Aligarh Journal of Statistics
Vol. 35(2015), 51-60

CHARACTERIZATIONS OF THE PARETO DISTRIBUTION IN THE PRESENCE OF OUTLIERS

U. J. Dixit and M. Jabbari Nooghabi

ABSTRACT

Here we have given some characterizations for the Pareto distribution in the presence of outliers. It is proved that a necessary and sufficient condition for $f(x)$ to be a Pareto density function in the presence of outliers is that the statistics $X(r)$ and $\frac{x(s)}{x(r)}(1 \le r < s \le n)$ are independent. Further, we have derived some another characterizations of the Pareto distribution in the presence of outliers.

1. INTRODUCTION

Ahsanullah and Kabir (1973) proved that necessary and sufficient condition for $f(x)$ to be a Pareto distribution is that the statistics $X(r)$ and $\frac{X(s)}{X(r)}(1 \le r < s \le n)$ are independent. Dallas (1976) proved that for a cumulative distribution function (CDF) $G(y)$ $y \ge \beta$, if $E(Y^r \mid Y > c) = E\left(\frac{Yc}{\beta}\right)^r$ holds then Y has a Pareto distribution.

In this paper, we assume that the random variables $X_1, X_2, ..., X_n$ are such that k of them are distributed with probability density function (pdf)

$$f_2(x;\alpha,\beta,\alpha) = \frac{\alpha(\beta\theta)\alpha}{x^{\alpha+1}}, < \beta\theta \le x, \alpha > 0, \beta > 1, \theta > 0, \tag{1}$$

and the remaining $(n-k)$ random variables are distributed as

$$f_1(x;\alpha,\theta) = \frac{\alpha\theta^{\alpha}}{x^{\alpha+1}}, 0 < \theta \le x, \alpha > 0 \tag{2}$$

One should note that sets of the observation (i.e. k and $n-k$) are independent. But $X_1, X_2, ..., X_n$ are not independent because of the model of outliers (for more details see Dixit (1989), Dixit and Jabbari Nooghabi (2011a) and Dixit and Jabbari

Nooghabi (2011b). Also, we may note that our assumptions are based on Dixit's model for the outliers problem and it is totaly different than the mixture models which considers $X_1, X_2, ..., X_n$ are independent.

Here, we have extended the approaches of Ahsanullah and Kabir (1973) and Dallas (1976) for the homogenous case of the Pareto distribution and derived some characterizations of thePareto distribution in the presence of outliers.

2. PREREQUISITE RESULTS

Assume that $X_{(1)}, X_{(2)} <,..., X_{(n)}$ be the order statistics of a random sample of size n such that k out of n are coming from pdf f_2 (or CDF F_2) and the remaining $n-k$ follow the pdf f_1 (or CDF F_1). The CDF and pdf of $r^{th}(1 \le r \le n)$ order statistic are

$$H_{X(r)}(x) = \sum_{i-r}^{n} \sum_{j-m_5}^{m_6} C(k, j)\left[F_2(x)\right]^j \left[1 - F_2(x)\right]^{k-j} \times C(n-k, i-j)\left[F_1(x)\right]^{i-j} \left[1 - F_1(x)\right]^{n-k-i+j} \quad (3)$$

where $m_5 = \max(0, i-n+k)$ and $m_6 = \min(k, i)$ and

$$h_{X(r)}(x) = kf_2(x)\sum_{j-m_1}^{m_2} \left\{ \begin{array}{l} C(k-1, j)\left[F_2(x)\right]^j \left[1 - F_2(x)\right]^{k-j-1} C(n-k, r-1-j) \\ \times\left[F_1(x)\right]^{r-j-1}\left[1 - F_1(x)\right]^{n-k-i+j+1} \end{array} \right\} + (n-k) f_1(x)\sum_{j=m_3}^{m_4} \left\{ \begin{array}{l} C(k, j)\left[F_2(x)\right]^j \times \left[1 - F_2(x)\right]^{k-j} C(n-k-1, r-1-j) \\ \left[F_1(x)\right]^{r-j-1}\left[1 - F_1(x)\right]^{n-k-r+j} \end{array} \right\} \quad (4)$$

where $m_1 = \max(0, k+r-n)$, $m_2 = \min(k-1, r-1)$, $m_3 = \max(0, k+r-n)$, $m_4 = \min(k, r-1)$, respectively (for more details see Dixit (1987, 1989, 1994) and Dixit and Jabbari Nooghabi (2011b)).

Further, the joint CDF and pdf of $(X_{(r)}, X_{(s)}\,(1 \le r < s \le n)$ are

$$H_{X(r),X(s)}(x, y) = \sum_{j=s}^{n} \sum_{i=r}^{j} \sum_{m=w_9}^{w_{10}} \sum_{l=t_9}^{t_{10}} \left\{ \begin{array}{l} C(k, m)C(k-m, l)\left[F_2(x)\right]^m \left[F_2(y) - F_2(x)\right]^l \\ \times\left[1 - F_2(y)\right]^{k-m-l} C(n-k, i-m) \\ \times C(n-k-i+m, j-i-l)[F_1(x)]^{i-m} \\ \times[F_1(y) - F_1(x)]^{j-i-1}[1 - F_1(y)]^{n-k-j+m+l} \end{array} \right\} \quad (5)$$

where; $w_9 = \max(0, 1-n+k)$, $w_{10} = \min(k, i)$, $t_9 = \max(0, j-n+k-m)$

$t_{10} = \min(k-m, j-1)$ and

$$h_{X(r),X(s)}(x,y) = k(k-1) f_2(x) f_2(y) \sum_{j=w_1}^{w_2} \sum_{i=t_1}^{t_2} \left\{ \begin{array}{l} C(k-2,j)C(n-k,r-1-j)[F_2(x)]^j [F_1(x)]^{r-1-j} \\ \times C(k-2-j,i)C(n-k-r+j+1,n-s-i) \\ \times [1-F_2(y)]^i \times [1-F_1(y)]^{n-s-i} [F_2(y)-F_2(x)]^{k-1-j-2} \\ \times [F_1(y)-F_1(x)]^{s-r-k+i+j+1} \end{array} \right\}$$

$$+ (n-k)(n-k-1) f_1(x) f_1(y) \sum_{j=w3}^{w4} \sum_{i=t3}^{t4} \left\{ \begin{array}{l} C(n-k-2,j)C(k,r-1-j)[F_1(x)]^j [F_2(x)]^{r-1-j} C(n-k-2-j,i) \\ \times C(k-r+j+1,n-s-i)[1-F_1(y)]^i [1-F_2(y)]^{n-s-i} [F_1(y)-F_1(x)]^{n-k-2-i-j} \\ \times [F_2(y)-F_2(x)]^{s-r+k-n+i+j+1} \end{array} \right\}$$

$$+ k(n-k) f_1(x) f_2(y) \times \sum_{j=w_5}^{w_6} \sum_{i=t_5}^{t_6} \left\{ \begin{array}{l} C(n-k-1,j)C(k-1,r-1-j)[F_1(x)]^j [F_2(x)]^{r-j-1} C(n-k-j-1,i) \\ \times C(k-r+j,n-s-i)[1-F_1(y)]^i [1-F_2(y)]^{n-s-i} \\ \times [F_1(y)-F_1(x)]^{n-k-i-j-1} [F_2(y)-F_2(x)]^{s-r-n+k+i+j} \end{array} \right\}$$

$$+ k(n-k) f_1(y) f_2(x) \sum_{j=w_7}^{w_8} \sum_{i=t_7}^{t_8} \left\{ \begin{array}{l} C(k-1,j)C(n-k-1,r-1-j)[F_2(x)]^j [F_1(x)]^{r-1-j} \\ \times C(k-j-1,i)C(n-k-r+j,n-s-i)[1-F_2(y)]^i \\ \times [1-F_1(y)]^{n-s-i} [F_2(y)-F_2(x)]^{k--j-1} [F_1(y)-F_1(x)]^{s-r-k+i+j} \end{array} \right\} \quad (6)$$

where $w_1 = \max(0, r-n+k-1)$, $w_2 = \min(k-2, r-1)$, $t_1 = \max(0, k-s+r-j-1)$ and $t_2 = \min(k-j-2, n-s)$, $w_3 = \max(0, r-k-1)$, $w_4 = \min(n-k-j-2, r-1)$, $t_3 = \max(0, n-s-k+r-j-1)$ and $t_4 = \min(n-k-j-2, n-s)$, $w_5 = \max(0, r-k)$, $w_6 = \min(n-k-1, r-1)$, $t_5 = \max(0, n-s-k+r-j-1)$ $t_6 = \min(n-k-jr-j-1, n-s)$, $w_7 = \max(0, r-n+k)$, $w_8 = \min(k-1, r-1)$, $t_7 = \max(0, k-s+r-j-1)$, $t_8 = \min(k-j-1, n-s)$, respectively.

One should note that if k=1 the joint pdf of $X_{(r)}, X_{(s)}$ is given in Sinha (1973). Also, if we put $f_1 = f_2$ and $F_1 = F_2$ then all pdfs and CDFs are reduced to homogeneous cases.

The following equations are named as Pexider's equations.

$$f(xy) = g(x) + h(y), \quad (7)$$

and

$$f(xy) = g(x)h(y), \quad (8)$$

For solving these equations, the following Theorem has taken from Aczel (1966) (Theorem 4. in p. 144) and Kuczma (2008) (Theorem 13.3.4. in p. 358).

Theorem 2.1. The general solutions, with f continuous in a point, of (7)and (8), respectively, both supposed for positive x and y, are

$$f(t) = c\ln(\alpha\beta t), g(t) = c\ln(\alpha t), h(t) = c\ln(\beta t), (\alpha > 0, t > 0) \quad (9)$$

and

$$f(t) = abt^c, g(t) = at^c, h(t) = bt^c, (t > 0) \quad (10)$$

respectively, supplemented with the following trivial solutions in case of (8).

$f(t)=c\ln(\alpha\beta t), g(t)=c\ln(\alpha t), h(t)=c\ln(\beta t), (\alpha>0, t>0)$

$$\begin{cases} f(t)=0, \\ g(t)=0, \\ h(t)\ arbitrary \end{cases} \quad and \quad \begin{cases} f(t)=0, \\ g(t)=arbitrary, \\ h(t)=0 \end{cases} \tag{11}$$

3. CHARACTERIZATION OF THE PARETO DISTRIBUTION IN THE PRESENCE OF OUTLIERS

Theorem 3.1. Let X be a random variable having an absolutely continuous CDF $F(x)$. A necessary and sufficient condition that X follows the Pareto distribution in the presence of outliers as given by (1) and (2) is that for some r and $s(1\le r<s\le n)$ the statistics $X_{(r)}$ and $\frac{X_{(s)}}{X_{(r)}}$ are independent.

Proof. Necessity: From (6) we can get the joint pdf of $X_{(r)}$ and $X_{(s)}$ Substituting $U=X_{(r)}$ and $V=\frac{X_{(s)}}{X_{(r)}}$ in (6), we can obtain the joint pdf of U and V as $h_{U,V}(u,v)=uh_{X_{(r)},X_{(s)}}(u,uv)$.

Then after some simplification

$$\begin{aligned} h(u,v)=\alpha^2\theta^{\alpha(n-r+1)}\beta^{\alpha k}v^{\alpha(s-n-1)-1}[1-v^{-\alpha}]^{s-r-1}u^{\alpha(r-n-1)-1} \\ \times\Bigg\{k(k-1)\sum_{j=w_1}^{w_2}\sum_{i=t_1}^{t_2}A_1\beta^{-\alpha j}\left[1-\left(\frac{\beta\theta}{u}\right)^{\alpha}\right]^j\left[1-\left(\frac{\theta}{u}\right)^{\alpha}\right]^{r-1-j} \\ +(n-k)(n-k-1)\times\beta^{-\alpha(r-1)}\sum_{j=w_3}^{w_4}\sum_{i=t_3}^{t_4}A_2\beta^{-\alpha j}\left[1-\left(\frac{\theta}{u}\right)^{\alpha}\right]^j\left[1-\left(\frac{\beta\theta}{u}\right)^{\alpha}\right]^{r-1-j} \\ +k(n-k)\beta^{-\alpha(r-1)}\times\sum_{j=w_5}^{w_6}\sum_{i=t_5}^{t_6}A_3\beta^{\alpha j}\left[1-\left(\frac{\theta}{u}\right)^{\alpha}\right]^j\left[1-\left(\frac{\beta\theta}{u}\right)^{\alpha}\right]^{r-1-j} \\ +k(n-k)\times\sum_{j=w_7}^{w_8}\sum_{i=t_7}^{t_8}A_4\beta^{-\alpha j}\left[1-\left(\frac{\beta\theta}{u}\right)^{\alpha}\right]^j\left[1-\left(\frac{\theta}{u}\right)^{\alpha}\right]^{r-1-j}\Bigg\} \end{aligned} \tag{12}$$

where

$$\begin{cases} A_1 = C(k-2, j)C(n-k, r-1-j)C(k-2-j, i)C(n-k-r+j+1, n-s-i), \\ A_2 = C(n-k-2, j)C(k, r-1-j)C(n-k-2-j, i)C(k-r+j+1, n-s-i), \\ A_3 = C(n-k-1, j)C(k-1, r-1-j)C(n-k-j-1, i)C(k-r+j, n-s-i), \\ A_4 = C(k-1, j)C(n-k-1, r-1-j)C(k-j-1, i)C(n-k-r+j, n-s-i). \end{cases} \tag{13}$$

Therefore, it establishes the independence of U and V.

Sufficiency: Here we assume that U and V are independent. The joint pdf of U and V is

$$\begin{aligned} h(u,v) = \Big\{ & k(k-1)uf_2(u) f_2(uv) \sum_{j=w_1}^{w_2} \sum_{i=t_1}^{t_2} \Big\{ A_1 [F_2(u)]^j [F_1(u)]^{r-1-j} [1-F_2(uv)]^i \\ & + [1-F_1(uv)]^{n-s-i} [F_2(uv)-F_2(u)]^{k-j-i-2} [F_1(uv)-F_1(u)]^{s-r-k+i+j+1} \Big\} \\ & + (n-k)(n-k-1)uf_1(u) f_1(uv) \sum_{j=w_3}^{w_4} \sum_{i=t_3}^{t_4} \Big\{ A_2 [F_1(u)]^j [F_2(u)]^{r-1-j} [1-F_1(uv)]^i \\ & [1-F_2(uv)]^{n-s-i} [F_1(uv)-F_1(u)]^{n-k-2-i-j} [F_2(uv)-F_2(u)]^{s-r+k-n+i+j-1} \Big\} \\ & + k(n-k)uf_1(u) f_2(uv) \sum_{j=w_5}^{w_6} \sum_{i=t_5}^{t_6} \Big\{ A_3 [F_1(u)]^j [F_2(u)]^{r-1-j} [1-F_1(uv)]^i \\ & [1-F_2(uv)]^{n-s-i} [F_1(uv)-F_1(u)]^{n-k-i-j-1} [F_2(uv)-F_2(u)]^{s-r+k-n+i+j} \Big\} \\ & + k(n-k)uf_2(u) f_1(uv) \sum_{j=w_7}^{w_8} \sum_{i=t_7}^{t_8} \Big\{ A_4 [F_2(u)]^j [F_1(u)]^{r-1-j} [1-F_2(uv)]^i \\ & \times [1-F_1(uv)]^{n-s-i} [F_2(uv)-F_2(u)]^{k-i-j-1} [F_1(uv)-F_1(u)]^{s-r-k+i+j} \Big\} \end{aligned} \tag{14}$$

where $A_1, A_2, A_3,$ and A_4 are given in (13).

By using some elementary algebra we have

$$\begin{aligned} & h(u,v) = \{k(k-1)uf_2(u) f_2(uv)[F_1(u)]^{r-1}[1-F_1(uv)]^{n-s}[F_1(uv)-F_1(u)]^{s-r-k+1}\} \\ & \times [F_2(uv)-F_2(u)]^{k-2} \sum_{j=w_1}^{w_2} \sum_{i=t_1}^{t_2} \Big\{ A_1 \left[\frac{F_2(u)}{F_1(u)}\right]^j \left[\frac{1-F_1(u)}{1-F_2(u)}\right]^j \left[\frac{1-F_2(uv)}{F_2(uv)-F_2(u)}\right]^i \\ & \times \left[\frac{F_1(uv)-F_1(u)}{1-F_1(uv)}\right]^i \left[\frac{F_1(uv)-F_1(u)}{1-F_1(u)}\right]^j \left[\frac{1-F_2(u)}{F_2(uv)-F_2(u)}\right]^j \Big\} \\ & + (n-k)(n-k-1)uf_1(u) f_1(uv)[F_2(u)]^{r-1}[1-F_2(uv)]^{n-s}[F_1(uv)-F_1(u)]^{n-k-2} \end{aligned}$$

$$\times[F_2(uv)-F_2(u)]^{s-r+k-n+1}\sum_{j=w_3}^{w_4}\sum_{i=t_3}^{t_4}\left\{A_2\left[\frac{F_1(u)}{F_2(u)}\right]^j\left[\frac{1-F_2(u)}{1-F_1(u)}\right]^j\left[\frac{1-F_1(uv)}{F_1(uv)-F_1(u)}\right]^i\right.$$

$$\left.\times\left[\frac{F_2(uv)-F_2(u)}{1-F_2(uv)}\right]^i\left[\frac{1-F_1(u)}{F_1(uv)-F_1(u)}\right]^j\left[\frac{F_2(uv)-F_2(u)}{1-F_2(u)}\right]^j\right\}$$

$$+\left\{k(n-k)uf_1(u)f_2(uv)[F_2(u)]^{r-1}[1-F_2(uv)]^{n-s}[F_1(uv)-F_1(u)]^{n-k-1}\right.$$

$$\times[F_2(uv)-F_2(u)]^{s-r+k+n}\sum_{j=w_5}^{w_6}\sum_{i=t_5}^{t_6}\{A_3\left[\frac{F_1(u)}{F_2(u)}\right]^j\left[\frac{1-F_2(u)}{1-F_1(u)}\right]^j\left[\frac{1-F_1(uv)}{F_1(uv)-F_1(u)}\right]^i$$

$$\left.\times\left[\frac{F_2(uv)-F_2(u)}{1-F_2(uv)}\right]^i\left[\frac{1-F_1(u)}{F_1(uv)-F_1(u)}\right]^j\left[\frac{F_2(uv)-F_2(u)}{1-F_2(u)}\right]^j\right\}$$

$$+\left\{k(n-k)uf_2(u)f_1(uv)[F_1(u)]^{r-1}[1-F_1(uv)]^{n-s}[F_2(uv)-F_2(u)]^{k-1}\right.$$

$$\times[F_1(uv)-F_1(u)]^{s-r-k}\sum_{j=w_7}^{w_8}\sum_{i=t_7}^{t_8}\{A_4\left[\frac{F_2(u)}{F_1(u)}\right]^j\left[\frac{1-F_1(u)}{1-F_2(u)}\right]^j\left[\frac{1-F_2(uv)}{F_2(uv)-F_2(u)}\right]^i$$

$$\left.\times\left[\frac{F_1(uv)-F_1(u)}{1-F_1(uv)}\right]^i\left[\frac{F_1(uv)-F_1(u)}{1-F_1(u)}\right]^j\left[\frac{1-F_2(u)}{F_2(uv)-F_2(u)}\right]^j\right\} \tag{15}$$

Also from (4) and after some simplification, the marginal pdf of $U = X_{(r)}$ is as $h_1(u)$.

$$h_1(u)=[1-F_2(u)]^{k-1}[F_1(u)]^{r-1}[1-F_1(u)]^{n-k-r+1}D, \tag{16}$$

where

$$D=kf_2(u)\sum_{j=m_1}^{m_2}B_1\left[\frac{F_2(u)}{F_1(u)}\right]^j\left[\frac{1-F_1(u)}{1-F_2(u)}\right]^j+(n-k)f_1(u)\left[\frac{1-F_2(u)}{1-F_1(u)}\right]$$

$$\times\sum_{j=m_3}^{m_4}B_2\left[\frac{F_2(u)}{F_1(u)}\right]^j\left[\frac{1-F_1(u)}{1-F_2(u)}\right]^j,$$

and

$$\begin{cases}B_1=C(k-1,j)C(n-k,r-1-j),\\ B_1=C(k,j)C(n-k-1,r-1-j).\end{cases} \tag{17}$$

Therefore from independency of U and V, we can write

$$h_2(v)=\frac{h(u,v)}{h_1(u)} \tag{18}$$

where $h_2(v)$ is pdf of V.

Letting $p = p(u,v) = \dfrac{1-F_1(uv)}{1-F_1(u)}$ and $q = q(u,v) = \dfrac{1-F_2(uv)}{1-F_2(u)}$,

we obtain

$$h_2(v) = -\Bigg\{ k(k-1)p^{n-s}[1-p]^{s-k-r+1}[1-q]^{k-2}\sum_{j=w_1}^{w_2}\sum_{i=t_1}^{t_2} A_1\left[\frac{F_2(u)}{F_1(u)}\right]^j\left[\frac{1-F_1(u)}{1-F_2(u)}\right]^j$$

$$\times p^{-i}[1-p]^{i+j}q^i[1-q]^{-i-j} f_2(u)\frac{\partial q}{\partial v} + (n-k)(n-k-1)\left[\frac{F_2(u)}{F_1(u)}\right]^{r-1}\left[\frac{1-F_1(u)}{1-F_2(u)}\right]^{r-2}$$

$$[1-p]^{n-k-2}q^{n-s}[1-q]^{k+s-r-n+1}\sum_{j=w_3}^{w_4}\sum_{i=t_3}^{t_4} A_2\left[\frac{F_1(u)}{F_2(u)}\right]^j\left[\frac{1-F_2(u)}{1-F_1(u)}\right]^j q^{-i}[1-q]^{i+j}$$

$$p^i[1-p]^{-i-j} f_1(u)\frac{\partial p}{\partial v} + k(n-k)\left[\frac{F_2(u)}{F_1(u)}\right]^{r-1}\left[\frac{1-F_1(u)}{1-F_2(u)}\right]^{r-2} q^{n-s}[1-q]^{k+s-r-n}$$

$$[1-p]^{n-k-1}\times\sum_{j=w_5}^{w_6}\sum_{i=t_5}^{t_6} A_3\left[\frac{F_1(u)}{F_2(u)}\right]^j\left[\frac{1-F_2(u)}{1-F_1(u)}\right]^j p^i[1-p]^{-i-j}q^{-i}[1-q]^{i+j} f_1(u)\frac{\partial q}{\partial v}$$

$$+k(n-k)p^{n-s}[1-p]^{s-k-r}[1-q]^{k-1}\sum_{j=w_7}^{w_8}\sum_{i=t_7}^{t_8} A_4\left[\frac{F_2(u)}{F_1(u)}\right]^j\left[\frac{1-F_1(u)}{1-F_2(u)}\right]^j \tag{19}$$

$$\times p^{-i}[1-p]^{i+j}q^i[1-q]^{-i-j} f_2(u)\frac{\partial p}{\partial v}\Bigg\} D^{-1}.$$

From the assumption, we know that U and V are independent. So $h_2(v)$ is independent of u and by using the lemma in Ahsanullah and Kabir (1973) $p = p(u,v) = g_1(v)$ and $q = q(u,v) = g_2(v)$ (we say functions of v only) and the remaining parts should be constant. Therefore

$$\begin{cases} 1-F_1(uv) = [1-F_1(u)]g_1(v), \theta \le u,\ 1<v,\ \theta>0, \\ 1-F_2(uv) = [1-F_2(u)]g_2(v), \beta\theta \le u,\ 1<v,\ \theta>0, \beta>1. \end{cases} \tag{20}$$

It is clear that these are version of Pexider's equation. So from Theorem 2.1 we can solve them. Since $F_1(x)$ and $F_2(x)$ are CDFs continuous for all $x \in [\theta,\infty)$ and $x \in [\beta\theta,\infty)$, respectively. We may conclude that

$$\begin{cases} 1-F_1(x) = c_1 x^{-\alpha}, \theta \le x,\ \theta>0, \\ 1-F_2(x) = c_2 x^{-\alpha}, \beta\theta \le x,\ \theta>0, \beta>1. \end{cases} \tag{21}$$

Where c_1, c_2 and α are constant.

After replacing these solutions in (19) and using some simplification we get

$$h_2(v) = \alpha v^{-\alpha(n-s+1)-1}[1-v^{-\alpha}]^{s-r-1} H[c_2 D]^{-1}, \tag{22}$$

where

$$H = k(k-1)c_2 \sum_{j=w_1}^{w_2}\sum_{i=t_1}^{t_2} A_1 \left[\frac{1-c_2u^{-\alpha}}{1-c_1u^{-\alpha}}\right]^j \left[\frac{c_1}{c_2}\right]^j + (n-k)(n-k-1)c_1$$
$$\times \sum_{j=w_3}^{w_4}\sum_{i=t_3}^{t_4} A_2 \left[\frac{1-c_2u^{-\alpha}}{1-c_1u^{-\alpha}}\right]^{r-1-j} \left[\frac{c_1}{c_2}\right]^{r-2-j} + k(n-k)c_1 \sum_{j=w_5}^{w_6}\sum_{i=t_5}^{t_6} A_3 \left[\frac{1-c_2u^{-\alpha}}{1-c_1u^{-\alpha}}\right]^{r-1-j} \left[\frac{c_1}{c_2}\right]^{r-2-j}$$
$$+ k(n-k)c_2 \sum_{i=t_7}^{t_8} A_4 \left[\frac{1-c_2u^{-\alpha}}{1-c_1u^{-\alpha}}\right]^j \left[\frac{c_1}{c_2}\right]^j$$

We know that $C(n, j) = 0$ if $j > n,$ then by using some elementary algebra $H[c_2 D]^{-1} = (n-r)C(n-r-1, n-s)$ and the right side of (22) is only depend on v and it is pdf of V.

Finally, from the property of CDF, $\alpha > 0$, $c_1 = \theta^\alpha$ and $c_2 = (\beta\theta)^\alpha$. Thus sufficiency is established and the proof is complete.

Theorem 3.2. Let X be a random variable with CDF $F(x) = bF_2(x) + \bar{b}F_1(x)$ such that $F_1(x)$ $(x \geq \theta)$ and $F_2(x)$ $(x \geq \beta\theta)$ are CDFs, where $b = \frac{k}{n}$, $\bar{b} = 1-b$, $\theta > 0$ and $\beta > 1$. If

$$E(X^\alpha \mid X > c) = bE_2\left(\frac{Xc}{\beta\theta}\right)^\alpha + \bar{b}E_1\left(\frac{Xc}{\theta}\right)^\alpha \tag{23}$$

holds for some $\alpha > 0$ then $F(x)$ is the Pareto distribution in the presence of outliers. We assume that $E(X^\alpha) < \infty$.

Proof. Proof is similar as given in Dallas (1976). In the process to prove the theorem, we should note that the solution of the differential equation $cP'(c) = -\gamma P(c)$ $(P(c) = 1 - F(c))$ is $P(c) = Ac^{-\alpha}$, where A is a constant, $\gamma = \alpha\delta/(\delta-1) > 0$ and

$$\delta = b\int_{\beta\theta}^{\infty}\left(\frac{X}{\beta\theta}\right)^\alpha dF_2(x) + \bar{b}\int_{\theta}^{\infty}\left(\frac{X}{\theta}\right)^\alpha dF_1(x). \tag{24}$$

Comparing the solution with the assumption imply that $A = b(\beta\theta)^\alpha + \bar{b}\,\theta^\alpha$ and the proof is complete.

4. AN ACTUAL EXAMPLE

Here, we have given an example of motor insurance company from Dixit and Jabbari Nooghabi (2011a). From the example, we know that the data follow the Pareto distribution in the presence of outliers. So by using Theorem 3.1, we can check the sufficiency. Assuming $r=3 \text{ and } s=12$, we have $x_{(r)}=63000$ and $\frac{x_{(s)}}{x_{(r)}}=2.857$. So, using the copula method and independent test by package 'copula' in **R**, the result is as follows:

Global Cramer-von Mises statistic: 0.03125 with p-value 0.9950495
Combined p-values from the Mobius decomposition:
0.9950495 from Fisher's rule,
0.9950495 from Tippett's rule.

Therefore, $X_{(r)}$ and $\frac{X_{(s)}}{X_{(r)}}$ are independent, because of the p-value is grater than 0.05, as significant level of the test. So, we can conclude that the data follow the Pareto distribution in the presence of outliers.

ACKNOWLEDGEMENTS

The authors are thankful to the two referees for their valuable comments and suggestions. Partial support from the Ordered and Spatial Data Center of Excellence of Ferdowsi University of Mashhad is acknowledged.

REFERENCES

Aczel, J. (1966). Lectures on Functional Equations and their Applications. Academic Press, New York.

Ahsanullah, M. and Kabir Lutful A. B. M. (1973). A Characterization of the Pareto Distribution. *The Canadian Journal of Statistics*, 1(1), 109–112.

Dallas, A. C. (1976). Characterizing the Pareto and power distributions. *Ann. Inst. Statist. Math.*, 28, 491–497.

Dixit, U. J. (1987). Characterization of the gamma distribution in the presence of k outliers. Bulletin Bombay Mathematical Colloquim, 4, 54–59.

Dixit, U. J. (1989). Estimation of parameters of the Gamma Distribution in the presence of Outliers. *Commun. Statist. Theory and Meth.*, 18, 3071–3085.

Dixit, U. J. (1994). Bayesian approach to prediction in the presence of outliers for

Weibull distribution, *Metrika*, 41, 127–136.
Dixit, U. J. and Jabbari Nooghabi, M. (2011a). Efficient estimation in the Pareto distribution with the presence of outliers, *Statistical Methodology*, 8(4), 340–355.
Dixit, U. J. and Jabbari Nooghabi, M. (2011b). Efficient Estimation of the parameters of the Pareto Distribution in the Presence of Outliers, *Communications of the Korean Statistical Society*, 18(6), 817
Kuczma M. (2008). An Introduction to the Theory of Functional Equations and Inequalities. Second edition, edited by Attila Gilányi, Birkhäuser Verlag AG, Berlin.
Sinha, S. K. (1973). Distributions of order statistics and estimation of mean life when an outlier may be present. *The Canadian Journal of Statistics*, 1(1), 119–121.

Received: 01.06.2013

Revised: 17.05.2014

U. J. Dixit[1] and M. Jabbari Nooghabi[2]
Department of Statistics[1]
University of Mumbai, Mumbai-India.
Department of Statistics[1],
Ferdowsi University of Mashhad,
Mashhad-Iran[2].
E-mail: *ulhasdixit@yahoo.co.in*,
jabbarinm@um.ac.ir;
jabbarinm@yahoo.com