UAB
Universitat Autònoma de Barcelona

Dipòsit digital de documents de la UAB

This is the **accepted version** of the journal article:

Castillo, Joan del; Serra, Isabel. «Likelihood inference for generalized Pareto distribution». *Computational statistics & data analysis*, Vol. 83 (March 2015), p. 116-128 DOI 10.1016/j.csda.2014.10.014

This version is available at https://ddd.uab.cat/record/338632

under the terms of the BY-NC-ND license.

Available online at www.sciencedirect.com
ScienceDirect
00 (2014) 1–15
Journal Logo

Likelihood inference for Generalized Pareto Distribution

Joan del Castillo[a], Isabel Serra[b]

[a] *Universitat Autònoma de Barcelona. castillo@mat.uab.cat*
[b] *Centre de Recerca Matemàtica. Corresponding author. Campus de Bellaterra, E-08193 Barcelona, Spain. iserra@mat.uab.cat. Phone: +34935811081. Fax: +34935812202*

Abstract

A new methodological approach that enables the use of the maximum likelihood method in the Generalized Pareto Distribution is presented. Thus several models for the same data can be compared under Akaike and Bayesian information criteria. The view is based on a detailed theoretical study of the Generalized Pareto Distribution submodels with compact support.

Keywords: Maximum likelihood estimation, Model selection, Akaike and Bayesian information criteria, Extreme value theory, Bilbao data

1. Introduction

The results produced by Balkema and DeHaan (1974) and Pickands (1975) have widened the use of the Generalized Pareto Distribution (GPD) in extreme values theory as a model for tails, see McNeil *et al.* (2005), Beirlant *et al.* (2004), Finkenstadt and Rootzén (2003), Coles (2001) and Embrechts *et al.* (1997). However, several authors have highlighted the estimation problems that arise from this distribution, see Hosking and Wallis (1987), Castillo and Hadi (1997), Zhang and Stephens (2009). Currently, together with the maximum likelihood estimation (MLE), the methods of Zhang (2010) and Song and Song (2012) are the best alternatives. The main issue with MLE is that for some datasets the likelihood function appears to have no local maximum.

The motivation for this study lies in the fact that, despite its weaknesses, we need the MLE in many inference procedures. The adjustments of various models to the same data are commonly compared with the Akaike (AIC) and Bayesian (BIC) information criteria and with the likelihood ratio test (LRT) in nested models, all of which are based on the MLE, see Clauset *et al.* (2009). It is also common to use the MLE in goodness of fit tests as in Choulakian and Stephens (2001). A summary of the theoretical results for MLE in the GPD can be found in Davison and Smith (1990). After that, many authors report that the MLE exists for large samples provided $\kappa < 1$ and is asymptotically efficient provided $\kappa < 0.5$. For $\kappa > 1$ with probability tending to 1 there are no local maximum. This article goes a step further, see Theorem 1 and Proposition 4.

In Section 2, we propose a fast, simple and stable algorithm, which always provides an estimate of the MLE when possible. This algorithm is compared with those of Zhang (ZSE) and Song and Song (SSE), in terms of efficiency and bias, where neither existing algorithm has been shown to dominate. According to the behavior of the probability density functions of the GPD, we can distinguish three submodels that are separated by the exponential distribution ($\kappa = 0$) and the uniform distribution ($\kappa = 1$). The study highlights the problem of misspecification of the submodel. Running a simulation from one of the three GPD submodels, the estimated parameters can easily belong to another, as shown in Table 1. The best correct classification results are always obtained with MLE or ZSE.

Section 3 gives some mathematical results that provide precise arguments to explain the anomalous behavior of the likelihood function when sampling from the GPD distribution for positive values of κ. Theorem 2 proves that there is

a global MLE for the submodel $0 \leq \kappa \leq 1$. Proposition 2 and Proposition 3 show that the likelihood function extends continuously up to the boundary and reaches the likelihood of the exponential and uniform distributions. Hence, the AIC and BIC criteria can be used in this case. However, it is proved that for $\kappa > 1$ the likelihood function is unbounded and reaches infinity, in a half-line of the parameter space. Hence, mathematically, neither AIC nor BIC can be used properly for $\kappa > 1$.

In Section 4, the numerical and analytical results of the previous sections lead us to suggest a new methodological approach that emphasizes the role of transition points $\kappa = 0$ and $\kappa = 1$. Two of the most commented examples of extreme values theory are discussed in Section 5. Finally, Section 6 shows a summary of the main conclusions.

2. Computing the maximum likelihood estimation

The Generalized Pareto Distribution (GPD), named by Pickands (1975), is a two-parameter family of distributions, with the cumulative distribution function given by

$$F(x;\, \kappa,\, \psi) = 1 - (1 - \kappa x/\psi)^{1/\kappa}, \tag{1}$$

where $\psi > 0$ and κ are the scale and shape parameters, respectively. For $\kappa > 0$ the range of x is $0 < x < \psi/\kappa$ and for $\kappa < 0$ the range is $x > 0$. Let us define σ as ψ/κ. The GPD becomes the uniform distribution for $\kappa = 1$ and the exponential distribution for $\kappa = 0$ (taken as the limit). The GPD has mean $\psi/(1+\kappa)$ and variance $\psi^2/[(1+\kappa)^2(1+2\kappa)]$ provided $\kappa > -0.5$. Note that the coefficient of variation is independent of the scale parameter, since it is given by

$$cv = 1/\sqrt{(1+2\kappa)}. \tag{2}$$

The probability density function of the GPD is given by

$$f(x;\, \kappa,\, \psi) = \psi^{-1}\,(1 - \kappa x/\psi)^{1/\kappa - 1}\,. \tag{3}$$

Figure 1 shows the importance of the uniform distribution ($\kappa = 1$) in the GPD, which is between the monotone decreasing probability density functions ($\kappa < 1$) and the monotone increasing probability density functions tending to infinity at the upper support ($\kappa > 1$).

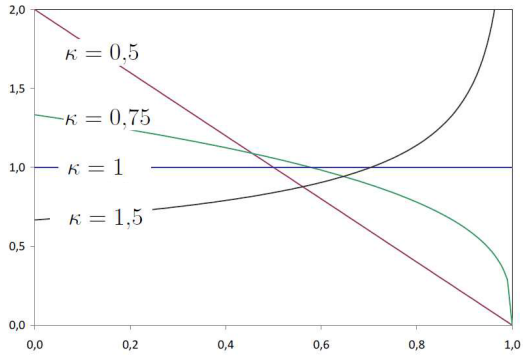


Figure 1. Example to clarify the heterogeneity of the probability density functions of the GPD family for $\kappa > 0$ and $\sigma = 1$.

Let $x_1, \ldots, x_n$ be a random sample from the GPD and let $x_{(1)} \leq \cdots \leq x_{(n)}$ be the order statistics of the sample. The log-likelihood is given by

$$l(\kappa, \psi) = n\left(-\log(\psi) + (1/\kappa - 1)\frac{1}{n}\sum_{i=1}^{n}\log(1 - \kappa x_i/\psi)\right), \tag{4}$$

where $\psi > 0$ for $\kappa \leq 0$ and $\psi > \kappa x_{(n)}$ for $\kappa > 0$.

The log-likelihood function parameterized by κ and σ is denoted by l^*, that is $l^*(\kappa, \sigma) = l(\kappa, \kappa\sigma)$, where $\sigma < 0$ for $\kappa \leq 0$ and $\sigma > x_{(n)}$ for $\kappa > 0$. Note that σ is the right endpoint for the distributions with bounded support. This

parameterizing is more efficient from a computational point of view. Note that the likelihood equation $\partial l^*(\kappa,\sigma)/\partial\sigma = 0$ is equivalent to $\kappa = \kappa(\sigma)$ where

$$\kappa(\sigma) = -\frac{1}{n}\sum \log(1 - x_i/\sigma). \tag{5}$$

Therefore, the corresponding profile-likelihood function is given by

$$l_p^*(\sigma) = n\left(-\log(\kappa(\sigma)\sigma) + \kappa(\sigma) - 1\right). \tag{6}$$

It is not always possible to compute the MLE using the likelihood equations. In fact, the MLE as a solution for likelihood equations exists in large samples provided that $\kappa < 1$ and if $\kappa < 0.5$ then the MLE is asymptotically normal and efficient, see Smith (1985), Davison and Smith (1990).

We use the following R code, see R Development Core Team (2010), as an approximation algorithm to obtain an optimal value of the MLE from a numerical point of view.

```
#to estimate the maximum likelihood (MLE) of a sample x by GPD(k,psi)
eGPD<-function(x){
       fk<-function(sigma) -mean(log(1-x/sigma))
       fp<-function(sigma) length(x)*(-log(fk(sigma)*sigma)+fk(sigma)-1)
       int<-c(-100*max(x),100*max(x))
       sigma<-optimize(fp,interval=int,maximum=T)$maximum
       list(k=fk(sigma),psi=fk(sigma)*sigma)}
```

This code searches for the value of σ to maximize the profile-likelihood equation (6). Note that the functions `fk` and `fp` in the code correspond to equations (5) and (6), respectively. For now, the possibility of the likelihood function having no local maximum is not taken into account in function `eGPD`. The code is used in this section to compare with other estimation methods. Our estimation method is fast to compute, very stable, and does not depend on initial values. Moreover, the simplicity of the code provides clear solutions.

The main point of the code is the R function:`optimize` which searches the interval from 'lower' to 'upper' for a minimum or maximum. This routine is one of the several optimization routines that the `stats` package offers for general purposes. The method used is a combination of golden section search and successive parabolic interpolation, and was designed to be used with continuous functions. The function is never evaluated at two points that are too close together. For further information see Brent (1973).

2.1. A summary of GPD estimation methods

Traditional methods for estimating GPD parameters were reviewed by Hosking and Wallis (1987). The most relevant alternatives to MLE are the method of moments, the probability-weighted moments and the likelihood moment estimations considered by Zhang (2007). More exhaustive analysis is given by Mackay *et al.* (2011), Juarez and Schucany (2004) and Zhang and Stephens (2009). Currently, the best alternatives to MLE are the method proposed by Zhang and Stephens (2009) improved by Zhang (2010), here denoted ZSE and, the Song and Song (2012) method, denoted SSE.

The interest in searching for an alternative estimation method to MLE for GPD is due to computation problems and correction of bias, while improving efficiency. The computation problems are solved with a suitable algorithm, for example, the previous routine in R: `eGPD`. On the other hand, the methods of estimation that correct the bias of MLE for GPD does so locally as stated by Zhang and Stephens (2009).

In fact, ZSE also works locally but in a larger interval. Precisely ZSE shows an improvement on the alternative methods in bias and efficiency for $-6 < \kappa < 0.5$, see Zhang (Fig. 1, 2010). However, the improvement on MLE in bias may not be relevant, since for κ or ψ greater than 1 the bias of ZSE increases quickly, see Figure 2.

ZSE improves efficiency measured as the ratio of the Cramér-Rao lower bound of the variance to the mean squared error (MSE) of the estimate. However, Figure 2 shows that the improvement on MLE is not significant if measured directly with MSE and for $\kappa > 1.5$ or $\psi > 1$ the MLE has significantly less MSE than ZSE. The same holds for the parameter σ, where the plots are in log-scale. Figure 2 corresponds to sample size 500.

For a small sample size, the bias and efficiency of the current ZSE, SSE and MLE methods are illustrated in detail in the shadow graphs in Figure 3. The shadow graphs use gray scale to represent the frequency observed for each

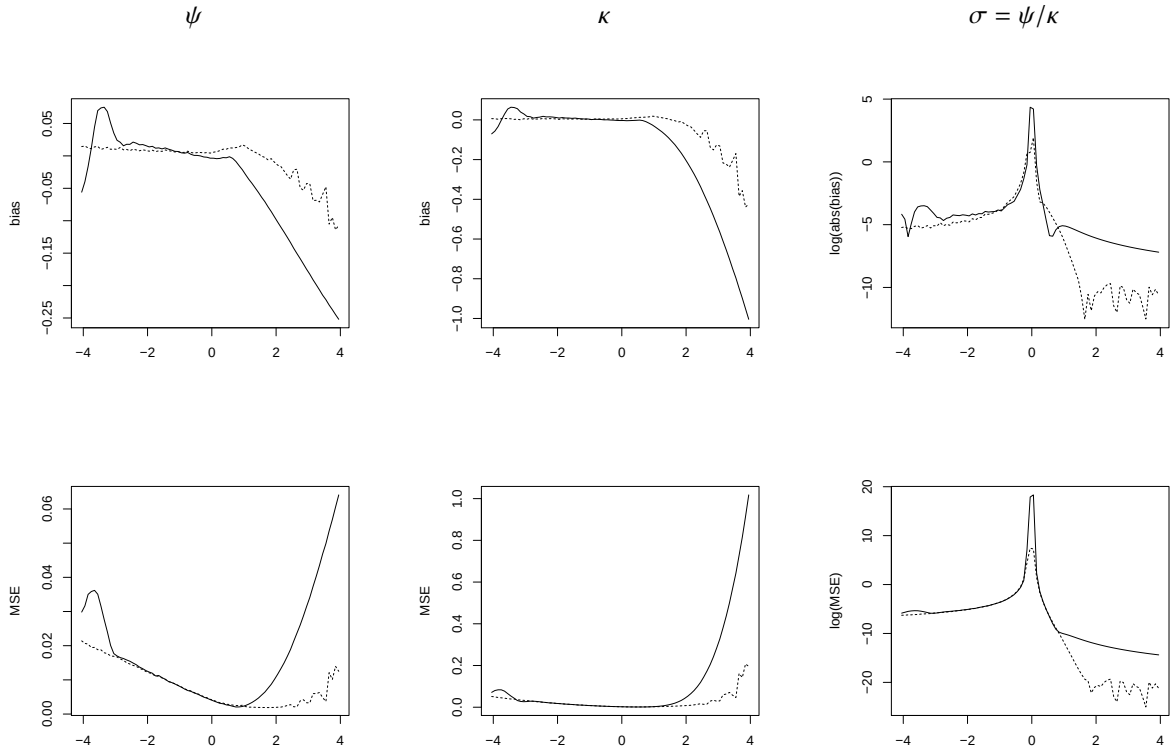


Figure 2. Bias and MSE of parameters estimation ψ, κ and σ. It is obtained from 10,000 simulations of sample size 500 with parameters $\psi = 1$ and κ between -4 and 4, displayed on the horizontal axis. The solid and dashed lines correspond to the ZSE and MLE methods, respectively.

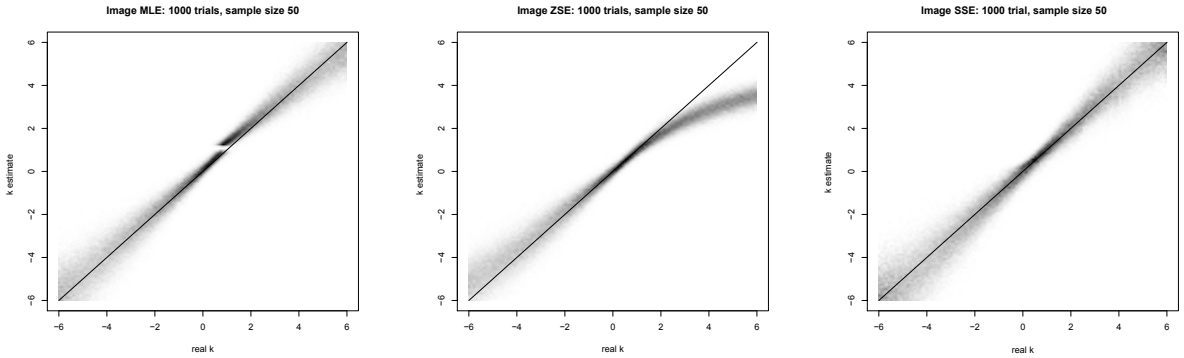


Figure 3. The horizontal axis represents the parameter κ with which we perform the simulation of samples and the vertical axis shows the value of the parameter κ estimated for each of the estimation methods: ML, SS and ZS. The grey intensity corresponds to the observed frequency in 1,000 trials of sample size 50 in each range of κ with: width of 0.1 and uniformly distributed.

interval in the partition of the range of the parameter κ, the x-axis represents the value of the parameter with which the simulation is made and the y-axis shows the estimated value obtained with each of the methods MLE, ZSE and SSE. Figure 3 is obtained from 1,000 trials for each interval width 0.1 of parameter κ of GPD sample size 50 with $\psi = 1$, the values of κ for simulation are uniformly distributed in each partition interval.

As shown in Figure 3, the bias of MLE is not very different from the other proposed methods. Observe that MLE

shows problems with small sample size and κ around 1, so these types of sample must be carefully studied, Section 4 proposes a new methodology to solve this. Note that ZSE does not work outside $-6 < \kappa < 1$. All this may suggest that the best method is SSE, but Section 2.2 shows that it is not.

Finally, we measured the computation times for the current methods: ZSE, SSE and our MLE algorithm; the slowest method considered provides estimates for a sample size of 10^5 in under one minute. The computation times of ZSE and MLE are comparable for small samples, but MLE becomes comparatively faster with the increase in sample size.

2.2. Misspecification in GPD

The GPD shape parameter includes two critical points corresponding to $\kappa = 0$ and $\kappa = 1$. The first is the exponential distribution and is well-studied. For $\kappa = 1$ corresponds to uniform distribution and is not sufficiently studied, see Section 3. Therefore, from a statistical modelling point of view, the GPD has to be considered as three models separated by the exponential and uniform distribution, respectively. Summarizing, for $\kappa < 0$ the Pareto submodel, for $\kappa = 0$ the exponential distribution, for $0 < \kappa < 1$ a family of decreasing probability densities functions with finite support $(0, \sigma)$, for $\kappa = 1$ the uniform distribution and for $\kappa > 1$ a family of increasing densities with finite support, see Figure 1. In Table 1, these three models, with their boundary points, are denoted A, B and C, respectively. A problem with any method of estimation of the GPD is the misspecification among these three models. Since they are highly heterogeneous, this problem cannot be ignored.

Table 1 shows a summary of the misspecification by the MLE, SSE and ZSE methods. Random samples were generated from the GPD for each combination of values of sample sizes n =15, 25, 50, 100 and shape parameters κ =-0.2, -0.1, 0.1, 0.2, 0.8, 0.9, 1.1, 1.2. The scale parameter was set to 1, since the model is invariant under scale changes. For each combination of values of n and κ, 10,000 random samples were generated from GPD, and the number of times the parameter estimates κ to be negative or between 0 and 1 or to be greater than 1 is reported. Observe that the algorithm to compute MLE from Section 2 made it possible to construct the table of classification since it converges for all samples.

Referring to the Pareto model, Table 1 shows the SSE as the best for small sample size, but when increasing the sample size the classification is not improved as the other methods show. For instance, if sample size 100 and $\kappa = -0.2$, the SSE has 16% misspecification, this is twice MLE.

Regarding, the case of $\kappa > 1$, MLE classifies better than the other methods, since the misspecification is about 4% for each sample size and for $\kappa = 1.2$ is around 2%. Moreover, SSE is slightly better than ZSE, but SSE classifies wrongly, as does the Pareto model. In fact, for sample size 15, around 15% are classified by SSE as $\kappa < 0$.

For κ between 0 and 1, SSE and ZSE tend to mis-specify as does Pareto. In fact, more than 30% are classified as $\kappa < 0$ for sample sizes less than 50 and $\kappa = 0.1$ and more than 10% for $\kappa = 0.2$. However, MLE tends to mis-specify by $\kappa > 1$, in fact for $\kappa \geq 0.8$ more than 40% of sample sizes less than 50 are classified wrongly as $\kappa > 1$. SSE and ZSE also classify wrongly as $\kappa > 1$, but less than MLE.

The main conclusion is that since no method is able to ensure the correct classification between these three different models A, B and C, then the estimation of parameters should be made subsequently to the choice of model and even more for small samples, see details in Section 4.

3. On the existence of the maximum likelihood estimator

Much of our knowledge of GPD estimation comes from numerical techniques such as those used in the previous section. However, doubts remain about the existence of local and global maxima and there are examples where estimation is controversial, see Castillo and Hadi (1997) and Zhang and Stephens (2009). In this section, we present some mathematical results that provide precise arguments to explain the anomalous behavior of the likelihood surface when sampling from the GPD distribution.

In summary, we show the central role of the uniform distribution between distributions with light tails. It is proven that for $0 \leq \kappa \leq 1$ the likelihood function is bounded, then the global MLE exists, see Theorem 2. Moreover, from Proposition 2 and Proposition 3, the likelihood function extends continuously up to the boundary and reaches the likelihood of the exponential and uniform distributions. Hence, the AIC and BIC criteria can be used in this submodel. However, it is proven that for $\kappa > 1$ the likelihood function is unbounded and reaches infinity, in a half-line of the parameter space. Hence, mathematically, neither AIC nor BIC can be used properly in this case.

Table 1. Percentages of the classification for each submodel for some sample size and values of κ. The model A corresponds to the submodel of GPD for $\kappa \leq 0$, the model B for $0 \leq \kappa \leq 1$ and the model C for $\kappa \geq 1$. The bold values correspond to well classified categories.

	A			**B**						**C**		
κ	-0.1			0.1			0.9			1.1		
n	**A**	B	C	A	**B**	C	A	**B**	C	A	B	**C**
ZSE												
15	**65.8**	34.1	0.1	44.1	**55.5**	0.4	0.9	**76.5**	22.6	0.4	57.7	**41.9**
25	**71.0**	29.0	0.0	41.7	**58.3**	0.0	0.1	**79.8**	20.1	0.0	52.7	**47.3**
50	**76.5**	23.6	0.0	30.9	**69.1**	0.0	0.0	**83.7**	16.3	0.0	40.5	**59.5**
100	**84.2**	15.8	0.0	19.9	**80.1**	0.0	0.0	**90.7**	9.3	0.0	31.0	**69.0**
SSE												
15	**85.3**	14.5	0.3	73.8	**25.6**	0.7	20.7	**50.0**	29.4	14.7	40.7	**44.6**
25	**73.4**	26.6	0.0	54.2	**45.8**	0.0	3.7	**61.0**	35.3	1.8	40.4	**57.8**
50	**67.5**	32.6	0.0	39.4	**60.6**	0.0	0.1	**69.3**	30.6	0.0	34.4	**65.6**
100	**70.1**	30.0	0.0	29.1	**70.9**	0.0	0.0	**77.1**	22.9	0.0	27.3	**72.7**
MLE												
15	**41.4**	50.2	8.4	18.8	**64.6**	16.6	0.1	**11.2**	88.8	0.0	4.1	**95.9**
25	**49.2**	49.9	0.9	17.4	**80.1**	2.5	0.0	**16.1**	83.9	0.0	4.8	**95.2**
50	**63.0**	37.0	0.0	15.2	**84.8**	0.0	0.0	**30.1**	69.9	0.0	4.8	**95.2**
100	**75.0**	25.0	0.0	9.7	**90.3**	0.0	0.0	**49.8**	50.2	0.0	4.4	**95.6**

Suppose $F(x)$ is an absolutely continuous cumulative distribution function of a non negative random variable X and $f(x) = F'(x)$. Let $F_u(x)$ be the cumulative distribution function of the excess of X at $u > 0$, the corresponding probability density function of the excess of X at $u > 0$ is given by $f_u(x) = f(x+u)/(1-F(u))$. The tails were classified for $\kappa < 0$, $\kappa = 0$ and $\kappa > 0$ as heavy tails, exponential tails and light tails, respectively. Hence, a distribution has a class of tail-distribution in the GPD family. Evidently, this is well defined since for any threshold $u > 0$, the shape parameter κ is invariant by the tail of GPD, that is

$$GPD_u(\kappa, \psi) = GPD(\kappa, \psi - \kappa u). \tag{7}$$

The following result shows the role of the uniform distribution in the study of tails.

Proposition 1. *Given an absolutely continuous random variable with probability density function $f(x)$ on the support $[0, x_F]$ left continuous in x_F and $0 < f(x_F) < \infty$, then its tail distribution is the uniform distribution.*

Proof. From McNeil *et al.*(2005), it must be shown that

$$\lim_{u \to x_F} \sup_{0 \leq x < x_F - u} |F_u(x) - G_u(x)| = 0, \tag{8}$$

where $G_u(x) = x/(x_F - u)$ denotes the uniform distribution on $[0, x_F - u]$ and $F_u(x) = F(x+u)/(1-F(u))$ is the cumulative distribution function of the excess at $u > 0$. Given ϵ_0 we will find u such that $\sup_{0 \leq x < x_F - u} |F_u(x) - G_u(x)| < \epsilon_0$. Let $c = f(x_F)$ be, since f is left continuous in x_F, if $\epsilon < c\epsilon_0/4$ and $\epsilon < c/2$ (equivalently, $2/(c-\epsilon) < 4/c$) there is u such that

$$c - \epsilon < f(s) < c + \epsilon, \tag{9}$$

for any $u \leq s < x_F$. Let $s = u + x$, with $x \geq 0$, the fundamental theorem of calculus in $[u, s]$ gives

$$x(c - \epsilon) < F(u + x) - F(u) < x(c + \epsilon). \tag{10}$$

In particular, for $s = x_F$,

$$(x_F - u)(c - \epsilon) < 1 - F(u) < (x_F - u)(c + \epsilon). \tag{11}$$

From the last two inequalities it follows that

$$\frac{x}{x_F-u}\left(\frac{c-\epsilon}{c+\epsilon}\right) < \frac{F(u+x)-F(u)}{1-F(u)} < \frac{x}{x_F-u}\left(\frac{c+\epsilon}{c-\epsilon}\right). \tag{12}$$

Hence,

$$|F_u(x)-G_u(x)| < \left(\frac{c+\epsilon}{c-\epsilon}-1\right)\frac{x}{x_F-u} \leq \frac{2\epsilon}{c-\varepsilon} < \frac{4}{c}\frac{c\epsilon_0}{4} = \epsilon_0.$$

□

Proposition 1 shows that if a family of right-truncated distributions (at a point where the density is different to 0) is used to model tails, then this model only contains a class of tail, the uniform distribution, for instance Aban *et al.* (2006). This is a motivation for choosing the model beforehand and once selected, to show that the existence of MLE is not a problem.

For $\kappa \leq 0$, Castillo and Daoudi (2009) show that the MLE exists for the model GPD for $\kappa \in (-\infty, 0]$, note that the global maximum holds inside the domain of κ if the empirical coefficient of variation is greater than 1. If the empirical coefficient of variation is less than 1, then in $\kappa = 0$ it has a local maximum and the authors notes that from an empirical point of view, it is global. See also Kozubowski *et al.* (2009).

3.1. The GPD family for $0 \leq \kappa \leq 1$.

Consider

$$\partial l/\partial\psi = -(n/\psi)(1-(1/\kappa-1)\frac{1}{n}\sum_{i=1}^{n}\frac{\kappa x_i/\psi}{1-\kappa x_i/\psi}, \tag{13}$$

for $\kappa \geq 0$. Simplifying and adding 1 to each addition it is found that $\partial l/\partial\psi = 0$ is equivalent to

$$1/(1-\kappa) = \frac{1}{n}\sum 1/(1-\kappa x_i/\psi). \tag{14}$$

Theorem 1. *Any local maximum of the likelihood equations for GPD always satisfies* $\kappa < 1$.

Proof. In the bounded support situation ($\kappa > 0$), $x_{(n)} \leq \sigma$ or equivalently, $\psi > \kappa x_{(n)}$, see the log-likelihood function (4). In particular $1-\kappa x_i/\psi > 0$ for all $1 \leq i \leq n$, then

$$\frac{1}{n}\sum_{i=1}^{n} 1/(1-\kappa x_i/\psi) > 0.$$

Hence, if the log-likelihood equation $\partial l/\partial\psi = 0$ holds, it is equivalent to the equation (14) and $\kappa < 1$ is obtained. □

For $0 \leq \kappa \leq 1$, Choulakian and Stephens (2001) show that given $\kappa < 1$ fixed a single solution exists for $\partial l/\partial\psi = 0$ and this is a maximum that we denote $\hat{\psi}(\kappa)$. We are going to explore this.

Proposition 2. *The function,* $\hat{\psi}(\kappa)$*, for* $0 < \kappa < 1$ *is continuous, monotone increasing and differentiable. Moreover,* $\bar{x} \leq \hat{\psi}(\kappa) \leq x_{(n)}$ *and the function extends continuously to the boundary with,*

$$\lim_{\kappa\to 0}\hat{\psi}(\kappa) = \bar{x} \quad \textit{and} \quad \lim_{\kappa\to 1}\hat{\psi}(\kappa) = x_{(n)}. \tag{15}$$

Consequently, the curve $(\kappa, \hat{\psi}(\kappa))$ *for* $0 \leq \kappa \leq 1$ *will be called the Choulakian-Stephens curve.*

Proof. The continuity and differentiability of the function is obtained from the implicit function theorem applied to $g(\kappa,\psi) = 1/(1-\kappa) - \frac{1}{n}\sum 1/(1-\kappa x_i/\psi) = 0$ defined in the product of open intervals $(0,1)\times(\bar{x}, x_{(n)})$. We merely have to prove that the partial derivatives are different to zero for each κ. We compute them:

$$g_1(\kappa,\hat{\psi}(\kappa)) = \frac{1}{\kappa}\left(\frac{\kappa}{(1-\kappa)^2} - \frac{1}{n}\sum\frac{\kappa x_i/\hat{\psi}(\kappa)}{(1-\kappa x_i/\hat{\psi}(\kappa))^2}\right), \tag{16}$$

$$g_2(\kappa, \hat{\psi}(\kappa)) = \frac{1}{\hat{\psi}(\kappa)} \frac{1}{n} \sum \frac{\kappa x_i / \hat{\psi}(\kappa)}{(1 - \kappa x_i / \hat{\psi}(\kappa))^2}. \tag{17}$$

Of course, g_2 is strictly positive for all κ. We are going to prove that $g_1 < 0$. From $(1 - a)^2$ develops, we obtain the equivalence

$$2\frac{a}{(1-a)^2} = \left(\frac{1}{1-a}\right)^2 + \left(\frac{a}{1-a}\right)^2 - 1, \tag{18}$$

for all a. On the other hand,

$$1/(1-\kappa) = \frac{1}{n} \sum \frac{1}{1 - \kappa x_i / \hat{\psi}(\kappa)} \Rightarrow \left(\frac{1}{1-\kappa}\right)^2 < \frac{1}{n} \sum \left(\frac{1}{1 - \kappa x_i / \hat{\psi}(\kappa)}\right)^2 \tag{19}$$

and

$$\kappa/(1-\kappa) = \frac{1}{n} \sum \frac{\kappa x_i / \hat{\psi}(\kappa)}{1 - \kappa x_i / \hat{\psi}(\kappa)} \Rightarrow \left(\frac{\kappa}{1-\kappa}\right)^2 < \frac{1}{n} \sum \left(\frac{\kappa x_i / \hat{\psi}(\kappa)}{1 - \kappa x_i / \hat{\psi}(\kappa)}\right)^2. \tag{20}$$

Both the equalities are direct consequence of $g(\kappa, \hat{\psi}(\kappa)) = 0$. Then,

$$2\frac{1}{n} \sum \frac{\kappa x_i / \hat{\psi}(\kappa)}{(1 - \kappa x_i / \hat{\psi}(\kappa))^2} = \frac{1}{n} \sum \left(\frac{1}{1 - \kappa x_i / \hat{\psi}(\kappa)}\right)^2 + \frac{1}{n} \sum \left(\frac{\kappa x_i / \hat{\psi}(\kappa)}{1 - \kappa x_i / \hat{\psi}(\kappa)}\right)^2 - 1 >, \tag{21}$$

$$> \left(\frac{1}{1-\kappa}\right)^2 + \left(\frac{\kappa}{1-\kappa}\right)^2 - 1 = 2\frac{\kappa}{(1-\kappa)^2}, \tag{22}$$

therefore $g_1 < 0$. Finally,

$$\hat{\psi}'(\kappa) = -\frac{g_1(\kappa, \hat{\psi}(\kappa))}{g_2(\kappa, \hat{\psi}(\kappa))} > 0, \tag{23}$$

for all κ, then the function is increasing for all $0 \leq \kappa \leq 1$, see Figure 6.

We are going to show that the limits in (15) hold and therefore $\hat{\psi}(\kappa)$ is continuous in the closed set $[0, 1]$. Note that $\hat{\psi}(\kappa)$ is implicitly defined by (14), then $1/(1-\kappa) = \frac{1}{n} \sum 1/(1 - \kappa x_i / \hat{\psi}(\kappa))$. If $\kappa = 1$ then $1/(1 - \kappa x_i / \hat{\psi}(\kappa)) = 0$ for some $i = 1 \cdots n$ hence $\hat{\psi}(1) = x_i$. Since $x_i \leq \psi$ for all $i = 1 \cdots n$, $\hat{\psi}(1) = x_{(n)}$. On the other hand, observe that

$$1/(1-\kappa) = \frac{1}{n} \sum 1/(1 - \kappa x_i / \hat{\psi}(\kappa)) \Rightarrow \kappa/(1-\kappa) - \frac{1}{n} \sum (\kappa x_i / \hat{\psi}(\kappa))/(1 - \kappa x_i / \hat{\psi}(\kappa)),$$

hence

$$1/(1-\kappa) = \frac{1}{n} \sum (x_i / \hat{\psi}(\kappa))/(1 - \kappa x_i / \hat{\psi}(\kappa)).$$

From the last equality, $\hat{\psi}(0) = \frac{1}{n} \sum x_i = \bar{x}$. □

Proposition 3. *The likelihood function on the Choulakian-Stephens curve is continuous. That is the curve defined by the profile-likelihood function*

$$l_p(\kappa) = l(\kappa, \hat{\psi}(\kappa)),$$

for $0 \leq \kappa \leq 1$. *In fact, the limits for* $\kappa = 0$ *and* $\kappa = 1$ *correspond to the log-likelihood values for the exponential and the uniform distributions,*

$$\lim_{\kappa \to 0} l_p(\kappa) = -n \log(\bar{x}) - n \quad \textit{and} \quad \lim_{\kappa \to 1} l_p(\kappa) = -n \log(x_{(n)}). \tag{24}$$

Proof. Consider the Choulakian-Stephens curve parameterized by $\sigma = \psi/\kappa$, for $\sigma \in [x_{(n)}, \infty)$. Then

$$l(\sigma) = n\left(-\log(\sigma\kappa(\sigma)) + \kappa(\sigma)\frac{1}{n} \sum \log(1 + x_i/\sigma)^{-1}\right),$$

note that $\kappa(\sigma) = 1 - \frac{1}{n}\sum(1 + x_i/\sigma)^{-1}$ since it is defined by the Choulakian-Stephens curve, see equation (14). Using the l'H$\hat{o}$pital rule once for each one, the next limits hold

$$\lim_{\sigma\to\infty} \sigma\kappa(\sigma) = \bar{x} \quad \text{and} \quad \lim_{\sigma\to x_{(n)}^{+}} \sigma\kappa(\sigma) = x_{(n)},$$

$$\lim_{\sigma\to\infty} \kappa(\sigma)\frac{1}{n}\sum \log(1 + x_i/\sigma)^{-1} = 1 \quad \text{and} \quad \lim_{\sigma\to x_{(n)}^{+}} \kappa(\sigma)\frac{1}{n}\sum \log(1 + x_i/\sigma)^{-1} = 0.$$

□

Theorem 2. *Consider the model GPD for* $0 \le \kappa \le 1$*, then the global MLE exists.*

Proof. To summarize, by the Proposition 2, the function, $\hat{\psi}(\kappa)$, defines the MLE for the parameter ψ for each κ fixed in $(0, 1)$. Given an arbitrary sample, for $\kappa = 0$ and $\kappa = 1$ the MLE for the parameter ψ is $\bar{x}$ and $x_{(n)}$, respectively. $\hat{\psi}(\kappa)$ is continuous in the closed set $[0, 1]$. The set $\{\, (\kappa, \hat{\psi}(\kappa)) \mid 0 \le \kappa \le 1 \,\}$ is the Choulakian-Stephens curve and is a closed subset of $[0, 1] \times [\bar{x}, x_{(n)}]$, so it is a compact set. The likelihood function on the Choulakian-Stephens curve is continuous, by Proposition 3, so there is a global MLE. □

Note that the corresponding values for the log-likelihood function on the extremes of the Choulakian-Stephens curve are the values of the log-likelihood function for exponential and uniform distributions, since the limits (24) and therefore Theorem 2 justifies that the AIC and BIC work. An sketched of the simulation tests show that the uniqueness of Theorem 2 may occur with probability one.

3.2. The GPD family for $\kappa \ge 1$.

In the GPD model for $\kappa \ge 1$ the likelihood function has no local maximum since Proposition 1. Moreover, the likelihood function is unbounded and reaches infinity, in a half-line of the parameter space.

Proposition 4. *For any* $\kappa > 1$*, there is* ψ *such that* $l(\kappa, \psi)$ *is larger than an arbitrary fixed value.*

Proof. Fixed $\kappa > 1$, the value $(1/\kappa - 1)$ is positive, however, if ψ tends to $\kappa x_{(n)}$ then $l(\kappa, \psi)$ tends to infinity, see the loglikelihood function in (4). □

The MLE in GPD for $\kappa > 1$ has important mathematical limitations. If $\kappa > 1$ then a local maximum of the likelihood function does not exist, see Proposition 1. The global maximum of the likelihood function fulfills $\sigma = x_{(n)}$. The parameters κ and ψ are not uniquely determined, since for every κ, $l(\kappa, \kappa x_{(n)}) = \infty$. The statistical methods for model comparison, AIC, BIC and LRT cannot be used, since for $\kappa > 1$ the likelihood is always infinite.

From a numerical point of view, the duly modified MLE provides satisfactory results. The likelihood function is defined for $\kappa\psi = \sigma > x_{(n)}$ and for each value of σ, the likelihood function has a local maximum that is global. Simply, assuming that there are rounding errors, the MLE for σ must be strictly greater than $x_{(n)}$ and we have a satisfactory solution for the estimation. In fact, some papers present conventional solutions for non-regular problems in MLE, see Cheng and Iles (1987), Cheng and Amin (1983), Smith (1985) and Hall and Wang (2005) for the methodology. The algorithm proposed in Section 2 gives us a solution. In general, all these solutions are similar, see Table 4 and Table 5.

4. Methodological approach

The analysis conducted in this article enables us to make some comments that lead to methodological recommendations. Section 3 has shown the importance of the uniform distribution for modeling tails. Many commonly used families of distributions, such as stable distributions or the Pareto distribution, are right-truncated in practice, so they have uniform tails, see Aban *et al.* (2006).

The exponential and uniform distributions make it possible to mark off three submodels by the GPD: the Pareto distribution (A: $\kappa \le 0$), the submodel of monotone decreasing densities with compact support (B: $0 \le \kappa \le 1$) and the submodel of monotone increasing densities with compact support (C: $\kappa \ge 1$). The errors of classification in these submodels are a problem in statistical modelling, see Table 1, since they qualitatively behave very differently. Therefore,

the first recommendation is to **specify *a priori* the submodel of the GPD** in order to estimate the parameters. Of course, this depends on the dataset analyzed, and does not provide an automatic algorithm.

The specification *a priori* of the GPD submodel is reasonable in many cases. If the basic model for financial data is the normal distribution (the values are not in a bounded interval), but there is evidence that the extreme values significantly exceed what is expected under this model, then the tail models with bounded support ($\kappa > 0$) have to be discarded. If it is assumed that the tail has to be compact support, as in hydrological problems, we believe it is necessary to consider *a priori* whether the tails can be as extremely light as in the case ($\kappa > 1$) where extreme values are more likely than values close to the mean, or whether it is better to limit the model to B, see Figure 1.

Table 2. Percentages of the classification for each submodel for some sample size, for values of κ and for percentages for the confidence interval by the exponential and uniform distribution. Model A corresponds to the submodel of GPD for $\kappa \leq 0$, model B for $0 \leq \kappa \leq 1$ and model C for $\kappa \geq 1$. The bold values correspond to well classified categories.

		A			**B**						**C**		
	κ	-0.1			0.1			0.9			1.1		
n	%	**A**	B	C	A	**B**	C	A	**B**	C	A	B	**C**
15													
	0%	**41.4**	50.2	8.4	18.8	**64.6**	16.6	0.1	**11.2**	88.8	0.0	4.1	**95.9**
	20%	**58.4**	33.6	8.0	18.8	**67.5**	13.7	0.11	**30.2**	69.7	0.0	4.0	**96.1**
	40%	**76.8**	17.6	5.6	18.9	**70.8**	10.2	0.07	**50.2**	49.8	0.0	4.3	**95.7**
	60%	**86.5**	10.3	3.1	11.1	**83.3**	5.6	0.03	**69.7**	30.3	0.0	5.7	**94.3**
50													
	0%	**63.0**	37.0	0.0	15.2	**84.8**	0.0	0.0	**30.1**	69.9	0.0	4.8	**95.2**
	20%	**77.6**	22.5	0.0	15.6	**84.4**	0.0	0.0	**45.0**	55.0	0.0	4.2	**95.8**
	40%	**84.7**	15.3	0.0	10.1	**89.9**	0.0	0.0	**60.2**	39.8	0.0	3.2	**96.9**
	60%	**91.3**	8.7	0.0	5.2	**94.8**	0.0	0.0	**74.5**	25.5	0.0	2.6	**97.4**

To estimate the parameters when the shape parameter is large and negative, MLE performs fine, as has been shown in many publications. If the submodel specified is A and the empirical coefficient of variation of the data is larger than 1, then a local maximum for the log-likelihood function exists, see Castillo and Daoudi (2009); otherwise the global maximum is attained at $\kappa = 0$. If the submodel specified is B, the MLE always exists as the global maximum of the likelihood function, see Theorem 2. If the submodel specified is C, in numerical methods it is necessary to avoid the estimations with $\sigma = x_{(n)}$; for instance, the routine eGPD in Section 2 provides an estimation considering the estimation of σ to be slightly larger than $x_{(n)}$.

The abovementioned submodels have boundary points that correspond to the exponential or the uniform distributions. These distributions appear on the tail limit of many common distributions, we have already mentioned the uniform case. Remember that in the Pickands-Balkema and DeHaan theorem the exponential is the limit of distributions as different as the normal distribution and the log-normal distribution, which some authors consider to be a heavy tail distribution. Hence, if the complete GPD model is considered, we recommend **assuming the exponential (E) or the uniform (U) distributions to be possible models**, i.e. assigning a weight to them in advance.

The usual way to do this is to consider the models E or U as the null hypothesis testing against GPD, see Kozubowski *et al*. (2009). Alternatively, one can consider the criteria AIC or BIC for model selection, see Clauset *et al.* (2009). In contrast, the GPD model assigns a zero weight to E and U. In the absence of any information about the value of the shape parameter a middle way is to give equal weight to the five models. The standard recommendation of statistics professionals is to consider **the simplest models**.

In order to illustrate how assigning weights to E and U modifies the misspecification of MLE in GPD we will run again the simulation in Table 1 for MLE, now assigning probabilities for E and U of 0.20, 0.40 and 0.60. That is, if an observation is in the confidence interval of E or U with the selected probability then the corresponding model is assigned. Otherwise, eGPD code is used to allocate models A, B or C. Many observations that were misclassified are now allocated to models E and U and, since these models are in the boundary, they can be correctly classified in A, B or C.

AIC, BIC and LRT cannot be used to compare model C with U, since the likelihood always tends to infinite in

C. However, exact confidence intervals for U and E are easily obtained by simulation of the empirical coefficient of variation. From (2) it is clear that in GPD the coefficient of variation is a one-to-one transformation of the shape parameter κ and does not depend on ψ, hence the simulation is straightforward.

Table 2 is a summary of the results of assigning weights to E and U. The percentages of the classification for submodels A, B and C for several sample sizes are reported in terms of the probabilities used for the confidence intervals for E and U (0.00, 0.20, 0.40 and 0.60), obtained by simulation of the empirical coefficient of variation. The bold values correspond to well-classified categories. The same results are obtained with several sample sizes. Clearly increasing the assigned weight improves the correct classification rates.

5. Real world examples

In this Section, two of the most commented examples of extreme values theory will be discussed. Note that they are both around the transition points E and U. The first are the exceedances of a threshold level by the River Nidd. The second is the zero-crossing hourly mean periods of the sea waves in Bilbao. We emphasize that as well as to estimating the GPD, there is the problem of selecting the threshold from which the model can be considered acceptable.

To find the threshold from which the GPD model is appropriate, Castillo *et al.* (2014) suggest use of the residual coefficient of variation, since it characterizes the distribution and does not depend on the scale parameter, see Gupta and Kirmani (2000). Given a sample $\{x_k\}$ of positive numbers, let $\{x_{(k)}\}$ be the ordered sample, so that $x_{(1)} \leq x_{(2)} \leq \ldots \leq x_{(n)}$. Defining a *CV-plot* as the representation of the empirical coefficient of variation (cv_n) of the conditional exceedances, given by

$$k \to cv\left(x_{(j)} - x_{(k)} : j \geq k\right). \tag{25}$$

Note that the residual coefficient of variation for GPD is the constant given in (2), since (7) shows that κ is unchanged taking conditional exceedances. The asymptotic confidence intervals in Figure 4 and Figure 5 are obtained from the delta method by

$$\sqrt{n}\,(cv_n - c) \to N\left(0, \sigma^2\right), \tag{26}$$

where $c = 1$ and $\sigma^2 = 1$ for E, and $c = 1/\sqrt{3}$ and $\sigma^2 = 8/45$ for U.

In the examples, the MLE is computed for several thresholds, using eGPD in Section 2. For the Nidd data, AIC and BIC criteria provide similar results to exact confidence intervals for U and E obtained by simulation of the empirical coefficient of variation. For the Bilbao data, where $\hat{\kappa}$ may be above or below 1, the AIC and BIC criteria are not used since the likelihood always tends to infinite in model C.

5.1. On Nidd exceedance data

The dataset consists of 154 exceedances of the threshold level 65 m3/s by the River Nidd at Hunsingore Weir from 1934 to 1969. Many authors have explored this dataset, from Hosking and Wallis (1987) and Davison and Smith (1990) to Papastathopoulos and Tawn (2012). The question is to decide whether the tail is exponential or not, and the latter authors solve the problem by changing the model for the tail.

Figure 4 shows the evolution of the empirical coefficient of variation (25) in terms of thresholds and quantiles for the Nidd data with asymptotic confidence intervals under exponentiality, given by (26), for 95% and 99% respectively. Above threshold 95, the empirical coefficient of variation is in the confidence interval of the exponential distribution. The qualitative behavior given by Figure 4 can be compared to the evolution of MLE for κ. Table 3 shows that $\hat{\kappa}$ evolves around $\kappa = 0$, which is almost reached at the threshold $u = 100$. Hence, asymptotic theory for MLE exists and LRT, AIC and BIC criteria can be used to model selection between Pareto ($\kappa < 0$) and exponential ($\kappa = 0$).

Table 3 shows whether we accept Pareto (P) or exponential (E) in terms of the threshold and selection criteria used. Although threshold 120 may suggest a model with compact support ($\kappa = 0.25$), our methodological approach leads us to reject this possibility. Above threshold 95 all methods conclude exponentially, BIC decides on exponential even above 85 which is also reasonable. In any case, the conclusion is to accept the exponential model at the tail.

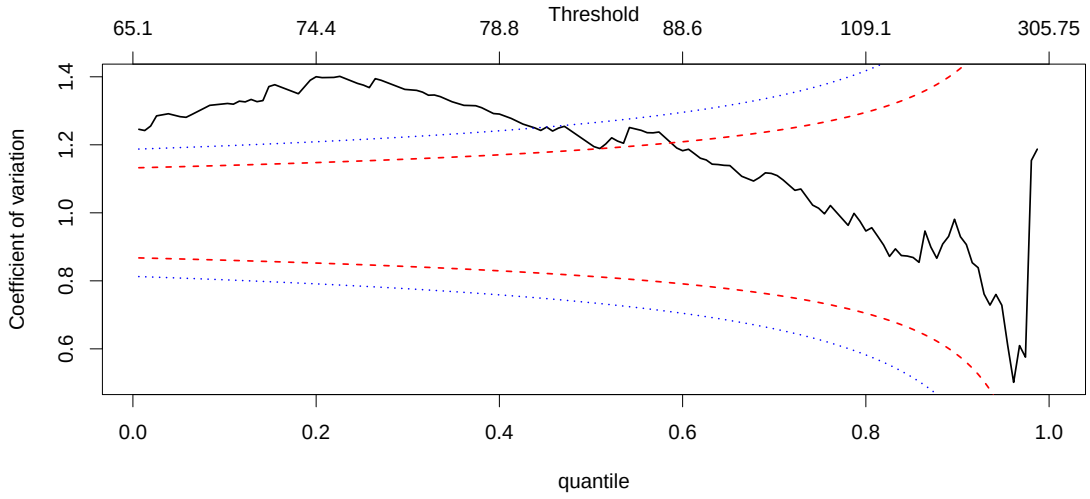


Figure 4. The solid line is the evolution of the empirical coefficient of variation in terms of thresholds and quantiles for the Nidd data. The dashed and dotted lines correspond to the asymptotic confidence intervals under exponentiality ($cv = 1$) for 95% and 99%, respectively.

Table 3. MLE for GPD parameters and selection criteria between Pareto (P) and exponential (E) for several thresholds.

u	n	κ	ψ	lv_{GPD}	lv_E	LRT		AIC		BIC		cv
65	154	-0.20	26.26	-688.4	-691.8	6.97	P	-4.97	P	-0.10	P	1.25
75	117	-0.47	18.16	-511.6	-520.6	17.92	P	-15.92	P	-11.32	P	1.38
85	72	-0.35	26.29	-332.8	-335.4	5.25	P	-3.25	P	0.86	E	1.24
95	49	-0.21	35.91	-235.0	-235.5	1.13	E	0.87	E	4.60	E	1.13
100	39	0.00	50.62	-192.2	-192.2	0.00	E	2.00	E	5.50	E	1.01
110	31	0.07	56.38	-153.8	-153.9	0.08	E	1.92	E	5.19	E	0.98
120	24	0.25	71.64	-120.6	-121.0	0.85	E	1.15	E	4.17	E	0.88

5.2. On Bilbao waves data

This real-world example, originally analyzed in Castillo and Hadi (1997), consists of the zero-crossing hourly mean periods (in seconds) of the sea waves measured in the Bilbao bay, Spain. Later on, this dataset was revisited in Luceño (2006) and in Zhang and Stephens (2009). Only the 197 observations with periods above 7 seconds were taken into consideration.

Figure 5 shows the evolution of the empirical coefficient of variation (25) in terms of thresholds and quantiles for the Bilbao data with asymptotic confidence intervals under uniformity, given by (26), for 95% and 99% respectively. We can observe that the coefficient of variation of the tail becomes constant for a threshold around 8 or 8.5. Moreover, the qualitative analysis provided by the CV-plot suggests that the uniform tail has to be considered to model the tail of the data and submodel A must be discarded.

Modeling these data by the GPD is done with thresholds at t = 7, 7.5, 8, 8.5, 9, 9.5 following the above mentioned authors. They all agree that the MLE does not exist for the latter three cases and effectively no local maximum exists. Indeed, the MLE is not a solution for the likelihood equations. However the algorithm in Section 2 provides a solution, see the underlined values in Table 4. Note that Grimshaw (1993) proposes a method to estimate parameters and if we apply it to the Bilbao wave data it gives these same results.

Table 4 shows, for every threshold of data, the estimations of the parameters obtained by SSE, ZSE and MLE. In every case, M denotes the corresponding maximum for the Bilbao waves obtained from the σ estimation and σ is obtained from the relation $\sigma = \psi/\kappa$. The underlined values in the MLE frame correspond to the cases of non-existing solutions from likelihood-equations. For SSE and ZSE the maximums are overestimated and/or do not seem stable in

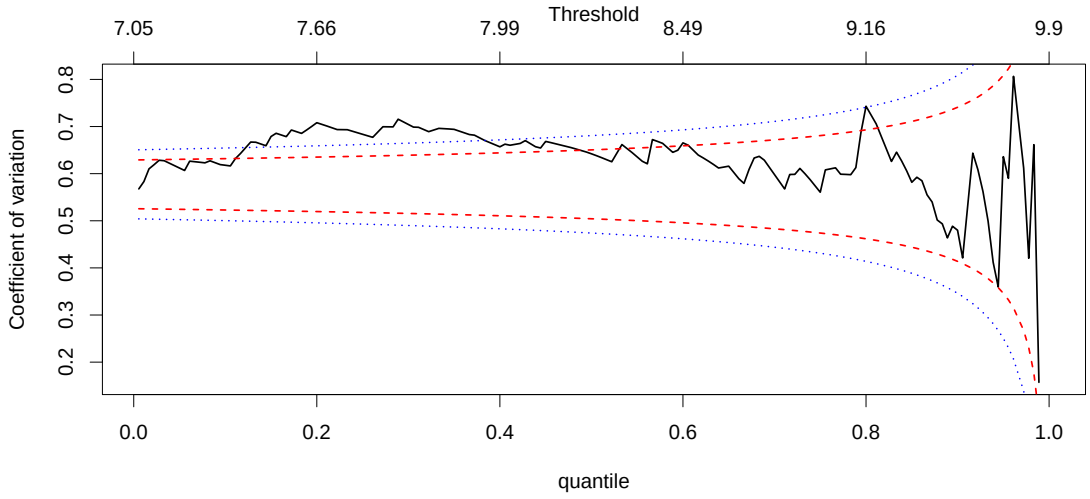


Figure 5. The solid line is the evolution of the empirical coefficient of variation in terms of thresholds and quantiles for the Bilbao data. The dashed and dotted lines correspond to the asymptotic confidence intervals under uniformity ($cv = 1/\sqrt{3}$) for 95% and 99%, respectively.

Table 4. For some thresholds of Bilbao wave data and every estimation method: ZSE, SSE and MLE, the values of parameter estimation are shown.

	SSE			ZSE			MLE		
u	κ	ψ	$\sigma\,(M)$	κ	ψ	$\sigma\,(M)$	κ	ψ	$\sigma\,(M)$
7	0.84	2.44	2.90 (9.90)	0.81	2.38	2.95 (9.95)	0.86	2.50	2.91 (9.91)
7.5	0.56	1.60	2.82 (10.32)	0.71	1.75	2.49 (9.99)	0.77	1.86	2.42 (9.92)
8	0.63	1.42	2.25 (10.25)	0.77	1.51	1.96 (9.96)	0.86	1.65	1.91 (9.91)
8.5	0.77	1.18	1.54 (10.04)	0.83	1.21	1.45 (9.95)	1.06	1.49	1.40 (9.90)
9	0.80	0.81	1.02 (10.02)	0.88	0.83	0.94 (9.94)	1.19	1.07	0.90 (9.90)
9.5	-0.63	0.22	-0.35 (9.15)	1.01	0.43	0.43 (9.93)	1.53	0.61	0.40 (9.90)

the tail.

The MLE problem for non-regular cases has been worked on intensely, the main results being in Smith (1985), Cheng and Amin (1983), Cheng and Iles (1987) and Hall and Wang (2005). Table 5 provides the estimated GPD parameters for the Bilbao wave data for each method of correction of MLE proposed by the corresponding authors. Furthermore, in each case we show a corresponding value of the support for the GPD distribution, σ, and a subsequent proposal for the maximum for the Bilbao waves, M. In every case, the values of estimations for κ are around the value 1. The uniform tail hypothesis therefore has to be checked, with a possible suggestion being, the use of a right-truncated model to model whole data, Proposition 1.

From the support of the distribution, σ, all estimations considered in Table 4 and Table 5 are inside the 95% confidence interval obtained from the uniformity hypothesis. For instance, from the distribution of maximum, a confidence interval for σ under the uniformity hypothesis can be computed as

$$x_{(n)} < \sigma < x_{(n)}/\alpha^{(1/n)}, \tag{27}$$

where α is the statistical significance. Therefore in each case we cannot reject the hypothesis of uniformity for the tail of data. But, from Figure 5 the tail of data should be taken from the thresholds 8 or 8.5. Effectively, for data above 8.5 the p-value for the hypothesis of uniformity based on the asymptotic result 26 is 0.22 whereas for threshold 8 the p-value is 0.04. Therefore, the hypothesis of uniformity cannot be rejected for the data for threshold 8.5. In any case, the conclusion is to accept the uniform model at the tail.

Table 5. For some thresholds of Bilbao waves data and some estimation methods for non-regular problems, the values of parameters estimation are showed.

	Smith			Cheng&Illes			Hall&Wang		
u	κ	ψ	$\sigma\,(M)$	κ	ψ	$\sigma\,(M)$	κ	ψ	$\sigma\,(M)$
7	0.84	2.43	2.90 (9.90)	0.86	2.50	2.91 (9.91)	0.87	2.51	2.90 (9.90)
7.5	0.77	1.85	2.40 (9.90)	0.77	1.86	2.42 (9.92)	0.80	1.92	2.40 (9.90)
8	0.83	1.58	1.90 (9.90)	0.86	1.65	1.91 (9.91)	0.87	1.66	1.90 (9.90)
8.5	0.90	1.26	1.40 (9.90)	1.02	1.42	1.40 (9.90)	0.96	1.34	1.40 (9.90)
9	0.95	0.85	0.90 (9.90)	1.13	1.01	0.90 (9.90)	1.04	0.93	0.90 (9.90)
9.5	1.07	0.43	0.40 (9.90)	1.38	0.55	0.40 (9.90)	1.22	0.49	0.40 (9.90)

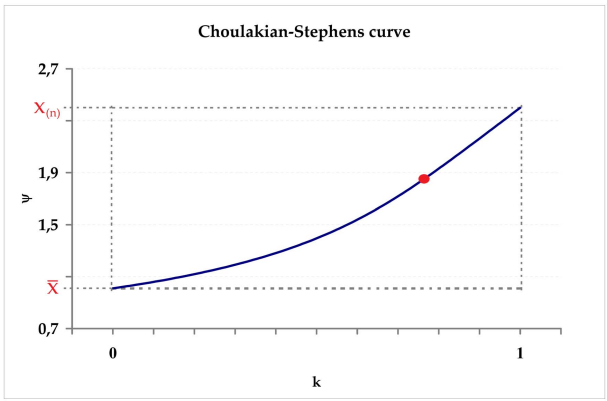


Figure 6. Example of Choulakian-Stephens curve for Bilbao wave data above 7.5. The large dot shows the MLE of GPD for this subset of data.

6. Concluding remarks

- The estimation of the GPD by MLE can be weak since for some datasets the likelihood function appears to have no local maximum, especially for small samples and $\kappa > 0$. In fact, for $\kappa > 1$ there is never exist a local maximum. The model has three submodels separated by the exponential ($\kappa = 0$) and the uniform distributions ($\kappa = 1$).
- Asymptotic theory for MLE exists only if $\kappa < 0.5$. Thus, the LRT cannot be used beyond this value.
- The AIC and BIC criteria can be used if $\kappa \leq 1$, since it has been shown that for $0 \leq \kappa \leq 1$ the likelihood function extends continuously to the boundary points.
- From a strictly mathematical point of view, for $\kappa > 1$ the likelihood is always infinite. Hence the LRT, BIC and AIC criteria cannot be used.
- There are algorithms for estimation of the GPD by MLE even when $\kappa > 1$, its values depend on the accuracy. Generally, they give stable and equivalent results.
- Exact confidence intervals for any value of $\kappa > -0.5$ are easily obtained by simulation of the empirical coefficient of variation, since it does not depend on ψ.

Acknowledgement

This work was supported by the Spanish Ministerio de Educación y Ciencia (MEC) under Grant Procesos Estocásticos Aplicados, MTM 2012-31118.

References

[1] Aban, B. I., Meerschaert, M.M. and Panorska, K. A. (2006). Parameter Estimation for the Truncated Pareto Distribution. *J. Amer. Statist. Assoc.*, 101, 270 - 277.

[2] Balkema, A. and de Haan, L. (1974). Residual life time at great age. *Annals of Probability*, 2, 792 804.

[3] Beirlant, J., Goegebeur, Y., Segers, J. and Teugels, J. (2004). *Statistics of Extremes: Theory and applications*. Wiley Series in Probability and Statistics.

[4] Brent, R. (1973). *Algorithms for Minimization without Derivatives*. Englewood Cliffs N.J.: Prentice-Hall.

[5] Castillo, E. and Hadi, A. S. (1997). Fitting the generalized Pareto distribution to data. *J. Amer. Statist. Assoc.*, 92, 1609 - 1620.

[6] Castillo, J. and Daoudi, J. (2009). Estimation of the generalized Pareto distribution. *Statistics & Probability Letters*, 79, 684 - 688.

[7] Castillo, J.; Daoudi, J. and Lockhart, R. (2014) Methods to distinguish between polynomial and exponential tails. *Scandinavian Journal of Statistics*, 41, 382 393.

[8] Cheng, R. C. H. and Amin, N. A. K. (1983). Estimating parameters in continuous univariate distributions with a shifted origin. *J. Roy. Statist. Soc. Ser. B*, 45, 394 - 403.

[9] Cheng, R. C. H. and Iles, T. C. (1987). Corrected maximum likelihood in nonregular problems. *J. Roy. Statist. Soc. Ser. B*, 49, 95 - 101.

[10] Choulakian, V. and Stephens, M. A. (2001). Goodness-of-Fit for the Generalized Pareto Distribution. *Technometrics*, 43, 478 - 484.

[11] Clauset A., Shalizi C. and Newman, M. (2009). Power-law distributions in empirical data. *SIAM Review*, 51, 661 - 703.

[12] Coles, S. (2001). *An Introduction to Statistical Modeling of Extreme Values.* Springer Series in Statistics.

[13] Davison, A. C. and Smith, R. L. (1990). Models for exceedances over high thresholds. With discussion and a reply by the authors. *J. Roy. Statist. Soc. Ser. B*, 52, 393 - 442.

[14] Embrechts, P.,Klüppelberg, C. and Mikosch., T. (1997). *Modelling Extremal Events for Insurance and Finance.* Springer, Applications of Mathematics, 33.

[15] Finkenstädt, B. and Rootzén, H. (2003). *Extreme Values in Finance, Telecommunications and the Enviroment*, Monographs on Statistics and Applied probability, 99.

[16] Grimshaw, S. D. (1993). Computing maximum likelihood estimates for the generalized Pareto distribution. *Technometrics* , 35, 185 - 191.

[17] Gupta, R and Kirmani, S. (2000). Residual coefficient of variation and some characterization results. *J. Stat. Plan. Infer.*, 91, 23 31.

[18] Hall, P. and Wang, J. Z. (2005). Bayesian likelihood methods for estimating the end point of a distribution. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 67, 717 - 729.

[19] Hosking, J.R.M. and Wallis, J.R. (1987). Parameter and quantile estimation for the generalized Pareto distribution. *Technometrics*, 29, 3, 339 - 349.

[20] Juarez, S. F. and Schucany, W. R. (2004). Robust and efficient estimation for the generalized pareto distribution. *Extremes*, 7, 237 - 251.

[21] Kozubowski, T. J.; Panorska, A. K.; Qeadan, F.; Gershunov, A. and Rominger, D. (2009) Testing exponentiality versus Pareto distribution via likelihood ratio. *Comm. Statist. Simulation Comput.*, 38, 118 - 139.

[22] Luceño, A. (2006). Fitting the generalized Pareto distribution to data using maximum goodness-of-fit estimators. *Comput. Statist. Data Anal.*, 51, 2, 904 - 917.

[23] Mackay, E.B.L, P.G. Challenor, A.S. Bahaj (2011). A comparison of estimators for the generalized Pareto distribution. *Ocean Engineering*, 38, 1338 - 1346.

[24] McNeil, A. J., Frey, R. and Embrechts P. (2005). *Quantitative Risk Management: Concepts, Techniques and Tools.* Princeton University Press.

[25] Papastathopoulos, I. and Tawn, J.A. (2012). Extended generalised Pareto models for tail estimation. *J. Statist. Plann. Inference*, 143, 131 - 143.

[26] Pickands, J. (1975). Statistical inference using extreme order statistics. *Annals of Statistics*, 3, 119 - 131.

[27] Smith, R. L. (1985). Maximum likelihood estimation in a class of non-regular cases. *Biometrika*, 72, 67 - 90.

[28] Song, J and Song, S. (2012). A quantile estimation for massive data with generalized Pareto distribution. *Comput. Statist. Data Anal.*, 56, 143 - 150.

[29] Zhang, J. (2007). Likelihood moment estimation for the generalized pareto distribution. *Australian & New Zealand Journal of Statistics*, 49, 69 77.

[30] Zhang, J. (2010). Improving on estimation for the generalized Pareto distribution. *Technometrics* 52, 335 - 339.

[31] Zhang, J. and Stephens, M. A. (2009). A new and efficient estimation method for the generalized Pareto distribution. *Technometrics* 51, 316 - 325.