Filomat 29:10 (2015), 2311–2324
DOI 10.2298/FIL1510311O

Published by Faculty of Sciences and Mathematics,
University of Niš, Serbia
Available at: http://www.pmf.ni.ac.rs/filomat

# On Asymptotic Efficiency of Goodness of Fit Tests for Pareto Distribution Based on Characterizations

**Marko Obradović**[a]

[a]*Faculty of Matemathics, University of Belgrade, Studentski trg 16*

**Abstract.** In this paper we present a characterization for Pareto distribution. We propose two new goodness of fit tests based on this characterization and two tests based on Rossberg's characterization. We calculate their Bahadur efficiencies against different alternatives and compare the tests. We also find a class of locally optimal alternatives for each test.

## 1. Introduction

The Pareto distribution plays a very important role in probability and statistics. It is used to model various quantities in economics, finance, actuaries, hydrology and many other fields. One of the most important tasks in those applications is to ensure that the Pareto distribution is the appropriate for modeling, since in many situations it is important to realize whether the underlying distribution is Pareto or some other strongly skewed to the right. This is usually checked using goodness of fit tests.

A characterization of a family of distributions is a property that is true only for that family. See [6] for more on characterizations and [7] for characterizations of Pareto distribution. Since the characterizations are a good way to distinguish one family from the others, they are useful in goodness-of fit testing. However, creating tests based on characterizations is relatively new and recently popular approach in goodness of fit testing theory. Such tests are often free of some parameters and thus suitable for testing composite hypotheses. Some examples of such tests can be found in [1], [4], [11], [17]. The asymptotic efficiency of the exponentiality tests based on a characterization have been studied in [13], [16], [23], and for power function distribution in [24]. Asymptotic efficiency of tests for Pareto distribution based on a characterization are studied in [18].

The structure of the paper is the following. In section 2 we present the characterization theorems and propose four test statistics. In section three we study the asymptotic behaviour of integral type test statistics. We find the Bahadur efficiency of the tests against some common alternatives. We also find a class of locally optimal alternatives. In section 4 we do the analogous study for Kolmogorov-Smirnov type statistics. In section 5 we compare the Bahadur efficiencies of our new tests among each other and with some other tests based on different characterization.

2010 *Mathematics Subject Classification*. Primary 62G10 ; Secondary 62G20, 62H05
*Keywords*. Pareto distribution, Bahadur efficiency, V-statistics, Goodness of fit tests
Received: 01 September 2014; Accepted: 14 December 2014
Communicated by Danijela Cvetković Ilić
*Email address:* marcone@matf.bg.ac.rs (Marko Obradović)

## 2. The characterizations and the test statistics

Let $\mathcal{P}$ be the family of Pareto distributions with the distribution function

$$F(x) = 1 - x^{-\alpha}, x \geq 1, \ \ \alpha > 0.$$

We are going to test the composite null hypohesis $F \in \mathcal{P}$ against $F \notin \mathcal{P}$.
Let $X_{(k;n)}$ be the $k$ th order statistic of the sample of size $n$.
We now present two characterization theorems we will use to make our tests.

**Theorem 2.1 (Rossberg,1972).** *Let $X_1, \ldots, X_n$ be non-negative i.i.d random sample. If for some $j$ statistics $X_{(j+1;n)}/X_{(j;n)}$ and $X_{(1;n-j)}$ are identically distributed, then X belongs to the family $\mathcal{P}$.*

The proof can be found in [20]. We will use this characterization for special case of $n = 3$ and $j = 1$.

**Theorem 2.2.** *Let $X_1, X_2$ and $X_3$ be i.i.d. non-negative absolutely continuous random variables with strictly monotone distribution function and monotonically increasing or decreasing hazard function. Then, $X_{(3;3)}/X_{(2;3)}$ and $(X_{(2;3)}/X_{(1;3)})^2$ have the same distribution if and only if the distribution of X belongs to the family $\mathcal{P}$.*

**Proof.** Let $Y_k = \ln X_k, \ k = 1, 2, 3$. Since the logarithm is a monotonous transformation, then the statement of the theorem can be reformulated that $Y_{(3;3)} - Y_{(2;3)}$ and $2(Y_{(2;3)} - Y_{(1;3)})$ have the same distribution. This is a particular case of theorem of [2] (for $k = 3, \ i = 1$ and $j = 2$), where it was proven that this property charaterizes the exponential distribution with some parameter $\alpha$. Thus our theorem characterizes the Pareto distribution with the same parameter $\alpha$. □

The reason for choosing these special cases of characterization theorems for building our test statistics is that they are the simplest and thus convenient for practical applications of the tests. We chose to present these two characterizations since they are similar in the sense that they are based on ratio of consecutive order statistics.

Let $(X_1, \ldots, X_n)$ be a sample from non-negative continuous distribution $F$. Let $F_n(t)$ be the usual empirical distribution function. Following Rossberg's characterization theorem 2.1, we introduce some so-called $V$-empirical distribution functions:

$$G_n(t) = n^{-3} \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{k=1}^{n} I\{X_{(2),X_i,X_j,X_k}/X_{(1),X_i,X_j,X_k} \leq t\}, t \geq 1$$

and

$$H_n(t) = n^{-2} \sum_{i=1}^{n} \sum_{j=1}^{n} I\{\min\{X_i, X_j\} \leq t\}, t \geq 1,$$

where $X_{(l),X_a,X_b,X_c}, l = 1, 2$ is the lth order statistic within sample $(X_a, X_b, X_c)$.
We now introduce two tests statistics:

$$I_n^{[1]} = \int_1^{\infty} (G_n(t) - H_n(t)) dF_n(t)$$

$$D_n^{[1]} = \sup_{t \geq 1} |G_n(t) - H_n(t)|.$$

Based on the characterization from theorem 2.2 we define in the analogous way the $V$-empirical distribution functions:

$$J_n(t) = n^{-3}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n} I\{X_{(3),X_i,X_j,X_k}/X_{(2),X_i,X_j,X_k} \le t\}, t \ge 1$$

and

$$K_n(t) = n^{-3}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n} I\{(X_{(2),X_i,X_j,X_k}/X_{(1),X_i,X_j,X_k})^2 \le t\}, t \ge 1,$$

and the corresponding test statistics:

$$I_n^{[2]} = \int_1^\infty (J_n(t) - K_n(t))dF_n(t)$$

$$D_n^{[2]} = \sup_{t\ge 1} |J_n(t) - K_n(t)|.$$

For all our tests we consider large values of test statistics to be significant.

## 3. Integral type statistics

In this section we examine the properties of integral statistics $I_n^{[1]}$ and $I_n^{[2]}$.

Without loss of generality we may assume that $\alpha = 1$. The statistics $I_n^{[1]}$ is a $V$-statistic with the following kernel:

$$\Upsilon^{[1]}(X_1, X_2, X_3, X_4) = \frac{1}{4}\sum_{\pi(j_1,j_2,j_3,j_4)} I\{X_{(2);X_{j_1},X_{j_2},X_{j_3}}/X_{(1);X_{j_1},X_{j_2},X_{j_3}} \le X_{j_4}\} - \frac{1}{12}\sum_{\pi(j_1,j_2,j_3)} I\{\min\{X_{j_1}, X_{j_2}\} \le X_{j_3}\}$$

where $\pi(j_1, j_2, j_3, j_4)$ is the set of all permutations of the set $\{1, 2, 3, 4\}$ and $\pi(j_1, j_2, j_3)$ is set of all 3-permutations of the same set.

The projection of this kernel under null Pareto hypothesis is

$$\begin{aligned} v^{[1]}(s) &= E(\Upsilon^{[1]}(X_1, X_2, X_3, X_4)|X_1 = s) \\ &= \frac{1}{4}P\{X_{(2);X_2,X_3,X_4}/X_{(1);X_2,X_3,X_4} \le s\} + \frac{3}{4}P\{X_{(2);s,X_2,X_3}/X_{(1);s,X_2,X_3} \le X_4\} \\ &- \frac{1}{4}P\{\min\{X_2, X_3\} \le X_4\} - \frac{1}{4}P\{\min\{X_2, X_3\} \le s\} - \frac{1}{2}P\{\min\{s, X_2\} \le X_3\}. \end{aligned}$$

The first and the fourth term are equal due to the theorem 2.1 so we get

$$\begin{aligned} v^{[1]}(s) &= \frac{3}{4}(2P\{s < X_2 < X_3, \frac{X_2}{s} \le X_4\} + 2P\{X_2 < s < X_3, \frac{s}{X_2} \le X_4\} + 2P\{X_2 < X_3 < s, \frac{X_3}{X_2} \le X_4\}) \\ &- \frac{1}{2}(1 - P\{s > X_2 > X_3\} - P\{X_2 > s > X_3\}) - \frac{1}{4}\cdot\frac{2}{3} \\ &= \frac{1}{4}\frac{3\ln s - \frac{1}{2}}{s^2} - \frac{1}{24}. \end{aligned}$$

It is easy to calculate that the mean of this projection is equal to zero. Its variance is

$$\sigma_1^2 = Var(v^{[1]}(X)) = \int_1^{\infty} (v^{[1]}(s))^2 \frac{1}{s^2} ds = \frac{13}{4500}. \quad (1)$$

Since the variance of this projection is positive, the kernel $\Upsilon(X_1, X_2, X_3, X_4)$ is non-degenerate, so we can apply Hoeffding theorem for $U$ and $V$ statistics with non-degenerate kernels, see [8]. We get the following asymptotic distribution:

$$\sqrt{n} I_n^{[1]} \xrightarrow{d} \mathcal{N}\Big(0, \frac{52}{1125}\Big).$$

The statistic $I^{[2]}$ is the $V$-statistic with the kernel

$$\begin{aligned}\Upsilon^{[2]}(X_1, X_2, X_3, X_4) &= \frac{1}{4} \sum_{\pi(j_1,j_2,j_3,j_4)} I\{X_{(3);X_{j_1},X_{j_2},X_{j_3}} / X_{(2);X_{j_1},X_{j_2},X_{j_3}} \le X_{j_4}\} \\ &- \frac{1}{4} \sum_{\pi(j_1,j_2,j_3,j_4)} I\{(X_{(2);X_{j_1},X_{j_2},X_{j_3}} / X_{(1);X_{j_1},X_{j_2},X_{j_3}})^2 \le X_{j_4}\}.\end{aligned}$$

Its projection under null hypothesis is

$$\begin{aligned} v^{[2]}(s) &= E(\Upsilon^{[2]}(X_1, X_2, X_3, X_4)|X_1 = s) \\ &= \frac{1}{4} P\{X_{(3);X_2,X_3,X_4} / X_{(2);X_2,X_3,X_4} \le s\} + \frac{3}{4} P\{X_{(3);s,X_2,X_3} / X_{(2);s,X_2,X_3} \le X_4\} \\ &- \frac{1}{4} P\{(X_{(2);X_2,X_3,X_4} / X_{(1);X_2,X_3,X_4})^2 \le s\} - \frac{3}{4} P\{(X_{(2);s,X_2,X_3} / X_{(1);s,X_2,X_3})^2 \le X_4\}.\end{aligned}$$

The first and the third term cancel each other out due to the theorem 2.2 so we get

$$\begin{aligned} v^{[2]}(s) &= \frac{3}{4}\Big(2P\{s < X_2 < X_3, \frac{X_3}{X_2} \le X_4\} + 2P\{X_2 < s < X_3, \frac{X_3}{s} \le X_4\} + 2P\{X_2 < X_3 < s, \frac{s}{X_3} \le X_4\}\Big) \\ &- \frac{3}{4}\Big(2P\{s < X_2 < X_3, \Big(\frac{X_2}{s}\Big)^2 \le X_4\} + 2P\{X_2 < s < X_3, \Big(\frac{s}{X_2}\Big)^2 \le X_4\} + 2P\{X_2 < X_3 < s, \Big(\frac{X_3}{X_2}\Big)^2 \le X_4\}\Big) \\ &= \frac{3}{4} \cdot \frac{2\ln s - 1}{s} + \frac{1}{s^3} - \frac{1}{4}.\end{aligned}$$

The expected value of this projection is zero and its variance is

$$\sigma_2^2 = \int_1^{\infty} (v^{[2]}(s))^2 \frac{1}{s^2} ds = \frac{19}{4200}.$$

As in the previous case, the kernel is non-degenerate and due to the Hoeffding's theorem the asymptotic distribution is

$$\sqrt{n} I_n^{[2]} \xrightarrow{d} \mathcal{N}\Big(0, \frac{38}{525}\Big).$$

Since the kernels are non-degenerate we can consider instead of V-statistics $I_n^{[1]}$ and $I_n^{[2]}$ the corresponding U-statistics with same kernels. They have the same limiting distribution and large deviation asymptotics but U-statistics are easier for calculation.

*3.1. Bahadur efficiency*

One way of measuring asymptotic efficiency is Bahadur efficiency. Its advantage over other types of asymptotic efficiency is that it is applicable in cases where the null distribution is not normal, e.g. Kolmogorov-Smirnov type tests.

The Bahadur theory is explained in [3], [12]. For two tests with the same null and alternative hypotheses the asymptotic relative Bahadur efficiency is defined as the ratio of sample sizes needed to reach the same test power when the level of significance approaches zero. It can be expressed as the ratio of Bahadur exact slopes, provided that these functions exist.

The Bahadur exact slope (see [12]) can be evaluated as

$$c_T(\theta) = 2f(b_T(\theta)), \tag{2}$$

where $T_n \xrightarrow[n\to\infty]{p} b_T(\theta)$ for $\theta \in \Theta_1$, $G_n(t) = \inf\{P_\theta\{T_n < t\}, \theta \in \Theta_0\}$ and $f(t) = -\lim\limits_{n\to\infty} \frac{1}{n} \ln(1 - G_n(t))$ for each $t$ from an open interval $I$ on which $f$ is continuous and $\{b_T(\theta), \theta \in \Theta_1\} \subset I$. For Bahadur exact slope the following inequality holds:

$$c_T(\theta) \leq 2K(\theta),$$

where $K(\theta)$ is the Kullback-Leibler information number which measures the statistical distance between the alternative and the null hypothesis. The absolute Bahadur efficiency is defined as

$$e_T(\theta) = \frac{c_T(\theta)}{2K(\theta)}. \tag{3}$$

In most cases the Bahadur efficiency is not computable for any alternative $\theta$. However, it is possible to calculate the limit of Bahadur efficiency when $\theta$ approaches some $\theta_0 \in \Theta_0$. This limit is called the local asymptotic Bahadur efficiency.

Let $G(x;\theta)$ be a family of distributions such that $G(x;0) \in \mathcal{P}$ and $G(x;\theta) \notin \mathcal{P}$ for $\theta \neq 0$. Then we can reformulate our null hypothesis to be $H_0 : \theta = 0$. For close alternatives, the local asymptotic Bahadur efficiency is

$$e_T = \lim_{\theta\to 0} \frac{c_T(\theta)}{2K(\theta)}. \tag{4}$$

In what follows we shall calculate the local asymptotic Bahadur efficiency for some alternatives and find locally optimal alternatives. Let $\mathcal{G} = \{G(x;\theta)\}$ be a class of alternatives that satisfy the condition that it is possible to differentiate along $\theta$ under integral sign in all appearing integrals. Let $g(x;\theta)$ be a density of a distribution which belongs to $\mathcal{G}$, and let $H(x) = G'_\theta(x;0)$ and $h(x) = g'_\theta(x;0)$. It it easy to see that $\int\limits_1^\infty h(x)dx = 0$.

Let us now calculate the Bahadur exact slope for test statistic $I_n^{[1]}$. The functions $f_{(t)}$ and $b_T(\theta)$ will be determined from the following lemmas.

**Lemma 3.1.** *Let $t > 0$. For statistic $I_n^{[1]}$ the function $f(t)$ is analytic for sufficiently small $t > 0$ and it holds*

$$f(t) = \frac{1125}{104}t^2 + o(t^2), \quad t \to 0.$$

**Proof.** Since the kernel $\Upsilon^{[1]}$ is bounded, centered, and non-degenerate, applying the theorem on large deviations for non-degenerate U-statistics (see [15], Theorem 2.3) we get the statement of the lemma. □

**Lemma 3.2.** *For a given alternative density $g(x;\theta)$ whose distribution belongs to $\mathcal{G}$ holds*

$$b_{I^{[1]}}(\theta) = 4\theta \int_1^\infty v^{[1]}(x)h(x)dx + o(\theta), \ \theta \to 0. \tag{5}$$

**Proof.** By the law of large numbers for U-statistics ([22]) we get that

$$b_{I^{[1]}}(\theta) = P_\theta\{X_{(2;3)}/X_{(1;3)} \leq X_4\} - P_\theta\{\min\{X_1, X_2\} \leq X_3\}$$

Then

$$\begin{aligned} b_{I^{[1]}}(\theta) &= \int_1^\infty \int_1^\infty \int_y^{yz} 6(1-G(x;\theta))g(x;\theta)g(y;\theta)g(z;\theta)dxdydz - \int_1^\infty \int_x^\infty 2(1-G(x;\theta))g(x;\theta)g(y;\theta)dydx \\ &= 3\Big(1 - \int_1^\infty \int_1^\infty (1-G(yz;\theta))^2)g(y;\theta)g(z;\theta)dydz\Big) - 2\int_1^\infty (1-G(x;\theta))^2 g(x;\theta)dx. \end{aligned}$$

Since the second term is a constant its derivative is equal to zero and the first derivative of $b_{I^{[1]}}(\theta)$ at $\theta = 0$ is

$$\begin{aligned} b'_{I^{[1]}}(0) &= 6\int_1^\infty \int_1^\infty \frac{H(yz)}{y^3z^3}dydz - 6\int_1^\infty \int_1^\infty \frac{h(y)}{y^2z^4}dydz \\ &= 6\Big(\int_1^\infty \int_1^x \int_{\frac{x}{y}}^\infty \frac{h(x)}{y^3z^3}dxdydz + \int_1^\infty \int_x^\infty \int_1^\infty \frac{h(x)}{y^3z^3}dxdydz - \frac{1}{3}\int_1^\infty \frac{h(y)}{y^2}dy\Big) \\ &= \int_1^\infty \frac{h(x)}{x^2}(3\ln x - \frac{1}{2})dx = 4\int_1^\infty v^{[1]}(x)h(x)dx. \end{aligned}$$

Since $b_{I^{[1]}}(0) = 0$ using Maclaurin expansion we obtain (5). □

**Example 3.3.** *Let the alternative hypothesis be the log-Weibull distribution with distribution function*

$$G(x;\theta) = 1 - e^{-(\ln x)^{\theta+1}}, \quad x \geq 1, \theta \in (0,1). \tag{6}$$

*The first derivative along $\theta$ of its density at $\theta = 0$ is*

$$h(x) = \frac{\alpha}{x^{\alpha+1}}(-\alpha \ln x \ln\ln x + \ln\ln x + 1). \tag{7}$$

*From* (5) *and* (7) *we get*

$$b_{I^{[1]}}(\theta) = 4\theta \int_1^\infty v^{[1]}(x)\frac{\alpha}{x^{\alpha+1}}(-\alpha \ln x \ln\ln x + \ln\ln x + 1)dx = \frac{2}{9}\theta.$$

*From ([18], Lemma 3.3), we get that the Kullback-Leibler upper bound can be calculated as*

$$2K(\theta) = \theta^2\Big(\int_1^\infty x^2h^2(x)dx - \Big(\int_1^\infty h(x)\ln x dx\Big)^2\Big) + o(\theta^2), \tag{8}$$

*and for log-Weibull distribution we get [18]*

$$2K(\theta) = \theta^2 \psi'(1), \tag{9}$$

*where $\psi(x)$ is digamma function. Hence, the Bahadur efficiency becomes*

$$e_{I^{[1]}} = \frac{c_{I^{[1]}}(\theta)}{2K(\theta)} = \frac{125}{117\psi'(1)} \approx 0.649.$$

We proceed with calculation of Bahadur exact slope for statistic $I_n^{[2]}$. We obtain the functions $f(t)$ and $b_T(\theta)$ from the following lemmas. Their proofs are analogous to those of lemmas 3.1 and 3.2 so we omit them here.

**Lemma 3.4.** *Let $t > 0$. For statistic $I_n^{[1]}$ the function $f(t)$ is analytic for sufficiently small $t > 0$ and it holds*

$$f(t) = \frac{525}{76}t^2 + o(t^2), \quad t \to 0.$$

**Lemma 3.5.** *For a given alternative density $g(x;\theta)$ whose distribution belongs to $\mathcal{G}$ holds*

$$b_{I^{[2]}}(\theta) = 4\theta \int_1^\infty v^{[2]}(x)h(x)dx + o(\theta), \ \theta \to 0. \tag{10}$$

**Example 3.6.** *Let the alternative distribution be again the log-Weibull distribution* (6).
*From* (10) *and* (7) *we get*

$$b_{I^{[2]}}(\theta) = 4\theta \int_1^\infty v^{[2]}(x)\frac{\alpha}{x^{\alpha+1}}(-\alpha \ln x \ln\ln x + \ln\ln x + 1)dx = \frac{3}{4}(1 - \log 2)\theta,$$

*and using* (9) *we get that the Bahadur efficiency is*

$$e_{I^{[1]}} = \frac{c_{I^{[2]}}(\theta)}{2K(\theta)} = \frac{4725(1-\log 2)^2}{608\psi'(1)} \approx 0.445.$$

*3.2. Locally optimal alternatives*

Here we study the problem of locally optimal alternatives, the alternatives for which our test statistics attain the maximal efficiency. We shall determine some of those alternatives in the following theorem.

**Theorem 3.7.** *Let $g(x;\theta)$ be a density from $\mathcal{G}$ which also satisfies the condition*

$$\int_1^\infty x^2 h^2(x)dx < \infty.$$

*The alternative densities*

$$g(x;\theta) = \frac{1}{x^2} + \theta\Big(C\frac{v^{[l]}(x)}{x^2} + D\frac{\ln x - 1}{x^2}\Big), \ x \geq 1, \ C > 0, \ D \in \mathbb{R},$$

*for small $\theta$ are asymptotically optimal for the test based on $I_n^{[l]}$, $l = 1, 2$.*

**Proof.** We shall prove the theorem for the first statistic ($l = 1$). For the second statistic it is equivalent. Denote

$$h_0(x) = h(x) - \frac{(\ln x - 1)}{x^2}\int_1^\infty h(s)\ln s ds. \tag{11}$$

It can be shown that this function satisfies the following equalities:

$$\int_1^\infty h_0^2(x)x^2dx = \int_1^\infty x^2h^2(x)dx - \Big(\int_1^\infty h(x)\ln xdx\Big)^2$$

$$\int_1^\infty v^{[1]}(x)h_0(x)dx = \int_1^\infty v^{[1]}(x)h(x)dx.$$

From lemmas 3.1 and 3.2, using (1), we get that the local asymptotic efficiency is

$$\begin{aligned}
e_{I^{[1]}} &= \lim_{\theta\to 0}\frac{c_{I^{[1]}}(\theta)}{2K(\theta)} = \lim_{\theta\to 0}\frac{2\cdot\frac{1125}{108}b^2_{I^{[1]}}(\theta)}{2K(\theta)} = \lim_{\theta\to 0}\frac{b^2_{I^{[1]}}(\theta)}{9\sigma_1^2 2K(\theta)}\\
&= \lim_{\theta\to 0}\frac{9\theta^2\Big(\int_1^\infty v^{[1]}(x)h(x)dx\Big)^2 + o(\theta^2)}{9\int_1^\infty (v^{[1]})^2(x)x^2dx\Big(\theta^2(\int_1^\infty x^2h^2(x)dx - \Big(\int_1^\infty h(x)\ln xdx\Big)^2) + o(\theta^2)\Big)}\\
&= \frac{\Big(\int_1^\infty v^{[1]}(x)h_0(x)dx\Big)^2}{\int_1^\infty (v^{[1]})^2(x)x^{-2}dx\int_1^\infty h_0^2(x)x^2dx}.
\end{aligned}$$

From Cauchy-Schwarz inequality we obtain that $e_{I^{[1]}} = 1$ if and only if $h_0(x) = Cv^{[1]}(x)\alpha x^{-2}$. Substituting this equality in (11) we get the expression for $h(x)$. Since $h(x)$ for our alternatives is of such form, we complete the proof. □

## 4. Kolmogorov-Smirnov type statistics

In this section we examine the asymptotic properties of Kolmogorov-Smirnov type statistics $D_n^{[1]}$ and $D_n^{[2]}$. As previously, we can assume $\alpha = 1$. For $t \geq 1$, the expression $G_n(t) - H_n(t)$ is a $V$ statistics with kernel

$$\begin{aligned}
\Xi^{[1]}(X_1, X_2, X_3, t) &= I\{X_{(2);X_1,X_2,X_3}/X_{(1);X_1,X_2,X_3} \leq t\} - \frac{1}{3}(I\{\min\{X_1, X_2\} \leq t\} + I\{\min\{X_2, X_3\} \leq t\}\\
&+ I\{\min\{X_1, X_3\} \leq t\}).
\end{aligned}$$

The projection of this kernel is

$$\begin{aligned}
\xi^{[1]}(s,t) &= E(\Xi^{[1]}(X_1, X_2, X_3, t)|X_1 = s)\\
&= P\{X_{(2);s,X_2,X_3}/X_{(1);s,X_2,X_3} \leq t\} - \frac{2}{3}P\{\min\{s, X_2\} \leq t\} - \frac{1}{3}P\{\min\{X_2, X_3\} \leq t\}\\
&= 2P\{X_2 \leq st, s \leq X_2 \leq X_3\} + 2P\{s \leq X_2t, X_2 \leq s \leq X_3\} + 2P\{X_3 \leq tX_2, X_2 \leq X_3 \leq s\}\\
&- \frac{2}{3}(1 - I\{s > t\}P\{X_2 > t\}) - \frac{1}{3}(1 - (P\{X_2 > t\})^2).
\end{aligned}$$

After some calculations we get

$$\xi^{[1]}(s,t) = \frac{t}{s^2} - \frac{1}{s^2t^2} + \frac{1}{3t^2} - \frac{1}{3t} + I\{s < t\}(\frac{1}{3t} - \frac{t}{s^2}).$$

It is easy to show that the expected value of this projection is zero. Its variance for fixed $t$ is equal to

$$\sigma_1^2(t) = \int_1^{\infty} (\xi^{[1]}(s,t))^2 s^{-2} ds = \frac{4}{45}(t^{-3} + t^{-4} - 2t^{-6}).$$

The function $\sigma_1^2(t)$ reaches its maximum for $t_1 = 1.245$ and $\sigma_1^2(t_2) = 0.0353$. Hence, the family of kernels $\Xi^{[1]}(X_1, X_2, X_3, t)$ is non-degenerate according to the argumentation [14]. The $V$-empirical process

$$\rho_n^{[1]}(t) = \sqrt{n}(G_n(t) - H_n(t)), \ t \geq 1$$

converges in distribution to some Gaussian process following the argumentation of [21]. The covariance of this process is calculable but complicated, while the distribution of statistics $D_n^{[1]}$ is unknown.

Similarly, for fixed $t \geq 1$, the expression $J_n(t) - K_n(t)$ is a $V$-statistic with the kernel

$$\Xi^{[2]}(X_1, X_2, X_3, t) = I\{X_{(3);X_1,X_2,X_3}/X_{(2);X_1,X_2,X_3} \leq t\} - I\{(X_{(2);X_1,X_2,X_3}/X_{(1);X_1,X_2,X_3})^2 \leq t\}.$$

Performing analogous calculations as in the previous case we derive that the projection of this kernel is

$$\xi^{[2]}(s,t) = \frac{1}{s^2t}(2 - 2s + s^2t^{\frac{1}{2}} - s^2t - t^{\frac{3}{2}} + 2st^2 - t^3 + I\{s < t^{\frac{1}{2}}\}t^{\frac{1}{2}}(t - s^2) + I\{s < t\}t(t-s)^2).$$

Its variance is

$$\sigma_1^2(t) = \frac{1}{15}(3t^{-1} - 2t^{-\frac{3}{2}} + 2t^{-2} - 14t^{-\frac{5}{2}} + 13t^{-3} - 2t^{-4})$$

which has maximum equal to 0.0265 for $t_2 = 3.160$. This kernel is also non-degenerate, and all the arguments and conclusions about the asymptotic distribution of the statistic $D_n^{[2]}$ are equivalent to those of $D_n^{[1]}$.

### *4.1. Bahadur efficiency*

We now proceed with calculations of Bahadur efficiency for statistics $D_n^{[1]}$ and $D_n^{[2]}$. The functions $f_1$ and $f_2$ from (2) can be determined from the following theorem.

**Theorem 4.1.** *Let $a \geq 0$. Then $f_1(a)$ and $f_2(a)$ are analytic for sufficiently small $a$ and it holds*

$$\begin{aligned} f_1(a) &= \frac{a^2}{18}\sigma_1^2(t_1) + o(a^2) \approx 1.58a^2, \ a \to 0 \\ f_2(a) &= \frac{a^2}{18}\sigma_2^2(t_2) + o(a^2) \approx 2.10a^2, \ a \to 0 \end{aligned}$$

The proof of this theorem can be found in [14]. The limit in probability of statistics $D_n^{[l]}$, $l = 1, 2$ is determined from the following lemma.

**Lemma 4.2.** *For a given alternative density $g(x;\theta)$ from $\mathcal{G}$ holds*

$$b_{D^{[l]}}(\theta) = 3\theta \sup_{t \geq 1} \left| \int_1^{\infty} \xi^{[l]}(x,t)h(x)dx \right| + o(\theta), \ \theta \to 0. \tag{12}$$

**Proof.** We shall prove the theorem for the first statistics ($l = 1$). The other case is equivalent. Using Glivenko-Cantelli theorem for $U$ and $V$ empirical distribution functions (see [9]) we get

$$b_{D^{[1]}}(\theta) = \sup_{t\geq 1}\left|P_\theta\{X_{(2;3)}/X_{(1;3)} \leq t\} - P_\theta\{\min\{X_1, X_2\} \leq t\}\right|.$$

Denote $a(\theta) = P_\theta\{X_{(2;3)}/X_{(1;3)} \leq t\} - P_\theta\{\min\{X_1, X_2\} \leq t\}$. Then we have

$$\begin{aligned} a(\theta) &= \int_1^\infty\int_1^{tx} 6(1 - G(x;\theta))g(x;\theta)g(y;\theta)dxdy - (1 - (P_\theta\{X_1 > t\})^2) \\ &= 3\int_1^\infty (1 - G(tx;\theta))^2 g(x;\theta)dx - (1 - (1 - G(t;\theta))^2). \end{aligned}$$

The first derivative of $a(\theta)$ at $\theta = 0$ is

$$\begin{aligned} a'(\theta) &= 6\int_1^\infty \frac{H(tx)}{tx^3}dx - 3\int_1^\infty \frac{h(x)}{t^2x^2}dx + \frac{2}{t}H(t) \\ &= \frac{6}{t}\int_1^t\int_1^\infty \frac{h(y)}{x^3}dxdy + \frac{6}{t}\int_t^\infty\int_{\frac{y}{t}}^\infty \frac{h(y)}{x^3}dxdy - 3\int_1^\infty \frac{h(x)}{t^2x^2}dx + \frac{2}{t}\int_1^t h(y)dy \\ &= \int_1^\infty h(y)(-\frac{3}{t}I\{y < t\} + \frac{3t}{y^2}I\{y > t\} - \frac{3}{t^2y^2} + \frac{2}{t}I\{y < t\})dy \\ &= 3\int_1^\infty \xi^{[1]}(y;t)h(y)dy. \end{aligned}$$

Applying the Maclaurin's expansion on the function $a(\theta)$ we get the statement of the theorem. □

As in the previous section in the following example we show the calculation of Bahadur efficiency for log-Weibull alternative.

**Example 4.3.** *For statistic* $D_n^{[1]}$ *from* (12) *and* (7) *we get*

$$b_{D^{[1]}}(\theta) = 3\theta\sup_{t\geq 1}\left|\int_1^\infty \xi^{[1]}\frac{1}{x^2}(-\ln x\ln\ln x + \ln\ln x + 1)dx\right| = 0.4088\theta$$

*and using* (9) *we get that the Bahadur efficiency is*

$$e_{D^{[1]}} = \frac{c_{D^{[1]}}(\theta)}{2K(\theta)} \approx 0.320.$$

*Similarly, for statistic* $D_n^{[2]}$ *we have*

$$b_{D^{[2]}}(\theta) = 3\theta \sup_{t\geq 1}\Big|\int_1^{\infty} \xi^{[2]}\frac{1}{x^2}(-\ln x \ln\ln x + \ln\ln x + 1)dx\Big| = 0.3309\theta,$$

*and*

$$e_{D^{[2]}} = \frac{c_{D^{[2]}}(\theta)}{2K(\theta)} \approx 0.280.$$

*4.2. Locally optimal alternatives*

As in the previous section, we shall determine some of locally optimal alternatives in the following theorem.

**Theorem 4.4.** *Let $g(x;\theta)$ be a density from $\mathcal{G}$ which also satisfies the condition*

$$\int_1^{\infty} x^2h^2(x)dx < \infty.$$

*The alternative densities*

$$g(x;\theta) = \frac{1}{x^2} + \theta\Big(C\frac{\xi^{[l]}(x;t_l)}{x^2} + D\frac{\ln x - 1}{x^2}\Big),\ x \geq 1,\ C > 0,\ D \in \mathbb{R},$$

*where $t_1 = 1.245$, $t_2 = 3.16$, for small $\theta$ are asymptotically optimal for the test based on $D_n^{[l]}$, $l = 1, 2$.*

The proof is analogous to that of theorem 3.7, so we omit it here.

## 5. Comparison of asymptotic efficiencies

In this section we compare the Bahadur efficiencies of our tests against each other and the statistics $T_n$ and $V_n$ from [18]. The common alternatives we consider for the comparison are:

- log-Weibull distribution with density

$$g_1(x;\theta) = (1+\theta)x^{-1}(\ln x)^{\theta}e^{-(\ln x)^{1+\theta}},\ \ x \geq 1, \theta \in (0,1),$$

- log-gamma distribution with density

$$g_2(x;\beta,\theta) = \frac{(\ln x)^{\theta}}{x^2\Gamma(1+\theta)},\ \ x \geq 1, \theta \in (0,1),$$

- distribution considered in [18] with density

$$g_3(x;\theta) = \frac{1}{x^2}(e^{-\theta(\ln x)^{\beta}} + \theta\beta(\ln x)^{\beta-1}e^{-\theta(\ln x)^{\beta}}),\ \ x \geq 1,\ \ \beta > 1,\ \ \theta \in (0,1)$$

  for $\beta = 1.5$ and $\beta = 2$,

- inverse-beta distribution with density

$$g_4(x;\theta) = \frac{1+\theta}{x^2}\Big(1 - \frac{1}{x}\Big)^{\theta},\ \ x \geq 1, \theta \in (0,1),$$

- Pareto distribution with so-called 'tilt' parameter (see [10]) with density

$$g_5(x;\theta) = \frac{1+\theta}{(x+\theta)^2}, \quad x \geq 1, \theta \in (0,1),$$

- mixture of two Pareto distributions with density

$$g_6(x;\beta,\theta) = \frac{1-\theta}{x^2} + \frac{\beta\theta}{x^{\beta+1}}, \quad x \geq 1, \quad \beta > 1, \quad \theta \in (0,1)$$

for $\beta = 1.5$, $\beta = 4$ and $\beta = 8$.

The Bahadur efficiencies are presented in table 1 for integral type statistics and in table 2 for Kolmogorov-Smirnov type.

As a general rule, we can see that, as usual, the efficiencies for integral type statistics are higher than those of Kolmogorov-Smirnov type ones. However, the integral type tests have a drawback in the sense that are not consistent against all alternatives and cannot be used in such circumstances. The example here is the mixture alternative $g_6$ against which neither of our integral type statistics is consistent since the values of $b(\theta)$ (limit in probability under alternative $g_6$) turn out to be negative and only (large) positive values of tests statistics are considered significant.

We can see that each test has at least one alternative against which it can be considered reasonably efficient. Comparing our two proposed integral type tests we can see that based on our results we cannot decide that one is better than the other. Log-Weibull, log-gamma and inverse-beta alternatives favour test based on $I_n^{[1]}$, while the others favour $I_n^{[2]}$. It is interesting to note that although the tests are based on similar order statistics ratio characterizations, the efficiencies of the corresponding tests are quite different.

The test based on $T_n$ seems to be the best overall, however all our tests have at least one common alternative for which they are most efficient of the three.

For Kolmogorov-Smirnov type tests we can notice the similar pattern, but the ordering of the efficiencies of the tests is not identical for every alternative. We can see that the test based on $D_n^{[2]}$ is now the most efficient for 'tilt' and $g_3$ for both value of the parameters. The alternative $g_6$ shows us an interesting phenomenon. Depending on the value of parameter $\beta$ the ordering of the tests changes and each test is the most efficient at some point. This shows that when we suspect a contamination to be present in the sample, we should choose the appropriate test based on the suspected ratio of the shape parameters of mixture components. If it is high, we should choose $D_n^{[1]}$, if low, $D_n^{[2]}$ and if moderate, $V_n$.

| | Alternative density | | | | | |
|---|---|---|---|---|---|---|
| Statistic | $g_1$ | $g_2$ | $g_3(1.5)$ | $g_3(2)$ | $g_4$ | $g_5$ |
| $I_n^{[1]}$ | 0.649 | 0.757 | 0.326 | 0.119 | 0.778 | 0.451 |
| $I_n^{[2]}$ | 0.445 | 0.277 | 0.610 | 0.486 | 0.244 | 0.737 |
| $T_n$ | 0.821 | 0.788 | 0.618 | 0.338 | 0.777 | 0.800 |

Table 1: Bahadur efficiencies for integral type statistics

| | Alternative density | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Statistic | $g_1$ | $g_2$ | $g_3(1.5)$ | $g_3(2)$ | $g_4$ | $g_5$ | $g_6(1.5)$ | $g_6(4)$ | $g_6(8)$ |
| $D_n^{[1]}$ | 0.320 | 0.414 | 0.141 | 0.047 | 0.436 | 0.207 | 0.124 | 0.462 | 0.632 |
| $D_n^{[2]}$ | 0.280 | 0.174 | 0.410 | 0.362 | 0.156 | 0.513 | 0.500 | 0.323 | 0.116 |
| $V_n$ | 0.437 | 0.448 | 0.331 | 0.192 | 0.455 | 0.473 | 0.359 | 0.631 | 0.581 |

Table 2: Bahadur efficiencies for Kolmogorov-Smirnov type statistics

## 6. Critical values and application to real data

In this section we present the critical values of our tests for small samples and give an example of real data to see their performance.

The critical values at level of significance 0.05 are given in the table 3. They have been obtained using Monte Carlo simulation with 10000 replicates.

| $n$ | $I_n^{[1]}$ | $I_n^{[2]}$ | $D_n^{[1]}$ | $D_n^{[2]}$ |
|---|---|---|---|---|
| 10 | 0.11 | 0.24 | 0.57 | 0.47 |
| 20 | 0.08 | 0.14 | 0.37 | 0.29 |
| 30 | 0.06 | 0.11 | 0.29 | 0.22 |
| 40 | 0.05 | 0.09 | 0.25 | 0.19 |
| 50 | 0.05 | 0.08 | 0.22 | 0.17 |
| 100 | 0.03 | 0.05 | 0.15 | 0.11 |

Table 3: Critical values of the tests at level of significance 0.05

In table 4 we present the data of Norwegian fire claims from 1975 (see [5], Appendix I). This is a well-known example of Pareto distributed data. It has often been used for demonstrating the quality of tests for Pareto distribution. The scale parameter is considered known and equal to 500 (see e.g. [19]).

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 500 | 550 | 586 | 620 | 680 | 798 | 927 | 1038 | 1291 | 1515 | 2497 | 4585 |
| 500 | 550 | 593 | 622 | 700 | 800 | 940 | 1041 | 1293 | 1519 | 2690 | 4810 |
| 500 | 551 | 596 | 632 | 725 | 800 | 940 | 1104 | 1298 | 1587 | 2760 | 6855 |
| 502 | 552 | 596 | 635 | 728 | 800 | 948 | 1108 | 1300 | 1700 | 2794 | 7371 |
| 515 | 557 | 600 | 635 | 736 | 826 | 957 | 1137 | 1305 | 1708 | 2886 | 7772 |
| 515 | 558 | 600 | 640 | 737 | 835 | 1000 | 1143 | 1327 | 1820 | 2924 | 7834 |
| 528 | 570 | 600 | 650 | 740 | 862 | 1002 | 1180 | 1387 | 1822 | 2953 | 13000 |
| 530 | 572 | 605 | 650 | 748 | 885 | 1009 | 1243 | 1455 | 1848 | 3289 | 13484 |
| 530 | 574 | 610 | 650 | 752 | 900 | 1013 | 1248 | 1475 | 1906 | 3860 | 17237 |
| 530 | 579 | 610 | 650 | 756 | 900 | 1020 | 1252 | 1479 | 2110 | 4016 | 52600 |
| 540 | 583 | 613 | 672 | 756 | 910 | 1024 | 1280 | 1485 | 2251 | 4300 | |
| 544 | 584 | 615 | 674 | 777 | 912 | 1033 | 1285 | 1491 | 2362 | 4397 | |

Table 4: Norwegian Fire Claims (1975) (1000 Norwegian Krones)

We test the data at level of significance 0.05. The p-values of our tests are given below.

| test | $I_n^{[1]}$ | $I_n^{[2]}$ | $D_n^{[1]}$ | $D_n^{[2]}$ |
|---|---|---|---|---|
| p-value | 0.58 | 0.59 | 0.66 | 0.96 |

As we can see, all our tests do not reject the null Pareto hypothesis.

## 7. Conclusion

In this paper we gave a characterization of Pareto distribution. We also proposed four new goodness of fit tests, two of them based on this characterization and two more based on Rossberg's characterization. All tests are free of shape parameter which enables us to test a composite null hypothesis without estimating it.

The Bahadur efficiencies for some common alternatives have been calculated and quite a few of them are reasonably high. Also, for each test we found a class of locally optimal alternatives against which the test is asymptotically optimal.

We compared our tests with two similar tests from [18] and we concluded that no test dominates the others. On the contrary, for each proposed there are common alternatives for which the test is better than its competitors, not even taking into account locally optimal alternatives.

The conclusion is that all proposed tests can be useful in practice. There is no simple rule in deciding which test to choose. If we do not suspect a particular alternative distribution, but only want to be sure of our null hypothesis, then we suggest to try all proposed tests. Otherwise we should choose the test that gives the best performance against suspected alternative distribution.

## References


[1] J. E. Angus, Goodness-of-fit Test for Exponentiality Based on Loss of Memory Type Functional Equation, Journal of Statistical Planning and Inference, 6 (1982) 241–251

[2] M. Ahsanullah, On a characterization of the exponential distribution by spacings, Annals of the Institute of Statistical Mathematics 30.1 (1978) 163–166.

[3] R. R. Bahadur, Some limit theorems in statistics, SIAM, Philadelphia, 1971.

[4] L. Baringhaus, N. Henze, Test of fit for exponetiality based on a characterization via the mean residual life function, Statistical Papers 41 (2001) 225–236.

[5] J. Beirlant, J.L. Teugles, P. Vynckier, Practical Analysis of Extreme Values, Leuven University Press, Leuven, Belgium, 1996.

[6] J. Galambos, S. Kotz, Characterizations of Probability Distributions, Springer-VErlag, Berlin-Heidelberg-New York, 1978.

[7] N. Hitha, Some Characterizations of Pareto and Related Populations, Ph.D. thesis, Cochin university of science and technology, India, 1991.

[8] W. Hoeffding, A class of statistics with asymptotically normal distribution, Annals of Mathetmatical Statistics 19 (1948) 293-325.

[9] P. L. Janssen, Generalized empirical distribution functions with statistical applications, Limburgs Universitair Centrum, Diepenbeek, 1988.

[10] A. W. Marshall, I. Olkin, Life distributions: Structure of nonparametric, semiparametric, and parametric families. Springer, New York, 2007.

[11] K. Morris, D. Szynal, Goodness of Fit Tests Based on Characterizations of Continuous Distributions, Applicationes Mathematicae 27.4 (2000) 475–488.

[12] Ya. Yu. Nikitin, Asymptotic efficiency of nonparametric tests, Cambridge University Press, 1995.

[13] Ya. Yu. Nikitin, K. Yu. Volkova, Asymptotic efficiency of exponentiality tests based on order statistics characterization, Georgian Mathematical Journal 17.4 (2010) 749–763.

[14] Ya. Yu. Nikitin, Large deviation of U-empirical Kolmogorov-Smirnov test and their efficiency, Journal of Nonparametric Statistics 22 (2010) 649–668.

[15] Ya. Yu. Nikitin, E. V. Ponikarov, Rough large deviation asymptotics of Chernoff type for von Mises functionals and U-statistics, Proceedings of Saint-Petersburg Mathematical Society 7 (1999) 124–167.

[16] Ya. Yu. Nikitin, Bahadur Efficiency of Test of Exponentiality Based on Loss of Memory Type Functional Equation, Journal of Nonparametric Statistics 6 (1996) 13–26.

[17] H. A. Noughabi, N. R. Arghami, Testing Exponentiality Based on Characterizations of the Exponential Distribution, Journal of Statistical Computation and Simulation, 81.11, (2011) 1641–1651.

[18] M. Obradović, M. Jovanović, B. Milošević, Goodness-of-fit tests for Pareto distribution based on a characterization and their asymptotics, Statistics (ahead-of-print) (2014) 1–16.

[19] M.L. Rizzo, New Goodness-of-fit Tests for Pareto Distribution, Astin Bulletin 39.2 (2009) 691–715.

[20] H.J. Rossberg, Characterization of the exponential and the Pareto distributions by means of some properties of the distributions which the differences and quotients of order statistics are subject to, Statistics, 3.3, (1972) 207–216.

[21] B. W. Silverman, Convergence of a class of empirical distribution function of dependent random variables, Annals of Probability 11 (1983) 745–751.

[22] R. J. Serfling, Approximation Theorems of Mathematical Statistics, John Wiley & Sons, New York, 2002.

[23] K. Yu. Volkova, On asymptotic efficiency of exponentiality tests based on Rossberg's characterization, Probability and Statistics, part 15, Zapiski nauchnikh seminarov POMI 368, POMI St. Peterburg (2009) 95–109.

[24] K. Yu. Volkova, Ya Yu Nikitin, Goodness-of-Fit Tests for the Power Function Distribution Based on the Puri-Rubin Characterization and Their Efficiences, Journal of Mathematical Sciences (2012) 1–9.