A Robust estimator for the Extreme Value Index of Pareto-Type Distributions

D. Vanpaemel[1], M. Hubert[1] and G. Dierckx[1 2]

[1] Department of Mathematics, Katholieke Universiteit Leuven, Celestijnenlaan 200 b, B-3001 Leuven
[2] Department of Mathematics, Hogeschool-Universiteit Brussel, Stormstraat 2, B-1000 Brussel

Keywords: Robustness, Extreme Value Index, Pareto-Type Distribution.

1 Extreme value statistics

In extreme value statistics, the Extreme Value Index (EVI), denoted by γ, is used to characterize the tail behavior of a distribution. This real-valued parameter helps to indicate the size and frequency of certain extreme events under a given probability distribution: the larger γ is, the heavier the tail of the distribution.

Consider $X_1, \ldots, X_n$ independent and identically distributed (i.i.d.) random variables with common distribution function F and quantile function Q. Denote the corresponding order statistics by $X_{1,n} \leq \cdots \leq X_{n,n}$ and suppose there exist sequences of constants $(a_n > 0)$ and $(b_n \in \mathbb{R})$ such that the properly centered and normed sample maxima $\frac{X_{n,n} - b_n}{a_n}$ converge in distribution to a non-degenerate limit H. It can be shown that the limit distribution H is always of the extreme value type:

$$H_\gamma = \begin{cases} \exp(-(1+\gamma x)^{-1/\gamma} & \text{for } 1 + \gamma x > 0, \gamma \neq 0 \\ \exp(-\exp(-x)) & \text{for } x \in \mathbb{R}, \gamma = 0, \end{cases} \tag{1}$$

and F is said to belong to the maximum domain of attraction of H_γ, denoted as $F \in \mathcal{D}(H_\gamma)$. Most common distributions satisfy this weak condition quite naturally. Distribution for which the Extreme Value Index $\gamma > 0$ are called Pareto-type (or heavy tailed) distributions, as the tail typically decays polynomially, i.e. $\bar{F}(x) = x^{-1/\gamma} l_F(x)$ with l_F a slowly varying function. This talk will concentrate on Pareto-type distributions. Examples of this class are Pareto, Fréchet and Burr distributions.

The most prominent classical extreme value estimators are constructed based on efficient maximum likelihood estimators, resulting in the Hill estimator

$$H_{k,n} = \frac{1}{k} \sum_{j=1}^{k} Z_j,$$

with k a well-chosen tuning parameter, and its second order refinement (Beirlant et al., 2004), which reduces the bias in the estimation. The fact that $Z_j := j(\log X_{n-j+1,n} - \log X_{n-j,n})$ are independent and exponentially distributed:

$$Z_j \overset{d}{\sim} \left(\gamma + \left(\frac{j}{k+1} \right)^\beta b_{n,k} \right) E_j, j = 1, \ldots, k \tag{2}$$

with $(E_1, E_2, \ldots)$ being standard exponentially distributed leads to a ML-estimator for γ that will be denoted as $\hat{\gamma}_{ML}$. Note that one also can derive ML-estimators for β and $b_{n,k}$. Maximum likelihood estimators, however, are not very robust, which makes them sensitive to few particular observations. Even in extreme value statistics, where the most extreme data usually receive the most attention, this can constitute a serious problem.

2 Robust extreme value estimator

In order to obtain robust estimates for the extreme value index, we approximate the distributions denoted in (2) by

$$Z_j \overset{d}{\sim} \exp\left(\log(\gamma) + \left(\frac{j}{k+1}\right)^{\beta} \frac{b_{n,k}}{\gamma}\right) E_j. \tag{3}$$

If we then define

$$\mathbf{x}_j = \left(1 \ \left(\tfrac{j}{k+1}\right)^{\beta}\right)^T \text{ and } \beta = \left(\log\gamma \ \tfrac{b_{n,k}}{\gamma}\right)^T$$

we can derive a Generalized Linear Model (GLM):

$$\log(E(Z_j)) = \mathbf{x}_j{}^T \beta. \tag{4}$$

Robust estimators for β (and hence for γ) can now be obtained by the robust GLM estimators, as introduced by Cantoni and Ronchetti (2001, 2006). The robust weights resulting from this procedure can now be appropriately transformed into weights for the Z_j. These weights are plugged into Hill-estimator and the $\hat{\gamma}_{ML}$ estimator to obtain robust estimators for γ.

The robustness towards contamination of the proposed estimator will be illustrated on both simulated and real data.

References

J. Beirlant et al. (2004). *Statistics of Extremes, Theory and Applications.* Wiley & Sons, England.

E. Cantoni, and E. Ronchetti (2001). Robust Inference for Generalized Linear Models. *Journal of the American Statistical Association*, 96, 1022-1030.

E. Cantoni, and E. Ronchetti (2006). A robust approach for skewed and heavy-tailed outcomes in the analysis of health care expenditure. *Journal of Health Economics*, 25, 198–213.