XII Congreso Galego de Estatística e Investigación de Operacións
Lugo, 22, 23 y 24 de octubre de 2015

On the Hill estimator: a comparison of methods

Marta Ferreira[1] and Márcio Rebelo[2]

[1]Center of Mathematics of Minho University and CEMAT
[2]Department of Mathematics and Applications of University of Minho

ABSTRACT

Extreme value theory (EVT) deals with the occurrence of extreme phenomena. The tail index is a very important parameter appearing in the estimation of the probability of rare events. Under a semiparametric framework, inference requires the choice of a number k of upper order statistics to be considered. This is the crux of the matter and there is no definite formula to do it, since a small k leads to high variance and large values of k tend to increase the bias. Several methodologies have emerged in literature, specially concerning the most popular Hill estimator (Hill, 1975). In this work we compare through simulation well-known procedures presented in Drees and Kaufmann (1998), Matthys and Beirlant (2000), Beirlant *et al.* (2002) and de Sousa and Michailidis (2004), with a heuristic scheme considered in Frahm *et al.* (2005) within the estimation of a different tail measure but with a similar context. We will see that the new method may be an interesting alternative.

Keywords: Extreme value theory, Tail index estimation Monte-Carlo simulations

1. INTRODUCTION

Extreme value statistics are being increasingly used in the environment, engineering, finance, among other sciences, given the need to account for the possible occurrence of extreme phenomena in the modeling.

Let $\{X_n\}_{n\geq 1}$ be a sequence of independent and identically distributed (iid) random variables (rv's), with common distribution function (df) F. If for some sequences of real constants $a_n > 0$ and b_n, $n \geq 1$, the limit

$$P(\max(X_1,\ldots,X_n) \leq a_n x + b_n) \underset{n\to\infty}{\longrightarrow} G_\xi(x)$$

exists for some non-degenerate function G_ξ with ξ real, then F is said to belong to the max-domain of attraction of G_ξ, denoted $F \in \mathcal{D}(G_\xi)$. This function is called generalized extreme value (GEV) and is given by

$$G_\xi(x) = \exp(-(1+\xi x)^{-1/\xi}), \quad 1+\xi x > 0,$$

($G_0(x) = \exp(-e^{-x})$). Parameter ξ is the so-called tail index and determines the shape and weight of the tail. Thus, $\xi > 0$ implies a heavy tail (Fréchet max-domain of attraction) with polynomial decay and infinite right-end-point, whenever null means an exponential tail (Gumbel max-domain of attraction) and $\xi < 0$ corresponds to a light tail (Weibull max-domain of attraction) with finite right-end-point.

The tail index plays a determinant role concerning the inference within rare events like, for instance, the estimation of an unusual high quantile (e.g. the Value-at-Risk in finance) or the dual problem of estimating the probability of exceeding a high level x, i.e., $p = 1 - F(x)$. The Hill estimator (Hill 1975),

$$\widehat{\xi}_{k,n} := \frac{1}{k}\sum_{i=1}^{k} \log \frac{X_{n-i+1:n}}{X_{n-k:n}},$$

with $X_{n:n} \geq X_{n-1:n} \geq ... \geq X_{1:n}$ being the order statistics of $\{X_n\}_{n\geq 1}$, is perhaps the most applied in literature and assumes that F belongs to max-domain of attraction of a heavy tail ($\xi > 0$). Observe that it depends on the k upper order statistics, where $k \equiv k_n \to \infty$ and $k/n \to 0$, as $n \to \infty$ (i.e., $k \equiv k_n$ is an intermediate sequence), in order to achieve consistency.

The choice of k is the estimation core, being somewhat ambiguous, since small values of k results in greater variance and large values lead to an increased bias. The plot $(k, \widehat{\xi}_{k,n})$, $1 \leq k < n$, usually shows a plateau region from which we may infer the values of k where $\widehat{\xi}_{k,n}$ approximates the true value (see Figure 1). This problem often accompanies extreme-values semi-parametric inference and has been much studied in the literature. See, e.g., Beirlant *et al.* (2004) for a survey.

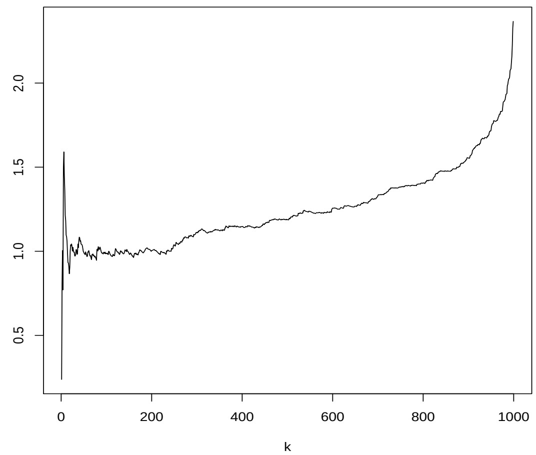

Figure 1: Hill plots of 1000 realizations of a Fréchet model with $\xi = 1$.

For non-null ξ, Beirlant *et al.* (2002) stated, under little restrictive conditions,

$$Y_i := (i+1)\log\frac{X_{n-i:n}\,\widehat{\xi}_{i,n}}{X_{n-(i+1):n}\,\widehat{\xi}_{i+1,n}} = \xi + b(n/k)\left(\frac{i}{k}\right)^{-\rho} + \epsilon_i,\ i = 1, ..., k, \tag{1}$$

where the error term ϵ_i is zero-centered and b is a positive function such that $b(x) \to 0$, as $x \to \infty$. The so called second order tail parameter ρ is usually replaced by $\rho = -1$. For more details, see Beirlant *et al.* 2004 and references therein. Least squares estimators of ξ and $b(n/k)$ derived from (1) are expressed as

$$\widetilde{\xi}^{LS}_{k,n} = \overline{Y}_k - \widetilde{b}^{LS}_{k,n}/(1-\rho) \quad \text{and} \quad \widetilde{b}^{LS}_{k,n} = \tfrac{(1-\rho)^2(1-2\rho)}{\rho^2}\tfrac{1}{k}\sum_{i=1}^{k}\left(\left(\tfrac{i}{k}\right)^{-\rho} - \tfrac{1}{1-\rho}\right)Y_i. \tag{2}$$

One method consists in finding the k value that corresponds to the smallest estimate of the Hill's asymptotic mean squared error (AMSE)

$$\text{AMSE}(\widehat{\xi}_{k,n}) = \tfrac{\xi^2}{k} + \left(\tfrac{b(n/k)}{1-\rho}\right)^2. \tag{3}$$

More precisely, for fixed $\rho = -1$, compute the least squares estimates of ξ and $b(n/k)$ by using (2). Obtain the Hill's AMSE in (3) by replacing ξ and $b(n/k)$ with the respective estimates. Take $\widehat{k}^1_{opt}$ as the value of k that minimizes the obtained estimates of the AMSE and estimate ξ as $\widehat{\xi}_{\widehat{k}^1_{opt},n}$.

Alternatively, the value of k for which the respective AMSE is minimal is

$$k_{opt} \sim b(n/k)^{-2/(1-2\rho)}k^{-2\rho/(1-2\rho)}\left(\tfrac{\xi^2(1-\rho)^2}{-2\rho}\right)^{1/(1-2\rho)} \tag{4}$$

Consider again the least squares estimates for ξ and $b(n/k)$ by using (2) and fixed $\rho = -1$, for $k = 3, ..., n$. Compute $\widehat{k}_{opt,k}$ according to the expression (4). Take $\widehat{k}^2_{opt} = median\{\widehat{k}_{opt,k},\ k = 3, ..., \lfloor\frac{n}{2}\rfloor\}$, where $\lfloor x\rfloor$ denotes the largest integer not exceeding x. Estimate ξ by $\widehat{\xi}_{\widehat{k}^2_{opt},n}$.

For more details about these two methods, see Beirlant *et al.* 2002 Matthys and Beirlant (2000) and references therein. We denote them, respectively, AMSE and Kopt.

The adaptive procedure of Drees and Kaufmann (1998), here denoted DK, is based on results for a class of Generalized Pareto models, known has Hall's class, where the maximum random fluctuation of $\sqrt{i}(\widehat{\xi}_{i,n} - \xi)$, $i = 1, ..., k-1$, with $k \equiv k_n$ an intermediate sequence, is of order $\sqrt{\log\log n}$. The derived stopping time criterion seeks the optimum k under which the bias starts to dominate the variance. More precisely, for $\rho = -1$:

1. State $r_n = 2.5 \times \widetilde{\xi} \times n^{0.25}$, where $\widetilde{\xi} = \widehat{\xi}_{2\sqrt{n},n}$.

2. Compute $\widetilde{k}(r_n) := \min\{k = 1, ..., n-1 : \max_{i<k} \sqrt{i}|\widehat{\xi}_{i,n} - \widehat{\xi}_{k,n}| > r_n\}$. If $\sqrt{i}|\widehat{\xi}_{i,n} - \widehat{\xi}_{k,n}| > r_n$ is not satisfied for any k, impute $0.9 \times r_n$ to r_n and repeat step 2. Else move to step 3.

3. For $\varepsilon \in (0,1)$, usually $\varepsilon = 0.7$, obtain

$$\widehat{k}_{\texttt{DK}} = \left\lfloor \frac{1}{3}(2\widetilde{\xi}^2)^{1/3} \left(\frac{\widetilde{k}(r_n^\varepsilon)}{(\widetilde{k}(r_n))^\varepsilon} \right)^{1/(1-\varepsilon)} \right\rfloor$$

The procedure considered in Sousa and Michailidis (2004) looks for an estimate of the optimal value k based on the Hill sum plot, (k, S_k), $k = 1, ..., n-1$, where $S_k = k\widehat{\xi}_{k,n}$. Since $E(S_k) = k\xi$, then the sumplot is expected to be approximately linear for the values of k where $\widehat{\xi}_{k,n} \approx \xi$. The slope of this linear part is an estimator of ξ. Thus, an algorithm was developed to establish where there is a breakdown of linearity. Considering the regression model $y = X\xi + \delta$, with $y = (S_1, ..., S_k)'$, $X = [1\, i]_{i=1}^k$ and δ the error term, the method is based on a sequential testing procedure discussed in McGee and Carleton (1970). This process checks the null hypothesis that a new point y_0 is adjacent to the left or to the right of the set of points $y = (y_1, ..., y_k)$, through the statistics

$$F = s^{-2}\left((y_0 - \widehat{y}_0^*)^2 + \sum_{i=1}^k (\widehat{y}_i - \widehat{y}_i^*)^2 \right),$$

where $*$ denotes the predictions based on $k+1$ and $s^2 = (k-2)^{-1}(y'y - \widehat{\xi}X'y)$. We have F approximately distributed by $F_{1,k-2}$ and thus the null hypothesis is rejected whenever F is larger than the $(1-\alpha)$-quantile, $F_{1,k-2;1-\alpha}$. The algorithm runs along the following steps:

1. Fit a least-squares regression line to the initial $k = \nu n$ upper observations, $y = [y_i]_{i=1}^k$ (usually $\nu = 0.02$).

2. Using the test statistic F, determine if a new point $y_0 = y_j$ for $j > k$, belongs to the original set of points y. Go adding points until the null hypothesis is rejected.

3. Consider $k_{\texttt{new}} = \max(0, \{j : F < F_{1,k-2;1-\alpha}\})$. If $k_{\texttt{new}} = 0$, no new points are added to y and thus move forward to step 4. Return to step 1. if $k_{\texttt{new}} > 0$ by considering $k = k_{\texttt{new}}$.

4. Estimate ξ by considering the obtained k.

Although in a different context concerning the estimation of a bivariate tail dependence measure, but with a similar problematic, Frahm *et al.* (2005) presented a heuristic procedure aiming to find the plane area of the estimator sample path. Ferreira (2014, 2015) adapted this method to the Hill estimator, whose algorithm is described below:

1. Smooth the Hill plot $(k, \widehat{\xi}_{k,n})$ by taking the means of $2b+1$ successive points, $\overline{\widehat{\xi}}_{1,n}, ..., \overline{\widehat{\xi}}_{n-2b,n}$, where the bandwidth $b = \lfloor 0.005 \times n \rfloor$ (and thus each moving average consists in 1% of the data).

2. Define the regions $p_k = (\overline{\overline{\xi}}_{k,n}, ..., \overline{\overline{\xi}}_{k+m-1,n})$, $k = 1, ..., n-2b-m+1$, with length $m = \lfloor\sqrt{n-2b}\rfloor$. The algorithm stops at the first region satisfying

$$\sum_{i=k+1}^{k+m-1} \left|\overline{\overline{\xi}}_{i,n} - \overline{\overline{\xi}}_{k,n}\right| \leq 2s,$$

where s is the empirical standard-deviation of $\overline{\overline{\xi}}_{1,n}, ..., \overline{\overline{\xi}}_{n-2b,n}$.

3. Consider the chosen plateau region p_k* and estimate ξ as the mean of the values of p_k* (consider the estimate zero if no plane region fulfills the stopping condition).

In the sequel it will be referred as plateau method.

In this work we compare through simulation the performance of these methods within the estimation of the tail index and the exceedance probability. A simple estimator under $\xi > 0$ is given by (Dijk and de Haan, 1992)

$$\widehat{p}_{k,n} := \frac{k}{n}\left(\frac{x}{X_{n-k:n}}\right)^{-1/\widehat{\xi}}.$$

In this measure, the plateau algorithm will be applied to the respective sample path $(k, \widehat{p}_{k,n})$ in a similar manner of the variance estimation considered in Ferreira (2015). More precisely, we choose the plane region in step 2 at the same position of the one found for the ξ estimation.

2. SIMULATION STUDY

We consider 1000 independent samples of sizes $n = 100, 1000$, generated from the following models:

- Pareto(ξ) with d.f. $F(x) = 1 - x^{-1/\xi}$, $x > 1$, $\xi > 0$: ($\xi = 1$);
- Cauchy with d.f. $F(x) = \frac{1}{2} + \frac{1}{\pi}\arctan x$, x real: ($\xi = 1$);
- Burr(β,τ,λ), with d.f. $F(x) = 1 - (\beta/(\beta + x^\tau))^\lambda)$, $x > 0$, $\beta, \tau, \lambda > 0$, $\xi = 1/(\tau\lambda)$: we take Burr(1,1,1) ($\xi = 1$)
- Fréchet(ξ) with d.f. $F(x) = \exp\left(-x^{-1/\xi}\right)$, $x > 0$, $\xi > 0$: ($\xi = 1$);
- Log-Gamma(τ,λ), with d.f. $F(x) = \int_1^x \frac{\lambda^\tau}{\Gamma(\tau)}(\log t)^{\tau-1}t^{-\lambda-1}dt$, $x \geq 0$, $\tau, \lambda > 0$, $\xi = 1/\lambda$: we take Log-Gamma(2,1) ($\xi = 1$).

The bias and root mean squared error (rmse) values obtained for the tail index are reported in Table 1. The results concerning the exceedance probabilities, $p = 0.001, 0.0001$, respectively in the cases $n = 100, 1000$, are reported in Table 2. Regarding the tail index, the AMSE method presents the smaller rmse, followed by the DK except in the Pareto model. Indeed, this latter methodology was developed for the so-called Hall's class that leaves out the simple Pareto case considered here. The plateau procedure has also a good performance, in particular, for small sample sizes. The worst performance for $n = 100$ is associated with the Kopt method. The sumplot methodology was developed under the assumption of a Pareto tail behavior and therefore it is not surprising its best performance within the Pareto model. However, it revealed to be the worst method in the remaining models, in particular, for $n = 1000$. In what concerns the exceedance probability, the best results for $n = 100$ lie in the plateau and Kopt methods. In the case $n = 1000$, the methods have a broadly similar performance, except the sumplot with the highest rmse.

3. Acknowledgments

The first author research was supported by the Research Centre CEMAT through the Project UID/Multi/04621/2013.

bias (n=100)	AMSE	Kopt	DK	Plateau	Sumplot
Burr	0.1136	0.1798	0.1338	0.0424	0.3085
Fréchet	0.0101	0.1489	0.0827	0.0362	0.2158
Cauchy	0.0070	0.1121	0.0282	-0.0306	0.2166
Log-Gamma	0.1512	0.2918	0.1803	0.2242	0.4240
Pareto	-0.0882	0.0349	-0.1396	-0.0139	-0.0311
rmse (n=100)	AMSE	Kopt	DK	Plateau	Sumplot
Burr	0.2848	0.4546	0.2887	0.3058	0.4243
Fréchet	0.2261	0.4610	0.2371	0.2606	0.3327
Cauchy	0.2953	0.5366	0.2991	0.3737	0.4118
Log-Gamma	0.3003	0.5708	0.3595	0.3523	0.5045
Pareto	0.1975	0.3665	0.4256	0.2010	0.1262
bias (n=1000)	AMSE	Kopt	DK	Plateau	Sumplot
Burr	0.0710	0.0475	0.0662	0.0142	0.3123
Fréchet	-0.0122	0.0620	0.0577	0.0107	0.2735
Cauchy	0.0199	0.0194	0.0064	-0.0147	0.1707
Log-Gamma	0.1496	0.1896	0.2646	0.2025	0.5249
Pareto	-0.0402	0.0187	-0.0901	-0.0002	-0.0002
rmse (n=1000)	AMSE	Kopt	DK	Plateau	Sumplot
Burr	0.1319	0.1303	0.1279	0.1465	0.3372
Fréchet	0.1032	0.2446	0.1069	0.1166	0.2927
Cauchy	0.1115	0.1470	0.1152	0.1765	0.2124
Log-Gamma	0.1957	0.3568	0.2833	0.2380	0.5404
Pareto	0.0860	0.2116	0.3648	0.0760	0.0321

Table 1: Bias and rmse of the tail index estimation.

REFERENCES

Beirlant, J., Dierckx, G., Guillou, A. and Stărică, C. (2002) On Exponential Representation of Log-Spacings of Extreme Order Statistics. Extremes, 5, 157-180.

Beirlant,J., Goegebeur, Y., Segers, J. and Teugels, J.L. (2004). Statistics of Extremes: Theory and Applications. J. Wiley & Sons.

de Sousa, B. and Michailidis, G. (2004) A Diagnostic Plot for Estimating the Tail Index of a Distribution. Journal of Computational and Graphical Statistics, 13(4), 1-22.

Dijk, V. and de Haan, L. (1992). On the estimation of the exceedance probability of a high level. Order statistics and nonparametrics: Theory and applications. In Sen, P.K. and Salama I.A. (Editors), 79-92. Elsevier, Amsterdam.

Drees, H. and Kaufmann, E. (1998) Selecting the optimal sample fraction in univariate extreme value estimation. Stochastic Process Appl., 75, 149-172.

Ferreira, M. (2014) A Heuristic Procedure to Estimate the Tail Index, Proceedings of the 14th International Conference in Computational Science and Its Applications - ICCSA 2014, June 30 - July 3 (2014), Guimares, Portugal, IEEE-Computer Society 4264a241, 241-245.

Ferreira, M. (2015) Estimating the tail index: Another algorithmic method. ProbStat Forum, 08, 45-53.

Frahm, G., Junker, M. and Schmidt R. (2005) Estimating the tail-dependence coefficient: properties and pitfalls. Insurance: Mathematics & Economics, 37(1), 80-100.

Hill, B.M. (1975). A Simple General Approach to Inference About the Tail of a Distribution. Ann. Stat., 3, 1163-1174.

Matthys, G. and Beirlant, J. (2000) Adaptive Threshold Selection in Tail Index Estimation. In: Extremes and Integrated Risk Management, ed. P. Embrechts, London: Risk Books.

bias (n=100)	AMSE	Kopt	DK	Plateau	Sumplot
Burr	0.0002	-0.0008	-0.0004	-0.0009	0.0056
Fréchet	0.0002	-0.0008	0.0001	-0.0006	0.0118
Cauchy	0.0017	-0.0004	0.0009	-0.0006	0.0174
Log-Gamma	0.0003	-0.0008	0.0025	-0.0004	0.0338
Pareto	-0.0004	-0.0007	-0.0008	-0.0004	0.0381
rmse (n=100)	AMSE	Kopt	DK	Plateau	Sumplot
Burr	0.0020	0.0009	0.0012	0.0009	0.0130
Fréchet	0.0011	0.0009	0.0015	0.0009	0.0241
Cauchy	0.0046	0.0012	0.0039	0.0011	0.0335
Log-Gamma	0.0011	0.0009	0.0053	0.0010	0.0878
Pareto	0.0008	0.0009	0.0009	0.0008	0.0603
bias (n=1000)	AMSE	Kopt	DK	Plateau	Sumplot
Burr	0.0001	-0.0001	-0.0001	-0.0001	0.0004
Fréchet	0.0001	-0.0001	-0.0001	-0.0001	0.0012
Cauchy	0.0000	-0.0001	-0.0007	-0.0001	0.0008
Log-Gamma	0.0000	-0.0001	0.0003	-0.0001	0.0037
Pareto	-0.0001	-0.0001	-0.0001	-0.0001	0.0503
rmse (n=1000)	AMSE	Kopt	DK	Plateau	Sumplot
Burr	0.0001	0.0001	0.0001	0.0001	0.0007
Fréchet	0.0001	0.0001	0.0001	0.0001	0.0016
Cauchy	0.0001	0.0001	0.0008	0.0001	0.0012
Log-Gamma	0.0001	0.0001	0.0005	0.0001	0.0051
Pareto	0.0001	0.0001	0.0001	0.0001	0.0729

Table 2: Bias and rmse of the exceedance probability estimation, considering $p = 0.001$ for $n = 100$ and $p = 0.0001$ for $n = 1000$.

McGee, V.E. and Carleton, W.T. (1970) Piecewise Regression. Journal of the American Statistical Association, 65, 11091124.