Annals of the Institute of Statistical Mathematics (2021) 73:1153–1185
https://doi.org/10.1007/s10463-020-00782-z

Efficient likelihood-based inference for the generalized Pareto distribution

Hideki Nagatsuka[1] · **N. Balakrishnan**[2]

Received: 1 June 2020 / Revised: 2 November 2020 / Accepted: 06 November 2020 /
Published online: 11 January 2021

Abstract
It is well known that inference for the generalized Pareto distribution (GPD) is a difficult problem since the GPD violates the classical regularity conditions in the maximum likelihood method. For parameter estimation, most existing methods perform satisfactorily only in the limited range of parameters. Furthermore, the interval estimation and hypothesis tests have not been studied well in the literature. In this article, we develop a novel framework for inference for the GPD, which works successfully for all values of shape parameter k. Specifically, we propose a new method of parameter estimation and derive some asymptotic properties. Based on the asymptotic properties, we then develop new confidence intervals and hypothesis tests for the GPD. The numerical results are provided to show that the proposed inferential procedures perform well for all choices of k.

Keywords Asymptotic normality · Interval estimation · Hypothesis testing · Non-regularity problem · Extreme value · Peaks over threshold

1 Introduction

1.1 Background

Statistical modeling of the largest or smallest values (extreme values) of certain natural phenomena (e.g., waves, floods, earthquakes, winds, temperatures, etc) is of importance

Supplementary Information The online version of this article (https://doi.org/10.1007/s10463-020-00782-z) contains supplementary material, which is available to authorized users.

✉ Hideki Nagatsuka
hideki@indsys.chuo-u.ac.jp

1 Department of Industrial and Systems Engineering, Chuo University, 1-13-27 Kasuga, Bunkyo-ku, Tokyo 112-8551, Japan

2 Department of Mathematics and Statistics, McMaster University, Hamilton, Ontario L8S 4K1, Canada

in various practical applications. For example, the distributions of high waves and large floods are important in the designs of dikes and dams, respectively. The traditional approach to the analysis of extreme values is based on the generalized extreme value distribution (GEVD), which is a limiting distribution for extreme values, comprising the Gumbel, Fréchet and Weibull distributions all as special cases (Coles 2001, Castillo et al. 2004, Beirlant et al. 2004, and de Haan and Ferreira 2006). Although the GEVD is suitable for fitting data on maxima, there have been some criticisms since using only maxima leads to loss of information contained in other values in the data. This problem is remedied by considering some largest values in the given period instead of the largest values alone, that is, considering all values larger than a given threshold (exceedances over the threshold). The generalized Pareto distribution (GPD), which is a limiting distribution for exceedances over the threshold, offers a unifying approach to modeling such values (Coles 2001, Castillo et al. 2004, Beirlant et al. 2004, and de Haan and Ferreira 2006). Using the GPD instead of the GEVD can be a solution to the above mentioned problem. This distribution was initially introduced by Pickands (1975) and has since been used widely to analyze exceedances over threshold in various areas (see, for example, Salvadori et al. 2007). The cumulative distribution function (cdf) of the GPD is given by

$$F(x;k,\sigma) = \begin{cases} 1-\left(1-k\dfrac{x}{\sigma}\right)^{1/k}, & k\neq 0, \\ 1-\exp\left(-\dfrac{x}{\sigma}\right), & k=0, \end{cases} \qquad x\in\mathscr{X}_{k,\sigma}, \tag{1}$$

where $k\in\mathbb{R}$ and $\sigma>0$ are the shape and scale parameters, respectively, and $\mathscr{X}_{k,\sigma}=\{x : 0<x<\infty, \text{ if } k\leq 0, \text{ or } , 0<x<\sigma/k, \text{ if } k>0\}$. The corresponding probability density function (pdf) is

$$f(x;k,\sigma) = \begin{cases} \dfrac{1}{\sigma}\left(1-k\dfrac{x}{\sigma}\right)^{1/k-1}, & k\neq 0, \\ \dfrac{1}{\sigma}\exp\left(-\dfrac{x}{\sigma}\right), & k=0, \end{cases} \qquad x\in\mathscr{X}_{k,\sigma}, \tag{2}$$

for $k\in\mathbb{R}$ and $\sigma>0$. The shape of the pdf varies over the shape parameter k (see Figure 1). The smaller the value of k is, the heavier the tail of the distribution becomes, that is, very large values can be observed. On the other hand, the larger the value of k becomes, lighter the tail of the distribution becomes. In fact, $-1/k$ is known as the tail index, and we can know the risk in a situation from the value of k or the tail index. For example, a small value of k (or the tail index) indicates that the events associated with large values occur with high probability. An important property of the GPD is that $X-u \sim GPD(k,\sigma-k\,u)$ if $X\sim GPD(k,\sigma)$, given that $X>u$ for every $u\in\mathbb{R}$, where $GPD(k,\sigma)$ is the GPD with shape parameter k and scale parameter σ (Castillo and Hadi 1997).

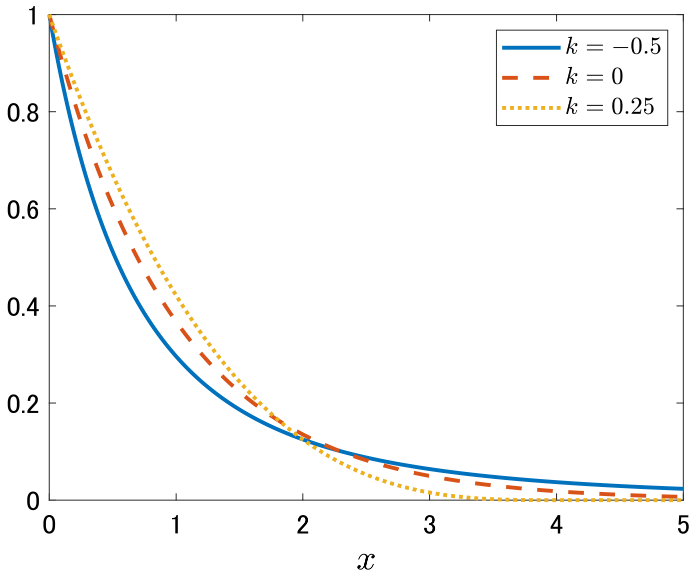

Fig. 1 Pdfs of GPD for different values of the shape parameter k, with $\sigma = 1$

1.2 Review of parameter estimation for GPD

We first review parameter estimation for the GPD. Although the GPD is useful for modeling exceedances over threshold, it is well known that parameter estimation for the GPD is a difficult problem. We refer the readers to de Zea Bermudez and Kotz (2010a, 2010b), Zhang and Stephens (2009), del Castillo and Serra (2015), and Chen et al. (2017) for pertinent details. For $k > 1$, the maximum likelihood estimators (MLEs) do not exist. For $k \le -1/r$, $r \in \mathbb{N}$, the rth moment does not exist, and therefore, all the moment-based estimators such as the method of moments (MOM) estimators, the probability weighted moments (PWM) estimators and the L-moments estimators proposed by Hosking (1990) exist only for k in certain ranges. The Hill estimator (Hill 1975) and the Pickands estimator (Pickands 1975) are well-known estimators of the tail index. These estimators are very simple and sometimes used for parameter estimation of k as an initial value for other methods. Two empirical Bayesian methods of parameter estimation have been proposed by Zhang and Stephens (2009) and Zhang (2010). They have shown good performances of their estimators for moderate or small values of k ($k \le 0.5$), through Monte Carlo simulations. However, for large values of k, their estimators have considerably large bias and RMSE even if the sample size is large (Chen et al. (2017)), also as shown later by Monte Carlo simulations. This reveals that their estimators may not have consistency property for large k. Recently, Chen et al. (2017) have proposed a new method, inspired by the minimum distance estimation and the M-estimation in the regression with ρ functions including Tukey's biweight function. Prior to this, Song and Song (2012) proposed a method that is close to the one proposed by Chen et al. (2017). This method is a special case of Chen et al.'s method when ρ is a constant function. The reason for using ρ functions in Chen et al.'s method is that the estimates obtained by minimizing the residual sum of squares in M-estimation without ρ functions are sensitive to the shape parameter k (Chen et al. 2017). These authors also provided a weighted version of their method, wherein the residuals are weighted by the inverse of the asymptotic variance of the residual. They proved that both methods are consistent for all k. Simulation studies showed that their two estimators were comparable to the Bayesian method of Zhang (2010) and EPM by Castillo and Hadi (1997) for $k < 1/2$, while they performed better for $k > 1/2$. Moreover, between their two methods, the weighted version performed slightly better than the original method.

1.3 Review of other inferential methods for GPD

As pointed out by Chen et al. (2017), interval estimation and hypothesis testing for GPD have not been discussed much in the literature. One of the reasons is the non-regularity problem in the maximum likelihood (ML) method for $k \geq 1/2$. The ML method for the GPD has been discussed by many authors including Smith (1984, 1985), Hosking and Wallis (1987), Davison and Smith (1990), Grimshaw (1993), de Zea Bermudez and Kotz (2010a), and del Castillo and Serra (2015). It is known from these works that the MLEs exist only for $k < 1$, and have asymptotic normality only for $k < 1/2$. Therefore, when $k \geq 1/2$, it is difficult to construct confidence intervals and hypothesis testing procedures based on the asymptotic properties of MLEs while, when $k < 1/2$, we can develop them readily using the asymptotic properties of MLEs. Instead, Castillo and Hadi (1997) and Chen et al. (2017) provided interval estimation by using bootstrap method (especially, bootstrap-t method) based on their proposed estimators.

1.4 Aims and outline of the present work

In spite of many papers dealing with the GPD, there does not appear to be any work wherein inferential procedure is established formally for all possible values of k. In this work, we develop a new framework for inference for the GPD, which works successfully for all k. The key idea is to use an unimodal likelihood function that is proposed here, instead of the usual likelihood function, to tackle the non-regularity problem in the usual ML method. We propose a new method of parameter estimation for the GPD and derive some asymptotic properties of the proposed estimators and related statistics (for $0.5 \leq k < 1$, the asymptotic properties of the estimator of σ have not been derived yet). We also prove that the estimates by the proposed method always exist uniquely for all choices of k while, in most existing methods, existence and uniqueness of the estimates are not guaranteed. Based on the asymptotic properties of the proposed estimators and related statistics, we also develop new methods of confidence intervals and hypothesis tests for the GPD.

The remainder of this article is organized as follows. In Sects. 2 and 3, we develop inferential methods for shape parameter k and scale parameter σ, respectively. Then, the performances of the proposed estimators of parameters, confidence intervals and hypothesis tests are assessed by Monte Carlo simulation in Sect. 4. To illustrate all the inferential methods proposed here, we apply them to analyze the zero-crossing hourly mean periods (in seconds) of the sea waves measured in Bilbao buoy, Spain, and the Kevlar/Epoxy strand lifetime (in hours) at 70% stress level, in Sect. 5. Finally, some concluding remarks are made in Sect. 6. Some technical proofs, remarks on the computation for the proposed inferential method and one of the illustrative examples are given in the supplementary materials. The Matlab codes are available on Github (https://www.github.com/HidekiNagatsuka/GPD/).

2 New likelihood-based inference for the shape parameter *k*

In this section and in the next section, we develop a new likelihood-based inference providing efficient methods for parameter estimation, interval estimation and hypothesis testing for k and σ. Unlike most existing methods which can estimate parameters only when their estimates of k is in a certain range and also their asymptotic properties being not known, the proposed method provides a successful inferential framework providing efficient estimators, confidence intervals and tests of hypothesis, based on their asymptotic properties of the estimators.

The proposed method consists of two parts: one is for k, and another is for σ. In this section, we develop the new likelihood-based inference for k and specifically provide an estimator for k. Next, we derive a score function for k and some asymptotic properties of statistics related to the score function. Finally, we construct confidence intervals and hypothesis tests for the shape parameter k.

2.1 New maximum likelihood estimator of *k*

Let $X_1, \dots, X_n$ be i.i.d. random variables from the GPD with cdf as in (1), and $X_{1:n} \le \cdots \le X_{n:n}$ be the order statistics obtained by arranging the above X_i's in increasing order of magnitude. For any fixed j, $1 \le j \le n$, we derive the joint density of $\boldsymbol{S}_n^{(j)}$, where
$\boldsymbol{S}_n^{(j)} = (S_{1:n}^{(j)}, \dots, S_{j-1:n}^{(j)}, S_{j+1:n}^{(j)}, \dots, S_{n:n}^{(j)})$, with $S_{i:n}^{(j)} = X_{i:n}/X_{j:n}$, $i \ne j$, $1 \le i \le n$. We note that $S_{i:n}^{(j)}$'s do not depend on σ.

Proposition 1 *For $k \in \mathbb{R}$ and any fixed j, $1 \le j \le n$, the joint density of $\boldsymbol{S}_n^{(j)}$ is given by*

$$\phi\left(\boldsymbol{s}_n^{(j)};k\right) = \begin{cases} n! \displaystyle\int_{\chi_k} \frac{1}{|k|}\left(\frac{u}{k}\right)^{n-1} \prod_{i=1}^{n} \left(1 - us_i\right)^{1/k-1} du, & k \ne 0, \\ \dfrac{n!\,(n-1)!}{\left(\sum_{i=1}^{n} s_i\right)^{n}}, & k = 0, \end{cases}$$

$$s_1 \le \cdots \le s_{j-1} \le 1 \le s_{j+1} \le \cdots \le s_n,$$

where $\chi_k = \{u : -\infty < u < 0,$ if $k < 0,$ or $, 0 < u < 1/s_n,$ if $k > 0\}$, $\boldsymbol{s}_n^{(j)} = (s_1, \dots, s_{j-1}, s_{j+1}, \dots, s_n)$, *and* $s_j = 1$.

Proof See "Appendix B.1." □

From Proposition 1, we can obtain the likelihood function for k based on $\boldsymbol{S}_n^{(j)}$ as

$$l\left(k;\boldsymbol{s}_n^{(j)}\right) = \phi\left(\boldsymbol{s}_n^{(j)};k\right), \tag{3}$$

where $\boldsymbol{s}_n^{(j)}$ is the vector consisting of the realized values of $S_{i:n}^{(j)}$, $i \neq j, 1 \leq i \leq n$. Then, the MLE of k based on $\boldsymbol{S}_n^{(j)}$, denoted by $\hat{k}$, is obtained by maximizing $l(k;\boldsymbol{s}_n^{(j)})$ with respect to k, replacing $\boldsymbol{S}_n^{(j)}$ by $\boldsymbol{s}_n^{(j)}$.

2.2 Properties of the proposed estimator of *k*

In this subsection, we provide some properties of the estimator of k proposed above, which can be used to derive confidence intervals and hypothesis tests for the parameter k.

First, it may be noted that the likelihood function $l(k;\boldsymbol{s}_n^{(j)})$ depends on j, and so we may wonder which j may be good to use. Fortunately, from Theorem 1, we note that we do not need to worry about the choice of j, in the ML method based on $\boldsymbol{S}_n^{(j)}$.

Theorem 1 *For any* $j, 1 \leq j \leq n$, *the MLE of* k *based on* $\boldsymbol{S}_n^{(j)}$, $\hat{k}$, *does not depend on* j.

Proof For any $j \neq t$, $1 \leq j, t \leq n$, there is a one-to-one transformation between $\boldsymbol{S}_n^{(j)}$ and $\boldsymbol{S}_n^{(t)}$ since we have

$$S_{i:n}^{(t)} = S_{i:n}^{(j)}/S_{t:n}^{(j)}, \qquad i \neq j,\ t, \quad i = 1, \dots, n.$$

This clearly implies that the MLE of k based on $\boldsymbol{S}_n^{(j)}$ is identical to the MLE based on $\boldsymbol{S}_n^{(t)}$, as required. □

We further observe that the ML method based on $\boldsymbol{S}_n^{(j)}$ is equivalent to the ML methods based on $X_{i:n}/X_{i+1:n}$'s and $X_{i+1:n}/X_{i:n}$'s in the sense of the following theorem, which once again reinforces that the ML method based on $\boldsymbol{S}_n^{(j)}$ does not depend on j.

Theorem 2 *For any fixed* i *and* j, $1 \leq i,j \leq n$, *the MLE of* k *based on* $X_{i:n}/X_{i+1:n}$*'s (and* $X_{i+1:n}/X_{i:n}$*'s) agrees with the MLE based on* $\boldsymbol{S}_n^{(j)}$.

Proof The proof is very similar to the proof of Theorem 1, and is therefore omitted. □

In the required computation, we need to choose a j. We recommend the choice of $\left[\frac{n}{2}\right] + 1$ for j, where $[z]$ is the greatest integer less than or equal to z, since the proposed estimator could be sensitive to the presence of outliers if we take j corresponding to lower or upper extremes.

We provide Propositions 8, 9 and 10 in "Appendix A." These propositions give the derivatives of $l(k;\boldsymbol{s}_n^{(j)})$, which are used for theorems, lemmas and propositions provided later, and their proofs are given in the supplementary materials.

Although, in most existing methods, existence and uniqueness of the estimates are not always guaranteed, the following theorem and the ensuing corollary imply that the estimate of k obtained by solving the equation $l'\left(k;\boldsymbol{s}_n^{(j)}\right) = 0$ always exists uniquely for all k.

Theorem 3 *For $k \in \mathbb{R}$, any fixed j, $1 \le j \le n$, and any given $\boldsymbol{s}_n^{(j)} = (s_1, \dots, s_{j-1}, s_{j+1}, \dots, s_n)$ such that $s_1 \le \cdots \le s_{j-1} \le 1 \le s_{j+1} \le \cdots \le s_n$, the likelihood equation*

$$l'\left(k;\boldsymbol{s}_n^{(j)}\right) = 0$$

always has a unique solution with respect to k.

Proof See "Appendix 3." □

Corollary 1 *For $k \in \mathbb{R}$, any fixed j, $1 \le j \le n$, and any given $\boldsymbol{s}_n^{(j)}$, the likelihood function $l\left(k;\boldsymbol{s}_n^{(j)}\right)$ is unimodal with respect to k.*

For proving the result that the estimator of k has consistency property, the following lemma is needed.

Lemma 1 *For any fixed $k \neq k_0$, where k_0 is the true value of the parameter k, and for any fixed j, $1 \le j \le n$,*

$$\lim_{n\to\infty} \Pr\left(l\left(k;\boldsymbol{S}_n^{(j)}\right) < l\left(k_0;\boldsymbol{S}_n^{(j)}\right)\right) = 1.$$

Proof See "Appendix 1." □

Theorem 4 *The estimator $\hat{k}$ is consistent for $k \in \mathbb{R}$.*

Proof Using Lemma 1, the proof is similar to that of Theorem 3.7 in Lehmann and Casella (1998, p.447), and is therefore not presented here. □

We can obtain the score function of k by differentiating the logarithm of the likelihood function in (3) with respect to k. We provide the asymptotic normality of the score function, denoted by $U\left(k;\boldsymbol{S}_n^{(j)}\right)$, which will be made use of in developing test of hypothesis later in Sect. 4.

Theorem 5 *For $k \in \mathbb{R}$ and $j \in \{1, \dots, n\}$,*

$$\frac{U\left(k;\boldsymbol{S}_n^{(j)}\right)}{\sqrt{I_{j,n}(k)}} \xrightarrow{d} N(0,1), \qquad \text{as } n \longrightarrow \infty,$$

where

$$U\left(k;\boldsymbol{S}_n^{(j)}\right) = \frac{\partial}{\partial k} \log l\left(k;\boldsymbol{S}_n^{(j)}\right)$$

is the score function of k based on $\boldsymbol{S}_n^{(j)}$, *and*

$$I_{j,n}(k) = -E\left(\frac{\partial^2}{\partial k^2} \log l\left(k;\boldsymbol{S}_n^{(j)}\right)\right),$$

is the Fisher information about k in $\boldsymbol{S}_n^{(j)}$.

Proof See "Appendix B.4." □

We also provide the asymptotic normality of the proposed estimator of k, which will be used later in Sect. 4 to develop interval estimation and hypothesis testing for the parameter k.

Theorem 6

$$\frac{\hat{k} - k}{\sqrt{I_{j,n}^{-1}(k)}} \xrightarrow{d} N(0,1), \qquad \text{as } n \longrightarrow \infty. \tag{4}$$

Proof See "Appendix B.5." □

From Theorems 5 and 6, the following Wald, score and likelihood ratio (LR)-type test statistics for k and their asymptotic properties can be readily presented.

Theorem 7 (*Wald-type test statistic for k*)

$$T_W(k) := \left(\hat{k} - k\right)^2 I_{j,n}(k) \xrightarrow{d} \chi_1^2, \qquad \text{as } n \longrightarrow \infty.$$

Theorem 8 (*Rao's score-type test statistic for k*)

$$T_S(k) := \frac{U\left(k;\boldsymbol{S}_n^{(j)}\right)^2}{I_{j,n}(k)} \xrightarrow{d} \chi_1^2, \qquad \text{as } n \longrightarrow \infty.$$

Theorem 9 (*Likelihood ratio* (*LR*) *test statistic for k*)

$$T_L(k) := 2\left(L\left(\hat{k};\boldsymbol{S}_n^{(j)}\right) - L\left(k;\boldsymbol{S}_n^{(j)}\right)\right) \xrightarrow{d} \chi_1^2, \qquad \text{as } n \longrightarrow \infty,$$

where χ_1^2 *is* χ^2 *distribution with degree of freedom 1, and L is the logarithm of the likelihood function of k.*

For interval estimation of k, we propose two methods based on $T_W(k)$ in Theorem 7 and $T_L(k)$ in Theorem 9. The method based on $T_S(k)$ in Theorem 8 is dropped here since the computational burden of interval estimation based on Rao's score test statistics may be heavier than those of the other two methods. For test of hypothesis for k, we propose three methods, based on $T_W(k)$ in Theorem 7, $T_S(k)$ in Theorem 8 and $T_L(k)$ in Theorem 9.

Finally, some remarks on the computation for the proposed inferential methods for k are presented in the supplementary materials.

3 Inference for the scale parameter σ

In this section, we discuss methods of inference for the scale parameter σ. We specifically provide the estimator of σ, the score function of σ and asymptotic properties of some statistics related to the score function. We then use them to construct confidence intervals and hypothesis tests for the parameter σ.

3.1 Estimation of σ

Once we obtain the estimate of k, by using the method outlined in the preceding section, we can adopt the usual ML method to obtain the estimate of σ, by replacing the shape parameter k by its estimate $\hat{k}$.

The usual log-likelihood function of k and σ, based on $\boldsymbol{X} = \left(X_1, \dots, X_n\right)$, is given by

$$L_{k,\sigma}(k,\sigma;\boldsymbol{X}) = \begin{cases} -n \ln \sigma + \left(\frac{1}{k} - 1\right) \sum_{i=1}^{n} \ln \left(1 - k\frac{X_i}{\sigma}\right), & k \neq 0, \\ -n \ln \sigma - \frac{\sum_{i=1}^{n} X_i}{\sigma}, & k = 0; \end{cases}$$

see Hosking and Wallis (1987) and de Zea Bermudez and Kotz (2010a). Then, the MLE of σ is given by, after replacing k by $\hat{k}$,

$$\hat{\sigma} = \begin{cases} \text{Solution of} \quad n - \left(\frac{1}{\hat{k}} - 1\right) \sum_{i=1}^{n} \frac{\hat{k} X_i}{\sigma - \hat{k} X_i} = 0, \ \hat{k} < 0, \quad 0 < \hat{k} < 1, \\ \frac{1}{n} \sum_{i=1}^{n} X_i, & \hat{k} = 0, \\ \hat{k} X_{n:n}, & \hat{k} \geq 1. \end{cases}$$

It is known that the usual MLE of σ possesses large bias (Giles et al. (2016)). For $k < 1/2$, Giles et al. (2016) provided an analytical expression for the bias, to $O\left(n^{-1}\right)$, of the MLE of σ as follows.

Proposition 2 (*Bias of MLE of σ for k<1/2* (Giles et al. 2016))

$$E(\hat{\sigma}) - \sigma = \sigma\left(3 - 5k - 4k^2\right)/\{n(1 - 3k)\} + O\left(n^{-2}\right).$$

Proof See Giles et al. (2016). □

We adopt the bias correction based on Proposition 2 to the proposed estimator of σ only when $\hat{k} < 0$.

For $k \geq 1$, we readily have the following result for the expectation of $X_{n:n}$.

Proposition 3 (*Expectation of $X_{n:n}$ for $k \geq 1$*)

$$E\left(X_{n:n}\right) = \frac{\sigma}{k}\left(1 - \frac{n!\,k!}{(n+k)!}\right).$$

Proof For $k \geq 1$,

$$E\left(X_{n:n}\right) = \int_0^{\sigma/k} 1 - \{F(x;k,\sigma)\}^n\,dx = 1 - \frac{n!\,k!}{(n+k)!}.$$

□

By making use of Propositions 2 and 3, we propose the new bias-corrected estimator of σ as follows:

$$\hat{\sigma}_{BC} = \begin{cases} \hat{\sigma}\left[1 - \left(3 - 5\hat{k} - 4\hat{k}^2\right)/\left\{n\left(1 - 3\hat{k}\right)\right\}\right], & \hat{k} < 0, \\ \dfrac{\hat{\sigma}}{1 - \frac{n!\,k!}{(n+k)!}}, & \hat{k} \geq 1, \\ \hat{\sigma}, & \text{otherwise.} \end{cases}$$

3.2 Properties of the estimator of σ

In this subsection, we establish some properties of the proposed estimator of σ.

Theorem 10 *For $k \in \mathbb{R}$, $\sigma > 0$ and any given realized values $x_1, \ldots, x_n$ of $X_1, \ldots, X_n$, the proposed estimate of σ, when $\hat{k}$ is substituted for its realized value, always exists uniquely.*

Theorem 11 *The proposed estimator $\hat{\sigma}$ is consistent for σ.*

Proofs of Theorems 10 and 11 are given in the supplementary materials.

We now present some asymptotic properties of $\hat{\sigma}$ which are subsequently used to develop confidence intervals and hypothesis tests for σ. We first introduce the usual score function and the Fisher information matrix about k and σ in $\boldsymbol{X}$ for GPD, which are as follows:

$$\boldsymbol{U}(k,\sigma) = \frac{\partial}{\partial (k\,\sigma)'} L_{k,\sigma}(k,\sigma;\boldsymbol{X}) = \begin{pmatrix} \frac{\partial}{\partial k} L_{k,\sigma}(k,\sigma;\boldsymbol{X}) \\ \frac{\partial}{\partial \sigma} L_{k,\sigma}(k,\sigma;\boldsymbol{X}) \end{pmatrix} = \begin{pmatrix} U_k(k,\sigma) \\ U_\sigma(k,\sigma) \end{pmatrix},$$

where

$$U_k(k,\sigma) = \begin{cases} -\frac{1}{k^2}\sum_{i=1}^{n} \ln\left(1 - k\frac{X_i}{\sigma}\right) - \left(\frac{1}{k} - 1\right)\sum_{i=1}^{n} \frac{X_i}{\sigma - kX_i}, & k \neq 0, \\ 0, & k = 0, \end{cases}$$

$$U_\sigma(k,\sigma) = \begin{cases} -\frac{n}{\sigma} + \frac{1-k}{\sigma}\sum_{i=1}^{n} \frac{X_i}{\sigma - kX_i}, & k \neq 0, \\ -\frac{n}{\sigma} + \frac{\sum_{i=1}^{n} X_i}{\sigma^2}, & k = 0, \end{cases}$$

and

$$I_{k,\sigma}(k,\sigma) = \begin{pmatrix} i_{11} & i_{12} \\ i_{21} & i_{22} \end{pmatrix} = \begin{pmatrix} \frac{2n}{(1-k)(1-2k)} & -\frac{n}{(1-k)(1-2k)\,\sigma} \\ -\frac{n}{(1-k)(1-2k)\,\sigma} & \frac{n}{(1-2k)\,\sigma^2} \end{pmatrix}, \quad k < \frac{1}{2}.$$

We also have the inverse of the Fisher information matrix as

$$I^{-1}_{k,\sigma}(k,\sigma) = \begin{pmatrix} j_{11} & j_{12} \\ j_{21} & j_{22} \end{pmatrix} = \begin{pmatrix} \frac{(1-k)^2}{n} & \frac{\sigma\,(1-k)}{n} \\ \frac{\sigma\,(1-k)}{n} & \frac{2\,\sigma^2\,(1-k)}{n} \end{pmatrix}, \quad k < \frac{1}{2}.$$

Note that the Fisher information matrix and the inverse of the Fisher information matrix exist only when $k < 1/2$, and the usual MLEs of k and σ have asymptotic normality only when $k < 1/2$ (Smith 1984, 1985). From these, we can get the following asymptotic normality of $\hat{\sigma}$, when $k < 1/2$ and is known.

Proposition 4 (*Asymptotic distribution of $\hat{\sigma}$ when $k< 1/2$ and is known*)

$$W_0(k,\sigma) = \frac{\hat{\sigma} - \sigma}{\sqrt{j_{22}}} \xrightarrow{d} N(0,1), \qquad \text{as } n \longrightarrow \infty.$$

Proof It can be established from Theorem 4, the asymptotic normality of the MLE of σ, and by the use of Slutsky's theorem. □

From the above facts and Proposition 4, the following Wald, score and likelihood ratio (LR)-type statistics for the case when k is known and their asymptotic properties can be readily provided as follows.

Proposition 5 (*Wald-type test statistic for σ, when k< 1/2 and is known*)

$$W_W(k,\sigma) = W_0^2(k,\sigma) = \frac{(\hat{\sigma}-\sigma)^2}{j_{22}} \xrightarrow{d} \chi_1^2, \quad \text{as } n \longrightarrow \infty.$$

Proposition 6 (*Rao's score-type test statistic for σ, when k< 1/2 and is known*)

$$W_S(k,\sigma) = \frac{U_\sigma(k,\sigma)^2}{i_{22}} \xrightarrow{d} \chi_1^2, \quad \text{as } n \longrightarrow \infty.$$

Proposition 7 (*Likelihood ratio (LR) test statistic for σ, when k < 1/2 and is known*)

$$W_L(k,\sigma) = 2\left(L_{k,\sigma}(k,\hat{\sigma};\boldsymbol{X}) - L_{k,\sigma}(k,\sigma;\boldsymbol{X})\right) \xrightarrow{d} \chi_1^2, \quad \text{as } n \longrightarrow \infty.$$

We now present two methods of interval estimation of σ when $\hat{k} < 1/2$, where $\hat{k}$ is the proposed estimator of k, by the use of Propositions 5 and 7, wherein k is replaced by $\hat{k}$. The method based on Proposition 6 is dropped here for the same reason as for the interval estimation of k. We also have three methods of hypothesis testing for σ when $\hat{k} < 1/2$, where $\hat{k}$ is the proposed estimator of k, by the use of the Propositions 5, 6 and 7, wherein k is replaced by $\hat{k}$.

Note that the usual MLEs do not exist for $k \geq 1$. Therefore, we cannot get the asymptotic properties of the usual MLE of σ and consequently the confidence intervals and hypothesis tests cannot be developed based on them. However, based on the estimator of k proposed in the preceding section that exists for every k in $\mathbb{R}$, we can obtain the following asymptotic normality of the proposed estimator of σ even for the case when $k \geq 1$.

Theorem 12 (*Asymptotic distribution of $\hat{\sigma}$ when k ≥1 and is known*)

$$W_1(k,\sigma) := \frac{\hat{\sigma}-\sigma}{X_{n:n} \times \sqrt{I_{j,n}^{-1}(k)}} \xrightarrow{d} N(0,1), \quad \text{as } n \longrightarrow \infty.$$

Proof It can be proved from Theorems 6 and 11, and by the use of Slutsky's theorem. □

We propose the construction of confidence intervals and hypothesis tests for σ, with the use of Theorem 12, wherein k is replaced by $\hat{k}$, when $\hat{k} \geq 1$.

It is important to mention that we have not derived asymptotic properties of $\hat{\sigma}$, and also methods of interval estimation and hypothesis tests have not been established so far, for $1/2 \leq k < 1$. We performed Monte Carlo simulations and found that the interval estimation and hypothesis testing methods based on Theorem 12 turn out to be best for the case when $1/2 \leq \hat{k} < 1$.

Furthermore, in our Monte Carlo simulations, we found the score- and LR-type methods for interval estimation and hypothesis tests for σ do not perform well when $0 < k < 1/2$. One of the reasons might be that the support of the GPD depends on the unknown parameters k and σ and the regularity condition in the usual ML method is thus violated, when $0 < k < 1/2$. Meanwhile, the Wald-type interval estimation and hypothesis testing methods based on $W_W(k, \sigma)$ turns out to be best for the case when $0 < \hat{k} < 1/2$.

The proposed methods of interval estimation and hypothesis testing for k and σ are all summarized in Table 1.

4 Empirical evaluation of proposed methods of inference

In this section, we conduct extensive Monte Carlo simulation studies to investigate the finite-sample performance of the proposed methods of inference and also compare them with some other methods available in the literature for the GPD. All programs in this numerical study were written and run in Matlab.

4.1 Point estimation

In this subsection, we compare the proposed method (Proposed) of estimation of parameters k and σ with the following prominent methods: the usual maximum likelihood method (ML), the Bayesian method proposed by Zhang and Stephens (2009) (ZS), the Bayesian method proposed by Zhang (2010) (Zj) and the weighted minimum distance estimation method by Chen et al. (2017) (WMD). Figures 2 and 3 depict the simulation results of bias and root mean squared error ($RMSE = \sqrt{variance + bias^2}$) of the estimators of k and σ by all these methods, based on 10,000 Monte Carlo runs, for $-4 \leq k \leq 4$, and $n = 20$ and 100. We set $\sigma = 1$ since all considered estimators are equivariant for σ. Although the ML estimators exist when $k \leq 1$, their behavior becomes unstable when k gets close to 1 (see the simulation results of Chen et al. (2017)). For this reason, we show the bias and RMSE of the ML estimators only for $k \leq 0$.

From these results, we observe that the proposed method is quite successful in providing estimates of k and σ and also demonstrates good performance in terms of both bias and RMSE, for all considered values of the shape parameter k.

Table 1 Proposed interval estimation and hypothesis testing procedures

Inference for k	
$\hat{k} \in \mathbb{R}$	Wald-type method based on $T_W(k)$
	(Rao's score-type method based on $T_S(k)$)
	LR-type method based on $T_L(k)$
Inference for σ	
$\hat{k} \leq 0$	Wald-type method based on $W_W(k, \sigma)$
	(Rao's score-type method based on $W_S(k, \sigma)$)
	LR-type method based on $W_L(k, \sigma)$
$0 < \hat{k} < 1/2$	Wald-type method based on $W_W(k, \sigma)$
$\hat{k} \geq 1/2$	The method based on $W_1(k, \sigma)$

The methods in parentheses are only for hypothesis testing

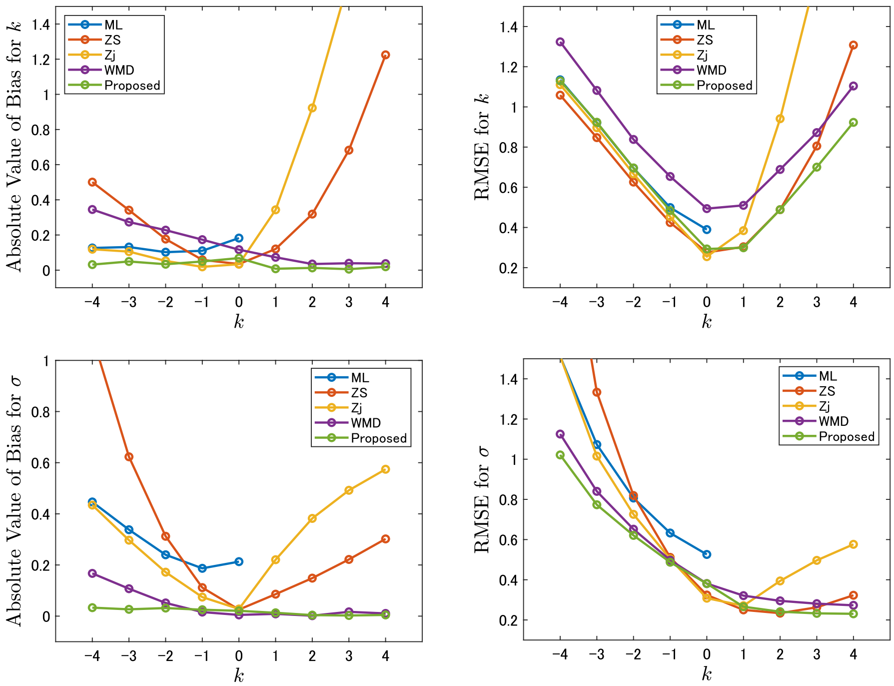

Fig. 2 Absolute value of bias and RMSE of estimators, based on 10,000 Monte Carlo simulations for $n = 20$ ($\sigma = 1$)

Furthermore, the proposed method of estimation of k and σ is seen to outperform in general the WMD method of Chen et al. (2017), based on both bias and RMSE. For estimation of k, the WMD method has higher RMSE than all other methods and does not perform very well in terms of bias, especially for negative k. Finally, it is seen that ML, ZS and Zj do not exist or do not have consistency property for certain ranges of k.

4.2 Interval estimation

In this subsection, we investigate the performances of the two proposed interval estimation methods based on Wald- and LR-type statistics (Proposed-Wald and Proposed-LR), detailed in Section 3.2. In this Monte Carlo simulation study, we compare the proposed confidence intervals (CIs) with the bootstrap-t CIs based on ML, ZS, Zj and WMD methods of parameter estimation. The number of bootstrap samples was set to 1,000. Figures 4 and 5 show the simulation results on the coverage probability (CP) and the average interval width of the 95% CIs of k and σ by all methods, based on 10,000 Monte Carlo runs, for the choices of $-4 \le k \le 4$, $\sigma = 1$, and $n = 20$ and 100.

Only the proposed methods (Proposed-Wald and Proposed-LR) and the boot-t method based on WMD successfully produce the CIs (in terms of CP) for all values

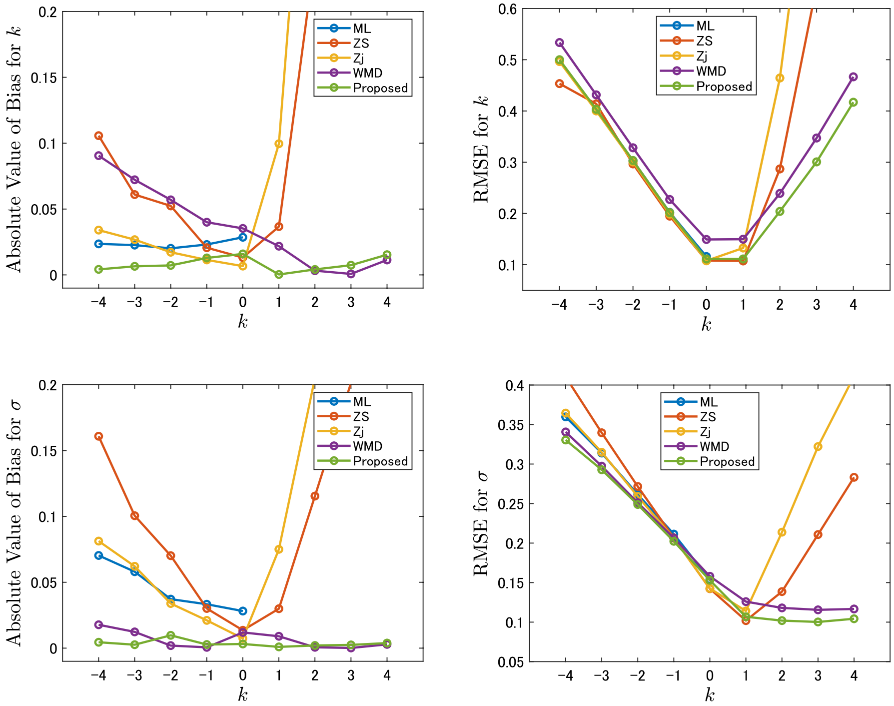

Fig. 3 Absolute value of bias and RMSE of estimators, based on 10,000 Monte Carlo simulations, for $n = 100$ ($\sigma = 1$)

of the shape parameter k. In addition, we see that the two proposed methods outperform the boot-t method based on WMD in terms of CP in general and the AIW (average interval width) in all considered cases. Furthermore, we observe that the Proposed-LR is superior to the Proposed-Wald method since the CPs of the former are better than those of the latter, while there is not much difference between the two methods in terms of AIW. Moreover, the boot-t methods based on ML, ZS and Zj do not work well when $k > 0$. From all these empirical results, we conclude that the Proposed-LR is the best method, in term of both these performance measures.

4.3 Hypothesis testing

In this subsection, we conduct simulation studies to evaluate the performance of the three proposed tests based on Wald (Proposed-Wald), score (Proposed-Score) and LR (Proposed-LR)-type statistics, introduced earlier in Section 3.2. In the Monte Carlo simulation studies, we estimated rejection probabilities based on 10,000 Monte Carlo runs. Unlike in the studies for interval estimation, we do not evaluate the hypothesis testing methods based on bootstrap for three reasons: (1) almost all authors who proposed parameter estimation methods did not suggest hypothesis tests based on bootstrap (see, for examples, de Zea Bermudez and Kotz (2010a,

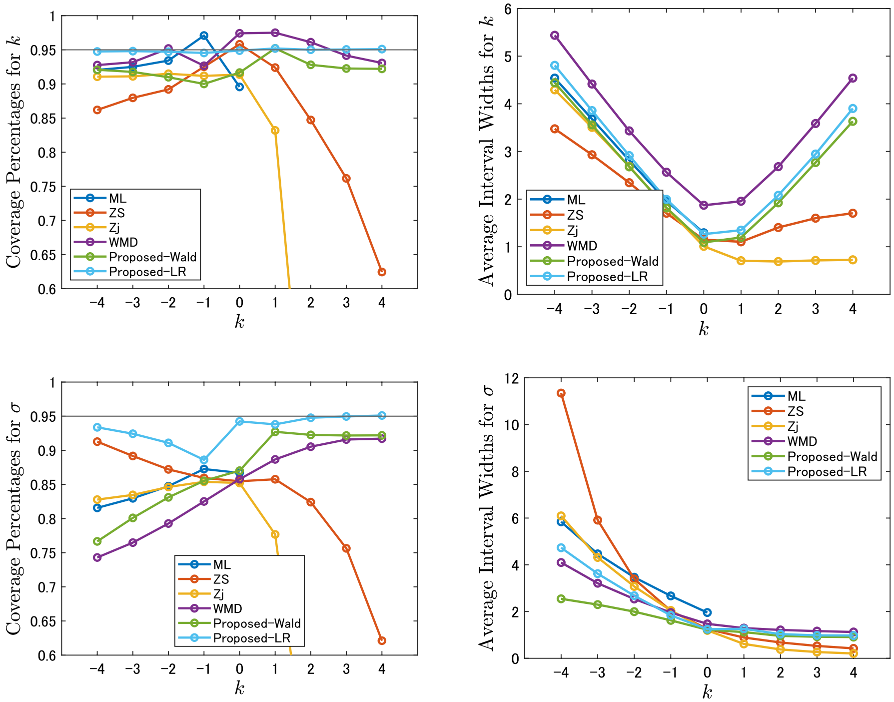

Fig. 4 The coverage probabilities and average interval widths for the parameters at nominal level 0.95, based on 10,000 Monte Carlo simulations, for $n = 20$ ($\sigma = 1$)

2010b), Zhang and Stephens (2009), Zhang (2010), del Castillo and Serra (2015), Chen et al. (2017)), (2) comprehensive Monte Carlo studies become infeasible due to the heavy computational burden of the bootstrap method, and (3) the results of hypothesis tests may be predicted reasonably from those of interval estimation, that is, the proposed methods would perform better than the bootstrap methods based on the existing parameter estimation methods. So, in this part of Monte Carlo studies, we focus on which method of hypothesis testing among the three proposed methods performs the best.

First, we conduct simulations for tests of hypothesis for k. Figures 6 and 7 show the rejection rates of the tests under $H_0 : k = a$ versus $H_1 : k \neq a$, where $a = -2, -1, 0, 1, 2$, for different true values of k, $n = 20$ and 100. We take the true values of k as $-4 \leq k \leq 4$. We set $\sigma = 1$, and the significance level to be 0.05.

We observe that all three tests have similar power, and also their levels are close to the nominal significance level. Yet, the Proposed-LR displays best performance with respect to the level among the three tests, and is therefore the one we recommend for testing the shape parameter k.

Next, we conduct simulations for tests of hypothesis for σ. Figures 8 and 9 show the rejection rates of the tests under $H_0 : \sigma = 100$ versus $H_1 : \sigma \neq 100$, for different true values of σ, $n = 20$ and 100. We take various true values of σ (see the horizontal axes in figures for details). All tests depend on the estimates of k. Therefore, we

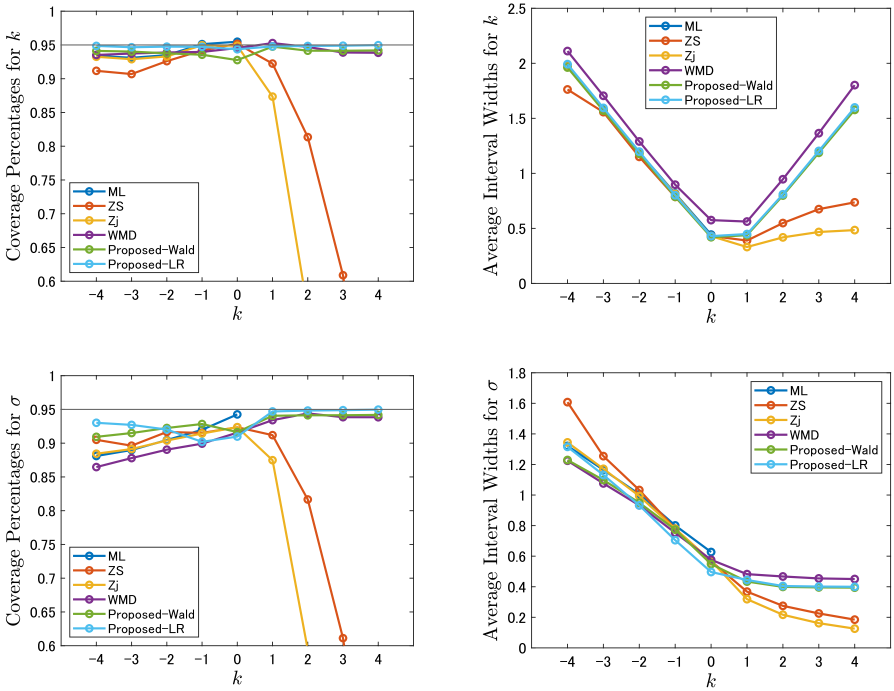

Fig. 5 The coverage probabilities and average interval widths for the parameters at nominal level 0.95, based on 10,000 Monte Carlo simulations, for $n = 100$ ($\sigma = 1$)

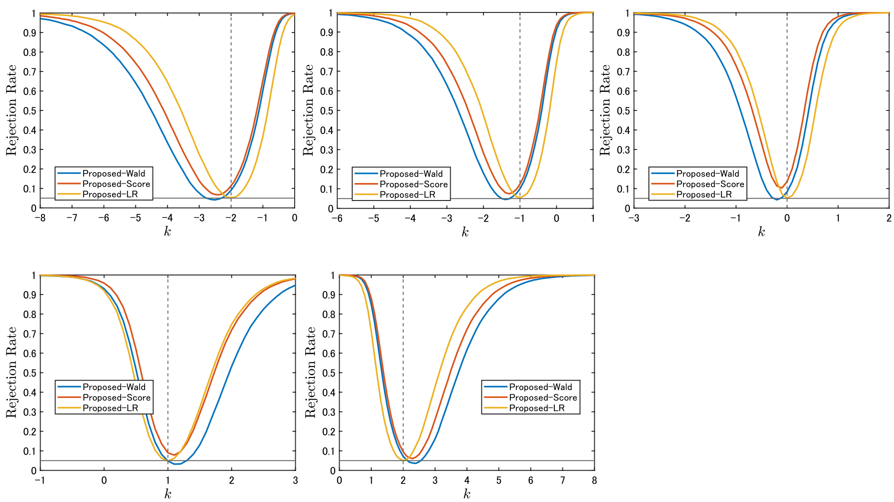

Fig. 6 Empirical power of the Wald-, score- and LR-type tests for shape parameter, under the stated hypotheses with significance level 5%, based on 10,000 Monte Carlo simulations, for $n = 20$

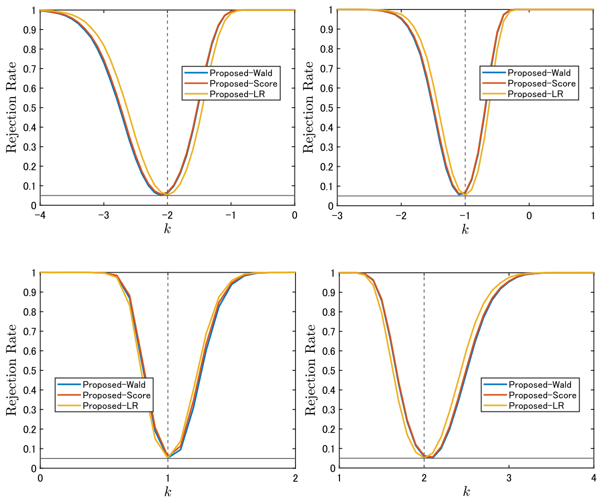

Fig. 7 Empirical power of the Wald-, score- and LR-type tests for shape parameter, under the stated hypotheses with significance level 5%, based on 10,000 Monte Carlo simulations, for $n = 100$

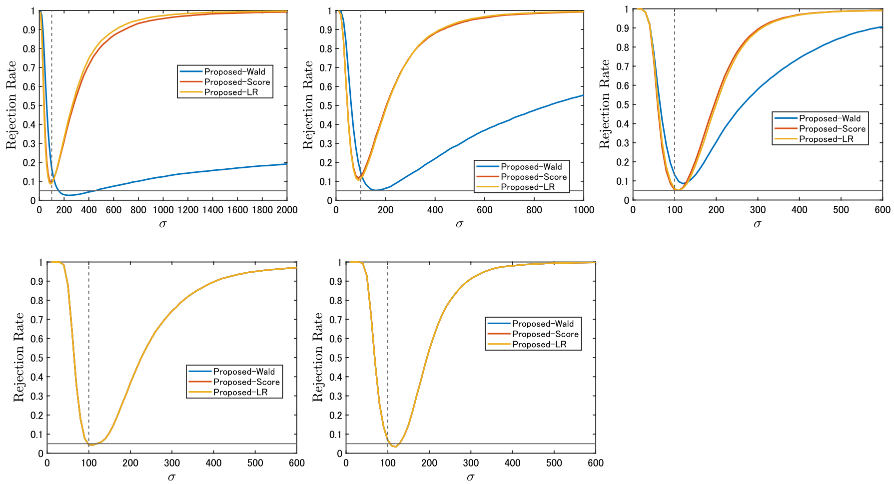

Fig. 8 Empirical power of the Wald-, score- and LR-type tests for scale parameter, based on 10,000 Monte Carlo simulations, for $n = 20$

conduct the simulation of hypothesis tests, for different true values of k, taken to be $-2, -1, 0, 1$, and 2. Once again, the significance level is set to be 0.05.

We find that the Proposed-Score and Proposed-LR have very similar performance in terms of power and levels, and their levels are close to the nominal level. Meanwhile, the performance of the Proposed-Wald test is worse than those of the other two when $k < 1$. Note that when $k = 1$ and 2, the performances of all three tests are same since the test statistics are the same for all the methods when $k > 0$.

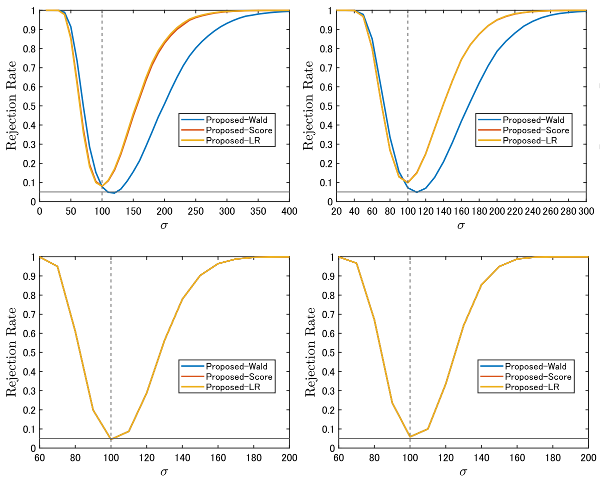

Fig. 9 Empirical power of the Wald-, score- and LR-type tests for scale parameter, based on 10,000 Monte Carlo simulations, for $n = 100$

Based on all the above findings, we recommend the use of the Proposed-Score and Proposed-LR tests for test concerning the scale parameter σ.

5 Illustrative examples

We demonstrate the proposed inferential procedures using two well-known real datasets: one is the Bilbao waves data, discussed by Castillo and Hadi (1997), and another is the fatigue data, by Andrews and Herzberg (1985). The latter example is presented in the supplementary materials.

5.1 Bilbao Waves Data

In the first example, we fit the GPD to the Bilbao waves data. These data gave the zero-crossing hourly mean periods (in seconds) of the sea waves measured in a Bilbao bay in January 1997, Spain, initially provided by Castillo and Hadi (1997) and later reanalyzed by many authors including Zhang and Stephens (2009), del Castillo and Serra (2015), and Chen et al. (2017). For these data, the GPD is found to provide a good fit, and all estimates of the shape parameter k are larger than 0.5. The purpose of analyzing these data is to study the influence of periods on beach morphodynamics and other problems related to the right tail, and so only observations above 7 seconds are taken, which are displayed in Table 2.

In Tables 3 and 4, the estimates of k and σ by the ML, ZS, Zj, WMD and the proposed method (Proposed) are presented for each threshold. The thresholds are taken as $u = 7.0, 7.5, 8.0, 8.5, 9.0$, and 9.5, same as those in Castillo and Hadi (1997). As several authors have reported, the ML estimates are not found when $u \geq 8.5$.

Table 2 The Bilbao waves data: the zero-crossing hourly mean periods (in seconds) of the sea waves measured in Bilbao Buoy in January 1997

7.05	7.12	7.15	7.18	7.19	7.20	7.20	7.20	7.20	7.25
7.26	7.27	7.28	7.30	7.31	7.31	7.32	7.33	7.37	7.40
7.46	7.46	7.47	7.48	7.48	7.52	7.54	7.55	7.55	7.58
7.59	7.59	7.61	7.63	7.65	7.66	7.66	7.67	7.67	7.68
7.69	7.72	7.72	7.72	7.72	7.72	7.77	7.77	7.79	7.79
7.82	7.83	7.83	7.83	7.84	7.85	7.85	7.88	7.88	7.90
7.90	7.91	7.93	7.93	7.93	7.94	7.95	7.95	7.97	7.97
7.97	7.99	8.00	8.03	8.03	8.05	8.06	8.06	8.07	8.10
8.11	8.12	8.15	8.15	8.15	8.18	8.18	8.18	8.19	8.20
8.21	8.23	8.23	8.30	8.30	8.31	8.31	8.32	8.32	8.33
8.40	8.41	8.42	8.43	8.43	8.45	8.48	8.49	8.50	8.50
8.51	8.52	8.53	8.54	8.56	8.58	8.59	8.59	8.60	8.65
8.69	8.71	8.72	8.74	8.74	8.74	8.74	8.79	8.81	8.84
8.85	8.86	8.88	8.88	8.94	8.98	8.98	8.99	9.01	9.03
9.06	9.12	9.16	9.17	9.17	9.18	9.18	9.18	9.21	9.22
9.23	9.24	9.27	9.29	9.30	9.32	9.33	9.36	9.38	9.43
9.46	9.47	9.59	9.59	9.60	9.61	9.62	9.63	9.66	9.74
9.75	9.78	9.79	9.79	9.80	9.84	9.85	9.89	9.90	

Table 3 Estimates of k for the Bilbao waves data: u is the threshold and m is the number of exceedances; — indicates nonexistence of the corresponding estimate

u	m	ML	ZS	Zj	WMD	Proposed
7.0	179	0.861	0.808	0.782	0.831	0.822
7.5	154	0.768	0.706	0.686	0.602	0.725
8.0	106	0.864	0.768	0.731	0.668	0.797
8.5	69	–	0.833	0.767	0.771	0.877
9.0	41	–	0.878	0.760	0.877	0.942
9.5	17	–	1.010	0.736	1.274	1.122

Table 4 Estimates of σ for the Bilbao waves data: u is the threshold and m is the number of exceedances; — indicates nonexistence of the corresponding estimate

u	m	ML	ZS	Zj	WMD	Proposed
7.0	179	2.501	2.382	2.331	2.436	2.415
7.5	154	1.860	1.753	1.722	1.621	1.781
8.0	106	1.648	1.508	1.462	1.406	1.549
8.5	69	–	1.208	1.146	1.165	1.256
9.0	41	–	0.826	0.756	0.836	0.873
9.5	17	–	0.430	0.361	0.515	0.468

To measure the overall goodness of fit, we use the average scaled absolute error (ASAE), defined in Castillo and Hadi (1997) as

$$\text{ASAE} = \frac{1}{n}\sum_{i=1}^{n}\frac{|x_{i:n}-\hat{x}_{i:n}|}{x_{n:n}-x_{1:n}},$$

where $\hat{x}_{i:n} = \hat{\sigma}\left[1-(1-i/(n+1))^{\hat{k}}\right]/\hat{k}$, and $\hat{k}$ and $\hat{\sigma}$ are the estimates of k and σ, respectively.

The values of ASAE for the ML, ZS, Zj, WMD and proposed methods are presented in Table 5.

The proposed method is comparable with the ZS, Zj and WMD methods with respect to ASAE, for all cases of u, although slightly superior in most of the cases. Tables 6 and 7 show the 95% the confidence intervals for k and σ by

Table 5 ASAE for the Bilbao waves data: u is the threshold and m is the number of exceedances; — indicates nonexistence of the corresponding estimate

u	m	ML	ZS	Zj	WMD	Proposed
7.0	179	0.0298	0.0257	0.0246	0.0275	0.0267
7.5	154	0.0262	0.0185	0.0169	0.0124	0.0203
8.0	106	0.0307	0.0181	0.0152	0.0132	0.0214
8.5	69	–	0.0200	0.0188	0.0179	0.0232
9.0	41	–	0.0334	0.0333	0.0326	0.0347
9.5	17	–	0.0698	0.0974	0.0629	0.0636

Table 6 CIs of k for the Bilbao waves data: u is the threshold and m is the number of exceedances

u	m	ZS	Zj	WMD	Proposed-Wald	Proposed-LR
7.0	179	(0.692, 0.946)	(0.697, 0.929)	(0.672, 1.056)	(0.682, 0.962)	(0.686, 0.971)
7.5	154	(0.587, 0.854)	(0.595, 0.838)	(0.456, 0.844)	(0.570, 0.879)	(0.564, 0.884)
8.0	106	(0.610, 0.961)	(0.629, 0.932)	(0.480, 0.962)	(0.604, 0.990)	(0.602, 1.004)
8.5	69	(0.625, 1.094)	(0.653, 1.035)	(0.503, 1.176)	(0.622, 1.132)	(0.629, 1.163)
9.0	41	(0.598, 1.258)	(0.646, 1.134)	(0.497, 1.528)	(0.583, 1.301)	(0.601, 1.368)
9.5	17	(0.519, 1.738)	(0.641, 1.368)	(0.297, 2.680)	(0.448, 1.796)	(0.539, 2.050)

Table 7 CIs of σ for the Bilbao waves data: u is the threshold and m is the number of exceedances

u	m	ZS	Zj	WMD	Proposed-Wald	Proposed-LR
7.0	179	(2.041, 2.767)	(2.052, 2.728)	(2.002, 2.898)	(1.978, 2.788)	(1.991, 2.817)
7.5	154	(1.475, 2.062)	(1.489, 2.035)	(1.292, 1.952)	(1.369, 2.109)	(1.354, 2.122)
8.0	106	(1.216, 1.836)	(1.248, 1.793)	(1.066, 1.750)	(1.148, 1.881)	(1.144, 1.908)
8.5	69	(0.920, 1.535)	(0.952, 1.470)	(0.815, 1.527)	(0.871, 1.584)	(0.880, 1.629)
9.0	41	(0.574, 1.117)	(0.620, 1.031)	(0.506, 1.196)	(0.525, 1.171)	(0.541, 1.231)
9.5	17	(0.240, 0.660)	(0.286, 0.555)	(0.142, 0.842)	(0.179, 0.718)	(0.215, 0.820)

Table 8 P-values under $H_0 : k \leq 0.5$ for the proposed methods for the Bilbao waves data: u is the threshold and m is the number of exceedances

u	m	Proposed-Wald	Proposed-Score	Proposed-LR
7.0	179	0.000	0.000	0.000
7.5	154	0.004	0.002	0.008
8.0	106	0.003	0.000	0.004
8.5	69	0.004	0.000	0.004
9.0	41	0.016	0.000	0.013
9.5	17	0.070	0.000	0.037

bootstrap-t methods based on the ZS, Zj, WMD and the proposed methods based on Wald-type statistic (Proposed-Wald) and LR-type statistic (Proposed-LR).

In addition, we demonstrate the three proposed tests for k. We will be naturally interested in testing whether $k > 0.5$ or not, since the asymptotic properties of MLEs are valid only when $k < 0.5$. So, we consider the hypothesis $H_0 : k \leq 0.5$ vs. $H_1 : k > 0.5$. Under the null hypothesis $H_0 : k \leq 0.5$, we carry out the three proposed tests. The p-value at $k = 0.5$ is the largest among $H_0 : k \leq 0.5$, for all the cases. The resulting p-values under $H_0 : k \leq 0.5$ are presented in Table 8 for each threshold level. All p-values, except for the Proposed-Wald when $u = 9.5$, are less than 0.05, providing evidence to k being larger than 0.5. So, there is a necessity for methods other than the ML method for analyzing these data.

6 Concluding remarks

The GPD is one of the most important distributions in modeling extreme values. However, in spite of many papers dealing with GPD, there has been little work done on inferential procedures established formally for all values of the shape parameter k. In this article, we develop a new framework for efficient inference for GPD, which works well for all possible k. Specifically, we propose a new likelihood-based method of parameter estimation for GPD and establish asymptotic properties of the estimators and some related statistics. We prove that the estimates by the proposed method always exist uniquely for all k. Based on the asymptotic properties of the proposed estimators and of related statistics, we develop confidence intervals and hypothesis tests for the GPD, based on Wald-, score- and LR-type statistics, and these are shown to work satisfactorily for all k. The simulation results show that the proposed methods, for both point and interval estimation, outperformed in general a number of prominent existing methods, and also that the proposed hypothesis tests perform well for all k, especially those based on score- and LR-type statistics, even for small sample sizes.

As mentioned in Sect. 3.2, for inference for σ, we have not provided the asymptotic properties of the proposed estimator, when $1/2 \leq \hat{k} < 1$, and so we adopted the results for $\hat{k} \geq 1$ in this case too. Although the simulation study shows that this adopted method works well for k in the range [1/2, 1), the asymptotic properties of the proposed estimator of σ remain open for this case. We have discussed here only the complete data case, and the approach for the case of censored data will be of great interest. Our approach can be generalized to other distributions possessing non-regularity problems such as the generalized extreme value distribution. We hope to consider these problems in our future research.

Acknowledgements The authors thank the Associate Editor and two referees for their incisive comments and suggestions which led to a great improvement in the paper. Hideki Nagatsuka was partially supported by the Grant-in-Aid for Scientific Research (C) 19K04890, Japan Society for the Promotion of Science, and Chuo University Grant for Special Research, while N. Balakrishnan was supported by an Individual Discovery Grant (RGPIN-2020-06733) from the Natural Sciences and Engineering Research Council of Canada.

Propositions for derivatives of the likelihood function of *k*

Proposition 8 *For $k \in \mathbb{R}$ and any given $\boldsymbol{s}_n^{(j)}$, where j, $1 \le j \le n$, is fixed, the derivative $l'(k;\boldsymbol{s}_n^{(j)}) = (\partial/\partial k)l(k;\boldsymbol{s}_n^{(j)})$ is given by*

$$
\begin{aligned}
&l'(k;\boldsymbol{s}_n^{(j)})\\
&=\begin{cases}
n!\displaystyle\int_{\chi_k}\left(-\frac{n}{k}-\frac{\sum_{i=1}^n \log\left(1-us_i\right)}{k^2}\right)\frac{1}{|k|}\left(\frac{u}{k}\right)^{n-1}\prod_{i=1}^n\left(1-us_i\right)^{1/k-1}du, & k\neq 0,\\
\left\{1-\dfrac{(n+1)\sum_{i=1}^n s_i^2}{2\left(\sum_{i=1}^n s_i\right)^2}\right\}\dfrac{(n!)^2}{\left(\sum_{i=1}^n s_i\right)^n}, & k=0,
\end{cases}\\
&\quad s_1\le\cdots\le s_{j-1}\le 1\le s_{j+1}\le\cdots\le s_n,\ \text{and } s_j=1.
\end{aligned}
\tag{5}
$$

Proposition 9 *For $k \in \mathbb{R}$ and any given $\boldsymbol{s}_n^{(j)}$, where j, $1 \le j \le n$, is fixed, the second derivative $l''(k;\boldsymbol{s}_n^{(j)}) = (\partial^2/\partial k^2)l(k;\boldsymbol{s}_n^{(j)})$ is given by*

$$
\begin{aligned}
&l''(k;\boldsymbol{s}_n^{(j)})\\
&=\begin{cases}
n!\displaystyle\int_{\chi_k}\left(\frac{n(n+1)}{k^2}+\frac{2(n+1)\sum_{i=1}^n\log\left(1-us_i\right)}{k^3}+\frac{\left\{\sum_{i=1}^n\log\left(1-us_i\right)\right\}^2}{k^4}\right)\\
\times\dfrac{1}{|k|}\left(\dfrac{u}{k}\right)^{n-1}\displaystyle\prod_{i=1}^n\left(1-us_i\right)^{1/k-1}du, & k\neq 0,\\
\left\{1-(n+1)\dfrac{\sum_{i=1}^n s_i^2}{\left(\sum_{i=1}^n s_i\right)^2}+\dfrac{(n+2)(n+3)}{4}\dfrac{\left(\sum_{i=1}^n s_i^2\right)^2}{\left(\sum_{i=1}^n s_i\right)^4}-\dfrac{2(n+2)}{3}\dfrac{\sum_{i=1}^n s_i^3}{\left(\sum_{i=1}^n s_i\right)^3}\right\}\\
\times\dfrac{n!\,(n+1)!}{\left(\sum_{i=1}^n s_i\right)^n}, & k=0,
\end{cases}\\
&s_1\le\cdots\le s_{j-1}\le 1\le s_{j+1}\le\cdots\le s_n,\ \text{and } s_j=1.
\end{aligned}
$$

Proposition 10 *For $k \in \mathbb{R}$ and any given $\boldsymbol{s}_n^{(j)}$, where j, $1 \le j \le n$, is fixed, the third derivative $l'''(k;\boldsymbol{s}_n^{(j)}) = (\partial^3/\partial k^3)l(k;\boldsymbol{s}_n^{(j)})$ is given by*

$$
\begin{aligned}
&l'''(k;\boldsymbol{s}_n^{(j)}) \\
&\quad= n! \int_{\chi_k} \left(-\frac{n(n+1)(n+2)}{k^3} - \frac{3(n+1)(n+2)\sum_{i=1}^n \log\left(1-us_i\right)}{k^4} \right. \\
&\qquad \left. -\frac{3(n+2)\left\{\sum_{i=1}^n \log\left(1-us_i\right)\right\}^2}{k^5} - \frac{\left\{\sum_{i=1}^n \log\left(1-us_i\right)\right\}^3}{k^6} \right) \\
&\qquad \times \frac{1}{|k|}\left(\frac{u}{k}\right)^{n-1} \prod_{i=1}^n \left(1-us_i\right)^{1/k-1} du, \quad \text{for } k \neq 0,
\end{aligned} \tag{6}
$$

and

$$
\begin{aligned}
&l'''(0;\boldsymbol{s}_n^{(j)}) \\
&\quad= \left\{ 1 - \frac{3(n+1)}{2}\frac{\sum_{i=1}^n s_i^2}{\left(\sum_{i=1}^n s_i\right)^2} - \frac{(n+3)(n+4)(n+5)}{8}\frac{\left(\sum_{i=1}^n s_i^2\right)^3}{\left(\sum_{i=1}^n s_i\right)^6} \right. \\
&\qquad - 2(n+2)\frac{\sum_{i=1}^n s_i^3}{\left(\sum_{i=1}^n s_i\right)^3} + (n+3)(n+4)\frac{\left(\sum_{i=1}^n s_i^2\right)\left(\sum_{i=1}^n s_i^3\right)}{\left(\sum_{i=1}^n s_i\right)^5} \\
&\qquad \left. + \frac{3(n+2)(n+3)}{4}\frac{\left(\sum_{i=1}^n s_i^2\right)^2}{\left(\sum_{i=1}^n s_i\right)^4} - \frac{3(n+3)}{2}\frac{\sum_{i=1}^n s_i^4}{\left(\sum_{i=1}^n s_i\right)^4} \right\} \frac{n!\,(n+2)!}{\left(\sum_{i=1}^n s_i\right)^n},
\end{aligned} \tag{7}
$$

where $s_1 \leq \cdots \leq s_{j-1} \leq 1 \leq s_{j+1} \leq \cdots \leq s_n$, *and* $s_j = 1$.

Proofs

Proof of Proposition 1

Denote the cdf and pdf of the GPD with $\sigma = 1$, $F(\cdot;k,1)$ and $f(\cdot;k,1)$, by $G(\cdot\,;k)$ and $g(\cdot\,;\beta)$, respectively, for simplicity. Suppose Z_i, $i = 1,\dots,n$, are n independent random variables from such a standard GPD with shape parameter k. For $i = 1,\dots,n$, let $Z_{i:n}$ be the i-th order statistic among $Z_1,\dots,Z_n$.

First, we assume that $k \neq 0$. Define $\boldsymbol{i}_{\boldsymbol{n}}^{(j)} = \{i | i = 1,\dots,j-1,j+1,\dots,n\}$. For a fixed positive integer value j, and for any $n-1$ real values $s_1 \leq \cdots \leq s_{j-1} \leq 1 \leq s_{j+1} \leq \cdots \leq s_n$, we consider

$$
\begin{aligned}
&\Pr\left(S_{i:n}^{(j)} \le s_i,\ i \in \boldsymbol{i}_{\boldsymbol{n}}^{(j)}\right) \\
&\quad = \Pr\left(\frac{Z_{i:n}}{Z_{j:n}} \le s_i,\ i \in \boldsymbol{i}_{\boldsymbol{n}}^{(j)}\right) \\
&\quad = \int_{\mathscr{X}_{k,1}} \Pr\left(Z_{i:n} \le us_i,\ i \in \boldsymbol{i}_{\boldsymbol{n}}^{(j)} \middle| Z_{j:n} = u\right) h_j(u;k)\, du \\
&\quad = \int_{\mathscr{X}_{k,1}} (j-1)! \prod_{i=1}^{j-1} \frac{G(us_i\,;k)}{G(u\,;k)} \times (n-j)! \prod_{i=j+1}^{n} \frac{G(us_i\,;k)}{1-G(u\,;k)} \\
&\qquad \times \frac{n!}{(j-1)!(n-j)!} \{G(u\,;k)\}^{j-1} \{1-G(u\,;k)\}^{n-j} g(u\,;k)\, du \\
&\quad = \int_{\mathscr{X}_{k,1}} n!\, g(u\,;k) \prod_{i \in \boldsymbol{i}_{\boldsymbol{n}}^{(j)}} G(us_i\,;k)\, du,
\end{aligned}
\tag{8}
$$

where $h_j(\cdot;k)$ is the pdf of $Z_{j:n}$.

We note that the integrand in Eq. (8) has its partial derivative with respect to s_i, $i \in \boldsymbol{i}_{\boldsymbol{n}}^{(j)}$, as $n!g(u\,;k) \prod_{i\in \boldsymbol{i}_{\boldsymbol{n}}^{(j)}} u\, g(us_i\,;k)$, and further that

$$
(j-1)! \prod_{i=1}^{j-1} \frac{u\, g(us_i\,;k)}{G(u\,;k)} \times (n-j)! \prod_{i=j+1}^{n} \frac{u\, g(us_i\,;k)}{1-G(u\,;k)},
\tag{9}
$$

is bounded above. From the boundedness of (9), we have

$$
n!g(u\,;k) \prod_{i \in \boldsymbol{i}_{\boldsymbol{n}}^{(j)}} u\, g(us_i\,;k) \le C_0\, h_j(u;k),
$$

where C_0 is a positive constant, and

$$
\int_{\mathscr{X}_{k,1}} n!g(u\,;k) \prod_{i \in \boldsymbol{i}_{\boldsymbol{n}}^{(j)}} u\, g(us_i\,;k)\, du \le \int_{\mathscr{X}_{k,1}} C_0\, h_j(u;k)\, du = C_0 < \infty.
$$

Then, upon using Part (ii) of Theorem 16.8 of Billingsley (1994), we can interchange the derivatives and the integration in (8), so that the partial derivative of it with respect to $s_i, i \in \boldsymbol{i}_{\boldsymbol{n}}^{(j)}$, as

$$
n! \int_{\mathscr{X}_{k,1}} g(u\,;k) \prod_{i \in \boldsymbol{i}_{\boldsymbol{n}}^{(j)}} u\, g(us_i\,;k)\, du.
$$

The result for $k = 0$ can be obtained by letting $k \to 0$ in the result for $k \ne 0$. After some simple algebra, the proof of Proposition 1 gets completed. □

Proof of Theorem 3

First, we shall show that the likelihood equation has at least one solution. Given $s_n^{(j)}$, the derivative of the likelihood function in (5) for $k \neq 0$ can be rewritten as

$$l'(k;\boldsymbol{s}_n^{(j)}) = n! \int_{\chi_k} \eta(k,u)\Lambda(k,u)\,du,$$

where $\eta(k,u) = -\frac{n}{k} - \frac{\sum_{i=1}^n \log(1-us_i)}{k^2}$, $\Lambda(k,u) = \frac{1}{|k|}\left(\frac{u}{k}\right)^{n-1}\prod_{i=1}^n \left(1-us_i\right)^{1/k-1}$ and $s_j = 1$. It follows from the facts that $\eta(k,u) = -\frac{1}{k}\left(n + \frac{\sum_{i=1}^n \log(1-us_i)}{k}\right) > 0$ for sufficiently small $k \in \mathbb{R}$, and $\eta(k,u) < 0$ for sufficiently large $k \in \mathbb{R}$, and $\Lambda(k,u) > 0$ for every $k \in \mathbb{R}$, for every $u \in \chi_k$, there exist real values δ_1 and δ_2 such that $l'(k;\boldsymbol{s}_n^{(j)}) > 0$ for every $k < \delta_1$, and $l'(k;\boldsymbol{s}_n^{(j)}) < 0$ for every $k > \delta_2$, respectively. In addition, we see from Proposition 8 that $l'(k;\boldsymbol{s}_n^{(j)})$ is continuous with respect to $k \in \mathbb{R}$. Thus, $l'(k;\boldsymbol{s}_n^{(j)}) = 0$ has at least one solution.

Next, we shall show that the number of solutions is exactly one. Let k^* be one of the solutions of $l'(k;\boldsymbol{s}_n^{(j)}) = 0$. We see that $\eta(k^*,u)$ is strictly increasing in u and takes on values over $(-\infty, -n/k^*)$ for $k^* < 0$. Thus, there exists a unique value of u such that $\eta(k^*,u) = 0$, which we denote by u_0. We also see that $\eta(k^*,u) < 0$ for $u < u_0$ and $\eta(k^*,u) > 0$ for $u > u_0$.

We have, for $k^* < 0$ and sufficiently small $\Delta k > 0$ such that $k^* + \Delta k < 0$,

$$\begin{aligned} l'(k^*+\Delta k;s_n^{(j)}) =& n! \int_{-\infty}^{0} \eta(k^*,u)\frac{\eta(k^*+\Delta k,u)}{\eta(k^*,u)}\Lambda(k^*+\Delta k,u)\,du \\ <& n! \int_{-\infty}^{u_0-} \eta(k^*,u)\left(1+\frac{\Delta k}{k^*}\right)^{-2}\Lambda(k^*+\Delta k,u)\,du \\ &+ n! \int_{u_0}^{0} \eta(k^*,u)\left(1+\frac{\Delta k}{k^*}\right)^{-2}\Lambda(k^*+\Delta k,u)\,du \\ =& \left(1+\frac{\Delta k}{k^*}\right)^{-2} n! \int_{-\infty}^{0} \eta(k^*,u)\Lambda(k^*+\Delta k,u)\,du, \end{aligned} \tag{10}$$

where $u_0- = \lim_{u\uparrow u_0} u$, and the inequality follows from the facts that

$$\frac{\eta(k^*+\Delta k,u)}{\eta(k^*,u)} = \left(1+\frac{\Delta k}{k^*}\right)^{-2}\left\{1 + \frac{\Delta k}{k^* + \frac{1}{n}\sum_{i=1}^n \log\left(1-us_i\right)}\right\}$$

is greater than $(1+\Delta k/k^*)^{-2}$ for $u \in (-\infty, u_0)$, is less than $(1+\Delta k/k^*)^{-2}$ for $u \in (u_0, 0)$, and $\eta\left(k^*+\Delta k, u_0\right) < 0 = \eta\left(k^*, u_0\right)(1+\Delta k/k^*)^{-2}$.

We further note that

$$\frac{\Lambda(k^*+\Delta k,u)}{\Lambda(k^*,u)} = \left(1+\frac{\Delta k}{k^*}\right)^{n}\prod_{i=1}^{n}\left(1-u\,s_i\right)^{\frac{1}{k^*+\Delta k}-\frac{1}{k^*}}$$

is strictly increasing in u and takes on value over $(0, (1 + \Delta k/k^*)^n)$. Then, it follows from (10) and by the mean value theorem that

$$\begin{aligned} l'(k^* + \Delta k; \boldsymbol{s}_n^{(j)}) <& \left(1 + \frac{\Delta k}{k^*}\right)^{-2} M\, n! \int_{-\infty}^{0} \eta(k^*, u)\Lambda(k^*, u)\, du \\ =& \left(1 + \frac{\Delta k}{k^*}\right)^{-2} M\, n!\, l'(k^*; \boldsymbol{s}_n^{(j)}) = 0, \end{aligned}$$

where $M = \Lambda\big(k^* + \Delta k, u'\big)/\Lambda\big(k^*, u'\big) \in (0, (1 + \Delta k/k^*)^n)$, for $u' \in (-\infty, 0)$.

We can also obtain the same results for $k^* \geq 0$. The proofs are very similar to the proof for $k^* < 0$ (for $k^* = 0$, by using the fact that $l'(0; \boldsymbol{s}_n^{(j)}) = \lim_{k \to 0} l'(k; \boldsymbol{s}_n^{(j)})$ and Lebesgue's dominated convergence theorem) and are therefore omitted here. The fact that $l'(k^* + \Delta k; \boldsymbol{s}_n^{(j)}) < 0$ for every $k^* \neq 0$ clearly implies that $l'(k; \boldsymbol{s}_n^{(j)})$ changes sign only once with respect to k.

From the above arguments, $l'(k; \boldsymbol{s}_n^{(j)}) = 0$ always has a unique solution with respect to k, and the proof of Theorem 3 thus gets completed. □

Proof of Lemma 1

Let $\boldsymbol{S}_{n,1}^{(j)} = \left(S_{1:n}^{(j)}, \ldots, S_{j-1:n}^{(j)}\right)$ and $\boldsymbol{S}_{n,2}^{(j)} = \left(S_{j+1:n}^{(j)}, \ldots, S_{n:n}^{(j)}\right)$. Then, by Theorem 2 of Iliopoulos and Balakrishnan (2009), conditional on $Z_{j:n} = u \in \lambda_k$, where $Z_{j:n} = X_{j:n}/\sigma_0$ and $\lambda_k = \{u : 0 < u < \infty,\ \text{if } k < 0,\ \text{or}\ 0 < u < 1/k,\ \text{if } k > 0\}$ as defined in the proof of Proposition 1, we see that $\boldsymbol{S}_{n,1}^{(j)}$ are distributed exactly as order statistics from a sample of size $j-1$ from the distribution with density $\psi_1(s; k_0, u) = u\, g(u\, s; k_0)/G(u; k_0),\ 0 \leq s \leq 1$, and $\boldsymbol{S}_{n,2}^{(j)}$ are distributed exactly as order statistics from a sample of size $n-j$ from the distribution with density $\psi_2(s; k_0, u) = u\, g(u\, s; k_0)/(1 - G(u; k_0)),\quad s \geq 1,$ where $g(\cdot; k) = f(\cdot; k, 1)$ and $G(\cdot; k) = F(\cdot; k, 1)$. We also have $\boldsymbol{S}_{n,1}^{(j)}$ and $\boldsymbol{S}_{n,2}^{(j)}$ to be conditionally independent. Hence, under the condition that $Z_{j:n} = u \in \lambda_k$, we have the joint density function of $\boldsymbol{S}_{n,1}^{(j)}$ and $\boldsymbol{S}_{n,2}^{(j)}$ to be

$$(j-1)! \prod_{i=1}^{j-1} \psi_1(s_i; k_0, u) \times (n-j)! \prod_{i=j+1}^{n} \psi_2(s_i; k_0, u), \tag{11}$$

denoted by $l_u\left(k_0; \boldsymbol{s}_n^{(j)}\right)$, where $\boldsymbol{s}_n^{(j)} = (s_1, \ldots, s_{j-1}, s_{j+1}, \ldots, s_n)$, for $0 \leq s_1 \leq \cdots \leq s_{j-1} \leq 1 \leq s_{j+1} \leq \cdots \leq s_n$. Equation (11) implies that $\boldsymbol{S}_{1*}^{(j)} = (S_1^{(j)}, \ldots, S_{j-1}^{(j)})$, which are the corresponding random variables to $\boldsymbol{S}_{n,1}^{(j)} = (S_{1:n}^{(j)}, \ldots, S_{j-1:n}^{(j)})$, are i.i.d. distributed with the conditional density function ψ_1, and $\boldsymbol{S}_{2*}^{(j)} = (S_{j+1}^{(j)}, \ldots, S_n^{(j)})$, which are the corresponding random variables to $\boldsymbol{S}_{n,2}^{(j)} = (S_{j+1:n}^{(j)}, \ldots, S_{n:n}^{(j)})$, are i.i.d. with the conditional density function ψ_2, given $Z_{j:n} = u$.

Let $Z'_{j:n} = X'_{j:n}/\sigma$, where $X'_{j:n}$ is the jth-order statistic from the GPD with parameters $k \neq k_0$ and $\sigma \neq \sigma_0$. Then, for any fixed $u \in \lambda_k$ and $u' \in \lambda_k$, and any $k \neq k_0$, conditional on $Z_{j:n} = u$ and $Z'_{j:n} = u'$, it follows that

$$\begin{aligned}\frac{1}{n-1}\log\frac{l_{u'}\big(k;\boldsymbol{S}_n^{(j)}\big)}{l_u\big(k_0;\boldsymbol{S}_n^{(j)}\big)} =& \frac{1}{n-1}\log\frac{l_{u'}\big(k;\boldsymbol{S}_{n*}^{(j)}\big)}{l_u\big(k_0;\boldsymbol{S}_{n*}^{(j)}\big)}\\ =& \frac{j-1}{n-1}\frac{1}{j-1}\sum_{i=1}^{j-1}\log\frac{\psi_1(S_i^{(j)};k,u')}{\psi_1(S_i^{(j)};k_0,u)}\\ &+\frac{n-j}{n-1}\frac{1}{n-j}\sum_{i=j+1}^{n}\log\frac{\psi_2(S_i^{(j)};k,u')}{\psi_2(S_i^{(j)};k_0,u)},\end{aligned} \tag{12}$$

where $\boldsymbol{S}_{n*}^{(j)} = (\boldsymbol{S}_{1*}^{(j)}, \boldsymbol{S}_{2*}^{(j)}) = (S_1^{(j)}, \ldots, S_{j-1}^{(j)}, S_{j+1}^{(j)}, \ldots, S_n^{(j)})$. By the weak law of large numbers, (12) converges in probability to

$$p\,E_1\left[\log\frac{\psi_1(S_1;k,u')}{\psi_1(S_1;k_0,u)}\right] + (1-p)\,E_2\left[\log\frac{\psi_2(S_2;k,u')}{\psi_2(S_2;k_0,u)}\right],$$

where S_1 and S_2 are random variables which are distributed with the conditional density functions $\psi_1(x;k_0,u)$ and $\psi_2(x;k_0,u)$, given $Z_{j:n} = u$, respectively, and $p = \lim_{n\to\infty} j/n$ $(0 \leq p \leq 1)$. E_1 and E_2 denote the conditional expectations with respect to ψ_1 and ψ_2, given $Z_{j:n} = u$, respectively. By Jensen's inequality, we have

$$\begin{aligned}&q\,E_1\left[\log\frac{\psi_1(S_1;k,u')}{\psi_1(S_1;k_0,u)}\right] + (1-q)\,E_2\left[\log\frac{\psi_2(S_2;k,u')}{\psi_2(S_2;k_0,u)}\right]\\ &< \log E_1\left[\frac{\psi_1(S_1;k,u')}{\psi_1(S_1;k_0,u)}\right] + \log E_2\left[\frac{\psi_2(S_2;k,u')}{\psi_2(S_2;k_0,u)}\right]\\ &= \log\int_0^1 \psi_1(t;k,u')\,dt + \log\int_1^\infty \psi_2(t;k,u')\,dt = 0.\end{aligned}$$

Hence, for any fixed $u, u' \in \lambda_k$, we have

$$\lim_{n\to\infty} P\left(\frac{1}{n-1}\log\frac{l_{u'}\big(k;\boldsymbol{S}_n^{(j)}\big)}{l_u\big(k_0;\boldsymbol{S}_n^{(j)}\big)} < 0 \,\middle|\, Z_{j:n} = u,\, Z'_{j:n} = u'\right) = 1,$$

or

$$\lim_{n\to\infty} P\Big(l_{u'}\big(k;\boldsymbol{S}_n^{(j)}\big) < l_u\big(k_0;\boldsymbol{S}_n^{(j)}\big) \mid Z_{j:n} = u,\, Z'_{j:n} = u'\Big) = 1.$$

Now, the density of $Z_{j:n}$, with $k_0 \in \mathbb{R}$, is given by

$$h_j\big(u;k_0\big) = \frac{n!}{(j-1)!(n-j)!}\big\{G\big(u;k_0\big)\big\}^{j-1} g\big(u;k_0\big)\big\{1 - G\big(u;k_0\big)\big\}^{n-j},$$

and thus we see that

$$\int_{\lambda_k}\int_{\lambda_{k_0}} P\Big(l_{u'}\big(k;\boldsymbol{S}_n^{(j)}\big) < l_u\big(k_0;\boldsymbol{S}_n^{(j)}\big) \mid Z_{j:n}=u,\, Z'_{j:n}=u'\Big)h_j\big(u;k_0\big)h_j\big(u';k\big)\,du\,du'$$
$$\leq \int_{\lambda_{k_0}} h_j\big(u;k_0\big)\,du \int_{\lambda_k} h_j\big(u';k\big)\,du' = 1,$$

since $P\Big(l_{u'}\big(k;\boldsymbol{S}_n^{(j)}\big) < l_u\big(k_0;\boldsymbol{S}_n^{(j)}\big) \mid Z_{j:n}=u,\, Z'_{j:n}=u'\Big)$ is bounded by 1. Then, by applying the dominated convergence theorem, we have

$$\lim_{n\to\infty} P\big(l\big(k;\boldsymbol{S}_n^{(j)}\big) < l\big(k_0;\boldsymbol{S}_n^{(j)}\big)\big)$$
$$= \int_{\lambda_k}\int_{\lambda_{k_0}} \lim_{n\to\infty} P\Big(l_{u'}\big(k;\boldsymbol{S}_n^{(j)}\big)$$
$$< l_u\big(k_0;\boldsymbol{S}_n^{(j)}\big) \mid Z_{j:n}=u,\, Z'_{j:n}=u'\Big)h_j\big(u;k_0\big)h_j\big(u';k\big)\,du\,du' = 1,$$

which completes the proof of Lemma 1. □

Proof of Theorem 5

Here, we shall use the shorthand notation $L\big(k;\boldsymbol{S}_n^{(j)}\big)$ for the log-likelihood function based on $\boldsymbol{S}_n^{(j)}$, $\log l\big(k;\boldsymbol{S}_n^{(j)}\big)$, and $L'\big(k;\boldsymbol{S}_n^{(j)}\big)$ and $L''\big(k;\boldsymbol{S}_n^{(j)}\big)$ for its derivatives with respect to k.

First, we assume that $k \neq 0$ and $k^* \neq 0$. We then have

$$\frac{1}{n}L'\big(k;\boldsymbol{S}_n^{(j)}\big) \xrightarrow{\mathscr{P}} \frac{1}{n}\frac{\int_{\lambda_k}\left(-\frac{n}{k}-\frac{\sum_{i=1}^{n}\log\left(1-k\,u\,S_{i:n}^{(j)}\right)}{k^2}\right)\psi_{n,j}(\boldsymbol{S}_n^{(j)},k,u)\,\delta(u-v)\,du}{\int_{\lambda_k}\psi_{n,j}(\boldsymbol{S}_n^{(j)},k,u)\,\delta(u-v)\,du}$$
$$= \frac{1}{n}\sum_{i=1}^{n}\left(-\frac{1}{k}-\frac{\log\left(1-k\,v\,S_i^{(j)}\right)}{k^2}\right)$$
$$\xrightarrow{\mathscr{P}} p\,\frac{1}{j-1}\sum_{i=1}^{j-1}\left(-\frac{1}{k}-\frac{\log\left(1-k\,v\,S_i^{(j)}\right)}{k^2}\right)+\frac{1}{n}C_1(k,v) \tag{13}$$
$$+(1-p)\,\frac{1}{n-j}\sum_{i=j+1}^{n}\left(-\frac{1}{k}-\frac{\log\left(1-k\,v\,S_i^{(j)}\right)}{k^2}\right), \quad \text{as } n\to\infty,$$

where $S_{j:n}^{(j)} = S_j^{(j)} = 1$, $C_1(k, v) = -\frac{1}{k} - \frac{\log(1-k\,v)}{k^2}$, $\delta(\cdot)$ is the Dirac delta function, $\psi_{n,j}(\boldsymbol{S}_n^{(j)}, k, u) = (j-1)! \prod_{i=1}^{j-1} \psi_1(S_{i:n}; k, u) \times (n-j)! \prod_{i=j+1}^{n} \psi_2(S_{i:n}; k, u)$, and ψ_1, ψ_2, h_j, $\boldsymbol{S}_{1*}^{(j)} = (S_1^{(j)}, \dots, S_{j-1}^{(j)})$ and $\boldsymbol{S}_{2*}^{(j)} = (S_{j+1}^{(j)}, \dots, S_n^{(j)})$ are all as defined in the proof of Lemma 1.

As with the likelihood function under regularity conditions (see Lemma 5.3 of Lehmann and Casella 1998), we obtain

$$E\big(L'\big(k; \boldsymbol{S}_n^{(j)}\big)\big) = \frac{\partial}{\partial k} 1 = 0, \tag{14}$$

$$\begin{aligned} Var\big(L'\big(k; \boldsymbol{S}_n^{(j)}\big)\big) &= E\Big(L'\big(k; \boldsymbol{S}_n^{(j)}\big)^2\Big) = E\bigg(-\frac{\partial^2}{\partial k^2} L\big(k; \boldsymbol{S}_n^{(j)}\big)\bigg) + \frac{\partial^2}{\partial k^2} 1 \\ &= -E\big(L''\big(k; \boldsymbol{S}_n^{(j)}\big)\big). \end{aligned} \tag{15}$$

Hence, it follows, from (13)–(15), with the use of central limit theorem, that

$$\frac{L'\big(k; \boldsymbol{S}_n^{(j)}\big)}{\sqrt{I_{j,n}(k)}} \xrightarrow{d} N(0, 1).$$

We can obtain the same results when $k = 0$ or $k^* = 0$, by noting that $L'(0; \boldsymbol{S}_n^{(j)}) = \lim_{k\to 0} L'(k; \boldsymbol{S}_n^{(j)})$ and $L''(0; \boldsymbol{S}_n^{(j)}) = \lim_{k\to 0} L''(k; \boldsymbol{S}_n^{(j)})$, and by Lebesgue's dominated convergence theorem. These details are therefore omitted for the sake of brevity. □

Proof of Theorem 6

Here, we shall use the shorthand notation $L\big(k; \boldsymbol{S}_n^{(j)}\big)$ for the log-likelihood function based on $\boldsymbol{S}_n^{(j)}$, $\log l\big(k; \boldsymbol{S}_n^{(j)}\big)$, and $L'\big(k; \boldsymbol{S}_n^{(j)}\big)$, $L''\big(k; \boldsymbol{S}_n^{(j)}\big)$ and $L'''\big(k; \boldsymbol{S}_n^{(j)}\big)$ for its derivatives with respect to k. By a Taylor expansion of $L'\big(\hat{k}; \boldsymbol{S}_n^{(j)}\big)$ around k, we obtain

$$\hat{k} - k = \frac{L'\big(k; \boldsymbol{S}_n^{(j)}\big)}{-L''\big(k; \boldsymbol{S}_n^{(j)}\big) - \frac{\hat{k}-k}{2} L'''\big(k^*; \boldsymbol{S}_n^{(j)}\big)}, \tag{16}$$

where k^* lies between k and $\hat{k}$.

By Theorem 1, we can take any $j \in \{1, \dots, n\}$ to treat $\hat{k}$ that is the MLE based on $\boldsymbol{S}_n^{(j)}$, without loss of generality. Here, we take j such as $Z_{j:n} \xrightarrow{\mathscr{P}} v \in \lambda_k$ as $n \to \infty$, where $\xrightarrow{\mathscr{P}}$ denotes convergence in probability, and $\lambda_k = \{u : 0 < u < \infty$, if $k < 0$, or $0 < u < 1/k$, if $k > 0\}$ as defined in the proof of Proposition 1, and let $p = \lim_{n\to\infty} j/n$.

From here, we shall show the following facts:

$$\frac{L'\left(k; \boldsymbol{S}_n^{(j)}\right)}{\sqrt{I_{j,n}(k)}} \xrightarrow{d} N(0,1), \tag{17}$$

$$-L''\left(k; \boldsymbol{S}_n^{(j)}\right) \xrightarrow{\mathscr{P}} I_{j,n}(k), \tag{18}$$

$$\frac{1}{n} L'''\left(k^*; \boldsymbol{S}_n^{(j)}\right) \text{ is bounded in probability }, \tag{19}$$

as $n \to \infty$.

We first note that (17) holds from Theorem 6. So, we shall show now (18), for which we assume that $k \neq 0$ and $k^* \neq 0$.

As in (13), we have

$$\begin{aligned}
&-\frac{1}{n} L''\left(k; \boldsymbol{S}_n^{(j)}\right) - \frac{1}{n} I_{j,n}(k) \\
&\xrightarrow{\mathscr{P}} p \frac{1}{j-1} \sum_{i=1}^{j-1} \left\{ -\frac{1}{k^2} - \frac{2 \log\left(1 - k\, v\, S_i^{(j)}\right)}{k^3} \right\} \\
&\quad + (1-p) \frac{1}{n-j} \sum_{i=j+1}^{n} \left\{ -\frac{1}{k^2} - \frac{2 \log\left(1 - k\, v\, S_i^{(j)}\right)}{k^3} \right\} \\
&\quad - pE\left(-\frac{1}{k^2} - \frac{2 \log\left(1 - k\, v\, S_1^{(j)}\right)}{k^3} \right) - (1-p)E\left(-\frac{1}{k^2} - \frac{2 \log\left(1 - k\, v\, S_{j+1}^{(j)}\right)}{k^3} \right) \\
&\xrightarrow{\mathscr{P}} 0.
\end{aligned}$$

The last convergence follows by weak law of large numbers.

Next, (19) holds since

$$
\begin{aligned}
&\left|\frac{1}{n}L'''\left(k^*;\boldsymbol{S}_n^{(j)}\right)\right| \\
&= \left|\frac{2}{n}\left\{\frac{l'(k^*;\boldsymbol{S}_n^{(j)})}{l(k^*;\boldsymbol{S}_n^{(j)})}\right\}^3 - \frac{3}{n}\frac{l'(k^*;\boldsymbol{S}_n^{(j)})l''(k^*;\boldsymbol{S}_n^{(j)})}{\left\{l(k^*;\boldsymbol{S}_n^{(j)})\right\}^2} + \frac{1}{n}\frac{l'''(k^*;\boldsymbol{S}_n^{(j)})}{l(k^*;\boldsymbol{S}_n^{(j)})}\right| \\
&\xrightarrow{\mathscr{P}} \left|p\frac{1}{j-1}\sum_{i=1}^{j-1}\left\{-\frac{2}{k^{*3}} - \frac{6\log\left(1-k^* v S_i^{(j)}\right)}{k^{*4}}\right\} + \frac{1}{n}C_2(k^*,v)\right. \\
&\quad \left. +(1-p)\frac{1}{n-j}\sum_{i=j+1}^{n}\left\{-\frac{2}{k^{*3}} - \frac{6\log\left(1-k^* v S_i^{(j)}\right)}{k^{*4}}\right\}\right| \\
&\xrightarrow{\mathscr{P}} \left|pE_1\left(-\frac{2}{k^{*3}} - \frac{6\log\left(1-k^* v S_1^{(j)}\right)}{k^{*4}}\right) + \frac{1}{n}C_2(k^*,v)\right. \\
&\quad \left. +(1-p)E_2\left(-\frac{2}{k^{*3}} - \frac{6\log\left(1-k^* v S_{j+1}^{(j)}\right)}{k^{*4}}\right)\right| \\
&\le \frac{2}{\left|k^{*3}\right|} + \frac{6p}{k^{*4}}E_1\left(\left|\log\left(1-k^* v S_1^{(j)}\right)\right|\right) + \frac{6(1-p)}{k^{*4}}E_2\left(\left|\log\left(1-k^* v S_{j+1}^{(j)}\right)\right|\right) \\
&\quad + \frac{1}{n}\left|C_2(k^*,v)\right| \\
&\le \frac{2}{\left|k^{*3}\right|} + \frac{6vp}{\left|k^{*3}\right|}E_1\left(S_1^{(j)}\right) + \frac{6v(1-p)}{\left|k^{*3}\right|}E_2\left(S_{j+1}^{(j)}\right) + \frac{1}{n}\left|C_2(k^*,v)\right| \\
&= \frac{2}{\left|k^{*3}\right|} + \frac{6vp}{\left|k^{*3}\right|(1+k^*)}\left[\frac{1}{1-(1-vk^*)^{-1/k^*}} + \frac{1}{v}\right] + \frac{6v(1-p)}{\left|k^{*3}\right|(1+k^*)}\left(1+\frac{1}{v}\right) \\
&\quad + \frac{1}{n}\left|C_2(k^*,v)\right| \\
&< \infty,
\end{aligned}
\tag{20}
$$

where $C_2(k^*, v) = -\frac{2}{k^{*3}} - \frac{6\log(1-k^* v)}{k^{*4}}$, and E_1 and E_2 are conditional expectations with respect to ψ_1 and ψ_2, given $Z_{j:n} = v$, respectively, as defined in the proof of Lemma 1.

The interchangeability of integrations, differentiations and limits in the proofs of (17), (18) and (19) can be justified by Lebesgue's dominated convergence theorem. These proofs are quite similar to the proof of interchangeability of differentiations and integration in Proposition 1 and are therefore omitted. We can further obtain the same results when $k = 0$ or $k^* = 0$, by noting that $L''(0;\boldsymbol{S}_n^{(j)}) = \lim_{k\to 0} L''(k;\boldsymbol{S}_n^{(j)})$ and $L'''(0;\boldsymbol{S}_n^{(j)}) = \lim_{k^*\to 0} L'''(k^*;\boldsymbol{S}_n^{(j)})$ and by the use of Lebesgue's dominated convergence theorem. These proofs are not presented here for the sake of brevity. Thus, the proof of Theorem 6 gets completed. □

References

Andrews, D. F., Herzberg, A. M. (1985). *Data: A Collection of Problems from Many Fields for the Student and Research Worker*. New York: Springer.

Beirlant, J., Goegebeur, Y., Segers, J., Teugels, J. (2004). *Statistics of Extremes: Theory and Applications*. Chichester, England: Wiley.

Billingsley, P. (1994). *Probability and Measure* 3rd ed. New York: Wiley.

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

Castillo, E., Hadi, A. S., Balakrishnan, N., Sarabia, J. M. (2004). *Extreme Value and Related Models with Applications in Engineering and Science*. Hoboken, New Jersey: Wiley.

Chen, P., Ye, Z., Zhao, X. (2017). Minimum distance estimation for the generalized Pareto distribution. *Technometrics*, *59*, 528–541.

Coles, S. (2001). *An Introduction to Statistical Modeling of Extreme Values*. London: Springer.

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

de Haan, L., Ferreira, A. (2006). *Extreme Value Theory: An Introduction*. New York: Springer.

de Zea, Bermudez P., Kotz, S. (2010a). Parameter estimation of the generalized Pareto distribution—Part I. *Journal of Statistical Planning and Inference*, *140*, 1353–1373.

de Zea, Bermudez P., Kotz, S. (2010b). Parameter estimation of the generalized Pareto distribution - Part II. *Journal of Statistical Planning and Inference*, *140*, 1374–1388.

del Castillo, J., Serra, I. (2015). Likelihood inference for generalized Pareto distribution. *Computational Statistics & Data Analysis*, *83*, 116–128.

Giles, D. E., Feng, H., Godwin, R. T. (2016). Bias-corrected maximum likelihood estimation of the parameters of the generalized Pareto distribution. *Communications in Statistics—Theory and Methods*, *45*, 2465–2483.

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

Hill, B. M. (1975). A simple general approach to inference about the tail of a distribution. *The Annals of Statistics*, *3*, 1163–1174.

Hosking, J. R. M. (1990). L-moments: Analysis and estimation of distributions using linear combinations of order statistics. *Journal of the Royal Statistical Society, Series B*, *52*, 105–124.

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

Iliopoulos, G., Balakrishnan, N. (2009). Conditional independence of blocked ordered data. *Statistics & Probability Letters*, *79*, 1008–1015.

Lehmann, E. L., Casella, G. (1998). *Theory of Point Estimation* 2nd ed. New York: Springer.

Pickands, J. (1975). Statistical inference using extreme order statistics. *The Annals of Statistics*, *3*, 119–131.

Salvadori, G., De Michele, C., Kottegoda, N. T., Rosso, R. (2007). *Extremes in Nature: An Approach Using Copulas*. Dordrecht: Springer.

Smith, R. L. (1984). Threshold methods for sample extremes. In J: Tiago de Oliveira (Ed.), *Statistical Extremes and Applications*, pp. 621–638. Dordrecht: Springer.

Smith, R. L. (1985). Maximum likelihood estimation in a class of nonregular cases. *Biometrika*, *72*, 67–90.

Song, J., Song, S. (2012). A quantile estimation for massive data with generalized Pareto distribution. *Computational Statistics & Data Analysis*, *56*, 143–150.

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

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