Research Article

Characterization and Goodness-of-Fit Test of Pareto and Some Related Distributions Based on Near-Order Statistics

Masoumeh Akbari

Department of Statistics, University of Mazandaran, Babolsar, Mazandaran, Iran

Correspondence should be addressed to Masoumeh Akbari; m.akbari@umz.ac.ir

Received 6 September 2019; Revised 23 May 2020; Accepted 23 June 2020; Published 15 July 2020

Academic Editor: Chin-Shang Li

Copyright © 2020 Masoumeh Akbari. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In this paper, a new definition of the number of observations near the k th order statistics is developed. Then some characterization results for Pareto and some related distributions are established in terms of mass probability function, first moment of these new counting random variables, and using completeness properties of the sequence of functions $\{x^n, 0 < x < 1, n \geq 1\}$. Finally, new goodness-of-fit tests based on these new characterizations for Pareto distribution are presented. And the power values of the proposed tests are compared with the power values of well-known tests such as Kolmogorov-Smirnov and Cramer-von Mises tests by Monte Carlo simulations.

1. Introduction

Let order statistics $X_{1:n} \leq X_{2:n} \leq \dots \leq X_{n:n}$ be the nondecreasing order of independent and identically distributed (iid) random variables $X_1, X_2, \dots, X_n$ with cumulative distribution function (cdf) $F_X(\cdot)$. The order statistics have important roles in different areas of statistics and probability. In reality, some kinds of order statistics are more applicable. For example, in actuarial science, the distribution of the minimum of the two lifespans of the couple is important for insurance policy to make decisions. In industries, specifically in reliability and survival analysis, order statistics are used to solve problems. Meteorology, hydrology, and so on are other fields of applications of order statistics. Interested readers can study the detail of theory and application of order statistics, for instance, in Arnold et al. [1]. Let $F_X(\cdot)$ be discrete cdf. For the first time, Eisenberg et al. [2] defined the quantity $K_n = \sum_{i=1}^n I(X_i = X_{n:n})$ as the number of winners in a golf competition. They studied sufficient conditions under which K_n converges to 1. After that Pakes and Steutel [3] considered similar notions for continuous cdf as follows:

$$K_n(a) = \sum_{j=1}^n I_{(X_{n:n}-a, X_{n:n})}(X_j), \quad (1)$$

where $a > 0$ is a constant. In fact, $K_n(a)$ counts the number of observations in the left-hand neighbourhood of the sample maximum with fixed distance “ a .” Later, $K_n(a)$ was developed to the number of observations near the k th-order statistics as follows:

$$K_-(n, k, a) = \#\{j \in \{1, 2, \dots, n\}; X_j \in (X_{k:n} - a, X_{k:n})\}, \quad (2)$$

where the support of $K_-(n, k, a)$ is $0, 1, \dots, k-1$. Similarly, the number of observations in the right-hand neighbourhood of the k th order statistics was defined as

$$K_+(n, k, a) = \#\{j \in \{1, 2, \dots, n\}; X_j \in (X_{k:n}, X_{k:n} + a)\}. \quad (3)$$

The support of $K_+(n, k, a)$ is $0, 1, \dots, n-k$. The probability mass function (pmf) of $K_-(n, k, a)$ is as follows:

$$P(K_-(n, k, a) = j) = \binom{k-1}{j} \int_{S_X} (\eta_L(x, a))^{k-j-1} (1 - \eta_L(x, a))^j dF_{k:n}(x), \quad (4)$$

where $\eta_L(x, a) = (F_X(x-a)/F_X(x))$ and $F_{k:n}(\cdot)$ is the cdf of $X_{k:n}$. Also, the pmf of $K_+(n, k, a)$ is given by

$$P(K_+(n, k, a) = j) = \binom{n-k}{j} \int_{S_X} (\eta_U(x, a))^{n-k-j} (1 - \eta_U(x, a))^j dF_{k:n}(x), \quad (5)$$

where $\eta_U(x, a) = (\bar{F}_X(x+a)/\bar{F}_X(x))$. For more details, one can refer to Dembińska et al. [4].

More results of the description of their distributions, asymptotic properties, and their generalization have been investigated e.g., Pakes and Li [5], Li [6], Pakes ([7,8]), Balakrishnan and Stepanov ([9,10]), Dembińska et al. [4], and Dembińska ([11–13]). So far, few researchers have addressed the issue of statistical inference based on near-order statistics, e.g., Müller [14], Hashorva and Hüsler [15], Akbari et al. [16], and Akbari and Akbari [17]. In the present paper, a new version of near-order statistics is first defined. Then some characterization results as a statistical tool in goodness-of-fit (GOF) test for some continuous distributions are obtained. The results of this paper are organized as follows. Section 2 contains preliminary results. The characterization results of the paper are included in Section 3. Finally, in Section 4 are introduced two tests for goodness-of-fit tests for Pareto distribution. The critical values of the proposed test statistics are computed by Monte Carlo simulations. Also, their power is compared with those computed by well-known tests such as Kolmogorov–Smirnov and Cramer–von Mises tests by simulations. All simulations are carried out by using R 3.6.3 and with 10000 replications.

2. Preliminary Results

Let $X_1, \dots, X_n$ be iid random variables from continuous cdf $F_X(\cdot)$ with support S_X , and F_X be one of the distribution functions Pareto, uniform, or power function. Constructing characterizations for such F_X by the pmfs or moments of some functions of random variables is a common way. But it is not possible by counting random variables $K_{\pm}$ as the number of observations on the location-type neighbourhood

of the certain order statistics defined in equations (4) and (5), because their pmfs do not have a closed-form expression. Therefore, new types of number of observations falling within the left-hand and right-hand of neighbourhood of the specific order statistics, as an extension to scale-type neighbourhood, are introduced, respectively, as follows:

$$K'_-(n, k, a) = \#\{j \in \{1, 2, \dots, n\}; X_j \in (aX_{k:n}, X_{k:n})\}, \quad (6)$$

where $0 < a < 1$ and

$$K'_+(n, k, b) = \#\{j \in \{1, 2, \dots, n\}; X_j \in (X_{k:n}, bX_{k:n})\}, \quad (7)$$

where $b > 1$.

Proposition 1. *Using the same arguments given in Dembińska et al. [4], it is concluded that the pmfs of new counting random variables $K'_-(n, k, a)$ and $K'_+(n, k, b)$ are the same as the pmfs of $K_-(n, k, a)$ and $K_+(n, k, b)$, respectively, with $\eta_L(x, a) = (F_X(ax)/F_X(x))$ and $\eta_U(x, a) = (\bar{F}_X(bx)/\bar{F}_X(x))$, that is,*

$$P(K'_-(n, k, a) = j) = \binom{k-1}{j} \int_{S_X} \left(\frac{F_X(ax)}{F_X(x)} \right)^{k-j-1} \left(1 - \frac{F_X(ax)}{F_X(x)} \right)^j dF_{k:n}(x), \quad (8)$$

$$P(K'_+(n, k, b) = j) = \binom{n-k}{j} \int_{S_X} \left(\frac{\bar{F}_X(bx)}{\bar{F}_X(x)} \right)^{n-k-j} \left(1 - \frac{\bar{F}_X(bx)}{\bar{F}_X(x)} \right)^j dF_{k:n}(x). \quad (9)$$

On the other hand, by simple algebra calculations, the first moment of $K'_-(n, k, a)$ and $K'_+(n, k, b)$ can be derived, respectively, as follows:

$$E(K'_-(n, k, a)) = (k-1) \left(1 - \frac{n!}{(k-1)!(n-k)!} \int_{S_X} F_X(ax) F_X^{k-2}(x) (1 - F_X(x))^{n-k} dF_X(x) \right) \quad (10)$$

$$= k-1 - n(k-1) \binom{n-1}{k-1} \int_0^1 F_X(aF_X^{-1}(u)) u^{k-2} (1-u)^{n-k} du,$$

$$E(K'_+(n, k, b)) = (n-k) - k(n-k) \binom{n}{k} \int_{S_X} \bar{F}_X(bx) F_X^{k-1}(x) \bar{F}_X^{n-k-1}(x) dF_X(x) \quad (11)$$

$$= (n-k) - k(n-k) \binom{n}{k} \int_0^1 \bar{F}_X(bF_X^{-1}(u)) u^{k-1} (1-u)^{n-k-1} du.$$

Here, some results as examples of special cases of Pareto and power function distributions are reported that will be useful for obtaining further results in the next section.

Example 1. Let $X_1, X_2, \dots, X_n$ be iid random variables from Pareto (α, β) . So their survival distribution functions is given by

$$\bar{F}(x) = \left(\frac{\beta}{x}\right)^\alpha, \quad x \geq \beta. \quad (12)$$

Then from equation (9), the pmf of $K'_+(n, k, b)$ related to this sequence is as follows:

$$\begin{aligned} P(K'_+(n, k, b) = j) \\ &= \binom{n-k}{j} \int_{\beta}^{\infty} \left(\frac{1}{b}\right)^{\alpha(n-k-j)} \left(1 - \left(\frac{1}{b}\right)^\alpha\right)^j dF_{k:n}(x) \\ &= \binom{n-k}{j} \left(\frac{1}{b}\right)^{\alpha(n-k-j)} \left(1 - \left(\frac{1}{b}\right)^\alpha\right)^j. \end{aligned} \quad (13)$$

Equation (13) shows that $K'_+(n, k, b)$ has binomial pmf with parameters $(n-k)$ and $1 - (1/b)^\alpha$. Thus,

$$E(K'_+(n, k, b)) = (n-k) \left(1 - \left(\frac{1}{b}\right)^\alpha\right), \quad (14)$$

which does not depend on scale parameter β .

Example 2. It is well-known that if $X \sim \text{Pareto}(\alpha, \beta)$, then random variable $Y = (1/X)$ has power function distribution with following cdf:

$$F_Y(y) = (\beta y)^\alpha, \quad 0 \leq y \leq \frac{1}{\beta}. \quad (15)$$

The notation power (α, β) is used for power function distribution with parameters α and β . It is also called generalized uniform distribution because it is standard uniform cdf at $\beta = 1$ and $\alpha = 1$. From (8), the pmf of $K'_-(n, k, a)$ for random variables $X_1, \dots, X_n$ that are distributed as power (α, β) is concluded as

$$\begin{aligned} P(K'_-(n, k, a) = j) \\ &= \binom{k-1}{j} \int_0^{1/\beta} a^{\alpha(k-j-1)} (1-a^\alpha)^j dF_{k:n}(x) \\ &= \binom{k-1}{j} a^{\alpha(k-j-1)} (1-a^\alpha)^j. \end{aligned} \quad (16)$$

According to (16), $K'_-(n, k, a)$ has binomial pmf with parameters $k-1$ and success probability $1-a^\alpha$. So

$$E(K'_-(n, k, a)) = (k-1)(1-a^\alpha). \quad (17)$$

As we know, the uniform distribution function on interval $(0, 1)$ is a special case of power $(1, 1)$ with following cdf:

$$F(x) = x, \quad 0 < x < 1. \quad (18)$$

Therefore from (16), the pmf of $K'_-(n, k, a)$ when $X_i \sim U(0, 1), i = 1, \dots, n$, equals

$$\begin{aligned} P(K'_-(n, k, a) = j) &= \binom{k-1}{j} \int_0^1 a^{(k-j-1)} (1-a)^j dF_{k:n}(x) \\ &= \binom{k-1}{j} a^{(k-j-1)} (1-a)^j, \end{aligned} \quad (19)$$

and its expectation is

$$E(K'_-(n, k, a)) = (k-1)(1-a). \quad (20)$$

3. Characterizaion Results

In this section, some characterization results based on distributional properties of near-order statistics $K'_-(n, k, a)$ and $K'_+(n, k, b)$ for some continuous distributions are established in terms of property of sequence of complete functions. Thus, in the sequel, some notions and theorems related to this theory are reminded.

Definition 1. A sequence $\{\phi_n\}_{n \geq 1}$ of elements of a Hilbert space $\mathcal{H}$ is called complete if the only element which is orthogonal to every ϕ_n is the null element, that is

$$\langle f, \phi_n \rangle = 0, \quad \text{for every } n \geq 1, \quad (21)$$

implies f null.

The notation $\langle \cdot, \cdot \rangle$ denotes the inner product of $\mathcal{H}$. In the present paper, the Hilbert space $L^2[0, 1]$ with the following inner product being considered:

$$\langle f, g \rangle = \int_0^1 f(x)g(x)dx, \quad (22)$$

where f and g are real-valued square integrable functions on $[0, 1]$. One of the sequences of complete functions in $L^2[0, 1]$ is $\{x^n, n \geq 1\}$ which is used in this paper.

The following theorem is known as Müntz theorem that states the necessary and sufficient condition for completeness of the subsequence $\{x^{n_j}, j \geq 1, n_1 < n_2 < \dots\}$.

Theorem 1 (Higgins [18], page 95). *Sequence $\{x^{n_1}, x^{n_2}, \dots, 1 \leq n_1 < n_2 < \dots\}$ in $L^2[0, 1]$ is complete if and only if*

$$\sum_{j=1}^{\infty} n_j^{-1} = \infty. \quad (23)$$

For more details about Hilbert space and complete sequences, refer to Higgins [18]. Pareto is one of the distributions that have many applications in economics and actuturial sciences. So far, a lot of properties and characterization of it based on order statistics or their functions have been obtained, for example, Lee and Chang [19], Afify [20], Ahsanullah and Shakil [21], Ahsanullah et al. [22], and Nofal and El Gebaly [23]. In the following theorem, some characterizations for Pareto law in terms of $K'_+(n, k, b)$ are established.

Theorem 2. *Let $X_1, X_2, \dots, X_n$ be iid random variables with continuous cdf $F_X(\cdot)$ whose support is $[\beta, \infty)$. $F_X(\cdot)$ is Pareto (α, β) cdf, if and only if one of the following conditions holds:*

- (a) *There exists $j_0 \in \{0, 1, \dots\}$ such that for all $n \geq j_0 + 1, b > 1$ and for $k = n - j_0$, we have*

$$P(K'_+(n, k, b) = j_0) = \left(1 - \left(\frac{1}{b}\right)^\alpha\right)^{j_0}. \quad (24)$$

(b) For all $n \geq k, b > 1$, and a fixed $k \geq 1$, we have

$$E(K'_+(n, k, b)) = (n - k) \left(1 - \left(\frac{1}{b} \right)^\alpha \right). \quad (25)$$

Proof. If X -sequence has Pareto (α, β) cdf, by the use of equations (13) and (14), one can easily obtain parts (a) and (b). Let part (a) hold. Using pmf of $K'_+(n, k, b)$ and the assumptions of (a), the equality in (a) can be rewritten as

$$\int_{\beta}^{\infty} \left\{ 1 - \frac{\bar{F}_X(bx)}{\bar{F}_X(x)} \right\}^{j_0} dF_{k:n}(x) dx = \left(1 - \left(\frac{1}{b} \right)^\alpha \right)^{j_0}. \quad (26)$$

The right-hand side of equation (26) can be expressed as

$$\left(1 - \left(\frac{1}{b} \right)^\alpha \right)^{j_0} \int_0^1 \frac{n!}{(k-1)!(n-k)!} (1-u)^{k-1} u^{n-k} du. \quad (27)$$

On the other hand, replacing $dF_{k:n}(x)$ with $(n!/(k-1)!(n-k)!)F_X^{k-1}(x)\bar{F}_X^{n-k}(x)dF_X(x)$, the equality in (26) can be stated as

$$\begin{aligned} \int_{\beta}^{\infty} \left\{ 1 - \frac{\bar{F}_X(bx)}{\bar{F}_X(x)} \right\}^{j_0} F_X^{k-1}(x) \bar{F}_X^{n-k}(x) dF_X(x) \\ = \left(1 - \left(\frac{1}{b} \right)^\alpha \right)^{j_0} \int_0^1 (1-u)^{k-1} u^{n-k} du. \end{aligned} \quad (28)$$

Taking the change of variable $u = 1 - F_X(x)$ in the left-hand side of equation (28), it is deduced:

$$\begin{aligned} \int_0^1 \left\{ 1 - \frac{\bar{F}_X(bF_X^{-1}(1-u))}{u} \right\}^{j_0} (1-u)^{k-1} u^{n-k} du \\ = \left(1 - \left(\frac{1}{b} \right)^\alpha \right)^{j_0} \int_0^1 (1-u)^{k-1} u^{n-k} du. \end{aligned} \quad (29)$$

By the assumption " $k = n - j_0$ " and after some algebra simplifications in the aforementioned equality, it is concluded that

$$\int_0^1 \left\{ [u - \bar{F}_X(bF_X^{-1}(1-u))]^{j_0} - \left[u \left(1 - \left(\frac{1}{b} \right)^\alpha \right) \right]^{j_0} \right\} (1-u)^{n-j_0-1} du = 0. \quad (30)$$

Since $|u - \bar{F}_X(bF_X^{-1}(1-u))| \leq 2$ and $u(1 - (1/b)^\alpha) \leq 2$, by Minkowski's inequality, we have

$$\left\{ [u - \bar{F}_X(bF_X^{-1}(1-u))]^{j_0} - \left[u \left(1 - \left(\frac{1}{b} \right)^\alpha \right) \right]^{j_0} \right\} \in L^2[0, 1]. \quad (31)$$

If (30) holds for all $n \geq j_0 + 1$, by the completeness property of the sequence $\{(1-u)^{n-j_0-1}, n \geq j_0 + 1\}$, the following identity can be derived

$$[u - \bar{F}_X(bF_X^{-1}(1-u))]^{j_0} - \left[u \left(1 - \left(\frac{1}{b} \right)^\alpha \right) \right]^{j_0} = 0. \quad (32)$$

Hence,

$$\frac{\bar{F}_X(bF_X^{-1}(1-u))}{u} = \left(\frac{1}{b} \right)^\alpha. \quad (33)$$

By taking $t = F_X^{-1}(1-u)$ in (33), it can be rewritten as

$$\frac{\bar{F}_X(bt)}{\bar{F}_X(t)} = \left(\frac{1}{b} \right)^\alpha. \quad (34)$$

If (34) holds for all $b > 1$ and $t \geq \beta$, by the use of the method of solution given in Aczél [24], it is concluded that function $\bar{F}_X(t) = ct^{-\alpha}$ is the general solution of (34). Because $t \geq \beta$ and $\bar{F}_X(\cdot)$ is a survival distribution function, the constant c will be β^α . So, the proof is completed.

Suppose that part (b) holds. Then from equation (11), it is deduced that

$$\frac{n!}{(k-1)!(n-k)!} \int_0^1 \bar{F}_X(bF_X^{-1}(u)) u^{k-1} (1-u)^{n-k-1} du = \left(\frac{1}{b} \right)^\alpha. \quad (35)$$

Since $(1/b)^\alpha = (1/b)^\alpha \int_0^1 (n!/(k-1)!(n-k)!) u^{k-1} (1-u)^{n-k} du$, the last equality, after some simplifications, can be expressed as

$$\int_0^1 u^{k-1} \left[\frac{\bar{F}_X(bF_X^{-1}(u))}{1-u} - \left(\frac{1}{b} \right)^\alpha \right] (1-u)^{n-k} du = 0. \quad (36)$$

The rest of the proof is similar to the proof of part (a).

So far, some results of characterization of power function distribution have been obtained. For example, it was characterized by Ahsanullah et al. [25] through lower records. Also Khan and Khan [26] and Lim and Lee [27] characterized it based on dependency property of lower records. Tavangar [28] presented a characterization of it using dual generalized order statistics. Now, in the next theorem new characterization results of power function distribution are proved. $\square$

Theorem 3. Suppose that $X_1, X_2, \dots, X_n$ are iid continuous random variables from cdf $F_X(\cdot)$ with support $[0, (1/\beta)]$. Then X_i 's have power function distribution with cdf (15) if and only if one of the following statements holds.

(a) For all $n \geq k$ and $0 < a < 1$, there exists $j_0 \in \{0, 1, \dots\}$ such that for $k = j_0 + 1$ and some $\alpha > 0$, we have

$$P(K'_-(n, k, a) = j_0) = (1 - a^\alpha)^{j_0}. \quad (37)$$

(b) For a fixed $k \geq 1$ and for all $n \geq k, 0 < a < 1$ and some $\alpha > 0$, we have

$$E(K'_-(n, k, a)) = (k-1)(1 - a^\alpha). \quad (38)$$

Proof. By supposing X_i 's have power function distribution with cdf (15), from equations (7) and (16), parts (a) and (b) can be easily concluded. Let condition (a) be satisfied, then

$$\begin{aligned}
0 &= P(K'_-(n, k, a) = j_0) - (1 - a^\alpha)^{j_0} \\
&= \int_0^{1/\beta} \left(1 - \frac{F_X(ax)}{F_X(x)}\right)^{j_0} dF_{k:n}(x) - (1 - a^\alpha)^{j_0} \\
&= \frac{n!}{(k-1)!(n-k)!} \int_0^{1/\beta} \left(1 - \frac{F_X(ax)}{F_X(x)}\right)^{j_0} F_X^{k-1}(x) (1 - F_X(x))^{n-k} dF_X(x) - (1 - a^\alpha)^{j_0} \\
&= \frac{n!}{(k-1)!(n-k)!} \int_0^1 \left(1 - \frac{F_X(aF_X^{-1}(u))}{u}\right)^{j_0} u^{k-1} (1-u)^{n-k} du - (1 - a^\alpha)^{j_0} \frac{n!}{(k-1)!(n-k)!} \int_0^1 u^{k-1} (1-u)^{n-k} du \\
&= \int_0^1 u^{k-1} \left[\left(1 - \frac{F_X(aF_X^{-1}(u))}{u}\right)^{j_0} - (1 - a^\alpha)^{j_0} \right] (1-u)^{n-k} du.
\end{aligned} \tag{39}$$

Since $|1 - (F_X(aF_X^{-1}(u))/u)| \leq 2$ and $|1 - a^\alpha| \leq 2$, the quantity $u^{k-1} \{ [1 - (F_X(aF_X^{-1}(u))/u)]^{j_0} - [1 - a^\alpha]^{j_0} \}$ belongs in $L^2[0, 1]$ by Minkowski's inequality. The completeness property of the sequence $\{(1-u)^{n-k}, n \geq k\}$ and equation (39) result in

$$\left\{ \left[1 - \frac{F_X(aF_X^{-1}(u))}{u} \right]^{j_0} - [1 - a^\alpha]^{j_0} \right\} = 0. \tag{40}$$

This is equivalent to

$$\frac{F_X(aF_X^{-1}(u))}{u} = a^\alpha. \tag{41}$$

Taking the change of variable $t = F_X^{-1}(u)$ in (41) gives

$$\frac{F_X(at)}{F_X(t)} = a^\alpha. \tag{42}$$

The function $F_X(x) = cx^\alpha$ is the general solution of above functional equation. This completes the proof of (a). In a similar way, if (b) holds, one can easily prove that the parent population is power function distribution.

The results of Theorem 3 can also be observed directly from Theorem 2 by noticing that $X \sim \text{power}(\alpha, \beta)$ if and only if $Y = (1/X) \sim \text{Pareto}(\alpha, \beta)$. Therefore, for $0 < a < 1$

$$\begin{aligned}
{}^X K'_-(n, k, a) &= \#\{j \in \{1, 2, \dots, n\}, X_j \in (aX_{k:n}, X_{k:n})\} \\
&= \#\left\{j \in \{1, 2, \dots, n\}, \frac{1}{Y_j} \in \left(a \frac{1}{Y_{n-k+1:n}}, \frac{1}{Y_{n-k+1:n}}\right)\right\} \\
&= \#\left\{j \in \{1, 2, \dots, n\}, Y_j \in \left(Y_{n-k+1:n}, \frac{1}{a} Y_{n-k+1:n}\right)\right\} \\
&= {}^Y K'_+(n, n-k+1, \frac{1}{a}),
\end{aligned} \tag{43}$$

where $K'_-(n, k, a)$ with superscript X presents the number of observations near the k th-order statistics related to X -sequence. According to relationship between distributions of power function and standard uniform that is mentioned before, from Theorem 3, the following results without proof are stated.

Corollary 1. Let $X_1, X_2, \dots, X_n$ be iid continuous random variables from cdf $F_X(\cdot)$ that is supported on $[0, (1/\beta)]$. Then, $F_X(\cdot)$ is standard uniform distribution if and only if one of the following statements holds:

(a) For all $n \geq k$ and $0 < a < 1$, there exists $j_0 \in \{0, 1, \dots\}$ such that for $k = j_0 + 1$, we have

$$P(K'_-(n, k, a) = j_0) = (1 - a^\alpha)^{j_0}. \tag{44}$$

(b) For a fixed $k \geq 1$ and for all $n \geq k$ and $0 < a < 1$, we have

$$E(K'_-(n, k, a)) = (k-1)(1-a). \tag{45}$$

Remark 1. According to Theorem 1, it is not necessary that the assumptions of Theorem 2 hold for all " $n \geq j_0 + 1$ " or " $n \geq k$." This fact is also true for Theorem 3 and Corollary 1. So, the results of Theorems 2 and 3 and Corollary 1 hold if their assumptions provide for any increasing subsequence $\{n_i \geq j_0 + 1, i \geq 1\}$ or $\{n_i \geq k, i \geq 1\}$ such that the equality (23) holds.

Remark 2. Some characterization results of two-parameter exponential distribution have been obtained based on counting random variable $K_+(n, k, a)$ by Akbari and Akbari [17]. Their results of characterizations in Section 2 can also be derived directly by Theorems 2 and 3. For considering this claim, suppose X be a random variable having two-parameter exponential distribution with parameters (μ, α) , denoted by $\text{Exp}(\mu, \alpha)$, and the following cdf:

$$\bar{F}_X(x) = e^{-\alpha(x-\mu)}, \quad x > \mu. \tag{46}$$

Since $X \sim \text{Exp}(\mu, \alpha)$ if and only if $Z = e^X \sim \text{Pareto}(\alpha, e^\mu)$, the following relationship holds between ${}^X K_+$ and ${}^Z K'_+$. For $a > 0$,

$$\begin{aligned}
{}^X K_+(n, k, a) &= \#\{j \in \{1, \dots, n\}, X_j \in (X_{k:n}, X_{k:n} + a)\} \\
&= \#\{j \in \{1, \dots, n\}, Z_j \in (Z_{k:n}, e^a Z_{k:n})\} \\
&= {}^Z K'_+(n, k, e^a).
\end{aligned} \tag{47}$$

Also, $X \sim \text{Exp}(\mu, \alpha)$ if and only if $T = e^{-X} \sim \text{power}(\alpha, e^\mu)$. So for $a > 0$,

$$\begin{aligned} {}^X K_+(n, k, a) &= \#\{j \in \{1, \dots, n\}, X_j \in (X_{k:n}, X_{k:n} + a)\} \\ &= \#\{j \in \{1, \dots, n\}, T_j \in (e^{-a} T_{n-k+1:n}, T_{n-k+1:n})\} \\ &= {}^T K'_-(n, n-k+1, e^{-a}). \end{aligned} \quad (48)$$

4. Goodness-of-Fit Test Results

So far, many results of GOF tests for different distributions using characterization results have been obtained. For example, Rizzo [29], Obradovi et al. [30], and Volkova [31] obtained GOF tests for Pareto distribution in terms of its different characteristic properties. According to Nikitin [32], tests based on characterization results are usually more efficient than other tests, because the unique feature of the same distribution has been used in constructing test statistics.

Let $X_1, \dots, X_n$ be iid random variables from continuous distribution function $F_X(x)$. For testing null hypothesis $H_0: F_X(x) = F_0(x) = 1 - (x/\beta)^\alpha, x \geq \beta$ against $H_1: F_X(x) \neq F_0(x)$ for some x , are presented two test statistics based on characterization results of Theorem 2. According to part (a) of this theorem, the null hypothesis H_0 will be rejected if there exists “ n ” such that for all $j \leq n-1$ and $k = n-j$, equation (24) is not satisfied, i.e., the value of quantity

$$\left| P(K'_+(n, k, b)) - \left(1 - \left(\frac{1}{b}\right)^\alpha\right)^j \right|, \quad (49)$$

to be large. From (9) and assumption $k = n-j$, the above expression is equivalent to

$$\int_0^1 F_X^{n-j-1}(x) \left| \{\bar{F}_X(x) - \bar{F}_X(bx)\}^j - \left\{ \bar{F}_X(x) \left(1 - \left(\frac{1}{b}\right)^\alpha\right) \right\}^j \right| dF_X(x). \quad (50)$$

Replacing $F_X(x)$ by $F_n(x)$, the empirical distribution function, a point estimator of (50), can be considered as

$$\begin{aligned} D_P(j, n) &= \frac{1}{n} \sum_{i=1}^n F_n^{n-j-1}(x_{i:n}) \left| \{F_n(bx_{i:n}) - F_n(x_{i:n})\}^j - \left\{ (1 - F_n(x_{i:n})) \left(1 - \left(\frac{1}{b}\right)^\alpha\right) \right\}^j \right| \\ &= \frac{1}{n} \sum_{i=1}^n \left(\frac{i}{n+1}\right)^{n-j-1} \left| \left\{ F_n(bX_{i:n}) - \frac{i}{n+1} \right\}^j - \left\{ \left(1 - \frac{i}{n+1}\right) \left(1 - \left(\frac{1}{b}\right)^\alpha\right) \right\}^j \right|. \end{aligned} \quad (51)$$

Therefore, the test statistic that its large value will be rejected, H_0 , is given by

$$D_P = \sum_{m=1}^n \sum_{j=0}^{n-1} D_P(j, m). \quad (52)$$

With the same discussion, the another test statistic based on the part (b) of Theorem 2 can be as

$$D_E = \sum_{m=1}^n \sum_{k=1}^m D_E(k, m), \quad (53)$$

where $D_E(k, n) = (1/n) \sum_{i=1}^n (i/(n+1))^{k-1} (1 - (i/(n+1)))^{n-k} |(1 - F_n(bX_{i:n})) - (1 - (i/(n+1))) (1/b)^\alpha|$.

It is obvious that D_P and D_E are free of scale parameter of Pareto distribution and their large values reject H_0 .

In the rest of this section, the power values of two test statistics D_P and D_E will be compared with well-known tests, namely, Kolmogorov–Smirnov and Cramer–von Mises tests which their statistics are, respectively,

$$D = \max\{D_+, D_-\}, \quad (54)$$

where $D_+ = \max_{1 \leq i \leq n} |(i/n) - F_0(X_{i:n})|$ and $D_- = \max_{1 \leq i \leq n} |F_0(X_{i:n}) - ((i-1)/n)|$, and

$$W^2 = \frac{1}{12n} + \sum_{i=1}^n \left(\frac{2i-1}{2n} - F_0(X_{i:n}) \right)^2. \quad (55)$$

Since Pareto distribution is long tail, following alternative distributions that are long tail on the right-hand side are considered for comparison of power values of statistics D_P, D_E, D , and W^2 .

- (i) The Weibull distribution with density $(a/b)(x/b)^{a-1} \exp(-(x/b)^a)$, denoted by $W(a, b)$
- (ii) The gamma distribution with density $(1/\lambda^r \Gamma(r)) x^{r-1} \exp(-x/\lambda)$, denoted by $G(r, \lambda)$
- (iii) The lognormal distribution with density $(1/\sqrt{2\pi} \sigma x) \exp(-((\ln x - \mu)^2/2\sigma^2))$, denoted by $LN(\mu, \sigma)$
- (iv) The uniform distribution with density $(1/10), 0 \leq x \leq 10$, denoted by U
- (v) The log-gamma distribution with density $(\exp(((x-a)/b) - \exp((x-a)/b))/b)$, denoted by $LG(a, b)$

Since it is not easy to find the null distribution of D_P, D_E, D , and W^2 , Monte Carlo simulations with 10000 replications are used for calculating their power values and critical values at 5 percent significance level. Tables 1 and 2 show the results for null distribution Pareto (1,2) and Pareto (2,2), respectively. Because the statistics D_P and D_E have the parameter b , one can choose an optimal b to maximize corresponding power values. So, these values are calculated and shown in the tables. Unfortunately there is no accurate

TABLE 1: Monte Carlo power estimates of the tests at 5 percent significant level with null distribution function Power (1,2).

Sample size	Test	LG (2,8)	LN (0,1)	G (3,1)	W (0.5,2)	U
5	D_E	0.9999 (0.0494) $b = 2.1$	0.892 (0.0475) $b = 1.66$	0.7619 (0.05) $b = 1.69$	0.8969 (0.048) $b = 1.68$	0.8698 (0.0498) $b = 1.67$
	D_P	0.9999 (0.05) $b = 2.22$	0.892 (0.0495) $b = 1.42$	0.7619 (0.05) $b = 1.39$	0.8969 (0.05) $b = 1.65$	0.8698 (0.05) $b = 1.39$
	D	0.757 (0.05)	0.752 (0.05)	0.3217 (0.05)	0.618 (0.05)	0.121 (0.05)
	W^2	0.428 (0.05)	0.588 (0.05)	0.3217 (0.05)	0.367 (0.05)	0.08 (0.05)
7	D_E	1 (0.05) $b = 2.3$	0.9626 (0.499) $b = 1.67$	0.8711 (0.05) $b = 1.61$	0.963 (0.05) $b = 1.71$	0.9528 (0.05) $b = 1.62$
	D_P	1 (0.05) $b = 2.42$	0.9626 (0.05) $b = 1.37$	0.8711 (0.05) $b = 1.88$	0.963 (0.05) $b = 1.381$	0.9528 (0.05) $b = 1.89$
	D	0.826 (0.05)	0.810 (0.05)	0.392 (0.05)	0.81 (0.05)	0.11 (0.05)
	W^2	0.833 (0.05)	0.766 (0.05)	0.496 (0.05)	0.764 (0.05)	0.094 (0.05)
10	D_E	1 (0.05) $b = 2.38$	1 (0.05) $b = 1.15$	1 (0.05) $b = 1.15$	1 (0.05) $b = 1.11$	0.8813 (0.05) $b = 2$
	D_P	1 (0.05) $b = 2.57$	1 (0.05) $b = 1.38$	1 (0.05) $b = 1.88$	1 (0.05) $b = 1.38$	0.8813 (0.05) $b = 2.01$
	D	0.945 (0.05)	0.9066 (0.05)	0.559 (0.05)	0.9118 (0.05)	0.129 (0.05)
	W^2	0.975 (0.05)	0.899 (0.05)	0.706 (0.05)	0.9071 (0.05)	0.121 (0.05)

TABLE 2: Monte Carlo power estimates of the tests at 5 percent significant level with null distribution function Power (2,2).

Sample size	Tests	LG (2,8)	LN (0,1)	G (3,1)	W (0.5,2)	U
5	D_E	0.9999 (0.049) $b = 2.10$	0.9578 (0.047) $b = 2.21$	0.933 (0.047) $b = 2.13$	0.997 (0.0476) $b = 2.32$	0.975 (0.0487) $b = 2.13$
	D_P	0.9999 (0.05) $b = 2.22$	0.9578 (0.05) $b = 2.24$	0.933 (0.048) $b = 2.04$	0.9969 (0.05) $b = 2.49$	0.9751 (0.05) $b = 2.18$
	D	0.759 (0.05)	0.916 (0.05)	0.2352 (0.05)	0.767 (0.05)	0.414 (0.05)
	W^2	0.4624 (0.05)	0.8089 (0.05)	0.1562 (0.05)	0.5238 (0.05)	0.365 (0.05)
7	D_E	1 (0.05) $b = 2.24$	0.9682 (0.05) $b = 2.26$	0.9589 (0.0499) $b = 2.08$	0.9985 (0.05) $b = 2.23$	0.9922 (0.05) $b = 2.08$
	D_P	1 (0.05) $b = 2.43$	0.9682 (0.0499) $b = 2.39$	0.9589 (0.0486) $b = 2.28$	0.9985 (0.05) $b = 2.47$	0.9922 (0.04) $b = 2.39$
	D	0.8 (0.05)	0.949 (0.05)	0.2243 (0.05)	0.809 (0.05)	0.491 (0.05)
	W^2	0.74 (0.05)	0.932 (0.05)	0.2049 (0.05)	0.763 (0.05)	0.471 (0.05)
10	D_E	1 (0.05) $b = 2.38$	0.9879 (0.05) $b = 2.11$	0.978 (0.05) $b = 2.14$	0.9997 (0.05) $b = 2.11$	0.9977 (0.05) $b = 2.15$
	D_P	1 (0.05) $b = 2.57$	0.9879 (0.05) $b = 2.57$	0.978 (0.05) $b = 2.56$	0.9997 (0.05) $b = 2.94$	0.9977 (0.05) $b = 2.55$
	D	0.9199 (0.05)	0.9883 (0.05)	0.283 (0.05)	0.912 (0.05)	0.647 (0.05)
	W^2	0.9286 (0.05)	0.9883 (0.05)	0.285 (0.05)	0.921 (0.05)	0.675 (0.05)

method to find these values and depend on the support of null and alternative distributions.

The values in parentheses in the tables refer to estimated significance level. According to the results of the two tables, it is concluded that proposed tests are always more powerful than the other tests. Even in small sample size, the proposed tests perform very well and better than the others.

In the following example, it is used real data set to illustrate how the proposed tests can be applied.

Example 3 (As an application to real data). The following data represent the time for break down of a type of electrical insulating material subject to a constant-voltage stress (Nelson [33]).

$$1.97, 0.59, 2.58, 1.69, 2.71, 25.50, 0.35, 0.99, 3.67, 2.07, 0.96, 5.35, 2.90, 13.77. \quad (56)$$

This data recently were used by Tiku and Akkaya [34]. They established that the null hypothesis where data come from exponential distribution cannot be rejected at 10 percent significance level. It is obvious that data come from a distribution with long tail on the right-hand side. So, Pareto distribution can be another suggested distribution for such data. For testing $H_0: X \sim \text{Pareto}$ versus $H_1: X \not\sim \text{Pareto}$, first the parameters of Pareto distribution are estimated with shape parameter $\alpha = 0.51$ and scale parameter $\beta = 0.35$. Then, based on data, the values of the proposed statistics (with $b = 1.2$), Kolmogorov-Smirnov, and Cramer-von Mises statistics have been obtained as follows:

$$\begin{aligned}
 D_P &= 0.563, \\
 D_E &= 1.958, \\
 D &= 0.286, \\
 W^2 &= 0.2989.
 \end{aligned}
 \tag{57}$$

Also, with 10000 replications based on estimated Pareto distribution, the critical values at 10 percent significance level have been calculated, respectively,

$$0.75, 2.3, 0.48, 1.22. \tag{58}$$

Hence, the null hypothesis that data come from Pareto distribution cannot be rejected using this data.

Data Availability

No data are included in the study.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

- [1] B. C. Arnold, N. Balakrishnan, and H. N. Nagaraja, *A First Course in Order Statistics*, SIAM, Philadelphia, PA, USA, 2008.
- [2] B. Eisenberg, G. Stengle, and G. Strang, "The asymptotic probability of a tie for first place," *The Annals of Applied Probability*, vol. 3, no. 3, pp. 731–745, 1993.
- [3] A. G. Pakes and F. W. Steutel, "On the number of records near the maximum," *Australian Journal of Statistics*, vol. 39, no. 2, pp. 179–192, 1997.
- [4] A. Dembińska, A. Stepanov, and J. Wesolowski, "How many observations fall in a neighbourhood of an order statistic?" *Communications in Statistics Theory and Methods*, vol. 36, pp. 1–17, 2007.
- [5] A. G. Pakes and Y. Li, "Limit laws for the number of near maxima via the Poisson approximation," *Statistics & Probability Letters*, vol. 40, no. 4, pp. 395–401, 1998.
- [6] Y. Li, "A note on the number of records near the maximum," *Statistics & Probability Letters*, vol. 43, no. 2, pp. 153–158, 1999.
- [7] A. G. Pakes, "The number and sum of near-maxima for thin-tailed populations," *Advances in Applied Probability*, vol. 32, no. 4, pp. 1100–1116, 2000.
- [8] A. G. Pakes, "Numbers of observations near order statistics," *Australian & New Zealand Journal of Statistics*, vol. 51, no. 4, pp. 375–395, 2009.
- [9] N. Balakrishnan and A. Stepanov, "A note on the number of observations near an order statistic," *Journal of Statistical Planning and Inference*, vol. 134, no. 1, pp. 1–14, 2005.
- [10] N. Balakrishnan and A. Stepanov, "Asymptotic properties of numbers of near minimum observations under progressive type-II censoring," *Journal of Statistical Planning and Inference*, vol. 138, no. 4, pp. 1010–1020, 2008.
- [11] A. Dembińska, "Asymptotic properties of numbers of observations in random regions determined by central order statistics," *Journal of Statistical Planning and Inference*, vol. 142, pp. 516–528, 2012.
- [12] A. Dembińska, "Limit theorems for proportions of observations falling into random regions determined by order statistics," *Australian & New Zealand Journal of Statistics*, vol. 54, pp. 199–210, 2012.
- [13] A. Dembińska, "Asymptotic normality of numbers of observations in random regions determined by order statistics," *Statistics*, vol. 48, pp. 508–523, 2014.
- [14] S. Müller, "Tail estimation based on numbers of near m-extremes," *Methodology and Computing in Applied Probability*, vol. 5, pp. 197–210, 2003.
- [15] E. Hashorva and J. Hüsler, "Estimation of tails and related quantities using the number of near-extremes," *Communications in Statistics: Theory and Methods*, vol. 34, no. 2, pp. 337–349, 2005.
- [16] M. Akbari, M. Fashandi, and J. Ahmadi, "Characterizations based on the numbers of near-order statistics," *Statistical Papers*, vol. 57, no. 1, pp. 21–30, 2016.
- [17] M. Akbari and M. Akbari, "Some applications of near-order statistics in two-parameter exponential distribution," *Journal of Statistical Theory and Applications*, vol. 19, no. 1, pp. 21–27, 2020.
- [18] J. R. Higgins, *Completeness and Basis Properties of Sets of Special Functions*, Cambridge University Press, New York, NY, USA, 2004.
- [19] M. Y. Lee and S. K. Chang, "Characterizations based on the independence of the exponential and Pareto distributions by record values," *Journal of Applied Mathematics and Computing*, vol. 18, pp. 497–503, 2005.
- [20] E. E. Afify, "Characterization of mixtures from exponential and Pareto distributions using left truncated moments," *IJSS*, vol. 5, pp. 1–18, 2006.
- [21] M. Ahsanullah and M. Shakil, "A note on the characterizations of Pareto distribution by upper record values," *Communications of the Korean Mathematical Society*, vol. 27, no. 4, pp. 835–842, 2012.
- [22] M. Ahsanullah, M. Shakil, and B. M. G. Kibria, "Characterizations of continuous distributions by truncated moment," *Journal of Modern Applied Statistical Methods*, vol. 15, no. 1, pp. 316–331, 2016.
- [23] Z. M. Nofal and Y. M. El Gebaly, "New characterizations of the Pareto distribution," *Pakistan Journal of Statistics and Operation Research*, vol. 13, no. 1, pp. 63–74, 2017.
- [24] J. Aczél, *Lectures on Functional Equations and Their Applications*, Academic Press, New York, NY, USA, 1966.
- [25] M. Ahsanullah, M. Shakil, and B. M. Golam Kibria, "Characterization of power function distribution based on lower records," *Probst Forum*, vol. 6, pp. 68–72, 2013.
- [26] M. I. Khan and M. A. R. Khan, "Characterization of generalized uniform distribution based on lower record values," *Probst Forum*, vol. 10, pp. 23–26, 2017.
- [27] E. H. Lim and M. Y. Lee, "A characterization of the power function distribution by independent property of lower record values," *Journal of the Chungcheong Mathematical Society*, vol. 26, pp. 269–273, 2013.
- [28] M. Tavangar, "Power function distribution characterized by dual generalized order statistics," *JIRSS*, vol. 10, pp. 13–27, 2011.
- [29] M. L. Rizzo, "New goodness-of-fit tests for Pareto distributions," *Astin Bulletin*, vol. 39, no. 2, pp. 691–715, 2009.
- [30] M. Obradovi, M. Jovanovic, and B. Milosevic, "Goodness-of-fit tests for Pareto distribution based on a characterization and its asymptotic," *Statistics*, vol. 49, pp. 1026–1041, 2015.
- [31] K. Y. Volkova, "On asymptotic efficiency of goodness-of-fit tests for the Pareto distribution based on its characterization," *Filomat*, vol. 29, pp. 2311–2324, 2015.

- [32] Y. Nikitin, "Test based on characterizations, and their efficiencies: a survey," *Acta et Commentationes Universitatis Tartuensis de Mathematica*, vol. 21, pp. 3–24, 2017.
- [33] W. B. Nelson, "Statistical methods for accelerated life test data—the inverse power law model," Technical report 71-C011, General Electric, Schenectady, NY, USA, 1970.
- [34] M. L. Tiku and A. D. Akkaya, *Robust Estimation and Hypothesis Testing*, New Age International Pvt. Ltd., New Dehli, India, 2004.