STATISTICA, anno LXXI, n. 3, 2011

RECORD VALUES FROM A FAMILY OF J-SHAPED DISTRIBUTIONS

Ahmad A. Zghoul

1. INTRODUCTION

A class of distributions called a family of J-shaped distributions was introduced by (Topp and Leone, 1955). The distribution function form of a J-shaped distribution with scale parameter β and shape parameter θ is

$$F(x)=\begin{cases}0 & ;x<0\\ \left(\frac{x}{\beta}\left(2-\frac{x}{\beta}\right)\right)^{\theta} & ;0\le x\le\beta\\ 1 & ;\beta<x\end{cases}$$

and hence the density has the form

$$f(x)=\frac{2\theta}{\beta}\left(1-\frac{x}{\beta}\right)\left(\frac{x}{\beta}\left(2-\frac{x}{\beta}\right)\right)^{\theta-1}\quad ;0<x<\beta,$$

with $\beta>0$ and $\theta>0$.

The appropriateness of the distribution for modeling life data was emphasized by (Nadarajah and Kotz, 2003). They highlighted the fact that its hazard rate function is bathtub shaped. Further they derived general formulas for the moments and the central moments of the distribution and provided the maximum likelihood estimator of the shape parameter θ when the scale β is known. The distribution of ordered statistics along with their moments and product moments were studied by (Zghoul, 2010).

The bathtub or U-shaped hazard rate functions have many applications in life time modeling. For example, in human populations the death rate is high due to birth defects or infant diseases at infant age, and then it remains constant up to the age of thirties where it increases again. This pattern is common in manufactured items. Most parametric models having U-shaped hazard rate function usually involve three or more parameters which in turns raise substantial problems in statistical inference unless large samples are available. An advantage of the J-shaped family of distributions, which has a U-shaped hazard rate function, is that it has only two parameters; namely the scale parameter β and the shape parameter θ.

We assume through this work, without loss of generality, that $\beta = 1$, in which case the distribution and the density functions are, respectively, reduced to

$$F(x) = \begin{cases} 0 & ; x < 0 \\ (x(2-x))^{\theta} & ; 0 \le x \le 1; 0 < \theta < 1 \\ 1 & ; 1 < x \end{cases} \tag{1}$$

and

$$f(x) = 2\theta(1-x)(x(2-x))^{\theta-1}; 0 < x < 1; 0 < \theta < 1 . \tag{2}$$

In section 2 of this paper we study the distribution of record values from this J-shaped family of distributions. Then, in section 3, moments, product moments, and recurrence relations are obtained. Also, bounds based on Jensens' inequality for the mean of record values are provided. In section 4, the maximum likelihood estimator for the shape parameter θ based on lower record values is derived and its properties are studied. A real life application is investigated in section 5. Finally, the conclusions and suggestions for further studies is given in section 6.

2. DISTRIBUTION OF RECORD VALUES

Let $X_1, X_2, \ldots$ be a sequence of independent and identically distributed random variables. Let $L(0) = 0$, $L(1) = 1$ and $L(n) = \min\{j : X_j > X_{L(n-1)}\}; n = 2, 3, \ldots$, then $R_n = X_{L(n)}, n \ge 1$ is the sequence of upper record values. If the inequality sign is reversed then $R'_n = X_{L(n)}$ is the sequence of lower record values.

The theory of record values, record times, and inter-record times were first introduced by (Chandler, 1952). Since then many research papers have been published in the subject. Detailed review of records and related bibliography can be found in (Nevzorov, 1987), (Nagaraja, 1988), (Ahsanullah, 1995), (Balakrishnan and Nagaraja, 1998), (Nevzorov and Balakrishnan, 1998), (Arnold *et al.*, 1998), to mention some.

Let $R_1, R_2, \ldots$ be a sequence of upper record values from the distribution (1), then the density of the nth record value is given by

$$f_{R_n}(r) = f(r)\frac{[-\log(1-F(r))]^n}{n!}$$

$$= \theta u'(r)(u(r))^{\theta-1}\frac{[-\log(1-(u(r))^{\theta})]^n}{n!}; 0 < r < 1,$$

where $u(r) = r(2-r)$ and hence $u'(r) = 2(1-r)$.

The joint probability distribution function (pdf) of the set of upper records $R_1, R_2, ..., R_n$ is

$$f_{R_1,...,R_n}(r_1,...,r_n) = \prod_{i=1}^{n-1}\frac{f(r_i)}{1-F(r_i)}f(r_n)$$

$$= \theta^n \prod_{i=1}^{n-1}\frac{u'(r_i)(u(r_i))^{\theta-1}}{1-(u(r_i))^{\theta}} u'(r_n)(u(r_n))^{\theta-1}.$$

3. MOMENTS AND PRODUCT MOMENTS OF RECORD VALUES

In this section the moments and product moments of record values are derived. Recurrence relations for the moments are given. Moreover, we derive a lower bound and an upper bound for the mean of upper records.

Theorem 1: For k=1,2,…, and n=1,2,…, the kth moment of the nth lower record value is given by:

$$\alpha_n'^{(k)} = \sum_j^{\infty} d(k,j)(1+i)^{-(n+1)},$$

where $d(k,0) = 2^{-k}$, $\mathrm{d}(k,1) = 2^{-k-2}k$, and

$$d(k,j) = \frac{2^{-k-2j}k}{j!}\frac{(k+2j-1)!}{(k+j)!}, \text{ for } j = 2,3,\ldots. \tag{3}$$

Proof: $\alpha_n'^{(k)} \equiv E(\mathrm{R}_n'^{k}) = \int_0^1 r^k \theta u'(r)(u(r))^{\theta-1}\frac{[-\log(u^{\theta}(r))]^n}{n!}dr$.

Put $s = u(r)$, then $r = u^{-1}(s) = 1-\sqrt{1-s}$, hence

$$\alpha_n'^{(k)} = \theta^{n+1}\int_0^1 (1-\sqrt{1-s})^k s^{\theta-1}\frac{[-\log(s)]^n}{n!}ds. \tag{4}$$

A power series expansion for $(1-\sqrt{1-x})^k$ given by (Gradstein and Ryzhik, 1980), page 21, is:

$$(1+\sqrt{1+x})^k = 2^k\left(1+\frac{k}{1!}\left(\frac{x}{4}\right)+\frac{k(k-3)}{2!}\left(\frac{x}{4}\right)^2+\frac{k(k-4)(k-5)}{3!}\left(\frac{x}{4}\right)^3 + \frac{k(k-5)(k-6)(k-7)}{4!}\left(\frac{x}{4}\right)^4+\dots\right),$$

for any real number k and $|x|<1$.

Which, alternatively can be written as:

$$(1+\sqrt{1+x})^k = \sum_{j=0}^{\infty} c(k,j)x^j ,$$

where $c(k,0)=2^k$, $c(k,1)=2^{k-2}k$, and $c(k,j)=\frac{2^{k-2j}k}{j!}\prod_{i=1}^{j-1}(k-j-i)$, for $j \geq 2$.

Since $(1+\sqrt{1-x})^k(1-\sqrt{1-x})^k = x^k$, we have

$$\begin{aligned}(1-\sqrt{1-x})^k &= x^k(1+\sqrt{1-x})^{-k} \\ &= x^k\sum_{j=0}^{\infty} c(-k,j)(-x)^j \\ &= \sum_{j=0}^{\infty} d(k,j)x^{k+j} ,\end{aligned} \tag{5}$$

where $d(k,0)=2^{-k}$, $d(k,1)=2^{-k-2}k$, and for $j \geq 2$

$$\begin{aligned}d(k,j) &= \frac{2^{-k-2j}k}{j!}\prod_{i=1}^{j-1}(k+j+i) \\ &= \frac{2^{-k-2j}k}{j!}\frac{(k+2j-1)!}{(k+j)!} \\ &= \frac{2^{-k-2j}k}{j(k+j)}B(k+j,j) .\end{aligned}$$

Therefore the r.h.s. of (4) expands to:

$$\sum_{j=0}^{\infty} d(k,j)\ \theta^{n+1}\int_0^1 s^{(k+j)+\theta-1}\frac{[-\log(s)]^n}{n!}ds\,.$$

Set $u=-\log(s)$ to get:

$$\alpha_n'^{(k)}=\sum_{j=0}^{\infty} d(k,j)\ \theta^{n+1}\int_0^1 e^{-(k+j+\theta)u}\frac{u^n}{n!}du$$

$$=\sum_{j=0}^{\infty} d(k,j)\left(\frac{\theta}{\theta+k+j}\right)^{n+1}.$$

In particular,

$$E[R_n']=\sum_{j=0}^{\infty} d(1,j)\left(\frac{\theta}{\theta+j+1}\right)^{n+1}$$

$$=\frac{1}{2}\left[\left(\frac{\theta}{\theta+1}\right)^{n+1}+\frac{1}{4}\left(\frac{\theta}{\theta+2}\right)^{n+1}+\frac{1}{8}\left(\frac{\theta}{\theta+3}\right)^{n+1}+\ldots\right].$$

Let $R_1', R_2', \ldots$ be a sequence of lower records from the distribution in (1), and let $\alpha_1'(k), \alpha_2'(k), \ldots$ be the corresponding sequence of the kth moments of $R_1', R_2', \ldots$, then we prove the following recurrence relation.

Theorem 2: For k=1,2,…, and n=1,2,…, the kth moments of the nth upper record value is given by:

$$\alpha_n^{(k)}=\sum_j^{\infty}\sum_i^{\infty} d(j,k)(-1)^i\binom{(k+j)/\theta}{i}(1+i)^{-(n+1)}.$$

where $d(j,k)$ are as defined in (5).

Proof: $\alpha_n^{(k)}\equiv E(R_n^k)=\int_0^1 r^k\theta u'(r)(u(r))^{\theta-1}\frac{[-\log(1-u^{\theta}(r))]^n}{n!}dr\,.$

Put $s=u(r)$, then $r=u^{-1}(s)=1-\sqrt{1-s}$, and

$$\alpha_n^{(k)}=\theta\int_0^1(1-\sqrt{1-s})^k s^{\theta-1}\frac{[-\log(1-s^{\theta})]^n}{n!}ds\,.$$

Expanding as in (5) and following steps similar to those in the proof of Theorem 1, the theorem is proved.

Theorem 3: For n=1,2,… and k=0,1,2,…

$$\alpha_n'^{(k+1)} = \frac{2(k+1)(k+\theta)}{k(k+2\theta+1)}\alpha_n'^{(k)} + \frac{2\theta}{(k+2\theta+1)}\alpha_{n-1}'^{(k+1)} - \frac{2\theta(k+1)}{k(k+2\theta+1)}\alpha_{n-1}'^{(k)} .$$

Proof: We have

$$2\alpha_n'^{(k)} - \alpha_n'^{(k+1)} = \frac{1}{n!}\int_0^1 x^k (2-x)(-\log(F(x))^n f(x)dx . \tag{6}$$

From (1) and (2) it is readily seen that $x\,(2–x)\,f(x) = 2\theta\,(1–x)\,F(x)$, so (6) becomes

$$2\alpha_n'^{(k)} - \alpha_n'^{(k+1)} = \frac{2\theta}{n!}\int_0^1 x^{k-1}(1-x)(-\log(F(x))^n F(x)dx . \tag{7}$$

Integrating by parts treating $x^{k-1}(1-x)$ for integration and the rest of the integrand for differentiation, the r.h.s. of (7) turns out to be

$$2\theta\int_0^1 (\frac{x^k}{k} - \frac{x^{k+1}}{k+1})\left\{\frac{1}{(n-1)!}(-\log(F(x))^{n-1} f(x) - \frac{1}{n!}(-\log(F(x))^n f(x)\right\}dx,$$

hence

$$2\alpha_n'^{(k)} - \alpha_n'^{(k+1)} = 2\theta\left(\frac{\alpha_{n-1}'^{(k)}}{k} - \frac{\alpha_{n-1}'^{(k+1)}}{k+1}\right) - 2\theta\left(\frac{\alpha_n'^{(k)}}{k} - \frac{\alpha_n'^{(k+1)}}{k+1}\right).$$

Arranging terms in the above equation, the theorem is proved.

The following theorem gives upper and lower bounds for the mean of R_n.

Theorem 4: For n =1, 2, …, and $0 < \theta < 1$*, we have* $l_n \le \alpha_n \le u_n$*, where*

$$l_n = 1 - \sqrt{1-(1-2^{-(n+1)})^{1/\theta}} \;\; \textit{and} \;\; u_n = 1-(2/3)^{n+1} .$$

Proof: As $g(x) = [x(2-x)]^\theta$ is concave, by Jensen's inequality we have

$$g(E(R_n)) \ge E(g(R_n)) ,$$

which implies that

$$[\alpha_n(2-\alpha_n)]^\theta \ge E[F(R_n)] = 1 - 2^{-(n+1)} .$$

Writing $[\alpha_n(2-\alpha_n)]^\theta$ as $[1-(1-\alpha_n)^2]^\theta$ and manipulating, the lower bound is obtained.

To derive the upper bound, we have

$$\alpha_n = \int_0^1 r f_{R_n}(r) dr$$

$$= \int_0^1 r f(r) \frac{[-\log(1-F(r))]^n}{n!} dr .$$

Use the transformation $u = -\log(1-F(r))$ to get

$$\alpha_n = \int_0^\infty \frac{u^n}{n!} [1 - \sqrt{1-(1-e^{-u})^{1/\theta}}] e^{-u} du .$$

For $0 < \theta < 1$, we have $(1-e^{-u})^{1/\theta} < 1 - e^{-u}$, which implies that

$$\alpha_n < \int_0^\infty \frac{u^n}{n!} (1 - e^{-u/2}) e^{-u} du = 1 - (2/3)^{n+1} .$$

Upon computing lower bounds for the first few records we found that they get tighter as either n or θ increases.

To derive the product moments of record values we first derive their joint distribution. The joint pdf of the *mth* and *nth* record values is given by:

$$f_{R_m, R_n}(r_m, r_n) = 4\theta^2 \frac{[-\log(1-u^\theta(r_m))]}{m!} \frac{\left[-\log \frac{1-u^\theta(r_n)}{1-u^\theta(r_m)}\right]^{n-m-1}}{(n-m-1)!}$$

$$\times \frac{u'(r_m)u'(r_n)u^{\theta-1}(r_m)}{1-u^\theta(r_m)}, 0 < r_m < r_n < 1.$$

Based on the expansion in (5) the joint moments of $R_m R_n$ are given in the following theorem.

Theorem 5: For k=1,2,…, l=1,2,…, and for m=1,2,…and n=1,2,…, the joint moments of $R_m R_n$ *are*

$$\alpha_{m,n}^{(k),(l)} = \sum_{i_1} \sum_{i_2} \sum_{i_3} \sum_{i_4} (-1)^{i_3+i_4} c_{i_1} c_{i_2} \binom{(l+i_2)/\theta}{i_3} \binom{(k+i_1)/\theta}{i_4} (1+i_3)^{-(n-m)} (1+i_3+i_4)^{-(m+1)},$$

where c_i are as defined in (5).

Proof: analogous to the proof of Theorem 1.

4. ESTIMATION BASED ON LOWER RECORD VALUES

The joint pdf of the set of lower records $R'_1, R'_2, ..., R'_n$ is

$$f_{R'_1,...,R'_n}(r'_1,...,r'_n) = \prod_{i=1}^{n-1} \frac{f(r'_i)}{F(r'_i)} f(r'_n)$$

$$= \theta^n \prod_{i=1}^{n-1} \frac{u'(r'_i)}{u(r'_i)} u'(r'_n)(u(r'_n))^{\theta-1}.$$

Thus the log likelihood function is given by:

$$\log L(r'_1,...,r'_n;\theta) \propto n \log \theta + (\theta - 1) \log u(r'_n).$$

From which the maximum likelihood estimator (MLE) of θ is found to be

$$\hat{\theta} = -n(\log u(R'_n))^{-1}. \tag{8}$$

To show that $\hat{\theta}$ is an unbiased and consistent estimator for θ, we have

$$E(\hat{\theta}) = n\theta \int_0^1 (-\log u(r'_n))^{-1} u'(r')(u(r'))^{\theta-1} \frac{[-\log(u(r'))^{\theta}]^n}{n!} dr'$$

$$= n\theta \int_0^{\infty} \frac{y^{n-1}}{n!} e^{-y} dy = \theta,$$

where the substitution $y = -\theta \log u(r'_n)$ is used.

If similar integration is carried out, one obtains

$$E(\hat{\theta}^2) = \frac{n^2}{n(n-1)} \theta^2,$$

and hence

$$Var(\hat{\theta}) = \frac{\theta^2}{(n-1)}; n \geq 2,$$

which implies that $\hat{\theta} = -n(\log u(r'_n))^{-1}$ is a consistent estimator for θ.

The density of $Y_n = u(R'_n) = R'_n(2 - R'_n)$ is:

$$f_{Y_n}(y) = \frac{[-\log(y^{\theta})]^n}{n!} \theta y^{\theta-1} = \theta^{n+1} \frac{(-\log(y))^n}{n!} y^{\theta-1}. \tag{9}$$

This is the density of lower record values from the power distribution.

It is clear from (9) that $u(R_n')$ is sufficient and complete statistic for θ. Therefore

$$\hat{\theta} = -n(\log u(r_n'))^{-1} \text{ is UMVUE for } \theta. \qquad (10)$$

5. AN APPLICATION

Lawless (1982) page 256 fitted a set of data, representing the numbers of cycles to failure for a group of 60 electrical appliances, to a mixture of two Weibull distributions. We reproduce the data (ordered failure times) here:

14	34	59	61	69	80	123	142	165	210
381	464	479	556	574	839	917	969	991	1064
1088	1091	1174	1270	1275	1355	1397	1477	1578	1649
1702	1893	1932	2001	2161	2292	2326	2337	2628	2785
2811	2886	2993	3122	3248	3715	3790	3857	3912	4100
4106	4116	4315	4510	4584	5267	5299	5583	6065	9701

The last observation is about 4 standard deviations larger than the mean so it may be considered as an outlier. We will ignore this data point in our analysis and the data is rescaled by dividing each observation by 7000. The maximum likelihood estimator (based on the whole sample) of θ is computed to be 0.77.

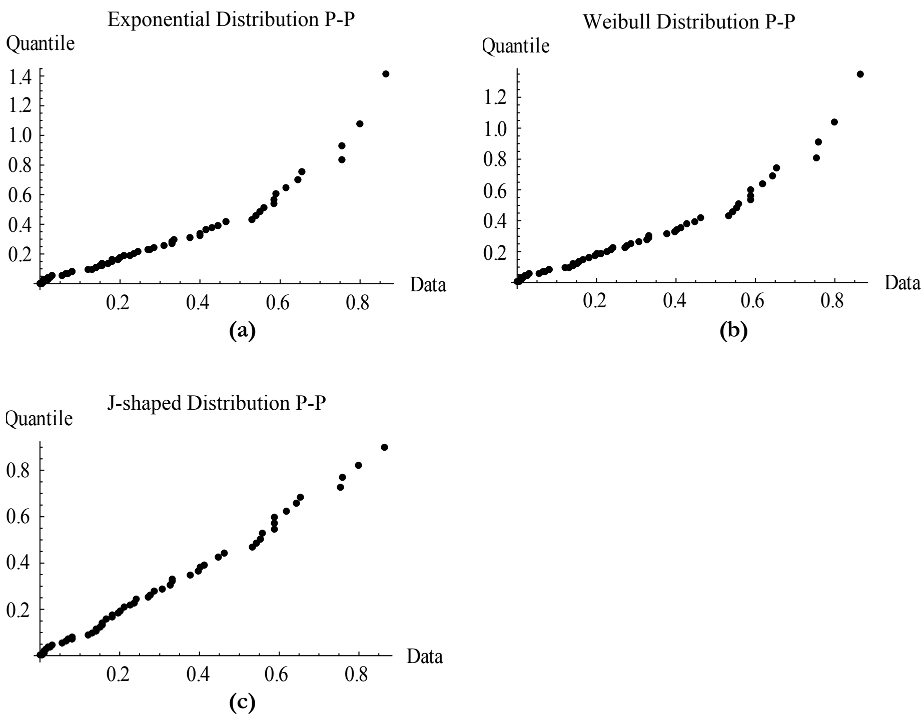

Figure 1(a)-(c) – Probability plots for failure data assuming exponential, Weibull, and J-shaped distributions.

The probability plots assuming that the data is exponential, Weibull, and J-shaped distribution, respectively are displayed in Figure 1 (a)-(c). These plots are nothing but the inverse of the empirical distribution function under each of the above assumed distributions. Clearly, the J-shaped distribution under study has the best fit compared to the exponential and the Weibull distributions.

The density, distribution, and the survival functions of the fitted J-shaped distribution are displayed in Figure 2. We notice that the hazard function is almost bathtub-shapes.

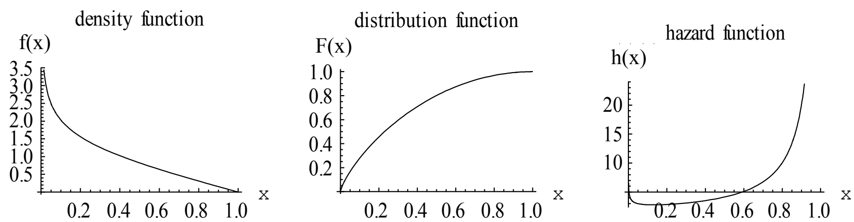

Figure 2 – The density, distribution, and survival functions of the fitted J-shaped distribution for the failure data.

Having fitted a J-shaped distribution with $\theta = 0.77$ to the data given above, all theoretical results can be applied to this model. In particular we may obtain the value of the MLE for θ based on a sequence of lower records obtained upon randomizing the above data. For example, one randomization produces the sequence 1702, 1091, 165, 142, 34, 14. Based on this sequence, applying (10), the MLE for θ is approximately 0.72. We note that this estimator is somewhat close to the MLE based on the whole sample obtained to be 0.77.

6. CONCLUSIONS AND FURTHER STUDIES

We have studied in this article the distributional properties of record values from a J-shaped family of distributions. Based on lower records, we derived recurrence relations and bounds for moments and product moments of record values. Moreover, the maximum likelihood estimator of the shape parameter θ was obtained and shown to be consistent, sufficient, complete and UMVUE estimator. Further a real life data has been fitted to this family of distributions.

In addition to further investigation of the pertinence of these distributions in survival and reliability analysis, one may study the prediction of future record values from this family of distributions. Also, studies to obtain the MLE estimator of the scale parameter, based on both complete and record samples, can be conducted.

Department of Mathematics
College of Sciences
The University of Jordan

AHMAD A. ZGHOUL

REFERENCES

M. AHSANULLAH, (1995), *Introduction to Record Statistics*, NOVA Science Publishers Inc., Hungtington, New York.

B.C. ARNOLD, N. BALAKRISHNAN, H.N. NAGARAJA, (1998), *Records*, John Wiley and Sons.

K.N. CHANDLER, (1952), *The distribution and frequency of record values*, "Journal of the Royal Statistical Society", Series B, 14, pp. 220-228.

I.S. GRADSHTEYN, I. M. RYZHIK, (1980), *Table of Integrals, Series, and Products*, Academic Press, Orlando, Florida, USA.

J.F. LAWLESS, (1982), *Statistical models and methods for lifetime data*, John Wiley and Sons.

S. NADARAJAH. S. KOTZ, (2003), *Moments of Some J-Shaped Distributions*, "Journal of Applied Statistics", 30(3), pp. 311-317.

H.N. NAGARAJA, (1988), *Record Values and Related Statistics- A review*, "Communications in Statistics- Theory and Methods ", 17, pp. 2223-2238.

V.B. NEVZOROV, (1987), *Records*, "Theory of Probability and Applications", 32(2), pp. 201-228.

V.B. NEVZOROV, N. BALAKRISHNAN, (1998). *A Record of Records*, "In Handbook of Statistics-16: Order Statistics: Theory and Methods", pp. 515-570, North-Holland, Amsterdam, The Netherlands.

C.W. TOPP, F.C. LEONE, (1955), *A family of J-Shaped Frequency Functions*, "Jounal of the American Statistical Association", 50, pp. 209-219.

A.A. ZGHOUL, (2010), *Order statistics from a family of J-shaped distributions,* to appear in "METRON", LXVIII, n. 2, pp. 127-136.

SUMMARY

Record values from a family of J-shaped distributions

A family of J-shaped distributions has applications in life testing modeling. In this paper we study record values from this family of distributions. Based on lower records, recurrence relations and bounds as well as expressions for moments and product moments of record values are obtained, the maximum likelihood estimator of the shape parameter is derived and shown to be consistent, sufficient, complete and UMVU estimator. In addition, an application in reliability is given.