A review of goodness of fit tests for Pareto distributions
by J. Chu, O. Dickin and S. Nadarajah
School of Mathematics, University of Manchester, Manchester, UK

Abstract: Pareto distributions are the most popular models in economics and finance. Hence, it is essential to have a wide spectrum of tools for checking their goodness of fit to a given data set. This paper provides the first review of known goodness of fit tests for Pareto distributions. Over twenty tests are reviewed. Their powers are compared by simulation.

Keywords: Economics; Finance; Power; Simulation

1 Introduction

Pareto distributions are the most popular models in economics, finance and related areas. In fact, the first Pareto distribution due to Pareto [50] was used to model the allocation of wealth among individuals. Since Pareto [50], several extended Pareto distributions have been proposed in the literature and have been applied in a wide variety of fields. The list of applications is too exhaustive, however, some recent applications have included: income modeling (Bhattacharya [10]); the wealth distribution in the Forbes 400 list (Klass et al. [40]); commercial fire loss severity in Taiwan (Lee [43]); city size distribution in the United States (Ioannides and Skouras [35]).

Pareto distributions are increasingly being used to model problems in economics and finance. Hence, it is essential to have tools to check the goodness of fit (GOF) of Pareto distributions. Several tests have in fact been proposed to check the GOF of Pareto distributions, however, we are not aware of any review covering all known tests for Pareto distributions. Such a review is essential for practitioners given the wide spread use of Pareto distributions and such a review could also encourage the development of more GOF tests.

The aim of this paper is to provide the first review of known GOF tests for Pareto distributions. We review the known tests for the generalized Pareto, Pareto type I and Pareto type II distributions in Section 3. Some variations of these tests are given in Section 4. Some preliminaries for stating the tests are given in Section 2. Arnold [4] provides a comprehensive account of all generalizations of the Pareto distribution and their applications.

The generalized Pareto distribution (GPD) in the general form has the CDF

\[ F(x) = \begin{cases} 
1 - \left[1 + \frac{\beta}{\sigma} (x - \mu)\right]^{-\frac{1}{\beta}}, & \text{if } \beta \neq 0, \\
1 - \exp \left[-\frac{x - \mu}{\sigma}\right], & \text{if } \beta = 0, 
\end{cases} \]

where \( \mu \in (-\infty, \infty) \), \( \sigma \in (0, \infty) \) and \( \beta \in (-\infty, \infty) \) are the location, scale and shape parameters, respectively. The domain of this CDF is \( x \geq \mu \) if \( \beta \geq 0 \) and \( \mu \leq x \leq \mu + \sigma/\beta \) if \( \beta < 0 \). We shall
denote a random variable \( X \) having the CDF (1) by \( \text{GPD}(\mu, \sigma, \beta) \). The PDF and the inverse CDF corresponding to (1) are

\[
f(x) = \frac{1}{\sigma} \left[ 1 + \frac{\beta}{\sigma} (x - \mu) \right]^{-\frac{1}{\beta} - 1}
\]

and

\[
F^{-1}(p) = \mu + \frac{\sigma}{\beta} \left[ (1 - p)^{-\beta} - 1 \right],
\]

respectively. Sometimes we shall write \( F(x) \), \( f(x) \) and \( F^{-1}(p) \) as \( F(x; \mu, \sigma, \beta) \), \( f(x; \mu, \sigma, \beta) \) and \( F^{-1}(p; \mu, \sigma, \beta) \), respectively, to make the dependence on the parameters explicit. The exponential distribution is the limiting case of the GPD for \( \beta \to 0 \). Also if \( X \) is a \( \text{GPD}(0, \sigma, \beta) \) random variable then

\[
Y = -(1/\beta) \log \left[ 1 - (\beta X/\sigma) \right]
\]

is an exponential random variable.

The Pareto type I (PI) distribution has the CDF

\[
F(x) = 1 - \left( \frac{\sigma}{x} \right)^{\beta}
\]

for \( x > \sigma \), where \( \sigma > 0 \) is the scale parameter and \( \beta > 0 \) is the shape parameter. We shall denote a random variable \( X \) having the CDF (3) by \( \text{PI}(\sigma, \beta) \). The PDF and the inverse CDF corresponding to (3) are

\[
f(x) = \frac{\beta \sigma^{\beta}}{x^{\beta+1}}
\]

and

\[
F^{-1}(p) = \sigma (1 - p)^{-1/\beta},
\]

respectively. Sometimes we shall write \( F(x) \), \( f(x) \) and \( F^{-1}(p) \) as \( F(x; \sigma, \beta) \), \( f(x; \sigma, \beta) \) and \( F^{-1}(p; \sigma, \beta) \), respectively, to make the dependence on the parameters explicit. If \( X \sim \text{PI}(1, \beta) \) then \( Y = \log(X) \sim \text{Exp}(\beta) \).

The Pareto type II (PII) distribution has the CDF

\[
F(x) = 1 - \left[ 1 + \frac{x - \mu}{\sigma} \right]^{-\frac{1}{\beta}}
\]

for \( x > 0 \), where \( \mu \) is the location parameter, \( \sigma > 0 \) is the scale parameter and \( \beta > 0 \) is the shape parameter. We shall denote a random variable \( X \) having the CDF (4) by \( \text{PII}(\mu, \sigma, \beta) \). The PDF and the inverse CDF corresponding to (4) are

\[
f(x) = \frac{1}{\sigma \beta} \left[ 1 + \frac{x - \mu}{\sigma} \right]^{-\frac{1}{\beta}-1}
\]

and

\[
F^{-1}(p) = \mu + \sigma \left[ (1 - p)^{-\beta} - 1 \right],
\]
respectively. Sometimes we shall write \( F(x), f(x) \) and \( F^{-1}(p) \) as \( F(x; \mu, \sigma, \beta), f(x; \mu, \sigma, \beta) \) and \( F^{-1}(p; \mu, \sigma, \beta) \), respectively, to make the dependence on the parameters explicit. Note that PII \((\mu, \sigma, \beta)\) reduces to PI \((\sigma, \beta)\) if \( \mu = 0 \).

In total, we review twenty one GOF tests in this paper: of these eleven are for the GPD, eight are for the PI distribution and two are for the PII distribution. A simulation study comparing the power of all these tests is given in Section 5.

2 Preliminaries

2.1 Notation

Throughout, we suppose \( X_1, X_2, \ldots, X_n \) is a complete random sample from the distribution specified by \( H_0 \); \( x_1, x_2, \ldots, x_n \) denote their observed values; \( X_{1:n} < X_{2:n} < \cdots < X_{n:n} \) denote the order statistics of \( X_1, X_2, \ldots, X_n \); \( x_{1:n} < x_{2:n} < \cdots < x_{n:n} \) denote the observed order statistics; \( z(i) = F(x_{i:n}) \) for a hypothesized CDF \( F \); \( F_n \) denotes the empirical CDF (ECDF) of the random sample; \( F_n = 1 - F_n \) denotes the empirical survival function; \( \Phi(\cdot) \) denotes the standard normal CDF; \( \Phi^{-1}(\cdot) \) denotes the standard normal inverse CDF; \( \hat{\theta}_n \) or \( \hat{\theta} \) denotes an estimator of a parameter \( \theta \) based on a sample of size \( n \); \( \alpha \) denotes the level of significance; \( I\{\cdot\} \) denotes the indicator function.

If \( X_1, X_2, \ldots, X_n \) is a random sample from (3), we define

\[
Z_i = i \left[ \log (X_{n-i+1:n}) - \log (X_{n-i:n}) \right]
\]

and

\[
H_{k,n} = \frac{1}{k} \sum_{i=1}^{k} Z_i.
\]

2.2 Estimators for the parameters

We now present maximum likelihood and other estimators for the three distributions. If \( X_1, X_2, \ldots, X_n \) is a random sample from GPD\((\mu, \sigma, \beta)\) the maximum likelihood estimators of \( \mu, \sigma \) and \( \beta \) are the simultaneous solutions of

\[
\sum_{i=1}^{n} \left[ 1 + \beta \frac{X_i - \mu}{\sigma} \right]^{-1} = 0,
\]

\[
n\sigma - (\beta + 1) \sum_{i=1}^{n} \left[ 1 + \beta \frac{X_i - \mu}{\sigma} \right]^{-1} (X_i - \mu) = 0
\]

and

\[
\sigma \sum_{i=1}^{n} \log \left[ 1 + \beta \frac{X_i - \mu}{\sigma} \right] - (\beta + 1) \sum_{i=1}^{n} \left[ 1 + \beta \frac{X_i - \mu}{\sigma} \right]^{-1} (X_i - \mu) = 0.
\]
If \( X_1, X_2, \ldots, X_n \) is a random sample from \( \text{GPD}(0, \sigma, \beta) \) the maximum likelihood estimators of \( \sigma \) and \( \beta \) are the simultaneous solutions of

\[
ns - (\beta + 1) \sum_{i=1}^{n} \left[ 1 + \frac{X_i}{\sigma} \right]^{-1} X_i = 0
\]

and

\[
\sigma \sum_{i=1}^{n} \log \left[ 1 + \frac{X_i}{\sigma} \right] - (\beta + 1) \sum_{i=1}^{n} \left[ 1 + \frac{X_i}{\sigma} \right]^{-1} X_i = 0.
\]

If \( X_1, X_2, \ldots, X_n \) is a random sample from \( \text{PI}(\sigma, \beta) \) the maximum likelihood estimators of \( \sigma \) and \( \beta \) are

\[
\hat{\beta} = \min (X_1, X_2, \ldots, X_n)
\]

and

\[
\hat{\beta} = n \left[ \sum_{i=1}^{n} \log X_i - n \log \hat{\sigma} \right]^{-1}.
\]

If \( X_1, X_2, \ldots, X_n \) is a random sample from \( \text{PII}(\mu, \sigma, \beta) \) the maximum likelihood estimators of \( \mu, \sigma \) and \( \beta \) are the simultaneous solutions of

\[
\sum_{i=1}^{n} \left[ 1 + \frac{X_i - \mu}{\sigma} \right]^{-1} = 0,
\]

\[
n\beta\sigma - (\beta + 1) \sum_{i=1}^{n} \left[ 1 + \frac{X_i - \mu}{\sigma} \right]^{-1} (X_i - \mu) = 0
\]

and

\[
n\beta - \sum_{i=1}^{n} \log \left[ 1 + \frac{X_i - \mu}{\sigma} \right] = 0.
\]

Finally, if \( X_1, X_2, \ldots, X_n \) is a random sample from \( \text{PII}(0, \sigma, \beta) \) the maximum likelihood estimators of \( \sigma \) and \( \beta \) are the simultaneous solutions of

\[
n\beta\sigma - (\beta + 1) \sum_{i=1}^{n} \left[ 1 + \frac{X_i}{\sigma} \right]^{-1} X_i = 0
\]

and

\[
n\beta - \sum_{i=1}^{n} \log \left[ 1 + \frac{X_i}{\sigma} \right] = 0.
\]

If \( X_1, X_2, \ldots, X_n \) is a random sample from \( \text{GPD}(0, \sigma, \beta) \) the method of moment estimators of \( \sigma \) and \( \beta \) are

\[
\hat{\beta} = \frac{1}{2} \left[ 1 - \frac{\bar{X}^2}{\bar{S}^2} \right]
\]
and
\[ \hat{\sigma} = \frac{\bar{X}}{2} \left[ 1 + \frac{\bar{X}^2}{S^2} \right], \]
where \( \bar{X} \) denotes the sample mean and
\[ S^2 = \frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X})^2. \]

If \( X_1, X_2, \ldots, X_n \) is a random sample from GPD(0, \( \sigma \), \( \beta \)) then asymptotic maximum likelihood estimators of \( \sigma \) and \( \beta \) are (Villaseñor-Alva and González-Estrada [60])
\[ \hat{\beta} = -W_{n-k+1} + \frac{1}{k} \sum_{j=1}^{k} W_{n-j+1} \]
and
\[ \hat{\sigma} = \hat{\beta} \exp \left[ W_{n-k+1} + \hat{\beta} \log(k/n) \right], \]
where \( W_j = \log X_{j:n} \) and \( 1 \leq k \leq n \). These estimators exist for \( 0 < \beta < 0.5 \). If \( X_1, X_2, \ldots, X_n \) is a random sample from GPD(0, \( \sigma \), \( \beta \)) then estimators \( \sigma \) and \( \beta \) based on a combination of the methods of moments and maximum likelihood are (Villaseñor-Alva and González-Estrada [60])
\[ \hat{\beta} = \frac{X}{\bar{X} - X_{n:n}} \]
and
\[ \hat{\sigma} = -\frac{XX_{n:n}}{\bar{X} - X_{n:n}}. \]
These estimators exist also for \( 0 < \beta < 0.5 \).

If \( X_1, X_2, \ldots, X_n \) is a random sample from PII(0, \( \sigma \), \( \beta \)) the method of probability weighted moments [30] estimators of \( \sigma \) and \( \beta \) are
\[ \hat{\beta} = 2 - \frac{\beta_0}{\beta_0 - 2\beta_1}, \]
\[ \hat{\sigma} = \frac{2\beta_0\beta_1}{\beta_0 - 2\beta_1}, \]
where
\[ \beta_0 = \bar{X}, \]
\[ \beta_1 = \frac{1}{n} \sum_{j=1}^{n} (1 - p_j) X_{j:n}, \]
\[ p_j = \frac{j - 0.35}{n}. \]
These estimators exist for \( 0 < \beta < 1 \). There may be cases where \( \hat{\beta} \) falls outside of the range of \( \beta \) (\( \beta \) is a fixed parameter, but \( \hat{\beta} \) is an estimate and thus can fall outside a specific range).
2.3 Bootstrap / simulation methodology

Suppose we wish to test $H_0 : X_1, X_2, \ldots, X_n$ is a random sample from one of the three distributions (the GPD, the PI distribution or the PII distribution) versus $H_1 : X_1, X_2, \ldots, X_n$ is a random sample not from the distribution. Let $T(X_1, X_2, \ldots, X_n)$ denote the corresponding test statistic. The rejection rule for the test can be determined by the following bootstrap scheme:

1. estimate the parameters of the distribution based on the observed sample $X_1, X_2, \ldots, X_n$;
2. simulate a random sample of size $n$ from the distribution based on the estimated parameters;
3. compute $T$ for the simulated sample;
4. repeat steps 2 and 3 one thousand times, resulting in the values $T_1, T_2, \ldots, T_{1000}$ of $T$ say;
5. compute the ECDF, say $\hat{F}$, of $T_1, T_2, \ldots, T_{1000}$;
6. reject $H_0$ with significance level $\alpha$ if $T(X_1, X_2, \ldots, X_n) > \hat{F}^{-1}(1 - \alpha)$.

The $p$-value is $1 - \hat{F}(T(X_1, X_2, \ldots, X_n))$.

3 Tests for the GPD, the PI distribution and the PII distribution

3.1 The intersection-union test (Villaseñor-Alva and González-Estrada [60])

Hypothesis

We wish to consider the hypothesis

$H_0 : F \in A$,
$H_1 : F \notin A$,

where $A = \{\text{CDFs of GPD}(\mu, \sigma, \beta), -\infty < \mu < \infty, \sigma > 0, -\infty < \beta < \infty\}$.

The proposed test considers the GPD over two distinct cases: the family of GPDs with positive shape parameter $\beta$ and the family of GPDs with negative $\beta$. The two cases are then combined using an intersection union test, see, for example, Villaseñor-Alva and González-Estrada [60]. The two separate hypotheses are:

$H_0^+ : F \in A^+$,
$H_1^+ : F \notin A^+$,

and

$H_0^- : F \in A^-$,
$H_1^- : F \notin A^-$,

where $A^+ = \{\text{CDFs of GPD}(\mu, \sigma, \beta), -\infty < \mu < \infty, \sigma > 0, \beta \geq 0\}$ and $A^- = \{\text{CDFs of GPD}(\mu, \sigma, \beta), -\infty < \mu < \infty, \sigma > 0, \beta < 0\}$. Obviously $A = A^+ \cup A^-$ and $A^+ \cap A^- = 0$. 

6
A test for $H_0^+$

We first note that if $X$ has the CDF (1), there is a linear relationship between the variables $Y = F(X)^{-\beta}$ and $X$. This relation is the basis behind the following test.

Define $Y_i = \left[ F_n(X_{i:n}) \right]^{-\hat{\beta}}$, $i = 1, 2, \ldots, n$. We also define $\hat{\beta} = \hat{\beta}_K$ as the estimator of $\beta$, which is found using the asymptotic maximum likelihood method over the $k$ largest order statistics, see Section 2.2.

For $0 \leq \beta < 0.5$ the second moment of the GPD is defined. Therefore, we define the sample correlation coefficient between $X_i$ and $Y_i$,

$$R_1 = \frac{\sum_{i=1}^{n} (X_i - \bar{X}) (Y_i - \bar{Y})}{n \sqrt{S_X^2 S_Y^2}},$$

as an estimator of the linear correlation between $X$ and $Y$. $\bar{X}$, $\bar{Y}$, $S_X^2$, $S_Y^2$ are the corresponding sample means and variances of $X_1, X_2, \ldots, X_n$ and $Y_1, Y_2, \ldots, Y_n$.

For $\beta > 0.5$ the second moment of the GPD is not necessarily defined, although if we define $X_i^*$ and $Y_i^*$ such that

$$X_i^* = \log(X_i),$$

$$Y_i^* = \log \left\{ \left[ F_n(X_i) \right]^{-\beta} - 1 \right\}, \quad i = 1, 2, \ldots, n,$$

then the second moment of $X^*$ is finite and so defined. Therefore, much like the case above, we define the sample correlation coefficient between $X_i^*$ and $Y_i^*$,

$$R_2 = \frac{\sum_{i=1}^{n} (X_i^* - \bar{X}^*) (Y_i^* - \bar{Y}^*)}{n \sqrt{S_X^2 S_Y^2}},$$

as an estimator of the linear correlation between $X^*$ and $Y^*$.

Hence, for testing $H_0^+$, we propose the test statistic

$$R^+ = \begin{cases} R_1, & \text{if } 0 \leq \beta < 0.5, \\ R_2, & \text{if } \beta \geq 0.5. \end{cases}$$

Null distribution for $H_0^+$

We use the parametric bootstrap to find the null distribution of $R^+$, see Section 2.3. The null distribution is dependent on the shape parameter $\beta$ and our chosen value of $k$. More details on the implications of this on our rejection region are given below.

Rule for rejecting $H_0^+$

As we expect a value close to 1 for $R^+$ under our null hypothesis, we reject $H_0^+$ if $R^+ < c_\alpha^+$, where $c_\alpha^+$ is the critical value defined to be the $100\alpha\%$ quantile of the distribution of $R^+$ over $H_0^+$. The
null distribution of $R^+$ is dependent on the shape parameter $\beta$, therefore to obtain a value for $c^+_\alpha$ we use the parametric bootstrap, see Section 2.3.

As the asymptotic maximum likelihood estimator (see Section 2.2) is dependent on the $k$ largest order statistics, the null distribution of $R^+$ depends on $k$. Hence we choose $k$ such that $R^+$ is an $\alpha$ level test. Furthermore, as $c^+_\alpha$ is dependent on $k$, we also choose $k$ so that it minimises $c^+_\alpha$ when drawing samples of size $n$ from the GPD with $\sigma = 1$ and $\beta = 0$.

A test for $H^-_0$

To test the null hypothesis $H^-_0$ we use the same method as that for $H^+_0$, although as we now have $\beta < 0$, we estimate $\beta$ by $\tilde{\beta}$, the estimator found by the combined method of the method of moments and the maximum likelihood method given in Section 2.2. Therefore, we use the sample correlation coefficient between $X_i$ and $Z_i = \left[ F_n(X_i) \right]^{-\tilde{\beta}}$, $i = 1, 2, \ldots, n$, given by

$$R^- = \sum_{i=1}^{n} \frac{(X_i - \bar{X})(Z_i - \bar{Z})}{n \sqrt{S^2_X S^2_Z}},$$

as our statistic to test $H^-_0$.

Null distribution for $H^-_0$

Let $|R^-|$ be the absolute value of $R^-$, then the distribution of $|R^-|$ can be calculated using the parametric bootstrap, see Section 2.3. Furthermore, the null distribution of $|R^-|$ is dependent on $\beta$, although it is independent of the choice of $k$.

Rule for rejecting $H^-_0$

Under the null hypothesis we would expect $|R^-|$ to have a value close to 1. Therefore, we reject $H^-_0$ if $|R^-| < c^-_\alpha$, where $c^-_\alpha$ is the 100$\alpha$% quantile of the null distribution of $|R^-|$. We obtain $c^-_\alpha$ using the same method described for $c^+_\alpha$, using $\tilde{\beta}$ in place of $\hat{\beta}$.

Rejection region of the intersection-union test

An intersection-union test for $H_0 : F \in A$ rejects $H_0$ when we reject both of our sub-hypotheses $H^+_0$ and $H^-_0$. Therefore the rejection region of the intersection-union test is the intersection of our two sub-hypotheses; to get a test of level $\alpha$, we test $H^+_0$ and $H^-_0$ with a test size equal to $\alpha$.

3.2 Empirical distribution function statistics

The following tests are applicable to a number of different distributions due to their formation, all being based on the ECDF of the data sample being tested.

Hypothesis

Let $F$ denote a CDF given by (1). We test the null hypothesis
$H_0 : X_1, \ldots, X_n$ is a random sample from $F$

against the alternative hypothesis

$H_0 : X_1, \ldots, X_n$ is a random sample not from $F$.

To test this hypothesis, we consider the following statistics:

### 3.2.1 Kolmogorov Smirnov test (Kolmogorov [41], Smirnov [59])

The Kolmogorov Smirnov statistic is given by

$$D = \sup_x |F_n(x) - F(x)|.$$  

Under $H_0$ we can estimate the parameters of the GPD using the method of probability weighted moments (PWM), see Section 2.2. In particular, for the negative shape parameter the combined method could be used to estimate the unknown parameters, and for the positive shape parameter the asymptotic maximum likelihood method can be used, see Section 2.2.

**Null distribution and critical values**

It is difficult to derive the distribution of the Kolmogorov Smirnov statistic under $H_0$. We can approximate its distribution using Section 2.3. Once we have calculated the parameter estimates, we are able to calculate the values of $D$. These values can then be used to calculate the critical values of our distribution. For more information on this, see [5] [51].

### 3.2.2 Cramer von Mises test (Cramer [20], von Mises [62])

The Cramer von Mises statistic is given by

$$W^2_n = n \int_{-\infty}^{\infty} [F(x) - F_n(x)]^2 dF(x).$$

Under the observed order statistics, the test statistic becomes

$$W^2 = \sum_{i=1}^{n} \left[ \frac{z(i) - 2i - 1}{2n} \right]^2 + \frac{1}{12n}.$$

Under $H_0$, supposing that the parameters are unknown, we can estimate the parameters using the maximum likelihood method, see Section 2.2. Whilst there is some possibility that the maximum likelihood estimator will not exist, this is rarely a problem in practice.

**Null distribution and critical values**

Like the Kolmogorov Smirnov statistic, it is difficult to find the distribution of the Cramer von Mises statistic. A good explanation of how we can find the approximate distribution of the Cramer von Mises statistic is given by [19]; it is shown that the asymptotic distribution of the Cramer von Mises statistic is a sum of weighted $\chi^2_1$ variables. From this, it is relatively simple to find the critical values of the Cramer von Mises statistic.
### 3.2.3 Anderson Darling test (Anderson and Darling [2], [3])

The Anderson Darling statistic is given by

\[
A_n^2 = n \int_{-\infty}^{\infty} \left[ \frac{F(x) - F_n(x)}{F(x)[1 - F(x)]} \right]^2 dF(x).
\]

Under the observed order statistics, the Anderson Darling statistic becomes

\[
A^2 = -n - \frac{1}{n} \sum_{i=1}^{n} (2i - 1) \left[ \log \left( z(i) \right) + \log \left( 1 - z(n+1-i) \right) \right] .
\]

The Anderson Darling statistic gives greater weighting to the tail of the distribution compared to the Cramer von Mises statistic and so is useful when we are trying to detect outliers.

**Null distribution and critical values**

Similar to the Cramer von Mises statistic, the Anderson Darling statistic can be found to have an asymptotic distribution given by a sum of weighted \( \chi_1^2 \) variables.

### 3.2.4 Modified Anderson Darling test (Anderson and Darling [2], [3])

The modified Anderson Darling test statistic is defined as

\[
AU_n^2 = n \int_{-\infty}^{\infty} \left[ F_n(x) - F(x) \right]^2 \psi^2(x) dF(x),
\]

where \( \psi(x) = \left[ 1 - F(x) \right]^{-1} \) is the weight function. For computations, the statistic can be expressed in the form

\[
AU_n^2 = \frac{n}{2} - 2 \sum_{i=1}^{n} z(i) - \sum_{i=1}^{n} \left[ 2 - \left( \frac{2i - 1}{n} \right) \right] \log \left[ 1 - z(i) \right] .
\]

### 3.3 Test based on transforms (Meintanis and Bassiakos [48])

**Hypothesis**

We test the null hypothesis

\[ H_0 : X_1, \ldots, X_n \text{ is a random sample from } \text{GPD}(0, \sigma, \beta) \]

against the alternative hypothesis

\[ H_1 : X_1, \ldots, X_n \text{ is a random sample not from } \text{GPD}(0, \sigma, \beta). \]

**Test statistic**

The test statistic is

\[
T_n = n \int_{0}^{\infty} D_n^2(t) w(t) dt,
\]
where \( w(t) \) is a non-negative weight function,
\[
D_n(t) = (1 + t)L_n(t) - 1 \quad \text{on} \quad [0, \infty)
\]
and
\[
L_n(t) = \frac{1}{n} \sum_{j=1}^{n} \exp \left( -t\tilde{Y}_j \right),
\]
where \( \tilde{Y}_j = -\left(1/\tilde{\beta}\right) \log \left[1 - \left(\tilde{\beta}X_j/\tilde{\sigma}\right)\right] \) are independent exponential random variables, see (2). \( \tilde{\sigma} \) and \( \tilde{\beta} \) are the method of moment estimators given in Section 2.2.

**Null distribution of test statistic**

There exists a Gaussian element \( \mathcal{W} \) such that
\[
T_n = n \int_0^{\infty} D_n^2(t)w(t)dt \to \|\mathcal{W}\|^2
\]
as \( n \to \infty \), see Theorem 2.1 in [48] for a detailed proof.

**Rejection criteria of \( H_0 \)**

We reject \( H_0 \) if the test statistic \( T_n > K_\alpha \), where \( K_\alpha \) is the critical value for \( \alpha \). The null distribution of \( T_n \) is dependent on the unknown value of the shape parameter \( \beta \), therefore Section 2.3 can be used to obtain the critical value of the test.

Expressions for moment and probability weighted moment estimators of \( \beta \) and \( \sigma \) are given in Section 2.2. The estimators, \( \hat{\beta}_n \) and \( \hat{\sigma}_n \), are regular for \( \beta < 1/4 \) and \( \beta < 1/2 \), respectively. The maximum likelihood estimators, also given in Section 2.2, are regular for \( \beta > -1/2 \).

### 3.4 LAN based Neyman smooth test (Falk et al. [25])

This GOF test is motivated by LeCam’s theory of local asymptotic normality (LAN). Let \( f(x, \xi) \) and \( F(x, \xi) \) denote, respectively, the PDF and the CDF of GPD(0, \( \sigma, \beta \)), where \( \xi = (\sigma, \beta)^T \in \Theta = (0, \infty) \times (-\infty, \infty) \). \( f(x, \xi) \) can be combined into a \( J \)-dimensional exponential PDF
\[
g_J(x, \theta, \xi) = f(x, \xi) \exp \left\{ \sum_{s=1}^{J} \theta_s F^s(x, \xi) - K(\theta) \right\}, \quad (5)
\]
where \( \theta = (\theta_1, \ldots, \theta_J)^T \), \( F(x, \xi) = 1 - F(x, \xi) \) is the survival function, and \( K(\theta) \) is the normalising constant
\[
K(\theta) = \log \left\{ \int_0^{1} \exp \left( \sum_{s=1}^{J} \theta_s t^s \right) dt \right\}.
\]
Hypothesis

Let $X_1, \ldots, X_n$ be a random sample from (5). We test the hypothesis

$H_0 : \theta = 0$,

$H_1 : \theta \neq 0$.

Under the null hypothesis, the distribution of the random sample is GPD.

Test statistic

The test statistic is

$$\Psi^2_J = Z_n^T \left( \hat{\xi}_n \right) \Sigma_J^{-1} \left( \hat{\xi}_n \right) Z_n \left( \hat{\xi}_n \right),$$

where

$$Z_n \left( \hat{\xi}_n \right) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left[ \left( 1 - F \left( X_i; \hat{\xi}_n \right) \right)^s - \frac{1}{s+1} \right]_{s=1,\ldots,J}$$

and

$$\Sigma_J(\xi) = \left[ \frac{uv}{(u+v+1)(u+1)(v+1)} - \frac{uv(1+\beta)(uv+\beta+(u+1)(v+1))}{(v+\beta+1)(u+\beta+1)(u+1)^2(v+1)^2} \right]_{u,v=1,\ldots,J}.$$  

Null distribution and rejection criteria

Under $H_0$, the test statistic $\Psi^2_J$ converges weakly to a chi-square distribution with $J$ degrees of freedom as $n \to \infty$. In addition, the statistic $Z_n \left( \hat{\xi}_n \right)$ converges weakly to the normal distribution $N \left( 0, \Sigma_J(\xi) \right)$. The null hypothesis is rejected with significance level $\alpha$ if $\Psi^2_J > \chi^2_{J,\alpha}$.

3.5 Generalized smooth test (De Boeck et al. [23])

Let $f(x; \xi)$ denote the PDF of GPD(0, $\sigma, \beta$), where $\xi = (\sigma, \beta)^T$. Define the family

$$g_k(x; \theta, \xi) = C(\theta, \xi) \exp \left[ \sum_{j=1}^{k} \theta_j h_j(x; \xi) \right] f(x; \xi),$$

(6)

where $\theta^T = (\theta_1, \ldots, \theta_k)$ and $C(\theta, \xi)$ is a normalising constant. The polynomials $h_j$ are of degree $j$ and $\{h_j(\cdot; \xi), j = 0, \ldots, k\}$ form a set of orthonormal polynomials with respect to $f(\cdot; \xi)$ satisfying

$$\int_{-\infty}^{+\infty} h_i(x; \xi) h_j(x; \xi) f(x; \xi) dx = \begin{cases} 1, & \text{for } i = j, \\ 0, & \text{for } i \neq j. \end{cases}$$

Hypothesis

Let $X_1, \ldots, X_n$ be a random sample from (6). We test the hypothesis

$H_0 : \theta = 0$,

$H_1 : \theta \neq 0$.  

12
Test statistic

The generalized smooth test statistic is

\[ \hat{S}_k = \hat{V}^T \hat{\Sigma}_0^{-1} \hat{V} \]

provided \( \hat{\beta} < 1/(2k) \), where \( \hat{V}^T (\hat{\xi}) = \left( V_3 (\hat{\xi}), \ldots, V_k (\hat{\xi}) \right) \), \( \hat{\Sigma}_0 = \Sigma_0 (\hat{\xi}) \) is the asymptotic variance covariance matrix of \( \hat{V} \), \( \hat{\Sigma}_0 = \Sigma_0 (\hat{\xi}) \) and

\[ V_j (\xi) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} h_j (X_i; \xi) \]

for \( j = 1, 2, \ldots, k \).

Note that \( \Sigma_0 \) has no convenient form, however an explicit formula for \( k = 4 \) can be found in Appendix B in [23].

Null distribution and rejection criteria

Under the null hypothesis (in the case of testing for a GPD), the test statistic \( \hat{S}_k \) is asymptotically chi-square distributed with \( k - 2 \) degrees of freedom. Hence, the test rejects \( H_0 \) with significance level \( \alpha \) if \( \hat{S}_k > \chi^2_{k-2, \alpha} \).

3.6 Zhang’s \( Z_C \) statistic (Zhang and Stephens [63])

Hypothesis

We test the null hypothesis

\( H_0 : X_1, \ldots, X_n \) is a random sample from GPD \( (0, \sigma, \beta) \)

against the alternative hypothesis

\( H_1 : X_1, \ldots, X_n \) is a random sample not from GPD \( (0, \sigma, \beta) \).

Test statistic

The test statistic is

\[ Z_C = \sum_{i=1}^{n} \left[ \log \left( \frac{z_{(i)} - 1}{n/(i - 0.5) - 1} \right) \right]^2, \]

where \( z_{(i)} = F \left( x_{i:n}; \hat{\sigma}, \hat{\beta} \right) \). The estimates \( \hat{\sigma} \) and \( \hat{\beta} \) are given by

\[ \hat{\beta} = -\frac{1}{n} \sum_{i=1}^{n} \log \left( 1 - \psi X_i \right) \]

and

\[ \hat{\sigma} = \frac{\hat{\beta}}{\psi}, \]
where
\[
\hat{\psi} = \sum_{j=1}^{m} \psi_j w(\psi_j),
\]
\[
w(\psi_j) = 1/\sum_{t=1}^{m} \exp \left[ \ell(\psi_t) - \psi(\psi_j) \right],
\]
\[
\ell(\psi) = n \left[ \log(\psi/k) + k - 1 \right],
\]
\[
k = \frac{1}{n} \sum_{i=1}^{n} \log(1 - \psi X_i),
\]
\[
\psi_j = \frac{1}{X_{n:n}} + \frac{1}{3X_{n:[n/4+0.5]}} \left[ 1 - \sqrt{\frac{m}{j - 0.5}} \right].
\]
\(\hat{\psi}\) is not sensitive to \(m\) provided that \(m > 20\).

Null distribution and critical values

With the Zhang statistic \(Z_C\), it is difficult to obtain the exact null distribution of \(Z_C\) for finite samples. Section 2.3 can be used to approximate the null distribution and the \(p\)-value.

3.7 The kernel statistic (Beirlant et al. [8])

Hypothesis

We consider the hypothesis test given by

\[ H_0 : \text{the upper tail of } F \text{ behaves as } PI(\sigma, \beta), \]
\[ H_1 : \text{the upper tail of } F \text{ does not behave as } PI(\sigma, \beta), \]

where \(\sigma\) and \(\beta\) are unknown parameters.

Test statistic

The kernel GOF statistic is
\[
\frac{1}{kH_{k,n}} \sum_{i=1}^{k} K \left( \frac{i}{k + 1} \right) Z_i,
\]
where \(K\) denotes a kernel function satisfying
\[
\int_0^1 K(u)du = 0.
\]
Null distribution of the test statistic

Suppose there exists a real constant $\rho \leq 0$ and a rate function $b$ satisfying $b(x) \to 0$ as $x \to \infty$, such that for all $\lambda \geq 1$,

$$\frac{F^{-1} (1 - \frac{1}{X})}{F^{-1} (1 - \frac{1}{x})} - 1 \sim b(x) \frac{\lambda^{\rho} - 1}{\rho}$$

as $x \to \infty$. Let

$$K(t) = \frac{1}{t} \int_0^t u(v)dv$$

for some function $u$ satisfying

$$\left| k \int_{(j-1)/k}^{j/k} u(t)dt \right| \leq f \left( \frac{j}{k+1} \right)$$

for some positive continuous function $f$ defined on $(0,1)$ such that

$$\int_0^1 \max (1/w, 1) f(w)dw < \infty$$

in case $\rho < 0$, and

$$\int_0^1 w^{-\xi} f(w)dw < \infty$$

for some small $\xi > 0$ in case $\rho = 0$,

$$\int_0^1 |K(w)|^{2+\delta} dw < \infty$$

for some $\delta > 0$ and

$$\frac{1}{\sqrt{k}} \sum_{j=1}^k K \left( \frac{j}{k+1} \right) \to 0$$

as $k \to \infty$. Then as $k \to \infty$, $n \to \infty$, $k/n \to 0$ and $\sqrt{k} b(n/k) \to c$,

$$\frac{\sqrt{k}}{H_{k,n}} \frac{1}{k} \sum_{i=1}^k K \left( \frac{i}{k+1} \right) Z_i \overset{d}{\to} N \left( c \beta \int_0^1 K(u)u^{-\rho}du, \int_0^1 K^2(u)du \right)$$

as $n \to \infty$.

**Rule for rejecting $H_0$**

We reject $H_0$ at the significance level $\alpha$ if

$$\sqrt{k} \left| \frac{1}{kH_{k,n}} \sum_{i=1}^k K \left( \frac{i}{k+1} \right) Z_i - \frac{b(n/k)}{\beta} \int_0^1 K(u)u^{-\rho}du \right| > \Phi^{-1} \left( 1 - \frac{\alpha}{2} \right) \sqrt{\int_0^1 K^2(u)du}.$$
This rule is hard to use as it depends on the unknown parameters $c$, $\beta$ and $\rho$. To overcome this, we choose $k$ to be small enough such that $\sqrt{\hat{k}}b(n/k) \approx 0$. This leads us to reject $H_0$ if

$$\sqrt{k} \left| \frac{1}{kH_{k,n}} \sum_{i=1}^{k} K \left( \frac{i}{k+1} \right) Z_i \right| > \Phi^{-1} \left( 1 - \frac{\alpha}{2} \right) \sqrt{\int_0^1 K^2(u)du}.$$ 

The kernel function has maximal power against Weibull-type alternatives when the kernel maximises

$$\left| \int_0^1 K(u) \log(1/u)du \right|$$

subject to

$$\int_0^1 K(u) = 0, \quad \int_0^1 K^2(u)du = 1.$$

This is precisely when $K(u) = -1 - \log(u)$, also known as the Jackson kernel function. This is a special case of the kernel GOF statistic, which will be considered along with the special case of the Lewis kernel function in Section 3.11.

### 3.8 Bias corrected statistic (Beirlant et al. [8])

#### Hypothesis

We test the same hypotheses as in Section 3.7, so we follow the same notation.

#### Test statistic

The test statistic is

$$\frac{\hat{\beta}_{LS,k}(\hat{\rho})}{k} \sum_{i=1}^{k} K_{BC} \left( \frac{i}{k+1} \right) \left[ Z_i - \hat{b}_{LS,k}(\rho) \left( \frac{i}{k+1} \right)^{-\rho} \right]$$

with kernel function

$$K_{BC}(u; \rho) = K(u) - \frac{(1-\rho)^2(1-2\rho)}{\rho^2} \left( u^{-\rho} - \frac{1}{1-\rho} \right) \int_0^1 K(v)v^{-\rho}dv,$$

where

$$\hat{\beta}_{LS,k}(\hat{\rho}) = \left[ \frac{1}{k} \sum_{i=1}^{k} Z_i - \hat{b}_{LS,k}(\rho) \right]^{-1}$$

and

$$\hat{b}_{LS,k}(\hat{\rho}) = \frac{(1-\rho)^2(1-2\rho)}{\rho^2} \frac{1}{k} \sum_{i=1}^{k} \left[ \left( \frac{i}{k+1} \right)^{-\rho} - \frac{1}{1-\rho} \right] Z_i.$$
Null distribution

Suppose the assumptions of Section 3.7 hold with \( \rho < 0 \). Then, as \( k \to \infty, n \to \infty, k/n \to 0 \) and \( \sqrt{k}b(n/k) \to c \),

\[
\sqrt{k} \hat{\beta}_{LS,k}(\rho) \frac{1}{k} \sum_{i=1}^{k} K_{BC} \left( \frac{i}{k+1}; \rho \right) Z_i \to N \left( 0, \int_0^1 K_{BC}^2(u; \rho) du \right)
\]
as \( n \to \infty \). The significant difference between this result and the corresponding one in Section 3.7 is that the limiting distribution is now centred at zero.

Rule for rejecting \( H_0 \)

We reject \( H_0 \) when

\[
\left| \sqrt{k} \hat{\beta}_{LS,k}(\rho) \frac{1}{k} \sum_{i=1}^{k} K_{BC} \left( \frac{i}{k+1}; \rho \right) Z_i \right| > \Phi^{-1} \left( 1 - \frac{\alpha}{2} \right) \sqrt{\int_0^1 K_{BC}^2(u; \rho) du}.
\]

Estimation of \( \rho \)

As can be seen from the above result, the rejection region of \( H_0 \) is dependent on the unknown parameter \( \rho \). There are a number of options available to solve this problem. A simple solution is to fix \( \rho \) for a specific value: the choice \( \rho = -1 \) suggested in [26] and the value \( \rho = -1/\beta \) [29] are just two such values that can be used. However, there are some drawbacks to this solution; by using a fixed value for \( \rho \), the bias correcting effect of the statistic will be lost, i.e. the bias corrected statistics would no longer be normally distributed with zero mean. A good explanation of this is given on page 15 of [8].

Another option is to use the estimator suggested in [26], given by

\[
\hat{\rho}_{k,n} = \left| 3 \frac{T_n^{(\tau)}(\bar{k}) - 1}{T_n^{(\tau)}(\bar{k}) - 3} \right|,
\]

where

\[
T_n^{(\tau)}(\bar{k}) = \frac{\left[ M_n^{(1)}(\bar{k}) \right]^{\tau} - \left[ M_n^{(2)}(\bar{k}) / 2 \right]^{\tau/2}}{\left[ M_n^{(2)}(\bar{k}) / 2 \right]^{\tau/2} - \left[ M_n^{(3)}(\bar{k}) / 6 \right]^{\tau/3}}
\]

and

\[
M_n^{(j)}(k) = \frac{1}{k} \sum_{i=1}^{k} \left[ \log (X_{n-i+1:n}) - \log (X_{n-i:n}) \right]^j
\]

for \( j \geq 1 \), as defined in [8]. One disadvantage of \( \hat{\rho}_{k,n} \) is its complexity and the unknown parameter \( \tau \). [26] recommends the following values for \( \tau \): 0, 0.5, 1, 2. [26] also provides a detailed explanation of the effect that the choice of \( \tau \) may have on the \( \rho \)-estimator and an overall explanation of \( \hat{\rho}_{k,n} \).
3.9 The Jackson kernel function (Jackson [36])

Hypothesis
This is a special case of the test in Section 3.7, the hypotheses are the same and we follow the same notation.

Test statistic
The Jackson statistic is

\[ T_J = \sqrt{\frac{k}{H_{k,n}}} \left( \frac{i}{k+1} \right) Z_i, \]

where \( K_J(u) = -1 - \log(u) \).

Null distribution of the test statistic
With the assumptions of Section 3.7, as \( k \to \infty, n \to \infty, k/n \to 0 \) and \( \sqrt{kb(n/k)} \to c \),

\[ \frac{1}{\sqrt{kH_{k,n}}} \sum_{j=1}^{k} \left( -1 - \log \frac{j}{k+1} \right) Z_j \overset{d}{\to} N \left( \frac{c\rho\beta}{(1-\rho)^2}, 1 \right) \]

as \( n \to \infty \).

Rule for rejecting \( H_0 \)
We reject \( H_0 \) when

\[ \sqrt{k} \left| \frac{1}{kH_{k,n}} \sum_{i=1}^{k} K \left( \frac{i}{k+1} \right) Z_i - \frac{b(n/k)(1-\rho)}{(1-\rho)^2} \right| > \Phi^{-1} \left( 1 - \frac{\alpha}{2} \right). \]

By choosing \( k \) small enough such that \( \sqrt{kb(n/k)} \approx 0 \), the expression above can be simplified so that we reject \( H_0 \) if

\[ \sqrt{k} \left| \frac{1}{kH_{k,n}} \sum_{i=1}^{k} K \left( \frac{i}{k+1} \right) Z_i \right| > \Phi^{-1} \left( 1 - \frac{\alpha}{2} \right). \]

3.10 Bias corrected Jackson kernel function (Jackson [36])

Hypothesis
This is a special case of the test in Section 3.8, the hypotheses are the same and we follow the same notation.
Test statistic

Using the special case of the Jackson statistic as a kernel function and the method of bias correction described in Section 3.8, we find the bias corrected Jackson statistic

\[ T_{BCJ} = \frac{\hat{\beta}_{LS,k}(\hat{\rho})}{k} \sum_{i=1}^{k} \left[ 1 - \log \left( \frac{i+1}{k+1} \right) \right] \left[ Z_i - \hat{b}_{LS,k}(\hat{\rho}) \left( \frac{i}{k+1} \right) \right], \]

where \( \hat{\rho} \) is a consistent estimator for \( \rho \).

Null distribution of the test statistic

With the assumptions of Section 3.7, as \( k \to \infty, n \to \infty, k/n \to 0 \) and \( \sqrt{k}(n/k) \to c \),

\[ \sqrt{k}(T_{BCJ} - 2) \xrightarrow{d} N \left( 0, \left( \frac{\rho}{1-\rho} \right)^2 \right) \]

as \( n \to \infty \). Note that the limiting distribution is centred at zero compared to the basic Jackson kernel statistic.

Rule for rejecting \( H_0 \)

We reject \( H_0 \) when

\[ \left| \sqrt{k}(T_{BCJ} - 2) \right| > \left( \frac{\rho}{1-\rho} \right) \Phi^{-1} \left( 1 - \frac{\alpha}{2} \right). \]

For estimation methods of the unknown parameter \( \rho \), see Section 3.8.

3.11 Lewis kernel function (Lewis [45])

Hypothesis

This is a special case of the test in Section 3.7, the hypotheses are the same and we follow the same notation.

Test statistic

The Lewis kernel statistic is

\[ T_L = \frac{1}{\sqrt{kH_{k,n}}} \sum_{i=1}^{k} K_L \left( \frac{i}{k+1} \right) Z_i, \]

where \( K_L(u) = u - 0.5 \).
Null distribution of the test statistic

With the assumptions of Section 3.7, as \( k \to \infty, n \to \infty, k/n \to 0 \) and \( \sqrt{k}b(n/k) \to c \),

\[
\frac{1}{\sqrt{k}H_{k,n}} \sum_{i=1}^{k} K_L \left( \frac{i}{k+1} \right) Z_i \xrightarrow{d} N \left( -\frac{c\rho\beta}{2(1-\rho)(2-\rho)}, \frac{1}{12} \right)
\]

as \( n \to \infty \).

Rule for rejecting \( H_0 \)

Using the null distribution, we reject \( H_0 \) when

\[
\sqrt{12k} \left| \frac{1}{k H_{k,n}} \sum_{i=1}^{k} K_L \left( \frac{i}{k+1} \right) Z_i + \frac{b(n/k)\rho\beta}{2(1-\rho)(2-\rho)} \right| > \Phi^{-1} \left( 1 - \frac{\alpha}{2} \right).
\]

3.12 Bias corrected Lewis kernel function (Lewis [45])

Hypothesis

This is a special case of the test in Section 3.8, the hypotheses are the same and we follow the same notation.

Test statistic

This is analogous to the bias corrected Jackson kernel statistic, and so the statistic is simply

\[
T_{BCL} = \frac{\hat{\beta}_{LS,k}(\hat{\rho})}{k} \sum_{i=1}^{k} \left( \frac{i}{k+1} - \frac{1}{2} \right) \left[ Z_i - \hat{\beta}_{LS,k}(\hat{\rho}) \left( \frac{i}{k+1} \right)^{-\hat{\rho}} \right].
\]

Null distribution of the test statistic

Suppose the assumptions of Section 3.7 hold with \( \rho \neq -1 \). Then, as \( k \to \infty, n \to \infty, k/n \to 0 \) and \( \sqrt{k}b(n/k) \to c \),

\[
\sqrt{k} (T_{BCL} - 2) \xrightarrow{d} N \left( 0, \frac{1}{12} \left( \frac{1+\rho}{2-\rho} \right)^2 \right)
\]

as \( n \to \infty \).

Rule for rejecting \( H_0 \)

We reject \( H_0 \) at significance level \( \alpha \) when

\[
\left| \sqrt{k} (T_{BCL} - 2) \right| > \frac{1}{\sqrt{12}} \left( \frac{1+\rho}{2-\rho} \right)^2 \Phi^{-1} \left( 1 - \frac{\alpha}{2} \right).
\]
3.13 Two tests based on a characterization of the Pareto distribution (Obradović et al. [49])

Hypothesis

We test the null hypothesis that
\[ H_0 : X_1, X_2, \ldots, X_n \text{ is a random sample from } \text{PI}(1, \beta) \]
against the alternative hypothesis
\[ H_1 : X_1, X_2, \ldots, X_n \text{ is a random sample not from } \text{PI}(1, \beta). \]

Test statistics

There are two test statistics: \( T(n) \) and \( V(n) \) given by
\[
T(n) = \int_1^\infty [M_n(t) - F_n(t)] dF_n(t)
\]
and
\[
V(n) = \sup_{t \geq 1} |M_n(t) - F_n(t)|,
\]
respectively, where
\[
M_n(t) = \left( \frac{n}{2} \right)^{-1} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} I \left\{ \max \left( \frac{X_i}{X_j} X_j, X_i \right) \leq t \right\}
\]
for \( t \geq 1 \).

Critical values of \( H_0 \)

For small sample sizes, we do not have an exact distribution for \( T(n) \). Therefore, the critical values of the test can be calculated using Section 2.3.

The asymptotic distribution of \( V(n) \) is unknown, although using [58] we are able to show that the random process \( \rho(t) = \sqrt{n} [M_n(t) - F_n(t)] \), \( t \geq 1 \), converges to some Gaussian process. Although, as for the Cramer von Mises and Anderson Darling statistics (see [19]), it is very difficult to calculate the covariance matrix for this process. Therefore, in most cases, it is more effective to calculate the approximate null distribution using Section 2.3.

3.14 Test based on spacings (Gulati and Shapiro [32])

Hypothesis

We test the null hypothesis that
\[ H_0 : X_1, X_2, \ldots, X_n \text{ is a random sample from } \text{PI}(\sigma, \beta) \]
against the alternative hypothesis
\[ H_1 : X_1, X_2, \ldots, X_n \text{ is a random sample not from } \text{PI}(\sigma, \beta). \]
Test statistic

The test statistic is
\[ \Lambda_0 = \Lambda_1^2 + \Lambda_2^2, \]
where
\begin{align*}
\Lambda_1 &= \sqrt{\frac{12}{n-1}} \left( \frac{U - 1}{2} \right), \\
\Lambda_2 &= \sqrt{\frac{5}{4(n+2)(n-1)(n-2)}} \left( n - 2 + 6nU - 12 \sum_{i=1}^{n-1} \frac{iU_i}{n-1} \right), \\
U &= \frac{\sum_{i=1}^{n-1} U_i}{n-1},
\end{align*}
where \( U_i = \frac{Y_i^*}{n}, Y_i^* = \sum_{j=1}^{i} Y_j \) and \( Y_i = (n - i + 1) [X_{i:n} - X_{i-1:n}] \) with the convention \( X_{0:n} = 0 \).

Null distribution of the test statistic

The limiting distribution of \( \Lambda_0 \) under the null hypothesis is the chi-squared distribution with two degrees of freedom, i.e. \( \chi^2_2 \).

Rule for rejecting \( H_0 \)

We reject the null hypothesis when \( \Lambda_0 > \chi^2_2(\alpha) \) for a chosen significance level \( \alpha \). Using the fact that \( \Lambda_0 \) also has an exponential distribution of mean 2, we are able to calculate the \( p \)-value.

3.15 Euclidean distances method (Rizzo [56])

Hypothesis

We test the null hypothesis that
\[ H_0 : X_1, X_2, \ldots, X_n \text{ is a random sample from } \text{PI}(\sigma, \beta) \]
against the alternative hypothesis
\[ H_1 : X_1, X_2, \ldots, X_n \text{ is a random sample not from } \text{PI}(\sigma, \beta). \]

Test statistic

The test statistic is
\[ Q_\gamma = n \left\{ \frac{2}{n} \sum_{i=1}^{n} E\|T_i - T\|^\gamma - E\|T - T'\|^\gamma - \frac{1}{n^2} \sum_{i,j=1}^{n} \|T_i - T_j\|^\gamma \right\}, \]
where $\| \cdot \|$ denotes the Euclidean norm, $T_j = \log X_j$, and $T$ and $T'$ are independent $\text{PI}(\sigma, \beta)$ random variables. The exponent $\gamma$ is chosen such that $X^\gamma$ has finite variance, so that $\gamma < \beta/2$.

The moments defined above can be calculated using the following formulas given in Section 3 of [56]:

$$
E \| t - T \|^\gamma =
\begin{cases}
  t + \frac{2\sigma^\beta t^{1-\beta} - \beta \sigma}{\beta - 1}, & \text{if } t \geq \sigma, \gamma = 1, \\
  (t - \sigma)^\beta + \sigma^\beta t, & \text{if } t \geq \sigma, \gamma = \beta - 1, \\
  (t - \sigma)^\beta - \sigma^\beta \frac{\gamma B_{y_0}(\gamma, 1-\beta) - \beta B(\beta - \gamma, \gamma + 1)}{\beta - \gamma}, & \text{if } t \geq \sigma, 0 < \gamma < \beta < 1,
\end{cases}
$$

and

$$
E \| T - T' \|^\gamma =
\begin{cases}
  \frac{2\sigma^\beta}{(\beta - 1)(2\beta - 1)}, & \text{if } t \geq \sigma, \gamma = 1, \\
  \frac{\beta + 1}{2\beta + \sigma^\beta - 1}, & \text{if } t \geq \sigma, \gamma = \beta - 1, \\
  \frac{2\beta^2 \sigma^\gamma B(\beta - \gamma, \gamma + 1)}{\beta - \gamma}, & \text{if } t \geq \sigma, 0 < \gamma < \beta < 1,
\end{cases}
$$

where $y_0 = (t - \sigma)/t$, $B(a, b)$ denotes the beta function, and $2F_1(a, b; c; z)$ denotes the Gauss hypergeometric function.

**Null distribution of the test statistic**

When $H_0$ is true and the variance of the Pareto distribution is finite, $Q_\beta$ converges to the quadratic form

$$
\sum_{i=1}^{\infty} \lambda_i \Omega_i^2
$$

as $n \to \infty$, where $\Omega_i$ are independent standard normal random variables. Notice here the similarity between this and the $\chi^2$ distribution.

**Rule for rejecting $H_0$**

The critical points of $Q_\beta$ can be determined by Section 2.3. The null hypothesis can be rejected if the observed $Q_\beta$ is greater than the critical point at a given significance level.
3.16 Test based on a property of order statistics (Volkova [61])

Hypothesis

We test the null hypothesis that

\[ H_0 : X_1, X_2, \ldots, X_n \text{ is a random sample from } \text{PI}(1, \beta) \]

against the alternative hypothesis

\[ H_1 : X_1, X_2, \ldots, X_n \text{ is a random sample not from } \text{PI}(1, \beta). \]

Test statistic

Two test statistics can be defined as

\[ I_n^{(k)} = \int_1^\infty [H_n(t) - F_n(t)] dF_n(t) \]

and

\[ D_n^{(k)} = \sup_{t \geq 1} |H_n(t) - F_n(t)|, \]

where

\[ H_n(t) = \binom{n}{k}^{-1} \sum_{1 \leq i_1 < \cdots < i_k \leq n} I \left\{ X_{(k, \{i_1, \ldots, i_k\})}/X_{(k-1, \{i_1, \ldots, i_k\})} < t \right\} \]

for \( t \geq 1 \) and \( X_{(s, \{i_1, \ldots, i_k\})} \) denotes the \( s \)-th order statistic of the subsample \( X_{i_1}, \ldots, X_{i_k} \).

Null distribution of the test statistic

Due to complications in the calculations for the general case, the specific cases of \( k = 3 \) and \( k = 4 \) are considered here as examples. Under \( H_0 \), the statistics \( \sqrt{n}I_n^{(3)} \) and \( \sqrt{n}I_n^{(4)} \) are asymptotically normal such that

\[ \sqrt{n}I_n^{(3)} \xrightarrow{d} N \left( 0, \frac{11}{120} \right) \]

and

\[ \sqrt{n}I_n^{(4)} \xrightarrow{d} N \left( 0, \frac{271}{2100} \right) \]

as \( n \to \infty \).

The limiting distributions of \( D_n^{(3)} \) and \( D_n^{(4)} \) are unknown, however, it can be shown that

\[ \eta_n(t) = \sqrt{n} [H_n(t) - F_n(t)] \]

weakly converges as \( n \to \infty \) to a centred Gaussian process \( \eta(t), t \geq 1 \) of which the covariance can be calculated. It follows that the sequence of statistics \( \sqrt{n}D_n^{(3)} \) and \( \sqrt{n}D_n^{(4)} \) converge in distribution to the random variable \( \sup_{t \geq 1} |\eta(t)| \). However it is not currently possible to find an explicit distribution for this. Therefore, critical values for \( \sqrt{n}D_n^{(3)} \) and \( \sqrt{n}D_n^{(4)} \) can be determined by Section 2.3.
Rejection criteria of $H_0$

In the case of the integral statistics, $I_n^{(3)}$ and $I_n^{(4)}$, we reject $H_0$ if

$$\left| \sqrt{n}I_n^{(3)} \right| > \Phi^{-1}\left(1 - \frac{\alpha}{2}\right) \sqrt{\frac{11}{120}}$$

and

$$\left| \sqrt{n}I_n^{(4)} \right| > \Phi^{-1}\left(1 - \frac{\alpha}{2}\right) \sqrt{\frac{271}{2100}},$$

respectively, at the $\alpha$ significance level.

In the case of the Kolmogorov type statistic $D_n^{(k)}$, the limiting distribution is not known. However, critical values for the statistic can be determined using Section 2.3.

3.17 Kullback-Leibler divergence (Lequesne [44])

Hypothesis

We test the null hypothesis that

$H_0 : X_1, X_2, \ldots, X_n$ is a random sample from $\text{PI}(\sigma, \beta)$

against the alternative hypothesis

$H_1 : X_1, X_2, \ldots, X_n$ is a random sample not from $\text{PI}(\sigma, \beta)$.

Test statistic

The test statistic is

$$T_{m,n} = \exp \left( V_{m,n} - 1 - \log \frac{\hat{\sigma}}{\hat{\beta}} - \frac{1}{\hat{\beta}} \right),$$

where

$$V_{m,n} = \frac{1}{n} \sum_{i=1}^{n} \log \left\{ \frac{n}{2m} [X_{i+m:n} - X_{i-m:n}] \right\}$$

is Vasicek’s entropy estimator, and $m$ is a window size taking a positive integer smaller than $n/2$. Estimators for $\sigma$ and $\beta$ are given in [44], which appear to be the same as the maximum likelihood estimators for $\sigma$ and $\beta$ given in Section 2.2.

Null distribution and critical values

The critical region of the test is $[0, C(\alpha, \sigma, \beta)]$, where $C(\alpha, \sigma, \beta)$ is the critical value determined by the $\alpha$th quantile of the distribution of $T_{m,n}$ under $H_0$, and $\alpha$ is the significance level. The distribution of $T_{m,n}$ under $H_0$ is unknown, however $C(\alpha, \sigma, \beta)$ can be computed by Section 2.3.
3.18 Weighted quantile correlation test for Pareto families (Csörgő and Szabó, [21])

Hypothesis

We test the null hypothesis that

\( H_0 : X_1, X_2, \ldots, X_n \) is a random sample from \( \text{PI} (\sigma, \beta) \)

against the alternative hypothesis

\( H_1 : X_1, X_2, \ldots, X_n \) is a random sample not from \( \text{PI} (\sigma, \beta) \).

We assume \( \sigma \) is known.

Test statistic

The test statistic is

\[
W_n = \frac{\pi^2}{6} - \frac{6}{n} \sum_{k=1}^{n} R_{k:n}^2 \int_{(k-1)/n}^{k/n} \log \frac{1}{t} \, dt - \left[ \sum_{k=1}^{n} R_{k:n} \int_{(k-1)/n}^{k/n} \log \frac{1}{t} \, dt \right]^2 - 2 \sum_{k=1}^{n} R_{k:n} \left[ \int_{(k-1)/n}^{k/n} \left( - \log \log \frac{1}{t} \right) \log \frac{1}{t} \, dt - (\gamma - 1) \int_{(k-1)/n}^{k/n} \log \frac{1}{t} \, dt \right],
\]

where \( \gamma \) is Euler’s constant, \( R_{1:n} \leq \cdots \leq R_{n:n} \) are the order statistics of \( -\log \log (X_i/\sigma) \), \( i = 1, \ldots, n \).

Null distribution and rejection criteria

Under the null hypothesis, we have

\[
nW_n - c_n^\circ \xrightarrow{D} W,
\]

where \( c_n^\circ = \log n + \gamma + o(1) \), \( \gamma \) is Euler’s constant and

\[
W \overset{D}{=} -1 + \sum_{k=2}^{\infty} \frac{\Omega_k^2 - 1}{k},
\]

where \( \Omega_1, \Omega_2, \ldots \) are independent standard normal random variables.

Define the limiting distribution function as \( H_1(x) = \Pr (W \leq x) \). \( H_1(\cdot) \) and the critical value can be determined by Section 2.3.

3.19 Empirical distribution function tests (Brazauskas and Serfling [12], [13])

Hypothesis

We test the null hypothesis that

\( H_0 : X_1, X_2, \ldots, X_n \) is a random sample from \( \text{PI} (\sigma, \beta) \)

against the alternative hypothesis
\[ H_1 : X_1, X_2, \ldots, X_n \text{ is a random sample not from } \text{PI}(\sigma, \beta). \]

We assume \( \sigma \) is known.

Test statistics

The test statistics are the Kolmogorov Smirnov, Cramer von Mises and Anderson Darling statistics defined by

\[
D_n = \max \left( D_n^+, D_n^- \right), \\
W_n^2 = \sum_{j=1}^{n} \left[ F_n(X_{j:n}) - \frac{2j - 1}{2n} \right]^2 + \frac{1}{12n}, \\
A_n^2 = -n - \frac{1}{n} \sum_{j=1}^{n} \left\{ (2j - 1) \log F_n(X_{j:n}) + (2n + 1 - 2j) \log \left[ 1 - F_n(X_{j:n}) \right] \right\},
\]

where

\[
D_n^+ = \max_{1 \leq j \leq n} \left[ \frac{j}{n} - F_n(X_{j:n}) \right], \\
D_n^- = \max_{1 \leq j \leq n} \left[ F_n(X_{j:n}) - \frac{j - 1}{n} \right].
\]

When the parameter \( \beta \) is estimated via the method of maximum likelihood (see Section 2.2) critical values and formulas for significance levels for \( D_n, W_n^2 \) and \( A_n^2 \) can be found from Tables 4.11 and 4.12 in [22].

3.20 Test based on maximum likelihood and probability weighted moments (Gulati and Shapiro [32])

Hypothesis

We test the null hypothesis that

\[ H_0 : X_1, X_2, \ldots, X_n \text{ is a random sample from } \text{PII}(0, \sigma, \beta) \]

against the alternative hypothesis

\[ H_1 : X_1, X_2, \ldots, X_n \text{ is a random sample not from } \text{PII}(0, \sigma, \beta). \]

Estimation and test procedure

As stated in Section 2.2, the maximum likelihood estimators of \( \sigma \) and \( \beta \) involve non-linear equations. The non-linear equations may not always yield a root. In this case, the method of probability weighted moments estimators stated in Section 2.2 can be used.

The test procedure is as follows

1. Set \( \sigma = 1 \) to start, the estimation procedure is not affected by the initial value;
2. Generate \( X_1, X_2, \ldots, X_n \) from the PII distribution;
3. Compute the maximum likelihood estimates of $\sigma$ and $\beta$, see Section 2.2;

4. If they do not exist, compute the probability weighted moment estimates of $\sigma$ and $\beta$, see Section 2.2;

5. If the probability weighted moment estimate \( \hat{\beta} \) is negative, then set \( \hat{\beta} = -0.005 \) and \( \tilde{\sigma} = 0.005/\hat{\sigma} \);

6. Transform using

\[
T_i = \log \left( \frac{1 + X_i}{\tilde{\sigma}} \right), \quad i = 1, 2, \ldots, n;
\]

7. Use \( T_i \)s as input to the test procedure of Section 3.14, substitute \( T_i \)s for the \( X_i \)s, and calculate the test statistic \( \Lambda_0 \).

Null distribution and rejection criteria

Under \( H_0 \), the test statistic \( \Lambda_0 \) is approximately chi-squared distributed with one degree of freedom \( (\chi^2_1) \). Note that the chi-squared distribution with one degree of freedom is the square of a standard normal distribution and so the \( p \)-value of the test can be found from normal tables.

3.21 Test based on the transformed sample Lorenz curve (Kang and Cho [37])

Hypothesis

We test the null hypothesis that

\( H_0 : X_1, X_2, \ldots, X_n \) is a random sample from PII(\( \mu, \sigma, \beta \))

against the alternative hypothesis

\( H_1 : X_1, X_2, \ldots, X_n \) is a random sample not from PII(\( \mu, \sigma, \beta \)).

Test statistic

The test statistic is based on the normalised sample Lorenz curve, expressed as

\[
TS = \frac{TSL(p)}{TSL'(p)},
\]

where \( p = i/n, i = 1, 2, \ldots, n \) and

\[
TSL(p) = \sum_{j=1}^{i} (X_{j:n} - X_{1:n}) - p + 1 \over \sum_{j=1}^{n} (X_{j:n} - X_{1:n})
\]
and

\[
TSL'(p) = \sum_{j=1}^{i} \left[ \left( 1 - \frac{j}{n+1} \right)^{\hat{\beta}} - \left( 1 - \frac{1}{n+1} \right)^{\hat{\beta}} \right] - p + 1.
\]

Null distribution and rejection criteria

Due to the exact distribution of \(TS\) being difficult to calculate, critical values of \(TS\) can be determined by Section 2.3.

4 Variations

Some of the tests in this section apply for progressively type II censoring data. The progressively type II censoring scheme can be described as follows: let \(n\) be the units in a lifetime study, and assume that \(m\) (\(\leq n\)) be fixed in advance. Assume also that \(m\) non-negative integers \(R_1, R_2, \ldots, R_m\) are fixed in advance, so that \(R_1 + \cdots + R_m + m = n\). Let \(X_{1:m:n}\) denote the time of \(i\)th failure. When the first failure occurs at time \(X_{1:m:n}\), \(R_1\) surviving units are removed at random. At the time of the second failure, \(X_{2:m:n}\), \(R_2\) surviving units are removed at random. This continues until the time of the \(m\)th failure, \(X_{m:m:n}\), when all the remaining \(R_m\) surviving units are removed.

4.1 Probability plot correlation coefficient (Kim et al. [38])

Hypothesis

We test the null hypothesis

\(H_0 : X_1, \ldots, X_n\) is a random sample from GPD \((\mu, \sigma, \beta)\)

against the alternative hypothesis

\(H_1 : X_1, \ldots, X_n\) is a random sample not from GPD \((\mu, \sigma, \beta)\).

Test statistic

Define

\[
M_i = \Phi^{-1}(m_i),
\]

where \(m_1 = 1 - (0.5)^{1/n}\), \(m_n = (0.5)^{1/n}\) and \(m_i = (i - 0.3175)/(n + 0.365)\), \(i = 2, 3, \ldots, n - 1\). Let \(r\) denote the correlation coefficient

\[
r = \frac{\sum_{i=1}^{n} (X_i - \bar{X}) (M_i - \bar{M})}{\sqrt{\sum_{i=1}^{n} (X_i - \bar{X})^2 \sum_{i=1}^{n} (M_i - \bar{M})^2}},
\]
where $\overline{X}$ and $\overline{M}$ are the mean values of $X_i$ and $M_i$, respectively.

The test statistic, denoted as $r_\alpha(n)$, is based on the correlation coefficient and is derived using the following method:

1. Generate $X_1, \ldots, X_n$ from the GPD, with given parameters;
2. Calculate $M_i$ using (7);
3. Calculate the correlation coefficient $r$ between the $X_i$ and $M_i$;
4. Repeat steps 1-3, 100,000 times to generate 100,000 values of $r$;
5. Choose the 100,000-th smallest $r$ as $r_\alpha$, where $\alpha$ is the significance level chosen.

Rejection criteria for $H_0$

We reject $H_0$ for a given sample (of size $n$) if $r_\alpha(n) > r_\alpha$ at the $\alpha$ level of significance.

4.2 Percentile residual (PR) plot (Brazauskas and Kleefeld [11])

Hypothesis

We test the null hypothesis

$H_0 : X_1, \ldots, X_n$ is a random sample from GPD ($\mu, \sigma, \beta$)

against the alternative hypothesis

$H_1 : X_1, \ldots, X_n$ is a random sample not from GPD ($\mu, \sigma, \beta$).

PR plot

The PR graph plots the empirical percentile levels $(j/n)100\%$ against the standardised residuals given by

$$R_{j,n} = \frac{X_{j:n} - \hat{F}^{-1}\left(\frac{j - 0.5}{n}\right)}{\text{standard deviation of } \hat{F}^{-1}\left(\frac{j - 0.5}{n}\right)}$$

for $j = 1, 2, \ldots, n$, where $\hat{F}^{-1}(p) = \hat{F}^{-1}\left(p; \hat{\mu}, \hat{\sigma}, \hat{\beta}\right)$.

Various estimation methods and their restrictions are discussed in [11]. Suppose $\sqrt{n}\left(\hat{\mu} - \mu, \hat{\sigma} - \sigma, \hat{\beta} - \beta\right) \rightarrow N_3\left(0, \Sigma\right)$ for a variance-covariance matrix $\Sigma$. By the delta method, the standard deviation in the denominator of (8) can be estimated by

$$\frac{1}{\sqrt{n}} \left[ \left( \frac{\partial \hat{F}^{-1}(p)}{\partial \hat{\mu}}, \frac{\partial \hat{F}^{-1}(p)}{\partial \hat{\sigma}}, \frac{\partial \hat{F}^{-1}(p)}{\partial \hat{\beta}} \right) \Sigma \left( \frac{\partial \hat{F}^{-1}(p)}{\partial \hat{\mu}}, \frac{\partial \hat{F}^{-1}(p)}{\partial \hat{\sigma}}, \frac{\partial \hat{F}^{-1}(p)}{\partial \hat{\beta}} \right)^T \right].$$
where \( p = (j - 0.5)/n \). Various expressions for \( \Sigma \) are discussed in [11]. Two of them are: if \( \beta > -1/2 \) then

\[
\Sigma_0 = (1 + \beta) \left( \begin{array}{cc} 2\sigma^2 & \sigma \\ \sigma & 1 + \beta \end{array} \right);
\]

if \( \beta < 1/4 \) then

\[
\Sigma_1 = \frac{(1 - \beta)^2}{(1 - 3\beta)(1 - 4\beta)} \left( \begin{array}{ccc} 2\sigma^2 (1 - 6\beta + 12\beta^2)/(1 - 2\beta) & \sigma (1 - 4\beta + 12\beta^2) \\ \sigma (1 - 4\beta + 12\beta^2) & (1 - 2\beta)(1 - \beta + 6\beta^2) \end{array} \right).
\]

Tolerance limits can be plotted above and below 0, to assist in determining GOF. A good fit of the GPD is indicated if the majority (or ideally all) of the points are distributed between the tolerance limits, for a given estimation method.

4.3 Trimmed mean absolute deviation (tMAD) (Brazauskas and Kleefeld [11])

Hypothesis

For this test, the hypothesis is the same as in the case for Section 4.2, the PR plot.

Test statistic and rejection rule

The trimmed mean absolute deviation measures the absolute distance between the fitted GPD quantiles and the observed data. The statistic is defined as

\[
\Delta_\delta = \frac{1}{[n\delta]} \sum_{i=1}^{[n\delta]} b_{i:n},
\]

where \( b_{i:n} \) is the \( i \)th smallest distance among

\[
\left| X_{j:n} - \hat{F}^{-1} \left( \frac{j - 0.5}{n} \right) \right|, \quad j = 1, 2, \ldots, n,
\]

and \( \hat{F}^{-1} \) is as defined in Section 4.2. The parameter \( \delta \) measures how far, on average, the \( 100\delta\% \) closest observations are from their corresponding fitted quantiles.

The critical points of \( \Delta_\delta \) can be determined by Section 2.3. We reject \( H_0 \) if the observed \( \Delta_\delta \) is greater than the critical point at a given significance level.

4.4 Test for exponentiality versus generalized Pareto (Brilhante [15])

Hypothesis

Let \( X_1, \ldots, X_n \) be a random sample from GPD(0, \( \sigma, \beta \)). We test the hypothesis

\[
H_0 : \beta = 0 \text{ (the sample is exponentially distributed)},
\]

\[
H_1 : \beta \neq 0 \text{ (the sample is Pareto distributed)}.
\]
Test statistic

The test statistic is given as

\[ T_n = \frac{X_{n-\{\frac{n}{4}\}+1:n} - X_{\frac{n}{4}:n}}{X_{\frac{n}{4}:n} - X_{\{\frac{n}{4}\}:n}}, \]

if \( n \) is odd, and

\[ T_n = \frac{X_{n-\{\frac{n}{4}\}+1:n} - \frac{1}{2}\left(X_{\frac{n}{4}:n} + X_{\frac{n}{4}+1:n}\right)}{\frac{1}{2}\left(X_{\frac{n}{4}:n} + X_{\frac{n}{4}+1:n}\right) - X_{\{\frac{n}{4}\}:n}}, \]

if \( n \) is even, where \( \{a\} \) denotes the integer closest to \( a \).

Null distribution and rejection criteria

For a large sample size, the test statistic \( T_n \) has a limiting normal distribution under \( H_0 \), and we have that

\[ \log\left(\frac{3}{2}\right)\sqrt{\frac{n}{2}}\left(T_n - \frac{\log 2}{\log\left(\frac{3}{2}\right)}\right) \rightarrow Z \sim N(0,1) \]

as \( n \to \infty \). We reject \( H_0 \) if

\[ \left|\log\left(\frac{3}{2}\right)\sqrt{\frac{n}{2}}\left(T_n - \frac{\log 2}{\log\left(\frac{3}{2}\right)}\right)\right| > \Phi^{-1}\left(1 - \frac{\alpha}{2}\right), \]

at the \( \alpha \) significance level, critical values can also be determined by Section 2.3.

4.5 Test procedure for the shape parameter of a GPD (Chaouche and Bacro [18])

Hypothesis

Let \( X_1, \ldots, X_n \) be a random sample from GPD\((0,\sigma,\beta)\). We test the hypothesis

\[ H_0 : \beta = \beta_0, \]
\[ H_1 : \beta \neq \beta_0. \]

Test statistic

Inference on \( \beta \) uses two test statistics which are invariant of \( \sigma \). Both are based on probability weighted moments, see Section 2.2. The first is

\[ T_0 \equiv \frac{(s+1)^2M_{10s}}{(s+1)M_{10s} - \max X_i}, \]
where 
\[ M_{10s} = \frac{\hat{\sigma}}{(1 + s) \left( 1 + s - \hat{\beta} \right)}. \]

The second test statistic is
\[ T_1 = \inf_{k \in K} (\beta_{1k}), \]

where
\[ K = \{k; X_k > 20M_{113}\}, \]
\[ M_{113} = \frac{\hat{\sigma}}{\hat{\beta}} \left[ B \left( 2, s - \hat{\beta} + 1 \right) - B(2, s + 1) \right], \]

and $\beta_{1k}$ is the smallest root of the following inequality
\[ [AM_{11s} - X_i] \beta^2 - (2s + 3) [AM_{11s} - X_i] \beta + A^2 M_{11s} > 0 \]
for all $i = 1, 2, \ldots, n$, where $A = (s + 1)(s + 2)$.

### Null distribution and rejection criteria

The distributions of the test statistics are unknown, however, the $p$-values can be obtained by Section 2.3.

#### 4.6 Test using the cumulative hazard function ([Saldaña-Zepeda et al. [57]])

Like many GOF tests for the GPD, this test is based on the relationship between the Pareto distribution and the exponential distribution, see Section 1. Furthermore, the test is only applicable to ungrouped data with type II right censoring; type II right censoring is the case where both the sample size and the number of censored observations are chosen in advance of the data being recorded.

**The cumulative hazard function (CHF)**

Let $X$ denote a random variable with CDF $F$. Its CHF is defined by
\[ H(x) = -\log [1 - F(x)]. \]

As explained in [57], an estimator of the CHF is
\[ \hat{H}(x) = \sum_{x_{i:n} \leq x} \frac{d_i}{Y_i}, \]

where $d_i$ denotes the number experiencing the event of interest at time $x_{i:n}$ and $Y_i$ is the number who are at risk immediately before $x_{i:n}$, i.e. those who have not yet experienced the event and have not been censored. This is known as the Nelson-Aalen (N-A) estimator. Under type II right censoring, the estimator becomes
\[ \hat{H}(x_{i:n}) = \sum_{j=1}^{i} \frac{1}{n - j + 1}. \]
Hypothesis

We test the null hypothesis

$$H_0 : X_1, X_2, \ldots, X_n \text{ is a random sample from } PI(\sigma, \beta)$$

versus the alternative hypothesis

$$H_1 : X_1, X_2, \ldots, X_n \text{ is a random sample not from } PI(\sigma, \beta).$$

Test statistic

The test makes the following transformations:

1. $$W(i) = \log(X_{i:n}), \ i = 1, 2, \ldots, r;$$

2. $$Z(i) = W(i+1) - W(1), \ i = 1, 2, \ldots, r - 1.$$

When $$H_0$$ is true, the ordered samples $$Z(i)$$ are distributed as a $$(r - 1)$$ sample of the $$\text{Exp}(\sigma)$$ distribution.

Using the above transformations, the test statistic is defined to be the sample correlation coefficient between the N-A estimator of the CHF and $$Z$$, given by

$$R_{N-A} = \frac{\sum_{i=1}^{r-1} (Z(i) - \bar{Z}) \left[ \hat{H}(Z(i)) - \bar{H} \right]}{\sqrt{\sum_{i=1}^{r-1} (Z(i) - \bar{Z})^2} \sqrt{\sum_{i=1}^{r-1} \left[ \hat{H}(Z(i)) - \bar{H} \right]^2}},$$

where $$\bar{Z}$$ and $$\bar{H}$$ are the means of $$Z(i)$$ and $$\hat{H}(Z(i))$$, respectively.

Null distribution of the test statistic

There is no definitive distribution of the test statistic under the null hypothesis, therefore we find the approximate distribution of $$R_{N-A}$$ by Section 2.3. Despite this, the distribution of $$R_{N-A}$$ is independent of the shape parameter $$\beta$$.

Define $$V(i) = \beta Z(i)$$, where $$Z(i)$$ is as defined above. Then the $$V(i)$$’s are distributed as order statistics from a standard exponential distribution. If we now consider $$R_{N-A}$$ in terms of $$V$$, we have

$$R_{N-A} = \frac{\sum_{i=1}^{r-1} (V(i) - \bar{V}) \left( \hat{H} \left( \frac{V(i)}{\beta} \right) - \bar{H} \right)}{\sqrt{\sum_{i=1}^{r-1} (V(i) - \bar{V})^2} \sqrt{\sum_{i=1}^{r-1} \left[ \hat{H} \left( \frac{V(i)}{\beta} \right) - \bar{H} \right]^2}}.$$

However, the distribution of $$R_{N-A}$$ is dependent on the percentage of censoring. For example, when $$H_0$$ is true and the censoring level is low, the distribution is closer to 1; as the censoring level

34
increases, the distribution becomes more dispersed. For greater detail on this, we refer the reader to [57], where Figure 1 shows the effect of the level of censoring on $R_{N-A}$, where the data sample is from the family of Pareto distributions with changing scale and shape parameters. In addition, Figure 1 gives further evidence that $R_{N-A}$ is independent of $\beta$.

**Rejection of $H_0$**

Under the null hypothesis $R_{N-A}$ should have a value close to 1. Therefore, we reject $H_0$ at the significance level $\alpha$ if $R_{N-A} < K_\alpha$, where $K_\alpha$ is such that $P[\text{Reject } H_0|H_0] = P[R_{N-A} < K_\alpha|H_0] \leq \alpha$. $K_\alpha$ can be determined numerically. Saldaña-Zepeda et al. [57] show that $K_{0.05} = 0.9568$ for $n = 30$ for example.

**4.7 A graphical test (Amin [1])**

**Hypothesis**

We test the null hypothesis that

$H_0 : X_1, X_2, \ldots, X_n$ is a random sample from $\Pi(\sigma, \beta)$

against the alternative hypothesis

$H_1 : X_1, X_2, \ldots, X_n$ is a random sample not from $\Pi(\sigma, \beta)$.

**Test statistic**

Let $h(t) = f(t)/[1 - F(t)]$ denote the hazard rate function. Under the null hypothesis, $\log h(t)$ should be linear against $\log t$, thus a test is to plot $\log h(t)$ versus $\log t$ and see if it is linear. Estimates of the hazard rate function can be found through various methods. One such method is the Kimball estimator [1] [39].

If the plot is approximately linear, then we are able to obtain an estimate for the parameter $\beta$. $\log(\beta)$ is the intercept on the graph, so we can estimate $\beta$ from the intercept using the least squares method.

**4.8 Kullback-Leibler information (Rad et al. [54])**

This test is based on Kullback-Leibler information, and is only applicable to progressively type II censored data.

**Hypothesis**

We test the null hypothesis that

$H_0 : X_1, X_2, \ldots, X_n$ is a random sample from $\Pi(\sigma, \beta)$

against the alternative hypothesis

$H_1 : X_1, X_2, \ldots, X_n$ is a random sample not from $\Pi(\sigma, \beta)$.
Test statistic

The test statistic is

\[ T(w, n, m) = -H(w, n, m) - \frac{1}{n} \left\{ \sum_{i=1}^{m} \log f(x_i; \hat{\sigma}, \hat{\beta}) + \sum_{i=1}^{m} R_i \log \left[ 1 - F(x_i; \hat{\sigma}, \hat{\beta}) \right] \right\}, \]

where

\[ H(w, n, m) = \frac{1}{n} \sum_{i=1}^{m} \log \left[ \frac{x_{i+w:m:n} - x_{i-w:m:n}}{E(X_{i+w:m:n}) - E(X_{i-w:m:n})} \right] - \left( 1 - \frac{m}{n} \right) \log \left( 1 - \frac{m}{n} \right) \]

and \( w \) is an optimal window size. The unknown parameters \( \sigma \) and \( \beta \) can be estimated by the method of maximum likelihood, see Section 2.2.

Null distribution of the test statistic

As the sampling distribution of the test statistic \( T(w, n, m) \) is difficult to deal with, percentiles can be determined by Section 2.3. The test statistic \( T(w, n, m) \) is a function of \( w \), in addition to \( n \) and \( m \). \( w \) in turn is dependent on \( n \) and \( m \), and is chosen optimally so that it gives the minimum critical value.

Rejection criteria of \( H_0 \)

By simulating progressively type II censored samples from the Pareto distribution, values of the distribution of \( T(w, n, m) \) and the critical value can be obtained, see Section 2.3.

4.9 Test based on multiply truncated samples (Marlin [47])

Hypothesis

We test the null hypothesis that

\( H_0 : X_1, X_2, \ldots, X_n \) is a random sample from \( \text{PI}(\sigma, \beta) \)

against the alternative hypothesis

\( H_1 : X_1, X_2, \ldots, X_n \) is a random sample not from \( \text{PI}(\sigma, \beta) \).

Test statistic

Suppose \( x_i \) would not have appeared in the sample had it not been less than the “truncation point” \( d_i > 0 \). Set

\[ y_i = \log \left( \frac{x_i}{d_i} \right), \quad i = 1, 2, \ldots, n \]

and let \( y_{0:n} = 0 \). Let

\[ t_i = (n + 1 - i) [y_{i:n} - y_{i-1:n}] \]
denote the normalized differences. Under $H_0$, the $t_i$'s are independent exponential random variables with parameter $\beta$. Split the sample of normalised differences into two subsets $S_1$ and $S_2$, containing $r_1$ and $r_2$ observations, respectively. Then the test statistic can be defined as

\[ Q = \frac{T_1}{r_1} / \frac{T_2}{r_2}, \]

where

\[ T_j = \sum_{i \in S_j} t_i, \quad j = 1, 2, \]

which are the sums of independent exponential random variables, and are gamma distributed with shape and scale parameters $r_j$ and $1/\beta$, respectively.

**Null distribution of the test statistic**

The test statistic $Q$ has an $F$ distribution with $v_1$ degrees of freedom in the numerator, and $v_2$ in the denominator, where $v_i = 2r_i$.

**Rejection criteria of $H_0$**

If the alternative hypothesis $H_1$ is unspecified, then $H_0$ is rejected at the $\alpha$ level of significance if

\[ Q < F(\alpha/2, v_1, v_2), \quad \text{or} \quad Q > F(1 - \alpha/2, v_1, v_2), \]

where $F(p, v_1, v_2)$ is the 100$p$th percentile of the $F$ distribution, with $v_1$ and $v_2$ degrees of freedom. If $H_1$ specifies a distribution with an increasing (decreasing) hazard rate, then this implies that the value of the test statistic $Q$ will be greater (less) than 1. For example, if the hazard rate specified by $H_1$ is increasing, then $H_0$ is rejected if $Q > F(1 - \alpha, v_1, v_2)$.

**4.10 Test for Pareto law based on the Lagrange multiplier (Goerlich [28])**

**Hypothesis**

We test the null hypothesis that

$H_0 : X_1, X_2, \ldots, X_n$ is a random sample from $PI(\sigma, \beta)$

against the alternative hypothesis

$H_1 : X_1, X_2, \ldots, X_n$ is a random sample from $PII(\mu, \sigma, \beta)$.

We assume $\sigma$ is known.

**Test statistic**

The test statistic is

\[ LM_P = n \frac{\left( \hat{\beta} + 2 \right) \left( \hat{\beta} + 1 \right)^4 \bar{z}^2}{\hat{\beta}}, \]
where

\[ \hat{\beta} = \left( \frac{1}{n} \sum_{i=1}^{n} \log \frac{x_i}{\sigma} \right)^{-1} \]

and

\[ z = \frac{\hat{\beta}}{\beta + 1} - \frac{1}{n} \sum_{i=1}^{n} \frac{\sigma}{x_i} \]

Null distribution and rejection criteria

Under the null hypothesis, the test statistic \( LM_P \) is asymptotically chi-square distributed with one degree of freedom \( \chi^2_1 \). \( H_0 \) is rejected if \( LM_P \) is larger than the critical point of \( \chi^2_1 \) at a given significance level.

4.11 Preliminary test for the Pareto distribution (Baklizi [7])

Hypothesis

Suppose \( X_1, X_2, \ldots, X_n \) is a random sample from PI \( (\sigma, \beta) \). We test the null hypothesis that

\[ H_0 : \beta = \beta_0 \]

against the alternative hypothesis

\[ H_1 : \beta \neq \beta_0. \]

Null distribution

The maximum likelihood estimator of \( \beta \), say \( \hat{\beta} \), is given in Section 2.2. It can be shown that

\[ \frac{2n\beta}{\hat{\beta}} \sim \chi^2_{2(n-1)} \]

Rejection criteria

We reject \( H_0 \) with significance level \( \alpha \) if \( \frac{2n\beta}{\hat{\beta}} > c_1 \), or \( \frac{2n\beta}{\hat{\beta}} < c_2 \), where \( c_1 \) and \( c_2 \) are such that

\[ \Pr \left( \chi^2_{2(n-1)} < c_1 \right) = \Pr \left( \chi^2_{2(n-1)} > c_2 \right) = \frac{\alpha}{2}. \]

5 Simulation study

We compare all of the tests in Section 3 by simulation. We have however excluded the following tests: the Cramer von Mises, Anderson Darling and modified Anderson Darling tests as they have the same spirit as the Kolmogorov Smirnov test; the bias corrected statistic, the Jackson kernel function, the bias corrected Jackson kernel function, the Lewis kernel function and the bias corrected Lewis kernel function tests as they are particular cases of the kernel statistic test. So, we compare six tests for the GPD, eight tests for the PI distribution and the two tests for the PII distribution.
The comparison is based on simulated power functions. The simulated power of the six tests for the GPD versus $\beta = -0.05, -0.049, \ldots, 0.05$ when the null distribution is GP(0, 1, 0) are plotted in Figure 1 for $n = 100, 200$. Also plotted in Figure 1 are the simulated power of the six tests for the GPD versus $\sigma = 0.01, 0.02, \ldots, 1$ when the null distribution is GP(0, 1, 0). The following abbreviations and coloring scheme have been used: the Kolmogorov Smirnov test abbreviated as KS and colored in black; the intersection union test abbreviated as Boot and colored in red; the test based on transforms abbreviated as Trans and colored in blue; the LAN based Neyman smooth test abbreviated as LAN and colored in green; the generalized smooth test abbreviated as Smooth and colored in brown; Zhang’s $Z_C$ statistic test abbreviated as Zhang and colored in pink.

The simulated power of the eight tests for the PI distribution versus $\beta = 0.01, 0.02, \ldots, 1$ when the null distribution is PI(1, 1) are plotted in Figure 2 for $n = 100, 200$. Also plotted in Figure 2 are the simulated power of the eight tests for the PI distribution versus $\sigma = 0.01, 0.02, \ldots, 1$ when the null distribution is PI(1, 1). The following abbreviations and coloring scheme have been used: the Kolmogorov Smirnov test abbreviated as KS and colored in black; the kernel statistics test abbreviated as Kernel and colored in red; tests based on a characterization of the Pareto distribution abbreviated as Char and colored in blue; the test based on spacings abbreviated as Space and colored in green; the Euclidean distances method test abbreviated as Eucl and colored in brown; the test based on property of order statistics abbreviated as Order and colored in pink; the Kullback-Leibler divergence test abbreviated as KL and colored in yellow; the weighted quantile correlation test abbreviated as Quantile and colored in orange.

The simulated power of the two tests for the PII distribution versus $\beta = 0.01, 0.02, \ldots, 1$ when the null distribution is PII(0, 1, 1) are plotted in Figure 3 for $n = 100, 200$. Also plotted in Figure 3 are the simulated power of the two tests for the PII distribution versus $\sigma = 0.01, 0.02, \ldots, 1$ when the null distribution is PII(0, 1, 1). The following abbreviations and coloring scheme have been used: the test based on maximum likelihood and probability weighted moments abbreviated as Weighted and colored in black; the test based on the transformed sample Lorenz curve abbreviated as Lorenz and colored in red. All computations were performed in the R statistical software (R Development Core Team [53]).

The simulated power functions were computed as follows:

1. set the parameter values (GPD, PI or PII);

2. simulate a random sample of size $n$ from the distribution (GPD, PI or PII) for the set parameter values;

3. estimate the parameters from the simulated sample, the method of maximum likelihood was used;

4. test the hypothesis that the sample comes from the null distribution at the five percent level of significance;

5. repeat steps 2 and 3 ten thousand times;

6. compute the power as the proportion of the number of times that the null distribution was rejected.

This procedure was repeated for every set of parameter values. The standard errors of the proportion were generally less than 0.01.
Figure 1: Power functions of tests for the GPD: versus $\beta$ when the null distribution is GPD($0, 1, 0$) and $n = 100$ (top left); versus $\beta$ when the null distribution is GPD($0, 1, 0$) and $n = 200$ (top right); versus $\sigma$ when the null distribution is GPD($0, 1, 0$) and $n = 100$ (bottom left); versus $\sigma$ when the null distribution is GPD($0, 1, 0$) and $n = 200$ (bottom right).
Figure 2: Power functions of tests for the PI distribution: versus $\beta$ when the null distribution is PI(1, 0) and $n = 100$ (top left); versus $\beta$ when the null distribution is PI(1, 0) and $n = 200$ (top right); versus $\sigma$ when the null distribution is PI(1, 0) and $n = 100$ (bottom left); versus $\sigma$ when the null distribution is PI(1, 0) and $n = 200$ (bottom right).
Figure 3: Power functions of tests for the PII distribution: versus $\beta$ when the null distribution is $\text{PII}(0, 1, 1)$ and $n = 100$ (top left); versus $\beta$ when the null distribution is $\text{PII}(0, 1, 1)$ and $n = 200$ (top right); versus $\sigma$ when the null distribution is $\text{PII}(0, 1, 1)$ and $n = 100$ (bottom left); versus $\sigma$ when the null distribution is $\text{PII}(0, 1, 1)$ and $n = 200$ (bottom right).

We can observe the following from Figure 1: the Kolmogorov Smirnov test gives the best performance; the intersection union gives the second best performance; the remaining four tests give an equal third best performance. We can observe the following from Figure 2: the Kolmogorov Smirnov test gives the best performance; the kernel statistic test gives the second best performance; the remaining six tests give an equal third best performance. The test based on maximum likelihood and probability weighted moments gives the better performance in Figure 3.

These observations are for the specified set of parameter values. The observations were similar for a wide range of other parameter values and a wide range of other non-null distributions (Weibull, gamma, lognormal, Burr, inverse Gaussian, etc). In particular, the Kolmogorov Smirnov test always
gave the best performance for the GPD and the PI distribution.

What sample sizes give reasonable approximations to the asymptotic distributions of the Kolmogorov Smirnov statistics? At what sample sizes is it better to simulate the critical values? What is the effect on the asymptotic critical values from using different methods of estimating the relevant parameters? Guidance on these questions can be found in Buning [16], Buning [17], Evans et al. [24], and Hrabakova and Kus [33].

**Acknowledgments**

The authors would like to thank the Editor and the two referees for careful reading and comments which greatly improved the paper.

**References**


