International Journal of Engineering Research and General Science Volume 2, Issue 4, June-July, 2014
ISSN 2091-2730

Record Values from Size-Biased Pareto Distribution and a Characterization

Shakila Bashir[1], Munir Ahmad[2]

[1]Assistant Professor, Kinnaird College for Women, Lahore

[2]Professor, National college of Business Administration & Economics (NCBA&E), Lahore

E-mail- shakilabashir15@gmail.com

Abstract— In this paper upper record values from the size-biased Pareto distribution(S-BPD) are studied. Several distributional properties of upper record values from the size-biased Pareto distribution, including probability density function(pdf), cumulative distribution function(cdf), moments, entropy, inverse/negative moments, relations between negative and positive moments, median, mode, joint and conditional pdfs, conditional mean and variance, have been derived. The reliability measures of the upper record values from the S-BPD such as survival function, hazard rate function, cumulative hazard rate function and reversed hazard rate are also discussed. A characterization of the S-BPD based on the conditional expectation of record values is given.

Keywords— S-BPD; distribution function; moments; record values; hazard function; entropy; mgf; cdf; pdf; characterization.

1. INTRODUCTION

Chandler (1952) introduced records as sequence of random variables such that random variable at *i*–th place is larger (smaller) than variable at (*i*–1)*th*place. He called random variables as upper (lower) records in a random sample of size *n* from some probability distribution. After the introduction of the field, number of researcher jumped into this area of statistics. Shorrock (1973) has comprehensively discussed about record values and record time in a sequence of random variables. Ahsanullah (1979) has characterized the exponential distribution by using the record values. Ahsanullah (1991) has also derived the distributional properties of records by using the Lomax distribution. Some moment properties of the records have been given by Ahsanullah (1992).Balakrishnan and Ahsanullah (1994) have established some recurrence relations satisfied by the single and double moments of upper record values from the standard form of the generalized Pareto distribution. Ahsanullah (1997) derived some properties and a characterization of upper record values from the classical Pareto distribution. Sultan and Moshref (2000) have obtained the best linear unbiased estimates for the location and scale parameters of record values from the generalized Pareto distribution. Ahsanullah (2010) considered several distributional properties of the upper records from the exponential distribution. Based on these distributional properties, some characterizations of the exponential distribution are also presented in this paper. Ahsanullah et al. (2013) discussed a new characterization of power function distribution based on lower record values.The pdf $f_n(x)$ of upper record values $X_{U(n)}$ is

$$f_n(x)=\frac{(R(x))^{n-1}}{\Gamma n} f(x), \qquad -\infty < x < \infty. \tag{1.1}$$

Thejoint pdf of $X_{U(j)}$ and $X_{U(i)}$ is

$$f_{i,j}(x,y)=\frac{[R(x)]^{i-1}}{\Gamma i} r(x)\frac{[R(y)-R(x)]^{j-i-1}}{\Gamma(j-i)} f(y), \qquad -\infty < x < y < \infty. \tag{1.2}$$

$$\therefore\ j > i$$

The conditional pdf of $X_{U(j)}/X_{U(i)} = x_i$ is

$$f\left(X_{U(j)}=y_j/X_{U(i)}=x_i\right)=\frac{(R(y)-R(x))^{j-i-1}}{(j-i-1)!}\frac{f(y)}{1-F(x)}, \qquad -\infty < x < y < \infty. \tag{1.3}$$

For $j=i+1$

IJERGS

$$f\left(y_{i+1}/X_{U(i)}=x_i\right)=\frac{f\left(y_{i+1}\right)}{1-F\left(x_i\right)}, \qquad -\infty<x_i<y_{i+1}<\infty. \tag{1.4}$$

1.1 SIZE-BIASED PARETO DISTRIBUTION

When an investigator records an observation by nature according to a certain stochastic model the recorded observation will not have the original distribution unless every observation is given an equal chance of being recorded. Patil and Rao (1978) examined some general models leading to weighted distributions with weight functions not essentially restricted by unity. The results were applied to the analysis of data relating to human populations and wildlife management. Sunoj and Maya (2006) introduced some fundamental relationships between weighted and
unique variables in the context of maintainability function and inverted repair rate. Furthermore, some characterization theorems for specific models such as power, exponential, Pareto II, beta, and Pearson system of distributions using the relationships between the original and weighted random variables was also established. Mir and Ahmad (2009) introduced some size-biased probability distributions and their generalizations. These distributions offer a joining approach for the problems where the observations fall in the non-experimental, non- replicated, and nonrandom categories.They introduced some of the possible uses of size- biased distribution theory to some real life data. A number of papers have been appeared during the last ten years implicitly using the concepts of weighted and size-biased sampling distributions.

The probability density function of weighted Pareto distribution is obtained by applying the weight $w(X)=x^m$

$$f(x)=(\beta-m)\alpha^{(\beta-m)}x^{m-\beta-1}, \qquad \alpha>0, \alpha<x<\infty. \tag{1.5}$$

Where $m=1$ or 2, these special cases are named as size-biased or length- biased distribution and area-biased distribution, respectively. We define the Size-Biased Pareto distribution, when $w(X)=x$. The probability density function of the size-biased Pareto Distribution is

$$f(x)=(\beta-1)\alpha^{\beta-1}x^{-\beta}, \qquad \alpha>0, \alpha<x<\infty. \tag{1.6}$$

The cumulative distribution function of moment Pareto distribution is

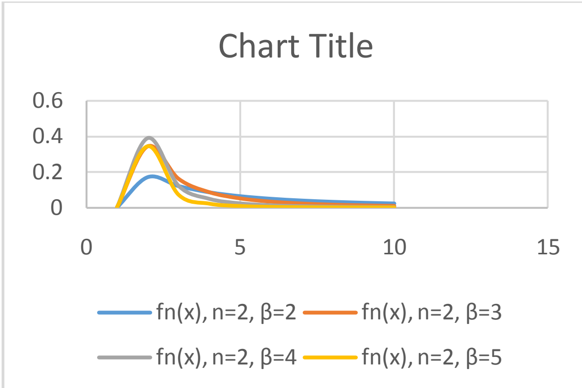
$$F(x)=1-\alpha^{\beta-1}x^{1-\beta}. \tag{1.7}$$

In this paper, upper record values from S-BPD have been derived and discussed various properties including characterization. Previously, no research work has been done on weighted distributions in the context of record values. So it is hoped that findings of this paper will be useful for researchers in different fields of applied sciences.

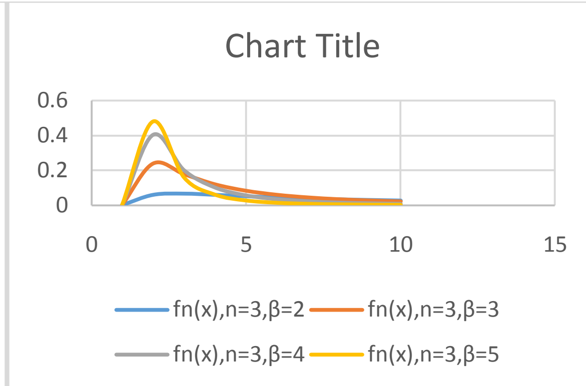
2. UPPER RECORD VALUES FROM SIZED-BIASED PARETO DISTRIBUTION (S-BPD)

Let $X_{U(1)}, X_{U(2)}, \cdots, X_{U(n)}$ denote the upper record values arising from the iid size-biased Pareto variables, then using equations (1.6) and (1.7), the probability density function of the nth upper record $X_{U(n)}$ is given by

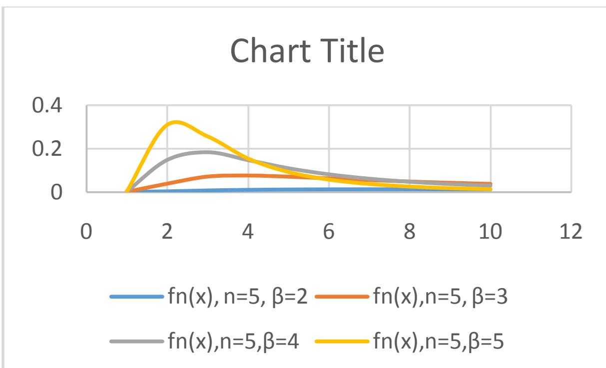
$$f_n(x)=\frac{(\beta-1)\alpha^{\beta-1}}{\Gamma(n)}x^{-\beta}\left(-\ln\left(\alpha^{\beta-1}x^{1-\beta}\right)\right)^{n-1}, \qquad \alpha>0, \quad \alpha<x<\infty. \tag{2.1}$$



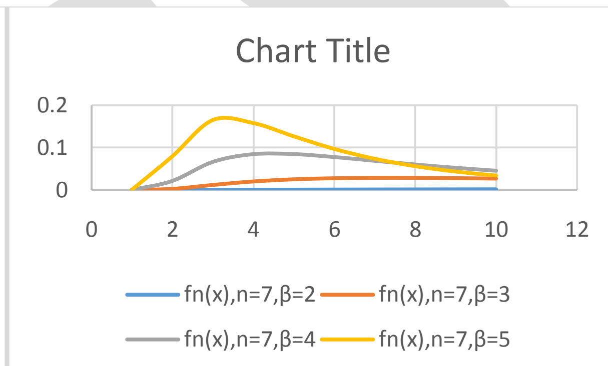
(a) $\alpha = 1, n = 2, \beta = 2,3,4,5$



(b) $\alpha = 1, n = 3, \beta = 2,3,4,5$



(c) $\alpha = 1, n = 5, \beta = 2,3,4,5$



(d) $\alpha = 1, n = 7, \beta = 2,3,4,5$

Fig. 1 pdf plots for upper record values from S-BPD

2.1 PROPERTIES

In this section some distributional properties of the upper record values from the S-BPD have been derived.

2.1.1 MOMENTS

The rth moment of the nth upper record value $X_{U(n)}$ by using (2.1), are

$$\mu'_{r(n)} = E\left(X_{(n)}^{r}\right) = \frac{(1-\beta)^n \alpha^r}{(1-\beta+r)^n}. \tag{2.2}$$

The mean and variance of the upper record values from the S-BPD are, respectively

$$\text{Mean} = \frac{(1-\beta)^n \alpha}{(2-\beta)^n}. \tag{2.3}$$

$$\text{Variance} = \frac{(1-\beta)^n \alpha^2}{(3-\beta)^n (2-\beta)^{2n}} \left[(2-\beta)^{2n} - (3-\beta)^n (1-\beta)^n\right]. \tag{2.4}$$

The mode of the upper record values from size-biased Pareto distribution is

$$x_{\text{mode}(n)} = \alpha \exp\left(\frac{-(n-1)}{\beta(1-\beta)}\right). \tag{2.5}$$

The inverse/negative moments of nth upper record values from the size-biased Pareto distribution are

$$\mu'_{-r(n)} = E\left(\frac{1}{X_{(n)}}\right)^r = \frac{(1-\beta)^n}{(1-\beta-r)^n \alpha^r}. \tag{2.6}$$

By using the equation (2.2) and (2.6), the relation between negative and positive moments of the nth upper record values from S-BPD is

$$\mu'_{-r(n)} = \left(\frac{1-\beta+r}{1-\beta-r}\right)^n \alpha^{-2r} \mu'_{r(n)}. \tag{2.7}$$

The moment generating function of the upper record values from size-biased Pareto distribution is

$$M_{x(n)}(t) = (1-\beta)^n \sum_{k=0}^{\infty} \frac{(t\alpha)^k}{k!(1-\beta+k)^n}. \tag{2.8}$$

2.1.2 ENTROPY

The entropy of the nth record values $X_{U(n)}$ from the S-BPD is

$$H(x) = -\ln\Gamma n - \frac{(n-1)}{\Gamma n}\sum_{k=1}^{\infty}\frac{(-1)^{k+1}}{k}\sum_{i=0}^{\infty}\binom{k}{i}(-1)^i\,\Gamma(n+k-i) - \ln\left[(\beta-1)\alpha^{\beta-1}\right] + \beta\ln\alpha - \frac{n\beta}{(1-\beta)}. \tag{2.9}$$

2.1.3 CUMULATIVE DISTRIBUTION FUNCTION

The cumulative distribution function of the upper record values from the S-BPD is

$$F_n(x) = 1 - \frac{\Gamma\left(n, \left(-\ln\left(\alpha^{\beta-1}x^{1-\beta}\right)\right)\right)}{\Gamma n}. \tag{2.10}$$

where $\Gamma(a,s) = \int_s^{\infty} x^{a-1}e^{-x}dx$ is the upper incomplete gamma function.

2.1.4 SURVIVAL AND HAZARD RATE FUNCTION

The survival function of upper record values from the S-BPD is

$$S_n(x)=\frac{\Gamma\left(n,\left(-\ln\left(\alpha^{\beta-1}x^{1-\beta}\right)\right)\right)}{\Gamma n}\ . \tag{2.11}$$

The Hazard rate function is

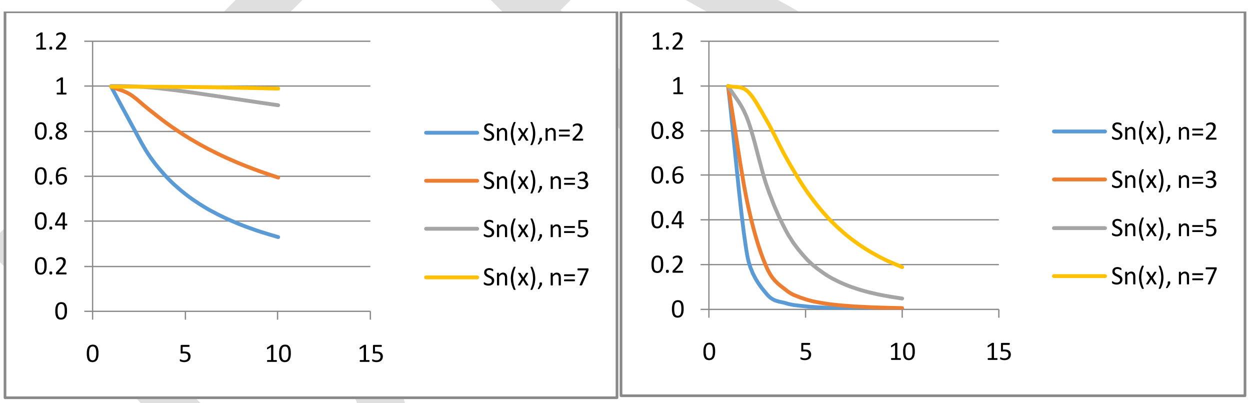
$$h_n(x)=\frac{(\beta-1)\alpha^{\beta-1}x^{-\beta}\left(-\ln\left(\alpha^{\beta-1}x^{1-\beta}\right)\right)^{n-1}}{\Gamma\left(n,\left(-\ln\left(\alpha^{\beta-1}x^{1-\beta}\right)\right)\right)}. \tag{2.12}$$

The cumulative hazard rate function is

$$H_n(x)=-\ln\left(\frac{\Gamma\left(n,\left(-\ln\left(\alpha^{\beta-1}x^{1-\beta}\right)\right)\right)}{\Gamma n}\right). \tag{2.13}$$

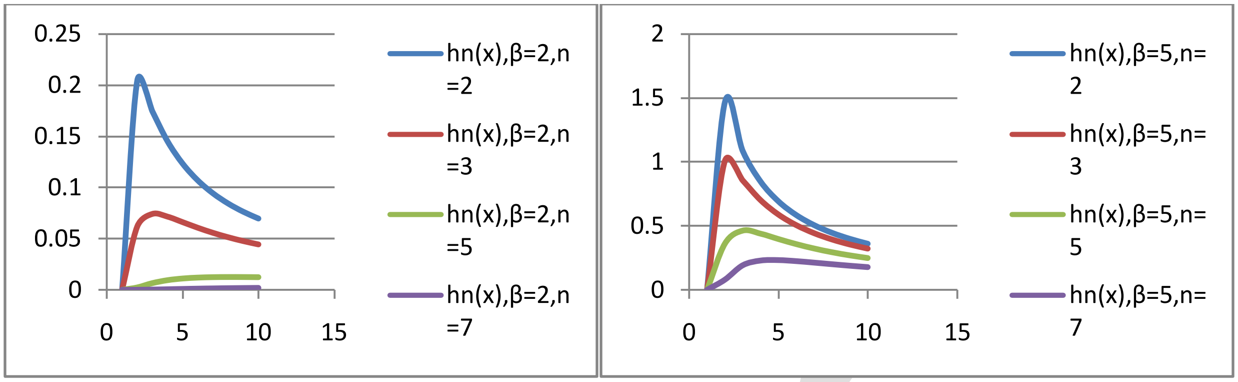
The reverse hazard rate function is

$$r_n(x)=\frac{(\beta-1)\alpha^{\beta-1}x^{-\beta}\left(-\ln\left(\alpha^{\beta-1}x^{1-\beta}\right)\right)^{n-1}}{\left[\Gamma n-\Gamma\left(n,\left(-\ln\left(\alpha^{\beta-1}x^{1-\beta}\right)\right)\right)\right]}\ . \tag{2.14}$$



(a) $\alpha=1,\beta=2,n=2,3,5,7$ (b) $\alpha=1,\beta=5,n=2,3,5,7$

Fig.2 Survival Plots for upper record values from S-BPD.



(a) $\alpha = 1, \beta = 2, n = 2,3,5,7$ (b) $\alpha = 1, \beta = 5, n = 2,3,5,7$

Fig.3 Hazard rate plots for upper record values from S-BPD.

3. JOINT AND CONDITIONAL DENSITY FUNCTIONS

The joint probability density function of $X_{U(i)}$ and $X_{U(j)}$ from S-BPD is

$$f_{i,j}(x,y) = \frac{(\beta-1)^2 \alpha^{\beta-1}}{\Gamma i \Gamma(j-i)} . \frac{y^{-\beta}}{x} \left(-\ln\left(\alpha^{\beta-1} x^{1-\beta}\right)\right)^{i-1} \left[\ln\left(\alpha^{\beta-1} x^{1-\beta}\right) + \ln\left(\alpha^{\beta-1} y^{1-\beta}\right)\right]^{j-i-1}, \quad \alpha < x < y < \infty. \because j > i \tag{3.1}$$

Thus the conditional pdf $f_{j/i}(y/x)$ of $X_{U(j)}/X_{U(i)} = x$ is given by

$$f_{j/i}(y/x) = \frac{(\beta-1)}{\Gamma(j-i)} \left(\frac{y}{x}\right)^{-\beta} \frac{1}{x} \left(-\ln\left(\frac{y}{x}\right)^{1-\beta}\right)^{j-i-1}, \qquad \alpha < x < y < \infty. \tag{3.2}$$

The mean of the conditional pdf $f_{j/i}(y/x)$ is

$$E\left(X_{U(j)}/X_{U(i)}\right) = E_{j/i}(y/x) = \frac{(1-\beta)^{j-i} x}{(2-\beta)^{j-i}} . \tag{3.3}$$

Variance of the conditional pdf $f_{j/i}(y/x)$ is

$$\operatorname{var}\left(X_{U(j)}/X_{U(i)}\right) = \operatorname{var}_{j/i}(y/x) = x^2 (1-\beta)^{j-i} \left[\frac{1}{(3-\beta)^{j-i}} - \frac{(1-\beta)^{j-i}}{(2-\beta)^{2(j-i)}}\right]. \tag{3.4}$$

4. CHARACTERIZATION

Using conditional pdf of $X_{L(n+1)}$ given $X_{L(n)}$ as given in equation (1.4), it can be shown that if $X \in P(\alpha, \beta)$, then

$$E\left[\left\{X_{L(n+1)} - X_{L(n)}\right\}^2 / X_{L(n)} = x\right] = \frac{2x^2}{(\beta-2)(\beta-3)}, \qquad \beta > 3.$$

The following theorem gives a characterization of the S-PBD using the above result.

Theorem 4.1

Let $\{X_n, \geq 1\}$ be iid random variables having absolutely continuous (with respect to Lebesgue measure) cdf $F(x)$. We will assume without loss of generality $F(0)=0$ and $F(1)=1$.We assume further that $F(x)$is twice differential and let $E(X_n^2) < \infty, n \geq 1$. If for $\beta > 3$,

$$E\left[\left\{X_{L(n+1)} - X_{L(n)}\right\}^2 / X_{L(n)} = x\right] = \frac{2x^2}{(\beta-2)(\beta-3)}, \qquad \beta > 3 \ . \tag{4.1}$$

and $X \in P(\alpha, \beta)$.

Proof. Using Equation (1.4), we get

$$\int_x^\infty (y-x)^2 f(y)dy = \frac{2x^2}{(\beta-2)(\beta-3)} \bar{F}(x) \ . \tag{4.2}$$

Differentiating both sides of equation (4.2), we get

$$\int_x^\infty -2(y-x)f(y)dy = \frac{4x\bar{F}(x)}{(\beta-2)(\beta-3)} - \frac{2x^2 f(x)}{(\beta-2)(\beta-3)} \ . \tag{4.3}$$

Now taking 2nd derivative of both sides of equation (4.3), we have

$$2\bar{F}(x) = \frac{4\bar{F}(x)}{(\beta-2)(\beta-3)} - \frac{8xf(x)}{(\beta-2)(\beta-3)} - \frac{2x^2 f'(x)}{(\beta-2)(\beta-3)}. \tag{4.4}$$

Substituting $y = \bar{F}(x), y' = -f(x), y'' = -f'(x)$, the equation (4.4) reduces to

$$x^2 y'' + 4xy' - (\beta^2 - 5\beta + 4)y = 0 \tag{4.5}$$

The equation (4.5) is the well-known Euler type equation. It has solution of the form $y = x^r$, where r must satisfy the equation

$$r(r-1) + 4r - (\beta^2 - 5\beta + 4) = 0$$

$$r^2 + 3r - (\beta^2 - 5\beta + 4) = 0$$

The roots of r , from the above equation are $r = 1 - \beta$ and $r = \beta - 4$. So the solutions are of the type

$$y = c_1 x^{1-\beta} \text{ and } y = c_2 x^{\beta-4} . \tag{4.6}$$

Where c_1 & c_2 are constants. Hence we assume $E(X^2)$ exists and $y = \bar{F}(x) = 1 - F(x)$, so we have

$$\lim_{x\to\infty} \bar{F}(x) = 0, \qquad \lim_{x\to\infty} x\bar{F}(x) = 0 \tag{4.7}$$

The solution $y = c_1 x^{1-\beta}$ satisfies the condition of equation (4.7) if $\beta > 1$, which contradicts the assumption that $\beta < 1$, and $y = c_2 x^{\beta-4}$ satisfies the condition of equation (4.7) if $\beta < 4$, which contradicts the assumption if $\beta > 4$. We must have

$$F(x) = 1 - \alpha^{\beta-1} x^{1-\beta}, \quad 0 < \alpha < x < \infty, \quad \beta > 0.$$

5. CONCLUSION

In this paper, we developed the distribution of upper record values from the size-biased Pareto distribution. The graphs show that the distribution of upper record values is positively skewed. For large values of n and β the pdf showing peaked and right tail longer while for smaller values of n and β the pdf is flattering. We derive the positive and negative moment of the upper record values from the size-biased Pareto distribution and developed a relation between them. The associated cdf, survival function, hazard function, entropy, mgf, median, mode, skewness and kurtosis have been derived. We derive the joint and conditional probability distribution functions of ith and jth upper record values from the size-biased Pareto distribution and find out conditional mean and variance of it. The cumulative hazard rate function and reverse hazard rate function for the record values from size-biased Pareto distribution have also been derived. The plot of survival function shows that the survival function is increasing for small n & β and decreasing and showing bathtub shape for large β. Plot of hazard function shows increasing trend with n = 2 while decreasing function when n increase. We hope this paper will contribute a valuable contribution for the enhancement of research in the theory of record values.

REFERENCES:

[1] Adamic, L.A. (2002). Zipf, Power-laws and Pareto-a ranking tutorial. Internet Ecologies area, Xerox Palo Alto Research Center, Palo Alto, CA 94304 (http://ginger.hpl.hp.com/shl/papers/ranking.html)

[2] Ahsanullah, M. (1979). Characterization of exponential distribution by record values, Sankhya, Vol. 41, 116–121.

[3] Ahsanullah, M. (1988). Introduction to Record Values. Ginn Press, Needham Heights, Massachusetts.

[4] Ahsanullah, M. (1991). Record values of Lomax distribution, Statisti. Nederlandica, Vol. 41(1), 21–29.

[5] Ahsanullah, M. (1992). Record values of independent and identically distributed continuous random variables, Pak. J. Statist. Vol. 8(2), 9–34.

[6] Ahsanullah, M. (1995). Record Statistics, Nova Science Publishers, USA.

[7] Ahsanullah, M (1997). On the record values of the classical Pareto distribution. Pak. J. Statist., Vol, 13(1), 9-15.

[8] Ahsanullah, M. (2010). Concomitants of Upper Record Statistics for Bivariate Pseudo–Weibull Distribution, J. Appl. Math. Vol. 5(10), 1379-1388.

[9] Ahsanullah, M. (2010). Some characterizations of exponential distribution by upper record values, Pak. J. Statist, Vol. 26(1), 69-75.

[10] Ahsanullah, M., Shakila, M., and GolamKibria, B. M. (2013). A characterization of power function distribution based on lower record values. ProbStat Forum, Vol, 6, 68-72.

[11] Arnold, B.C., Balakrishnan, N., and Nagaraja, H.N. (1992). A First Course in Order Statistics, John Wiley and Sons, New York.

[12] Balakrishnan, N. and Balasubramanian, K. (1995). A characterization of geometric distribution based on record values, J. Appl. Statist. Science, Vol. 2(1), 73–87.

[13] Candler, K.N. (1952). The distribution and frequency of record values, J. Roy. Statist. Soc., Vol.14, 220–228.

[14] Gustafson, G. & Fransson, A. (2005). The use of the Pareto distribution for fracture transmissivity assessment, Hydrogeology Journal, Vol. 14, 15-20.

[15] Johnson, N.L., Kotz, S., and Balakrishnan, N. (1995). Continuous Univariate Distributions, Vol. 2, Second edition, John Wiley & Sons, New York.

[16] Mir, K. A., and Ahmad, M. (2009). Size-biased distributions and their applications, Pak. J. Statist, Vol. 25(3), 283-294.

[17] Patil, G. P., and Rao, C. R. (1978). Weighted Distributions and Size-Biased Sampling with Applications to Wild life Populations and Human Families, Biometrics, Vol. 34, 179-189.

[18] Sultan, K. S., and Moshref, M. E. (2000). Record values from generalized Pareto distribution and associated inference, Metrika, Vol. 51, 105-116.

[19] Sultan, K.S. (2007). Record Values from the Modified Weibull Distribution and Applications, International Mathematical Forum, Vol. 2,(41), 2045 – 2054.

Sunoj, S. M., and Maya, S. S. (2006). Some Properties of Weighted Distributions in the Context of Repairable Systems, Communications in Statistics—Theory and Methods, Vol. 35, 223–228