



On Maximum Likelihood Estimates for the Shape Parameter of the Generalized Pareto Distribution

Kouider Mohammed Ridha

Department of Mathematics, Applied Mathematics Laboratory, University of Mohamed Khider, Biskra, Algeria

Email address:

mohakouider@gmail.com

To cite this article:

Kouider Mohammed Ridha. On Maximum Likelihood Estimates for the Shape Parameter of the Generalized Pareto Distribution. *Science Journal of Applied Mathematics and Statistics*. Vol. 7, No. 5, 2019, pp. 89-94. doi: 10.11648/j.sjams.20190705.15

Received: April 14, 2019; **Accepted:** May 24, 2019; **Published:** October 26, 2019

Abstract: The general Pareto distribution (GPD) has been widely used a lot in the extreme value for example to model exceedance over a threshold. Feature of The GPD that when applied to real data sets depends substantially and clearly on the parameter estimation process. Mostly the estimation is preferred by maximum likelihood because have a consistent estimator with lowest bias and variance. The objective of the present study is to develop efficient estimation methods for the maximum likelihood estimator for the shape parameter or extreme value index. Which based on the numerical methods for maximizing the log-likelihood by introduce an algorithm for computing maximum likelihood estimate of The GPD parameters. Finally, a numerical examples are given to illustrate the obtained results, they are carried out to investigate the behavior of the method.

Keywords: Extreme Value Index, Generalized Pareto Distributions, Excesses Over High Thresholds, Maximum Likelihood, The Modified Bisection Method Algorithm

1. Introduction

The GPD are the only "stable" ones, i.e. the only ones for which the conditional distribution of an exceedance is a scale transformation of the original distribution. The GPD distribution has many good properties, (for more details see, for instance, Embrechts, Klüppelberg and Mikosch, 1997, Section 3.4, [4], and Reiss and Thomas, Section 1.4, [11]).

The maximum likelihood estimate of the general Pareto distribution (GPD) parameters are the values which maximize the likelihood function which is defined as below formula (6).

The asymptotic behavior properties of maximum likelihood estimator of the GPD parameter have been studied in many articles including the important works of Davison [2] and R.L Smith [12], the maximum likelihood estimators have a consistent estimator of the variance and he used it to replace the asymptotic variance of unknown parameters. The maximum likelihood estimates must be derived numerically for the GPD because there is no obvious simplification of the nonlinear likelihood equation as we defined in (8).

From a statistical perspective, the threshold is loosely defined such that the population tail can be well approximated by an extreme value model (e.g., the

generalized Pareto distribution), and can be used the GPD maximum likelihood estimates for estimating extreme value index.

There is many numerical techniques for computes the GPD maximum likelihood estimates for the shape parameters $\gamma \in \mathbb{R}$ have been proposed in many articles including the important works like Hosking and Wallis [10] for the parameter space to $(1/2) < \gamma < (-1/2)$ and Grimshaw [7] for extreme value index, the shape parameter $\gamma \leq 1$, and we propose an algorithm, to estimate the extreme value index, the shape parameter $\gamma \geq -1$. For this we introduce the approach maximum likelihood of the extreme value index. Then, in section (2) we gives the numerical techniques for computes the GPD maximum likelihood estimates. And in section (3) we will be given a numerical example is based on two real data found in the literature, he is in order to illustrate the problem with estimation of the shape parameter $\gamma \in \mathbb{R}$ in small samples by the maximum likelihood estimator.

Let X_1, X_2, \dots, X_n be a sequence of independent and identically distribution (i. i. d). random variables from some unknown distribution function (d.f.) F . Denote the upper endpoint of F by τ_F where $\tau_F = \sup\{x : F(x) < 1\} \leq \infty$ and

let

$$F_t(x) = P(X \leq t+x | X > t) = \frac{F(x+t) - F(t)}{1 - F(t)} \quad (1)$$

with $1 - F(t)$, $t < \tau_F$ and $x > 0$, be the conditional d.f. of $X - t$ given $X > t$.

For an i.i.d. sample of size n , we denote the ascending order statistics by $X_{1,n} \leq X_{2,n} \leq \dots \leq X_{n,n}$. The weak convergence of the centered and standardized maxima $X_{n,n}$ implies the existence of sequences of constants $a_n > 0$, $b_n > 0$ and a df $\Phi(x)$ such that:

$$\lim_{n \rightarrow \infty} P\left(\frac{X_{n,n} - a_n}{b_n} \leq x\right) = \Phi(x) \quad (2)$$

The work of Fisher and Tippett (1928, [5]), Gnedenko (1943, [6]) and de Haan (1970, [8]) answered the question on the possible limits and characterized the classes of distribution functions F having a certain limit in (2).

This convergence result is our main assumption. Up to location and scale, the possible limiting dfs $\Phi(x)$ in (2) are given by the so-called extreme value distributions G_γ , defined by:

$$G_\gamma(x) = \begin{cases} \exp\left(- (1 + \gamma x)^{-1/\gamma}\right), & \text{if } \gamma \neq 0 \\ \exp(-\exp(-x)), & \text{if } \gamma = 0 \end{cases} \quad (3)$$

Then it is well known [see Balkema and de Haan (1974, [1]) and Pickands (1975, [14])] that up to scale and location transformations with the shape parameter the generalized Pareto d.f. given by

$$H_\gamma(x) = 1 - (1 + \gamma x)^{-1/\gamma},$$

for all $x > 0$, if $\gamma \geq 0$ and $0 \leq x \leq -1/\gamma$ if $\gamma < 0$, and for $\gamma = 0$ read $H_\gamma(x)$ as

$$H_\gamma(x) = 1 - \exp(-x).$$

More precisely, it has been proved that there exists a normalizing function $\sigma(t) > 0$, such that for all x

$$\lim_{t \rightarrow \tau_F} \sup_{0 < x < \tau_F - t} |F_t(x) - H_\gamma(x/\sigma(t))| = 0, \quad (4)$$

if and only if F is in the maximum domain of attraction of the corresponding extreme value d.f. i.e $F \in D(G_\gamma)$ where the parameter $\gamma \in \mathbb{R}$ is the extreme value index and is the same in both G_γ .

In view of (4) with τ_F denote the right endpoint of the support of F we can expect that observations above some

high threshold are approximately generalized Pareto. This motivates that inferences on γ should be based on some set of high order statistics, say $(X_{n-k,n}, X_{n-k+1,n}, \dots, X_{n,n})$, or, equivalently, on

$$C_0 = X_{n-k,n}, C_1 = X_{n-k+1,n} - X_{n-k,n}, \dots, C_k = X_{n,n} - X_{n-k,n}, \quad (5)$$

where in the asymptotic setting $k = k_n$ is an intermediate sequence, that is, $k_n \rightarrow \infty$ and $k_n/n \rightarrow 0$ as $n \rightarrow \infty$. Hence, in view of convergence (4), the conditional distribution of the (C_0, C_1, \dots, C_k) given $C_0 = c_0$ can be approximated by the distribution of an ordered sample of k i.i.d. generalized Pareto random variables with d.f. $H_{GP}(x) : x \rightarrow H_\gamma(x/\sigma(t))$. This suggests to estimate the unknown parameters γ and σ by a maximum likelihood estimator in the approximating generalized Pareto model; that is, given the sample (X_1, X_2, \dots, X_n) denoting the number of absolute excesses over t by k for rather the largest observations $(X_{n-k,n}, \dots, X_{n,n})$, we maximize:

$$\ell(C_0, C_1, \dots, C_k) = \prod_{i=1}^k \ell(C_i) = \prod_{i=1}^k h_{GP}(C_i), \quad (6)$$

where $h_{GP}(x) = (1/\sigma)(1 + (\gamma/\sigma)x)^{-(1/\gamma)-1}$ with

$$C_i = X_{n-k+i,n} - X_{n-k,n}, \text{ for } 1 \leq i \leq k$$

Proposition 1: If the random variable X has a generalized Pareto distribution, then the conditional distribution of $X - t$ given $X \geq t$ is also generalized Pareto, with the same shape parameters γ .

The log-likelihood is given by

$$\log(\ell(C_i)) = \sum_{i=1}^k \left(\log\left(\frac{1}{\sigma}\right) - \left(1 + \frac{1}{\gamma}\right) \log\left(1 + \frac{\gamma}{\sigma} C_i\right) \right) \quad (7)$$

The range for σ is $\sigma > 0$ for $\gamma > 0$ and $\sigma > -\gamma C_{k:k}$ for $\gamma \leq 0$. If $\gamma < -1$, there is no maximum likelihood estimate since, for any $\gamma < -1$,

$$\lim_{\frac{\sigma}{\gamma} \rightarrow C_{k:k}^+} \log(\ell(C_i)) = +\infty$$

If $\gamma = -1$ thus done that $\sigma > C_{k:k}$ then

$$\log(\ell(C_i)) = \sum_{i=1}^k \log\left(\frac{1}{\sigma}\right)$$

Therefore, denoted $\hat{\gamma} = -1$, $\hat{\sigma} = C_{k:k}$. The problem is complicated by the optimization being taken over an open set, but it is treated as a maximum taken over a closed set.

The GPD maximum likelihood estimates, denoted by $(\hat{\gamma}; \hat{\sigma})$ is then given by the local maximum if

$\log(\ell(C_i; (\hat{\gamma}, \hat{\sigma}))) > -\sum_{i=1}^k \log(C_{k:k})$ and is given by the

boundary maximum if $\log(\ell(C_i; (\hat{\gamma}, \hat{\sigma}))) < -\sum_{i=1}^k \log(C_{k:k})$.

If no local maximum is found, then there is no GPD maximum likelihood estimate and the alternative estimators given by Hosking and Wallis (1987, [10]) are recommended.

To obtain a finite maximum of the GPD log-likelihood, the constraint $\gamma \geq -1$ must be imposed. Therefore, computing the GPD maximum likelihood estimates is an optimization on the constrained space $A = \{\sigma > 0, \gamma > 0\} \cup \{-1 \leq \gamma < 0, \sigma / \gamma < -C_{k:k}\}$.

There are two values of (γ, σ) that must be investigated to compute the GPD maximum likelihood estimate. The first is the local maximum of the log-likelihood on the space A . The second is at the boundary of A , where $\gamma = -1$. The likelihood equations from (7) are then given in terms of the partial derivatives have been studied in many articles including the important works of Hann and Ferreira [9]; p 91 and Drees et al [3]. The resulting likelihood equations in terms of the excesses $C_i = X_{n-k+i,n} - X_{n-k,n}$, are as follows:

$$\begin{cases} \left(\frac{1}{k} \sum_{i=1}^k \log\left(1 + \frac{\gamma}{\sigma} C_i\right) + 1 \right) \cdot \left(\frac{1}{k} \sum_{i=1}^k \left(\frac{1}{1 + \frac{\gamma}{\sigma} C_i} \right) \right) - 1 = 0 & \text{for } \gamma \neq 0 \\ \hat{\sigma} = (1/\bar{C}) & \text{for } \gamma = 0 \end{cases} \quad (8)$$

Therefore, to compute the maximum likelihood estimates, consider θ in terms of γ and σ by $\theta = -(\gamma/\sigma)$ which is one-to-one in the space A .

Then, For $C_i = X_{n-k+i,n} - X_{n-k,n}$, the maximum likelihood estimator can be approximate in the following procedure:

1. Find the root $\hat{\theta}$ of $\psi_k(\hat{\theta}) = 0$ where:

$$\psi_k(\theta) = \left(\frac{1}{k} \sum_{i=1}^k \log(1 - \theta C_i) + 1 \right) \cdot \left(\frac{1}{k} \sum_{i=1}^k (1 - \theta C_i)^{-1} \right) - 1, \quad (9)$$

2. Compute $\hat{\gamma}$ by

$$\hat{\gamma} = \frac{1}{k} \sum_{i=1}^k \log(1 - \theta C_i),$$

$$\hat{\theta} = -(\hat{\gamma} / \hat{\sigma}) \text{ then } \hat{\sigma} = -(\hat{\gamma} / \hat{\theta}).$$

Now, we go to find the solution of $\psi_k(\theta) = 0$ which is on (9) so must be computed numerically on the space $B = \{\theta < 1/C_{k:k}, \theta \neq 0\}$.

The following theorem states several properties of $\psi_k(\theta)$ that are useful in formulating an algorithm for determining her zeros

Theorem 1: Consider the function $\psi_k(\theta)$ given in (9) defined on the space B. Then

$$\begin{aligned} \lim_{\theta \rightarrow 1/C_{k:k}} \psi_k(\theta) &= -\infty \\ \lim_{\theta \rightarrow 0} \psi_k(\theta) &= 0 \end{aligned} \quad (10)$$

$$\psi_k(\theta) < 0 \text{ for all } \theta_L = \frac{2(C_1 - \bar{C})}{(C_1)^2} \quad (11)$$

where $\bar{C} = \frac{1}{k} \sum_{i=1}^k C_i$ and $C_1 < C_2 < \dots < C_k$ the order statistics correspondence.

With the second term in formula (8) we have $\gamma \rightarrow 0$ implied $\theta \rightarrow 0$ with $\sigma = (1/\bar{C})$; as similar result we can get that $\gamma = 10^{-p}$ [Note that p is a natural number] as result we take an $\varepsilon = 10^{-p} / \bar{C}$ around the zero.

Since the upper bound $1/C_{k:k} \notin B$. The algorithm will use $\theta_U - \varepsilon$ and $\theta_L - \varepsilon$ for some $\varepsilon > 0$ as the upper bound and lower bound respectively. Then the space B will be define by $B = \{\theta \in [\theta_L, -\varepsilon] \cup [\varepsilon, \theta_U], \theta \neq 0\}$ where $\theta_U = 1/C_{k:k} - \varepsilon$ and $\varepsilon = 10^{-p} / \bar{C}$ with p is a natural number, required for modified Bisection algorithm to search the multi-roots of $\psi_k(\theta)$; which we present in section (2) which rely with the numerical method that proposed by Tanakan, (2013, [13]), and best of them as well. For this in section (3) we use a two real data utilized the first collected from the exceedance of the threshold given by Grimshaw [7], and the second we use model extreme value by the real Danish fire data.

2. Modified Bisection Algorithms

A modified bisection algorithms is much more efficient than the bisection method. Furthermore, it is faster than the Newton method, and don not count the derivative of a function at the reference point, which is not always easy. In the practice, the initial solution is really important for the Newton method. But some initial solutions can make the method Newton diverges. Hence, by the intermediate value theorem a modified bisection algorithms, it works when f is a continuous function on $[\alpha; \beta]$ where $(\alpha, \beta) \in \mathbb{R}^2$ with the initial condition $f(\alpha).f(\beta) < 0$ existing with each iteration. A modified bisection algorithms can reduced the number of iterations which less than the iteration number of the bisection method and nearby to the iteration number of Newton method, for numerical results see Tanakan ([13], section [3]).

In this work took the error are less than the tolerance, which is linked to condition $|\theta_{n+1} - \theta_n|$ less than the tolerance because we computed numerically on the space B.

We can state an alternative modified bisection algorithms

for multi-roots as the following,

1. Step 1: Divide the interval into several intervals by $h = (b - a) / t$ where $t \in \mathbb{N}$ is the number of intervals for example $t = 10$.
2. Step 2: For $i = 1, \dots, 10$, and tolerance 1×10^{-7} ; and $\alpha_1 = \alpha_i, \beta_1 = \beta_{i+1}$ where $f(\alpha_i) \cdot f(\beta_{i+1}) < 0$.
3. Step 3: For $n \geq 1$, compute $\theta_n^* = (\alpha_n + \beta_n) / 2$.
4. Step 4: Compute for a sub-interval $[\alpha_n^*, \beta_n^*]$ by

$$[\alpha_n^*, \beta_n^*] = \begin{cases} [\alpha_n, \theta_n^*] & \text{if } f(\alpha_n) \cdot f(\theta_n^*) < 0 \\ [\theta_n^*, \beta_n] & \text{if } f(\theta_n^*) \cdot f(\beta_n) < 0 \end{cases}$$

5. Step 5: Compute,

$$\theta_{n+1} = \beta_n^* - f(\beta_n^*) \frac{\beta_n^* - \alpha_n^*}{f(\beta_n^*) - f(\alpha_n^*)} \text{ or } \theta_{n+1} = \alpha_n^* - f(\alpha_n^*) \frac{\beta_n^* - \alpha_n^*}{f(\beta_n^*) - f(\alpha_n^*)}$$

6. Step 6: If $|\theta_{n+1} - \theta_n^*| < 1 \times 10^{-7}$, then stop program (i.e. $\theta_{n+1} = \theta_n^*$) the zero is θ_{n+1} . ELSE

$$[\alpha_{n+1}^*, \beta_{n+1}^*] = \begin{cases} [\alpha_n^*, \theta_{n+1}] & \text{if } f(\alpha_n^*) \cdot f(\theta_{n+1}) < 0 \\ [\theta_{n+1}, \beta_n^*] & \text{if } f(\theta_{n+1}) \cdot f(\beta_n^*) < 0 \end{cases}$$

and set $n = n + 1$, GOTO Step 3.

To alert, this the Modified Bisection Algorithms is based on two method: Bisection method and Modified Bisection method (MB) or method of false position (Regula false). We noticed that the algorithm at the top, we find that we calculate θ_n^* by Bisection method and θ_{n+1} by false position method.

Theorem 2: Let f be a continuous function and defined on $[\alpha_i; \beta_{i+1}] \subset [\alpha; \beta]$ with the initial condition $f(\alpha_i) \cdot f(\beta_{i+1}) < 0$ existing with each iteration. The modified bisection algorithm generates a sequence $\{\theta_{n+1}\}_{n=1}^\infty$

and $\{\theta_n^*\}_{n=1}^\infty$ approximating a zero $\theta^{(0)}$ of f with

$$|\theta_{n+1} - \theta_n^*| < \frac{5(\beta_{i+1} - \alpha_i)}{2^{n+2}} \text{ for } n \geq 1 \tag{12}$$

where $i = 1, \dots, t$ and t is the number sub-intervals of $[\alpha; \beta]$.

Proof: We have the theoretical result for Bisection method that

$$|\theta_n^* - \theta^{(0)}| < \frac{(\beta_{i+1} - \alpha_i)}{2^n} \text{ for } n \geq 1 \tag{13}$$

and for Modified bisection agrees Tanakan [13] we have also

$$|\theta_{n+1} - \theta^{(0)}| < \frac{(\beta_{i+1} - \alpha_i)}{2^{n+2}} \text{ for } n \geq 1 \tag{14}$$

We can show that: $|\theta_{n+1} - \theta_n^*| = |\theta_{n+1} - \theta^{(0)} - (\theta_n^* - \theta^{(0)})|$.

And it's easy to prove the theorem (2) by using a mathematical induction.

The result (13) and result (14) proved that the sequence $\{\theta_n^*\}_{n=1}^\infty$ converges to $\theta^{(0)}$ with rate of convergence $O\left(\frac{1}{2^n}\right)$

that is, $\theta_n^* = \theta^{(0)} + O\left(\frac{1}{2^n}\right)$ and the sequence $\{\theta_{n+1}\}_{n=1}^\infty$

converges to $\theta^{(0)}$ with rate of convergence $O\left(\frac{1}{2^{n+2}}\right)$ that

is, $\theta_{n+1} = \theta^{(0)} + O\left(\frac{1}{2^{n+2}}\right)$ respectively.

An Algorithm for the GPD Maximum Likelihood Estimates

In the data sets used investigating the GPD maximum likelihood estimates, it appears that there exists either no zero or two zeros on each interval. For this we preferred an algorithm that computes the GPD maximum likelihood estimates for search the zero of the function given on (9) by modified bisection algorithm. The algorithm that computes the GPD maximum likelihood estimates is given by the following:

1. Choose an ε , for example let $\varepsilon = 10^{-8} / \bar{C}$.
2. Compute the lower and upper bounds for zeros of $\psi_k(\theta)$ given by

$$\theta_L = \frac{2(C_1 - \bar{C})}{(C_1)^2} \text{ and } \theta_U = \frac{1}{C_{k:k}} - \varepsilon$$

3. To determine the zero of $\psi_k(\theta)$ on the bounded intervals $[\theta_L; -\varepsilon]$ and $[\varepsilon; \theta_U]$, use the modified bisection method algorithm given in section (2) with each interval.

4. For each $\theta_j^{(0)}$, compute γ_j given by

$$\gamma_j = \frac{1}{k} \sum_{i=1}^k \log(1 - \theta_j^{(0)} C_i),$$

5. compute $\sigma_j = -\left(\gamma_j / \theta_j^{(0)}\right)$.

6. Let γ_j denote the result of algorithm included on space A ; the local maximum log-likelihood parameter estimation of the GPD parameters is (γ_j, σ_j) which have a maximum $\log(\ell(C_i; (\gamma_j, \sigma_j)))$. The algorithm is over.

The maximum likelihood estimator (MLE's) are taken to

be the values $(\hat{\gamma}, \hat{\sigma})$ which yield a local maximum. The Fisher information matrix gives the asymptotic variance-covariance of the maximum likelihood estimator, which can be calculated by

$$\begin{bmatrix} \text{var}(\hat{\gamma}) & \text{cov}(\hat{\gamma}, \hat{\sigma}) \\ \text{cov}(\hat{\gamma}, \hat{\sigma}) & \text{var}(\hat{\sigma}) \end{bmatrix} = \begin{bmatrix} \frac{\partial^2 \log(\ell(C_i))}{\partial^2 \hat{\gamma}} & \frac{\partial^2 \log(\ell(C_i))}{\partial \hat{\gamma} \partial \hat{\sigma}} \\ \frac{\partial^2 \log(\ell(C_i))}{\partial \hat{\gamma} \partial \hat{\sigma}} & \frac{\partial^2 \log(\ell(C_i))}{\partial^2 \hat{\sigma}} \end{bmatrix}^{-1}$$

Then we conclude $100(1-\kappa) \%$ confidence intervals constructing by using

$$\hat{\gamma} \pm t_{\kappa/2} \sqrt{\text{var}(\hat{\gamma})} \text{ and } \hat{\sigma} \pm t_{\kappa/2} \sqrt{\text{var}(\hat{\sigma})}$$

where $t_{\kappa/2}$ denotes the $1-\kappa/2$ quantile of the standard normal distribution is symmetric about 0.

3. Numerical Examples

In the first example, we use the tensile-strength fiber data presented by Grimshaw (1993, [7]). In his study, the GPD maximum likelihood estimates are $\hat{\sigma} = 0.283040$ and $\hat{\gamma} = 0.117698$.

The data are collected from the exceedance of the threshold in the tensile-strength testing for a random value

Table 2. An account the values (γ_3, σ_3) and (γ_4, σ_4) .

$\theta_3^{(0)} = 0.4158348$	$\gamma_3 = -0.1176979$	$\sigma_3 = 0.2830401$	$\log(\ell(C_i; (\gamma_3, \sigma_3))) = 5.6979682$
$\theta_4^{(0)} = 1.141497$	$\gamma_4 = -0.9543413$	$\sigma_4 = 0.8360436$	$\log(\ell(C_i; (\gamma_4, \sigma_4))) = 2.0012366$

Next, we have $\sigma_3 / \gamma_3 < -0.876 \in A$ and $\sigma_4 / \gamma_4 < -0.876 \in A$ are the local maximum of the GPD log-likelihood on A for $C_{1515} = 0.876$. The boundary maximum, $(\hat{\gamma} = -1, \hat{\sigma} = 0.876)$ given by $\log(\ell(C_i; (-1, 0.876))) = 1.9858378$.

Since $\log(\ell(C_i; (\gamma_3, \sigma_3))) > \log(\ell(C_i; (\gamma_4, \sigma_4)))$ the GPD maximum likelihood estimates for the data $(C_1, C_2, \dots, C_{15})$ of 15 nylon carpet fibers, is $(\hat{\gamma} = -0.1176979, \hat{\sigma} = 0.2830401)$.

The asymptotic variance-covariance matrix of the parameters (γ, σ) is given by the inverse of the Fisher information matrix. Then, their corresponding 95% confidence intervals is $-0.7257669 < \gamma < 0.4903711$ and $0.05976156 < \sigma < 0.50631869$.

For a real second example application we use model extreme value. Let us consider the real Danish fire data, these data describe large fire insurance claims in Denmark from Thursday 3rd January 1980 until Monday 31st December 1990. A numeric vector containing 2167 observations. They offer many possibilities for modeling and have been used by

$(C_1, C_2, \dots, C_{15})$ of 15 nylon carpet fibers, listed in an increasing order (only observation is that we have not given us the value of threshold for the test is unknown for proprietary reasons):

Table 1. Exceedance (in kg/mm²) of the Testing Threshold in Tensile-Strength Tests for a Random Sample of $n = 15$ Nylon Carpet Fibers.

0.011	0.030	0.051	0.056	0.092
0.100	0.140	0.184	0.200	0.286
0.338	0.365	0.518	0.561	0.876

To begin the search for zeros of $\psi_k(\theta)$ compute the bounds given by $\theta_L = -3994.49$, $\theta_U = 1.141552$ and $\varepsilon = 3.957784 \times 10^{-8}$. The algorithm will search for two zeros on $[\theta_L; -\varepsilon]$ and $[\varepsilon; \theta_U]$. We begin with the interval $[\theta_L; -\varepsilon]$, by the modified Bisection algorithm we find that no zero of $\psi_k(\theta)$ exists on this interval. i.e $\theta_1^{(0)}$ and $\theta_2^{(0)}$ do not exist. We go to the interval $[\varepsilon; \theta_U]$. By the modified Bisection algorithm as we set in section (2), we will find that he converged to two roots are $\theta_3^{(0)} = 0.4158348$ and $\theta_4^{(0)} = 1.141497$. So we finished the work with the algorithm. And we calculate the values correspond to $\theta_3^{(0)}$ and $\theta_4^{(0)}$.

many researchers to illustrate their methods, see McNeil (1997) and Resnick (1997). These data can be found in evir package of the R software (Ihaka and Gentleman cite R).

In this study, we are concerned with confidence bounds for the 73 monthly maximum losses during the mentioned from the given 2167 observations, listed in an increasing order.

Table 3. Monthly maximum losses during the mentioned from Danish data for a random sample of $n = 73$ fire insurance claims.

4.450262	10.011123	14.394581	20.826733	32.467532
4.5	10.072303	14.678899	20.863675	34.141547
4.625	10.5	15.284653	20.969856	38.154392
4.657070	10.820452	16.441659	22.258226	42.091448
4.821601	10.998350	17.068467	23.283859	46.5
4.867987	11.123471	17.569546	24.970273	47.019521
5.469679	11.595547	17.739274	25.288376	50.065531
5.6	11.890606	17.743491	25.953860	56.225426
6.773920	12.376238	18.322083	26.214641	57.410636
6.798457	12.523191	18.424135	27.338066	56.707491
7.425743	12.631813	18.646484	27.829314	152.413209
7.613470	13.348165	19.070278	29.026037	144.657591
7.643979	14.013548	19.162304	29.037106	263.250366
7.863696	14.239000	19.4	31.055901	
8.678881	14.293194	20.049941	32.387807	

To begin the search for zeros of $\psi_k(\theta)$ we compute the bounds $(\varepsilon, \theta_L, \theta_U)$ for $k = 50$.

Then $\varepsilon = 2.817172 \times 10^{-9}$ and other bounds given by $\theta_L = -3.018889 \times 10^{-1}$, and $\theta_U = 3.798662 \times 10^{-3}$. The algorithm will search for two zeros on $[\theta_L; -\varepsilon]$ and $[\varepsilon; \theta_U]$ for $\psi_k(\theta)$.

Using our program yields we found The GPD maximum likelihood estimates for over threshold $k > 50$ monthly maximum losses during the mentioned from the given 73 observations for the Danish data is $\hat{\gamma} = 0.08586831$, $\hat{\sigma} = 32.31397400$.

Then, their corresponding 95 % confidence intervals is $-0.1249036 < \gamma < 0.2966402$ and $21.08807 < \sigma < 43.53988$.

4. Conclusion

Since, the value of $\gamma \in \square$ the shape parameter or extreme value index though it is often referred to as "tail behavior"; in model of extreme values distribution is dominant which are the three standard family see Haan and Ferreira [9] (either Gumbel ($\gamma = 0$), Fréchet ($\gamma > 0$), or Weibull ($\gamma < 0$)). by determining tail behavior. In this paper we introduce a good numerical method with less condition and we avoided many numerical and the negatives of Newton method, which is not always easy in the practice, and we have reduced time and effort, for maximum likelihood estimate without signs γ the shape parameter.

Acknowledgements

Special thanks are due to Dr, Fateh Ben Atia and Dr Djabrane Yahia for his valuable help and the important remarks.

Appendix

The second derivatives of the log-likelihood are computations for $\theta = -(\gamma/\sigma)$ by:

$$\begin{cases} \frac{\partial^2 \log(\ell(C_i))}{\partial^2 \gamma} = \left(\frac{1}{\gamma}\right)^3 \sum_{i=1}^k \left[2 \left(1 - \frac{1}{1-\theta C_i}\right) + (1+\gamma) \left(1 - \frac{1}{1-\theta C_i}\right)^2 - 2 \log(1-\theta C_i) \right] \\ \frac{\partial^2 \log(\ell(C_i))}{\partial^2 \sigma} = \left(\frac{\theta}{\gamma}\right)^2 \sum_{i=1}^k \left[1 - \left(1 + \frac{1}{\gamma}\right) \cdot \left(2 \left(1 - \frac{1}{1-\theta C_i}\right) - \left(1 - \frac{1}{1-\theta C_i}\right)^2\right) \right] \\ \frac{\partial^2 \log(\ell(C_i))}{\partial \gamma \partial \sigma} = \frac{\theta}{\gamma^2} \sum_{i=1}^k \left[\left(1 - \frac{1}{1-\theta C_i}\right) - \left(1 + \frac{1}{\gamma}\right) \cdot \left(1 - \frac{1}{1-\theta C_i}\right)^2 \right] \end{cases}$$

References

- [1] Balkema, August A., and Laurens De Haan., 1974. "Residual life time at great age, The Annals of probability,". 792-804.
- [2] Davison, Anthony C., 1984. "Behavior Modelling excesses over high thresholds, with an application. Statistical extremes and applications,". Springer Netherlands, 461-482.
- [3] Drees, Holger, Ana Ferreira, and Laurens De Haan., 2004. "On maximum likelihood estimation of the extreme value index," Annals of Applied Probability, 1179-1201.
- [4] Embrechts, P., and C. Klüppelberg., T. Mikosch., 1997. "Gas Modelling Extremal Events for Insurance and Finance, Applications of Mathematics-Stochastic Modelling and Applied Probability,". Springer New York, No. 33.
- [5] Fisher, Ronald Aylmer, and Leonard Henry Caleb Tippett., 1928. "Limiting forms of the frequency distribution of the largest or smallest member of a sample," Mathematical Proceedings of the Cambridge Philosophical Society, Cambridge University Press, Vol. 24. No. 02.
- [6] Fachinger, Gnedenko, Boris., 1943. "Sur la distribution limite du terme maximum d'une serie aleatoire Annals of mathematics,". 423-453.
- [7] Grimshaw, Scott D., 1993. "Computing maximum likelihood estimates for the generalized Pareto distribution,". Technometrics 35.2, 185-191.
- [8] Haan, L. DE., 1970. "On regular variation and its application to the weak convergence of sample extremes,". Mathematical Centre Tracts 32.
- [9] Haan, L de. and Ferreira, A., 2006. "Extreme Value Theory: An Introduction," Springer.
- [10] Hosking, Jonathan RM, and James R. Wallis., 1987. "Parameter and quantile estimation for the generalized Pareto distribution," Technometrics 29.3, 339-349.
- [11] Reiss, Rolf-Dieter, Michael Thomas, and R. D. Reiss., 2007. "Statistical analysis of extreme values," Basel, Birkhäuser, Vol. 2.
- [12] Smith, Richard L., 1985. "Computing Maximum likelihood estimation in a class of nonregular cases," Biometrika 72.1, 67-90.
- [13] Tanakan, S., 2013. "A New Algorithm of Modified Bisection Method for Nonlinear Equation," Applied Mathematical Sciences 7.123, 6107-6114.
- [14] Pickands III, James., 1975. "A Statistical inference using extreme order statistics," the Annals of Statistics 119-131.