$$\lim_{t \rightarrow +\infty} \left(\left(\frac{1}{t(1-F(a(t) \cdot +b(t)))} \right)^{\leftarrow} \right)^{\leftarrow} (x) = D^{\leftarrow}(x),$$

onde $D^{\leftarrow}(x) = \left(\left(\frac{1}{-\log G} \right)^{\leftarrow} \right)^{\leftarrow} (x)$. Então atendendo a (2.2)

$$\lim_{t \rightarrow +\infty} \left(\frac{1}{t(1-F(a(t) \cdot +b(t)))} \right)^{-} (x) = \left(\frac{1}{-\log G} \right)^{-} (x).$$

Aplicando o Lema 2.1.2, obtemos

$$\frac{U(tx) - b(t)}{a(t)} \xrightarrow{t \rightarrow \infty} D^{\leftarrow}(x),$$

ou seja

$$\lim_{t \rightarrow +\infty} \left(\frac{1}{t(1-F(a(t) \cdot +b(t)))} \right)^{+} (x) = \left(\frac{1}{-\log G} \right)^{+} (x),$$

o que permite concluir o limite (2.4) uma vez que F e G são contínuas à direita. Fica provado que 3. implica 2..

Notemos que (2.3) é equivalente a

$$\lim_{n \rightarrow +\infty} n \log(1 - (1 - F(a_n x + b_n))) = -\log G(x),$$

e como $\log(1-x) \sim -x$, $x > 0$, provamos que (2.3) é equivalente a

$$\lim_{n \rightarrow +\infty} n(1 - F(a_n x + b_n)) = -\log G(x), \quad (2.6)$$

pois $\log(1-x) \sim -x$, $x > 0$. Então, das relações anteriores (substituindo t por n) concluimos que (2.6) é equivalente a

$$\lim_{n \rightarrow +\infty} \frac{U(nx) - b_n}{a_n} = G^{\leftarrow}(e^{-1/x}). \quad (2.7)$$

Então, para provar as três equivalências do teorema, é suficiente provar que (2.7) implica 3. Seja x um ponto de continuidade D . Para $t \geq 1$, como U é não decrescente vem que

$$\frac{U([t]x) - b_{[t]}}{a_{[t]}} \leq \frac{U(tx) - b_{[t]}}{a_{[t]}} \leq \frac{U([t]x(1 + \frac{1}{[t]})) - b_{[t]}}{a_{[t]}}.$$

O lado direito da inequação será menor que $D(x')$ para qualquer ponto de continuidade $x' > x$ com $D(x') > D(x)$.

Como D é contínua em x , obtemos $\lim_{t \rightarrow \infty} \frac{U(tx) - b_{[t]}}{a_{[t]}} = D(x)$, provando assim a equivalência de 1 e 2. Assim, o Teorema 2.1.1 está provado. \square

A classe de f.d. G , limite em (2.3), é usualmente chamada de classe de f.d. de valor extremo. O Teorema de Gnedenko identifica tal classe.

Teorema 2.1.2 (Teorema de Gnedenko) *A classe das f.d. de valor extremo é $G_\gamma(\frac{x-b}{a})$, com $a > 0$, $b \in \mathbb{R}$, $\gamma \in \mathbb{R}$, onde*

$$G_\gamma(x) = \begin{cases} \exp(-e^{-x}) & x \in \mathbb{R}, \quad \gamma = 0 \\ \exp(-(1 + \gamma x)^{-1/\gamma}) & 1 + \gamma x > 0, \quad \gamma \neq 0 \\ 0 & 1 + \gamma x \leq 0, \quad \gamma > 0 \\ 1 & 1 + \gamma x \leq 0, \quad \gamma < 0. \end{cases} \quad (2.8)$$

Prova: Vamos considerar a classe de funções limite D em (2.5). Suponhamos que $x = 1$ é ponto de continuidade de D . Note-se que, para $x > 0$,

$$\lim_{t \rightarrow \infty} \frac{U(tx) - U(t)}{a(t)} = \lim_{t \rightarrow \infty} \frac{U(tx) - b_{[t]}}{a(t)} - \frac{U(t) - b_{[t]}}{a(t)} = D(x) - D(1) = E(x).$$

Para $y > 0$, escrevemos

$$\frac{U(txy) - U(t)}{a(t)} = \frac{U(txy) - U(ty)}{a(ty)} \frac{a(ty)}{a(t)} + \frac{U(ty) - U(t)}{a(t)}. \quad (2.9)$$

Suponhamos que não existem os limites $(t \rightarrow +\infty)$ de $\frac{U(ty) - U(t)}{a(t)}$ e de $\frac{a(ty)}{a(t)}$. Nesse caso, existem duas subseqüências de reais t' e t'' para as quais aquelas duas funções têm limites distintos $B_1(y)$ e $B_2(y)$ e $A_1(y)$ e $A_2(y)$. Então, de (2.9) obtemos

$$E(xy) = E(x)A_i(y) + B_i(y), \quad (2.10)$$

$i = 1, 2$, onde x é um ponto de continuidade de $E(\cdot)$ e de $E(\cdot y)$. Para um valor de x arbitrário, consideramos uma sucessão $\{x_n\}$ de pontos de continuidade de $E(\cdot)$ e de $E(\cdot y)$ tal que $x_n \uparrow x$ ($n \rightarrow +\infty$). Então

$$E(x_n y) \rightarrow E(xy), \quad n \rightarrow +\infty$$

e

$$E(x_n) \rightarrow E(x), n \rightarrow +\infty,$$

pois E é contínua à esquerda. Assim, (2.10) verifica-se para x e y positivos. Também de (2.10) se prova que

$$E(x)(A_1(y) - A_2(y)) = B_2(y) - B_1(y).$$

Uma vez que E não pode ser independente de x (G é não degenerada), então a igualdade anterior só pode acontecer se $A_1(y) = A_2(y)$ e $B_1(y) = B_2(y)$. Assim, para $y > 0$, existe $A(y) = \lim_{t \rightarrow \infty} \frac{a(ty)}{a(t)}$ e para $x, y > 0$ tem-se que $E(xy) = E(x)A(y) + E(y)$.

Portanto, para $s = \log x$, $t = \log y$ ($x, y \neq 1$) e $H(x) = E(e^x)$ temos

$$\begin{aligned} H(t+s) &= E(e^{t+s}) = E(xy) = E(x)A(y) + E(y) \\ &= E(x)A(e^t) + E(e^t) = H(s)A(e^t) + H(t) \end{aligned} \quad (2.11)$$

Como $H(0) = 0$, podemos escrever

$$\frac{H(t+s) - H(t)}{s} = \frac{H(s)A(e^t) + H(t) - H(t)}{s} = \frac{H(s) - H(0)}{s} A(e^t). \quad (2.12)$$

Se assumirmos que t é um ponto de diferenciabilidade de H , de (2.12) provamos que $H'(0)$ existe e portanto, como (2.12) é válido para qualquer t , também provamos que H é diferenciável em qualquer $t \neq 0$. Então

$$H'(t) = H'(0)A(e^t). \quad (2.13)$$

Seja $Q(t) = \frac{H(t)}{H'(0)}$. Note-se que $H'(0)$ não pode ser zero porque H não pode ser constante, uma vez que G é não degenerada. Temos então que $Q(0) = 0$ (pois $H(0) = E(1) = 0$) e $Q'(0) = 1$.

Por (2.11), vem

$$Q(t+s) - Q(t) = \frac{H(t+s) - H(t)}{H'(0)} = \frac{H(s)A(e^t)}{H'(0)} = Q(s)A(e^t)$$

e por (2.13), temos

$$Q(t+s) - Q(t) = Q(s)A(e^t) = Q(s) \frac{H'(t)}{H'(0)} = Q(s)Q'(t). \quad (2.14)$$

Então, temos $Q(t+s) - Q(t) = Q(s)Q'(t)$. Subtraindo a mesma expressão com t e s trocados vem

$$-Q(t) + Q(s) = Q(s)Q'(t) - Q(t)Q'(s) \Leftrightarrow -Q(t) + Q(t)Q'(s) = Q(s)Q'(t) - Q(s)$$

$$\Leftrightarrow Q(t)(Q'(s) - 1) = Q(s)(Q'(t) - 1)$$

$$\Leftrightarrow Q(t) \frac{Q'(s)-1}{s} = Q(s) \frac{Q'(t)-1}{s}.$$

Portanto, para $s \rightarrow 0$, temos $Q(t)Q''(0) = Q'(t) - 1$. Segue-se então que, Q é duas vezes diferenciável e vem que $Q'(t)Q''(0) = Q''(t)$ ou seja $Q''(0) = \frac{Q''(t)}{Q'(t)}$.

Então,

$$(\log Q')'(t) = Q''(0) = \gamma \in \mathbb{R}, \forall t.$$

Começemos por admitir que $\gamma \neq 0$. Como $Q'(0) = 1$, vem que $Q'(t) = e^{\gamma t}$ e como $Q(0) = 0$ vem também que

$$Q(t) = \int_0^t e^{\gamma s} ds = \frac{e^{\gamma t} - 1}{\gamma}$$

se $\gamma \neq 0$ e $Q(t) = t$ se $\gamma = 0$. Posto isto, se $\gamma \neq 0$, temos que $H(t) = H'(0) \left(\frac{e^{\gamma t} - 1}{\gamma} \right)$ e

$$D(t) = D(1) + E(t) = D(1) + H(\ln t) = D(1) + H'(0) \left(\frac{t^\gamma - 1}{\gamma} \right),$$

do que resulta

$$D^{\leftarrow}(x) = \left(1 + \gamma \frac{x - D(1)}{H'(0)} \right)^{1/\gamma}. \quad (2.15)$$

Se $\gamma = 0$, temos $(\log Q')(t) = c$ (constante) donde $Q'(t) = e^c$. Como $Q'(0) = 1$, obtemos $c = 0$. Então $Q(t) = \int_0^t 1 ds = t$ e de $Q(t) = \frac{H(t)}{H'(0)}$ vem $H(t) = tH'(0)$. Tem-se então

$$D(t) = D(1) + H(\ln t) = D(1) + \ln t H'(0)$$

o que implica $D^{\leftarrow}(t) = \exp\left(\frac{x - D(1)}{H'(0)}\right)$. As expressões obtidas para $D(t)$ mostram que se trata de uma função contínua e estritamente crescente.

Para $\gamma = 0$, obtemos de igual modo, $D^{\leftarrow}(x) = \exp\left(-\frac{x - D(1)}{H'(0)}\right)$.

Agora, temos $D(x) = G^{\leftarrow}(e^{\frac{-1}{x}})$, donde $D^{\leftarrow}(x) = \frac{1}{-\log G(x)}$ e isto combinado com (2.15) dá a prova do teorema como terminada.

Se $x = 1$ não for ponto de continuidade de D , considera-se a função $U(tx_0)$ onde x_0 é ponto de continuidade de D e retoma-se a demonstração. \square

A γ chamamos índice de cauda de G . Se $\gamma = 0$, G_0 denota a f.d. de Gumbel

$$\Lambda(x) = \exp(-e^{-x}), \quad x \in \mathbb{R}.$$

No caso de $\gamma > 0$, G_γ denota a f.d. de Fréchet Φ_α ($\alpha = \frac{1}{\gamma} > 0$) dada por

$$\Phi_\alpha(x) = G_\gamma\left(\frac{x-1}{\gamma}\right) = \begin{cases} 0, & x < 0 \\ \exp(-x^{-\alpha}), & x \geq 0. \end{cases} \quad (2.16)$$

Para $\gamma < 0$, G_γ denota a f.d. de Weibull Ψ_α ($\alpha = \frac{1}{\gamma} < 0$) dada por

$$\Psi_\alpha(x) = G_\gamma\left(-\frac{x+1}{\gamma}\right) = \begin{cases} \exp(-(-x)^\alpha), & x < 0 \\ 1, & x \geq 0. \end{cases} \quad (2.17)$$

A classe das f.d.'s de valor extremo coincide com a classe das f.d.'s max-estáveis ([10]), isto é, f.d.'s G que verificam (1.1). Em particular

$$\forall n \in \mathbb{N}, \exists a_n > 0, b_n \in \mathbb{R} : G^n(a_n x + b_n) = G(x), x \in \mathbb{R}.$$

Esta última equação mostra que qualquer f.d. de valor extremo pertence ao seu próprio domínio de atração.

Passamos a apresentar o esboço do gráfico da densidade das três distribuições de valor extremo, para valores concretos dos seus parâmetros.

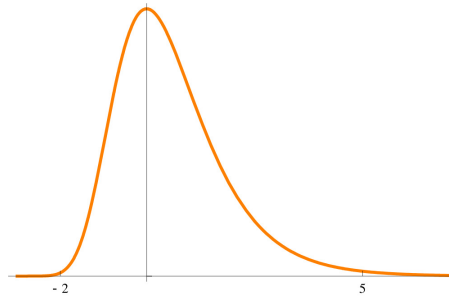


Fig. 2.1 Esboço do gráfico da densidade da lei de Gumbel com $\lambda = 1$ e $\delta = 0$.

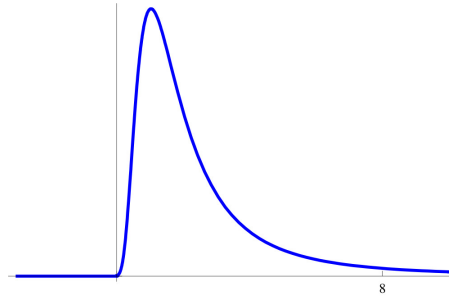


Fig. 2.2 Esboço do gráfico da densidade da lei de Fréchet ($\lambda = 0, \delta = 2, \alpha = 2$).

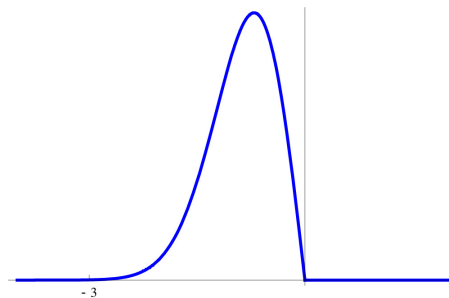


Fig. 2.3 Esboço do gráfico da densidade da lei de Weibull ($\lambda = 0, \delta = 1, \alpha = 2$).

O Teorema de Khintchine, que apresentamos de seguida, estabelece as alterações que se podem fazer às constantes a_n e b_n e identifica as consequências dessas alterações na f.d. limite G .

Assim, dada uma f.d. G_γ , chamamos domínio de atração de G_γ ao conjunto das f.d.'s F para as quais existem conjuntos $a_n > 0$ e b_n tais que se verifica (2.3).

Teorema 2.1.3 (Teorema de Khintchine) *Seja $\{F_n\}$ uma sucessão de f.d.'s, G uma f.d. não degenerada e $\{a_n\}$, $a_n > 0$, e $\{b_n\}$ sucessões de reais tais que*

$$\lim_{n \rightarrow +\infty} F_n(a_n x + b_n) = G(x).$$

Então existe uma f.d. não degenerada G^ e existem sucessões reais $\{\alpha_n\}$, $\alpha_n > 0$, e $\{\beta_n\}$ a satisfazer*

$$\lim_{n \rightarrow +\infty} F_n(\alpha_n x + \beta_n) = G^*(x)$$

se e só se $\lim_{n \rightarrow +\infty} \frac{\alpha_n}{a_n} = \alpha$ e $\lim_{n \rightarrow +\infty} a_n^{-1}(\beta_n - b_n) = \beta$, com $\alpha > 0$ e β finitos e $G^(x) = G(\alpha x + \beta)$.*

Exemplo 2.1.1 ([10]) *Seja F a f.d. da lei exponencial definida por*

$$F(x) = (1 - e^{-x})\mathbb{I}_{[0, +\infty[}(x), x \in \mathbb{R}.$$

Com $a_n = 1$ e $b_n = \log n$, obtemos $n(1 - F(x + \log n)) = ne^{-x - \log n} = e^{-x}$, $x \in \mathbb{R}$, para n tal que $x + \log n > 0$. Então, atendendo a (2.6), F pertence ao domínio de atração da f.d. de Gumbel.

Exemplo 2.1.2 ([10]) *As três f.d. de valor extremo pertencem ao seu próprio domínio de atração. Por exemplo, a f.d. de Fréchet com $\gamma = 1$, $a = 1$ e $b = 0$ é da forma $G_1(x) = \exp(-x^{-1})$, para $x > 0$, e $G_1(x) = 0$ para $x \leq 0$. Com $a_n = n$ e $b_n = 0$, obtemos*

$$G_1^n(a_n x + b_n) = \exp(-(n x)^{-1}) = \exp(-x^{-1}) = G_1(x), x > 0.$$

Para $x < 0$, tem-se $a_n x + b_n < 0$ e então $G_1(a_n x + b_n) = 0$.

Exemplo 2.1.3 ([10]) *Seja $F(x) = (1 - kx^{-\alpha})\mathbb{I}_{[k^{1/\alpha}, +\infty[}(x)$, $x \in \mathbb{R}$, a f.d. da lei de Pareto, onde k e α são constantes reais positivas. Se tomarmos $b_n = 0$ e $a_n = (kn)^{1/\alpha}$, obtemos*

$$n(1 - F(a_n x + b_n)) = n(1 - F((kn)^{1/\alpha} x)) = nk((kn)^{1/\alpha} x)^{-\alpha} = x^{-\alpha}, x > 0$$

e n tal que $n^{\frac{1}{\alpha}} x > 1$. Então, F pertence ao domínio de atração da f.d. de Fréchet.

Exemplo 2.1.4 *Seja F a f.d. da lei de Pareto definida no exemplo 2.1.3, com $\alpha = 1$. Defina-se a f.d. discreta H com suporte \mathbb{N} e tal que*

$$H(x) = F([x]), x \geq 1,$$

e $H(x) = 0$ para $x < 1$. Considerando $b_n = 0$ e $a_n = n$ obtemos, para n tal que $nx \geq 1$,

$$\begin{aligned} n(1 - H(a_n x + b_n)) &= n(1 - F([a_n x])) \\ &= n(1 - 1 + \frac{1}{[a_n x]}) = \frac{n}{[a_n x]}. \end{aligned}$$

Uma vez que $a_n x - 1 \leq [a_n x] \leq a_n x$, concluímos que $\frac{1}{a_n x} \leq n(1 - H(a_n x)) \leq \frac{n}{a_n x - 1}$ donde $\lim_{n \rightarrow +\infty} n(1 - H(a_n x)) = x^{-1}$, $x > 0$. Então, H pertence ao domínio de atração da f.d. de Fréchet.

Exemplo 2.1.5 Seja F uma f.d. com lei $\mathcal{U}(0, 1)$. Uma vez que $F(y) = y, y \in [0, 1]$, com $a_n = \frac{1}{n}$ e $b_n = 1$, obtemos

$$F^n\left(\frac{1}{n}x + 1\right) = \left(1 + \frac{x}{n}\right)^n \rightarrow e^x, n \rightarrow +\infty,$$

para $x < 0$. Então $G(x) = \exp(-(-x)^1), x < 0$, ou seja F pertence ao domínio de atração da f.d. de Weibull.

Exemplo 2.1.6 ([10]) Seja $\{X_n\}$ uma sucessão de v.a.'s normais i.i.d com média zero e variância 1. Vamos provar que,

$$P(a_n^{-1}(M_n - b_n) \leq x) \xrightarrow{n \rightarrow \infty} \exp(-e^{-x}), x \in \mathbb{R},$$

com

$$a_n = (2 \log n)^{-1/2}, \quad b_n = (2 \log n)^{1/2} - \frac{1}{2}(2 \log n)^{-1/2}(\log \log n + \log 4\pi). \quad (2.18)$$

Representemos por Φ e ϕ a f.d. e a função densidade da lei $\mathcal{N}(0, 1)$, respetivamente. Considere-se a sucessão $\{u_n\}$ real definida por

$$n(1 - \Phi(u_n)) = e^{-x}, x \in \mathbb{R}, \quad (2.19)$$

que verifica $u_n \xrightarrow{n \rightarrow \infty} +\infty$. Uma vez que $1 - \Phi(x) = \frac{\phi(x)}{x}$, de (2.19) vem que

$$\frac{1}{n} e^{-x} \frac{u_n}{\phi(u_n)} \rightarrow 1, n \rightarrow +\infty.$$

Aplicando logaritmos, obtemos $-\log n - x + \log u_n - \log \phi(u_n) \xrightarrow{n \rightarrow \infty} 0$, e como $\phi(u_n) = \frac{1}{\sqrt{2\pi}} e^{-\frac{u_n^2}{2}}$, concluímos que

$$-\log n - x + \log u_n + \frac{1}{2} \log 2\pi + \frac{u_n^2}{2} \rightarrow 0, n \rightarrow +\infty. \quad (2.20)$$

Dividindo por $\frac{u_n^2}{2}$ vem, $\frac{2 \log n}{u_n^2} \xrightarrow{n \rightarrow \infty} 1$, o que implica

$$\log 2 + \log \log n - 2 \log u_n \rightarrow 0, n \rightarrow +\infty.$$

Juntando este limite com (2.20), vem

$$-\log n - x + \frac{1}{2} \log 2 + \frac{1}{2} \log \log n + \frac{1}{2} \log 2\pi + \frac{u_n^2}{2} \xrightarrow{n \rightarrow \infty} 0.$$

Consequentemente,

$$u_n^2 = 2(x + \log n - \frac{1}{2} \log \log n - \frac{1}{2} \log 4\pi + o_n(1)),$$

ou seja,

$$u_n = (2 \log n)^{1/2} \left(1 + \frac{x - \frac{1}{2} \log \log n - \frac{1}{2} \log 4\pi}{\log n} + o(1/\log n) \right)^{\frac{1}{2}}.$$

Usando a igualdade $(1+x)^{1/2} = 1 + \frac{x}{2} + o(x)$, $x \rightarrow 0$, obtemos

$$u_n = (2 \log n)^{1/2} \left(1 + \frac{x - \frac{1}{2} \log \log n - \frac{1}{2} \log 4\pi}{2 \log n} + o(1/\log n) \right).$$

ou ainda

$$u_n = (2 \log n)^{-1/2} x + (2 \log n)^{1/2} - \frac{1}{2} (2 \log n)^{-1/2} (\log \log n + \log 4\pi) + o((\log n)^{-1/2}).$$

Temos assim, $u_n = a_n x + b_n + o(a_n)$ com a_n e b_n definidos em (2.18). Finalmente, por (2.19), obtemos

$$P(M_n \leq a_n x + b_n + o(a_n)) \xrightarrow[n \rightarrow \infty]{} \exp(-e^{-x}), x \in \mathbb{R},$$

ou seja,

$$P(a_n^{-1}(M_n - b_n) \leq x + o_n(1)) \xrightarrow[n \rightarrow \infty]{} \exp(-e^{-x}), x \in \mathbb{R}.$$

O Teorema de Khintchine garante o mesmo limite para a sucessão $v_n = a_n x + b_n$, concluindo assim a prova. Provamos assim que a f.d. da lei normal pertence ao domínio de atração da f.d. de Gumbel.

No caso em que X_n tem lei normal de média m e variância σ^2 (não necessariamente 0 e 1) basta considerar as constantes de normalização $\alpha_n = \sigma a_n$ e $\beta_n = \sigma b_n + m$.

2.2 Domínios de atração

Nesta secção apresentamos resultados que permitem caracterizar os domínios de atração de Fréchet, Weibull e Gumbel. Isto é, são apresentados resultados que, para os três casos $\gamma > 0$, $\gamma < 0$, $\gamma = 0$, permitem caracterizar as f.d.'s F para as quais existem constantes reais $a_n > 0$ e b_n tal que (2.3) se verifica.

O teorema seguinte estabelece condições necessárias e suficientes para que (2.3) se verifique.

Teorema 2.2.1 Para $\gamma \in \mathbb{R}$, as afirmações seguintes são equivalentes:

1. Existem constantes reais $a_n > 0$ e $b_n \in \mathbb{R}$ tal que

$$\lim_{n \rightarrow \infty} F^n(a_n x + b_n) = G_\gamma(x) = \exp(-(1 + \gamma x)^{-\frac{1}{\gamma}}), \quad (2.21)$$

para todo o x com $1 + \gamma x > 0$;

2. Existe uma função positiva $a(t)$, tal que, para $x > 0$,

$$\lim_{t \rightarrow \infty} \frac{U(tx) - U(t)}{a(t)} = D_\gamma(x) = \begin{cases} \frac{x^\gamma - 1}{\gamma}, & \gamma \neq 0 \\ \log x, & \gamma = 0. \end{cases} \quad (2.22)$$

3. Existe uma função positiva tal que

$$\lim_{t \rightarrow \infty} t(1 - F(a(t)x + U(t))) = (1 + \gamma x)^{-1/\gamma}, \quad (2.23)$$

para todo o x com $1 + \gamma x > 0$;

4. Existe uma função positiva f tal que

$$\lim_{t \uparrow x_F} \frac{1 - F(t + xf(t))}{1 - F(t)} = (1 + \gamma x)^{-1/\gamma}, \quad (2.24)$$

todo o x tal que $1 + \gamma x > 0$, onde $x_F = \sup \{x : F(x) < 1\}$.

Temos ainda que (2.21) é satisfeito para $b_n := U(n)$ e $a_n := a(n)$ e o resultado (2.24) é satisfeito com $f(t) = a(1/(1 - F(t)))$.

Não apresentamos a demonstração deste teorema uma vez que a equivalência entre 1. 2. e 3. já foi estabelecida no Teorema 2.1.1 e a prova de que 2. é equivalente a 4. usa os argumentos da demonstração do Teorema 2.2.2.

Exemplo 2.2.1 ([4]) Uma variável aleatória tem distribuição de Fréchet Negativa, com parâmetro $\beta > 0$, tem f.d. $F(x) = 1 - \exp(-(x_F - x)^{-\beta})$, $x \leq x_F$, $\beta > 0$ ($F(x) = 1$, $x > x_F$). A função U associada é

$$U(t) = F^{-1}(1 - 1/t) = x_F - (\log t)^{-1/\beta}, t \geq 1.$$

Então, considerando $a(t) = \frac{1}{\beta}(\log t)^{-1/\beta-1} \rightarrow 0$, $t \rightarrow \infty$, verifica-se (2.5) com $D(x) = \log x$.

Com efeito, como $(1+x)^\alpha = 1 + \alpha x + o(x)$, $x \rightarrow 0$, temos

$$\begin{aligned} \frac{U(tx) - U(t)}{a(t)} &= \frac{x_F - (\log t + \log x)^{-1/\beta} - x_F + (\log t)^{-1/\beta}}{\frac{1}{\beta}(\log t)^{-1/\beta-1}} \\ &= \frac{(\log t)^{-1/\beta}(1 - (1 + \frac{\log x}{\log t})^{-1/\beta})}{\frac{1}{\beta}(\log t)^{-1/\beta-1}} \\ &= \beta \frac{1 - (1 - \frac{1}{\beta} \frac{\log x}{\log t} + o(\frac{1}{\log t}))}{(\log t)^{-1}} \\ &= \log x + (\log t) o\left(\frac{1}{\log t}\right) \xrightarrow[t \rightarrow +\infty]{} \log x. \end{aligned} \quad (2.25)$$

Exemplo 2.2.2 ([4]) Seja $F(x) = 1 - \exp\{-\tan(x/\beta)\}$, $0 \leq x < \beta\pi/2 > 0$. Então, $U(t) = \beta \arctan(\log t)$, $t \geq 1$. Assim U satisfaz (2.5) com $D(x) = \log x$, com $a(t) = \beta/(\log^2 t + \beta^{-2})$ e $U(\infty) = \beta\pi/2 = x_F$.

De facto

$$\frac{U(tx) - U(t)}{a(t)} = \frac{\beta \arctan(\log tx) - \beta \arctan(\log t)}{\beta/(\log^2 t + \beta^{-2})}.$$

Usando a regra de L'Hospital

$$\begin{aligned}
 \frac{\frac{\beta_t^{\frac{1}{t}}}{1+\log^2(tx)} - \frac{\beta_t^{\frac{1}{t}}}{1+\log^2 t}}{\frac{-2\beta \log(t)^{\frac{1}{t}}}{(\log^2 t + \beta^{-2})^2}} &= \frac{1+\log^2 t - 1 - \log^2(tx)}{(1+\log^2 tx)(1+\log^2 t)} \frac{(\log^2 t + \beta^{-2})^2}{-2\log t} \\
 &= \frac{-2\log t \log x - \log^2 x}{1+\log^2 t + \log^2 tx + \log^2 tx \log^2 t} \frac{(\log^2 t + \beta^{-2})^2}{-2\log t} \\
 &= \frac{\log x (1 + \frac{\log^2 x}{2\log t})}{\log^4 t + o(\log^4 t)} (\log^4 t + o(\log^4 t)) \\
 &\rightarrow \log x, \quad t \rightarrow +\infty.
 \end{aligned}$$

No Teorema 2.2.2, são caracterizados os três domínios de atração da f.d. G_γ , correspondentes aos casos $\gamma = 0$, $\gamma > 0$ e $\gamma < 0$. Para a sua demonstração precisamos dos seguintes resultados.

Lema 2.2.1 *Suponhamos que (2.22) se verifica. Então são válidas as afirmações seguintes.*

1. Se $\gamma > 0$, então $\lim_{t \rightarrow \infty} U(t) = \infty$ e

$$\lim_{t \rightarrow \infty} \frac{U(t)}{a(t)} = \frac{1}{\gamma}. \quad (2.26)$$

2. Se $\gamma < 0$, então $\lim_{t \rightarrow \infty} U(t) < \infty$ e, com $U(\infty) := \lim_{t \rightarrow \infty} U(t)$,

$$\lim_{t \rightarrow \infty} \frac{U(\infty) - U(t)}{a(t)} = -\frac{1}{\gamma}. \quad (2.27)$$

Em particular, este limite implica que $\lim_{t \rightarrow \infty} a(t) = 0$.

3. Se $\gamma = 0$, então

$$\lim_{t \rightarrow \infty} \frac{U(tx)}{U(t)} = 1, \quad (2.28)$$

para todo $x > 0$ e $\lim_{t \rightarrow \infty} \frac{a(t)}{U(t)} = 0$. Temos ainda que, se $U(\infty) < \infty$

$$\lim_{t \rightarrow \infty} \frac{U(\infty) - U(tx)}{U(\infty) - U(t)} = 1, \quad (2.29)$$

para $x > 0$ e $\lim_{t \rightarrow \infty} \frac{a(t)}{U(\infty) - U(t)} = 0$. Além disso, para $x > 0$, temos

$$\lim_{t \rightarrow \infty} \frac{a(tx)}{a(t)} = 1. \quad (2.30)$$

Corolário 2.2.1 1. Para $\gamma > 0$, a relação (2.22) é equivalente a

$$\lim_{t \rightarrow \infty} \frac{U(tx)}{U(t)} = x^\gamma, \quad x > 0. \quad (2.31)$$

2. Para $\gamma < 0$ a relação (2.22) é equivalente a $U(\infty) < \infty$ e

$$\lim_{t \rightarrow \infty} \frac{U(\infty) - U(tx)}{U(\infty) - U(t)} = x^\gamma, \quad x > 0. \quad (2.32)$$

Lema 2.2.2 Suponhamos que para todo o $x > 0$, (2.22) é satisfeita com $\gamma = 0$, isto é,

$$\lim_{t \rightarrow \infty} \frac{U(tx) - U(t)}{a(t)} = \log(x), \quad (2.33)$$

onde $a(t)$ é uma função positiva adequada. Então, para todo o $\varepsilon > 0$, existem $c > 0$ e $t_0 > 1$ tais que para $x \geq 1$, $t \geq t_0$,

$$\frac{U(tx) - U(t)}{a(t)} \leq cx^\varepsilon. \quad (2.34)$$

Corolário 2.2.2 Se (2.22) é válido para $\gamma = 0$, então $\int_1^\infty \frac{U(s)}{s^2} ds < \infty$ e

$$\lim_{t \rightarrow 0} \frac{U^*(t) - U(t)}{a(t)} = 0 \quad (2.35)$$

com $U^*(t) := \frac{t}{e} \int_{\frac{t}{e}}^\infty U(s) \frac{ds}{s^2} = \int_1^\infty U\left(\frac{st}{e}\right) \frac{ds}{s^2}$. Note-se que U^* é contínua e estritamente crescente.

Corolário 2.2.3 Se $F \in D(G_0)$, então existe uma f.d. F_0 , contínua e estritamente crescente, tal que

$$\lim_{t \uparrow x_F} \frac{1 - F_0(t)}{1 - F(t)} = 1 \quad (2.36)$$

e (2.35) é válido.

Passamos então à caracterização dos três domínios de atração de G_γ .

Teorema 2.2.2 A f.d. F está no domínio de atração da distribuição de valor extremo G_γ , se e só se

1. Para $\gamma > 0$: $x_F = \sup\{x : F(x) < 1\}$ é infinito e

$$\lim_{t \rightarrow \infty} \frac{1 - F(tx)}{1 - F(t)} = x^{-1/\gamma}, \quad (2.37)$$

para todo $x > 0$.

2. Para $\gamma < 0$: x_F é finito

$$\lim_{t \downarrow 0} \frac{1 - F(x_F - tx)}{1 - F(x_F - t)} = x^{-1/\gamma}, \quad (2.38)$$

para todo $x > 0$.

3. Para $\gamma = 0$, x_F pode ser finito ou infinito e

$$\lim_{t \uparrow x_F} \frac{1 - F(t + xf(t))}{1 - F(t)} = e^{-x}, \quad (2.39)$$

para todo o $x \in \mathbb{R}$, onde f é uma função positiva adequada. Se (2.39) é satisfeita para alguma função f , então $\int_1^{x_F} (1 - F(s))ds < \infty$ para $t < x_F$ e (2.39) é satisfeita com

$$f(t) := \frac{\int_t^{x_F} (1 - F(s))ds}{1 - F(t)}. \quad (2.40)$$

Prova: Vamos começar por provar o teorema para $\gamma \neq 0$. A prova é análoga para $\gamma > 0$ e $\gamma < 0$. Assim, vamos apenas considerar $\gamma > 0$. Uma vez que U é uma função não decrescente, para $\varepsilon > 0$, temos

$$U\left(\frac{1 - \varepsilon}{1 - F(t)}\right) \leq U\left(\frac{1}{1 - F(t)}\right) \leq U\left(\frac{1 + \varepsilon}{1 - F(t)}\right) \quad (2.41)$$

Por outro lado, pela definição de U , vem

$$U\left(\frac{1}{1 - F(t)}\right) = \inf\left\{y : \frac{1}{1 - F(t)} \geq \frac{1}{1 - F(y)}\right\} = \inf\{y : F(y) \geq F(t)\} \leq t$$

e

$$U\left(\frac{1 + \varepsilon}{1 - F(t)}\right) = \inf\left\{y : \frac{1}{1 - F(y)} \geq \frac{1 + \varepsilon}{1 - F(t)}\right\} = \inf\left\{y : 1 - F(y) \leq \frac{1 - F(t)}{1 + \varepsilon}\right\} \geq t$$

donde (2.41) é equivalente a

$$U\left(\frac{1 - \varepsilon}{1 - F(t)}\right) \leq t \leq U\left(\frac{1 + \varepsilon}{1 - F(t)}\right).$$

Portanto,

$$\frac{U\left(\frac{x}{1 - F(t)}\right)}{U\left(\frac{1 + \varepsilon}{1 - F(t)}\right)} \leq \frac{U\left(\frac{x}{1 - F(t)}\right)}{t} \leq \frac{U\left(\frac{x}{1 - F(t)}\right)}{U\left(\frac{1 - \varepsilon}{1 - F(t)}\right)} \quad (2.42)$$

Supondo que (2.22) é válido, ou seja, $\lim_{t \rightarrow \infty} \frac{U(tx) - U(t)}{a(t)} = \frac{x^\gamma - 1}{\gamma}$, pelo Corolário 2.2.1, temos (2.31), ou seja, $\lim_{t \rightarrow \infty} \frac{U(tx)}{U(t)} = x^\gamma$, $x > 0$.

Então,

$$\lim_{t \rightarrow \infty} \frac{U\left(\frac{x}{1 - F(t)}\right)}{U\left(\frac{1 - \varepsilon}{1 - F(t)}\right)} = \lim_{t \rightarrow \infty} \frac{U\left(\frac{x(1 - \varepsilon)}{(1 - F(t))(1 - \varepsilon)}\right)}{U\left(\frac{1 - \varepsilon}{1 - F(t)}\right)} = \left(\frac{x}{1 - \varepsilon}\right)^\gamma$$

e

$$\lim_{t \rightarrow \infty} \frac{U\left(\frac{x}{1 - F(t)}\right)}{U\left(\frac{1 + \varepsilon}{1 - F(t)}\right)} = \lim_{t \rightarrow \infty} \frac{U\left(\frac{x(1 + \varepsilon)}{(1 - F(t))(1 + \varepsilon)}\right)}{U\left(\frac{1 + \varepsilon}{1 - F(t)}\right)} = \left(\frac{x}{1 + \varepsilon}\right)^\gamma.$$

Assim, os lados direito e esquerdo de (2.42) convergem para $(\frac{x}{1 + \varepsilon})^\gamma$ e $(\frac{x}{1 - \varepsilon})^\gamma$, respetivamente.

Posto isto, se a relação (2.42) é válida para todo o $\varepsilon > 0$, então,

$$\lim_{t \rightarrow \infty} \frac{1}{t} U\left(\frac{x}{1 - F(t)}\right) = x^\gamma. \quad (2.43)$$

Aplicando agora o Lema 2.1.1 na relação (2.43), temos

$$\lim_{t \rightarrow \infty} \left(\frac{1}{t} U \left(\frac{x}{1-F(t)} \right) \right)^{\leftarrow} = x^{1/\gamma}.$$

Pela definição de inversa generalizada U , tem-se que

$$U \left(\frac{y}{1-F(t)} \right) = \inf \left\{ x : \frac{1}{1-F(x)} \geq \frac{y}{1-F(t)} \right\} = \inf \left\{ x : \frac{1-F(t)}{1-F(x)} \geq y \right\} = \left(\frac{1-F(t)}{1-F(\cdot)} \right)^{\leftarrow} (y),$$

bem como

$$U^{\leftarrow}(y) = \left(\left(\frac{1}{1-F(\cdot)} \right)^{\leftarrow} \right)^{\leftarrow} (y) = \frac{1}{1-F^-(y)}.$$

Assim, uma vez que U é monótona não decrescente, obtemos

$$\begin{aligned} \left(\frac{1}{t} U \left(\frac{x}{1-F(t)} \right) \right)^{\leftarrow} &= \inf \left\{ y : \frac{1}{t} U \left(\frac{y}{1-F(t)} \right) \geq x \right\} = \inf \left\{ y : U \left(\frac{y}{1-F(t)} \right) \geq xt \right\} \\ &= \inf \left\{ z(1-F(t)) : U(z) \geq xt \right\} = (1-F(t)) \inf \left\{ z : U(z) \geq xt \right\} \\ &= (1-F(t)) U^{\leftarrow}(tx) = \frac{1-F(t)}{1-F^-(tx)}. \end{aligned}$$

Então, provámos que $\lim_{t \rightarrow \infty} \frac{1-F(t)}{1-F^-(tx)} = x^{1/\gamma}$, $x > 0$. Pelo Lema 2.1.2 concluímos que $\lim_{t \rightarrow \infty} \frac{1-F(t)}{1-F(tx)} = x^{1/\gamma}$, isto é, $\lim_{t \rightarrow \infty} \frac{1-F(tx)}{1-F(t)} = x^{-1/\gamma}$.

Está assim provado o Teorema 2.2.2 para $\gamma \neq 0$, mais concretamente para $\gamma > 0$.

Passamos a provar o Teorema 2.2.2 para $\gamma = 0$.

Começemos por observar que o limite (2.39) toma a forma de (2.24), para $\gamma = 0$, o qual, pelo Teorema 2.2.1, é equivalente a (2.21). Assim, provámos que (2.39) é necessário e suficiente para que $F \in G_0$.

Uma vez que F é monotóna e e^{-x} é contínua, então o limite (2.39) é localmente uniforme. Portanto, para provar que (2.39) é válido com a função f de (2.40), é suficiente provar que se (2.39) é válido, então

$$f(t) \sim \frac{\int_t^{x_F} (1-F(s)) ds}{1-F(t)}. \quad (2.44)$$

Consideremos as funções U_0 e F_0 dos corolários 2.2.2 e 2.2.3. Pelo Corolário 2.2.3, temos que $\lim_{t \uparrow x_F} \frac{1-F_0(t+xf(t))}{1-F(t+xf(t))} = 1$ e $\lim_{t \uparrow x_F} \frac{1-F(t)}{1-F_0(t)} = 1$. Atendendo a que

$$\frac{1-F_0(t+xf(t))}{1-F_0(t)} = \frac{(1-F_0(t+xf(t))) (1-F(t+xf(t)))}{1-F(t+xf(t))} \frac{1-F(t)}{1-F_0(t)},$$

provamos que o limite (2.39) é válido com F substituído por F_0 , isto é,

$$\lim_{t \uparrow x_F} \frac{1-F_0(t+xf(t))}{1-F_0(t)} = e^{-x}.$$

Ora, usando agora a regra de L'Hôpital, obtemos

$$\frac{\left(-\int_{x_F}^t (1-F_0(s))ds\right)'}{\left(-\int_{x_F}^t (1-F(s))ds\right)'} = \frac{1-F_0(t)}{1-F(t)} \rightarrow 1, t \uparrow x_F,$$

do que resulta

$$\int_t^{x_F} (1-F_0(s))ds \sim \int_t^{x_F} (1-F(s))ds, t \uparrow x_F.$$

Então, basta provar o resultado para F_0 . Sejam $A_0(t) = \int_t^{x_F} (1-F_0(s))ds$ e $A(t) = \int_t^{x_F} (1-F(s))ds$.

Note-se que,

$$\lim_{t \uparrow x_F} \frac{A_0(t)}{1-F_0(t)} = \lim_{t \uparrow x_F} \frac{A_0(t)}{A(t)} \frac{A(t)}{1-F(t)} \frac{1-F(t)}{1-F_0(t)},$$

em que, pelo Corolário 2.2.3, $\lim_{t \uparrow x_F} \frac{A_0(t)}{A(t)} = 1$ e $\lim_{t \uparrow x_F} \frac{1-F(t)}{1-F_0(t)} = 1$. Então, temos

$$\frac{\int_t^{x_F} (1-F_0(s))ds}{1-F_0(t)} \sim \frac{\int_t^{x_F} (1-F(s))ds}{1-F(t)}, t \uparrow x_F,$$

e assim é suficiente provar a relação (2.43) para F_0 em vez de F , uma vez que passamos as propriedades de F para F_0 .

Seja $U_0 = \left(\frac{1}{1-F_0}\right)^{\leftarrow}$. Fazendo a mudança de variável $s = zx$, temos

$$\begin{aligned} \lim_{z \rightarrow +\infty} \frac{z \int_1^{\infty} U_0(s) \frac{ds}{s^2} - U_0(z)}{a(z)} &= \lim_{z \rightarrow +\infty} \frac{z \int_1^{\infty} U_0(zx) \frac{zdx}{x^2 z^2} - U_0(z)}{a(z)} = \lim_{z \rightarrow +\infty} \frac{\int_1^{\infty} U_0(zx) \frac{1}{x^2} dx - U_0(z)}{a(z)} = \\ &= \lim_{z \rightarrow +\infty} \frac{1}{a(z)} \int_1^{\infty} U_0(zx) \frac{1}{x^2} dx - \frac{U_0(z)}{a(z)} \int_1^{\infty} \frac{1}{x^2} dx = \lim_{z \rightarrow +\infty} \frac{1}{a(z)} \int_1^{\infty} \frac{(U_0(zx) - U_0(z))}{x^2} dx, \end{aligned}$$

onde se usa $\int_1^{\infty} \frac{1}{x^2} dx = 1$.

Usando o Lema 2.2.2 e o Teorema da convergência dominada ([2], pag.213) obtemos

$$\lim_{z \rightarrow \infty} \int_1^{\infty} \frac{U_0(zx) - U_0(z)}{a(z)x^2} dx = \lim_{z \rightarrow \infty} \int_1^{\infty} \log(x) \frac{1}{x^2} dx = 1, \quad (2.45)$$

bem como

$$\lim_{z \rightarrow +\infty} \frac{z \int_1^{\infty} U_0(s) \frac{ds}{s^2} - U_0(z)}{a(z)} = 1.$$

Agora se substituirmos s por $\frac{1}{1-F_0(u)}$, ou seja, $u = U_0(s)$ (desde que F_0 seja contínua e estritamente crescente) e $z = \frac{1}{1-F_0(t)}$, então temos que

$$\begin{cases} U_0(z) = t \\ \frac{ds}{du} = F_0'(u)(1-F_0(u))^{-2} \\ u = t \implies s = z \\ u = x_F \implies s = +\infty \end{cases}$$

pois $\frac{1}{1-F_0(u)} \rightarrow \infty$ quando $F_0(u) \rightarrow 1$, o que acontece quando $x_F = u$. Assim,

$$\begin{aligned} \lim_{z \rightarrow \infty} \frac{z \int_z^\infty U_0(s) \frac{ds}{s^2} - U_0(z)}{a(z)} &= \lim_{t \uparrow x_F} \frac{\frac{1}{1-F_0(t)} \int_t^{x_F} u F_0'(u) (1-F_0(u))^{-2} \frac{du}{(1-F_0(u))^{-2}} - t}{a((1-F_0(t))^{-1})} \\ &= \lim_{t \uparrow x_F} \left(\frac{\int_t^{x_F} u F_0'(u) du}{1-F_0(t)} - t \right) \times \frac{1}{a((1-F_0(t))^{-1})}. \end{aligned} \quad (2.46)$$

Por outro lado, como $F_0(u) - 1$ é uma primitiva de F_0' , temos também

$$\int_t^{x_F} u F_0'(u) du = [(F_0(u) - 1)u]_t^{x_F} - \int_t^{x_F} (F_0(u) - 1) du = (1 - F_0(t))t + \int_t^{x_F} (1 - F_0(u)) du,$$

e assim,

$$\lim_{t \uparrow x_F} \frac{\frac{\int_t^{x_F} u F_0'(u) du}{1-F_0(t)} - t}{a((1-F_0(t))^{-1})} = \lim_{t \uparrow x_F} \frac{\frac{(1-F_0(t))t + \int_t^{x_F} (1-F_0(u)) du}{1-F_0(t)} - t}{a((1-F_0(t))^{-1})} = \lim_{t \uparrow x_F} \frac{\int_t^{x_F} (1-F_0(u)) du}{(1-F_0(t))a(\frac{1}{1-F_0(t)})}.$$

Posto isto, uma vez que se tem (2.45) e (2.46) chegamos a

$$\lim_{t \uparrow x_F} \frac{\int_t^{x_F} (1-F_0(u)) du}{(1-F_0(t))a(\frac{1}{1-F_0(t)})} = 1.$$

Assim, (2.44) é obtido pelo Teorema 2.2.1. O Teorema 2.2.2 para $\gamma = 0$ está provado. \square

Capítulo 3

Estimação do índice de cauda e escolha de domínios de atração

3.1 Estimadores do índice de cauda

3.1.1 Estimador de Hill

Seja (X_1, X_2, \dots, X_n) uma amostra aleatória de X e sejam $X_{1:n}, X_{2:n}, \dots, X_{n:n}$ as estatísticas de ordem associadas a esta amostra. O estimador de Hill é definido por

$$\hat{\gamma}_H := \frac{1}{k} \sum_{i=0}^{k-1} \log X_{n-i,n} - \log X_{n-k,n},$$

com $k \in \{0, \dots, n-1\}$.

Apresentamos de seguida um lema que contém resultados fundamentais às demonstrações que se seguem.

Lema 3.1.1 *Seja X uma v.a. com f.d. F . Seja (X_1, X_2, \dots, X_n) uma amostra aleatória de X e sejam $X_{1:n}, X_{2:n}, \dots, X_{n:n}$ as estatísticas de ordem associadas a esta amostra. Seja (Y_1, Y_2, \dots, Y_n) uma amostra aleatória da v.a. Y com f.d. de Pareto definida por $H(y) = (1 - \frac{1}{y})1_{[1, +\infty[}(y)$ e $Y_{1:n}, Y_{2:n}, \dots, Y_{n:n}$ as estatísticas de ordem n associadas. Então, com $k = k(n) \rightarrow \infty$, $\frac{k}{n} \rightarrow 0$, $n \rightarrow \infty$, temos*

1. $X_{n-k,n} \rightarrow x_F, n \rightarrow +\infty (q.c.)$;
2. $Y_{n-k,n} \rightarrow +\infty, n \rightarrow +\infty (q.c.)$;
3. $\frac{1}{k} \sum_{i=0}^{k-1} \log \left(\frac{Y_{n-i,n}}{Y_{n-k,n}} \right) \xrightarrow{P} 1, n \rightarrow +\infty$;
4. $(U(Y_{1,n}), \dots, U(Y_{n,n})) \stackrel{d}{=} (X_{1,n}, \dots, X_{n,n})$.

O teorema seguinte estabelece a consistência do estimador de Hill.

Teorema 3.1.1 *Sejam X_1, X_2, \dots, X_n variáveis aleatórias i.i.d. com f.d. F . Suponhamos que $F \in \mathcal{D}(G_\gamma), \gamma > 0$. Assim, com $k = k(n) \rightarrow \infty$, $\frac{k}{n} \rightarrow 0$, $n \rightarrow \infty$, então $\hat{\gamma}_H \xrightarrow{P} \gamma, n \rightarrow \infty$.*

Prova: Pelo Corolário 2.2.1, $F \in \mathcal{D}(G_\gamma)$, com $\gamma > 0$ implica que,

$$\lim_{t \rightarrow \infty} \frac{U(tx)}{U(t)} = x^\gamma, x > 0.$$

Então, para $x \geq 1$, $t \geq t_0$, temos $(1 - \varepsilon)x^{\gamma - \varepsilon'} < \frac{U(tx)}{U(t)} < (1 + \varepsilon)x^{\gamma + \varepsilon'}$, para quaisquer ε e ε' positivos.

Usando logaritmos, obtemos o seguinte

$$\log(1 - \varepsilon) + (\gamma - \varepsilon') \log x < \log U(tx) - \log U(t) < \log(1 + \varepsilon) + (\gamma + \varepsilon') \log x. \quad (3.1)$$

Sejam Y_1, Y_2, \dots, Y_n variáveis aleatórias i.i.d. com f.d. $F(y) = (1 - \frac{1}{y})1_{[1, +\infty[}(y)$, $y \in \mathbb{R}$. Sejam $Y_{1,n}, Y_{2,n}, \dots, Y_{n,n}$ as estatísticas de ordem n associadas. Então, com $k = k(n)$ nas condições do enunciado, temos pelo Lema 3.1.1 $\lim_{n \rightarrow \infty} Y_{n-k,n} = \infty$, q.c.. Atendendo ao ponto 4 do Lema 3.1.1 é suficiente provar o resultado para

$$\hat{\gamma}_H := \frac{1}{k} \sum_{i=0}^{k-1} \log U(Y_{n-i,n}) - \log U(Y_{n-k,n}).$$

Consideremos $t = Y_{n-k,n}$ e $x = \frac{Y_{n-i,n}}{Y_{n-k,n}}$ para qualquer i fixado em $\{0, \dots, k-1\}$. Uma vez que $Y_{n-k,n} \rightarrow \infty$, $n \rightarrow \infty$ (q.c), temos a partir de (3.1)

$$\log(1 - \varepsilon) + (\gamma - \varepsilon') \log \left(\frac{Y_{n-i,n}}{Y_{n-k,n}} \right) < \log U(Y_{n-i,n}) - \log U(Y_{n-k,n}) < \log(1 + \varepsilon) + (\gamma + \varepsilon') \log \left(\frac{Y_{n-i,n}}{Y_{n-k,n}} \right),$$

para $i = 0, \dots, k-1$. Então,

$$\log(1 - \varepsilon) + (\gamma - \varepsilon') \frac{1}{k} \sum_{i=0}^k \log \left(\frac{Y_{n-i,n}}{Y_{n-k,n}} \right) < \hat{\gamma}_H < \log(1 + \varepsilon) + (\gamma + \varepsilon') \frac{1}{k} \sum_{i=0}^k \log \left(\frac{Y_{n-i,n}}{Y_{n-k,n}} \right).$$

Uma vez que pelo Lema 3.1.1 se tem

$$\frac{1}{k} \sum_{i=0}^k \log \left(\frac{Y_{n-i,n}}{Y_{n-k,n}} \right) \xrightarrow{P} 1, n \rightarrow \infty,$$

e considerando ε e ε' arbitrariamente pequenos, a prova fica concluída. \square

3.1.2 Estimador de Pickands

O estimador mais simples para o parâmetro γ é o estimador de Pickands, definido por

$$\hat{\gamma}_P = \frac{1}{\log 2} \log \frac{X_{n-k,n} - X_{n-2k,n}}{X_{n-2k,n} - X_{n-4k,n}},$$

com $1 \leq k \leq \frac{n-1}{4}$.

Começamos por apresentar um lema fundamental às demonstrações que se seguem.

Lema 3.1.2 *Sejam Y_1, Y_2, \dots, Y_n variáveis aleatórias i.i.d. com f.d. $F(x) = (1 - \frac{1}{x})1_{[1, +\infty[}(x)$, $x \in \mathbb{R}$. Assim, com $n \rightarrow \infty$, $k = k(n) \rightarrow \infty$, $\frac{k}{n} \rightarrow 0$, a sucessão de vetores*

$$\sqrt{2k} \left(\frac{1}{2} \frac{Y_{n-k,n}}{Y_{n-2k,n}} - 1, \frac{\sqrt{2}}{2} \frac{Y_{n-2k,n}}{Y_{n-4k,n}} - \sqrt{2} \right) \quad (3.2)$$

converge em distribuição para um vetor aleatório normal bivariado com margens centradas e reduzidas.

Teorema 3.1.2 *Seja X_1, X_2, \dots, X_n variáveis aleatórias i.i.d. com f.d. F . Suponhamos $F \in \mathcal{D}(G_\gamma)$, com $\gamma \in \mathbb{R}$. Então, com $k = k(n) \rightarrow \infty$, $\frac{k}{n} \rightarrow 0$, $n \rightarrow \infty$, temos*

$$\hat{\gamma}_P \xrightarrow{P} \gamma, n \rightarrow +\infty.$$

Prova: Seja $\gamma \neq 0$. Recordemos que $F \in \mathcal{D}(G_\gamma)$ é equivalente a

$$\lim_{t \rightarrow \infty} \frac{U(tx) - U(t)}{a(t)} = \frac{x^\gamma - 1}{\gamma}, x > 0, \quad (\log x, \text{ se } \gamma = 0),$$

onde a convergência é localmente uniforme, uma vez que U é monótona. Em consequência, para $0 < x, y < +\infty$, tem-se a relação

$$\lim_{t \rightarrow \infty} \frac{U(tx) - U(t)}{U(ty) - U(t)} = \frac{x^\gamma - 1}{y^\gamma - 1}. \quad (3.3)$$

Sejam $Y_{1,n}, \dots, Y_{n,n}$ como definidas na demonstração do Teorema 3.1.1, isto é, com os Y_i 's variáveis aleatórias i.i.d com f.d. $F(x) = (1 - \frac{1}{x})1_{[1, +\infty[}(x)$.

Sejam $t = Y_{n-4k,n}$, $x = \frac{Y_{n-k,n}}{Y_{n-4k,n}}$ e $y = \frac{Y_{n-2k,n}}{Y_{n-4k,n}}$. Observemos que

$$\begin{aligned} \frac{X_{n-k,n} - X_{n-2k,n}}{X_{n-2k,n} - X_{n-4k,n}} &= \frac{X_{n-k,n} - X_{n-2k,n}}{X_{n-2k,n} - X_{n-4k,n}} + \frac{X_{n-4k,n}}{X_{n-2k,n} - X_{n-4k,n}} - \frac{X_{n-4k,n}}{X_{n-2k,n} - X_{n-4k,n}} = \\ &= \frac{X_{n-k,n} - X_{n-4k,n}}{X_{n-2k,n} - X_{n-4k,n}} - \left(\frac{X_{n-2k,n} - X_{n-4k,n}}{X_{n-2k,n} - X_{n-4k,n}} \right) = \frac{X_{n-k,n} - X_{n-4k,n}}{X_{n-2k,n} - X_{n-4k,n}} - 1 = \\ &\stackrel{d}{=} \frac{U(Y_{n-k,n}) - U(Y_{n-4k,n})}{U(Y_{n-2k,n}) - U(Y_{n-4k,n})} - 1 = \frac{U(\frac{Y_{n-k,n}}{Y_{n-4k,n}} Y_{n-4k,n}) - U(Y_{n-4k,n})}{U(\frac{Y_{n-2k,n}}{Y_{n-4k,n}} Y_{n-4k,n}) - U(Y_{n-4k,n})} - 1, \end{aligned}$$

onde se usou o ponto 4 do Lema 3.1.1. Pelo Lema 3.1.2, como $Y_{n-4k,n} \rightarrow +\infty$, $n \rightarrow +\infty$ temos que

$$\frac{Y_{n-k,n}}{Y_{n-4k,n}} \xrightarrow{P} 4, n \rightarrow +\infty, \quad \frac{Y_{n-2k,n}}{Y_{n-4k,n}} \xrightarrow{P} 2, n \rightarrow +\infty. \quad (3.4)$$

Assim, por (3.3) e (3.4) vem que

$$\frac{X_{n-k,n} - X_{n-2k,n}}{X_{n-2k,n} - X_{n-4k,n}} \stackrel{d}{=} \frac{U\left(\frac{Y_{n-k,n}}{Y_{n-4k,n}} Y_{n-4k,n}\right) - U(Y_{n-4k,n})}{U\left(\frac{Y_{n-2k,n}}{Y_{n-4k,n}} Y_{n-4k,n}\right) - U(Y_{n-4k,n})} - 1 \stackrel{P}{\rightarrow} \frac{4^\gamma - 1}{2^\gamma - 1} - 1 = 2^\gamma.$$

Portanto, $\hat{\gamma}_P = \frac{1}{\log 2} \log \frac{X_{n-k,n} - X_{n-2k,n}}{X_{n-2k,n} - X_{n-4k,n}} \stackrel{P}{\rightarrow} \frac{1}{\log 2} \log 2^\gamma = \frac{1}{\log 2} \gamma \log 2 = \gamma.$

Se $\gamma = 0$, o limite em (3.3) é 1 e portanto $\hat{\gamma}_P \stackrel{P}{\rightarrow} 0$. Então, está provado o resultado. \square

3.1.3 Estimador dos Momentos

Queremos desenvolver um estimador semelhante ao estimador de Hill, mas que pode ser usado para $\gamma \in \mathbb{R}$ geral, e não apenas para $\gamma > 0$. Olhe-se então para o comportamento do estimador de Hill para um γ geral.

Lema 3.1.3 *Sejam X_1, X_2, \dots, X_n variáveis aleatórias com f.d. F e suponhamos que $F \in \mathcal{D}(G_\gamma)$, $x^* = U(\infty)$. Definimos para $j = 1, 2$,*

$$M_n^{(j)} := \frac{1}{k} \sum_{i=0}^{k-1} (\log X_{n-i,n} - \log X_{n-k,n})^j. \quad (3.5)$$

Então para $k = k(n) \rightarrow \infty$, $\frac{k}{n} \rightarrow 0$, $n \rightarrow \infty$, tem-se

$$\frac{M_n^{(j)}}{\left(a(\frac{n}{k})/U(\frac{n}{k})\right)^j} \stackrel{P}{\rightarrow} \prod_{i=1}^j \frac{i}{1-i\gamma_-}, \quad \gamma_- = \min(0, \gamma). \quad (3.6)$$

\square

Para $\gamma \leq 0$, por (2.27) concluímos $a(n/k)/U(n/k) \rightarrow 0$, $n \rightarrow +\infty$ e assim, como (3.6) implica $\frac{M_n^{(1)}}{a_n} \stackrel{P}{\rightarrow} \frac{1}{1-\gamma_-} = \frac{1}{1-\gamma_-}$, concluímos que o estimador de Hill, $M_n^{(1)}$, converge em probabilidade para 0, para $\gamma \leq 0$. Então este estimador é não informativo neste intervalo.

Note-se que, por (3.6) obtemos

$$\frac{\left(M_n^{(1)}\right)^2}{M_n^{(2)}} = \frac{\left(\frac{M_n^{(1)}}{a(n/k)/U(n/k)}\right)^2}{\frac{M_n^{(2)}}{\left(a(n/k)/U(n/k)\right)^2}} \stackrel{P}{\rightarrow} \frac{\left(\frac{1}{1-\gamma_-}\right)^2}{\left(\frac{1}{1-\gamma_-}\right)\left(\frac{2}{1-2\gamma_-}\right)} = \frac{1}{1-\gamma_-} = \frac{1-2\gamma_-}{2(1-\gamma_-)}. \quad (3.7)$$

Assim, sob estas condições, o Lema 3.1.3 permite encontrar um estimador de γ , para $\gamma < 0$. Pelo Teorema 3.1.1, sabemos que $\hat{\gamma}_H \stackrel{P}{\rightarrow} \gamma_+$, onde $\gamma_+ = \max\{0, \gamma\}$.

Isto leva à seguinte combinação de resultados do estimador de Hill e da estatística em (3.7), a que chamamos estimador dos momentos, definido por

$$\begin{aligned}\hat{\gamma}_M &:= M_n^{(1)} + 1 - \frac{1}{2} \left(1 - \frac{(M_n^{(1)})^2}{M_n^{(2)}} \right)^{-1} \xrightarrow{P} \gamma_+ + 1 - \frac{1}{2} \left(1 - \left(\frac{1 - 2\gamma_-}{2(1 - \gamma_-)} \right) \right)^{-1} = \\ &= \gamma_+ + 1 - \frac{1}{2} \left(\frac{1}{2(1 - \gamma_-)} \right)^{-1} = \gamma_+ + \gamma_- = \gamma.\end{aligned}$$

Acabámos então de provar a consistência de $\hat{\gamma}_M$ como estimador de $\gamma \in \mathbb{R}$ e a consistência de

$$\hat{\gamma}_- = 1 - \frac{1}{2} \left(1 - \frac{(M_n^{(1)})^2}{M_n^{(2)}} \right)^{-1}$$

como estimador de γ_- . Estes resultados apresentam-se no teorema seguinte.

Teorema 3.1.3 *Sejam X_1, X_2, \dots, X_n variáveis aleatórias com f.d. F e suponhamos que $F \in \mathcal{D}(G_\gamma)$, $x^* > 0$. Então, com $k = k(n) \rightarrow \infty$, $\frac{k}{n} \rightarrow 0$, $n \rightarrow \infty$, tem-se*

1. $\hat{\gamma}_M \xrightarrow{P} \gamma$, para $\gamma \in \mathbb{R}$,
2. $\hat{\gamma}_- \xrightarrow{P} \gamma_-$, para $\gamma < 0$.

3.1.4 Estimador da mistura de momentos

Em Fraga Alves et al (2009)[7] é introduzido o estimador de γ dado por

$$\hat{\gamma}_{MM}(k) = \frac{\hat{\gamma}_n(k) - 1}{1 + 2 \min(\hat{\gamma}_n(k) - 1, 0)}$$

onde

$$\hat{\gamma}_n(k) = \frac{M_n^{(1)}(k) - \lambda_n(k)}{(\lambda_n(k))^2}$$

com

$$\lambda_n(k) = \frac{1}{k} \sum_{i=1}^n \left(1 - \frac{X_{n-k,n}}{X_{n-i+1,n}} \right).$$

Aqueles autores provam que se $x_F = U(+\infty)$ é positivo (finito ou infinito) e F satisfaz a condição de primeira ordem (2.20) para algum $\gamma \in \mathbb{R}$, então

$$\hat{\gamma}_{MM}(k) \xrightarrow{P} \gamma, n \rightarrow +\infty,$$

onde $k = k(n)$ é tal que $\frac{k}{n} \rightarrow 0$, $n \rightarrow +\infty$ e $k_n \rightarrow +\infty$, $n \rightarrow +\infty$.

Omitimos a prova deste resultado.

3.2 Finitude e estimação de x_F

Quando $\gamma > 0$, isto é, F pertence ao domínio de atração de Fréchet, então x_F é infinito. No caso $\gamma < 0$, tem-se sempre x_F finito. Quando F pertence ao domínio de atração da f.d. de Gumbel podem ocorrer às duas situações $x_F = +\infty$ ou x_F finito como vimos nos exemplos 2.1.1 e 2.2.1, respectivamente.

Muitos dos resultados que incluímos a partir de agora necessitam da chamada condição de segunda ordem.

Uma vez que a condição de primeira ordem (2.22) é equivalente a $F^n(a_n x + b_n) \rightarrow G_\gamma(x)$, $n \rightarrow +\infty$, (Teorema 2.2.1), a velocidade de convergência da sucessão de máximos para o seu limite em distribuição, G_γ , é controlada por uma condição de segunda ordem. Em particular, assumimos que existe uma função positiva ou negativa A_0 com $\lim_{t \rightarrow \infty} A_0(t) = 0$ tal que para cada $x > 0$,

$$\lim_{t \rightarrow \infty} \frac{\frac{U(tx) - U(t)}{a_0(t)} - \frac{x^\gamma - 1}{\gamma}}{A_0(t)} = \psi_{\gamma, \rho}(x), \quad (3.8)$$

onde ρ é um parametro não positivo e com

$$\psi_{\gamma, \rho}(x) := \begin{cases} \frac{x^{\gamma+\rho} - 1}{\gamma + \rho}, & \gamma + \rho \neq 0, \rho < 0, \\ \log x, & \gamma + \rho = 0, \rho < 0, \\ \frac{1}{\gamma} x^\gamma \log x, & \rho = 0 \neq \gamma, \\ \frac{1}{2} (\log x)^2, & \gamma = \rho = 0, \end{cases}$$

$$a_0(t) := \begin{cases} a(t)(1 - A_0(t)), & \rho < 0, \\ a(t)(1 - A_0(t)/\gamma), & \rho = 0 \neq \gamma, \\ a(t), & \gamma = \rho = 0. \end{cases}$$

Além disso, A_0 verifica $\lim_{t \rightarrow +\infty} \frac{A_0(tx)}{A_0(t)} = x^\rho$ e $\lim_{t \rightarrow \infty} \frac{\frac{a_0(tx) - x^\gamma}{a_0(t)} - x^\gamma}{A_0(t)} = x^\gamma \frac{x^\rho - 1}{\rho}$, para todo o $x > 0$.

Denota-se $U(\infty) := \lim_{t \rightarrow \infty} U(t) = x_F$. Se a condição de segunda ordem (3.8) é satisfeita com $\gamma < 0$ então, sendo $x = x(t) \rightarrow \infty$,

$$\lim_{t \rightarrow \infty} \frac{\frac{U(\infty) - U(t)}{a_0(t)} + \frac{1}{\gamma}}{A_0(t)} = \psi_{\gamma, \rho} := -\frac{1}{\gamma + \rho} 1_{\{\rho < 0\}} \quad (3.9)$$

Citando [11] incluímos alguns exemplos de f.d.'s que verificam esta condição de segunda ordem. A f.d. de Cauchy $F(x) = \frac{1}{2} + \frac{1}{\pi} \arctan x$, $x \in \mathbb{R}$, apresenta $(\gamma, \rho) = (1, -2)$ e a f.d. logística $F(x) = 1 - \frac{2}{1+e^x}$, $x > 0$, admite $(\gamma, \rho) = (0, -1)$. Os mesmos valores para a f.d. de Gumbel. Também a f.d. normal verifica (3.8) com $\gamma = \rho = 0$ e a f.d. $\mathcal{U}(0, 1)$ com $\gamma = \rho = -1$.

Com o objetivo de averiguar se x_F é finito ou infinito, em [12] é introduzido o teste de hipóteses

$$H_0 : F \in \mathcal{D}(G_0), x_F = +\infty, \quad \text{vs} \quad H_1 : F \in \mathcal{D}(G_\gamma)_{\gamma \leq 0}, x_F < +\infty,$$

com estatística de teste

$$E_{n,k} = \sqrt{k} \log k \frac{1}{k} \sum_{i=1}^k \frac{X_{n-i,n} - X_{n-k,n} - \hat{a}(n/k)}{X_{n,i,n} - X_{n-k,n}}.$$

Em [12] prova-se que se F satisfaz as condições de primeira e de segunda ordem com $\sqrt{k}A_0(n/k) \rightarrow 0$, $n \rightarrow +\infty$, então, admitindo H_0 verdadeira ($F \in \mathcal{D}(G_0)$ e $x_F = +\infty$) tem-se

$$E_{n,k} \rightarrow Z, n \rightarrow +\infty,$$

onde $Z \sim N(0, 1)$. Assim, observações que levam à rejeição de H_0 são tais que $|E_{n,k}| > z_{1-\frac{\alpha}{2}}$.

Perante a rejeição de H_0 , há que estimar x_F .

Para f.d.'s no domínio de atração da f.d. de Weibull, isto é com $\gamma < 0$ e $U(\infty) = x_F < +\infty$, o limite (2.27) estabelece que

$$\lim_{t \rightarrow +\infty} \frac{x_F - U(t)}{a(t)} = -\frac{1}{\gamma}.$$

Este limite motiva o estimador de x_F dado por $\hat{x}_F = \hat{U}(n/k) - \frac{1}{\hat{\gamma}}\hat{a}(n/k)$ ou seja

$$\hat{x}_F = X_{n-k,n} - \frac{1}{\hat{\gamma}}\hat{a}(n/k) \quad (3.10)$$

para qualquer estimador de $\gamma < 0$. Notamos que para $\gamma > 0$ se tem $x_F = +\infty$. O estimador de $a(n/k)$ introduzido Fraga Alves et al (2009)[7], definido por

$$\hat{a}(n/k) = X_{n-k,n} \frac{M_n^{(1)}}{2} \left(1 - \frac{(M_n^{(1)})^2}{M_n^{(2)}} \right)$$

permite assim estimar x_F usando (3.10). Assim, diferentes escolhas de estimadores de $\gamma < 0$ geram diferentes estimadores de x_F a partir de (3.10).

O caso em que x_F é finito e $\gamma = 0$ é considerado a seguir.

Em Fraga Alves [4](2014) é introduzido um estimador de x_F , para o caso em que x_F é finito e $\gamma = 0$, isto é $F \in \mathcal{D}(G_0)$. Este estimador não depende da estimação prévia de γ . Nos exemplos 2.2.2 e 2.1.3 encontramos f.d's deste tipo. As autoras mostram que a condição de primeira ordem

$$\lim_{t \rightarrow +\infty} \frac{U(tx) - U(t)}{a_0(t)} = \log x, x > 0, \quad (3.11)$$

é equivalente a $U(t) = U(t_0) + \int_{t_0}^t a(s) \frac{1}{s} ds + O(a(t))$, $t \rightarrow +\infty$, onde a função $a(t)$ verifica $a(t) \sim a_0(t)$, $t \rightarrow +\infty$, bem como $\lim_{t \rightarrow +\infty} \frac{a(tx)}{a(t)} = 1$. Desta igualdade resulta

$$U(\infty) = U(t) + \int_t^{+\infty} a(s) \frac{1}{s} ds + O(a(t)), t \rightarrow +\infty. \quad (3.12)$$

Por outro lado, a regra de L'Hospital e o limite (3.11) permitem concluir que, para $x > 0$, se tem

$$\lim_{t \rightarrow +\infty} \frac{\int_{tx}^{+\infty} U(s) \frac{1}{s} ds - \int_t^{+\infty} U(s) \frac{1}{s} ds}{\int_t^{+\infty} a(s) \frac{1}{s} ds} = \log x.$$

Seja $q(t) = \int_t^{+\infty} a(s) \frac{1}{s} ds$. Uma vez que $\int_{tx}^{+\infty} U(s) \frac{1}{s} ds = \int_0^1 U(tx/s) \frac{1}{s} ds$ e $\int_t^{+\infty} U(s) \frac{1}{s} ds = \int_0^1 U(t/s) \frac{1}{s} ds$ e o mesmo para $q(t) = \int_0^1 a(t/s) \frac{1}{s} ds$, obtemos

$$\lim_{t \rightarrow +\infty} \frac{\int_0^1 (U(tx/s) - U(t/s)) \frac{1}{s} ds}{q(t)} = \log x.$$

Se no limite anterior considerarmos $x = \frac{1}{2}$ e $t = \frac{n}{k}$, obtemos

$$\lim_{t \rightarrow +\infty} \frac{\int_0^1 (U(n/2ks) - U(n/ks)) \frac{1}{s} ds}{q(n/k)} = -\log 2. \quad (3.13)$$

Atendendo a que $U(n/k) = F^{\leftarrow}(1 - \frac{k}{n})$, temos $\hat{U}(n/k) = X_{n-k,n}$. Assim, o limite (3.13), de acordo com Fraga Alves [4], dá origem ao estimador

$$\hat{q}(n/k) = -\frac{1}{\log 2} \int_0^1 (X_{n-[2ks],n} - X_{n-[ks],n}) \frac{1}{s} ds.$$

Voltando a (3.12) e recordando que $U(\infty) = x_F$, podemos então escrever

$$\hat{x}_F = \hat{U}(n/k) + \hat{q}(n/k) = X_{n-k,n} - \frac{1}{\log 2} \int_0^1 (X_{n-[2ks],n} - X_{n-[ks],n}) \frac{1}{s} ds.$$

Para $s \in [0, 1/2k[$, tem-se $[2ks] = [ks] = 0$, pelo que $\int_0^{1/2k} (X_{n-[2ks],n} - X_{n-[ks],n}) \frac{1}{s} ds = 0$ e assim

$$\hat{x}_F = X_{n-k,n} - \frac{1}{\log 2} \int_{1/2k}^1 (X_{n-[2ks],n} - X_{n-[ks],n}) \frac{1}{s} ds.$$

Por outro lado, para $s \in [1/2k, 1/k[$ tem-se $[ks] = 0$, donde

$$\begin{aligned}
 & \int_{1/2k}^1 (X_{n-[2ks],n} - X_{n-[ks],n}) \frac{1}{s} ds = \\
 & = \int_{1/2k}^{1/k} (X_{n-[2ks],n} - X_{n,n}) \frac{1}{s} ds + \int_{1/k}^1 (X_{n-[2ks],n} - X_{n-[ks],n}) \frac{1}{s} ds = \\
 & = \int_{1/2k}^{1/k} X_{n-[2ks],n} \frac{1}{s} ds - X_{n,n} \int_{1/2k}^{1/k} \frac{1}{s} ds + \int_{1/k}^1 (X_{n-[2ks],n} - X_{n-[ks],n}) \frac{1}{s} ds \\
 & \stackrel{z = \frac{s}{2}}{=} \int_{1/2k}^{1/k} X_{n-[2ks],n} \frac{1}{s} ds - X_{n,n} \int_{1/2k}^{1/k} \frac{1}{s} ds + \int_{1/k}^1 X_{n-[2ks],n} \frac{1}{s} ds - \int_{1/2k}^{1/2} X_{n-[2kz],n} \frac{1}{z} dz \\
 & = \int_{1/2}^1 X_{n-[2ks],n} \frac{1}{s} ds - X_{n,n} \log 2
 \end{aligned}$$

Então,

$$\hat{x}_F = X_{n-k,n} + X_{n,n} - \frac{1}{\log 2} \int_{1/2}^1 X_{n-[2ks],n} \frac{1}{s} ds.$$

Como dissémos atrás o estimador \hat{x}_F foi estabelecido sob a hipótese de $F \in \mathcal{D}(G_0)$. Contudo, em Fraga Alves [8] as autoras provam que \hat{x}_F converge quase certamente para x_F para qualquer $\gamma \leq 0$. Assim, \hat{x}_F passou a ser um estimador fortemente consistente de x_F (no caso em que este é finito), para f.d.'s F no domínio de atração da f.d. de Gumbel e da f.d. de Weibull.

Proposição 3.2.1 *Admitamos que F verifica a condição de primeira ordem (2.22) para $\gamma \leq 0$, com x_F finito e que a sucessão $\{k_n\}$ satisfaz as condições $k_n \rightarrow 0$ e $k_n \rightarrow +\infty$. Então $\hat{x}_F \rightarrow x_F$, $n \rightarrow +\infty$, q.c.*

Prova: Atendendo a que $X_{n-[2ks],n}$ é decrescente como função de s , temos

$$(\log 2)X_{n-2k,n} \leq \int_{1/2}^1 X_{n-[2ks],n} \frac{1}{s} ds \leq (\log 2)X_{n-k,n}$$

donde

$$X_{n-k,n} + X_{n,n} - \frac{1}{\log 2} X_{n-k,n} \log 2 < \hat{x}_F < X_{n-k,n} + X_{n,n} - \frac{1}{\log 2} X_{n-2k,n} \log 2,$$

o que implica

$$X_{n,n} - x_F < \hat{x}_F - x_F < X_{n,n} - x_F + X_{n-k,n} - X_{n-2k,n}. \quad (3.14)$$

Uma vez que para qualquer $t \geq 0$, pelo Lema 3.1.1, com $\{k_n\}$ nas condições do enunciado, se tem $X_{n-[kt]} \xrightarrow{n \rightarrow \infty} x_F$, q.c., de (3.14) decorre que $\hat{x}_F \xrightarrow{n \rightarrow \infty} x_F$, q.c. \square

O estimador \hat{x}_F pode ser reescrito na forma que passamos a apresentar. Com efeito,

$$\begin{aligned}
 \hat{x}_F &= X_{n-k,n} + X_{n,n} - \frac{1}{\log 2} \int_{1/2}^1 X_{n-[2ks],n} \frac{1}{s} ds \\
 &= X_{n-k,n} + X_{n,n} - \frac{1}{\log 2} \int_{k/2k}^{2k/2k} X_{n-[2ks],n} \frac{1}{s} ds \\
 &= X_{n-k,n} + X_{n,n} - \frac{1}{\log 2} \sum_{i=0}^{k-1} \int_{\frac{k+i}{2k}}^{\frac{k+i+1}{2k}} X_{n-[2ks],n} \frac{1}{s} ds.
 \end{aligned} \tag{3.15}$$

Por outro lado, para k nas condições anteriores e $i \in \{0, \dots, k-1\}$ e s a satisfazer a desigualdade $2k \frac{k+i}{2k} \leq 2ks < 2k \frac{k+i+1}{2k}$, o que implica $[2ks] = [k+i] = k+i$, a última expressão de (3.15) dá lugar a

$$\begin{aligned}
 \hat{x}_F &= X_{n-k,n} + X_{n,n} - \frac{1}{\log 2} \sum_{i=0}^{k-1} X_{n-k-i,n} \int_{\frac{k+i}{2k}}^{\frac{k+i+1}{2k}} \frac{1}{s} ds \\
 &= X_{n-k,n} + X_{n,n} - \frac{1}{\log 2} \sum_{i=0}^{k-1} \log \left(1 + \frac{1}{k+i} \right) X_{n-k-i,n}.
 \end{aligned} \tag{3.16}$$

Terminamos esta secção com a estimação de quantis de ordem elevada da f.d. F , isto é, valores reais que têm uma probabilidade muito pequena de ser excedidos, isto é $U(\frac{1}{p})$, com p próximo de zero. Atendendo a que (2.21) e (2.22) são equivalentes, considerando $t = \frac{n}{k}$ em (2.22), obtemos

$$U\left(\frac{1}{p}\right) \simeq U\left(\frac{n}{k}\right) + a\left(\frac{n}{k}\right) D_\gamma\left(\frac{k}{np}\right)$$

onde $k := k_n$ verifica $\frac{k}{n} \rightarrow 0$ e $k_n \rightarrow +\infty$. Então, como $\hat{U}\left(\frac{n}{k}\right) = X_{n-k,n}$ e

$$\hat{a}\left(\frac{n}{k}\right) = X_{n-k,n} M_{n,k}^{(1)} (1 - \hat{\gamma}_{n,k}^-)$$

temos

$$\hat{U}\left(\frac{1}{p}\right) = X_{n-k,n} + \hat{a}\left(\frac{n}{k}\right) D_{\hat{\gamma}}\left(\frac{k}{np}\right)$$

onde

$$D_{\hat{\gamma}}\left(\frac{k}{np}\right) = \begin{cases} \frac{\left(\frac{k}{np}\right)^{\hat{\gamma}} - 1}{\hat{\gamma}}, & \gamma \neq 0 \\ \log\left(\frac{k}{np}\right), & \gamma = 0. \end{cases}$$

3.3 Escolha de domínios de atração

A escolha do domínio de atração a que a f.d. F pertence, iniciada com o conhecimento do sinal de $\hat{\gamma}$, é feita de forma mais robusta realizando um teste de hipóteses

$$H_0 : F \in \mathcal{D}(G_0) \quad vs \quad H_1 : F \in \mathcal{D}(G_\gamma)_{\gamma \neq 0} \quad (3.17)$$

ou

$$H_0 : F \in \mathcal{D}(G_0) \quad vs \quad H_1 : F \in \mathcal{D}(G_\gamma)_{\gamma > 0} \quad (3.18)$$

ou

$$H_0 : F \in \mathcal{D}(G_0) \quad vs \quad H_1 : F \in \mathcal{D}(G_\gamma)_{\gamma < 0} \quad (3.19)$$

Na literatura da Estatística de Extremos podemos encontrar os dois testes mais "populares" apropriados a este contexto: o teste de Greenwood e o teste de Hasofer & Wang (ver por exemplo [6], [8], [5]). A estatística de teste de Greenwood é dada por

$$G_{n,k} = \sqrt{\frac{k}{4}} \left(\frac{M_n^{(2)}}{(M_n^{(1)})^2} - 2 \right)$$

e prova-se que, sob a validade de $F \in \mathcal{D}(G_0)$ e com F a satisfazer as condições de primeira e segunda ordem se tem

$$G_{n,k} \xrightarrow{d} Z, n \rightarrow +\infty,$$

onde $Z \sim \mathcal{N}(0, 1)$. Assim, para testar as hipóteses (3.17) usamos a região crítica constituída pelas amostras para as quais se tem $|G_{n,k}| > z_{1-\frac{\alpha}{2}}$, onde α representa o nível de significância do teste e z_α representa o quantil de ordem $1 - \frac{\alpha}{2}$ da lei $\mathcal{N}(0, 1)$. Se optarmos por testar as hipóteses em (3.18) ou (3.19) a região de H_0 surge se $G_{n,k} > z_{1-\alpha}$ ou $G_{n,k} < -z_{1-\alpha}$ (respetivamente).

O teste de Hasofer & Wang tem estatística de teste

$$HW_{n,k} = \sqrt{\frac{k}{4}} \left(\frac{(M_n^{(1)})^2}{M_n^{(2)} - (M_n^{(1)})^2} - 1 \right).$$

Sob as mesmas condições para F e com $\sqrt{k}A_0(n/k) \rightarrow \lambda \in \mathbb{R}$, prova-se que

$$HW_{n,k} \xrightarrow{d} Z, n \rightarrow +\infty.$$

Uma vez que $HW_n(k)$ tem o sinal contrário ao de γ , isto é, assume valores negativos (positivos) quando F pertence ao domínio de atração da f.d. de Fréchet (Weibull, respetivamente), para testar (3.18) ou (3.19) rejeita-se H_0 se $HW_{n,k} < -z_{1-\alpha}$ ($HW_{n,k} > z_{1-\alpha}$, respetivamente).

Novamente com o objetivo de realizar um teste de hipótese nula $H_0 : F \in \mathcal{D}(G_0)$, contra as hipóteses alternativas $H_1 : F \in \mathcal{D}(G_\gamma)_{\gamma < 0}$, $H_1 : F \in \mathcal{D}(G_\gamma)_{\gamma > 0}$ ou $H_1 : F \in \mathcal{D}(G_\gamma)_{\gamma \neq 0}$, em Fraga Alves [8](2017) é introduzida a estatística

$$G_{n,k}^* = \log 2G_{n,k} - \left(\log k + \frac{\log 2}{2} \right),$$

onde

$$G_{n,k} = \frac{\hat{x}_F - X_{n-k,n}}{X_{n-k,n} - X_{n-2k,n}}.$$

A distribuição assintótica de $G_{n,k}^*$ é estabelecida no Teorema 3.3.1, cuja demonstração necessita dos lemas seguintes.

Lema 3.3.1 *Suponhamos que U satisfaz a condição de segunda ordem (3.8) com $\gamma \in \mathbb{R}$ e $\rho \leq 0$. Se $k = k_n$ é sucessão tal que $\sqrt{k}A_0(n/k) = O(1)$, então*

$$\sqrt{k}(P_n, Q_n) := \sqrt{k} \left(\int_{1/2}^1 \frac{X_{n-[2ks],n} - U(\frac{n}{2ks})}{a_0(\frac{n}{k})} \frac{ds}{s}, \frac{X_{n-k,n} - U(\frac{n}{k})}{a_0(\frac{n}{k})} \right) \quad (3.20)$$

converge em distribuição para o vetor normal bivariado (P, Q) , com média zero e covariância dada por $E(Q^2) = 1$,

$$E(P^2) = \begin{cases} \frac{2}{\gamma} \left(\frac{2^{-(2\gamma+1)} - 1}{2\gamma+1} - \frac{2^{-(\gamma+1)} - 1}{\gamma+1} \right), & \gamma \neq 0, \\ 1 - \log 2, & \gamma = 0 \end{cases}, \quad E(PQ) = \begin{cases} -\frac{1}{\sqrt{2}} \frac{2^{-\gamma} - 1}{\gamma}, & \gamma \neq 0 \\ \frac{\log 2}{\sqrt{2}}, & \gamma = 0. \end{cases}$$

Lema 3.3.2 *Sob as condições de primeira e de segunda ordem, tem-se no caso $\gamma = 0$,*

1. $\frac{X_{n,n} - U(n/k)}{a_0(n/k)} \stackrel{d}{=} \left(\log\left(\frac{Y_{n,n}}{n}\right) + \log k \right) (1 + o_p(1))$
2. $\frac{X_{n-2k,n} - U(n/k)}{a_0(n/k)} \stackrel{d}{=} \log(1/2) + O_p(1/\sqrt{k})$
3. $\frac{1}{\log 2} \int_{1/2}^1 \frac{U(n/2ks) - U(n/k)}{a_0(n/k)} \frac{1}{s} ds = -\frac{\log 2}{2} + O_p(A_0(n/k)).$

Teorema 3.3.1 *Suponhamos que $F \in \mathcal{D}(G_\gamma)$, para algum $\gamma \in \mathbb{R}$, e que são satisfeitas as condições de primeira e segunda ordem (2.12) e (3.8). Se $\{k_n\}$ nas condições anteriores verifica $\sqrt{k}A_0(n/k) \rightarrow \lambda^* \in \mathbb{R}$, $n \rightarrow +\infty$, então se $\gamma = 0$,*

$$G_{n,k}^* \xrightarrow{d} Z, n \rightarrow +\infty,$$

onde Z é uma variável com distribuição de Gumbel.

Prova: Começamos por observar que

$$\begin{aligned} G_{n,k} &= \frac{\hat{x}_F - X_{n-k,n}}{X_{n-k,n} - X_{n-2k,n}} = \frac{\hat{x}_F - X_{n-k,n} - U(n/k) + U(n/k)}{X_{n-k,n} - X_{n-2k,n}} \\ &= \frac{\hat{x}_F - X_{n-k,n} - U(n/k) + U(n/k)}{a_0(n/k)} \frac{1}{\frac{X_{n-k,n} - X_{n-2k,n} + U(n/k) - U(n/k)}{a_0(n/k)}}. \end{aligned} \quad (3.21)$$

Sejam, $Q_n = \frac{X_{n-k,n} - U(n/k)}{a_0(n/k)}$ e $P_n = \int_{1/2}^1 \frac{X_{n-[2ks],n} - U(n/2ks)}{a_0(n/k)} \frac{1}{s} ds$, as sucessões do Lema 3.3.1.

Deste lema, concluímos que se $\gamma = 0$,

$$\sqrt{k}P_n \xrightarrow{d} W \quad \text{e} \quad \sqrt{k}Q_n \xrightarrow{d} Z,$$

onde $W \sim N(0, 1 - \log 2)$ e $Z \sim N(0, 1)$. Pelos Lemas A.0.1(4.) e Lema A.0.4, concluímos que $P_n = O_p(1/\sqrt{k})$ e $Q_n(1/\sqrt{k})$.

Note-se que $U(n/k) = \frac{1}{\log 2} \int_{1/2}^1 U(n/k) \frac{1}{s} ds$. Então, (3.21) é igual a

$$\begin{aligned}
 & \left(X_{n,n} - U(n/k) + \frac{1}{\log 2} \int_{1/2}^1 U(n/k) \frac{1}{s} ds - \frac{1}{\log 2} \int_{1/2}^1 (X_{n-[2ks],n}) \frac{1}{s} ds \right) \times \\
 & \quad \times \frac{1}{\frac{X_{n-k,n} - X_{n-2k,n} + U(n/k) - U(n/k)}{a_0(n/k)}} = \\
 & = \left(\frac{X_{n,n} - U(n/k) + \frac{1}{\log 2} \int_{1/2}^1 (U(n/k) - X_{n-[2ks],n}) \frac{1}{s} ds}{a_0(n/k)} \right) \frac{1}{Q_n - \frac{X_{n-2k,n} - U(n/k)}{a_0(n/k)}} \\
 & = \left(\frac{X_{n,n} - U(n,k)}{a_0(n/k)} + \frac{1}{\log 2} \left(\frac{\int_{1/2}^1 (U(n/k) - X_{n-[2ks],n}) \frac{1}{s} ds}{a_0(n/k)} \right) \right) \frac{1}{Q_n - \frac{X_{n-2k,n} - U(n/k)}{a_0(n/k)}} \\
 & = \left(\frac{X_{n,n} - U(n,k)}{a_0(n/k)} - \frac{P_n}{\log 2} - \frac{1}{\log 2} \left(\int_{1/2}^1 \frac{(U(n/2ks) - U(n/k)) \frac{1}{s} ds}{a_0(n/k)} \right) \right) \frac{1}{Q_n - \frac{X_{n-2k,n} - U(n/k)}{a_0(n/k)}}
 \end{aligned}$$

Atendendo ao Lema 3.3.2 a última expressão é igual em distribuição a

$$\begin{aligned}
 & \stackrel{d}{=} \left(\left(\log \left(\frac{Y_{n,n}}{n} \right) + \log k \right) (1 + o_p(1)) - \frac{O_p(1/\sqrt{k})}{\log 2} - \frac{1}{\log 2} \int_{1/2}^1 \frac{U(n/2ks) - U(n/k)}{a_0(n/k)} \frac{1}{s} ds \right) \\
 & \quad \times \frac{1}{O_p(\frac{1}{\sqrt{k}}) + \log 2 + O_p(\frac{1}{\sqrt{k}})} \\
 & = \left(\left(\log \left(\frac{Y_{n,n}}{n} \right) + \log k \right) (1 + o_p(1)) - \frac{O_p(1/\sqrt{k})}{\log 2} + \frac{\log 2}{2} + O_p(A_0(n/k)) \right) \frac{1}{O_p(\frac{1}{\sqrt{k}}) + \log 2} \\
 & = \left(\left(\log \left(\frac{Y_{n,n}}{n} \right) + \log k \right) (1 + o_p(1)) + o_p(1) + \frac{\log 2}{2} + o_p(1) \right) \frac{1}{o_p(1) + \log 2} \\
 & = \frac{1}{\log 2} \left(\left(\log \left(\frac{Y_{n,n}}{n} \right) + \log k \right) (1 + o_p(1)) + \frac{\log 2}{2} + o_p(1) \right) (1 + o_p(1)) \\
 & = \frac{1}{\log 2} \left(\left(\log \left(\frac{Y_{n,n}}{n} \right) + \log k \right) (1 + o_p(1)) + \frac{\log 2}{2} (1 + o_p(1)) \right) (1 + o_p(1)) \\
 & = \frac{1}{\log 2} \left(\log \left(\frac{Y_{n,n}}{n} \right) + \log k + \frac{\log 2}{2} \right) (1 + o_p(1)),
 \end{aligned}$$

pois $O_p(A_0(n/k)) = o_p(1)$, $n \rightarrow +\infty$, e $O_p(\frac{1}{\sqrt{k}}) = o_p(1)$, $n \rightarrow +\infty$, uma vez que $k \rightarrow +\infty$ e $A_0(n/k) \rightarrow 0$, $n \rightarrow +\infty$.

Uma vez que $Y_{n,n}$ representa o máximo amostral associado a uma variável aleatória com f.d. de Pareto, pelo Exemplo 2.1.3, sabemos que $\frac{Y_{n,n}}{n}$ converge em distribuição para uma variável aleatória com f.d. de Fréchet. Por outro lado, se X é uma variável aleatória com f.d. de Fréchet então $\log X$ tem f.d. de Gumbel. Com efeito $P(\log X \leq x) = P(X \leq e^x) = \exp(-e^{-x})$, $\forall x \in \mathbb{R}$. Assim como a função logaritmo é contínua pelo Lema A.0.2 tem-se que

$$\log \left(\frac{Y_{n,n}}{n} \right) \xrightarrow{d} Z, n \rightarrow +\infty,$$

onde Z uma v.a. com f.d. de Gumbel. Então pela Proposição A.0.5, concluímos que

$$\log 2G_{n,k} - \log k - \frac{\log 2}{2} \xrightarrow{d} Z, n \rightarrow +\infty.$$

□

Por fim, provámos que a estatística $G_{n,k}^*$ converge em distribuição para uma variável aleatória com lei de Gumbel. Assim, denotando por g_ε o quantil de ordem ε da variável Z , ao testar as hipóteses em (3.17), rejeitamos H_0 se $G_{n,k}^* > g_{1-\alpha}$ ou $G_{n,k}^* < g_\alpha$. Para testar (3.18) ou (3.19) a rejeição de H_0 ocorre quando $G_{n,k}^* > g_{1-\alpha}$ ou $G_{n,k}^* < g_\alpha$ respetivamente.

3.4 Estimação dos parâmetros e escolha da função de distribuição G_γ

3.4.1 Método dos momentos ponderados

A estimação dos parâmetros da f.d. G_γ , usando o método dos Momentos Ponderados, foi introduzida em [16].

Os momentos ponderados de uma variável aleatória X com f.d. F são definidos

$$M_{q,r,s} = E[X^q \{F(X)\}^r \{1 - F(X)\}^s],$$

onde q, r, s são números reais. Quando F é invertível, podemos escrever

$$M_{q,r,s} = \int_0^1 \{Q(p)\}^q p^r (1-p)^s dp,$$

onde $Q(p)$ representa o quantil de ordem p da f.d. F .

Neste contexto, consideramos

$$\beta_r = M_{1,r,0} = E[X \{F(X)\}^r], r = 0, 1, 2.$$

Sejam $X_{1:n} \leq X_{2:n} \leq \dots \leq X_{n:n}$, as estatísticas ordinais associadas à amostra aleatória (X_1, X_2, \dots, X_n) .

A estatística

$$b_r = \frac{1}{n} \sum_{j=1}^n \frac{(j-1)(j-2)\dots(j-r)}{(n-1)(n-2)\dots(n-r)} X_{j:n}$$

é um estimador cêntrico de β_r ([9]).

Vamos determinar os quantis da f.d. $G_\gamma(\frac{x-b}{a})$. Com efeito, para $p \in]0, 1[$, o quantil de ordem p , denotado por $Q(p)$ verifica $p = G_\gamma(Q(p))$.

Para $\gamma \neq 0$, tem-se

$$\begin{aligned} p &= \exp\left(-\left(1 + \gamma \frac{Q(p)-b}{a}\right)^{-\frac{1}{\gamma}}\right) \Leftrightarrow (-\ln p)^{-\gamma} = \left(1 + \gamma \frac{Q(p)-b}{a}\right) \\ &\Leftrightarrow Q(p) = b + \frac{a}{\gamma} ((-\ln p)^{-\gamma} - 1). \end{aligned}$$

Para $\gamma = 0$, tem-se

$$p = \exp(-e^{-\frac{Q(p)-b}{a}}) \Leftrightarrow Q(p) = b - a \ln(-\ln p).$$

Estabelecemos de seguida os momentos ponderados da distribuição G_γ , para $\gamma \neq 0$ e $\gamma < 1$. Com efeito, temos

$$\begin{aligned} \beta_r = M_{1,r,0} &= \int_0^1 \left[b - a \left(\frac{(1 - (-\log p)^{-\gamma})}{\gamma} \right) \right] p^r dp = \int_0^\infty \left[b - \frac{a}{\gamma} (1 - u^{-\gamma}) \right] e^{-(r+1)u} du \\ &= (b - \frac{a}{\gamma}) \int_0^\infty e^{-(r+1)u} du + \frac{a}{\gamma} \int_0^\infty u^{-\gamma} e^{-(r+1)u} du \\ &= \frac{1}{r+1} (b - \frac{a}{\gamma} (1 - (r+1)^\gamma \Gamma(1-\gamma))), \end{aligned}$$

desde que $\gamma < 1$. Quando $\gamma \geq 1$, β_0 é a média da distribuição e os restantes valores de β_r não existem.

Pela expressão de β_r obtemos

$$\begin{cases} a = \frac{(2\beta_1 - \beta_0)\gamma}{\Gamma(1-\gamma)(2^\gamma - 1)} \\ b = \beta_0 + \frac{a(1 - \Gamma(1-\gamma))}{\gamma} \\ \frac{1 - 3^\gamma}{1 - 2^\gamma} = \frac{3\beta_2 - \beta_0}{2\beta_1 - \beta_0} \end{cases}$$

Assim, os estimadores $\hat{\gamma}$, \hat{b} e \hat{a} são soluções do sistema

$$\begin{cases} b_0 = \hat{b} - \frac{\hat{a}(1 - \Gamma(1-\hat{\gamma}))}{\hat{\gamma}} \\ 2b_1 - b_0 = \hat{a}\Gamma(1-\hat{\gamma}) \frac{2^{\hat{\gamma}} - 1}{\hat{\gamma}} \\ \frac{3b_2 - b_0}{2b_1 - b_0} = \frac{1 - 3^{\hat{\gamma}}}{1 - 2^{\hat{\gamma}}} \end{cases}$$

o qual terá de ser resolvido com métodos iterativos. Dado $\hat{\gamma}$ os parâmetros de escala e localização podem ser estimados por $\hat{a} = \frac{(2b_1 - b_0)\hat{\gamma}}{\Gamma(1-\gamma)(2^\gamma - 1)}$ e $\hat{b} = b_0 - \frac{\hat{a}(\Gamma(1-\hat{\gamma}) - 1)}{\hat{\gamma}}$.

O caso $\gamma = 0$ obtém-se de forma similar (ver por exemplo [9]). Os estimadores \hat{a} e \hat{b} são dados por $\hat{a} = \frac{2b_1 - b_0}{\log 2}$ e $\hat{b} = b_0 + 0.5772\hat{a}$.

3.4.2 Método da Máxima Verosimilhança

Os estimadores da Máxima Verosimilhança dos parâmetros da f.d. de Gumbel, $G_0(\frac{x-b}{a})$, são solução de

$$\begin{cases} \hat{b} = \frac{\hat{a}}{n} \ln \left\{ \sum_{i=1}^n \exp(-x_i/\hat{a}) \right\} \\ \hat{a} = \bar{x} - \frac{\sum_{i=1}^n x_i \exp(-x_i/\hat{a})}{\sum_{i=1}^n \exp(-x_i/\hat{a})} \end{cases} .$$

Os estimadores da Máxima Verosimilhança dos parâmetros da f.d. $G_\gamma(\frac{x-b}{a})$ ($\gamma > -1$, $\gamma \neq 0$), são solução de

$$\begin{cases} \frac{1}{\hat{a}} \sum_{i=1}^n \left[\frac{1 + \hat{\gamma} - (y_i)^{-\frac{1}{\hat{\gamma}}}}{y_i} \right] = 0 \\ \frac{-n}{\hat{a}} + \frac{1}{\hat{a}} \sum_{i=1}^n \left[\frac{1 + \hat{\gamma} - (y_i)^{-\frac{1}{\hat{\gamma}}}}{y_i} \left(\frac{x_i - \hat{b}}{\hat{a}} \right) \right] = 0 \\ -\frac{1}{\hat{\gamma}^2} \sum_{i=1}^n \left\{ \ln(y_i) \left[1 + \gamma - (y_i)^{-\frac{1}{\gamma}} \right] - \frac{1 + \hat{\gamma} - (y_i)^{-\frac{1}{\hat{\gamma}}}}{y_i} \hat{\gamma} \left(\frac{x_i - \hat{b}}{\hat{a}} \right) \right\} = 0 \end{cases} ,$$

onde $y_i = \frac{1 + \gamma/a}{x_i - b}$. A restrição $\gamma > -1$ e o facto de nem sempre os métodos iterativos convergirem, legitimam a procura de outros métodos de estimação, como o que foi apresentado na subsecção anterior.

3.4.3 Teste de Gumbel

Como sabemos, a classe das f.d.'s G_γ tem como sub-classes as f.d.'s Gumbel, Fréchet ou Weibull, as quais são definidas a partir do conhecimento de $\gamma = 0$, $\gamma > 0$, $\gamma < 0$, respectivamente.

Com o Teste de Gumbel, introduzido em [14], testam-se as hipóteses

$$H_0 : \gamma = 0 \quad \text{vs} \quad H_1 : \gamma \neq 0 \quad (\gamma < 0 \text{ ou } \gamma > 0).$$

A estatística de teste utilizada é

$$GS_n = \log(\log n) \left(\frac{X_{n,n} - X_{[n/2]+1,n}}{X_{[n/2]+1,n} - X_{n,1}} - \frac{\log n + \log \log 2}{\log(\log n) - \log(\log 2)} \right).$$

Admitindo H_0 verdadeira, prova-se que $GS_n \xrightarrow{d} W$, $n \rightarrow +\infty$, onde W é uma v.a. com lei de Gumbel.

Assim, as observações levam à rejeição de H_0 se $GS_n < g_{\alpha/2}$ ou $GS_n > g_{1-\alpha/2}$. Se considerarmos $H_1 : \gamma < 0$, esta condição é $GS_n < g_\alpha$ e se considerarmos o teste com $H_1 : \gamma > 0$, os dados conduzem à rejeição de H_0 desde que $GS_n > g_{1-\alpha}$. Recordamos que g_ε representa o quantil de ordem ε da f.d. de Gumbel, definido por $g_\varepsilon = -\ln(-\ln \varepsilon)$.

Observamos que no contexto da Estatística de valores extremos a amostra utilizada terá de ser uma amostra de valores máximos associados a uma variável aleatória.

Capítulo 4

Aplicações

4.1 Aplicações ao triplo salto feminino

Começemos com um exemplo de máximos de triplo salto feminino (outdoor). Entre os dados disponíveis, considerámos as 700 maiores observações respeitantes às marcas pessoais de cada atleta (em metros), não havendo repetição de marcas de um atleta. As marcas contidas na amostra foram obtidas entre 31/12/1899 até pelo menos ao dia 15/04/2022. Estes valores foram retirados de: <https://www.worldathletics.org/records/all-time-toplists/jumps/triple-jump/outdoor/women/senior>.

O tratamento estatístico destes dados foi realizado com o software R studio.

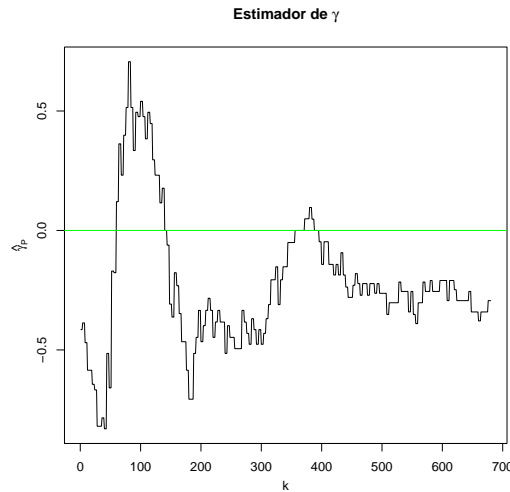
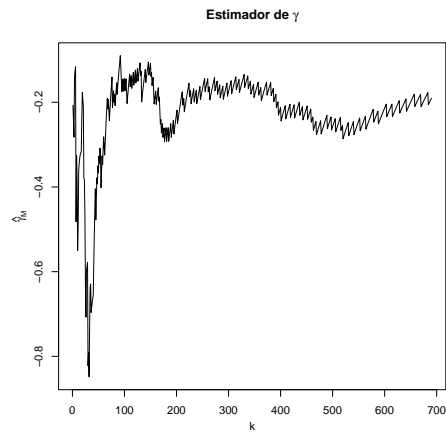
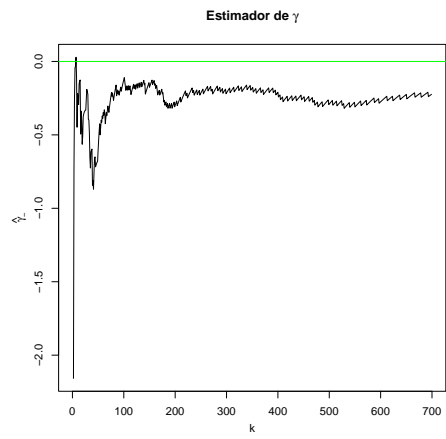
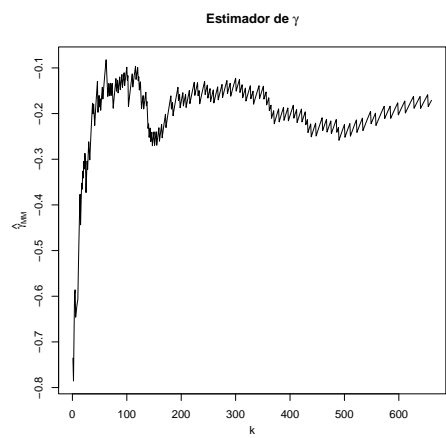


Fig. 4.1 Valores de $\hat{\gamma}_p$ versus k .

As Figuras 4.1, 4.2, 4.3, e 4.4 dizem respeito às trajetórias amostrais dos estimadores de γ , apresentados na Secção 3.2, em função de k . A primeira conclusão que retiramos destas figuras é a evidência de se ter $\gamma \leq 0$, ou seja, a f.d. F das marcas do triplo salto feminino pertence ao domínio de atração de Weibull ou eventualmente de Gumbel.

Fig. 4.2 Valores de $\hat{\gamma}_M$ versus k .Fig. 4.3 Valores de $\hat{\gamma}_-$ versus k .Fig. 4.4 Valores de $\hat{\gamma}_{MM}$ versus k .

Depois da informação fornecida pelos valores dos estimadores de γ , em função de k , continuámos a análise estatística realizando os testes de hipóteses

$$H_0 : F \in \mathcal{D}(G_0) \quad \text{vs} \quad H_1 : F \in \mathcal{D}(G_\gamma)_{\gamma < 0}.$$

Na Figura 4.5, referente ao teste de Greenwood, a linha a vermelho corresponde ao valor do quantil de ordem 0.05 da f.d. da lei $\mathcal{N}(0, 1)$.

Seguindo o critério proposto em [6], consideramos k_{opt} como o valor de k para o qual $G_{n,k}$ interseeta a reta $y = z_{0.05} = -1.645$. Neste caso, tem-se $k_{opt}^{(1)} = 165$. Notamos que para $k \geq k_{opt}^{(1)}$ se tem a rejeição de H_0 .

A Figura 4.6 relativa à trajetória da estatística de Hasofer & Wang, denotada por, $HW_{n,k}$, em função de k , permite concluir que $k_{opt}^{(2)} = 160$. Assim estas 160 maiores observações remetem para o domínio de atração de Weibull.

Por fim, na Figura 4.7 encontram-se os valores da estatística de Fraga Alves [6], $G_{n,k}^*$, em função de k ($k = 10, \dots, 350$). A linha a vermelho corresponde à reta $y = g_{0.05} = -\ln(-\ln(0.05)) = -1.097$. Neste caso, como $k_{opt}^{(3)}$ é muito inferior a $\max(k_{opt}^{(1)}, k_{opt}^{(2)})$, não foi necessário determinar o seu valor e concluímos que estes dados conduzem à rejeição de H_0 .

Uma vez que as f.d.'s F que pertencem ao domínio de atração da f.d. de Weibull têm x_F finito, as conclusões que temos até agora parecem bastante adequadas à natureza deste tipo de desporto.

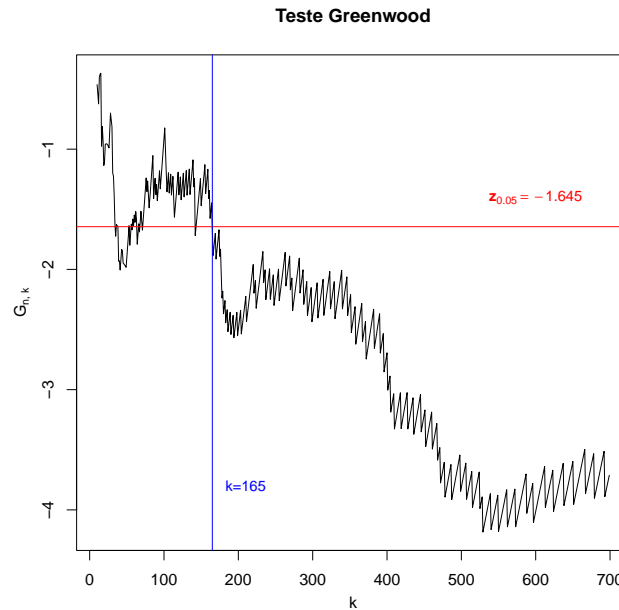
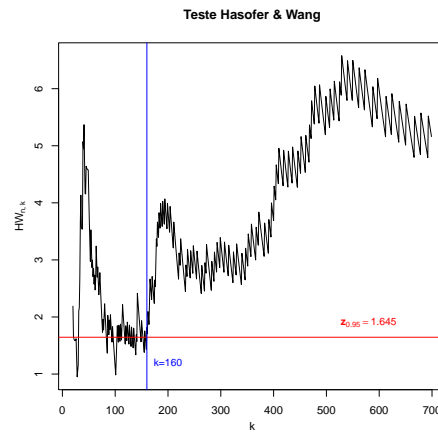
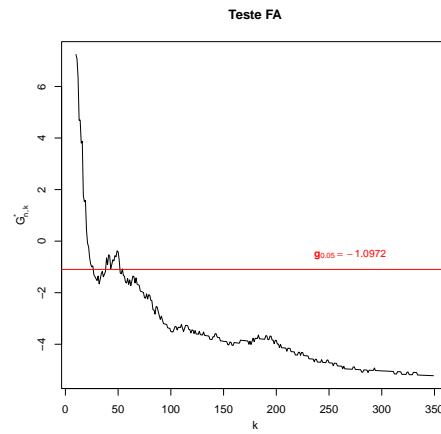
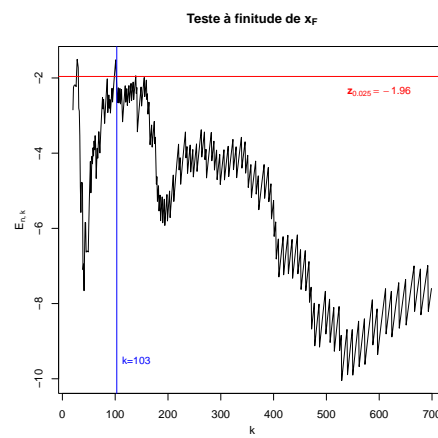


Fig. 4.5 Valores de $G_{n,k}$ versus k .

Fig. 4.6 Valores de $HW_{n,k}$ versus k .Fig. 4.7 Valores de $G_{n,k}^*$ versus k .Fig. 4.8 Valores de $E_{n,k}$ versus k .

No sentido de complementar a análise estatística feita anteriormente realizámos o teste à finitude de x_F [12], isto é, o teste de hipóteses

$$H_0 : F \in \mathcal{D}(G_0), \quad x_F = +\infty \quad \text{vs} \quad H_1 : F \in \mathcal{D}(G_\gamma)_{\gamma \leq 0}, \quad x_F < +\infty.$$

A partir da Figura 4.8, fica claro que os valores da estatística de teste, $E_{n,k}$, se encontram "quase todos" no intervalo $] -\infty, z_{0.025}] =] -\infty, -1.96]$. Fica também claro que para $k \geq 103$ se tem $E_{n,k} < -1.96$. Isto leva à clara rejeição de H_0 , o que vem sustentar as conclusões anteriores.

Então, perante estes dados devemos concluir que F pertence ao domínio de atração da f.d. de Weibull, restando apenas a estimação de x_F .

Considerámos o estimador de x_F apresentado na fórmula (3.10), com $\hat{\gamma}$ substituído por $\hat{\gamma}_-$, $\hat{\gamma}_M$ e $\hat{\gamma}_{MM}$. A substituição de $\hat{\gamma}$ por $\hat{\gamma}_p$ mostrou-se inconclusiva. Os resultados encontram-se na Tabela 4.1. O estimador de \hat{x}_F introduzido em [4] e estudado em [8], para estes dados gera a estimativa $\hat{x}_F = 15.9445$, o que também incluímos na Tabela 4.1.

As figuras 4.9, 4.10 e 4.11 mostram as trajetórias de \hat{x}_F em função de k para estas três substituições de $\hat{\gamma}$ em (3.10).

Depois de considerarmos $k_{opt} = 165$ (valor máximo de todos os k_{opt} anteriores) calculámos as estimativas de γ e de x_F geradas pelos estimadores referidos.

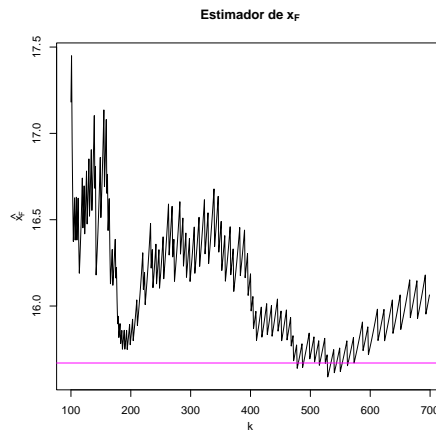


Fig. 4.9 Valores de \hat{x}_F versus k , considerando o estimador $\hat{\gamma}_-$ (expressão 3.10).

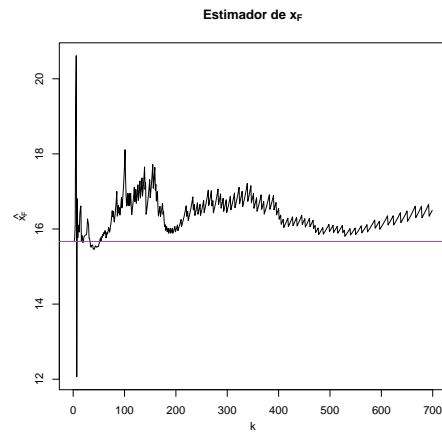


Fig. 4.10 Valores de \hat{x}_F versus k , considerando o estimador $\hat{\gamma}_M$ (expressão 3.10).

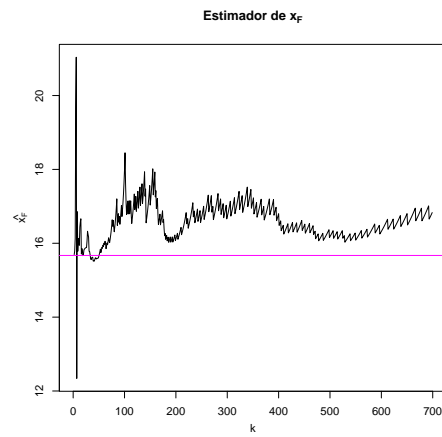


Fig. 4.11 Valores de \hat{x}_F versus k , considerando o estimador $\hat{\gamma}_{MM}$ (expressão 3.10).

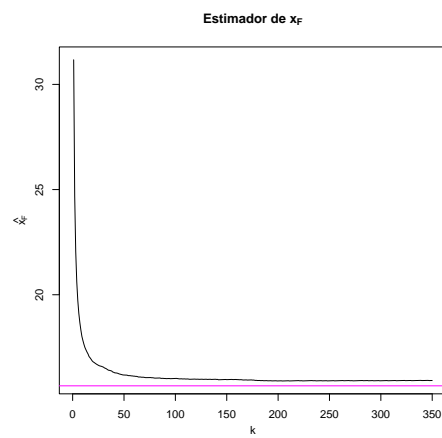


Fig. 4.12 Valores de \hat{x}_F versus k (expressão 3.16).

	$\hat{\gamma}$	\hat{x}_F
$\hat{\gamma}_P$	-0.0614	
$\hat{\gamma}_-$	-0.1725	16.3369
$\hat{\gamma}_M$	-0.1958	16.6123
$\hat{\gamma}_{MM}$	-0.1597	16.7964
FA		15.9445

Tabela 4.1 Estimativas de γ e de x_F – Triplo salto feminino

Na nossa opinião a estimativa $\hat{x}_F = 15.9445$ é a mais realista (facilmente atingível) uma vez que o máximo desta amostra é $x_{\max} = 15.67$ m.

Procedemos da mesma forma para obter a análise estatística do mesmo número de observações (700) do triplo salto masculino (outdoor).

Estes valores foram retirados de:

'<https://www.worldathletics.org/records/all-time-toplists/jumps/triple-jump/outdoor/man/senior>'.

As conclusões retiradas para os estimadores e a partir da realização dos mesmos testes são em tudo semelhantes ao caso do triplo salto feminino. Tem-se $k_{opt} = 223$. Uma vez que considerámos desnecessário apresentar as figuras deste estudo, optámos por incluir apenas o resumo na Tabela 4.2. Neste caso, $x_{\max} = 18.29$ m.

	$\hat{\gamma}$	\hat{x}_F
$\hat{\gamma}_P$	-0.2224	
$\hat{\gamma}_-$	-0.1989	18.7406
$\hat{\gamma}_M$	-0.1838	18.8704
$\hat{\gamma}_{MM}$	-0.1759	18.9545
FA		18.5347

Tabela 4.2 Estimativas de γ e de x_F – Triplo salto masculino

A nossa aplicação foi também feita aos dados do salto em comprimento feminino retirados de

'<https://www.worldathletics.org/records/all-time-toplists/jumps/long-jump/outdoor/women/senior>'.

Mais uma vez se assumiu que F pertence ao domínio de atração da f.d. de Weibull. O resumo encontra-se na Tabela 4.3. O máximo desta amostra é $x_{\max} = 7.52$ m.

Recordamos que os dados do salto em comprimento masculino foram estudados em [6].

	$\hat{\gamma}$	\hat{x}_F
$\hat{\gamma}_P$	0.1155	
$\hat{\gamma}_-$	-0.1656	7.8138
$\hat{\gamma}_M$	-0.1912	7.9797
$\hat{\gamma}_{MM}$	-0.1535	8.0776
FA		7.6217

Tabela 4.3 Estimativas de γ e de x_F – Salto em comprimento feminino

4.2 Aplicação aos tempos do cubo de Rubik

O tempo mínimo que o cubo de Rubik leva a fazer é a nova variável de estudo.

Uma vez que só apresentámos resultados para o máximo, usamos a igualdade

$$\min(x_1, \dots, x_n) = -\max(-x_1, \dots, -x_n)$$

para chegar ao que pretendemos. Consideremos uma amostra de 100 valores, correspondentes aos 100 tempos mínimos de resolução do cubo de Rubik (em segundos), correspondentes a 100 atletas diferentes. Os valores com que trabalhamos são os simétricos dos valores da amostra, uma vez que neste caso falamos de mínimos e

$$\begin{aligned} P(\min(X_1, X_2, \dots, X_n) \leq x) &= P(-\max(-X_1, -X_2, \dots, -X_n) \leq x) \\ &= 1 - P(\max(-x_1, \dots, -x_n) \leq -x). \end{aligned}$$

Estes valores foram retirados de '<https://www.worldcubeassociation.org/results/rankings/333/single>'.

O tratamento estatístico destes dados foi realizado com o software R studio.

Recorde-se que do limite (2.1) se obtém

$$P(M_n \leq y) \simeq G_\gamma \left(\frac{y - b_n}{a_n} \right).$$

Assim, assumindo n suficientemente grande mas fixo, podemos considerar que dispomos de uma amostra de dimensão 100 de uma variável aleatória (máximo) com f.d. aproximadamente $G_\gamma(\frac{y-b}{a})$. Então, basta estimar os parâmetros de forma, localização e escala, γ, b, a .

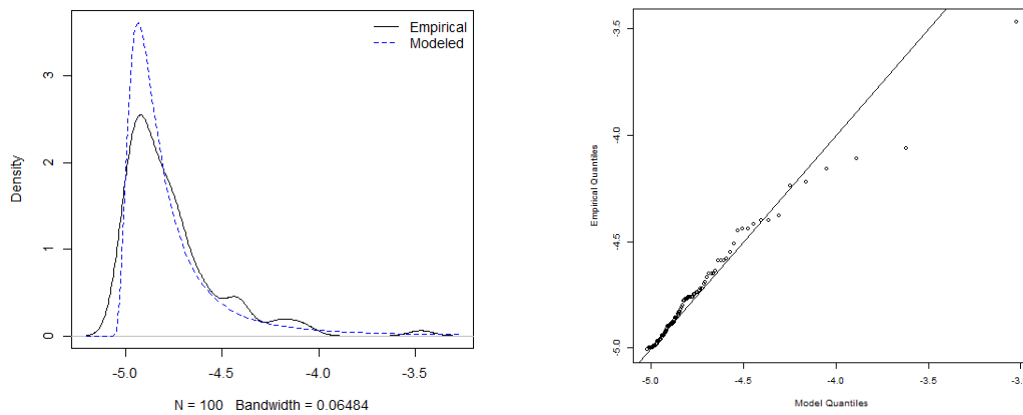
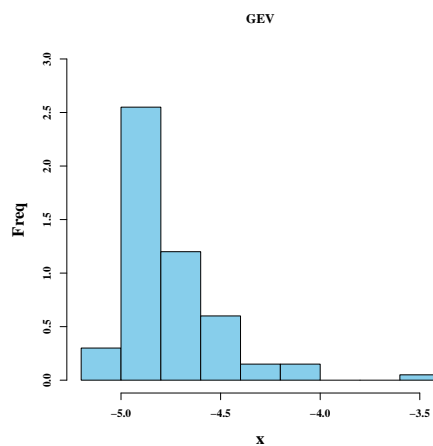
Fig. 4.13 Ajustamento de uma densidade e QQ-Plot (G_γ)

Fig. 4.14 Histograma

O histograma presente na Figura 4.14 sugere um ajustamento por uma f.d. de Fréchet ou de Gumbel, o que é confirmado pelo gráfico da Figura 4.13.

Realizámos o Teste de Gumbel (subsecção 3.5.3) de hipóteses

$$H_0 : \gamma = 0 \quad \text{vs} \quad H_1 : \gamma > 0$$

onde, para estes dados, a estatística de teste é $GS_{100} = 8.06$.

Como $g_{1-0.05} = 2.97$ (e $g_{1-0.001} = 6.9$), o valor de GS_{100} conduz sempre à rejeição de H_0 . Então, de acordo com estes dados, podemos admitir que a variável aleatória em estudo (simétrico de mínimos) dos tempos de resolução do cubo de Rubik tem lei de Fréchet.

Pelos métodos da Máxima Verosimilhança e dos Momentos Ponderados, foram estimados os parâmetros de forma, localização e dispersão. Vejamos a tabela 4.4 onde as estimativas de $\hat{\gamma}$ positivas confirmam a f.d. de Fréchet.

	$\hat{\gamma}$	\hat{a}	\hat{b}
MLE	0.4756	0.1120	-4.898
PWM	0.2685	0.1301	-4.8868

Tabela 4.4 Estimativas dos parâmetros da f.d. G_γ

	p	Estimados	Empíricos
MLE	0.95	-4.1666	-4.239
	0.98	-3.6268	-4.109
	0.99	-3.0355	-4.054
PWM	0.95	-4.2958	-4.239
	0.98	-3.9905	-4.109
	0.99	-3.7059	-4.054

Tabela 4.5 Quantis de ordem p estimados e empíricos

Pelos valores da Tabela 4.5, concluímos que, tendo em conta as estimativas da Máxima Verosimilhança, há uma probabilidade de aproximadamente 0.02 de algum atleta vir a resolver o cubo em menos de 3.63 segundos e uma probabilidade de aproximadamente 0.01 de algum atleta vir a registar um tempo menor do que 3.04 segundos.

Se considerarmos as estimativas geradas pelo método dos Momentos Ponderados, vemos que há uma probabilidade de aproximadamente 0.02 de algum atleta vir a fazer um tempo melhor que 3.99 segundos e uma probabilidade de aproximadamente 0.01 de algum atleta vir a resolver o cubo em menos de 3.70 segundos.

Observamos que o mínimo desta amostra é 3.47 segundos.

Bibliografia

- [1] Alexanderian, A. (2015). *Some notes on asymptotic theory in probability*, The University of Texas at Austin, pre publication.
- [2] Billingsley, P. (1986). *Probability and Measure*, volume 2. John Wiley Sons, New York.
- [3] de Haan, L. and Ferreira, A. (2006). *Extreme Value Theory: An Introduction*. Springer Series in Operations Research and Financial Engineering. Springer.
- [4] Fraga Alves, Isabel; Neves, C. (2014). Estimation of the finite right endpoint in the gumbel domain. *Statistica Sinica*, 24:1811–1835.
- [5] Fraga Alves, M.I.; Neves, C. (2008). Testing extreme value conditions - an overview and recent approaches. *REVSTAT - Statistical Journal*, 6:83–100.
- [6] Fraga Alves, Isabel; de Haan, L. N. C. (2013). How far can man go? *Advances in theoretical and applied statistics Stud. Theor. Appl. Stat. Sel. Papers Stat. Soc.*, Springer, Heidelberg, pages 187–197.
- [7] Fraga Alves, M. Isabel; Gomes, M. I. d. H. L. N. C. (2009). Mixed moment estimator and location invariant alternatives. *Springer Science + Business Media*, 12:149–185.
- [8] Fraga Alves, Isabel; Rosário, P. N. C. (2017). A general estimator for the right endpoint with an application to supercentenarian women's records. *Springer Science, Business Media New York 2016*.
- [9] Landwehr, J.M., N. a. W. J. (1973). "Probability Weighted Moments Compared With Some Traditional Techniques in Estimating Gumbel Parameters and Quantiles". *Water Resources Research*, 15, 1055-1064.
- [10] Leadbetter, M.R., L. G. R. H. (1983). "Extremes and Related Properties of Random Sequences and Processes". Springer Verlag, Berlin.
- [11] Neves, Cláudia; Picek, J. F. A. M. (2006). *The contribution of the maximum to the sum of excesses for testing max-domains of attracting*. *Journal of Statistical Planning and Inference* 136.
- [12] Pereira, António; Neves, C. (2010). Detecting finiteness in the right endpoint of light-tailed distributions. *Statistics and Probability Letters*, 80:437–444.
- [13] Resnick, S. (1987). *Extreme Values, regular variation and point processes*. Springer-Verlag, N.Y.
- [14] Tiago de Oliveira, J., G. M. (1984). *Two test statistics for choice of univariate extreme models*. Statistical Extremes and Applications, D.Reidel, Dordrecht, Holland, Ch.
- [15] Vicente, S. (2012). Extreme value theory: an application to sports. Master's thesis, Universidade de Lisboa, Faculdade de Ciências, Departamento de Estatística e Investigação Operacional.

- [16] Wallis, J. R. M. H. J. R. and Wood, E. F. (1985). *Estimation of the Generalized Extreme-Value Distribution by the Method of Probability-Weighted Moments*. Taylor Francis. Ltd. on behalf of American Statistical Association and American Society for Quality.

Anexo A

Convergência em probabilidade e as notações $o_p(n)$, $O_p(n)$

Começemos por apresentar a definição de convergência em probabilidade.

Definição A.0.1 *Seja $\{X_n\}$ uma sucessão de variáveis aleatórias reais. Dizemos que X_n converge em probabilidade para X , escrevendo $X_n \xrightarrow{P} X$, se para todo $\varepsilon > 0$,*

$$P(|X_n - X| > \varepsilon) \rightarrow 0, \quad n \rightarrow \infty.$$

No caso particular em que $X = 0$, isto é, em que X_n converge em probabilidade para zero, escrevemos $X_n \xrightarrow{P} 0$ ou $X_n = o_p(1)$.

Passamos a definir a noção de sucessão limitada em probabilidade.

Definição A.0.2 *Seja $\{X_n\}$ uma sucessão de variáveis aleatórias reais. Dizemos que X_n é limitada em probabilidade, escrevendo $X_n = O_p(1)$, se para todo $\varepsilon > 0$, existe $M_\varepsilon > 0$ tal que,*

$$P(|X_n| > M_\varepsilon) < \varepsilon, \quad \forall n.$$

Proposição A.0.1 *Se $X_n \xrightarrow{P} 0$, então X_n é limitada em probabilidade.*

Prova:

Suponhamos que $X_n \xrightarrow{P} 0$. Com $\varepsilon > 0$, sabemos que $P(|X_n| > 1) \rightarrow 0, n \rightarrow \infty$, portanto, existe algum n_0 tal que

$$P(|X_n| > 1) < \varepsilon, \quad n \geq n_0.$$

Agora escolhemos M_0 suficientemente grande de forma a que $P(|X_i| > M_0) < \varepsilon$, para $i = 1, \dots, n_0 - 1$. Assim, para $M = \max(1, M_0)$ temos que $P(|X_n| > M) < \varepsilon, \forall n$, que completa a prova de que $X_n = O_p(1)$. \square

Lema A.0.1 *Seja $\{X_n\}$ uma sucessão de variáveis aleatórias reais e seja $\{a_n\}$ uma sucessão de reais estritamente positivos.*

1. X_n converge em probabilidade para X se e só se $X_n - X = o_p(1)$.
2. $X_n = o_p(a_n)$ se e só se $a_n^{-1}X_n = o_p(1)$.

3. $X_n = O_p(a_n)$ se e só se $a_n^{-1}X_n = O_p(1)$.

Lema A.0.2 *Sejam $\{X_n\}$ e $\{Y_n\}$ sucessões de variáveis aleatórias reais e sejam $\{a_n\}$ e $\{b_n\}$ sucessões de reais estritamente positivos.*

1. Se $X_n \xrightarrow{P} X$ e $Y_n \xrightarrow{P} Y$ então,

$$(a) X_n Y_n \xrightarrow{P} XY.$$

$$(b) X_n \pm Y_n \xrightarrow{P} X \pm Y.$$

2. Suponhamos $X_n = o_p(a_n)$ e $Y_n = o_p(b_n)$. Então,

$$(a) X_n Y_n = o_p(a_n b_n).$$

$$(b) X_n + Y_n = o_p(\max(a_n, b_n))$$

3. $|X_n|^r = o_p(a_n)$ e $Y_n = O_p(b_n)$, então $X_n Y_n = o_p(a_n b_n)$, para $r > 0$.

A proposição seguinte estabelece que se $X_n = O_p(a_n)$ e $a_n \rightarrow 0, n \rightarrow +\infty$, então $X_n = o_p(1)$.

Proposição A.0.2 *Seja $\{X_n\}$ uma sucessão de variáveis aleatórias tais que $X_n = O_p(r_n)$, onde $0 < r_n \rightarrow 0$. Então, $X_n = o_p(1)$, isto é, $X_n \xrightarrow{P} 0$.*

Prova: Seja $\varepsilon > 0$ fixado arbitrariamente. Como $X_n = O_p(r_n)$, existe $M_\varepsilon \in (0, \infty)$ tal que $P(r_n^{-1}|X_n| > M_\varepsilon) < \varepsilon$, para todo o n . Agora, tomemos $n_0(\varepsilon)$ suficientemente grande, tal que para $n > n_0(\varepsilon)$, se tenha $r_n M_\varepsilon < \delta$, para algum $\delta > 0$. Então, temos, para $n > n(\varepsilon)$,

$$P(|X_n| > \delta) \leq P(|X_n| > M_\varepsilon r_n) < \varepsilon.$$

□

Definição A.0.3 *Seja $\{X_n\}$ uma sucessão de variáveis aleatórias, com funções de distribuição F_{X_n} . Diz-se que X_n converge em distribuição para X se,*

$$\lim_{n \rightarrow \infty} F_{X_n}(x) = F_X(x),$$

para todo o $x \in C$ onde C é o conjunto de pontos de continuidade de F_X .

Usa-se a notação $X_n \xrightarrow{d} X$ para a convergência em distribuição.

Proposição A.0.3 *Suponhamos que $X_n \xrightarrow{d} X$. Então $X_n = O_p(1)$.*

Prova: Para $\varepsilon > 0$ escolhe-se M_0 suficientemente grande tal que $P(|X| > M_0) < \varepsilon$. Agora, pela convergência em distribuição sabemos que $P(|X_n| > M_0) \rightarrow P(|X| > M_0), n \rightarrow \infty$. Assim, existe algum n_0 tal que para $n \geq 0$, se tem $P(|X_n| > M_0) < \varepsilon$. Seguidamente, escolhe-se M_1 suficientemente grande tal que $P(|X_i| > M_1) < \varepsilon$, para $i = 1, \dots, n_0 - 1$. Agora, para $M = \max(M_0, M_1)$, temos $P(|X_n| > M) < \varepsilon, \forall n$. □

Lema A.0.3 Se $X_n \xrightarrow{d} X, n \rightarrow +\infty$, e g é uma função real contínua, então $g(X_n) \xrightarrow{d} g(X), n \rightarrow +\infty$.

Lema A.0.4 Seja $\{X_n\}$ uma sucessão de variáveis aleatórias. Suponhamos que $X_n \xrightarrow{P} X$. Então,

1. $E [|e^{itX_n} - e^{itX}|] \rightarrow 0, \quad n \rightarrow \infty, \forall n \in \mathbb{R}.$
2. $X_n \xrightarrow{d} X.$

Proposição A.0.4 Seja $\{X_n\}$ uma sucessão de variáveis aleatórias reais tais que $X_n \xrightarrow{d} c$, onde c é uma constante. Então, temos que $X_n \xrightarrow{P} c$.

Prova: Seja $\delta > 0$, arbitrariamente fixo. Note-se que,

$$P(|X_n - c| > \delta) \leq P(X_n - c > \delta) + P(X_n - c \leq -\delta) = 1 - F_{X_n}(c + \delta) + F_{X_n}(c - \delta) \rightarrow 0, \quad n \rightarrow \infty.$$

□

Proposição A.0.5 Sejam $\{X_n\}$ e $\{Y_n\}$ sucessões de variáveis aleatórias reais. Se

1. $X_n - Y_n = o_p(1)$ e $X_n \xrightarrow{d} X$, então $Y_n \xrightarrow{d} X$;
2. $X_n = Y_n(1 + o_p(1))$ e $X_n \xrightarrow{d} X$, então $Y_n \xrightarrow{d} Y$.

Prova: Vamos provar que

$$|\Phi_{Y_n}(t) - \Phi_{X_n}(t)| \rightarrow 0, \quad n \rightarrow \infty, \quad t \in \mathbb{R}.$$

Isto resulta de

$$\begin{aligned} |\Phi_{Y_n}(t) - \Phi_{X_n}(t)| &= |E[e^{itY_n}] - E[e^{itX_n}]| \\ &= |E[e^{itY_n} - e^{itX_n}]| \\ &\leq E[|e^{itY_n} - e^{itX_n}|] \\ &= E[|e^{itY_n}| |1 - e^{it(X_n - Y_n)}|] \\ &= E[|1 - e^{it(X_n - Y_n)}|] \rightarrow 0, \quad n \rightarrow \infty, \end{aligned}$$

onde a última conclusão vem do Lema A.0.4, pois $X_n - Y_n \xrightarrow{P} 0$. Então, para qualquer $t \in \mathbb{R}$,

$$|\Phi_{Y_n}(t) - \Phi_Y(t)| \leq |\Phi_{Y_n}(t) - \Phi_{X_n}(t)| + |\Phi_{X_n}(t) - \Phi_X(t)| \rightarrow 0, \quad n \rightarrow \infty,$$

o que finaliza a demonstração da primeira parte. Para provar a segunda parte vejamos que $X_n = Y_n(1 + o_p(1))$ e $Y_n = O_p(1)$ implicam $X_n - Y_n = Y_n o_p(1) = O_p(1) o_p(1) = o_p(1)$. □

Anexo B

Códigos R

B.1 Código Triplo Salto feminino

Libraries

```
library(readxl)
library(latex2exp)
library(Rcmdr)
```

```
# leitura do ficheiro Excel com os dados
marcas <- readXL("C:/Users/Admin/Desktop/TJ_fem.xlsx",rownames=FALSE, header=TRUE, na="",
sheet="Folha1", stringsAsFactors=FALSE)
View(marcas)
```

```
# Introduzir as marcas num vetor x
x <- marcas[, "Mark"]
x
```

Estimadores de γ

Estimador de Pickands

```
# Estimador de Pickands
gammahat_pickands<-function(k){
  1/log(2)*log((x[700-floor((k+1)/4)+1]-x[700-2*floor((k+1)/4)+1])/(x[700-2*floor((k+1)/4)+1]-
x[700-4*floor((k+1)/4)+1]))}
xp<-seq(20,699,1)
P_est<-sapply(xp,gammahat_pickands)
pdf('estimador_Pickands.pdf')
plot(P_est,type="l",xlab="k",ylab=expression(hat(gamma)[list(P)]),main=TeX(r'(Estimador de  $\gamma$ '),
```

```
bold=TRUE),mgp=c(2.2,1,0))
abline(h=0, col="green")
dev.off()
```

Estimador dos Momentos

```
# Estimador dos Momentos
Mnr<-function(k,r) {
  y<-NULL
  for(j in 1:k) {y[j]<-(log(x[700-j+1])-log(x[700-k]))^r}
  (1/k)*sum(y)}

gammahat_momentos<-function(k) Mnr(k,1)+0.5*(1-(Mnr(k,2)/(Mnr(k,1))^2-1)^-1)
xm<-seq(10,699,1)
M_est<-sapply(xm,gammahat_momentos)
pdf('estimador_momentos.pdf')
plot(M_est,type="l",xlab="k",ylab=expression(hat(gamma)[list(M)]),main=TeX(r'(Estimador de $\gamma$)',
bold=TRUE),mgp=c(2.2,1,0))
abline(h=0,col="green")
dev.off()
```

Estimador de γ_-

```
# Estimador Gamma Menos
gammahat_menos<-function(k) 1-0.5*(1-((Mnr(k,1))^2)/Mnr(k,2))^-1

x_menos<-seq(1,699,1)
Estimador_Gamma_menos<-sapply(x_menos,gammahat_menos)
pdf('estimador_gama_menos.pdf')
plot(Estimador_Gamma_menos,type="l",xlab="k",ylab=expression(hat(gamma)[list("-)]),main=TeX(r'(Estimador
de $\gamma$)', bold=TRUE),mgp=c(2.2,1,0))
abline(h=0,col="green")
dev.off()
```

Estimador de Mixed Moments

```
# Estimador MM
Lnk<-function(k) {
  y<-NULL
  for(j in 1:k) { y[j]<-(1-(x[700-k]/x[700-j+1]))}

  (1/k)*sum(y)}
```

```

phi<-function(k) (Mnr(k,1)-Lnk(k))/(Lnk(k))^2
gammahat_mm<-function(k) (phi(k)-1)/(1+2*min(phi(k)-1,0))
xmm<-seq(40,699,1)
MM_est<-sapply(xmm,gammahat_mm)
pdf('estimador_gama_mm.pdf')
plot(MM_est,type="l",xlab="k",ylab=expression(hat(gamma)[list(MM)]),main=TeX(r'(Estimador de
 $\hat{\gamma}$ '), bold=TRUE),mgp=c(2.2,1,0))
abline(h=0,col="green")
dev.off()

```

Teste de Greenwood

```

# Teste1 Gamma=0
TMnr<-function(k,r) {
  y<-NULL
  for(j in 1:k) {y[j]<-(x[700-j+1]-x[700-k])^r}
  (1/k)*sum(y)
}

G<-function(k) TMnr(k,2)/(TMnr(k,1))^2
EstaTeste1<-function(k) (G(k)-2)*(sqrt(k/4))
xtm<-seq(10,699,1)
Tteste1<-sapply(xtm,EstaTeste1)
pdf('teste1_gama=zero.pdf')
plot(xtm,Tteste1,type="l",xlab="k",ylab=expression(G[list(n,k)]), main=TeX(r'(Teste Greenwood  $\hat{\gamma}=0$ '),
bold=TRUE),mgp=c(2.2,1,0))
abline(h=-1.645, col="red")
abline(v=165, col="blue")
text(210,-3.8,"k=165",cex=0.9,col="blue")
text(600,-1.4,TeX(r'( $z_{0.05}$ )=-1.645'),bold=TRUE),cex=0.9,col="red")
dev.off()

```

Teste de Hasofer e Wang

```

# Teste2 Gamma=0
H<-function(k) 1/(k*(G(k)-1))
EstaTeste2<-function(k) (k*H(k)-1)*(sqrt(k/4))
pvalor2<-function(k) 1-pnorm(EstaTeste2(k))
xtm<-seq(20,699,1)
Tteste2<-sapply(xtm,EstaTeste2)
pdf('teste2_gama=zero_hasoferwang.pdf')
plot(xtm,Tteste2,type="l",xlab="k",ylab=expression(HW[list(n,k)]), main=TeX(r'(Teste Hasofer &

```

```

Wang  $\gamma=0$ '), bold=TRUE),mgp=c(2.2,1,0))
abline(h=1.645, col="red")
abline(v=160, col="blue")
text(200,1.2,"k=160",cex=0.9,col="blue")
text(580,1.9,TeX(r'( $z_{0.95}$ )=1.645)'),bold=TRUE),cex=0.9,col="red")
dev.off()

```

Teste à finitude de x_F

```

# Teste da finitude do extremo superior do suporte
aux <- function(k) Mnr(k,1)^2/Mnr(k,2)
ank <- function(k) x[700-k]*0.5*Mnr(k,1)*(1-aux(k))^-1
Tn1<-function(k) {
  y<-NULL
  for(j in 1:k) {y[j]<-(x[700-j]-x[700-k]-ank(k))/(x[700]-x[700-k])}
  (1/k)*sum(y)
}
Tn1A <- function(k) sqrt(k)*log(k)*Tn1(k)
u<-seq(20,699,1)
TestFin<-sapply(u,Tn1A)
pdf('teste_finitude_xF.pdf')
plot(u,TestFin,type="l",xlab="k",ylab=expression(E[list(n,k)]),main=TeX(r'(Teste à finitude de  $x_{\{F\}}$ '),
bold=TRUE),mgp=c(2.2,1,0))
abline(h=-1.96, col="red")
abline(h=1.96, col="red")
abline(v=103, col="blue")
text(140,-9.5,"k=103",cex=0.9,col="blue")
text(600,-2.35,TeX(r'( $z_{0.025}$ )=-1.96)'),bold=TRUE),cex=0.9,col="red")
dev.off()

```

Estimadores de x_F

x_F com γ_-

Extremo superior do Suporte 1

```
XFhat1<-function(k)
```

```

  x[700-k]-(x[700-k]*Mnr(k,1)*(1-gammahat_menos(k)))/gammahat_menos(k)

```

```
xF1<-seq(100,699,1)
```

```
xFplot1<-sapply(xF1,XFhat1)
```

```
pdf('xF1.pdf')
```



```
plot(xF1,xFplot1,type="l",xlab="k",ylab=expression(hat(x[list(F)])),main=TeX(r'(Estimador de  $x_{\{F\}}$ '),
bold=TRUE),mgp=c(2.2,1,0))
abline(h=max(x),col="magenta")
dev.off()
```

x_F **com** γ_M

Extremo superior do Suporte2

```
XFhat3<-function(k)
```

```
  x[700-k]-(x[700-k]*Mnr(k,1)*(1-gammahat_menos(k)))/gammahat_momentos(k)
```

```
xF2<-seq(1,699,1)
```

```
xFplot2<-sapply(xF2,XFhat2)
```

```
pdf('xF2.pdf')
```

```
plot(xF2,xFplot2,type="l",xlab="k",ylab=expression(hat(x[list(F)])),main=TeX(r'(Estimador de  $x_{\{F\}}$ '),
```

```
bold=TRUE),mgp=c(2.2,1,0))
```

```
abline(h=max(x),col="magenta")
```

```
dev.off()
```

x_F **com** γ_-

Extremo superior do Suporte3

```
XFhat3<-function(k)
```

```
  x[700-k]-(x[700-k]*Mnr(k,1)*(1-gammahat_menos(k)))/gammahat_mm(k)
```

```
xF3<-seq(1,699,1)
```

```
xFplot3<-sapply(xF3,XFhat3)
```

```
pdf('xF3.pdf')
```

```
plot(xF3,xFplot3,type="l",xlab="k",ylab=expression(hat(x[list(F)])),main=TeX(r'(Estimador de  $x_{\{F\}}$ '),
```

```
bold=TRUE),mgp=c(2.2,1,0))
```

```
abline(h=max(x), col="magenta")
```

```
dev.off()
```

x_F **de FA**

Extremo superior do Suporte 4 - Artigo Fraga Alves

```
XFhat4A<-function(k) {
```

```
  y<-NULL
```

```
  for(j in 0:k-1) {y[j]<-(x[700-k-j])*log(1+1/(j+k))}
```

```
  (1/log(2))*sum(y)
```

```
}
```

```
XFhat4<-function(k) x[700]+x[700-k]-XFhat4A(k)
```

```
xF4<-seq(1,350,1)
```

```
xFplot4<-sapply(xF4,XFhat4)
```

```
pdf('xF4.pdf')
plot(xF4,xFplot4,type="l",xlab="k",ylab=expression(hat(x[list(F)])),main=TeX(r'(Estimador de  $\hat{x}_{\{F\}}$ '),
bold=TRUE),mgp=c(2.2,1,0))
abline(h=max(x),col="magenta")
dev.off()
```

Teste $\gamma = 0$ - Fraga Alves [8]

```
# Teste Gamma=0 Fraga Alves2017
TesteFA<-function(k) (XFhat4(k)-x[700-k])/(x[700-k]-x[700-2*k])
EstTesteFA<-function(k) (log(2))*TesteFA(k)-(log(k)+0.5*log(k))
En<-seq(10,350,1)
ETn<-sapply(En,EstTesteFA)
pdf('Teste_gamma=0_Fraga_Alves2017.pdf')
plot(En,ETn,type="l",xlab="k",ylab=expression(G[list(n,k)]^{"*"}),main=TeX(r'(Teste FA)'), bold=TRUE),
mgp=c(2.2,1,0))
abline(h=-1.0972,col="red")
text(270,-0.5,TeX(r'( $\hat{g}_{\{0.05\}}=-1.0972$ )'),
bold=TRUE),cex=0.9,col="red")
dev.off()
```

Estimação de x_F e γ com k ótimo

```
kopt=165 #TJ female
#Kopt=223 #TJ male
#kopt=310 #LJ female
```

```
gammahat_pickands(kopt)
gammahat_momentos(kopt)
gammahat_menos(kopt)
gammahat_mm(kopt)
```

```
XFhat1(kopt)
XFhat2(kopt)
XFhat3(kopt)
XFhat4(kopt)
```

B.2 Código Cubo de Rubik

```
library(extRemes)
library(evd)
library(fExtremes)
library(Rcmdr)
library(readxl)
library(latex2exp)

# Leitura do ficheiro excel
# Diretorio do ficheiro excel no meu PC marcas <- readXL("C:/Users/Admin/Desktop/cubo
magico.xlsx",rownames=FALSE, header=TRUE, na="", sheet="Sheet1", stringsAsFactors=FALSE)
View(marcas)

x<- marcas[, "Simmetric_Mark"]
x

#pdf('histograma.pdf')
hist(x,prob=T,ylim=c(0,3),xlab="x",main="GEV",ylab="Freq",font=2,family="serif",font.lab=2,
cex.lab=1.5,col='skyblue')
#dev.off()

fit1 = fevd(x,type="GEV")
fit1 #-> (Usa estimadores MV)

#png(file='fit1.png',width=1000, height=1000,res=100)
plot(fit1)
#dev.off()

out1<-gevFit(x,type="pwm")
out1

#Quantis de probabilidade para ambos os métodos de ordem 0.95,0.98,0.99

qevd(0.99,-4.8981079,0.1119791,0.4756207,type="GEV")
qevd(0.98,-4.8981079,0.1119791,0.4756207,type="GEV")
qevd(0.95,-4.8981079,0.1119791,0.4756207,type="GEV")

qevd(0.99,-4.8868204,0.1300748,0.2685176,type="GEV")
qevd(0.98,-4.8868204,0.1300748,0.2685176,type="GEV")
qevd(0.95,-4.8868204,0.1300748,0.2685176,type="GEV")
```

Teste de Gumbel

```
n<-100
GS<-(max(x)-x[floor(n/2)+1])/(x[floor(n/2)+1]-min(x))
bn<-(log(n)+log(log(2)))/(log(log(n))-log(log(2)))
an<-1/(log(log(n)))
Gstar<-(GS-bn)/an
Gstar
```