



SOCIEDADE PORTUGUESA  
DE ESTATÍSTICA

# ESTATÍSTICA: A CIÊNCIA DA INCERTEZA

ATAS DO XXI CONGRESSO DA SOCIEDADE PORTUGUESA DE ESTATÍSTICA

**Editores:**

Isabel Pereira • Adelaide Freitas • Manuel Scotto  
Maria Eduarda Silva • Carlos Daniel Paulino



SOCIEDADE PORTUGUESA  
DE ESTATÍSTICA

**ESTATÍSTICA:**  
**A CIÊNCIA DA INCERTEZA**

Atas do XXI Congresso Anual da  
Sociedade Portuguesa de Estatística

Aveiro, 29 de novembro a 2 de dezembro de 2013

**Editores**

*Isabel Pereira*

*Adelaide Freitas*

*Manuel Scottó*

*Maria Eduarda Silva*

*Carlos Daniel Paulino*

Dezembro, 2014

**Edições SPE**

© 2014, Sociedade Portuguesa de Estatística

**Editores:** Isabel Pereira, Adelaide Freitas, Manuel Scotto, Maria Eduarda Silva, Carlos Daniel Paulino

**Título:** Estatística: A Ciência da Incerteza.  
Atas do XXI Congresso da Sociedade Portuguesa de Estatística

**Editora:** Sociedade Portuguesa de Estatística

**Conceção Gráfica da Capa:** Carina Sousa

**Impressão:** Instituto Nacional de Estatística

**Tiragem:** 250 Exemplares

**ISBN:** 978-972-8890-35-3

**Depósito Legal:** 385476/14

## Índice

<b>Transectos lineares em ungulados de montanha: um estudo de simulação</b>	<b>1</b>
<i>Anabela Afonso, Russell Alpizar-Jara e Jesús M. Pérez</i>	
<b>Da utilização de cadeias de Markov multivariadas enquanto regressores num problema de previsão</b>	<b>11</b>
<i>Bruno Damásio e João Nicolau</i>	
<b>Modelos de cura aplicados ao cancro da mama</b>	<b>25</b>
<i>Carina Alves e Ana Maria Abreu</i>	
<b>Uma abordagem não paramétrica à previsão da dose individualizada de <i>atracurium</i></b>	<b>35</b>
<i>Conceição Rocha, Maria Eduarda Silva e Teresa Mendonça</i>	
<b>Aplicação do método dos excessos de nível a valores extremos de precipitação na ilha da Madeira</b>	<b>43</b>
<i>Délia Gouveia-Reis, Luiz Guerreiro Lopes e Sandra Mendonça</i>	
<b>Propriedade de Taylor em processos autorregressivos</b>	<b>51</b>
<i>Esmeralda Gonçalves, Cristina M. Martins e Nazaré Mendes-Lopes</i>	
<b>Propriedade de Taylor no modelo TGARCH(1,1)</b>	<b>65</b>
<i>Esmeralda Gonçalves, Joana Leite e Nazaré Mendes-Lopes</i>	
<b>Modelos GARCH de valores inteiros associados a leis infinitamente divisíveis</b>	<b>77</b>
<i>Esmeralda Gonçalves, Nazaré Mendes-Lopes e Filipa Silva</i>	
<b>Metodologias estatísticas para estudo da interacção</b>	<b>89</b>

**genótipo×ambiente em clones de videira***Elsa Gonçalves e Antero Martins*

**Análise bayesiana semiparamétrica de resposta binária com covariável contínua sujeita a omissão não aleatória** 105  
*Frederico Z. Poletto, Carlos Daniel Paulino, Julio M. Singer e Geert Molenberghs*

**Dependência extremal: risco de contágio de valores extremos** 119  
*Helena Ferreira e Marta Ferreira*

**Estimação do índice de valores extremos em ambiente R - as abordagens paramétrica e semi-paramétrica** 129  
*Helena Penalva, Sandra Nunes e Manuela Neves*

**Diagnóstico em regressão binária** 141  
*Isabel Natário e Sílvia Shruballs*

**Distâncias de Mahalanobis, variáveis originais e componentes principais** 155  
*Jorge Cadima*

**Generalized linear models, generalized additive models and generalized estimating equations to capture-recapture closed population models** 169  
*Md. Abdus Salam Akanda e Russell Alpizar-Jara*

**Intervalos de amostragem adaptativos inicialmente predefinidos para um risco cumulativo constante** 183  
*Manuel do Carmo, Paulo Infante e Jorge M. Mendes*

**Modelo Bayesiano de equações simultâneas para a estimação dos parâmetros da área basal e da mortalidade** 197

*Marco Marto, Isabel Pereira e Margarida Tomé*

**Análise da fiabilidade de centros de maquinação -  
um caso de estudo** 205

*Maria João Dias, Adelaide Freitas e Constantino Pinto*

**Sobrevivência a longo prazo de doentes com cancro do  
cólon e do reto** 213

*Mariana Rodrigues, Carina Alves e Ana Maria Abreu*

**On the protection of  $\alpha$ -thalassaemia from malaria  
infection in northeast Tanzania** 223

*Nuno Sepúlveda, Alphaxard Manjurano, Chris J Drakeley e  
Taane G Clark*

**Porque duram tanto tempo algumas dissertações de  
Mestrado?** 235

*Rita Freitas, Paulo Infante, Gonçalo Jacinto,  
Fernanda Figueiredo e João Dias*

**Sobrevivência relativa do cancro colo-rectal e do  
estômago no sul de Portugal** 245

*Ricardo São João, Ana Luisa Papoila e Ana Miranda*

**Um estudo de simulação para avaliar a performance de  
estimadores para a taxa de prevalência usando testes  
compostos** 253

*Ricardo Sousa, Rui Santos e João Paulo Martins*

**Medidas para avaliar a utilização de testes compostos** 267

*Rui Santos, João Paulo Martins e Miguel Felgueiras*

**Modelação de grandes incêndios em Portugal** 279

*Alexandra Ramos*

## Estimação do índice de valores extremos em ambiente R - as abordagens paramétrica e semi-paramétrica

Helena Penalva

Escola Superior de Ciências Empresariais do Instituto Politécnico de Setúbal e CEAUL, *helen.penalva@esce.ips.pt*

Sandra Nunes

Escola Superior de Ciências Empresariais do Instituto Politécnico de Setúbal, CEAUL e CMA/FCT/UNL, *sandra.nunes@esce.ips.pt*

Manuela Neves

Instituto Superior de Agronomia e CEAUL, Universidade de Lisboa, *manela@isa.ulisboa.pt*

**Palavras-chave:** Estimação paramétrica e semi-paramétrica, índice de valores extremos, *software* R, teoria de valores extremos

**Resumo:** Este trabalho tem como objectivos principais: ilustrar a utilização do *software* R numa análise de dados de valores extremos e estimação do índice de valores extremos,  $\xi$ , utilizando as abordagens paramétrica e semi-paramétrica; fazer uma discussão dos resultados obtidos por diferentes procedimentos em cada uma daquelas abordagens. A ilustração foi feita num conjunto de dados de níveis médios diários de caudais de um rio, registados durante 50 anos.

## 1 Introdução

A Teoria de Valores Extremos (EVT, do inglês “*Extreme Value Theory*”) estuda acontecimentos que poderão ser mais extremos do que aqueles que alguma vez já foram observados. Tem vindo a afirmar-se como uma das áreas da Estatística de grande relevo em várias ciências aplicadas onde se observa a ocorrência de valores extremos que é necessário modelar. Nesta modelação é necessário estimar pa-

râmetros de entre os quais tem primordial relevância o parâmetro de forma,  $\xi$ , que descreve o comportamento da cauda direita,  $1 - F$ , do modelo subjacente aos dados. A sua estimação precisa é muito importante e de enorme influência na estimação de outros parâmetros de valores extremos, tais como *quantis elevados* ou *período de retorno de quantis elevados*. Pretende-se utilizar e construir, quando não exista, procedimentos em R para estimar  $\xi$ .

## 2 Resultados preliminares

Os estudos em teoria assintótica de valores extremos tiveram início com Fréchet [8], mas ficaram definitivamente estabelecidos com Gnedenko [9] ao formular que a distribuição limite do máximo de variáveis aleatórias (v.a.'s) independentes e identicamente distribuídas (i.i.d.), convenientemente normalizado, se existir, é do tipo,

$$EV_{\xi}(x) = \begin{cases} \exp[-(1 + \xi x)^{-1/\xi}], & 1 + \xi x > 0, \text{ se } \xi \neq 0; \\ \exp[-\exp(-x)], & x \in \mathbb{R}, \text{ se } \xi = 0. \end{cases} \quad (1)$$

O parâmetro de forma,  $\xi$ , é usualmente designado por *índice de valores extremos* e a sua estimação é da maior importância em EVT. A função de distribuição (f.d.)  $EV_{\xi}$  incorpora as três leis limite: a lei de Gumbel:  $\Lambda(x) = \exp(-\exp(-x)) = EV_0(x)$ ,  $x \in \mathbb{R}$ ,  $\xi = 0$ , com cauda direita do tipo exponencial; a lei de Fréchet:  $\Phi_{\xi}(x) = \exp(-x^{-1/\xi}) = EV_{\xi}(\frac{x-1}{\xi})$ ,  $x > 0$ ,  $\xi > 0$  com cauda direita pesada, do tipo exponencial negativo; a lei de Weibull:  $\Psi_{\xi}(x) = \exp(-(-x)^{1/\xi}) = EV_{\xi}(\frac{-x-1}{\xi})$ ,  $x < 0$ ,  $\xi < 0$  com cauda direita curta. A f.d.  $EV_{\xi}$  pode também incluir parâmetros de localização,  $\lambda \in \mathbb{R}$ , e de escala,  $\delta > 0$ , denotando-s por  $EV_{\xi}(x; \lambda, \delta) = EV_{\xi}((x - \lambda)/\delta)$ . Em vez se pensar no máximo de uma amostra, Balkema and de Haan [1] e Pickands [19] propuseram considerar todos os valores que excedem um nível elevado  $u$ . Mostraram que, nas condições de existência de (1), a f.d. condicional dos excessos do nível  $u$ , quando  $u \rightarrow \infty$ , i.e. de  $Y = (X - u)|X > u$  é bem aproximada pela Generalizada de Pareto (GP), cuja f.d. é definida como,

$H_\xi(y) = 1 - (1 + \xi y/\tilde{\delta})^{-1/\xi}$ ,  $\{y : y > 0 \text{ e } (1 + \xi y/\tilde{\delta}) > 0\}$ , onde o parâmetro de forma,  $\xi$ , é o mesmo em ambas as distribuições, mas relativamente ao parâmetro de escala tem-se  $\tilde{\delta} = \delta + \xi(u - \lambda)$ .

### 3 Estimação

A estimação em EVT começou por ser realizada numa abordagem paramétrica, baseada na propriedade de max-estabilidade. Nesta abordagem, por limitações de espaço, faremos apenas referência à metodologia de Gumbel conhecida como *Modelo dos Máximos Anuais* (MMA) e à metodologia dos excessos acima de um limiar, conhecida como POT (do inglês *‘Peaks over Threshold’*). Em cada uma das metodologias a estimação dos parâmetros foi feita usando o método de máxima verosimilhança (ML) do inglês *‘Maximum Likelihood’* e o método dos momentos ponderados de probabilidade (PWM) do inglês *‘Probability Weighted Moments’*.

Na década de 70 os procedimentos de estimação começaram a ser efectuados considerando uma abordagem semi-paramétrica, na qual não se adopta um modelo limite, mas se admite apenas que a função  $F$  está no domínio de atracção da  $EV_\xi$ , para um valor  $\xi$  adequado. A estimação de  $\xi$  é, nestes procedimentos, baseada nas  $k$  maiores estatísticas ordinais, em que se admite  $k \rightarrow \infty$  e  $k/n \rightarrow 0$  quando  $n \rightarrow \infty$ . Nesta abordagem vários estimadores têm sido propostos. Para ilustrar a utilização do *software* R vamos referir apenas os estimadores clássicos de Hill [15], de Pickands [19] e dos Momentos, Dekkers *et al.* [6], e ainda um estimador mais recente, de viés reduzido e com variância mínima, Caeiro *et al.* [3].

Sendo  $X_{1:n} \leq X_{2:n} \leq \dots \leq X_{n:n}$  as estatísticas ordinais associadas à amostra  $(X_1, X_2, \dots, X_n)$ , estes estimadores são assim definidos:

$$\text{Hill } (\xi > 0) \quad \hat{\xi}_{k,n}^H := \frac{1}{k} \sum_{i=1}^k \ln(X_{n-i+1:n}) - \ln(X_{n-k:n}) \quad (2)$$

$$\text{Pickands } (\xi \in \mathbb{R}) \quad \hat{\xi}_{k,n}^P := \frac{1}{\ln 2} \ln \left( \frac{X_{n-k+1:n} - X_{n-2k+1:n}}{X_{n-2k+1:n} - X_{n-4k+1:n}} \right) \quad (3)$$

$$\text{Momentos } (\xi \in \mathbb{R}) \quad \widehat{\xi}_{k,n}^M := M_{k,n}^{(1)} + 1 - \frac{1}{2} \left( 1 - \frac{(M_{k,n}^{(1)})^2}{M_{k,n}^{(2)}} \right)^{-1} \quad (4)$$

onde  $M_{k,n}^{(r)} := \frac{1}{k} \sum_{i=1}^k [\ln(X_{n-i+1:n}) - \ln(X_{n-k:n})]^r$ .

A classe de estimadores de viés reduzido e com variância mínima, MVRB (do inglês *minimum-variance reduced-bias*), introduzida por Caeiro, et al. [3] revelou um desempenho superior ao dos estimadores clássicos, em contexto de caudas pesadas ( $\xi > 0$ ). Esta classe de estimadores tem a forma funcional

$$\overline{H}(k) \equiv \overline{H}_{\hat{\beta}, \hat{\rho}}(k) := H(k)(1 - \hat{\beta}(n/k)^{\hat{\rho}}/(1 - \hat{\rho})), \quad (5)$$

com  $H(k) \equiv \widehat{\xi}_{k,n}^H$  o estimador de Hill, e  $\hat{\beta}$  e  $\hat{\rho}$  estimadores consistentes dos parâmetros de escala e forma de segunda ordem,  $\beta$  e  $\rho$ , ver Caeiro et al. [3] e Gomes et al. [14], para mais detalhes. Sobre estimação de viés reduzido podem referir-se os trabalhos de Gomes et al. [12] e Beirlant et al. [2], entre outros.

Para a estimação de  $\rho$  consideraremos um membro particular da classe dos estimadores de  $\rho$  introduzidos em Fraga Alves, et al. [7]. Esta classe, parametrizada num parâmetro de controlo  $\tau \in \mathbb{R}$  que aqui tomaremos como  $\tau = 0$ , ver Gomes et al. [14], é definida como:

$$\hat{\rho}(k) \equiv \hat{\rho}_0(k) := \min \left( 0, \frac{3(T_n^{(0)}(k)-1)}{T_n^{(0)}(k)-3} \right), \text{ sendo, } T_n^{(0)}(k) \text{ assim definido,}$$

$$T_n^{(0)}(k) := \left[ \ln \left( M_{k,n}^{(1)} \right) - \frac{1}{2} \ln \left( M_{k,n}^{(2)}/2 \right) \right] / \left[ \frac{1}{2} \ln \left( M_{k,n}^{(2)}/2 \right) - \frac{1}{3} \ln \left( M_{k,n}^{(3)}/6 \right) \right],$$

com  $M_{k,n}^{(j)}(k)$ ,  $j = 1, 2, 3$ , definido acima. Para a estimação do parâmetro de escala de segunda-ordem,  $\beta$ , vamos considerar  $\hat{\beta}_{\hat{\rho}}(k) := \left( \frac{k}{n} \right)^{\hat{\rho}} [d_{\hat{\rho}}(k) D_0(k) - D_{\hat{\rho}}(k)] / [d_{\hat{\rho}}(k) D_{\hat{\rho}}(k) - D_{2\hat{\rho}}(k)]$ , com  $\hat{\rho} = \hat{\rho}_0(k)$ , e para  $\alpha \leq 0$ ,  $d_{\alpha}(k) := \frac{1}{k} \sum_{i=1}^k \left( \frac{i}{k} \right)^{-\alpha}$  e  $D_{\alpha}(k) := \frac{1}{k} \sum_{i=1}^k \left( \frac{i}{k} \right)^{-\alpha} U_i$ , com  $U_i := i [\ln(X_{n-i+1:n}/X_{n-i:n})]$ ,  $1 \leq i \leq k$ .

De modo a não termos um aumento de variância, no estimador  $\overline{H}(k)$  devem considerar-se os estimadores  $\hat{\rho}_0(k)$  e  $\hat{\beta}_{\hat{\rho}}(k)$ , calculados em  $k = k_1$ , com  $k_1 = \lfloor n^{1-\epsilon} \rfloor$ ,  $\epsilon = 0.001$ , ver Gomes e Martins [11], Gomes et al. [12] e Caeiro et al. [5], para mais detalhes. Estimadores

alternativos para  $\beta$  podem ser vistos em Caeiro e Gomes [4] e Gomes *et al.* [13].

## 4 Aplicação do *software* R na modelação de dados e na estimação de $\xi$

O software R, linguagem de código aberto para computação estatística e tratamento de dados, possui vários *packages* estatísticos e permite a implementação pelos utilizadores de *packages* adicionais. Para a modelação e estimação em EVT, quer na abordagem paramétrica quer na semi-paramétrica, podemos referir alguns *packages* que o *software* R possui: `evd`; `ismev`; `evir`; `fExtremes`; `POT`; `evd-bayes`; `copula`; `SpatialExtremes`.

Na ilustração que iremos apresentar serão considerados dados de níveis médios diários do caudal ( $m^3/s$ ) do rio Paiva, medidos na estação hidrométrica de Fragas da Torre, entre 1946/47 a 1995/96. Os 50 anos observados correspondem exatamente ao período entre 1 de Outubro de 1946 e 30 de Setembro de 1996. O estudo do fluxo deste rio é extremamente importante uma vez que é uma das alternativas ao rio Douro como fonte de abastecimento de água para a região sul do Porto.

Gomes [10] utilizou os dados relativos aos meses entre Novembro e Fevereiro. Contudo, como o interesse é analisar os valores extremos, verificámos que na maioria dos anos, os meses de Março e Abril apresentaram valores muito elevados de caudal, pelo que decidimos considerar aqueles meses na amostra a estudar. Ficamos então com uma amostra de 6 meses (em cada ano) de dados diários durante 50 anos, num total de 9050 dados. O teste *Augmented Dickey-Fuller* foi aplicado à amostra, tendo-se obtido um *p-value* inferior a 0.01 pelo que iremos admitir a estacionaridade dos dados, hipótese esta também admitida em Gomes [10].

A Figura 1 apresenta o cronograma dos dados em estudo (direita) e o histograma associado (esquerda).

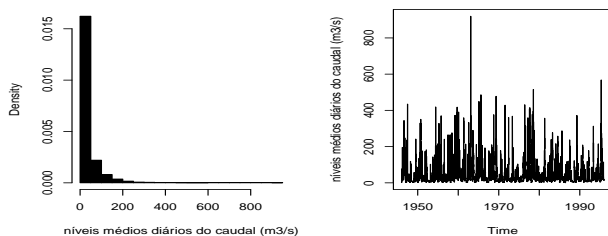


Figura 1: Histograma (esquerda) e cronograma (direita).

As principais medidas descritivas, das quais referimos a  $Skewness=4.12$  e  $Kurtosis=27.60$ , obtidas com recurso ao *package* *fBasics*, conjuntamente com o histograma, indicam uma cauda muito mais pesada que a do modelo de Gauss.

### Abordagem paramétrica

Para utilizarmos a metodologia MMA, é necessário obter os máximos de cada bloco, o que facilmente se consegue com *package* *evir* e o comando `gev(dados, block=181)`, onde  $block=181$ , indica o tamanho de cada bloco. O gráfico da autocorrelação parcial, Figura 2, mostra que os máximos são observações fracamente correlacionadas. Portanto faz sentido ajustar uma distribuição  $EV$  aos máximos anuais. Obtivemos as estimativas ML para a distribuição  $EV_{\xi}$ , os intervalos de confiança de Wald e os intervalos determinados com recurso à log-verosimilhança de perfil para todos os parâmetros, recorrendo ao *package* *evd* com os seguintes comandos no R:

```
fgev(gev(dados, block=181)$data);
confint(fgev(gev(dados, block=181)$data)),level=0.95);
confint(profile(fgev(gev(dados, block=181)$data)),
         level=0.95).
```

O método PWM também foi aplicado para determinar as estimativas

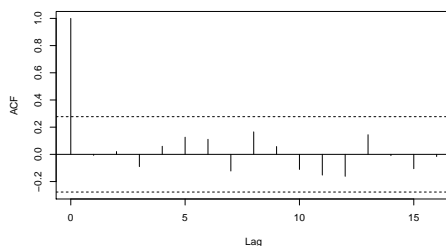


Figura 2: Função de autocorrelação parcial.

dos parâmetros. Neste caso utilizou-se o *package* `fExtremes` e o comando

```
gevFit(dados, block=181, type="pwm").
```

A aplicação da metodologia POT baseia-se na modelação da distribuição de excessos acima de um nível elevado, na qual se restringe a nossa atenção às observações que excedem esse nível. O desafio desta análise é exatamente a escolha deste limiar. Esta escolha pode ser feita a partir do gráfico da vida residual média, e outros dois gráficos, um referente ao parâmetro de escala e outro ao parâmetro de forma, onde se ajusta a distribuição *GP* a um conjunto de limiares.

Os gráficos foram construídos recorrendo ao *package* POT com os comandos : `mrlplot(dados)` e `tcplot(dados, c(100,500))`.

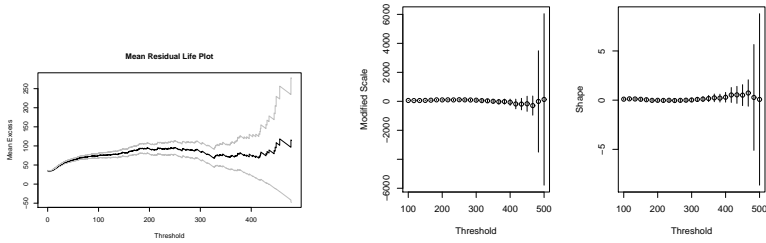


Figura 3: Vida residual média (esquerda) e ajustamento da distribuição  $GP$  a um conjunto de limiares (centro e direita)

A Figura 3. sugere que se considere um limiar perto de 300. As estimativas ML e as estimativas PWM para os parâmetros da distribuição  $GP$  podem ser obtidas utilizando o *package* POT e executando os seguintes comandos:

```
fitgpd(dados, 300, est="mle")
fitgpd(dados, 300, est="pwm")
```

Os valores das estimativas, e respectivos desvios padrão (d.p.), fornecidos pela aplicação dos métodos de estimação referentes à abordagem paramétrica encontram-se resumidos na Tabela 1.

Tabela 1. Estimativas dos parâmetros e intervalo de confiança

Métodos Estimação	$\hat{\xi}$	$\hat{\lambda}$	$\hat{\delta}$
MMA - ML (d.p.)	-0.02 (0.09)	227.35 (19.34)	122.73 (13.87)
I.C. de perfil	(-0.16, 0.21)	(190.53, 267.22)	(99.55, 155.46)
MMA - PWM	-0.08	231.25	129.78
	$\hat{\xi}$		$\hat{\delta}$
POT - ML (d.p.)	0.02 (0.09)		83.24 (12.98)
POT - PWMB (d.p.)	-0.19 (0.16)		100.98 (19.39)

Como o intervalo de confiança para  $\xi$  inclui o zero, não se rejeita a hipótese nula,  $\xi = 0$ . Assim a distribuição Gumbel é uma possível

candidata para modelar os máximos dos dados referentes aos níveis médios diários do rio.

### Abordagem semi-paramétrica

Na estimação semi-paramétrica são consideradas as  $k$  maiores estatísticas ordinais associadas à totalidade das observações, não se definindo nenhum modelo paramétrico, como dissémos. Neste trabalho foram utilizados o estimador de Hill [15], o estimador de Pickands [19], o estimador dos momentos Dekkers *et al.* [6] e o estimador de viés reduzido com variância mínima (MVRB), Caeiro *et al.* [3]. Estes estimadores foram programados em linguagem R. Por questões de espaço não é possível incluir neste trabalho os comandos do R contruídos.

A Figura 4. apresenta as trajectórias das estimativas provenientes daqueles estimadores para  $1 \leq k \leq 5000$ . Verifica-se a dificuldade, já mencionada atrás, da escolha do nível  $k$ . Este assunto, apesar de já abordado por vários autores dos quais citamos Neves *et al.* [16] continua a ser assunto de investigação.

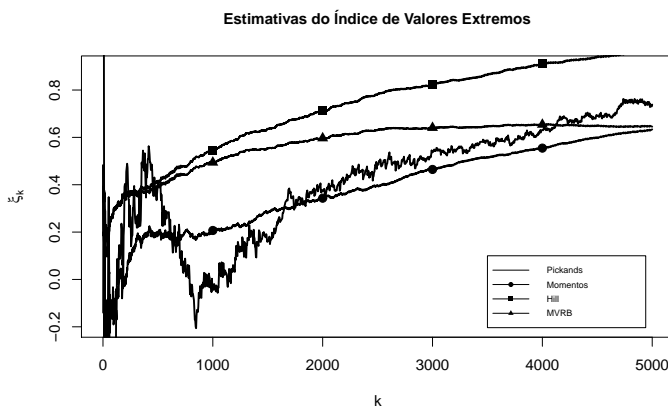


Figura 4: Estimativas de  $\xi$ , com os estimadores de Hill, Pickands, Momentos e MVRB.

**Notas:** Chama-se a atenção para maior estabilidade das estimativas MVRB, como é conhecido da literatura. É no entanto de salientar que, mesmo sem efectuar a escolha de  $k$  correspondente à zona de maior estabilidade, parece haver concordância no valor da estimativa apontada pelos métodos dos momentos e MVRB (e acompanhada pelas estimativas de Pickands, apesar da volatilidade deste estimador). No entanto parece ser apontada como estimativa para  $\xi$  um valor bem superior ao que as abordagens paramétricas obtiveram.

## 5 Comentários e trabalho futuro

Neste trabalho procurou-se continuar a explorar os procedimentos de modelação e inferência de dados de valores extremos com o R, iniciados em trabalhos anteriores, Penalva *et al.* [17] e Penalva *et al.* [18]. Alguns estimadores foram por nós programados em R, havendo outros em fase de programação. As estimativas de  $\xi$  apresentaram valores bastante diferentes consoante as abordagens, o que motiva trabalho futuro de pesquisa destas disparidades.

## Agradecimentos

Investigação parcialmente suportada por fundos nacionais através da FCT-Fundação para a Ciência e a Tecnologia, projectos PEst-OE/MAT/UI006/2011,2014 (CEAUL) e PEst-OE/MAT/UI0297/2011, 2014(CMA/FCT/UNL).

## Referências

- [1] Balkema, A.A., de Haan, L. (1974). Residual life time at great age. *Annals of Probability* 2, 792–804.
- [2] Beirlant, J., Caeiro, F., Gomes, M.I. (2012). An overview and open research topics in the field of statistics of univariate extremes. *REVSTAT - Statistical Journal* 10, 1–31.

- [3] Caeiro, F., Gomes, M.I., Pestana, D. (2005). Direct reduction of bias of the classical Hill estimator. *REVSTAT - Statistical Journal* 3, 111–136.
- [4] Caeiro, F., Gomes, M.I. (2006). A new class of estimators of a “scale” second order parameter. *Extremes* 9, 193–211.
- [5] Caeiro, F., Gomes, M.I., Henriques-Rodrigues, L. (2009). Reduced-bias tail index estimators under a third order framework. *Communications in Statistics - Theory and Methods* 38, 1019–1040.
- [6] Dekkers, A.L.M., Einmahl, J.H.J., de Haan, L. (1989). A moment estimator for the index of an extreme-value distribution. *Annals of Statistics* 17, 1833–1855.
- [7] Fraga Alves, M.I., Gomes M.I., de Haan, L. (2003). A new class of semi-parametric estimators of the second order parameter. *Portugaliae Mathematica* 60, 194–213.
- [8] Frechet, M. (1927). Sur la loi de probabilité de l’écart maximum. *Annales de la Société Polonaise de Mathématique (Cracovie)* 6, 93–116.
- [9] Gnedenko, B.V. (1943). Sur la distribution limite d’une série aléatoire. *Annals of Mathematics* 44, 423–453.
- [10] Gomes, M.I. (1993). on the estimation parameters of rare events in environmental times series. *Statistics for the Environment*, 226–241.
- [11] Gomes, M.I., Martins, M.J. (2002). “Asymptotically unbiased” estimators of the tail index based on external estimation of the second order parameter. *Extremes* 5, 5–31.
- [12] Gomes, M.I., de Haan, L., Henriques-Rodrigues, L. (2008). Tail Index estimation for heavy-tailed models: accommodation of bias in weighted log-excesses. *Journal of the Royal Statistical Society: Series B* 70, 31–52.
- [13] Gomes, M.I., Henriques-Rodrigues, L., Pereira, H., Pestana, D. (2010). Tail index and second order parameters’ semi-parametric estimation based on the log-excesses. *Journal of Statistical Computation and Simulation* 80, 653–666.
- [14] Gomes, M.I., Martins, M.J. and Neves, M.M. (2013) Generalized Jackknife-based estimators for univariate extreme-value modeling. *Communications in Statistics - Theory and Methods* 42, 1227–1245.

- [15] Hill, B.M. (1975). A simple general approach to inference about the tail of a distribution. *Annals of Statistics* 3, 1163–1174.
- [16] Neves, M.M., Gomes, M.I., Figueiredo, F., Prata Gomes, D. (2015). Modeling extreme events: sample fraction adaptive choice in parameter estimation. *Journal of Statistical Theory and Practice* 9, 184–199.
- [17] Penalva, H., Neves, M.M., Nunes, S. (2013). Topics in Data Analysis Using R in Extreme value Theory. *Advances in Methodology and Statistics / Metodoloski zvezki* 10, 17–29.
- [18] Penalva, H., Nunes, S., Neves, M.M. Statistical Modeling and Inference in Extremes: Applications with R. *Biometrie und Medizinische Informatik Greifswalder Seminarberichte, Statistical and Biometrical Challenges, Shaker Verlag*, 281–309.
- [19] Pickands, J. (1975). Statistical inference using extreme order statistics. *Annals of Statistics* 3, 119–131.