

Hampel Estimation for Extreme Value Distributions¹

Stima Secondo Hampel nelle Distribuzioni dei Valori Estremi

Alessandra R. Brazzale

LADSEB

Consiglio Nazionale delle Ricerche

Corso Stati Uniti 4, 35127 Padova

alessandra.brazzale@ladseb.pd.cnr.it

Laura Ventura

Dipartimento di Scienze Statistiche

Università degli Studi di Padova

Via C. Battisti 241, 35121 Padova

ventura@stat.unipd.it

Riassunto: In questo lavoro si considera uno stimatore robusto per i parametri delle distribuzioni dei valori estremi. Ciò permette di costruire procedure inferenziali parametriche con un comportamento asintotico analogo alle procedure basate sulla verosimiglianza, ma con proprietà di robustezza rispetto a valori influenti o alla specificazione scorretta. Una applicazione a dati reali è presentata.

Keywords: Generalized extreme value distribution, Hampel estimator, Maximum likelihood, Regression, Robustness.

1. Introduction

There is a long history of the use of “extreme value distributions” in the statistical analysis of extreme value data such as river heights, sea levels, rainfall, air pollutants or, in the reliability context, strengths and failure times. A classical reference is Gumbel (1958), while a recent comprehensive survey is provided by Coles (2001). The problem may be described, in its broadest terms, as how to make inference about the extreme values in a population. A traditional approach consists of fitting a linear regression model $y = X\beta + \sigma\varepsilon$ to the response y , where X is a fixed $n \times p$ matrix, $\beta \in \mathbb{R}^d$ a regression coefficient, $\sigma > 0$ a scale parameter, and ε an n -dimensional vector of independent and identically distributed errors. The most common choice for the error distribution consists in the Generalized Extreme Value (GEV) distribution, with cumulative distribution function

$$F(\varepsilon; \kappa) = \exp \left\{ -[1 + \kappa\varepsilon]^{-1/\kappa} \right\}, \quad \text{for } 1 + \kappa\varepsilon > 0. \quad (1)$$

Gumbel’s (1958) Types I, II and III distributions are special cases corresponding to $\kappa = 0$, $\kappa > 0$ and $\kappa < 0$, respectively. These arise in extreme value theory as the three possible stable limiting distributions for the maximum order statistic. In practice, the parameters β, σ and κ must be estimated from the data. The most common way is by maximum likelihood, using either a Newton or quasi-Newton optimizer.

The aims of this paper are to discuss the consequences of model misspecifications or the presence of influential observations on the classical inferential procedures used to estimate (β, σ, κ) , and to derive robust estimators. Davison and Smith (1990, Section 5) and Brazzale and Ventura (2001) show that deviations from the assumed model can have drastic effects on likelihood based procedures such as the classical maximum likelihood estimator (MLE). Anomalous observations may be dealt with by a preliminary screening of the data, but this is not possible with influential observations which can only be detected

¹The present paper is partially supported by grants from MIUR, Italy.

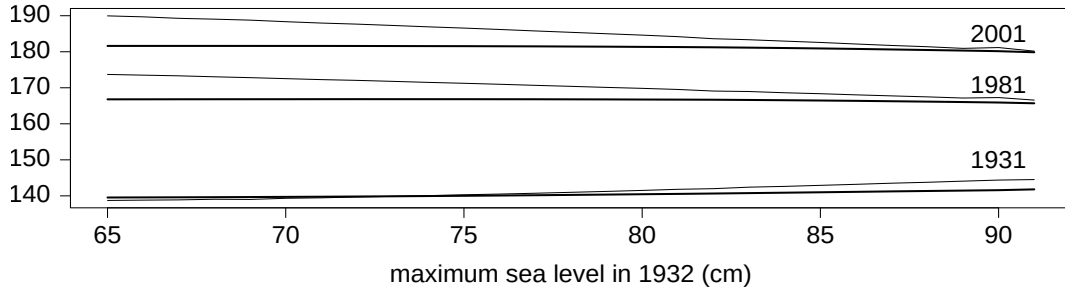


Figure 1: 20-year return level for the years 1931, 1981 and 2001 obtained from the MLE (solid line) and OBRE (bold line) GEV model fits to the Venice data.

once the model has been fitted. In the light of the amount of data available nowadays and the automated procedures used to analyze them, robust techniques may be preferable as they automatically take possible deviations into account. Moreover, the diagnostic information provided by these techniques can be used by the analyst to identify deviations from the model or from the data.

In this paper we follow the approach based on the influence function (IF) which describes the effect of a small contamination at a fixed observation on an estimator (Hampel et al. 1986). We furthermore focus on the GEV distribution, although the same procedures could be applied to other extreme value models such as the Generalized Pareto distribution. Moreover, the methods can be extended to situations that consider the r -largest order statistic instead of the maximum.

2. Hampel's estimator

The MLEs of the parameters (β, σ, κ) of the GEV regression model discussed in the previous section are obtained by maximizing the log-likelihood function

$$\ell(\beta, \sigma, \kappa) = -n \log \sigma - \left(1 + \frac{1}{\kappa}\right) \sum_{i=1}^n \log \left(1 + \frac{\kappa}{\sigma} (y_i - x_i^\top \beta)\right) - \sum_{i=1}^n \left(1 + \frac{\kappa}{\sigma} (y_i - x_i^\top \beta)\right)^{-1/\kappa},$$

where $x_i^\top$ is the i th row of X . This class of estimators is, however, not *B-robust*. In fact, it can easily be shown that the score functions for the parameters (β, σ, κ) and, consequently, the IFs of the MLEs, are unbounded. This means that the MLEs are not robust against outlying observations and long-tailed distributions and that a small contamination on an observation can have a drastic effect.

If we suspect possible deviations from the assumed model, it may be preferable to resort to robust techniques. The offer of robust alternatives to the MLE is wide (Hampel et al. 1986, Chapters 2 and 4). Here we focus on the optimal B-robust estimator (OBRE) defined as the solution to the problem of finding the most efficient estimator among all estimators that are B-robust (Hampel, Ronchetti, Rousseeuw and Stahel 1986, Chapter 4). The OBRE is optimal in the sense that it minimizes the trace of the asymptotic covariance matrix under the constraint of a bounded IF. It belongs to the wider class of M-estimators which are generalizations of the MLE. Under broad conditions, it can be shown that the OBRE is consistent and asymptotically normal. Indeed, large-sample tests and confidence regions for the parameters can be constructed in the standard way using an estimate of the asymptotic covariance matrix.

There are several versions of the OBRE, depending on the way one decides to bound the IF (Hampel *et al.* 1986, Section 4.3). Here we use the standardized OBRE. Given the bound $k = c\sqrt{p}$, where p is the size of the vector parameter $\theta = (\beta, \sigma, \kappa)$, the OBRE is defined as the solution to

$$\sum_{i=1}^n \psi(y_i; \theta) = \sum_{i=1}^n \{\ell_{\theta}(\theta; y_i) - a(\theta)\} w_i(y_i; \theta) = 0, \quad (2)$$

where ℓ_{θ} represents the gradient of the log-likelihood with respect to θ and

$$w_i(y; \theta) = \min \left\{ 1, \frac{k}{\|\ell_{\theta}(\theta; y) - a(\theta)\|_{A(\theta)}} \right\}, \quad (3)$$

$\|x\|_{A(\theta)} = [x^T A(\theta)^{-1} x]^{1/2}$. Expressions (2) and (3) are easily interpreted. For efficiency reasons, the OBRE has to be as close as possible to the MLE for values of y that lie in the bulk of data, i.e. which are non-influential. For those values the ψ -function equals the score function ℓ_{θ} . On the other hand, since the IF is proportional to the ψ -function, to obtain a bounded IF one has to truncate ℓ_{θ} where the bound k is exceeded. This is achieved by means of the weights (3). In applications, these weights are worth looking at because they identify the extent to which the observations have been considered more or less far from the bulk of the data. In particular, this allows one to approximately determine the degree of contamination. The $p \times p$ matrix $A(\theta)$ and the $p \times 1$ vector $a(\theta)$ must satisfy the conditions

$$A(\theta) = \frac{1}{n} E[\psi(y; \theta) \psi(y; \theta)^T] \quad \text{and} \quad E[\psi(y; \theta)] = 0. \quad (4)$$

They can be viewed as Lagrange multipliers for the constraints resulting from a bounded IF and Fisher consistency. In practice, $\hat{\theta}$, $A(\hat{\theta})$, $a(\hat{\theta})$ and $w_i(y; \hat{\theta})$ must be found simultaneously by solving (2), (3) and (4). An iterative procedure based on the Newton-Raphson algorithm can be used; see Carroll and Ruppert (1988, Chapter 6) and Bellio and Ventura (2001).

The constant k is the bound on the IF and can be interpreted as the regulator between robustness and efficiency: for small k one gains robustness but loses efficiency, and vice versa for large k . The most robust estimator can be obtained by choosing the lower bound $k = \sqrt{p}$. On the other hand, $k \rightarrow \infty$ gives the MLE, i.e. the most efficient, but non robust, estimator. The choice of a particular k depends in general on the model to be fitted.

3. Application: Venice data

The OBRE was applied to a well-known data set which concerns the maximum sea levels measured in Venice from 1931 to 1981 (Smith 1986). The fitted model is a regression model with GEV distributed errors. The linear predictor includes a linear trend in the years plus a 19-year seasonal component due to periodic tidal fluctuations (Smith 1986, model 2.6). The MLEs and OBREs with $c = 2$ are given in Table 1 together with their standard errors. The weights (3) returned from the OBRE model fit point toward observation 2 as an influential observation. The last line of Table 1 gives the MLEs one obtains when this observation is omitted from the fit. The MLEs of the parameters β_2 and β_3 , associated with the seasonal fluctuation, and of the form parameter κ change remarkably.

Parameter	β_0	β_1	β_2	β_3	σ	κ
MLE	95.9 (4.2)	32.4 (7.1)	-2.42 (2.75)	5.46 (3.10)	14.0 (1.6)	-0.0239 (0.0920)
OBRE	97.1 (3.8)	31.2 (5.8)	-1.15 (2.07)	5.51 (2.51)	13.2 (1.2)	-0.0246 (0.0612)
MLE[-2]	97.0 (4.9)	31.9 (8.1)	-1.62 (2.85)	6.58 (3.23)	13.8 (1.7)	0.000 (0.0952)

Table 1: MLE, OBRE and MLE[-2] parameter estimates for the GEV model fitted to the maximum sea level data for the years 1931–1981.

A similar behaviour results if we focus on quantities of interest that are functions of the original parameters, as for instance the 20-year return level. To assess this we run an empirical sensitivity analysis during which the sea level of 78cm observed in 1932 was perturbed and allowed to vary in the range 65cm to 90cm. As shown in Figure 1, the 20-year return levels for the years 1931, 1981 and 2001 are much more stable if computed from the OBRE than the MLE model fit.

References

- Bellio R. and Ventura L. (2001) Robust estimation of stochastic frontier models, in: *Modelli Complessi e Metodi Computazionali Intensivi per la Stima e la Previsione*, Cleup, Padova, 175–180.
- Brazzale A.R. and Ventura L. (2001) How robust is conditional inference?, in: *New Trends in Statistical Modelling*, Odense, July 2–6, 2001, 417–420, proceedings of the 16th International Workshop on Statistical Modelling.
- Carroll R.J. and Ruppert D. (1988) *Transformation and Weighting in Regression*, Chapman & Hall, New York.
- Coles S. (2001) *An Introduction to Statistical Modeling of Extreme Values*, Springer-Verlag, UK.
- Davison A.C. and Smith R.L. (1990) Models for exceedances over high thresholds, *Journal of the Royal Statistical Society Series B*, 52, 393–442.
- Gumbel E.J. (1958) *Statistics of Extremes*, Columbia University Press, New York.
- Hampel F.R., Ronchetti E.M., Rousseeuw P.J. and Stahel W.A. (1986) *Robust Statistics. The Approach Based on Influence Functions*, Wiley, New York.
- Smith R.L. (1986) Extreme value theory based on the r largest annual events, *Journal of Hydrology*, 86, 27–43.