`aste`: An R package for the adaptive estimation the right tail

Frederico Caeiro[a]

[a]Faculdade de Ciências e Tecnologia & CMA
Universidade Nova de Lisboa
2829-516 Caparica, Portugal
fac@fct.unl.pt

Keywords : Extreme value index, semi-parametric estimation, bias reduction.

Let us consider the common set-up of independent, identically distributed (i.i.d.) random variables (r.v.'s) $X_1, X_2, \cdots, X_n$, with a common distribution function (d.f.) F and denote the associated ascending order statistics (o.s.) by $X_{1:n} \leq X_{2:n} \leq \cdots \leq X_{n:n}$. Let us assume that there exist sequences of real constants $\{a_n > 0\}$ and $\{b_n \in \mathbb{R}\}$ such that the maximum, linearly normalized, i.e., $(X_{n:n} - b_n)/a_n$ converges in distribution towards a non-degenerate r.v. Then F belongs to the max-domain of attraction of an *extreme value* (EV) d.f.,

$$EV_\gamma(x) = exp(-(1+\gamma x)^{-1/\gamma}), \quad 1+\gamma x > 0, \quad \gamma \in \mathbb{R}.$$

and we write $F \in \mathcal{D}_\mathcal{M}(EV_\gamma)$. The parameter γ is the *extreme value index* (EVI), the primary parameter of extreme events, with a low frequency, but a high impact. This index measures the heaviness of the right *tail function* $\overline{F} := 1 - F$, and the heavier the tail, the larger γ is.
The EVI needs to be estimated in a precise way, because such an estimation is one of the basis for the estimation of other parameters of extreme and large events, like a *high quantile* of probability $1-p$, with p small, the *right endpoint* of the model F underlying the data, $x^F := \sup\{x : F(x) < 1\}$, whenever finite, and the *return period* of a high level, among others.
We will work with the $k+1$ top o.s.'s associated to the n available observations, assuming only that the model F underlying the data is in $\mathcal{D}_\mathcal{M}(G_\gamma)$. Most of the classical semi-parametric estimators of any parameter of extreme events have a strong bias for moderate up to large values of k, the number of top o.s.'s involved in the estimation, including the optimal k, in the sense of minimal mean squared error (MSE). Accommodation of bias of classical estimators of parameters of extreme events has been deeply considered in the recent literature. For the estimation of a negative or eventually zero EVI ($\gamma \leq 0$), we refer the recent *negative moment* estimator (Caeiro and Gomes, 2010),

$$\hat{\gamma}_{k,n}^{NM(\theta)} := \frac{1}{2}\left\{1 - \left(M_{k,n}^{(2)}/(M_{k,n}^{(1)})^2 - 1\right)^{-1}\right\} + \theta\, M_{k,n}^{(1)}, \quad \theta \in \mathbb{R}.$$

with

$$M_{k,n}^{(j)} := \frac{1}{k}\sum_{i=1}^{k} \{\ln X_{n-i+1:n} - \ln X_{n-k:n}\}^j, \qquad j \geq 1, \qquad X_{n-k:n} > 0.$$

Apart from the usual integer parameter k, related with the number of top order statistics involved in the estimation, the estimator depend on an extra real parameter θ, which makes it flexible and possibly second-order unbiased for a large variety of models in $\mathcal{D}_\mathcal{M}(EV_\gamma)_{\gamma<0}$.
The package `aste` (adaptive short tail estimation) provides the *Algorithm* in Gomes *et al.* (2013) for the adaptive choice of the *tuning* parameters θ and k in the semi-parametric estimation of the EVI through such an estimator. The package also covers the estimation of high quantiles and the right endpoint of the model F underlying the data.

Example

As an example, we apply the *Algorithm* to the analysis of a set of environmental data, the daily average wind speeds in knots (one nautical mile per hour), collected in Dublin airport, in the period 1961-1978. Due to the seasonality of wind data, we restrict ourselves to the Autumn season data of size $n = 1602$. Spring and Summer data were already analyzed in Gomes *et al.* (2013).

Figure 1 (left) illustrates, for $\theta = 0$, 1, 1.5 and 2, the behaviour of $\hat{\gamma}_{k,n}^{NM(\theta)}$ as function of k. Notice that the parameter θ has a big influence on the sample path of $\hat{\gamma}_{k,n}^{NM(\theta)}$, as function of k. The application of the adaptive choice of θ, proposed in Gomes et al. (2013) led us to $\hat{\theta} = 1.604$. Figure 1 (right) shows the estimates $\hat{\gamma}_{k,n}^{NM(\theta)}$ and the 95% confidence limit, as function of k, with the adaptative $\hat{\theta} = 1.604$.

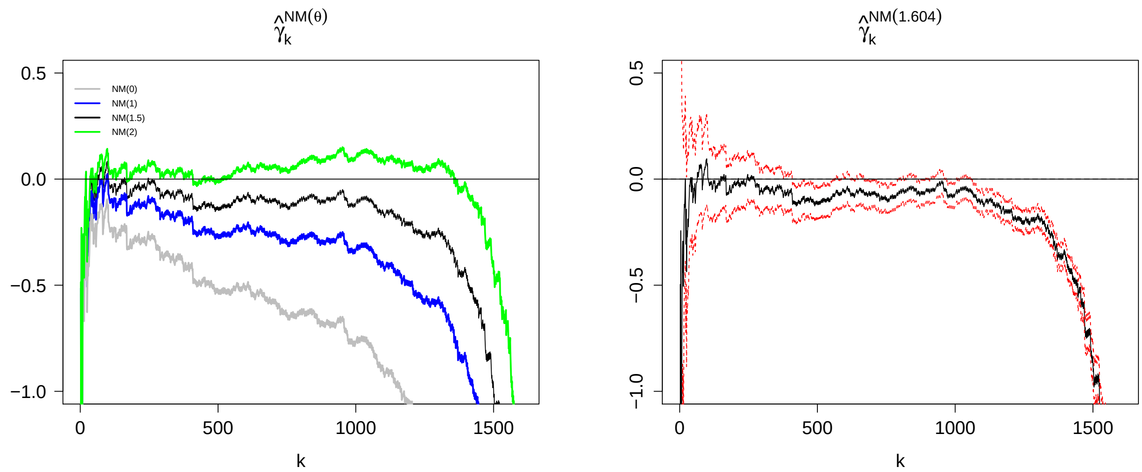

Figure 1: Left: Estimates of the EVI as function of k provided by the estimator under consideration, with several values of θ, for the Autumn wind speed dataset. Right: Estimates of the EVI (black) and 95% confidence limit (red), provided by the estimator under consideration with $\hat{\theta} = 1.604$.

The use of the Algorithm for the choice of k led us to $\hat{k} = 1048$ and to the adaptive EVI-estimate $\hat{\gamma} = -0.046$, and a 95% confidence interval $(-0.1045; 0.0129)$.

Acknowledgements. Research partially supported by National Funds through FCT— Fundação para a Ciência e a Tecnologia, projects PEst-OE/MAT/UI0297/2011 (CMA/UNL) and EXTREMA, PTDC/MAT/101736/2008.

References

[1] Caeiro, F. and Gomes, M.I. (2010). An asymptotically unbiased moment estimator of a negative extreme value index. Discussiones Mathematica: Probability and Statistics **30**(1), 5–19.

[2] Gomes, M.I., Henriques-Rodrigues, L. and Caeiro, F. (2013). Refined Estimation of a Light Tail: an Application to Environmental Data. Accepted in Torelli, N., Pesarin, F., Bar-Hen, A. (Eds.), Advances in Theoretical and Applied Statistics, Springer.