

Alguns resultados adicionais sobre a variância de um estimador de viés reduzido do índice de cauda

Frederico Caeiro

DM e CMA, Universidade Nova de Lisboa - fac@fct.unl.pt

M. Ivette Gomes

DEIO e CEAUL, Universidade de Lisboa - ivette.gomes@fc.ul.pt

Dinis Pestana

DEIO e CEAUL, Universidade de Lisboa - dinis.pestana@fc.ul.pt

Resumo: Muitos dos estimadores clássicos do índice de cauda positivo γ , como por exemplo o estimador de Hill, são muito sensíveis à escolha do número de observações de topo, k . Nos estimadores de segunda ordem, de variância mínima e de viés reduzido (MVRB), a estimação dos parâmetros de segunda ordem do viés é usualmente feita num nível k_1 , de ordem superior ao nível k usado para estimar o índice de cauda. Deste modo, os estimadores MVRB mantêm a variância do estimador de Hill, nos níveis k onde podemos assegurar a normalidade assintótica do estimador de Hill. Contudo, para atingirmos o erro quadrático médio mínimo, dos estimadores MVRB, devemos utilizar níveis k e k_1 da mesma ordem. Neste trabalho apresentamos um estudo de simulação de Monte-Carlo de uma possível escolha de k e k_1 , bem como do comportamento de um estimador MVRB, como função de k , para essa escolha de k_1 .

Palavras-chave: Estatística de Extremos, Índice de cauda, Estimação semi-paramétrica de viés reduzido, Simulação de Monte-Carlo.

Abstract: Classical tail index estimators, like the Hill estimator, are known to be quite sensitive to the number of top order statistics k used in the estimation. In the recent minimum-variance reduced-bias (MVRB) tail index estimators, the estimation of the second order parameters in the bias has been performed at a level k_1 of a larger order than that of the level k at which we compute the tail index estimators. Such a procedure enables us to keep the asymptotic variance of the new estimators equal to the asymptotic variance of the Hill estimator, for all k at which we can guarantee the asymptotic normality of the Hill statistics. To reach the minimal mean squared error of these MVRB estimators, we need to work with levels k and k_1 of the same order. In this paper perform a Monte-Carlo simulation study of a possible choice of k and k_1 , as well as the behaviour of a MVRB-estimator, as a function of k , for such a choice of k_1 .

Keywords: Statistic of Extremes, Tail Index, Semi-parametric estimation, Monte-Carlo simulation.

2 Caeiro, F. *et al.*/Variância de um estimador do índice de cauda

1 Introdução

Dizemos que a função de distribuição F tem cauda pesada, se a função cauda $\bar{F} = 1 - F$ é de variação regular de índice $-1/\gamma$ (Gnedenko, 1943), e usamos a notação $\bar{F} \in RV_{-1/\gamma}$ onde RV_α representa a classe de funções de variação regular de índice α . O parâmetro $\gamma (> 0)$ é o índice de cauda, o parâmetro a estimar. Se considerarmos a função quantil $U(t) := F^{\leftarrow}(1 - 1/t)$, $t > 1$ em que F^{\leftarrow} representa a função inversa generalizada de F , definida por $F^{\leftarrow}(x) := \inf\{y : F(y) \geq x\}$, então dizermos que F tem cauda pesada é equivalente a admitirmos que

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

isto é, $U \in RV_\gamma$.

Para obtermos a distribuição assintótica de muitos dos estimadores semi-paramétricos do índice de cauda, necessitamos uma condição de segunda ordem. Neste trabalho vamos considerar a seguinte condição de segunda ordem:

$$\lim_{t \rightarrow \infty} \frac{\ln U(tx) - \ln U(t) - \gamma \ln x}{A(t)} = \frac{x^\rho - 1}{\rho}, \quad \forall x > 0, \quad (2)$$

onde $\rho \leq 0$ é um parâmetro de segunda ordem que mede a velocidade de convergência de $U(tx)/U(t)$ para x^γ e $|A| \in RV_\rho$ (Geluk and de Haan, 1987). De modo a obter informação sobre o termo dominante de viés dos estimadores de variância mínima e viés reduzido (MVRB), precisamos de assumir a seguinte condição de terceira ordem,

$$\lim_{t \rightarrow \infty} \frac{\frac{\ln U(tx) - \ln U(t) - \gamma \ln x}{A(t)} - \frac{x^\rho - 1}{\rho}}{B(t)} = \frac{x^{\rho+\rho'} - 1}{\rho + \rho'}, \quad \forall x > 0, \quad (3)$$

onde $\rho' \leq 0$ é um parâmetro de terceira ordem e $|B| \in RV_{\rho'}$. Iremos também admitir que a condição anterior é válida com:

$$A(t) = \gamma \beta t^\rho, \quad B(t) = \beta' t^{\rho'}, \quad \beta, \beta' \neq 0, \quad \rho, \rho' < 0. \quad (4)$$

As condições (2), (3) e (4) são válidas para muitos dos modelos de cauda pesada, como por exemplo o Fréchet, o Burr, o Generalizado de Pareto e o t de Student.

Baseando a inferência nas k estatísticas ordinais de topo e admitindo que k é uma sucessão intermédia, isto é, uma sucessão de valores inteiros $k = k_n \in [1, n]$, verificando

$$k = k_n \rightarrow \infty \quad \text{e} \quad k_n/n \rightarrow 0, \quad \text{quando} \quad n \rightarrow \infty, \quad (5)$$

vamos considerar os log-excessos, acima de um nível elevado, isto é,

$$V_{ik} := \ln X_{n-i+1:n} - \ln X_{n-k:n}, \quad 1 \leq i \leq k < n$$

e os log espaçamentos escalados,

$$W_i := i\{\ln X_{n-i+1:n} - \ln X_{n-i:n}\}, \quad 1 \leq i < n,$$

onde $X_{i:n}$, $1 \leq i \leq n$, representa a i -ésima estatística ordinal ascendente associada à amostra aleatória (X_1, X_2, \dots, X_n) . Para modelos de cauda pesada, Hill (1975), propôs o seguinte estimador semi-paramétrico do índice de cauda,

$$H(k) = \hat{\gamma}_n^H(k) := \frac{1}{k} \sum_{i=1}^k V_{ik} = \frac{1}{k} \sum_{i=1}^k W_i, \quad (6)$$

consistente para a estimação do índice de cauda, sob a validade da condição de primeira ordem (1) e k intermédia. Sob a condição de segunda ordem, (2), e para k intermédia, podemos assegurar a representação assintótica:

$$H(k) \stackrel{d}{=} \gamma + \frac{\gamma}{\sqrt{k}} Z_k + \frac{A(n/k)}{1-\rho} (1 + o_p(1)),$$

onde Z_k é uma variável aleatória com distribuição normal padrão. Se $A(t)$ verificar a representação indicada em (4), o nível k “ótimo”, isto é, o nível k que minimiza o erro quadrático médio é

$$k_0^H = \left(\frac{(1-\rho)n^{-\rho}}{\beta\sqrt{-2\rho}} \right)^{2/(1-2\rho)}. \quad (7)$$

O mais simples estimador MVRB do índice de cauda, é o estimador de Hill “corrigido” (Caeiro *et al.*, 2005), com expressão funcional,

$$\bar{H}_{\hat{\beta}, \hat{\rho}}(k) := H(k) \left(1 - \frac{\hat{\beta}}{1-\hat{\rho}} (n/k)^{\hat{\rho}} \right), \quad (8)$$

onde $H(k)$ representa o estimador de Hill e $(\hat{\beta}, \hat{\rho})$ estimadores adequados dos parâmetros de segunda ordem (β, ρ) . Para garantirmos a consistência de $(\hat{\beta}, \hat{\rho})$, precisamos de usar k_1 estatísticas ordinais de topo, tais que $\sqrt{k_1}A(n/k_1) \rightarrow \infty$.

Na Secção 2 apresentamos o comportamento assintótico do estimador MVRB em (8), trabalhando com níveis tais que $k = o(k_1)$, e indicamos algumas condições que nos permitem manter a sua variância assintótica igual a γ^2 . Fazemos ainda uma revisão dos estimadores dos parâmetros de forma e escala, ρ e β , respectivamente, e apresentamos o comportamento assintótico de $\sqrt{k}(\bar{H}_{\hat{\beta}, \hat{\rho}}(k) - \gamma)$, quando $\hat{\beta}$ e $\hat{\rho}$ são os estimadores dos parâmetros de segunda ordem calculados num nível “ótimo” $k_1 = k_1^{opt}$ e a estimação do índice de cauda é feita num nível k , da mesma ordem de k_1 . Finalmente, na Secção 3, apresentamos um estudo, através de simulação de Monte-Carlo, duma possível escolha de k e k_1 , bem como do comportamento do estimador MVRB em (8) como função de k , para essa escolha de k_1 .

4 Caeiro, F. *et al.*/Variância de um estimador do índice de cauda

2 Propriedades assintóticas

Apresentamos a seguir uma generalização do Teorema 3.1 de Caeiro *et al.* (2005).

Teorema 1 (Gomes *et al.*, 2009). *Para a classe de modelos verificando (3), com A e B escolhidas conforme indicado em (4), k intermédia, e admitindo que conhecemos o valor dos parâmetros de segunda ordem β e ρ ,*

$$\bar{H}_{\beta,\rho}(k) \stackrel{d}{=} \gamma + \frac{\gamma}{\sqrt{k}} Z_k - A(n/k) \left(\frac{A(n/k)}{\gamma(1-\rho)^2} - \frac{B(n/k)}{1-\rho-\rho'} + O_p\left(\frac{1}{\sqrt{k}}\right) \right) (1 + o_p(1)).$$

Se escolhermos k intermédia tal que $\sqrt{k}A(n/k) \rightarrow \infty$, $\sqrt{k}A^2(n/k) \rightarrow \lambda_A$ e $\sqrt{k}A(n/k)B(n/k) \rightarrow \lambda_B$, com λ_A e λ_B finitos,

$$\sqrt{k}(\bar{H}_{\beta,\rho}(k) - \gamma) \xrightarrow[n \rightarrow \infty]{d} \text{Normal}(b_{\bar{H}}, \gamma^2),$$

com

$$b_{\bar{H}} = b_{\bar{H}}(\gamma, \rho, \rho') = -\frac{\lambda_A}{\gamma(1-\rho)^2} + \frac{\lambda_B}{1-\rho-\rho'} =: \lambda_A u_{\bar{H}} + \lambda_B v_{\bar{H}}. \quad (9)$$

Sejam $(\hat{\beta}, \hat{\rho})$ estimadores consistentes dos parâmetros de segunda ordem (β, ρ) . Se $\hat{\rho} - \rho = o_p(1/\ln n)$, quando $n \rightarrow \infty$, então,

$$\bar{H}_{\hat{\beta}, \hat{\rho}}(k) - \bar{H}_{\beta, \rho}(k) \underset{p}{\approx} -\frac{A(n/k)}{1-\rho} \left\{ \frac{\hat{\beta} - \beta}{\beta} + (\hat{\rho} - \rho) \left(\ln(n/k) + \frac{1}{1-\rho} \right) \right\}. \quad (10)$$

Consequentemente, $\sqrt{k}(\bar{H}_{\hat{\beta}, \hat{\rho}}(k) - \gamma)$ é assintoticamente normal com valor médio nulo e variância γ^2 , não só quando $\sqrt{k}A(n/k) \rightarrow 0$, mas também quando $\sqrt{k}A(n/k) \rightarrow \lambda$, finito. O mesmo resultado mantém-se válido para níveis tais que $\sqrt{k}A(n/k) \rightarrow \infty$, desde que também se verifique, $\sqrt{k}A(n/k)B(n/k) \rightarrow 0$, $\sqrt{k}A^2(n/k) \rightarrow 0$, $\hat{\beta} - \beta = o_p(1/(\sqrt{k}A(n/k)))$ e $(\hat{\rho} - \rho) \ln n = o_p(1/(\sqrt{k}A(n/k)))$.

2.1 Estimadores dos parâmetros de segunda ordem

A estimação do parâmetro de segunda ordem, ρ será feita através da classe de estimadores de ρ (Fraga Alves *et al.*, 2003),

$$\hat{\rho}_\tau(k) := -\left| \frac{3(T_n^{(\tau)}(k) - 1)}{T_n^{(\tau)}(k) - 3} \right|, \quad T_n^{(\tau)}(k) := \frac{(M_n^{(1)}(k))^\tau - (M_n^{(2)}(k)/2)^{\tau/2}}{(M_n^{(2)}(k)/2)^{\tau/2} - (M_n^{(3)}(k)/6)^{\tau/3}}, \quad (11)$$

para $\tau \neq 0$, com o usual prolongamento por continuidade para $\tau = 0$, e com $M_n^{(\alpha)}(k) := \frac{1}{k} \sum_{i=1}^k V_{ik}^\alpha$. Consideramos também o estimador de β introduzido em Gomes e Martins (2002), baseado nos log espaçamentos, W_i ,

$$\hat{\beta}(k; \hat{\rho}) := \frac{1}{n^{\hat{\rho}}} \frac{(\sum_{i=1}^k i^{-\hat{\rho}})(\sum_{i=1}^k W_i) - k(\sum_{i=1}^k i^{-\hat{\rho}} W_i)}{(\sum_{i=1}^k i^{-\hat{\rho}})(\sum_{i=1}^k i^{-2\hat{\rho}} W_i) - k(\sum_{i=1}^k i^{-\hat{\rho}} W_i)}. \quad (12)$$

Apresentamos a seguir alguns resultados, referentes ao comportamento assintótico destes estimadores dos parâmetros de segunda ordem.

Teorema 2 (Fraga Alves *et al.*, 2003). *Admitindo válida a condição de segunda ordem (2), com $\rho < 0$, k intermédia, isto é, verifica-se (5) e $\sqrt{k} A(n/k) \rightarrow \infty$, então $\hat{\rho}_\tau(k)$ em (11) converge em probabilidade para ρ , quando $n \rightarrow \infty$. Sob a condição de terceira ordem em (3),*

$$\hat{\rho}_\tau(k) - \rho \stackrel{d}{=} \frac{\gamma \sigma_\rho W_k^\rho}{\sqrt{k} A(n/k)} + (u_{\rho, \tau} A(n/k) + v_{\rho, \rho'} B(n/k))(1 + o_p(1)), \quad (13)$$

com W_k^ρ variável aleatória com distribuição assintótica normal padrão, $\sigma_\rho = \frac{(1-\rho)^3}{\rho} \sqrt{(2\rho^2 - 2\rho + 1)}$, e

$$u_{\rho, \tau} = \frac{\rho[\tau(1-2\rho)^2(3-\rho)(3-2\rho) + 6\rho(4(2-\rho)(1-\rho)^2 - 1)]}{12\gamma(1-\rho)^2(1-2\rho)^2}, \quad v_{\rho, \rho'} = \frac{\rho'(\rho+\rho')(1-\rho)^3}{\rho(1-\rho-\rho')^3}. \quad (14)$$

Consequentemente, se $\sqrt{k} A^2(n/k) \rightarrow \lambda_A$ e $\sqrt{k} A(n/k)B(n/k) \rightarrow \lambda_B$, finitos, então $\sqrt{k} A(n/k)(\hat{\rho}_\tau(k) - \rho)$ é assintoticamente normal com valor médio $\lambda_A u_{\rho, \tau} + \lambda_B v_{\rho, \rho'}$ e variância σ_ρ^2 .

Vamos agora considerar $k_1 = k_1^{opt}$, um nível “óptimo” para estimação de ρ . Então, $\sqrt{k_1} A^2(n/k_1) \rightarrow \lambda_{A_1}$ e $\sqrt{k_1} A(n/k_1)B(n/k_1) \rightarrow \lambda_{B_1}$, também finitos, sendo pelo menos um deles não nulo. Então, $k_1 = O(n^{-2(\rho+\rho')/(1-2(\rho+\rho'))})$ e $\hat{\rho}_\tau(k_1) - \rho$ é em probabilidade da ordem $1/(\sqrt{k_1} A(n/k_1)) = O(n^{\rho'/(1-2(\rho+\rho'))}) = o(1/\ln n)$, uma condição fundamental do Teorema 1.

Baseando-nos no Teorema 3 de Gomes *et al.*, (2008) podemos assegurar o seguinte: Admitindo válida a condição de segunda ordem (2), com $\rho < 0$, k intermédia, $A(t) = \gamma \beta t^\rho$, $\rho < 0$, $\sqrt{k} A(n/k) \rightarrow \infty$, e $\ln(n/k) = o(\sqrt{k} A(n/k))$ então, $\hat{\beta}(k; \hat{\rho})$ é consistente para a estimação de β . Se também se verificar $\sqrt{k} A^2(n/k) \rightarrow \lambda_A$ e $\sqrt{k} A(n/k)B(n/k) \rightarrow \lambda_B$ para $\hat{\rho}_\tau(k)$, podemos garantir que

$$\hat{\beta}_\rho(k) - \beta \stackrel{\mathcal{L}}{\rightarrow} -\beta \ln(n/k)(\hat{\rho}_\tau(k) - \rho). \quad (15)$$

6 Caeiro, F. et al./Variância de um estimador do índice de cauda

2.2 Estimadores MVRB

Seja $k_1 = k_1^{opt}$ o nível óptimo para a estimação de ρ e defina-se para qualquer τ real,

$$\tilde{H}_\tau(k, k_1) := \overline{H}_{\hat{\beta}(k_1; \hat{\rho}_\tau(k_1)), \hat{\rho}_\tau(k_1)}(k). \quad (16)$$

com $\overline{H}_{\hat{\beta}, \hat{\rho}}(k)$, $\hat{\rho}_\tau(k)$ e $\hat{\beta}(k; \hat{\rho})$ definidos em (8), (11) e (12), respectivamente. Com base nos Teoremas anteriores, podemos apresentar os seguintes resultados:

Teorema 3. *Para a classe de modelos verificando (3), com A e B escolhidas conforme indicado em (4) e se considerarmos $k_1 = k_1^{opt}$, óptimo para a estimação de ρ , podemos assegurar os seguintes resultados:*

1. *Se $k = o(k_1)$, $\sqrt{k}(\tilde{H}_\tau(k, k_1) - \gamma)$ é assintoticamente normal com valor médio nulo e variância igual a γ^2 .*
2. *Considere agora que $\sqrt{k}A(n/k) \rightarrow \infty$ e com λ_A e λ_B finitos, $\sqrt{k}A^2(n/k) \rightarrow \lambda_A$ e $\sqrt{k}A(n/k)B(n/k) \rightarrow \lambda_B$. Então $k/k_1 \rightarrow q > 0$, finito, e*

$$\sqrt{k}(\tilde{H}_\tau(k, k_1) - \gamma) \xrightarrow[n \rightarrow \infty]{d} \text{Normal}(b_q, \sigma_q^2),$$

com

$$b_q = \lambda_A(u_{\overline{H}} + u_{\rho, \tau} q^\rho a_{\overline{H}}(\ln q - a_{\overline{H}})) + \lambda_B(v_{\overline{H}} + v_{\rho, \rho'} q^{\rho'} a_{\overline{H}}(\ln q - a_{\overline{H}})),$$

e

$$\sigma_q^2 = \gamma^2(1 + q^{1-2\rho}(\ln q + 1/(1-\rho))^2(1-\rho)^4(2\rho^2 - 2\rho + 1)/\rho^2)$$

com $a_{\overline{H}} = -1/(1-\rho)$, $(u_{\overline{H}}, v_{\overline{H}})$ e $(u_{\rho, \tau}, v_{\rho, \rho'})$ definidos em (9) e (14), respectivamente. Obtemos assim a mesma velocidade de convergência, da ordem $1/\sqrt{k}$, para $\tilde{H}_\tau(k, k_1)$ com viés não nulo e variância assintótica dependente de q . A variância assintótica, σ_q^2 , é igual a γ^2 para $q = 0$ e $q = q_0 = \exp(-1/(1-\rho))$, com $e^{-1} < q_0 < 1$, e é uma função estritamente crescente para $q > q_0$.

Observação 1. *O comportamento de σ_q^2 , enquanto função de q , não depende de (γ, ρ) , e está ilustrado na Figura 1, para $(\gamma, \rho) = (1, -1)$. Quando $q \rightarrow 0$, σ_q^2 converge para γ^2 . A variância começa por aumentar até um valor ligeiramente superior a γ^2 , depois decresce novamente para γ^2 em $q_0 = \exp(-1/(1-\rho))$, e finalmente aumenta rapidamente a partir de q_0 , tomando o valor $\sigma_1^2 = \gamma^2(1 + ((1-\rho)/\rho)^2 - 2(1-\rho)^3/\rho)$ para $q = 1$.*

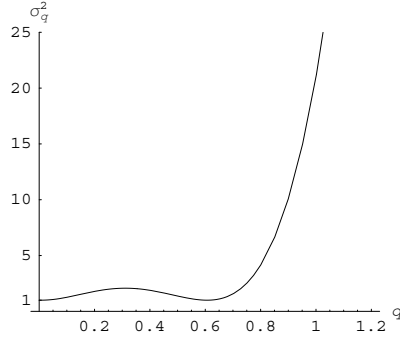


Figura 1: Padrão de $\sigma^2(q; \gamma, \rho)$, enquanto função de q , para $(\gamma, \rho) = (1, -1)$.

3 Estudo de simulação

Na prática, o Teorema 1 e a primeira parte do Teorema 3 são muito mais relevantes do que a segunda parte do Teorema 3. É mais fácil trabalhar com $k = o(k_1)$, com k_1 não necessariamente ótimo, mas verificando $(\hat{\rho}(k_1) - \rho) = o_p(1/\ln n)$ e estimar o índice de cauda, γ num nível que pode eventualmente ser a estimativa de k_0^H em (7), ou um nível ligeiramente superior, mas da mesma ordem. Do ponto de vista teórico, a segunda parte do Teorema 3 é mais interessante, pois permite-nos reduzir o erro quadrático médio do estimador MVRB. Precisamos apenas de conseguir estimar o nível ótimo $k_1 = k_1^{opt}$ do estimador do parâmetro de segunda ordem ρ . Depois com a estimativa \hat{k}_1 de k_1 e $\hat{\rho}$ de ρ , calculamos a estimativa do índice de cauda, γ , usando o estimador \bar{H} , no nível,

$$\hat{k} = \hat{k}_1^{opt} \times \exp(-1/(1 - \hat{\rho})). \quad (17)$$

Observação 2. A escolha do nível k_1 não é crucial para estimarmos os parâmetros de segunda ordem (β, ρ) . Devido à elevada estabilidade de $\hat{\rho}_\tau(k)$, junto ao valor de ρ , para k elevado e τ convenientemente escolhido, podemos estimar o parâmetro de segunda ordem ρ num nível de ordem $k_1^{(\epsilon)} = n^{1-\epsilon}$, com ϵ pequeno. Os resultados teóricos e de simulação em Fraga Alves et al. (2003), e a utilização de $\hat{\rho}$ em Caeiro and Gomes (2008), levaram-nos a considerar $\epsilon = 0.05$, isto é,

$$k_1^{(0.05)} = n^{0.995},$$

e a utilizar os parâmetros de controlo $\tau = 0$ para $\rho \geq -1$ e $\tau = 1$ para $\rho < -1$.

Uma vez que não conseguimos ainda estimar convenientemente k_1 , mas sabemos que para muitos dos modelos de cauda pesada, $\rho = \rho'$ e k_1 é da ordem $n^{-4\rho/(1-4\rho)}$, decidimos ignorar o factor de escala em $k_1 = C n^{-4\rho/(1-4\rho)}$ e considerar a estimação do índice de cauda em

$$\hat{k} = n^{-4\hat{\rho}/(1-4\hat{\rho})} \times \exp(-1/(1 - \hat{\rho})), \quad \text{com} \quad \hat{\rho} = \hat{\rho}_\tau(k_1^{(0.05)}). \quad (18)$$

8 Caeiro, F. *et al.*/Variância de um estimador do índice de cauda

Realizámos um estudo de simulação, através do método de Monte Carlo, para as seguintes distribuições com $\rho = \rho'$:

1. *t* de student com ν graus de liberdade com função de distribuição

$$F(x) = \frac{\Gamma((\nu+1)/2)}{\Gamma(\nu/2)\sqrt{\pi\nu}} \int_{-\infty}^x (1 + z^2/\nu)^{-(\nu+1)/2} dz, \quad x \in \mathbb{R},$$

onde Γ é a função Gama. Neste modelo $\gamma = 1/\nu$, e $\rho = \rho' = -2/\nu$.

2. *Burr*, parametrizada em função de γ e ρ , isto é, com função de distribuição

$$F(x) = 1 - (1 + x^{-\rho/\gamma})^{1/\rho}, \quad x > 0.$$

Também simulámos a distribuição de valores extremos com função de distribuição,

$$F(x) = \exp(-(1 + \gamma x)^{-1/\gamma}), \quad x \geq -1/\gamma, \quad \text{com } \gamma = 0.75.$$

Para este valor do índice de cauda temos $\rho = -\gamma = -0.75$ e $\rho' = \gamma - 1 = -0.25 \neq \rho$.

Para cada valor da dimensão da amostra $n \in \{100, 200, 500, 1000, 2000, 5000, 10000, 20000\}$, gerámos 10×5000 amostras de cada modelo. Depois calculámos as estimativas de $(\hat{\beta}, \hat{\rho}) = (\hat{\beta}(k_1^{(0.05)}, \hat{\rho}_\tau(k_1^{(0.05)})), \hat{\rho}_\tau(k_1^{(0.05)}))$ de (β, ρ) e simulámos o comportamento dos seguintes estimadores:

$$H(k), \quad \bar{H}_{\hat{\beta}, \hat{\rho}}(k), \quad H(\hat{k}_0^H), \quad \bar{H}_{\hat{\beta}, \hat{\rho}}(\hat{k}_0^H), \quad \bar{H}_{\hat{\beta}, \hat{\rho}}(\hat{k})$$

com $H(k)$, $\bar{H}_{\hat{\beta}, \hat{\rho}}(k)$, k_0^H e \hat{k} definidos em (6), (8), (7) e (18) respectivamente.

3.1 Valores médios e erros quadráticos médios como função de k

Apresentamos, para os modelos em estudo, os padrões de valores médios e erros quadráticos médios simulados de $H(k)$ e $\bar{H}_{\hat{\beta}, \hat{\rho}}(k)$ em (6) e (8), respectivamente, como função de k . Os resultados são apresentados nas Figuras 2, 3, 4 e 5, para $n = 1000$. Concluimos que o estimador \bar{H} apresenta para qualquer nível k , viés e erro quadrático médio inferior aos respectivos valores do estimador de Hill.

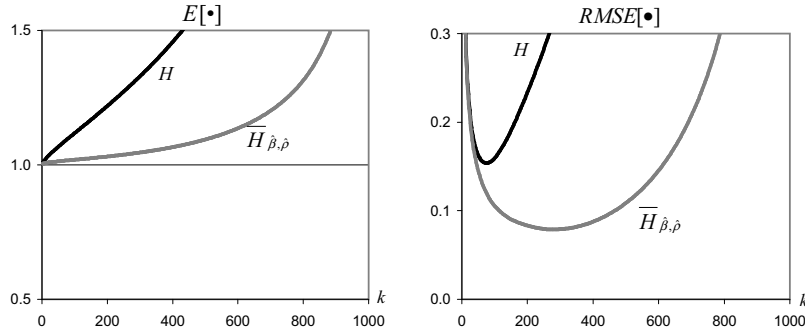


Figura 2: Valor médio e raiz quadrada do erro quadrático médio simulado, para amostras de dimensão $n = 1000$ do modelo Burr com $(\gamma, \rho) = (1, -1)$.

3.2 Valores médios e indicadores de eficiência dos estimadores adaptativos

Nas Tabelas 1 e 2 apresentamos os valores médios simulados (E), e o indicador de eficiência relativa ($REFF$) dos estimadores, $\bar{H}_{\hat{\beta}, \hat{\rho}}(\hat{k}_0^H)$ e $\bar{H}_{\hat{\beta}, \hat{\rho}}(\hat{k})$, relativamente a $H(\hat{k}_0^H)$. O indicador de eficiência de $\hat{\gamma}_1$ relativamente a $\hat{\gamma}_2$ é definido por:

$$REFF_{\hat{\gamma}_1|\hat{\gamma}_2} = \sqrt{MSE(\hat{\gamma}_2)/MSE(\hat{\gamma}_1)}.$$

Quanto maior que 1 for o valor do indicador REFF, melhor será o desempenho de $\hat{\gamma}_1$, relativamente a $\hat{\gamma}_2$.

Observação 3. *Os resultados apresentados permitem-nos fazer os seguintes comentários:*

1. *Nos modelos e dimensões de amostra considerados, os valores médios simulados de $\bar{H}_{\hat{\beta}, \hat{\rho}}(\hat{k}_0^H)$ e $\bar{H}_{\hat{\beta}, \hat{\rho}}(\hat{k})$ estão muito mais próximos do verdadeiro valor de γ , do que o valor médio simulado de $H(\hat{k}_0^H)$. Se analisarmos apenas o valor médio do estimador \bar{H} , concluímos que $|Viés(\bar{H}_{\hat{\beta}, \hat{\rho}}(\hat{k}_0^H))| < |Viés(\bar{H}_{\hat{\beta}, \hat{\rho}}(\hat{k}))|$, um resultado natural que é justificado por $\hat{k}_0^H = o(\hat{k})$.*
2. *Verificamos que o indicador de eficiência relativa é sempre superior a 1, em todos os modelos simulados. Isto significa que \bar{H} nos níveis \hat{k}_0 e \hat{k} é mais eficiente do que o estimador de Hill no seu nível ótimo. O estimador $\bar{H}_{\hat{\beta}, \hat{\rho}}(\hat{k})$ é o que apresenta maior eficiência, para amostras de dimensão maior ou igual a 1000 dos modelos t_2 , Burr e EV.*

10 Caeiro, F. *et al.*/Variância de um estimador do índice de cauda

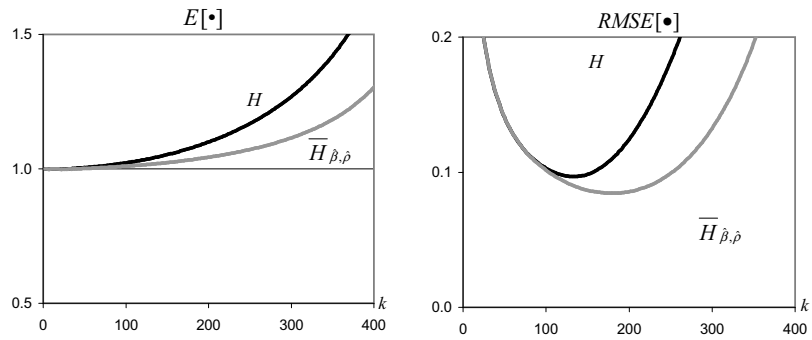


Figura 3: Valor médio e raiz quadrada do erro quadrático médio simulado, para amostras de dimensão $n = 1000$ do modelo t_1 .

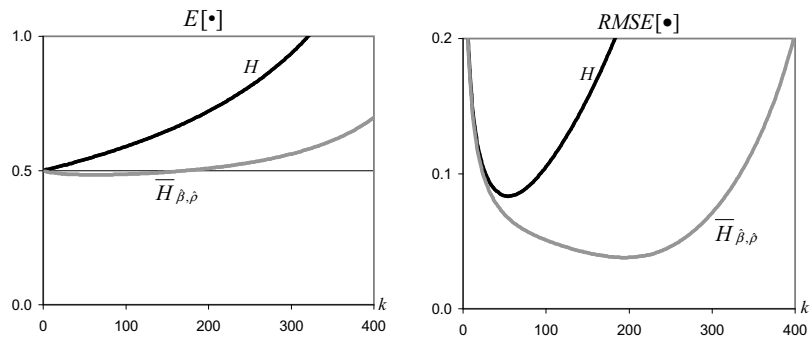


Figura 4: Valor médio e raiz quadrada do erro quadrático médio simulado, para amostras de dimensão $n = 1000$ do modelo t_2 .

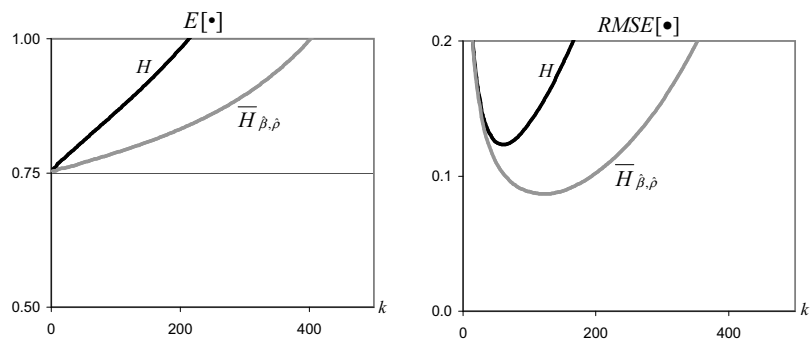


Figura 5: Valor médio e raiz quadrada do erro quadrático médio simulado, para amostras de dimensão $n = 1000$ do modelo EV com $\gamma = 0.75$.

Tabela 1: Valores médios simulados.

R	100	200	500	1000	2000	5000	10000	20000
t₁ (Cauchy): $\gamma = 1, \rho = -2, \rho' = -2$								
$H(\hat{k}_0^H)$	1.201	1.141	1.093	1.069	1.052	1.036	1.027	1.020
$\overline{H}_{\hat{\beta}, \hat{\rho}}(\hat{k}_0^H)$	1.092	1.061	1.040	1.029	1.022	1.015	1.011	1.009
$\overline{H}_{\hat{\beta}, \hat{\rho}}(\hat{k})$	1.067	1.053	1.043	1.036	1.031	1.025	1.021	1.018
t₂: $\gamma = 0.5, \rho = -1, \rho' = -1$								
$H(\hat{k}_0^H)$	0.633	0.592	0.560	0.543	0.532	0.522	0.517	0.513
$\overline{H}_{\hat{\beta}, \hat{\rho}}(\hat{k}_0^H)$	0.497	0.488	0.484	0.484	0.485	0.486	0.488	0.490
$\overline{H}_{\hat{\beta}, \hat{\rho}}(\hat{k})$	0.491	0.486	0.484	0.484	0.484	0.485	0.486	0.487
t₄: $\gamma = 0.25, \rho = -0.5, \rho' = -0.5$								
$H(\hat{k}_0^H)$	0.464	0.415	0.372	0.348	0.331	0.313	0.303	0.294
$\overline{H}_{\hat{\beta}, \hat{\rho}}(\hat{k}_0^H)$	0.359	0.337	0.318	0.307	0.299	0.290	0.284	0.280
$\overline{H}_{\hat{\beta}, \hat{\rho}}(\hat{k})$	0.343	0.330	0.318	0.310	0.304	0.297	0.293	0.289
Burr: $\gamma = 1, \rho = -0.75, \rho' = -0.75$								
$H(\hat{k}_0^H)$	1.237	1.184	1.134	1.107	1.086	1.064	1.052	1.042
$\overline{H}_{\hat{\beta}, \hat{\rho}}(\hat{k}_0^H)$	1.032	1.025	1.019	1.016	1.014	1.011	1.010	1.008
$\overline{H}_{\hat{\beta}, \hat{\rho}}(\hat{k})$	1.027	1.023	1.020	1.019	1.017	1.014	1.013	1.012
EV: $\gamma = 0.75, \rho = -0.75, \rho' = -0.25$								
$H(\hat{k}_0^H)$	0.935	0.890	0.850	0.828	0.812	0.797	0.787	0.780
$\overline{H}_{\hat{\beta}, \hat{\rho}}(\hat{k}_0^H)$	0.758	0.753	0.751	0.751	0.751	0.751	0.751	0.751
$\overline{H}_{\hat{\beta}, \hat{\rho}}(\hat{k})$	0.753	0.752	0.751	0.751	0.752	0.752	0.752	0.752

Tabela 2: Indicador de eficiência (REFF).

REFF	100	200	500	1000	2000	5000	10000	20000
t₁ (Cauchy): $\gamma = 1, \rho = -2, \rho' = -2$								
$\overline{H}_{\hat{\beta}, \hat{\rho}}(\hat{k}_0^H) H(\hat{k}_0^H)$	1.310	1.277	1.260	1.250	1.246	1.240	1.236	1.236
$\overline{H}_{\hat{\beta}, \hat{\rho}}(\hat{k}) H(\hat{k}_0^H)$	1.322	1.284	1.262	1.252	1.237	1.218	1.194	1.158
t₂: $\gamma = 0.5, \rho = -1, \rho' = -1$								
$\overline{H}_{\hat{\beta}, \hat{\rho}}(\hat{k}_0^H) H(\hat{k}_0^H)$	1.559	1.434	1.319	1.246	1.189	1.130	1.096	1.067
$\overline{H}_{\hat{\beta}, \hat{\rho}}(\hat{k}) H(\hat{k}_0^H)$	1.390	1.337	1.312	1.302	1.297	1.292	1.286	1.262
t₄: $\gamma = 0.25, \rho = -0.5, \rho' = -0.5$								
$\overline{H}_{\hat{\beta}, \hat{\rho}}(\hat{k}_0^H) H(\hat{k}_0^H)$	1.693	1.645	1.585	1.547	1.509	1.466	1.437	1.412
$\overline{H}_{\hat{\beta}, \hat{\rho}}(\hat{k}) H(\hat{k}_0^H)$	1.766	1.691	1.588	1.508	1.426	1.323	1.244	1.164
Burr: $\gamma = 1, \rho = -0.75, \rho' = -0.75$								
$\overline{H}_{\hat{\beta}, \hat{\rho}}(\hat{k}_0^H) H(\hat{k}_0^H)$	1.518	1.481	1.451	1.436	1.425	1.415	1.417	1.411
$\overline{H}_{\hat{\beta}, \hat{\rho}}(\hat{k}) H(\hat{k}_0^H)$	1.430	1.451	1.506	1.552	1.609	1.683	1.758	1.817
EV: $\gamma = 0.75, \rho = -0.75, \rho' = -0.25$								
$\overline{H}_{\hat{\beta}, \hat{\rho}}(\hat{k}_0^H) H(\hat{k}_0^H)$	1.488	1.443	1.395	1.369	1.352	1.339	1.334	1.327
$\overline{H}_{\hat{\beta}, \hat{\rho}}(\hat{k}) H(\hat{k}_0^H)$	1.360	1.371	1.406	1.446	1.496	1.577	1.651	1.737

Referências

- [1] Caeiro, F. and Gomes, M.I. (2008). Minimum-variance reduced-bias tail index and high quantile estimation. *Revstat* 6(1), 1-20.
- [2] Caeiro, F., Gomes, M.I. and Pestana, D.D. (2005). Direct reduction of bias of the classical Hill estimator. *Revstat* 3(2), 113-136.
- [3] Fraga Alves, M. I., Gomes, M. I. and de Haan, L. (2003). A new class of semi-parametric estimators of the second order parameter. *Portugaliae Mathematica* 60(1), 193-213.
- [4] Geluk, J. and de Haan, L. (1987). *Regular Variation, Extensions and Tauberian Theorems*. Tech. Report CWI Tract 40, Centre for Mathematics and Computer Science, Amsterdam, Netherlands.
- [5] Gnedenko, B. V. (1943). Sur la distribution limite du terme maximum d'une série aléatoire. *Ann. Math.* 44, 423-453.
- [6] Gomes, M.I., and Martins, M. J. (2002). "Asymptotically unbiased" estimators of the tail index based on external estimation of the second order. *Extremes* 5(1), 5-31.
- [7] Gomes, M.I., Pestana, D.D. and Caeiro, F. (2009). A Note on the Asymptotic Variance at Optimal Levels of a Bias-Corrected Hill Estimator. *Statistics and Probability Letters*, 79(3), 295-303.
- [8] Gomes, M.I., de Haan, L. and Henriques Rodrigues, L. (2008). Tail index estimation for heavy heavy-tailed models: Accomodation of bias in the weighted log-excesses. *Journal of Royal Statistical Society* B70(1), 31-52.
- [9] Hill, B.M. (1975). A simple general approach to inference about the tail of a distribution. *Ann. Statist.* 3, 1163-1174.