

Redução de viés em estimadores do índice de cauda

Frederico Caeiro

Faculdade de Ciências e Tecnologia - Universidade Nova de Lisboa, Departamento de Matemática e C.M.A.

Resumo: Este trabalho tem por objectivo apresentar uma nova classe de estimadores semiparamétricos do índice de cauda positivo, γ , que generalizam alguns dos já conhecidos, incluindo o estimador de Hill. Esta nova classe além de ser consistente para γ têm dois parâmetro de controlo, α e θ , que sendo devidamente escolhidos, permite obter um estimador de θ com componente dominante do viés assintótico nula e mais eficiente que o clássico estimador de Hill.

Palavras-chave: Estatística de Extremos, Índice de cauda

Abstract: We shall consider here a new class of consistent semi-parametric estimators of a positive tail index γ , which generalizes other estimators already known (including Hill's estimator). This new class is parameterized with two tuning parameter, α and θ , which enables us to have access, for any available sample, to an asymptotically unbiased estimator of γ , with a high efficiency relatively to the classical Hill estimator.

Keywords: Statistic of Extremes, Tail index

AMS Classification: 62G20, 62G32.

1 Introdução

Sejam $X_1, X_2, \dots, X_n$ v.a. independentes com f.d. comum F . Admita-se que o modelo F pertence ao domínio de atracção para máximos da distribuição de valores extremos, EV (do inglês "Extreme Value"):

$$G(x) = G_\gamma(x) = \begin{cases} \exp[-(1 + \gamma x)^{-\frac{1}{\gamma}}], & 1 + \gamma x > 0 & \text{se } \gamma \neq 0 \\ \exp(-e^{-x}), & x \in \mathbb{R} & \text{se } \gamma = 0 \end{cases}$$

representando-se por $F \in D(G_\gamma)$, onde γ é o parâmetro de forma da distribuição. Este parâmetro está directamente relacionado com o peso da cauda do modelo F , isto é, quanto maior o valor de γ mais pesada é a cauda $1 - F$. No que se segue vai-se admitir que F pertence ao domínio de atracção para máximos da distribuição EV com $\gamma > 0$. Neste caso sabe-se que:

$$F \in D(G_\gamma), \gamma > 0 \text{ se e só se } 1 - F \in RV_{-1/\gamma} \text{ se e só se } U \in RV_\gamma, \quad (1)$$

onde $U(t) := F^\leftarrow(1 - 1/t)$, $t \geq 1$, e RV_α representa a classe das funções de variação regular em infinito de índice α , isto é, a classe das funções positivas mensuráveis $f(\cdot)$ tais que: $\lim_{t \rightarrow \infty} \frac{f(tx)}{f(t)} = x^\alpha$, $\forall x > 0$.

2 Estimação de γ

2.1 Estimador de Hill

No caso de se trabalhar com modelos de cauda pesada ($\gamma > 0$), o estimador de Hill (Hill, 1975) é um dos primeiros e mais conhecido estimador do índice de cauda, ou índice de variação regular:

$$\gamma_n^H(k) := \frac{1}{k} \sum_{i=1}^k (\ln X_{n-i+1:n} - \ln X_{n-k:n}),$$

onde $X_{i:n}$ é a i -ésima estatística ordinal ascendente de $(X_1, X_2, \dots, X_n)$. Sob determinadas condições, que serão posteriormente introduzidas, prova-se que o estimador de Hill não é assintoticamente centrado. Em termos práticos, dada uma amostra de dimensão n de um modelo F , verifica-se que a estimativa de γ varia bastante consoante o valor de k . Como exemplo deste problema, é apresentada a figura 1. Pode-se constatar que existe elevada instabilidade nas trajectórias amostrais das estimativas. No entanto numa das amostras existe alguma estabilidade junto ao verdadeiro valor de γ no início da trajectória, mas tal não se verifica para a outra amostra.

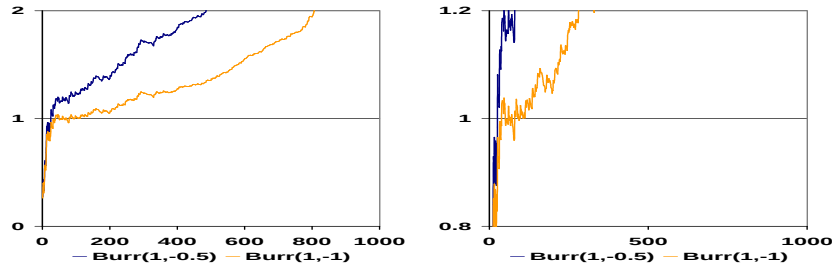


Figura 1: Trajectórias de $\gamma_n^H(k)$ como função de k , para duas amostras de dimensão 1000 de modelos com $\gamma = 1$. Os dois gráficos apresentam as mesmas trajectórias, embora com escalas diferentes.

A escolha do nível óptimo, nível que minimize o erro quadrático médio (EQM), não é fácil porque se por um lado a variância assintótica diminui à medida que k aumenta, por outro, o quadrado do viés aumenta. É assim natural que se procure encontrar outros estimadores que sejam uma alternativa ao estimador de Hill.

2.2 Nova classe de estimadores de γ

A classe de estimadores semiparamétricos de $\gamma > 0$ aqui introduzida é:

$$\gamma_n^{(\theta, \alpha)}(k) = \frac{\Gamma(\alpha)}{M_n^{(\alpha-1)}(k)} \times \left(\frac{M_n^{(\theta, \alpha)}(k)}{\Gamma(\theta\alpha + 1)} \right)^{\frac{1}{\theta}} \quad \alpha \geq 1, \theta > 0, \quad (2)$$

parametrizada através de α e θ , $M_n^{(0)}(k) \equiv 1$ e

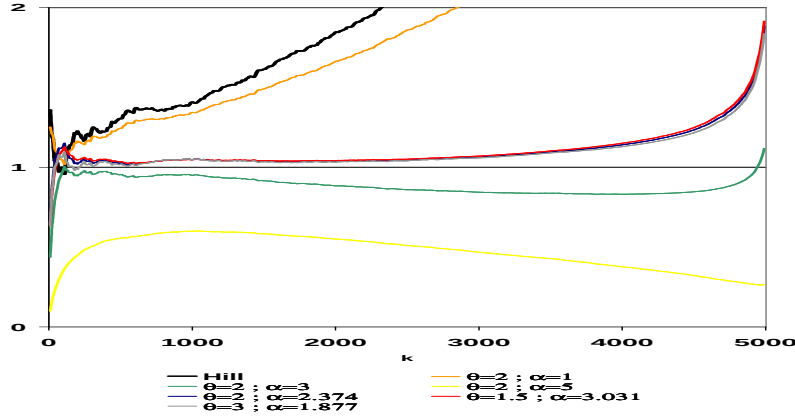
$$M_n^{(\alpha)}(k) := \frac{1}{k} \sum_{i=1}^k [\ln X_{n-i+1:n} - \ln X_{n-k:n}]^\alpha, \quad \alpha > 0$$

são estimadores consistentes de $\Gamma(\alpha + 1)\gamma^\alpha$ se $k = k(n)$ for uma sucessão intermédia, isto é, se:

$$k \rightarrow \infty, \text{ e } k = o(n), \text{ quando } n \rightarrow \infty. \quad (3)$$

$M_n^{(1)}(k)$, também aqui representado por $\gamma_n^H(k)$ é o estimador de Hill. Esta nova classe aqui introduzida, generaliza entre outros estimadores, o de Hill, $(\theta, \alpha) = (1, 1)$, o estimador estudado em Caeiro e Gomes (2001) que por sua vez generaliza $\gamma_n^{M_2}(k) := \sqrt{M_n^{(2)}(k)}/2$, e se $\alpha = 1$, o estimador estudado em Gomes et al. (2000).

No entanto, sendo objectivo deste trabalho obter estimadores assintoticamente centrados, é possível verificar que para isso apenas é necessário um parâmetro de controlo. Assim é preferível apenas utilizar e estimar um parâmetro de controlo. Apenas serão estudadas algumas subclasses de $\gamma_n^{(\theta, \alpha)}(k)$, com θ fixo e α parâmetro de controlo. Optou-se, por razões relacionadas com a escolha de α "ótimo", pelas subclasses de estimadores $\gamma_n^{(1.5, \alpha)}(k)$, $\gamma_n^{(2, \alpha)}(k)$ e $\gamma_n^{(3, \alpha)}(k)$. A seguinte figura, apresenta algumas trajectórias amostrais destas subclasses de um modelo com $\gamma = 1$. Para alguns valores de α existe estabilidade das trajectórias perto do verdadeiro valor de γ , o que significa que se reduziu o viés relativamente ao estimador de Hill.



3 Propriedades assintóticas da nova classe de estimadores

Nesta secção serão obtidas propriedades assintóticas de $\gamma_n^{(\theta, \alpha)}(k)$ que também serão válidas para as subclasses em estudo. Sob as condições $U(t) \in RV_\gamma$ e (3) relativa ao nível k , o estimador $\gamma_n^{(\theta, \alpha)}(k)$ é consistente para γ . Para se conhecer a distribuição assintótica de $\gamma_n^{(\theta, \alpha)}(k)$ considera-se uma condição adicional verificada por grande parte dos modelos de cauda pesada. Trata-se de uma condição de variação regular de segunda ordem que mede a velocidade de convergência de $U(tx)/U(t)$ para x^γ através do parâmetro de segunda ordem $\rho \leq 0$, isto é, quanto maior o valor de $|\rho|$, maior será essa velocidade. A condição de segunda ordem relativa à velocidade de convergência é a seguinte: existe uma função $A(\cdot)$ tal que:

$$\lim_{t \rightarrow \infty} \frac{\frac{U(tx)}{U(t)} - x^\gamma}{A(t)} = x^\gamma \frac{x^\rho - 1}{\rho}, \quad x > 0, \quad (4)$$

onde $|A(\cdot)| \in RV_\rho$ (Geluk e de Haan (1987)). Se $\rho = 0$, então $\frac{x^\rho - 1}{\rho} \equiv \ln(x)$ e admite-se que $A(\cdot)$ é assintoticamente equivalente a uma função decrescente e $\lim_{t \rightarrow \infty} A(t) = 0$.

Proposição 3.1 *Admitindo válidas as condições de variação regular de primeira e segunda ordem e que $k = k(n)$ é uma sucessão intermédia, verificando (3), $M_n^{(\alpha)}(k)$ tem a seguinte representação distribucional:*

$$M_n^{(\alpha)}(k) \stackrel{d}{=} \gamma^\alpha \Gamma(\alpha+1) + \frac{\gamma^\alpha \sigma_\alpha}{\sqrt{k}} Z_n^{(\alpha)} + \gamma^{\alpha-1} \Gamma(\alpha+1) \frac{1 - (1-\rho)^\alpha}{\rho(1-\rho)^\alpha} A(n/k)(1+o_p(1)),$$

onde $\sigma_\alpha := \sqrt{\Gamma(2\alpha+1) - \Gamma^2(\alpha+1)}$ e $Z_n^{(\alpha)} = \frac{\sqrt{k}}{\sigma_\alpha} \frac{1}{k} \sum_{i=1}^k [E_i^\alpha - \Gamma(\alpha+1)]$, com

E_i v.a.'s i.i.d. com distribuição $Exp(1)$, é uma sucessão de v.a. assintoticamente normais padrão. Por continuidade, o termo de viés $\frac{1-(1-\rho)^\alpha}{\rho(1-\rho)^\alpha}$ deverá ser substituído por α se $\rho = 0$.

Nota 1: A justificação desta proposição pode ser consultada em Gomes e Martins (2001) ou em Caeiro (2001).

Nota 2: Como $M_n^{(1)}(k) \equiv \gamma_n^H(k)$, o resultado anterior permite concluir, que sob as mesmas condições:

$$\gamma_n^H(k) \stackrel{d}{=} \gamma + \frac{\gamma}{\sqrt{k}} Z_n^{(1)} + \frac{1}{1-\rho} A(n/k)(1 + o_p(1)).$$

isto é, a variância e o termo dominante do viés assintótico do estimador de Hill são $\frac{\gamma^2}{k}$ e $\frac{1}{1-\rho} A(n/k)$, respectivamente.

Se também se verificar que $\sqrt{k}A(n/k) \xrightarrow{n \rightarrow \infty} \lambda$, então:

$$\sqrt{k}(\gamma_n^H(k) - \gamma) \xrightarrow{d} N\left(\frac{\lambda}{1-\rho}, \gamma^2\right), \quad n \rightarrow \infty$$

Teorema 3.1 Admitindo como válidas as condições da Proposição anterior:

1.

$$\gamma_n^{(\theta, \alpha)}(k) \stackrel{d}{=} \gamma + \frac{\gamma \sigma_{\theta, \alpha}}{\sqrt{k}} Z_n^{(\theta, \alpha)} + b_\theta(\alpha) A(n/k) + o_p(A(n/k)) + o_p(1/\sqrt{k})$$

onde:

$$\begin{aligned} Z_n^{(\theta, \alpha)} &= \frac{1}{\sigma_{\theta, \alpha}} \left(\frac{\sigma_{\theta, \alpha} Z_n^{(\theta \alpha)}}{\theta \Gamma(\theta \alpha + 1)} - \frac{\sigma_{\alpha-1} Z_n^{(\alpha-1)}}{\Gamma(\alpha)} \right) \xrightarrow{d} N(0, 1), \\ \sigma_{\theta, \alpha}^2 &= \frac{1}{\theta^2} \left[\frac{2\Gamma(2\theta\alpha)}{\theta\alpha\Gamma^2(\theta\alpha)} + \frac{\theta^2\Gamma(2\alpha-1)}{\Gamma^2(\alpha)} - \frac{2\Gamma((\theta+1)\alpha)}{\alpha\Gamma(\theta\alpha)\Gamma(\alpha)} - (\theta-1)^2 \right], \\ b_\theta(\alpha) &= \begin{cases} \frac{1}{\theta\rho} [(1-\rho)^{-\theta\alpha} - \theta(1-\rho)^{1-\alpha} + (\theta-1)] & \text{se } \rho < 0 \\ 1 & \text{se } \rho = 0 \end{cases} \end{aligned}$$

2. Para $\rho < 0$ e $\theta > 1$ (fixo) existe sempre um valor α tal que $b_\theta(\alpha) = 0$, ou seja, existe um valor que anula a componente dominante do viés de $\gamma_n^{(\theta, \alpha)}(k)$.

Demonstração:

A primeira parte do teorema não é muito difícil de provar. Basta usar-se a proposição 3.1 e os seguintes resultados:

$$1. \quad \frac{1}{1+x} = 1 - x + o(x), \quad \text{quando } x \rightarrow 0;$$

2. $(1+x)(1+y) = 1+x+y+o(xy)$, quando $x, y \rightarrow 0$.
3. Seja $P_n^{(\alpha)} = \sigma_\alpha Z_n^{(\alpha)}$. Tem-se que: $(P_n^{(\alpha)}, P_n^{(\beta)})$ é um par aleatório com distribuição assintótica binormal com vector médio nulo e matriz de covariâncias:

$$\sum_{(\alpha, \beta)} = \begin{bmatrix} \Gamma(2\alpha+1) - \Gamma^2(\alpha+1) & \Gamma(\alpha+\beta+1) - \Gamma(\alpha+1)\Gamma(\beta+1) \\ \Gamma(\alpha+\beta+1) - \Gamma(\alpha+1)\Gamma(\beta+1) & \Gamma(2\beta+1) - \Gamma^2(\beta+1) \end{bmatrix}$$

Como:

$$\begin{aligned} \left(\frac{M_n^{(\theta\alpha)}(k)}{\Gamma(\theta\alpha+1)} \right)^{\frac{1}{\theta}} &\stackrel{d}{=} \left[\gamma^{\theta\alpha} \left(1 + \frac{\sigma_{\theta\alpha}}{\sqrt{k}\Gamma(\theta\alpha+1)} Z_n^{(\theta\alpha)} + \frac{1-(1-\rho)^{\theta\alpha}}{\gamma\rho(1-\rho)^{\theta\alpha}} A(n/k)(1+o_p(1)) \right) \right]^{\frac{1}{\theta}} \\ &\stackrel{d}{=} \gamma^\alpha \left(1 + \frac{\sigma_{\theta\alpha}}{\sqrt{k}\theta\Gamma(\theta\alpha+1)} Z_n^{(\theta\alpha)} + \frac{1-(1-\rho)^{\theta\alpha}}{\theta\gamma\rho(1-\rho)^{\theta\alpha}} A(n/k)(1+o_p(1)) \right) \end{aligned}$$

e

$$\begin{aligned} \frac{\Gamma(\alpha)}{M_n^{(\alpha-1)}(k)} &\stackrel{d}{=} \frac{1}{\gamma^{\alpha-1} \left(1 + \frac{\sigma_{\alpha-1}}{\sqrt{k}\Gamma(\alpha)} Z_n^{(\alpha-1)} + \frac{1-(1-\rho)^{\alpha-1}}{\gamma\rho(1-\rho)^{\alpha-1}} A(n/k)(1+o_p(1)) \right)} \\ &\stackrel{d}{=} \gamma^{1-\alpha} \left(1 - \frac{\sigma_{\alpha-1}}{\sqrt{k}\Gamma(\alpha)} Z_n^{(\alpha-1)} - \frac{1-(1-\rho)^{\alpha-1}}{\gamma\rho(1-\rho)^{\alpha-1}} A(n/k)(1+o_p(1)) + o_p\left(\frac{1}{\sqrt{k}}\right) \right) \end{aligned}$$

resulta que

$$\begin{aligned} \gamma_n^{(\theta, \alpha)}(k) &= \frac{\Gamma(\alpha)}{M_n^{(\alpha-1)}(k)} \times \left(\frac{M_n^{(\theta\alpha)}(k)}{\Gamma(\theta\alpha+1)} \right)^{\frac{1}{\theta}} \\ &\stackrel{d}{=} \gamma + \frac{\gamma}{\sqrt{k}} \left(\frac{\sigma_{\theta\alpha} Z_n^{(\theta\alpha)}}{\theta\Gamma(\theta\alpha+1)} - \frac{\sigma_{\alpha-1} Z_n^{(\alpha-1)}}{\Gamma(\alpha)} \right) + o_p\left(\frac{1}{\sqrt{k}}\right) + b_\theta(\alpha) A(n/k)(1+o_p(1)) \end{aligned}$$

Usando o terceiro resultado conclui-se que $\left(\frac{\sigma_{\theta\alpha} Z_n^{(\theta\alpha)}}{\theta\Gamma(\theta\alpha+1)} - \frac{\sigma_{\alpha-1} Z_n^{(\alpha-1)}}{\Gamma(\alpha)} \right) \sim N\left(0, \sigma_{\theta, \alpha}^2\right)$ e portanto:

$$\gamma_n^{(\theta, \alpha)}(k) \stackrel{d}{=} \gamma + \frac{\gamma\sigma_{\theta, \alpha}}{\sqrt{k}} Z_n^{(\theta, \alpha)} + b_\theta(\alpha) A(n/k) + o_p(A(n/k)) + o_p\left(\frac{1}{\sqrt{k}}\right).$$

A segunda parte do teorema não é difícil de demonstrar. Tendo em conta que:

$$\begin{aligned} \lim_{\alpha \rightarrow 1} b_\theta(\alpha) &= \frac{1-(1-\rho)^\theta}{\theta\rho(1-\rho)^\theta} \geq 0, & \rho < 0, \quad \theta > 0, \\ \lim_{\alpha \rightarrow +\infty} b_\theta(\alpha) &= \frac{\theta-1}{\rho\theta} < 0, & \rho < 0, \quad \theta > 1, \end{aligned}$$

e o facto de $b_\theta(\alpha)$ ser contínua pode-se assegurar para $\theta > 1$, a existência de *pelo menos um valor* $\alpha_0 \in [1; +\infty[$ para o qual $b_\theta(\alpha_0) = 0$. Para provar a unicidade é ainda necessário estudar a monotonia da função $b_\theta(\alpha)$. Como para $\theta > 1$, $b_\theta(\alpha)$ como função de α é estritamente decrescente, fica provada a unicidade de α_0 . Como não há garantias que $\sqrt{k}A(n/k) \rightarrow \lambda \neq 0$ (finito) conclui-se que o termo dominante do viés assintótico de $\gamma_n^{(\theta, \alpha)}(k)$ poderá ser nulo, dependendo para isso de $b_\theta(\alpha)$. A existência de um valor α_0 tal que $b_\theta(\alpha_0) = 0$ permite obter um estimador assintoticamente centrado, o que significa que fazendo $\alpha = \alpha_0$ provavelmente haverá trajectórias amostrais de " $\gamma_n^{(\theta, \alpha)}(k)$ vs k " mais estáveis do que as do estimador de Hill.

A seguinte proposição é uma consequência imediata do teorema anterior, e apenas será demonstrada para $\theta = 2$.

Proposição 3.2 *Para os casos particulares $\theta = 1.5, 2, 3$ e 4 , isto é para as classes de estimadores em estudo $\gamma_n^{(1.5, \alpha)}(k)$, $\gamma_n^{(2, \alpha)}(k)$ e $\gamma_n^{(3, \alpha)}(k)$ e para a classe $\gamma_n^{(4, \alpha)}(k)$ (não estudada) tem-se:*

$$\begin{aligned}\alpha_0^{(1.5)} &= \alpha_0^{(1.5)}(\rho) = -\frac{2 \ln \left[(1-\rho) \left(\frac{1}{2} + \cos \left(\frac{\arctan(-\frac{1}{8}((1-\rho)^3-2), \frac{1}{4}\sqrt{(1-\rho)^3-1})+4\pi}{3} \right) \right) \right]}{\ln(1-\rho)} \\ \alpha_0^{(2)} &= \alpha_0^{(2)}(\rho) = -\frac{\ln \left[1 - \rho - \sqrt{(1-\rho)^2 - 1} \right]}{\ln(1-\rho)} \\ \alpha_0^{(3)} &= \alpha_0^{(3)}(\rho) = -\frac{\ln \left[2\sqrt{1-\rho} \cos \left(\frac{\arctan(-1, \sqrt{(1-\rho)^3-1})+4\pi}{3} \right) \right]}{\ln(1-\rho)} \\ \alpha_0^{(4)} &= \alpha_0^{(4)}(\rho) = -\frac{\ln \left[\frac{\sqrt{u}}{2} \left(1 - \sqrt{2\sqrt{1-12/u^2}-1} \right) \right]}{\ln(1-\rho)} \\ u &= 2(1-\rho)^{2/3} \left[\sqrt[3]{1 + \sqrt{1-(1-\rho)^{-4}}} + \sqrt[3]{1 - \sqrt{1-(1-\rho)^{-4}}} \right]\end{aligned}$$

Nota: $\arctan(x, y) \in [0, 2\pi[$ é o argumento do número complexo $x + iy$.

Demonstração:

Quando $\theta = 2$ a resolução da equação $b_2(\alpha_0) = 0$ é bastante simples uma vez que consiste em obter as raízes de um polinómio de segundo grau na variável $x = (1-\rho)^{-\alpha_0}$:

$$\begin{aligned}b_2(\alpha_0) = 0 &\Leftrightarrow (2\rho)^{-1} [(1-\rho)^{-2\alpha_0} - 2(1-\rho)^{1-\alpha_0} + 1] = 0 \\ &\Leftrightarrow \alpha_0 = -\frac{\ln[(1-\rho) \pm \sqrt{(1-\rho)^2 - 1}]}{\ln(1-\rho)}\end{aligned}$$

Como só interessa a solução positiva, opta-se pelo sinal " $-$ " na expressão anterior. Para os outros valores de θ a demonstração faz-se de forma análoga através da resolução de equações polinomiais de grau 3 e 4.

Conclui-se assim que para obter α_0 será necessário primeiro conhecer o valor do parâmetro de segunda ordem, ρ , que deverá ser convenientemente estimado. A estimação de ρ não será abordada neste trabalho, embora já existam alguns estimadores razoáveis para este parâmetro. Na tabela 1 é apresentada para os valores de $\theta = 1.5, 2, 3$ e 4 e para alguns valores de ρ a solução da equação $b_\theta(\alpha) = 0$.

Tabela 1: $\alpha_0 : b_\theta(\alpha_0) = 0$ como função do parâmetro de 2ª ordem, ρ .

	-0.01	-0.1	-0.25	-0.5	-1	-1.25	-1.5	-2	-2.5	-3	$-\infty$
$\alpha_0^{(1.5)}(\rho)$	19.755	6.259	4.070	3.031	2.355	2.195	2.081	1.927	1.827	1.756	1
$\alpha_0^{(2)}(\rho)$	14.201	4.654	3.106	2.374	1.900	1.789	1.710	1.605	1.536	1.488	1
$\alpha_0^{(3)}(\rho)$	10.211	3.470	2.385	1.877	1.557	1.483	1.432	1.364	1.321	1.291	1
$\alpha_0^{(4)}(\rho)$	8.428	2.933	2.057	1.654	1.405	1.350	1.311	1.261	1.229	1.207	1

Conclui-se assim, para qualquer dos estimadores em estudo, que conhecendo-se o valor do parâmetro ρ é possível escolher o valor de α de modo a anular o termo dominante do viés assintótico. Sendo desejável escolher teoricamente como k ótimo o valor k que minimize o EQM, e sendo o EQM igual à variância mais o quadrado do viés seria ainda interessante comparar a variância assintótica das várias classes de estimadores em estudo. No entanto sendo a distribuição assintótica do mesmo tipo, basta apenas comparar a constante $\sigma_{\theta, \alpha_0}^2$ que corresponde à constante que diferencia as variâncias das classes de estimadores. Da análise da tabela 2 conclui-se que dependendo do valor do parâmetro ρ , os estimadores $\gamma_n^{(1.5, \alpha_0)}(k)$ e $\gamma_n^{(2, \alpha_0)}(k)$ são os que apresentam variâncias mais pequenas.

Tabela 2: Tabela com $\sigma_{\theta, \alpha_0}^2$ para alguns valores de θ .

ρ	$\theta = 1.5$		$\theta = 2$		$\theta = 3$		$\theta = 4$	
	α_0		α_0		α_0		α_0	
-0.1	6.259	32561.053	4.654	17833.760	3.470	35399.104	2.933	118115.452
-0.25	4.070	368.022	3.106	277.828	2.385	459.256	2.057	1085.987
-0.5	3.031	43.638	2.374	38.752	1.877	61.245	1.654	128.191
-0.75	2.600	18.087	2.071	17.211	1.671	27.117	1.493	54.975
-1	2.355	11.028	1.900	10.919	1.557	17.281	1.405	34.675
-1.25	2.195	8.002	1.789	8.133	1.483	12.951	1.350	25.909
-1.5	2.081	6.380	1.710	6.609	1.432	10.588	1.311	21.178
-2	1.927	4.721	1.605	5.019	1.364	8.128	1.261	16.295
-2.5	1.827	3.894	1.536	4.211	1.321	6.878	1.229	13.830
-3	1.756	3.402	1.488	3.725	1.291	6.125	1.207	12.349
$-\infty$	1.000	1.065	1.000	1.250	1.000	2.111	1.000	4.313

4 Propriedades dos estimadores para amostras de dimensão finita (Estudo computacional)

No que se segue apenas será estudado, para amostras de dimensão finita, o estimador: $\gamma_n^{(\theta, \alpha)}(k)$ com $\theta = 1.5, 2$ e 3 e α escolhido de forma a anular o termo

dominante do viés assintótico. Para estudar o estimador no respectivo nível óptimo, realizou-se o seguinte estudo de simulação:

Foram geradas 1000 amostras de dimensão n e obtiveram-se as estimativas:

$$\hat{\gamma}_{n,(i)}(k), \quad k = 1, 2, \dots, n-1, \quad i = 1, 2, \dots, 1000,$$

que depois permitiram obter:

$$\widehat{EQM}[\hat{\gamma}_n(k)] = \frac{1}{1000} \sum_{i=1}^{1000} (\hat{\gamma}_{n,(i)}(k) - \gamma)^2, \quad \widehat{Viés} = \frac{1}{1000} \sum_{i=1}^{1000} (\hat{\gamma}_{n,(i)}(k) - \gamma),$$

$$k_0 = \arg \min_k \widehat{EQM}[\hat{\gamma}_n(k)], \quad \widehat{Efic}_{\gamma_n(k), \gamma_n^H(k)} = \sqrt{\frac{\widehat{EQM}[\hat{\gamma}_n^H(k)]}{\widehat{EQM}[\hat{\gamma}_n(k)]}}.$$

Este processo foi ainda repetido 10 vezes de forma a, sempre que possível, obter intervalos de confiança para as várias estimativas. O processo foi também repetido para diversos valores de n .

4.1 Modelo Fréchet

O primeiro modelo usado foi o *Fréchet*, $F(x) = \exp(-x^{-1/\gamma})$, $x > 0, \gamma > 0$ com dimensões de amostras compreendidas entre 200 e 3500. É fácil de verificar que:

$$\frac{U(tx)}{U(t)} = x^\gamma \left[1 + \left(\frac{x^{-1} - 1}{-1} \right) \frac{\gamma}{2t} \right] + o(t^{-1})$$

e portanto, neste modelo, o parâmetro de segunda ordem é sempre -1 e $A(t) \sim \frac{\gamma}{2t}$ e por isso $\gamma_n^{(1.5,2.4)}(k)$, $\gamma_n^{(2,1.9)}(k)$ e $\gamma_n^{(3,1.6)}(k)$ serão os estimadores a estudar. Os resultados do estudo aparecem nas figuras 2 e 3 e na tabela 3.

Tabela 3: Fracção óptima da amostra, valores esperados, erro quadrático médio e eficiência de $\gamma_n^H(k)$, $\gamma_n^{(1.5,2.4)}(k)$, $\gamma_n^{(2,1.9)}(k)$ e $\gamma_n^{(3,1.6)}(k)$ para o modelo *Fréchet* com $\gamma = 1$.

n	200	500	1000	2000	3500
k_0^H	0.2760 ± .0130	0.2162 ± .0051	0.1810 ± .0113	0.1462 ± .0079	0.1300 ± .0049
$k_0^{(1.5,2.4)}$	0.9950 ± .0000	0.9980 ± .0000	0.9985 ± .0000	0.9985 ± .0000	0.9981 ± .0001
$k_0^{(2,1.9)}$	0.9950 ± .0000	0.9980 ± .0000	0.9978 ± .0000	0.9970 ± .0002	0.9963 ± .0002
$k_0^{(3,1.6)}$	0.9950 ± .0000	0.9980 ± .0000	0.9990 ± .0000	0.9995 ± .0000	0.9997 ± .0000
E^H	1.0836 ± .0046	1.0623 ± .0023	1.0507 ± .0037	1.0402 ± .0024	1.0350 ± .0014
$E_{(1.5,2.4)}$	0.9945 ± .0015	1.0057 ± .0016	1.0084 ± .0031	1.0063 ± .0007	1.0038 ± .0005
$E_{(2,1.9)}$	1.0013 ± .0015	1.0136 ± .0015	1.0109 ± .0013	1.0072 ± .0008	1.0054 ± .0004
$E_{(3,1.6)}$	0.9574 ± .0017	0.9707 ± .0019	0.9778 ± .0009	0.9839 ± .0007	0.9883 ± .0005
$Ef_{(1.5,2.4),H}$	1.5496 ± .0408	1.7848 ± .0397	1.9449 ± .0243	2.1468 ± .0308	2.3171 ± .0456
$Ef_{(2,1.9),H}$	1.6290 ± .0404	1.8188 ± .0386	1.9586 ± .0264	2.1299 ± .0319	2.2837 ± .0451
$Ef_{(3,1.6),H}$	1.3501 ± .0344	1.4609 ± .0327	1.5897 ± .0168	1.7541 ± .0311	1.9126 ± .0371

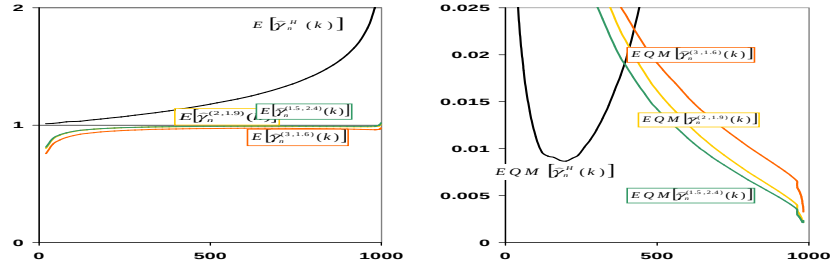


Figura 2: Valor esperado, erro quadrático médio de $\gamma_n^H(k)$, $\gamma_n^{(1.5,2.4)}(k)$, $\gamma_n^{(2,1.9)}(k)$ e $\gamma_n^{(3,1.6)}(k)$ para amostras de dimensão 1000 do modelo *Fréchet*.

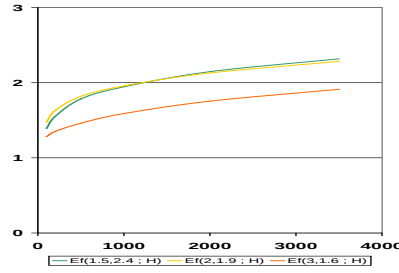


Figura 3: Eficiência de $\gamma_n^{(1.5,2.4)}(k)$, $\gamma_n^{(2,1.9)}(k)$ e $\gamma_n^{(3,1.6)}(k)$ relativamente ao estimador de Hill para amostras de várias dimensões do modelo *Fréchet*.

4.2 Modelo Burr

O modelo *Burr*, definido por $F(x) = 1 - (1 + x^{-\rho/\gamma})^{1/\rho}$, $x \geq 0$ com $\gamma = 1$ e $\rho < 0$. Facilmente se verifica que:

$$\frac{U(tx)}{U(t)} = x^\gamma \left[1 + \left(\frac{x^\rho - 1}{\rho} \right) \gamma t^\rho \right] + o(t^\rho)$$

O que permite concluir que neste modelo os valores dos parâmetros γ e ρ são independentes. Tendo em conta o facto de os resultados até agora conhecidos indicarem que das três classes em estudo, $\gamma_n^{(3,\alpha)}(k)$ é a que apresenta piores resultados (devido à maior variância), só são apresentados resultados para os estimadores $\gamma_n^{(1.5,\alpha_0)}(k)$ e $\gamma_n^{(2,\alpha_0)}(k)$, para as dimensões de amostras iguais a 100, 200, 500 e 1000. Os resultados do estudo são os apresentados na tabela 4 e figura 4.

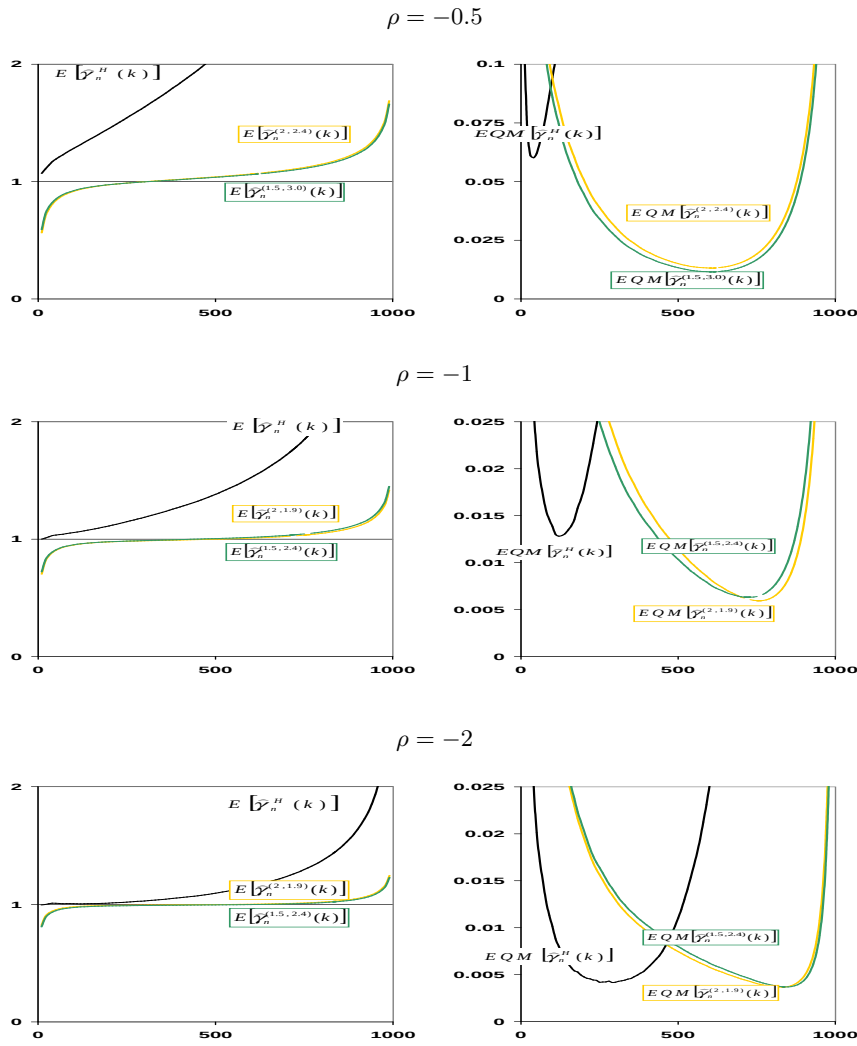


Figura 4: Valor esperado, erro quadrático médio de $\gamma_n^H(k)$, $\gamma_n^{(1.5, \alpha_0)}(k)$, e $\gamma_n^{(2, \alpha_0)}(k)$ para amostras de dimensão 1000 do modelo *Burr* com $\gamma = 1$ e $\rho = -0.5, -1$ e -2 .

Tabela 4: Fração óptima da amostra de $\gamma_n^H(k)$, $\gamma_n^{(1.5, \alpha_0)}(k)$, $\gamma_n^{(2, \alpha_0)}(k)$ e respectivas eficiências relativas, para o modelo *Burr* com $\gamma = 1$ e $\rho = -0.5, -1, -2$.

$\alpha \setminus n$	100	200	500	1000
Burr: $\gamma = 1, \rho = -0.5$				
$k_0(H)/n$	$0.0980 \pm .0094$	$0.0780 \pm .0061$	$0.0486 \pm .0029$	$0.0386 \pm .0042$
$k_0(1.5, 3.0)/n$	$0.7550 \pm .0091$	$0.7155 \pm .0049$	$0.6530 \pm .0036$	$0.6013 \pm .0060$
$k_0(2, 2.4)/n$	$0.7590 \pm .0071$	$0.7150 \pm .0056$	$0.6528 \pm .0036$	$0.6013 \pm .0061$
$Ef(1.5, 3.0), H$	$2.2609 \pm .0643$	$2.1509 \pm .0491$	$2.0806 \pm .0343$	$2.0426 \pm .0463$
$Ef(2, 2.4), H$	$2.3863 \pm .0616$	$2.2840 \pm .0473$	$2.2217 \pm .0336$	$2.1829 \pm .0475$
Burr: $\gamma = 1, \rho = -1$				
$k_0(H)/n$	$0.2240 \pm .0159$	$0.1785 \pm .0103$	$0.1386 \pm .0055$	$0.1182 \pm .0083$
$k_0(1.5, 2.4)/n$	$0.8670 \pm .0083$	$0.8385 \pm .0056$	$0.7896 \pm .0044$	$0.7513 \pm .0036$
$k_0(2, 1.9)/n$	$0.8430 \pm .0035$	$0.8185 \pm .0059$	$0.7648 \pm .0053$	$0.7189 \pm .0038$
$Ef(1.5, 2.4), H$	$1.5485 \pm .0457$	$1.4963 \pm .0400$	$1.4787 \pm .0187$	$1.4983 \pm .0225$
$Ef(2, 1.9), H$	$1.5476 \pm .0417$	$1.4813 \pm .0382$	$1.4484 \pm .0177$	$1.4543 \pm .0233$
Burr: $\gamma = 1, \rho = -2$				
$k_0(H)/n$	$0.4240 \pm .0203$	$0.3715 \pm .0133$	$0.3134 \pm .0079$	$0.2818 \pm .0097$
$k_0(1.5, 1.9)/n$	$0.9130 \pm .0035$	$0.8935 \pm .0041$	$0.8542 \pm .0037$	$0.8180 \pm .0029$
$k_0(2, 1.6)/n$	$0.9180 \pm .0045$	$0.9025 \pm .0019$	$0.8696 \pm .0046$	$0.8372 \pm .0031$
$Ef(1.5, 1.9), H$	$1.1272 \pm .0149$	$1.0953 \pm .0178$	$1.0764 \pm .0124$	$1.0700 \pm .0112$
$Ef(2, 1.6), H$	$1.1363 \pm .0157$	$1.0993 \pm .0191$	$1.0749 \pm .0133$	$1.0682 \pm .0124$

5 Conclusões

Os resultados até agora apresentados parecem ser promissores, embora neste trabalho não se tenha tratado o problema da estimação do parâmetro de segunda ordem, ρ (que continua a ser um problema ainda não totalmente resolvido). É necessário continuar com o estudo computacional, quer para outros valores de (γ, ρ) , quer para outros modelos de cauda pesada. No entanto com o trabalho até agora realizado é possível apresentar as seguintes conclusões:

- Quer a nível assintótico (devido à elevada variância), quer para amostras de dimensão finita (nos modelos *Fréchet* e *Burr*), o estimador $\gamma_n^{(3, \alpha_0)}(k)$ apresenta sempre piores resultados, que $\gamma_n^{(1.5, \alpha_0)}(k)$ e $\gamma_n^{(2, \alpha_0)}(k)$. Por outro lado, os resultados de $\gamma_n^{(1.5, \alpha_0)}(k)$ e $\gamma_n^{(2, \alpha_0)}(k)$ foram muito semelhantes.
- Da análise das tabelas conclui-se que qualquer dos estimadores apresenta, no nível óptimo, eficiências superiores a 1. Num problema prático não há possibilidades de saber qual o nível óptimo. Foi por esta razão que foram incluídos os gráficos das trajectórias amostrais das várias estimativas, e que nos permitem verificar que mesmo para uma vasta região de valores de k , a estimativa de γ não se afasta muito da estimativa obtida no nível óptimo. Isto significa que para estes estimadores, a escolha do nível k não é tão determinante, quanto para o estimador de Hill.
- Para o modelo *Fréchet* ($\rho = -1$), é necessário para estimar γ ir praticamente até ao mínimo da amostra. Os gráficos apresentados para este modelo sugerem que os estimadores em estudo praticamente não têm viés.
- No modelo *Burr*, para valores de ρ pequenos, como por exemplo $\rho = -2$,

o desempenho de $\gamma_n^{(1.5, \alpha_0)}(k)$ e $\gamma_n^{(2, \alpha_0)}(k)$, em termos de EQM , não é significativamente melhor que o desempenho de $\gamma_n^H(k)$. Contudo, analisando as figuras conclui-se que existe uma enorme estabilidade na trajectória do valor esperado de $\gamma_n^{(1.5, \alpha_0)}(k)$ e $\gamma_n^{(2, \alpha_0)}(k)$, junto ao verdadeiro valor de γ .

Referências

- [1] Abramowitz, M., e I. Stegun (1970). *Handbook of Mathematical Functions*. Dover.
- [2] Caeiro, F. (2001). *Generalizações de estimadores clássicos do índice de variação regular*. Tese de mestrado, D.E.I.O., Faculdade de Ciências da Universidade de Lisboa.
- [3] Caeiro, F. e M. I. Gomes (2001). A class of asymptotically unbiased semi-parametric estimators of the tail index. *Revista de Estatística*, vol II 2º Quadrimestre de 2001, 77-78.
- [4] Caeiro, F. e M. I. Gomes (2001). A class of asymptotically unbiased semi-parametric estimators of the tail index. *Test* **11**, nº2, 345-364.
- [5] Csörgő S., e D. Mason (1985). Central limit theorems for sums of extreme values. *Math. Proc. Cambridge Philos. Soc.* **98**, 547-558.
- [6] Geluk, J. e L. de Haan (1987). *Regular Variation, Extensions and Tauberian Theorems*. Tech. Report CWI Tract 40, Centre for Mathematics and Computer Science, Amsterdam, Netherlands.
- [7] Gomes, M. I. e Caeiro, F. (2002). Bias reduction in the semi-parametric estimation of parameters of rare events. Skopenko, V. V. et al (eds), International Gnedenko Conference Abstracts, 68.
- [8] Gomes, M. I., e M. J. Martins (2001). Generalizations of the Hill estimator - asymptotic versus finite sample behaviour. *Journal of Statistical Planning and Inference* 93, 161-180.
- [9] Gomes, M. I., M. J. Martins, e M. Neves (2000). Alternatives to a Semi-Parametric Estimator of Parameters of Rare Events - The Jackknife Methodology. *Extremes* **3**:3, 207-229.
- [10] Gomes, M. I., e M. J. Martins (1999). *Asymptotic efficiency of generalised Jackknife semi-parametric estimators of a heavy tail index*. Notas e Comunicações 16/99, CEAUL.
- [11] Guillou, A. e P. Hall (2001). A diagnostic for selecting the threshold in extreme value analysis. *J. R. Statist. Soc.* **B63**, 293-305.
- [12] Haan, L. de e L. Peng (1998). Comparison of Tail Index Estimators. *Statistica Neerlandica* **52**, 60-70.
- [13] Haeusler, E. e J. L. Teugels (1985). On Asymptotic Normality of Hill's Estimator for the Exponent of Regular Variation. *Ann. Statist.* **13**, no. 2, 743-756.
- [14] Hill, B. M. (1975). A Simple General Approach to Inference About the Tail of a Distribution. *Ann. Statist.* **3**, no. 5, 1163-1174.
- [15] Peng, L. (1998). *Second Order Condition and Extreme Value Theory*. Ph.D. thesis, Erasmus University, Rotterdam.