A Default Bayesian analysis of the Nidd River data

S. Cabras[1] and M. E. Castellanos[2]
[1]Department of Statistics, University of Florence (Italy)
[2]Department of Statistics, Rey Juan Carlos University (Spain)

Introduction

The problem of modelling extreme values is of great interest in many environmental studies. Among these we have certain natural phenomena e.g. waves, winds, temperatures or earthquakes. In this work we study the annual maximum flood of the River Nidd at Hunsingore Weir (England). Data consist on 154 exceedances over the level 65m^3/sec. from 1934 to 1969 (35 years) (source: Natural Environment Research, 1975). This data set is one of the most studied in statistical extreme value approach, and among various work on this data set we compare our results with those in Davison and Smith (1990). In this work we propose a default Bayesian procedure in order to estimate future high levels of the River Nidd.

Methods

The most known model to study extreme-value data is based on the family of generalized extreme value distributions (see Galambos 1981). This family of distributions is appropriate when the data only consist of a set of maxima. For the data set we study there is a loss of information if only maxima were considered. We concentrate on modelling the exceedances over a fixed threshold u. Pickands (1975) showed that the Generalized Pareto Distribution (GPD) is the distribution for the exceedances over a u, when u is sufficiently large. The GPD is governed by two parameters, a scale and a shape parameter. This distribution generalizes different families of better known distributions according to the values assumed by the shape parameter. For example when the shape parameters is 0 the GPD is the Exponential distribution, when the shape parameter is -1 the GPD is the Uniform distribution and when it is positive the GPD is the Pareto distribution.

It is interesting to study the sign of the shape parameter, because negative values indicate that the exceedances may have an upper bound.

The estimates of both scale and shape parameters are generally difficult to obtain due to the fact that the support of the likelihood depends on the observed data. Smith (1984) showed that when the shape parameter is larger than -0.5 the likelihood is regular, in the sense that exists the Fisher information matrix.

For the set of values where the Likelihood is regular we propose a Bayesian approach to estimate the posterior distribution of the scale and shape parameters of GPD. In particular we calculate the Jeffreys's prior (improper for the scale parameter) and show that the posterior is always proper. In this way the predictive distribution of the exceedances or the posterior distribution of the shape parameter can be used to estimate future high levels of the river.

Bayesian methods for estimating the GPD parameters have not been excessively explored probably due to the irregularity of the likelihood. The most relevant work, in a non informative context, is contained in de Zea Bermudez and Amaral Turkman (2003), while Arnold and Press (1989) only studied Bayesian inference for the Pareto distribution. de Zea Bermudez and Amaral Turkman (2003) proposed to use the posterior means using two independent priors for negative and positive values of the shape parameter. This lead to a procedure which is always biased when the shape parameter is zero and it does not allow to test the sign of the shape parameter.

Two problems are not touched in this work: the first is how to assess the goodness-of-fit of GPD and the second is the optimal choice of initial threshold. Therefore we assume the GPD is a reasonable model for the observed exceedances and that the uncertainty is left only on GPD's parameters.

We compare the posterior means and medians with other Bayesian approach (de Zea Bermudez and Amaral Turkman, 2003) as well as non Bayesian procedures to estimate the unknown parameters of GPD. In particular we consider maximum likelihood estimators (MLE) (Grimshaw, 1993), probability-weighted moments estimators (Hosking and Wallis, 1987) and estimators based on the elemental percentile method of Castillo and Hadi (1997).

Results

We provide a Markov Chain Monte Carlo algorithm to estimate the posterior distribution. The algorithm uses a moving proposal distribution for a Metropolis-Hastings step nested in two Gibbs steps.

We use a parametric bootstrap in order to estimate the BIAS and Mean Squared Error (MSE) of the posterior means and medians and compare them with other estimators. We obtain basically the same BIAS and MSE of other commonly used estimators.

We apply the procedure at different thresholds (higher than 65m3/sec.) of the water level of the Nidd River. We predict future observations using the K-year return level, defined as that level which is exceeded on average once in K years (we consider K=25, 50 and 100 years). Using a Bayesian approach it is straightforward to obtain the posterior distribution of the K-year return level. We found that this posterior distribution have larger variability than those obtained using a frequentist approach

(Davison and Smith, 1990). This is reasonable because we are taking into account parameter uncertainty which is, of course, not considered in a MLE approach such as those in Davison and Smith (1990). Finally we calculate the posterior odd on the existence of an upper bound for the Nidd River levels, provided that the prior odd is 1. In particular we find that the higher is the threshold, the higher is the evidence for the existence of an upper bound.

Conclusions

We think that the problem of the choice of the prior distribution on shape and scale parameter of GPD is still open, because expert elicitation is very difficult to obtain due to the lack of any physical interpretation of GPD parameters. Furthermore non-informative priors makes prediction of future observations quite problematic for two reasons: a) the prior variability of GPD parameters is larger than those induced by an informative prior, moreover b) the sample size is usually very small because extreme values are usually regarded as rare events.

Referencias

Arnold, B. y Press, S. J. (1989). Bayesian estimation and prediction for pareto data. *Journal of the American Statistical Association* 84:1079–1084.

Castillo, E. y Hadi, A. S. (1997). Fitting the generalized pareto distribution to data. *Journal of the American Statistical Association* 92:1609–1620.

Council, N. E. R. (1975). *Flood Studies Report*. Natural Enviroment Research Council, London.

Davison, A. C. y Smith, R. L. (1990). Models for exceedances over high thresholds. *Journal of the Royal Statistical Society, Series B* 52:393–442.

Galambos, J. (1981). Extreme value theory in applied probability. *Journal of the American Statistical Association* 6:13–26.

Grimshaw, S. D. (1993). Computing maximum likelihood estimates for the generalized pareto distribution. *Technometrics* 35:185–191.

Hosking, J. R. M. y Wallis, J. R. (1987). Parameter and quantile estimation for the generalized pareto distribution. *Technometrics* 29:339–349.

Pickands, J. (1975). Statistical inference using extreme order statistics. *The Annals of Statistics* 3:199–131.

Smith, R. L. (1984). Threshold methods for sample extremes. J. T. de Oliveira y D. Reidel (eds.), *Statistical Extremes and Applications*, 621–638. Dordrecht, Reidel.

de Zea Bermudez, P. y Amaral Turkman, M. A. (2003). Bayesian approach to parameter estimation of the generalized pareto distribution. *Test* 12:259–277.