

Estimação de um índice de valores extremos positivo através de médias generalizadas e em ambiente de não-regularidade

M. Ivette Gomes

Centro de Estatística e Aplicações (CEAUL), Faculdade de Ciências,
Universidade de Lisboa, Portugal, *ivette.gomes@fc.ul.pt*

Lígia Henriques-Rodrigues

Centro de Investigação em Matemática e Aplicações (CIMA), Uni-
versidade de Évora, Portugal, *ligiahr@uevora.pt*

Dinis Pestana

CEAUL, Faculdade de Ciências, Universidade de Lisboa, Portugal,
dinis.pestana@fc.ul.pt

Palavras-chave: Caudas pesadas; Estimação semi-paramétrica;
Leis estáveis para extremos e somas; Soma de Paretos

Resumo: O estimador de Hill é o estimador mais popular de um *índice de valores extremos* (EVI, do inglês ‘*extreme value index*’) positivo, denotado por ξ . Trata-se de uma média aritmética, sendo conseqüentemente o logaritmo da média geométrica, i.e. da média-de-ordem-0, de estatísticas adequadas, função das estatísticas ordinais de topo associadas a uma amostra aleatória. Podemos mais geralmente considerar a *média-de-ordem-p* (MO_p) dessas estatísticas para $p \in \mathbb{R}$, e construir uma classe de estimadores- MO_p do EVI. Sob condições de regularidade adequadas, o comportamento assintoticamente normal dos estimadores MO_p foi obtido para $p < 1/(2\xi)$, tendo-se consistência para $p < 1/\xi$. Discutiremos aqui o comportamento teórico degenerado destes estimadores na região $p \geq 1/\xi$, sugerido por estudos de simulação, parcialmente apresentados.

1 Introdução e preliminares

O *índice de valores extremos* (EVI, do inglês ‘*extreme value index*’), denotado por ξ , é um dos parâmetros fundamentais em *teoria de valores extremos* (EVT, do inglês ‘*extreme value theory*’), e mede o peso da cauda-direita de F , $\bar{F}(x) := 1 - F(x)$, quando $x \rightarrow +\infty$, peso esse que aumenta com o crescimento de ξ .

Face a uma amostra aleatória, (X_1, \dots, X_n) , proveniente de uma *função de distribuição* (f.d.) F , designemos por $(X_{1:n}, \dots, X_{n:n})$ a amostra de estatísticas ordinais ascendentes associada. Estamos interessados em valores elevados, ou mais especificamente, no comportamento da sucessão de máximos, $\{X_{n:n}\}_{n \geq 1}$. A f.d. de $X_{n:n}$ é $\mathbb{P}(X_{n:n} \leq x) = F^n(x)$, expressão esta de pouca utilidade caso F seja desconhecida, e com um comportamento degenerado quando $n \rightarrow \infty$. Como estamos frequentemente interessados no máximo de um grande número de variáveis aleatórias (v.a.’s), modelamos esse máximo usando um argumento assintótico, que recorre a uma normalização linear desse máximo, de forma semelhante ao que se passa no caso da teoria assintótica para somas e no *teorema limite central* (TLC). É pois sensato colocar a pergunta: será possível encontrar constantes reais $\{a_n\}_{n \geq 1}$ ($a_n > 0$) e $\{b_n\}_{n \geq 1}$ tais que

$$\frac{X_{n:n} - b_n}{a_n} \xrightarrow[n \rightarrow \infty]{d} Y \iff F^n(a_n x + b_n) \xrightarrow[n \rightarrow \infty]{} G(x),$$

com $G(x)$ f.d. não degenerada? Se existirem tais sucessões de constantes, elas são designadas por coeficientes de atracção de F para G , e dizemos que F pertence ao max-domínio de atracção da lei G . O *teorema dos tipos extremais* (Gnedenko, [4]) permite-nos garantir que G é do tipo da distribuição geral de valores extremos

$$G(x) \equiv G_\xi(x) \equiv \text{EV}_\xi(x) = \exp \left\{ - [1 + \xi x]_+^{-1/\xi} \right\}, \quad (1)$$

com $x_+ := \max(0, x)$, sendo $\xi \in \mathbb{R}$ o EVI.

Para modelos de tipo-Pareto, i.e. modelos com uma cauda-direita de tipo polinomial negativo, e consequentemente, com um EVI positivo,

os estimadores clássicos do EVI são os estimadores de Hill (Hill, [9]). O estimador de Hill (H) é a média dos excessos das log-observações, i.e. de

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

O viés do estimador de Hill pode ser muito pequeno, moderado ou grande, e aumenta quando k aumenta. O problema de *trade-off* viés-variância existente na escolha do nível k motivou a necessidade de procurar estimadores alternativos para o índice de valores extremos. Note-se que podemos escrever

$$\begin{aligned} H(k) &:= \frac{1}{k} \sum_{i=1}^k V_{ik} =: \sum_{i=1}^k \ln U_{ik}^{1/k} = \ln \left(\prod_{i=1}^k U_{ik} \right)^{1/k}, \\ U_{ik} &= \frac{X_{n-i+1:n}}{X_{n-k:n}}, \quad 1 \leq i \leq k < n. \end{aligned} \quad (2)$$

Este estimador $H(k)$ pode pois ser encarado como o logaritmo da média geométrica, i.e. da média-de-ordem-0, das estatísticas U_{ik} , $1 \leq i \leq k < n$, em (2). Podemos pois mais geralmente considerar a *média-de-ordem-p* (MO_p) dessas estatísticas para $p \in \mathbb{R}$, e construir uma classe de estimadores- MO_p do EVI. A classe de funcionais em jogo é então

$$\begin{aligned} H_p(k) &\equiv MO_p(k) \\ &:= \begin{cases} \left(1 - \left(\frac{1}{k} \sum_{i=1}^k U_{ik}^p \right)^{-1} \right) / p, & \text{se } p \neq 0 \\ H(k), & \text{se } p = 0. \end{cases} \end{aligned} \quad (3)$$

Esta classe de funcionais MO_p , em (3), depende deste *parâmetro de controlo* p . A escolha adequada de p permite obter uma nova classe de estimadores do EVI com viés e erro quadrático médio (MSE, do inglês '*mean square error*') inferiores aos do estimador de Hill, tornando esta classe extraordinariamente flexível, tal como se pode ver

em Brillhante, Gomes e Pestana [2], Paulauskas e Vaičiulis [10, 11], Beran, Schell e Stehlik [1], Gomes e Caeiro [6], Caeiro, Gomes, Beirlant e de Wet [3], entre outros. Sob condições de regularidade adequadas, o comportamento assintoticamente normal de $MO_p(k)$ foi demonstrado para $p < 1/(2\xi)$, tendo-se obtido consistência para $p < 1/\xi$. O comportamento simulado destes estimadores na região $1/(2\xi) \leq p \leq 1/\xi$, parcialmente considerado em [2], leva-nos à consideração, na **Secção 2**, de um estudo simulação dos estimadores MO_p na região $p \geq 1/(2\xi)$, incluindo também a região $0 \leq p < 1/(2\xi)$. Na **Secção 3** demonstramos a consistência do estimador $MO_p(k)$ para $p = 1/\xi$, e finalizamos o artigo com alguns comentários, na **Secção 4**.

2 O método de Monte-Carlo como laboratório em Estatística

2.1 Simulação do estimador em situações de regularidade e de não regularidade

Foram implementadas simulações de Monte-Carlo multi-amostra, com dimensões 5000×20 , para amostras de dimensão n , com n a variar desde 100 até 5000, provenientes de diversos modelos de cauda-direita pesada, com $\xi > 0$, de entre os quais destacamos o modelo de valores extremos, com f.d. $F(x) = G_\xi(x) \equiv EV_\xi(x)$, definida em (1), o modelo generalizado de Pareto, com f.d. $F(x) \equiv GP_\xi(x) = 1 + \ln EV_\xi(x) = 1 - (1 + \xi x)^{-1/\xi}$ e o modelo t -de-Student com ν graus de liberdade ($\xi = 1/\nu$). Para detalhes sobre simulação multi-amostra, veja-se Gomes e Oliveira [7], entre outros.

2.1.1 Valores médios e MSE simulados, em função de k

Com base na primeira réplica de 5000 amostras de um modelo $EV_{0.25}$, começamos por ilustrar, na Figura 1, o comportamento do valor

médio (E) e do root MSE (RMSE) de $H(k) \equiv H_0(k)$ e de $H_p(k)$, $p = \ell/(10\xi)$, $\ell = 4, 9, 10$.

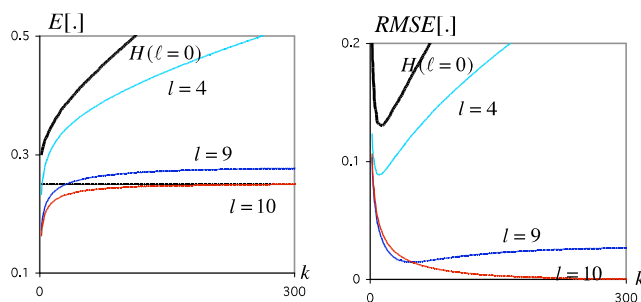


Figura 1: Valores médios (*esquerda*) e RMSEs simulados (*direita*), como função de k para modelos $EV_{0.25}$, $n = 1000$ e de $H_p(k)$ para $p = \ell/(10\xi)$, $\ell = 0, 4, 9, 10$ ($p = 0, 1.6, 3.6, 4$).

Para todos os outros modelos simulados, obtiveram-se figuras semelhantes à Figura 1.

2.1.2 Valores médios simulados em níveis ótimos

Apresentamos em seguida os valores médios simulados em níveis ótimos, no sentido de MSE mínimo, dos estimadores $H(k) \equiv H_0(k)$ e $H_p(k)$, $p = \ell/(10\xi)$, $\ell = 4, 5, 9, 10, 12$, bem como indicação sobre esses níveis ótimos. Não foram aqui incluídos valores negativos de p , que, em níveis ótimos, conduzem a MSE's mais elevados do que o MSE do estimador de Hill, mas que podem levar a estimadores mais robustos, como se pode ver em [1]. Esta questão levanta a necessidade de encontrar um indicador, que permita abordar de forma adequada o dilema eficiência versus robustez, tópico ainda em aberto. Fornecemos, nas tabelas 1, 2 e 3 esses valores médios simulados, respectivamente para modelo subjacente EV_ξ , $\xi = 0.1, 0.25, 0.5$, GP_ξ , $\xi = 0.1, 0.25, 0.5$ e Student- t_ν , $\nu = 4, 2$ ($\xi = 1/\nu = 0.25, 0.5$).

Assinalamos a **cheio** os valores médios simulados que conduzem a um viés mínimo, em valor absoluto.

De forma algo surpreendente, $p = 1/\xi$ ($\ell = 10$) conduz-nos a resultados muito interessantes. Mas se escolhermos $p > 1/\xi$, somos levados a valores médios perto de $1/p < \xi$, resultados que carecem de explicação, parcialmente fornecida na **Secção 3**. Previamente, apresentamos na Figura 2, uma ilustração gráfica dos valores simulados associados à Tabela 2, $\xi = 0.5$.

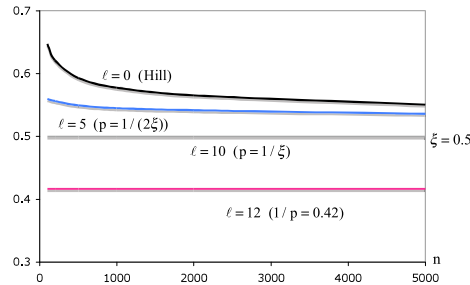


Figura 2: Valores médios simulados em níveis ótimos de $H_p(k)$ para $p = \ell/(10\xi)$, $\ell = 0, 5, 10, 12$ ($p = 0, 1, 2, 2.4$), em modelos $GP_{0.5}$.

2.1.3 Eficiência relativa simulada

Apresentamos ainda, nas tabelas 3–4–5, e de forma equivalente ao que fizemos nas tabelas 1–2–3 para valores médios simulados, um indicador de eficiência relativa: Face ao cálculo de H_{p0} , o estimador $H_p(k)$ calculado no valor estimado de $k_{0|p} := \arg \min_k \text{MSE}(H_p(k))$, o indicador em jogo é definido por

$$\text{REFF}_{p|0} = \frac{\text{RMSE}(H_{00})}{\text{RMSE}(H_{p0})} := \sqrt{\frac{\text{MSE}(H_{00})}{\text{MSE}(H_{p0})}} =: \frac{\text{RMSE}_{00}}{\text{RMSE}_{p0}}.$$

Sempre que este indicador for superior a 100, usamos a notação ∞ .

Tabela 1: Valores médios simulados em níveis ótimos (com intervalos de confiança a 95% associados), e parte inteira dos valores médios simulados desses níveis ótimos, colocados entre parêntesis na segunda linha, dos estimadores $H(k) \equiv H_0(k)$ e $H_p(k)$, $p = \ell/(10\xi)$, $\ell = 4, 5, 9, 10, 12$, para modelos EV_ξ , $\xi = 0.1, 0.25, 0.5$.

n	100	1000	5000
Modelo EV_ξ , $\xi = 0.1$			
$H (\ell = p = 0)$	0.334 ± 0.0009 (5)	0.223 ± 0.0016 (5)	0.195 ± 0.0011 (9)
$\ell = 4$	0.197 ± 0.0002 (5)	0.160 ± 0.0003 (5)	0.144 ± 0.0003 (5)
$\ell = 5 (p = 1/(2\xi))$	0.171 ± 0.0002 (5)	0.145 ± 0.0002 (5)	0.132 ± 0.0002 (5)
$\ell = 9$	0.108 ± 0.0001 (6)	0.107 ± 0.0001 (11)	0.106 ± 0.0001 (17)
$\ell = 10 (p = 1/\xi)$	0.100 ± 0.0001 (44)	0.100 ± 0.0001 (420)	0.100 ± 0.0001 (1756)
$\ell = 12$	0.083 ± 0.0000 (42)	0.083 ± 0.0000 (240)	0.083 ± 0.0000 (890)
Modelo EV_ξ , $\xi = 0.25$			
$H (\ell = p = 0)$	0.427 ± 0.0012 (5)	0.348 ± 0.0012 (13)	0.321 ± 0.0010 (27)
$\ell = 4$	0.336 ± 0.0007 (5)	0.301 ± 0.0013 (11)	0.294 ± 0.0008 (22)
$\ell = 5 (p = 1/(2\xi))$	0.312 ± 0.0006 (5)	0.289 ± 0.0011 (11)	0.285 ± 0.0007 (23)
$\ell = 9$	0.263 ± 0.0001 (14)	0.258 ± 0.0002 (48)	0.256 ± 0.0001 (115)
$\ell = 10 (p = 1/\xi)$	0.250 ± 0.0001 (44)	0.250 ± 0.0000 (574)	0.250 ± 0.0000 (3035)
$\ell = 12$	0.208 ± 0.0000 (44)	0.208 ± 0.0000 (561)	0.208 ± 0.0000 (2827)
Modelo EV_ξ , $\xi = 0.5$			
$H (\ell = p = 0)$	0.654 ± 0.0032 (9)	0.579 ± 0.0016 (35)	0.551 ± 0.0010 (86)
$\ell = 4$	0.578 ± 0.0022 (9)	0.556 ± 0.0014 (41)	0.541 ± 0.0005 (108)
$\ell = 5 (p = 1/(2\xi))$	0.559 ± 0.0017 (10)	0.546 ± 0.0011 (44)	0.536 ± 0.0008 (118)
$\ell = 9$	0.515 ± 0.0004 (29)	0.509 ± 0.0001 (172)	0.506 ± 0.0001 (601)
$\ell = 10 (p = 1/\xi)$	0.492 ± 0.0008 (44)	0.500 ± 0.0000 (574)	0.500 ± 0.0000 (3035)
$\ell = 12$	0.414 ± 0.0003 (44)	0.417 ± 0.0001 (574)	0.417 ± 0.0000 (3035)

Tabela 2: Valores médios simulados em níveis ótimos (com intervalos de confiança a 95% associados), e parte inteira dos valores médios simulados desses níveis ótimos, colocados entre parêntesis na segunda linha, dos estimadores $H(k) \equiv H_0(k)$ e $H_p(k)$, $p = \ell/(10\xi)$, $\ell = 4, 5, 9, 10, 12$, para modelos GP_ξ , $\xi = 0.1, 0.25, 0.5$.

n	100	1000	5000
Modelo GP_ξ , $\xi = 0.1$			
$H (\ell = p = 0)$	0.326 ± 0.0009 (5)	0.224 ± 0.0015 (5)	0.195 ± 0.0010 (9)
$\ell = 4$	0.195 ± 0.0003 (5)	0.160 ± 0.0004 (5)	0.144 ± 0.0003 (5)
$\ell = 5 (p = 1/(2\xi))$	0.170 ± 0.0002 (5)	0.145 ± 0.0003 (5)	0.132 ± 0.0002 (5)
$\ell = 9$	0.108 ± 0.0002 (7)	0.107 ± 0.0001 (11)	0.106 ± 0.0001 (17)
$\ell = 10 (p = 1/\xi)$	0.100 ± 0.0000 (75)	0.100 ± 0.0000 (534)	0.100 ± 0.0000 (2179)
$\ell = 12$	0.083 ± 0.0000 (51)	0.083 ± 0.0000 (277)	0.083 ± 0.0000 (978)
Modelo GP_ξ , $\xi = 0.25$			
$H (\ell = p = 0)$	0.419 ± 0.0024 (5)	0.347 ± 0.0016 (14)	0.320 ± 0.0011 (27)
$\ell = 4$	0.331 ± 0.0007 (5)	0.302 ± 0.0012 (11)	0.295 ± 0.0010 (23)
$\ell = 5 (p = 1/(2\xi))$	0.308 ± 0.0006 (5)	0.290 ± 0.0012 (11)	0.285 ± 0.0007 (23)
$\ell = 9$	0.262 ± 0.0003 (16)	0.258 ± 0.0002 (50)	0.256 ± 0.0001 (119)
$\ell = 10 (p = 1/\xi)$	0.250 ± 0.0000 (99)	0.250 ± 0.0000 (974)	0.250 ± 0.0000 (4781)
$\ell = 12$	0.208 ± 0.0000 (93)	0.208 ± 0.0000 (868)	0.208 ± 0.0000 (4223)
Modelo GP_ξ , $\xi = 0.5$			
$H (\ell = p = 0)$	0.647 ± 0.0043 (10)	0.578 ± 0.0012 (39)	0.551 ± 0.0011 (93)
$\ell = 4$	0.578 ± 0.0023 (11)	0.556 ± 0.0010 (45)	0.542 ± 0.0009 (118)
$\ell = 5 (p = 1/(2\xi))$	0.560 ± 0.0020 (12)	0.545 ± 0.0009 (48)	0.536 ± 0.0005 (130)
$\ell = 9$	0.514 ± 0.0004 (37)	0.508 ± 0.0002 (205)	0.506 ± 0.0001 (702)
$\ell = 10 (p = 1/\xi)$	0.500 ± 0.0000 (99)	0.500 ± 0.0000 (999)	0.500 ± 0.0000 (4998)
$\ell = 12$	0.417 ± 0.0001 (99)	0.417 ± 0.0000 (996)	0.417 ± 0.0000 (4975)

Tabela 3: Valores médios simulados em níveis ótimos (com intervalos de confiança a 95% associados), e parte inteira dos valores médios simulados desses níveis ótimos, colocados entre parêntesis na segunda linha, dos estimadores $H(k) \equiv H_0(k)$ e $H_p(k)$, $p = \ell/(10\xi)$, $\ell = 4, 5, 9, 10, 12$, para modelos Student- t_ν , $\nu = 4, 2$.

n	100	1000	5000
Modelo Student- t_4 , $\xi = 1/4 = 0.25$			
$H (\ell = p = 0)$	0.361 ± 0.0009 (5)	0.306 ± 0.0013 (20)	0.286 ± 0.0007 (48)
$\ell = 4$	0.296 ± 0.0005 (5)	0.284 ± 0.0007 (21)	0.277 ± 0.0004 (56)
$\ell = 5 (p = 1/(2\xi))$	0.285 ± 0.0026 (6)	0.277 ± 0.0006 (23)	0.272 ± 0.0005 (59)
$\ell = 9$	0.260 ± 0.0004 (16)	0.255 ± 0.0001 (85)	0.254 ± 0.0000 (279)
$\ell = 10 (p = 1/\xi)$	0.249 ± 0.0002 (31)	0.250 ± 0.0000 (442)	0.250 ± 0.0000 (2366)
$\ell = 12$	0.208 ± 0.0001 (31)	0.208 ± 0.0000 (442)	0.208 ± 0.0000 (2344)
Modelo Student- t_2 , $\xi = 1/2 = 0.5$			
$H (\ell = p = 0)$	0.602 ± 0.0039 (10)	0.544 ± 0.0008 (54)	0.526 ± 0.0005 (166)
$\ell = 4$	0.555 ± 0.0014 (12)	0.533 ± 0.0005 (67)	0.518 ± 0.0004 (206)
$\ell = 5 (p = 1/(2\xi))$	0.543 ± 0.0008 (13)	0.528 ± 0.0007 (75)	0.515 ± 0.0004 (234)
$\ell = 9$	0.511 ± 0.0013 (30)	0.506 ± 0.0001 (229)	0.503 ± 0.0001 (954)
$\ell = 10 (p = 1/\xi)$	0.474 ± 0.0018 (31)	0.498 ± 0.0002 (440)	0.500 ± 0.0001 (2367)
$\ell = 12$	0.406 ± 0.0009 (31)	0.416 ± 0.0001 (440)	0.417 ± 0.0000 (2367)

3 Uma breve justificação teórica

O modelo Pareto com parâmetro de forma unitário, ou seja, uma v.a. Y com f.d. dada por $F_Y(y) = 1 - y^{-1}$, $y \geq 1$, goza de papel fundamental em *estatística de extremos*, em situações muito variadas, mas também aquando da obtenção das propriedades de estimadores de parâmetros de acontecimentos raros, em contexto semi-paramétrico. Pensemos no funcional $MO_p(k)$, definido em (2).

Com a notação $F^{\leftarrow}(t) := \inf\{x : F(x) \geq t\}$ para a inversa generali-

Tabela 4: RMSEs simulados de H_{00}/ξ (primeira linha) e indicadores REFF de H_p , $p = \ell/(10\xi)$, $\ell = 4, 5, 9, 10, 12$ (com intervalos de confiança a 95%), para modelos EV_ξ , $\xi = 0.1, 0.25, 0.5$

n	100	1000	5000
Modelo EV_ξ , $\xi = 0.1$			
RMSE ₀₀	0.268 ± 0.1738	0.151 ± 0.1581	0.114 ± 0.1465
$\ell = 4$	2.581 ± 0.0062	2.052 ± 0.0045	1.856 ± 0.0062
$\ell = 5$ ($p = 1/(2\xi)$)	3.536 ± 0.0093	2.700 ± 0.0064	2.390 ± 0.0079
$\ell = 9$	26.790 ± 0.1198	17.123 ± 0.0798	14.030 ± 0.0641
$\ell = 10$ ($p = 1/\xi$)	∞	∞	∞
$\ell = 12$	16.054 ± 0.0586	9.065 ± 0.0367	6.816 ± 0.0306
Modelo EV_ξ , $\xi = 0.25$			
RMSE ₀₀	0.246 ± 0.1698	0.133 ± 0.1527	0.092 ± 0.1379
$\ell = 4$	1.879 ± 0.0048	1.463 ± 0.0066	1.300 ± 0.0040
$\ell = 5$ ($p = 1/(2\xi)$)	2.381 ± 0.0071	1.757 ± 0.0077	1.509 ± 0.0052
$\ell = 9$	12.793 ± 0.0681	8.989 ± 0.0493	7.331 ± 0.0242
$\ell = 10$ ($p = 1/\xi$)	∞	∞	∞
$\ell = 12$	5.906 ± 0.0313	3.183 ± 0.0157	2.213 ± 0.0086
Modelo EV_ξ , $\xi = 0.5$			
RMSE ₀₀	0.256 ± 0.1622	0.122 ± 0.1481	0.077 ± 0.1301
$\ell = 4$	1.530 ± 0.0044	1.221 ± 0.0054	1.123 ± 0.0048
$\ell = 5$ ($p = 1/(2\xi)$)	1.825 ± 0.0057	1.362 ± 0.0071	1.189 ± 0.0062
$\ell = 9$	8.554 ± 0.037	5.732 ± 0.0267	4.330 ± 0.0278
$\ell = 10$ ($p = 1/\xi$)	25.403 ± 2.2199	∞	∞
$\ell = 12$	2.991 ± 0.0202	1.463 ± 0.0059	0.923 ± 0.0036

zada de F e $U(t) := F^{\leftarrow}(1 - 1/t)$, podemos escrever

$$\begin{aligned} \frac{1}{k} \sum_{i=1}^k \left(\frac{X_{n-i+1:n}}{X_{n-k:n}} \right)^p &= \frac{1}{k} \sum_{i=1}^k \left(\frac{U(Y_{n-i+1:n})}{U(Y_{n-k:n})} \right)^p \\ &= \frac{1}{k} \sum_{i=1}^k Y_{k-i+1:k}^{\xi p} (1 + o_p(1)), \end{aligned}$$

com o termo $o_p(1)$ uniforme em i , $1 \leq i \leq k$ (veja-se [3]).

Tabela 5: RMSEs simulados de H_{00}/ξ (primeira linha) e indicadores REFF de H_p , $p = \ell/(10\xi)$, $\ell = 4, 5, 9, 10, 12$ (com intervalos de confiança a 95%), para modelos GP_ξ , $\xi = 0.1, 0.25, 0.5$

n	100	1000	5000
Modelo GP_ξ , $\xi = 0.1$			
RMSE ₀₀	0.259 ± 0.1790	0.151 ± 0.1580	0.113 ± 0.1464
$\ell = 4$	2.535 ± 0.0065	2.046 ± 0.0058	1.855 ± 0.0058
$\ell = 5$ ($p = 1/(2\xi)$)	3.467 ± 0.0097	2.691 ± 0.0085	2.390 ± 0.0078
$\ell = 9$	26.030 ± 0.1134	17.04 ± 0.1031	14.011 ± 0.0708
$\ell = 10$ ($p = 1/\xi$)	∞	∞	∞
$\ell = 12$	15.553 ± 0.0560	9.022 ± 0.0506	6.801 ± 0.0247
Modelo GP_ξ , $\xi = 0.25$			
RMSE ₀₀	0.237 ± 0.1756	0.131 ± 0.1525	0.092 ± 0.1377
$\ell = 4$	1.850 ± 0.0046	1.455 ± 0.0046	1.293 ± 0.0045
$\ell = 5$ ($p = 1/(2\xi)$)	2.333 ± 0.0065	1.747 ± 0.0053	1.501 ± 0.0056
$\ell = 9$	12.510 ± 0.0593	8.927 ± 0.0305	7.296 ± 0.0334
$\ell = 10$ ($p = 1/\xi$)	∞	∞	∞
$\ell = 12$	5.693 ± 0.0274	3.152 ± 0.0128	2.202 ± 0.0082
Modelo GP_ξ , $\xi = 0.5$			
RMSE ₀₀	0.239 ± 0.1751	0.118 ± 0.1479	0.075 ± 0.1294
$\ell = 4$	1.484 ± 0.0047	1.210 ± 0.0044	1.110 ± 0.0055
$\ell = 5$ ($p = 1/(2\xi)$)	1.761 ± 0.0058	1.345 ± 0.0056	1.175 ± 0.0073
$\ell = 9$	8.404 ± 0.0342	5.646 ± 0.0197	4.277 ± 0.0230
$\ell = 10$ ($p = 1/\xi$)	∞	∞	∞
$\ell = 12$	2.865 ± 0.0123	1.415 ± 0.0072	0.902 ± 0.0045

Conseqüentemente,

$$H_p(k) \equiv MO_p(k) \stackrel{d}{=} \frac{1}{p} \left(1 - \left(\frac{1}{k} \sum_{i=1}^k Y_i^{\xi p} (1 + o_p(1)) \right)^{-1} \right). \quad (4)$$

Como $\mathbb{E}(Y^a) = 1/(1 - a)$ se $a < 1$, a lei dos grandes números permite-nos garantir que se $p < 1/\xi$, $H_p(k)$ converge fracamente para ξ , quando $n \rightarrow \infty$. Para $p = 1/\xi$, temos pois em (4) uma soma de Paretos unitárias, como termo dominante. Com base no TLC generalizado ([5, 12]), podemos garantir que $\frac{1}{k} \sum_{i=1}^k Y_i - \ln k$ converge para uma estável para somas com um coeficiente de estabilidade

Tabela 6: RMSEs simulados de H_{00}/ξ (primeira linha) e indicadores REFF de H_p , $p = \ell/(10\xi)$, $\ell = 4, 5, 9, 10, 12$ (com intervalos de confiança a 95%), para modelos Student- t_ν , $\nu = 4, 2$

Modelo Student - t_4 , $\xi = 1/4 = 0.25$			
n	100	1000	5000
RMSE ₀₀	0.183 ± 0.1537	0.085 ± 0.1340	0.054 ± 0.1157
$\ell = 4$	1.750 ± 0.0065	1.329 ± 0.0050	1.182 ± 0.0045
$\ell = 5$ ($p = 1/(2\xi)$)	2.133 ± 0.0089	1.534 ± 0.0060	1.300 ± 0.0057
$\ell = 9$	10.390 ± 0.0650	7.023 ± 0.0207	5.344 ± 0.0294
$\ell = 10$ ($p = 1/\xi$)	77.870 ± 8.4682	∞	∞
$\ell = 12$	4.349 ± 0.0277	2.050 ± 0.0071	1.285 ± 0.0043
Modelo Student - t_2 , $\xi = 1/2 = 0.5$			
RMSE ₀₀	0.203 ± 0.5207	0.084 ± 0.1295	0.047 ± 0.1093
$\ell = 4$	1.433 ± 0.0052	1.196 ± 0.0032	1.184 ± 0.0038
$\ell = 5$ ($p = 1/(2\xi)$)	1.669 ± 0.0068	1.295 ± 0.0044	1.276 ± 0.0046
$\ell = 9$	6.734 ± 0.0836	4.441 ± 0.0197	3.725 ± 0.0185
$\ell = 10$ ($p = 1/\xi$)	6.339 ± 0.3947	35.243 ± 2.5342	∞
$\ell = 12$	2.149 ± 0.0285	0.997 ± 0.0050	0.564 ± 0.0023

$\alpha = 1$, ou seja uma v.a. de tipo Cauchy. Podemos pois escrever $H_p(k) = \xi(1 + O_p(1/\ln k))$, e consequentemente $H_p(k)$ continua a ser consistente para a estimação de ξ , quando $p = 1/\xi$.

Mas se $p > 1/\xi$ ($p\xi > 1$), como $\sum_{i=1}^k Y_i^{p\xi}/k = O_p(k^{p\xi-1}) \rightarrow \infty$, quando $k \rightarrow \infty$, $MO_p(k)$ converge para $1/p < \xi$, i.e. $MO_p(k)$ deixa de ser consistente para a estimação de ξ . Para detalhes sobre o comportamento não degenerado de $MO_p(k)$ na zona de não regularidade, i.e. quando $1/(2\xi) < p \leq 1/\xi$, veja-se [8].

4 Alguns comentários finais

É bem conhecido o facto de o estimador de Hill conduzir frequentemente a uma supra-estimação de ξ , mesmo quando se considera a escolha de k em níveis ótimos, no sentido de erro quadrático médio mínimo. A utilização de um parâmetro de controlo adicional, $p \in \mathbb{R}$ e da metodologia- MO_p leva-nos a uma mais fiável estimação

do EVI, através de estimadores assintoticamente normais, sempre que $p < 1/(2\xi)$. Sabemos agora que podemos avançar até $p = 1/\xi$, mas com um comportamento estável para somas e não normal, sempre que $1/(2\xi) < p \leq 1/\xi$, com índice de estabilidade $\alpha = 1/(p\xi)$. Vários métodos alternativos para a escolha de (p, k) têm sido delineados. Mas o tópico continua ainda parcialmente em aberto, devido ao perigo de a escolha de p poder ultrapassar $1/\xi$, o que levaria a situações de inconsistência.

Agradecimentos

Investigação financiada por Fundos Nacionais através da **FCT** – Fundação para a Ciência e a Tecnologia, projecto UIDB /MAT /00006 /2020 (CEA/UL) e UIDB/MAT/04674/2020 (CIMA).

Referências

- [1] Beran, J., Schell, D., Stehlik, M. (2014). The harmonic moment tail index estimator: asymptotic distribution and robustness. *Ann. Inst. Statist. Math.* 66, 193–220.
- [2] Brilhante, F., Gomes, M.I., Pestana, D. (2013). A simple generalization of the Hill estimator. *Computational Statistics and Data Analysis* 57:1, 518–535.
- [3] Caeiro, F., Gomes, M.I., Beirlant, J., de Wet, T. (2016). Mean-of-order- p reduced-bias extreme value index estimation under a third-order framework. *Extremes* 19:4, 561–589.
- [4] Gnedenko, B.V. (1943). Sur la distribution limite du terme maximum d’une série aléatoire. *Annals of Mathematics* 44:6, 423–453.
- [5] Gnedenko, B.V., Kolmogorov, A.N. (1954). *Limit Distributions for Sums of Independent Random Variables*, Addison-Wesley, Reading, MA.

- [6] Gomes, M.I., Caeiro, F. (2014). Efficiency of partially reduced-bias mean-of-order- p versus minimum-variance reduced-bias extreme value index estimation. In Gilli, M. *et al.* (eds.). *Proceedings of COMPSTAT 2014*, ISI/IASC, 289–298.
- [7] Gomes, M.I., Oliveira, O. (2001). The bootstrap methodology in Statistics of Extremes: choice of the optimal sample fraction. *Extremes* 4:4, 331–358.
- [8] Gomes, M.I., Henriques-Rodrigues, L., Pestana, D. (2020). *Non-regular Frameworks and the Mean-of-order p Extreme Value Index Estimation*. Submitted preprint available at: <https://doi.org/10.13140/RG.2.2.28347.64804>
- [9] Hill, B.M. (1975). A simple general approach to inference about the tail of a distribution. *Ann. Statist.* 3, 1163–1174.
- [10] Paulauskas, V., Vaičiulis, M. (2013). On the improvement of Hill and some others estimators. *Lith. Math. J.* 53, 336–355.
- [11] Paulauskas, V., Vaičiulis, M. (2017). A class of new tail index estimators. *Ann. Instit. Statist. Math.* 69, 661–487.
- [12] Samorodnitsky, G., Taqqu, M. (1994). *Stable Non-Gaussian Random Processes—Stochastic Models with Infinite Variance*, Chapman & Hall, USA.