REVSTAT – Statistical Journal
Volume 21, Number 3, July 2023, Supplementary Material, 1–14

Supplementary Material for "A Computational Approach to Confidence Intervals and Testing for Generalized Pareto Index Using the Greenwood Statistic"

Authors: Marek Arendarczyk
– Mathematical Institute, University of Wroclaw, Poland
marendar@math.uni.wroc.pl

Tomasz J. Kozubowski
– Department of Mathematics and Statistics, University of Nevada Reno, USA
tkozubow@unr.edu

Anna K. Panorska ✉
– Department of Mathematics and Statistics, University of Nevada Reno, USA
ania@unr.edu

Received: April 2020 Revised: January 2022 Accepted: January 2022

This supplement to the main paper includes:

1. Appendix T: Table 1 with power of our new test for several sample sizes and values of α and comparison of power for tests based on T_n and R_n in Tables 2–5.
2. Appendix A: A review of selected key works on estimation for the GPD family, particularly maximum likelihood procedure, for which there are numerous errors and inaccuracies in the literature.
3. References for Appendix A.

✉ Corresponding author.

1. Appendix T

Table 1: The table contains estimates of power for the test of exponential vs. Pareto distribution. Values of α are in the first column and sample sizes are in the first row of the table. The entries are power values for the corresponding combinations of α and sample size.

alpha=w	**n=20**	**n=50**	**n=100**	**n=250**	**n=500**	**n=1000**	**n=5000**	**n=10000**
2	0.9842	1	1	1	1	1	1	1
1	0.8546	0.9923	0.9997	1	1	1	1	1
0.833333	0.7867	0.9796	0.9996	1	1	1	1	1
0.666667	0.674	0.9362	0.9969	1	1	1	1	1
0.555556	0.5924	0.8839	0.9865	1	1	1	1	1
0.5	0.532	0.8366	0.9759	0.9999	1	1	1	1
0.333333	0.3644	0.6424	0.8645	0.9929	0.9999	1	1	1
0.25	0.2726	0.4845	0.7093	0.9577	0.9984	1	1	1
0.2	0.216	0.3808	0.5815	0.8723	0.9872	0.9998	1	1
0.166667	0.1889	0.3141	0.4797	0.7781	0.954	0.9988	1	1
0.142857	0.1584	0.2591	0.3994	0.693	0.8974	0.9896	1	1
0.125	0.138	0.2332	0.344	0.6131	0.8311	0.9748	1	1
0.111111	0.1328	0.2097	0.3096	0.5342	0.7586	0.9443	1	1
0.1	0.12	0.1893	0.2738	0.476	0.6903	0.9064	1	1
0.090909	0.1161	0.1708	0.2428	0.4095	0.6255	0.8601	1	1
0.083333	0.1046	0.1547	0.2278	0.3778	0.5822	0.8119	0.9999	1
0.076923	0.1032	0.1485	0.2074	0.3462	0.5354	0.7625	0.9993	1
0.071429	0.0933	0.1435	0.1943	0.3121	0.4822	0.7107	0.999	1
0.066667	0.0958	0.1276	0.1793	0.2869	0.4524	0.6643	0.9983	1

Table 2: Power estimates for testing $H_0\colon \alpha \leq 0$ vs. $H_1\colon \alpha > 0$ based on the test statistics T_n and R_n. Values of α are in the first column and the sample sizes n are in the first row of the table. The entries are approximate power values for various combinations of α and n.

$H_0 : \alpha \leq 0$ **vs.** $H_1 : \alpha > 0$										
Greenwood T_n						R_n				
	n					**n**				
α	**5**	**10**	**20**	**50**	**100**	**5**	**10**	**20**	**50**	**100**
0.5	0.195	0.34	0.545	0.843	0.975	0.199	0.302	0.472	0.707	0.877
1	0.351	0.613	0.859	0.994	1	0.344	0.554	0.777	0.953	0.994
1.5	0.467	0.781	0.968	1	1	0.448	0.715	0.907	0.996	1
2	0.58	0.868	0.983	1	1	0.557	0.799	0.948	1	1
2.5	0.687	0.905	0.995	1	1	0.655	0.862	0.981	1	1
3	0.735	0.948	0.996	1	1	0.702	0.912	0.983	1	1
3.5	0.761	0.972	0.997	1	1	0.727	0.932	0.99	1	1
4	0.809	0.976	1	1	1	0.776	0.944	0.994	1	1
4.5	0.84	0.976	1	1	1	0.804	0.945	0.998	1	1
5	0.85	0.988	1	1	1	0.818	0.965	1	1	1

Table 3: Power estimates for testing $H_0\colon \alpha \geq 0$ vs. $H_1\colon \alpha < 0$ based on the test statistics T_n and R_n. Values of α are in the first column and the sample sizes n are in the first row of the table. The entries are approximate power values for various combinations of α and n.

$H_0 : \alpha \geq 0$ **vs.** $H_1 : \alpha < 0$										
Greenwood T_n						R_n				
	n					**n**				
α	**5**	**10**	**20**	**50**	**100**	**5**	**10**	**20**	**50**	**100**
-0.5	0.108	0.232	0.465	0.886	0.998	0.112	0.234	0.478	0.935	1
-1	0.241	0.535	0.847	0.999	1	0.271	0.606	0.952	1	1
-1.5	0.35	0.72	0.971	1	1	0.437	0.854	0.998	1	1
-2	0.475	0.839	0.988	1	1	0.592	0.931	1	1	1
-2.5	0.535	0.895	0.999	1	1	0.689	0.975	1	1	1
-3	0.611	0.94	1	1	1	0.757	0.993	1	1	1
-3.5	0.692	0.963	1	1	1	0.841	0.996	1	1	1
-4	0.701	0.969	1	1	1	0.859	1	1	1	1
-4.5	0.735	0.971	1	1	1	0.889	1	1	1	1
-5	0.749	0.985	1	1	1	0.92	1	1	1	1

Table 4: Power estimates for testing $H_0\colon \alpha \le 1$ vs. $H_1\colon \alpha > 1$ based on the test statistics T_n and R_n. Values of α are in the first column and the sample sizes n are in the first row of the table. The entries are approximate power values for various combinations of α and n.

$H_0 : \alpha \le 1$ **vs.** $H_1 : \alpha > 1$										
Greenwood T_n						R_n				
	n					**n**				
α	**5**	**10**	**20**	**50**	**100**	**5**	**10**	**20**	**50**	**100**
1.5	0.134	0.13	0.157	0.186	0.2	0.129	0.132	0.157	0.184	0.194
2	0.219	0.244	0.295	0.331	0.317	0.211	0.242	0.287	0.33	0.32
2.5	0.276	0.326	0.35	0.399	0.46	0.276	0.322	0.347	0.4	0.455
3	0.37	0.393	0.462	0.5	0.531	0.367	0.381	0.462	0.494	0.521
3.5	0.417	0.433	0.53	0.553	0.613	0.414	0.436	0.52	0.541	0.602
4	0.466	0.51	0.593	0.637	0.663	0.465	0.506	0.588	0.63	0.64
4.5	0.509	0.522	0.612	0.643	0.668	0.509	0.53	0.602	0.63	0.65
5	0.548	0.606	0.646	0.695	0.748	0.548	0.6	0.645	0.683	0.727

Table 5: Power estimates for testing $H_0\colon \alpha \ge -1$ vs. $H_1\colon \alpha < -1$ based on the test statistics T_n and R_n. Values of α are in the first column and the sample sizes n are in the first row of the table. The entries are approximate power values for various combinations of α and n.

$H_0 : \alpha \ge -1$ **vs.** $H_1 : \alpha < -1$										
Greenwood T_n						R_n				
	n					**n**				
α	**5**	**10**	**20**	**50**	**100**	**5**	**10**	**20**	**50**	**100**
−1.5	0.115	0.152	0.228	0.38	0.625	0.142	0.224	0.393	0.716	0.95
−2	0.178	0.27	0.417	0.722	0.926	0.227	0.433	0.727	0.984	1
−2.5	0.2	0.357	0.594	0.906	0.994	0.304	0.615	0.92	0.999	1
−3	0.304	0.438	0.719	0.976	1	0.422	0.743	0.965	1	1
−3.5	0.378	0.513	0.814	0.987	1	0.518	0.821	0.991	1	1
−4	0.437	0.575	0.87	0.995	1	0.602	0.884	0.995	1	1
−4.5	0.461	0.691	0.898	1	1	0.663	0.926	0.999	1	1
−5	0.472	0.701	0.925	1	1	0.684	0.959	0.999	1	1

2. Appendix A: Estimation of Lomax/GPD parameters and the Greenwood statistic T_n

The purpose of this review section is to note and discuss various challenges and errors in the literature about estimation of parameters for GPD, and illustrate connections with the statistic T_n. There is a substantial literature on this topic, with several excellent reviews, including de Zea Bermudez and Kotz (2010ab). We also note that there are two alternative parameterizations of the GPD family in the literature, which differ in the sign of the main shape parameter. We review them in two subsections below, following a relatively straightforward special case of Pareto II (Lomax) distribution.

2.1. Parameter estimation: the Lomax case

Here we show how the Greenwood statistic T_n comes up in maximum likelihood estimation of the tail index α based on the random sample $X_1, ..., X_n$ from Pareto II (Lomax) distribution given by the survival function (SF)

$$\mathbb{P}(X_i > x) = (1 + \alpha\beta x)^{-1/\alpha}, \;\; x \in \mathbb{R}_+, \tag{2.1}$$

with $\alpha \geq 0$. To maximize the the log-likelihood function

$$l(\alpha, \beta) = n \log \beta - \left(\frac{1}{\alpha} + 1\right) \sum_{i=1}^{n} \log(1 + \alpha\beta x_i) \tag{2.2}$$

with respect to $(\alpha, \beta) \in \Theta$, where $\Theta = \{(\alpha, \beta) : \alpha \geq 0, \beta > 0\}$ is the parameter space, we first show that there exists a unique $\beta(\alpha)$ that maximizes $l(\alpha, \beta)$ with respect to $\beta > 0$ for each fixed $\alpha \geq 0$. To this end, we note that the condition

$$\frac{\partial}{\partial \beta} l(\alpha, \beta) > 0 \tag{2.3}$$

is equivalent to $h(\beta) < 1$, where

$$h(\beta) = \frac{1 + \alpha}{n} \sum_{i=1}^{n} \frac{\beta x_i}{1 + \alpha\beta x_i}.$$

Since the function $h(\cdot)$ is continuous and monotonically increasing on the interval $(0, \infty)$ with

$$\lim_{\beta \to 0^+} h(\beta) = 0 \;\; \text{and} \;\; \lim_{\beta \to \infty} h(\beta) = 1 + \frac{1}{\alpha} > 1,$$

there exists unique $\beta(\alpha)$ so that condition (2.3) is fulfilled if and only if $\beta < \beta(\alpha)$. This shows that $\beta(\alpha)$ is a unique value that maximizes the function $l(\alpha, \beta)$ with respect to $\beta > 0$, as desired. The following result provides crucial properties of the function

$$v(\alpha) = l(\alpha, \beta(\alpha))/n, \;\; \alpha \geq 0, \tag{2.4}$$

which needs to be maximized in order to find the MLE of the tail index α.

Lemma 2.1. *In the above setting, we have the following:*

(i) $$v(0) = -1 - \log(\bar{X}), \quad \lim_{\alpha\to\infty} v(\alpha) = -\infty.$$

(ii) *If*

$$\frac{n}{2}\frac{\sum_{i=1}^n X_i^2}{\left(\sum_{i=1}^n X_i\right)^2} > 1 \tag{2.5}$$

then $v(\alpha)$ is increasing in a (right) neighborhood of $\alpha = 0$, and if

$$\frac{n}{2}\frac{\sum_{i=1}^n X_i^2}{\left(\sum_{i=1}^n X_i\right)^2} < 1 \tag{2.6}$$

then $v(\alpha)$ is decreasing in a (right) neighborhood of $\alpha = 0$.

Proof: The result follows from Proposition 3.2 in Arendarczyk et al. (2018), where similar problem is discussed in a more general setting. Indeed, we note that the log-likelihood function (2.2) is the same as $ng(\alpha, \beta)$, where g is the function defined by (3.4) in Arendarczyk et al. (2018) with $N_i = 1$ for $i = 1, ...m$ and $m = n$. With these choices of m and N_i, the function $v(\alpha)$ defined by (3.7) in Arendarczyk et al. (2018) is exactly the same as our $v(\alpha)$ in (2.4). Thus, Part (i) of Lemma 2.1 follows directly from Parts (i) and (iii) of Proposition 3.2 of Arendarczyk et al. (2018) while Part (ii) follows from Part (ii) of Proposition 3.2 of Arendarczyk et al. (2018), since the relations (3.8) and (3.9) of that proposition simplify to (2.5) and (2.6), respectively. □

The above result shows that the MLE of α is closely related to the value of the statistic T_n. Positive values of $nT_n/2 - 1$ (so that condition (2.5) is fulfilled) point towards a strictly positive estimate $\hat{\alpha}$, whose computation requires a numerical search. On the other hand, for negative values of $nT_n/2 - 1$, when we have (2.6), the function $v(\cdot)$ may have a maximum value at its boundary, so that $\hat{\alpha} = 0$ corresponding to exponentiality (in which case we also have $\hat{\beta} = 1/\bar{X}$). In particular, it follows that the MLEs of α and β always exist and are relatively straightforward to obtain (cf. Castillo and Daoudi, 2009; Kozubowski et al., 2009).

2.2. Parameter estimation: the GPD case

In contrast with the Lomax case, parameter estimation via maximum likelihood provides a considerable challenge in the extended GPD family, with some inconsistencies in the literature regarding the existence of the estimates and convergence of the numerical algorithms that find them. We discuss the main issues below, following a brief account of the method of moments, where the statistic T_n plays an important role. We start with the two parameterizations of the GPD family that are often found in the literature. One version (see, e.g., Chaouche and Bacro, 2006; Castillo and Daoudi, 2009) is essentially the same as (2.1) with the reciprocal of the scale parameter. Here, the PDF of the GPD model is of the form

$$f(x) = \frac{1}{\beta}\left(1 + \frac{\xi x}{\beta}\right)^{-1/\xi - 1} \tag{2.7}$$

where the domain is the set $(0, \infty)$ for $\xi \geq 0$ (Pareto-exponential case) and $(0, -\beta/\xi)$ when $\xi < 0$ (light-tail case). Alternatively, some authors (see, e.g., Grimshaw, 1993; Choulakian and Stephens, 2001; Castillo and Serra, 2015) use the following form of the PDF:

$$f(x) = \frac{1}{\sigma}\left(1 - \frac{\kappa x}{\sigma}\right)^{1/\kappa - 1}, \tag{2.8}$$

with the domain of $(0, \infty)$ for $\kappa \leq 0$ (Pareto-exponential case) and $(0, \sigma/\kappa)$ for $\kappa > 0$ (light-tail case). While the two parameterizations are clearly equivalent (with the corresponding parameters connected via $\xi = -\kappa$ and $\beta = \sigma$), the domains are often not properly accounted for in the literature.

2.2.1. GPD: the method of moments

Estimation of the two parameters via the method of moments is relatively straightforward. As is well-known, the mean and the variance of the GPD random variable X in the (2.8) parametrization are

$$\mathbb{E}(X) = \frac{\sigma}{1+\kappa} \ (\kappa > -1) \ \text{ and } \ \mathbb{V}ar(X) = \frac{\sigma^2}{(1+\kappa)^2(1+2\kappa)} \ (\kappa > -1/2), \tag{2.9}$$

and lead directly to the method of moment estimators of the form (see, e.g., de Zea Bermudez and Kotz, 2010a),

$$\hat{\kappa} = \frac{1}{2}\left(\frac{\overline{X}^2}{S^2} - 1\right) \ \text{ and } \ \hat{\sigma} = \frac{1}{2}\overline{X}\left(\frac{\overline{X}^2}{S^2} + 1\right) \ (\kappa > -1/2), \tag{2.10}$$

where the quantities $\overline{X}$ and S^2 are the sample mean and variance, respectively. Thus, whenever $\overline{X}^2/S^2 - 1 < 0$ the estimator of κ is negative, pointing towards the heavy-tail Pareto case. Since this is equivalent to $T_n > 2/n$, the Greenwood statistic emerges naturally in the method of moments estimation of the GPD shape parameter.

2.2.2. GPD: maximum likelihood

In contrast with the method of moments, maximum likelihood estimation for the GPD family is a challenging problem, both theoretically and computationally. While there is a substantial literature on this topic, so far the existence and uniqueness of the solution to the likelihood equations has not been established, and there are some disagreements regarding the convergence of the algorithms (see, e.g., Hosking and Wallis, 1987, Ashkar and Tatsambon, 2007). Additionally, some analyses of this problem in the literature are not quite correct. Our brief (and selective!) journey through this mathematically interesting and practically relevant problem will be limited to several highly cited key papers, presenting different perspectives on this problem.

We start with the log-likelihood function for a random sample $X_1, ..., X_n$ from the the GPD model given by (2.8), which is given by

$$l(\kappa, \sigma) = -n \log \sigma + \left(\frac{1}{\kappa} - 1\right)\sum_{i=1}^{n} \log(1 - \kappa X_i/\sigma), \tag{2.11}$$

where the parameters are restricted as follows: $-\infty < \kappa < \infty$ and $\sigma > \max(0, \kappa X_{(n)})$.

In particular, we have $\sigma > 0$ for $\kappa < 0$ (Pareto case) and for $\kappa = 0$ (exponential case). Moreover, by a continuous extension, in the exponential case the log-likelihood reduces to that for an exponential sample:

$$l(\kappa,\sigma) = -n\log\sigma - \frac{1}{\sigma}\sum_{i=1}^{n} X_i, \quad \kappa = 0. \tag{2.12}$$

Two standard (and related) approaches for finding the pair of estimators that maximize the above function are: (i) solving the likelihood equations obtained by taking the partial derivatives (see, e.g. Chaouche and Bacro, 2006) and (ii) dealing with one parameter at a time by considering an appropriate profile log-likelihood (see, e.g., Grimshaw, 1993; Choulakian and Stephens, 2001; Castillo and Serra, 2015). We start with the latter, and first consider the problem of maximizing the function in (2.11) with respect to σ when the parameter κ is held fixed. By straightforward differentiation, we find the partial derivative with respect to σ to be

$$\frac{\partial l(\kappa,\sigma)}{\partial\sigma} = \frac{1}{\sigma}\left\{-n + (1-\kappa)\sum_{i=1}^{n}\frac{X_i}{\sigma - \kappa X_i}\right\}, \quad \sigma > \max(0, \kappa X_{(n)}). \tag{2.13}$$

A close examination of the above derivative and elementary calculations lead to several important facts, observed by numerous authors and summarized below.

Lemma 2.2. *When $\kappa \geq 1$, the function $l_p(\sigma) = l(\kappa,\sigma)$ is strictly decreasing in σ on its domain, while for $\kappa < 1$ the function $l_p(\sigma)$ is increasing in σ for $\max(0,\kappa X_{(n)}) < \sigma < \hat\sigma$ and is decreasing in σ for $\hat\sigma < \sigma < \infty$. In the latter case, the unique mode $\hat\sigma$ satisfies the equation*

$$\frac{1}{n}\sum_{i=1}^{n}\frac{X_i}{\sigma - \kappa X_i} = \frac{1}{1-k}. \tag{2.14}$$

Moreover, we have $\lim_{\sigma\to\infty} l_p(\sigma) = -\infty$ while

$$\lim_{\sigma\to\max(0,\kappa X_{(n)})} l_p(\sigma) = \begin{cases} -\infty & \text{for } \kappa < 1 \\ -n\log X_{(n)} & \text{for } \kappa = 1 \\ \infty & \text{for } \kappa > 1. \end{cases} \tag{2.15}$$

In view of the above result, it is clear that maximum likelihood estimators cannot be found if the parameter space allows for any $\kappa > 1$, as, due to the monotonic behavior of the profile log-likelihood $l_p(\sigma)$, there are no critical numbers in the region $\kappa > 1$, $\sigma > X_{(n)}\kappa$ of the parameter space and the likelihood is infinite along its lower boundary. Thus, one needs to restrict the parameter space to the values $\kappa \leq 1$, $\sigma > \max(0, \kappa X_{(n)})$ when searching for the MLEs[1]. To find them, it is enough to examine the log-likelihood along the curve $(\kappa, \hat\sigma(\kappa))$, referred to as the *Choulakian-Stephens curve* by Castillo and Sierra (2015), where one can find further details of this approach. Under the above restrictions on the parameters, the MLEs can be shown to exist, although their uniqueness is still an open question.

[1]There is no universal agreement that the parameters should be restricted in any way, as the data at hand may actually connect with the parameters that are beyond such artificial restrictions (see, e.g., Castillo and Hadi, 1997).

An alternative process of finding the MLEs, considered by several authors (see, e.g., Grimshaw, 1993; Choulakian and Stephens, 2001; Ashkar and Tatsambon, 2007) follows Davison (1984), where an appropriate re-parameterization allows to reduce the problem to a one-dimensional search (see also Kozubowski et al., 2009, for the exponential-Pareto case). This method was extensively studied in Grimshaw (1993), which is a key reference in this connection. However, as explained below, there are some issues with the analysis in Grimshaw (1993). Under the parameterization in (2.8), this approach can be described as follows. Consider the (entire) parameter space as being composed of half straight lines that originate at the origin. The slopes of these lines are written as $1/\theta$ where $-\infty < \theta < 1/X_{(n)}$. There are three distinct cases of these lines:

- **(i)** $-\infty < \theta < 0$, $\kappa < 0$ and $\sigma = \kappa/\theta$ (Pareto case);
- **(ii)** $\theta = 0$, $\sigma > 0$ (exponential case);
- **(iii)** $0 < \theta < 1/X_{(n)}$, $\kappa > 0$, and $\sigma = \kappa/\theta$ (light-tail case).

The analysis of case (iii) should be further restricted to $\kappa \leq 1$ as discussed above (and done in Grimshaw, 1993). Now, in cases (i) and (iii), the log-likelihood (2.11) evaluated along these lines can be written as

$$l(\kappa, \kappa/\theta) = n \log|\theta| - n \log|\kappa| + \left(\frac{1}{\kappa} - 1\right) \sum_{i=1}^{n} \log(1 - \theta X_i). \tag{2.16}$$

Elementary calculus shows that under case (i), the function in (2.16) is increasing in κ on the interval $(-\infty, \hat{\kappa})$ while it is decreasing on the interval $(\hat{\kappa}, 0)$, where

$$\hat{\kappa} = -\frac{1}{n} \sum_{i=1}^{n} \log(1 - \theta X_i), \tag{2.17}$$

see Grimshaw (1993, p. 188). Thus, the largest value of the log-likelihood (2.16) is equal to

$$L(\theta) = l(\hat{\kappa}, \hat{\kappa}/\theta) = -n \log\left(\frac{\hat{\kappa}}{\theta}\right) - n + n\hat{\kappa}. \tag{2.18}$$

Remark 2.1. This is exactly the same function as that given by equation (2.1) in Grimshaw (1993) or equation (7) in Choulakian and Stephens (2001). Elementary calculus again shows that at the boundary of case (i), when $\theta \to 0^-$, the function $L(\theta)$ above converges to

$$-n(\log \overline{X} + 1), \tag{2.19}$$

which is the largest likelihood corresponding to case (ii) where we maximize the function (2.12). In particular, when we are only interested in Pareto/exponential case $\kappa \leq 0$, finding the MLEs is equivalent to maximizing the function $L(\theta)$ given above with respect to $\theta \in (-\infty, 0]$. Note that in this case the function L in (2.18) can be written as

$$L(\theta) = n(\log(-\theta) + \hat{\kappa}) - n(\log(-\hat{\kappa}) + 1). \tag{2.20}$$

Moreover, as $\theta \to -\infty$ then $-\hat{\kappa}$ converges to infinity. In addition, we have

$$\log(-\theta) + \hat{\kappa} = -\frac{1}{n} \sum_{i=1}^{n} \log \frac{1 - \theta X_i}{-\theta} \to -\frac{1}{n} \sum_{i=1}^{n} \log X_i \;\text{ as } \theta \to -\infty.$$

Consequently, $L(\theta)$ approaches $-\infty$ as $\theta \to -\infty$. Since the function $L(\theta)$ is continuous with a finite limit (2.19) at $\theta = 0$, this shows the MLEs exist in this special sub-case.

The analysis connected with the case (iii) is very similar, and the conclusion is essentially the same: the function in (2.16) is increasing in κ on the interval $(0,\hat{\kappa})$ while it is decreasing on the interval $(\hat{\kappa},\infty)$, where $\hat{\kappa}$ is again given by (2.17). However, this quantity is not necessarily restricted to the interval $(0,1)$. Consequently, if we are restricting the parameter space so that $\kappa \le 1$, we need to pay a close attention to the magnitude of $\hat{\kappa}$. Let us make several observations regarding the function $\hat{\kappa}=\hat{\kappa}(\theta)$ defined through (2.17):

(A) $\lim_{\theta\to 0}\hat{\kappa}(\theta)=0=\hat{\kappa}(0)$, $\lim_{\theta\to -\infty}\hat{\kappa}(\theta)=-\infty$; $\lim_{\theta\to 1/X_{(n)}^-}\hat{\kappa}(\theta)=\infty$;

(B) $\frac{d}{d\theta}\hat{\kappa}(\theta)=\frac{1}{n}\sum_{i=1}^n\frac{X_i}{1-\theta X_i}>0$ and $\lim_{\theta\to 0}\frac{d}{d\theta}\hat{\kappa}(\theta)=\overline{X}$;

(C) $\frac{d^2}{d\theta^2}\hat{\kappa}(\theta)=\frac{1}{n}\sum_{i=1}^n\left(\frac{X_i}{1-\theta X_i}\right)^2>0$ and $\lim_{\theta\to 0}\frac{d^2}{d\theta^2}\hat{\kappa}(\theta)=\frac{1}{n}\sum_{i=1}^n X_i^2$.

It follows that the function $\hat{\kappa}(\theta)$ is increasing in θ. In particular, under case (iii) with positive θ and κ, there is a unique $\theta^*\in(0,1/X_{(n)})$ such that $\hat{\kappa}(\theta^*)=1$ (while $\hat{\kappa}(\theta)<1$ for $\theta\in(0,\theta^*)$ and $\hat{\kappa}(\theta)>1$ for $\theta\in(\theta^*,1/X_{(n)})$. As a consequence, under case (iii) with the restriction $\kappa\le 1$, the largest value of the log-likelihood (2.16) becomes $l(\hat{\kappa},\hat{\kappa}/\theta)$ whenever $0\le\theta\le\theta^*$ but then turns into $l(1,1/\theta)$ when $\theta^*\le\theta\le 1/X_{(n)}$. Thus, maximizing the log-likelihood over the region with $\kappa\in[0,1]$ reduces to finding $\hat{\theta}$ that maximizes the function

$$g(\theta)=n\log\theta+\begin{cases}-n\log\hat{\kappa}-n+n\hat{\kappa} & \text{for } 0\le\theta<\theta^*\\ 0 & \text{for } \theta^*\le\theta\le 1/X_{(n)}.\end{cases} \tag{2.21}$$

Note that $g(\theta)$ is a continuous function on a closed interval, so a global maximum of g on $[0,1/X_{(n)}]$ always exists. Moreover, we have $g(0)=-n(\log\overline{X}+1)$, $g(1/X_{(n)})=-n\log(X_{(n)})$, and g is strictly increasing on the interval $(\theta^*\le\theta\le 1/X_{(n)})$.

Remark 2.2. Let us note that the analysis in Grimshaw (1993) that begins on p. 188 and onward does not seem to be quite correct, for two reasons. First, the value of $\hat{\kappa}$ is not restricted there to be less than 1, and the function (2.18) is studied over its entire domain $[0,1/X_{(n)}]$, which likely results with "complications" with multiple local maxima discussed on p. 188 and in subsequent numerical routines. Second, the derivative $h(\theta)$ of the profile log-likelihood presented in the appendix (see also (2.2) on p. 188) does not seem to be quite correct as well, which propagates into the subsequent analysis and numerical routines. In particular, the theorem on top of p. 188 in Grimshaw (1993) seems to be in error. This derivative, that plays a crucial role in the analysis, should be as follows:

$$\frac{d}{d\theta}g(\theta)=-n\left\{-\frac{d}{d\theta}\hat{\kappa}(\theta)-\frac{1}{\theta}+\frac{\frac{d}{d\theta}\hat{\kappa}(\theta)}{\hat{\kappa}(\theta)}\right\},\quad \theta<\theta^*. \tag{2.22}$$

Going back to the function $g(\theta)$ in (2.21), we have the following explicit expression for its derivative:

$$\frac{d}{d\theta}g(\theta)=\frac{n}{\theta}+\sum_{i=1}^n\frac{X_i}{1-\theta X_i}+n\sum_{i=1}^n\frac{X_i}{1-\theta X_i}\Big/\sum_{i=1}^n\log(1-\theta X_i),\quad \theta<\theta^*. \tag{2.23}$$

An important property of the derivative is its value at zero, presented below.

Proposition 2.1. *Let g be defined as in (2.23). Then*

$$\lim_{\theta\to 0}\frac{d}{d\theta}g(\theta)=\sum_{i=1}^{n}X_i\left(1-\frac{n}{2}T_n\right). \tag{2.24}$$

This property, which is similar to the analysis of the Pareto-exponential case discussed in Section 2.1, makes it possible to conclude that when the derivative is negative (so that $T_n > 2/n$, that is the statistic T_n is relatively large) then likelihood will be larger at some Pareto case with $\kappa < 0$, as compared with its value at zero, corresponding to exponentiality.

Finally, let us turn to another common approach to maximum likelihood estimation, consisting of solving the likelihood equations. This approach was revisited in Chaouche and Bacro (2006), who worked with the (2.7) parameterization of the GPD family. Here, the log-likelihood takes on the form

$$l(\xi,\beta)=-n\log\beta-\left(\frac{1}{\xi}+1\right)\sum_{i=1}^{n}\log(1+\xi X_i/\beta), \tag{2.25}$$

where the parameters are restricted as follows: $-\infty<\xi<\infty$ and $\beta>\max(0,-\xi X_{(n)})$, so that $\beta>0$ for $\xi>0$ (Pareto case) and for $\xi=0$ (exponential case). Similarly to the parameterization (2.8), the parameters should be restricted to the case $\xi\geq -1$, as otherwise the log-likelihood has no local extrema and is infinite along the lower boundary of its domain. By setting the partial derivatives of the log-likelihood to zero, followed by some algebra, Chaouche and Bacro (2006) obtained the following system of equations:

$$\begin{cases}\frac{1}{n}\sum_{i=1}^{n}\log\left(1+\xi X_i/\beta\right)=\xi\\ \frac{1}{n}\sum_{i=1}^{n}\frac{1}{1+\xi X_i/\beta}=\frac{1}{1+\xi}.\end{cases} \tag{2.26}$$

Simple algebra shows that an alternative way of stating (2.26) is to say that both partial derivatives are zero iff ξ and $\rho=\xi/\beta$ satisfy the system

$$\begin{cases}\frac{1}{n}\sum_{i=1}^{n}\log\left(1+\rho X_i\right)=\xi\\ h(\rho)=0,\end{cases} \tag{2.27}$$

where the function $h(\rho)$ is given by[2]

$$h(\rho)=1+\frac{1}{n}\sum_{i=1}^{n}\log\left(1+\rho X_i\right)-\left(\frac{1}{n}\sum_{i=1}^{n}\frac{1}{1+\rho X_i}\right)^{-1}. \tag{2.28}$$

Clearly, one can find the critical points of the log-likelihood function by solving the equation $h(\rho)=0$ for the parameter ρ and then converting to the parameters ξ and β via $\beta=\xi/\rho$ and the first equation in (2.27) (provided that $\rho\neq 0$). This is the approach undertaken in Chaouche and Bacro (2006), who focus on the function $h(\rho)$ and its basic properties. It is easy to see that one of these roots is equal to zero and the limits of the function $h(\rho)$ at the boundaries of its domain, $-1/X_{(n)}<\rho<\infty$, are both $-\infty$, as noted by Chaouche and Bacro (2006). However, some further claims of that paper do not seem to be quite correct, as explained below. In particular, the authors claim that there is a negative root of this equation very close to the left boundary of the domain, which they denote by ρ_1. Additionally, based

[2] In Chaouche and Bacro (2006), this function appears in eq. (7), where it is defined through a special notation related to the geometric and harmonic means.

on their simulations, the authors make a conjecture that there are actually three roots of this equation, with the third root (besides 0 and ρ_1) being negative or positive according to the sign of the statistic (2.31). Indeed, our analysis shows that there is a close connection between that statistic (and ultimately, the statistic T_n) and the roots of the equation $h(\rho) = 0$, which can be deduced from the following simple lemma.

Lemma 2.3. *Let $h(\rho)$ be the function defined in (2.28). Then, we have $h(0) = h'(0) = 0$ while*

$$h''(0) = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - 2\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right)^2 . \tag{2.29}$$

Proof: The equation $h(0) = 0$ is trivial, while the property $h'(0) = 0$ follows by standard calculation of the derivative (and can also be deduced from Proposition A.1 in Chaouche and Bacro, 2006). In turn, routine calculations produce

$$h''(\rho) = -\frac{1}{n}\sum_{i=1}^{n}\frac{1}{(1+\rho X_i)^2} - 2\left\{\frac{\left(\frac{1}{n}\sum_{i=1}^{n}\frac{X_i}{(1+\rho X_i)^2}\right)^2}{\left(\frac{1}{n}\sum_{i=1}^{n}\frac{1}{1+\rho X_i}\right)^3} - \frac{\frac{1}{n}\sum_{i=1}^{n}\frac{X_i^2}{(1+\rho X_i)^3}}{\left(\frac{1}{n}\sum_{i=1}^{n}\frac{1}{1+\rho X_i}\right)^2}\right\}, \tag{2.30}$$

which leads to (2.29). □

It follows from the above lemma that the second derivative of h at zero is positive if and only if the expression on the right-hand-side in (2.29) is positive, which in turn is equivalent to the statistic

$$S_n = \frac{n}{2}T_n - 1 \tag{2.31}$$

being positive. Consequently, if this is the case, then the graph of the function $h(\rho)$ is concave up at $\rho = 0$. Since we also have $h(0) = h'(0) = 0$, under this condition there must be at least three roots of the equation $h(\rho) = 0$: a negative root, zero, and a positive root. In particular, large values of the statistic T_n point towards the existence of the positive root, which may lead to the positive MLE of the shape parameter ξ connected with Pareto II (heavy-tail) GPD case. However, when the statistic (2.31) is negative, then nothing really can be deduced from the concavity of the graph of h at zero regarding the existence of further roots of the equation $h(\rho) = 0$. We agree with the comments of Chaouche and Bacro (2006) that further results on the function $h(\rho)$ are needed to resolve the issue of the existence and the nature of its zeroes.

Remark 2.3. Let us note that while $\rho = 0$ is a root of the equation $h(\rho) = 0$, it is actually not relevant in maximization of the log-likelihood function. While it is true that $\xi = 0$, which corresponds to $\rho = 0$, satisfies the equations in (2.26), when these equations are formally derived by setting both partial derivatives of the log-likelihood to zero, one needs to actually assume that $\xi \neq 0$. When $\xi = 0$, these derivatives require a subtle calculation, which, when properly done, shows that they cannot both equal zero unless we have $(1/n)\sum_{i=1}^{n} X_i^2 = 2\overline{X}^2$ (or, equivalently, $T_n = 2/n$)[3]. Since this occurs with probability zero, the case $\xi = 0$, and the corresponding special exponential case of the GPD, can be eliminated from consideration.

[3]This was pointed out by Grimshaw (1993), although the two relevant equations on top of p. 187 in that paper are not quite correct; the right-hand-side of the first equation should be multiplied by -1.

While the analysis involving the derivatives of the log-likelihood function undertaken in Chaouche and Bacro (2006) is an alternative approach to that of Grimshaw (1993), discussed above, it has common elements with the latter. Consider again the quantity ρ_1, which is defined in Chaouche and Bacro (2006) as a unique solution of the equation

$$\frac{1}{n}\sum_{i=1}^{n}\log(1+\rho X_i) = -1. \tag{2.32}$$

Observe that the first equation in (2.27) provides the value of $\xi = \xi(\rho)$ that maximizes the log-likelihood function along the straight line $\beta = \xi/\rho$. As in the analysis of Grimshaw (1993), the log-likelihood along this line is increasing in ξ as ξ varies from $-\infty$ to the value given by the left-hand-side of the first equation in (2.27), $\xi = \xi(\rho)$, and then, passed that value, it is decreasing. Moreover, as a function of ρ, this $\xi = \xi(\rho)$ is increasing from $-\infty$ to 0 as ρ varies in $-1/X_{(n)} < \rho < 0$. Note that as ρ varies in that interval, the lines $\beta = \xi/\rho$ become steeper and steeper, and when $\rho = -1/X_{(n)}$ we get the line $\beta = -X_{(n)}$, which is the lower boundary of the parameters in the 2nd quadrant, while for $\rho = 0$ we obtain the vertical line $\xi = 0$. Thus, solving the equation (2.32) leads to that value of ρ for which we have $\xi(\rho) = -1$. This ρ is precisely what the authors of Chaouche and Bacro (2006) denote by ρ_1. This ρ_1 cannot be a solution of the equation $h(\rho) = 0$, because if it was, then the pair $\xi = -1$, $\beta = -1/\rho_1$ would be a solution of the system (2.27). However, equation (2.26), which is equivalent to (2.27), would not be satisfied. One fundamental reason for this is the fact that the log-likelihood function is strictly decreasing in β along any vertical line $\xi = c$ with $c \le 1$, so the partial derivative with respect to β cannot equal zero at any point on such a line. This shows that we cannot have $h(\rho_1) = 0$. Additionally, we cannot have $h(\rho^*) = 0$ for any ρ^* such that $-1/X_{(n)} < \rho^* < \rho_1$ either for precisely the same reason. Indeed, if we had, then this would imply that

$$\xi^* = \frac{1}{n}\sum_{i=1}^{n}\log\left(1+\rho^* X_i\right), \quad \beta^* = \xi^*/\rho^*$$

would satisfy the system (2.26). However, by the monotonic nature of the function $\xi(\rho)$ discussed above, we would necessarily have $\xi^* < -1$, and we already know that partial derivative of the log-likelihood function cannot be zero for any point in the interior of the parameter space for which $\xi < -1$. The above analysis shows that the claims made in Chaouche and Bacro (2006) regarding the existence of a root of the equation $h(\rho) = 0$ that is close to $-1/X_{(n)}$ are not quite correct. To the contrary, any root of this equation must necessarily be greater than ρ_1, where ρ_1 satisfies equation (2.32).

REFERENCES

[1] Arendarczyk, M.; Kozubowski, T.J. and Panorska, A.K. (2018). A bivariate distribution with Lomax and geometric margins, *J. Korean Statist. Soc.*, **47**(4), 405–422.

[2] Ashkar, F. and Tatsambon, C.N. (2007). Revisiting some estimation methods for the generalized Pareto distribution, *J. Hydrology*, **346**, 136–143.

[3] Castillo, E. and Hadi, A.S. (1997). Fitting the generalized Pareto distribution to data, *J. Amer. Statist. Assoc.*, **92**(440), 1609–1620.

[4] Castillo, J. and Daoudi, J. (2009). Estimation of the generalized Pareto distribution, *Statist. Probab. Lett.*, **79**, 684–688.

[5] Castillo, J. and Serra, I. (2015). Likelihood inference for generalized Pareto distribution, *Comput. Statist. Data Anal.*, **83**, 116–128.

[6] Chaouche, A. and Bacro, J.-N. (2006). Statistical inference for the generalized Pareto distribution: Maximum likelihood revisited, *Comm. Statist. Theory Methods*, **35**(5), 785–802.

[7] Choulakian, V. and Stephens, M.A. (2001). Goodness-of-fit tests for the generalized Pareto distribution, *Technometrics*, **43**(4), 478–484.

[8] Davison, A.C. (1984). *Modelling excesses over high thresholds, with an application.* In: "Statistical Extremes and Applications" (J.T. de Oliveira, Ed.), Reidel, Dordrecht, pp. 461–482.

[9] de Zea Bermundez, P. and Kotz, S. (2010a). Parameter estimation of the generalized Pareto distribution – Part I, *J. Statist. Plann. Inference*, **140**, 1353–1373.

[10] de Zea Bermundez, P. and Kotz, S. (2010b). Parameter estimation of the generalized Pareto distribution – Part II, *J. Statist. Plann. Inference*, **140**, 1374–1388.

[11] Grimshaw, S.D. (1993). Computing maximum likelihood estimates for the generalized Pareto distribution, *Technometrics*, **35**(2), 185–191.

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

[13] Kozubowski, T.J.; Panorska, A.K.; Qeadan, F.; Gershunov, A. and Rominger, D. (2009). Testing exponentiality versus Pareto distribution via likelihood ratio, *Comm. Statist. Sim. Comput.*, **38**(1), 118–139.