

Pseudo Completion of Censored sample in Burr type X Distribution

R.R.L.Kantam¹, and M.S.Ravikumar^{2*}

¹Retd. Professor, Department of Statistics, Acharya Nagarjuna University, Nagarjunanagar-522510, Guntur-Andhra Pradesh, INDIA, E-mail: kantam.rrl@gmail.com

²Asst.Professor, Department of Community Medicine, KIMS & Research Foundation / General Hospital, Amalapuram-533201, Andhra Pradesh, INDIA, *E-mail: msrk.raama@gmail.com

Abstract: The two parameter Burr type X distribution is considered. The well-known classical method – maximum likelihood (ML) estimation of both the parameters is from complete sample attempted. Two modifications to overcome the iterative estimation of the scale parameter are suggested. The suggested methods are found to be efficient. A censored sample is considered and a method of filling its missing observations is suggested in order to a Pseudo Complete Sample (PCS). The proposed modified maximum likelihood methods are applied to PCS/True Complete Sample (TCS) in an illustration. The estimate from PCS/TCS are found to differ little indicating admissibility of suggested PCS method.

Keywords: Burr Type X Distribution, MLE, Order Statistics, Censored Sample

I. INTRODUCTION

The natural phenomenon in reliability studies is “The aging concept” – indicated by increasing instantaneous failure probability with age of the product. A specific case of Weibull distribution exhibiting aging effect with an integer valued shape parameter is known as “The Rayleigh distribution”. Its cumulative distribution function (CDF), probability density function (PDF) and hazard function are given by

$$F(x) = 1 - e^{-x^2} \dots\dots\dots (1.1)$$

$$f(x) = 2xe^{-x^2} \dots\dots\dots (1.2)$$

$$h(x) = 2x \dots\dots\dots (1.3)$$

If, $F(x)$ is the cumulative distribution function of a positive valued random variable, then $[F(x)]^k$, $k > 0$ also satisfies all the requirements for the cumulative distribution function of another positive valued random variable. If k is an integer, it can be interpreted as the failure probability of a parallel system of k - components whose life times are independently and identically distributed random variables, each with a common CDF – $F(x)$. Exploring this concept to non-integer values of k also, many researchers in the recent past made extensive studies on models of the type $[F(x)]^k$ generated by a baseline model $F(x)$. Such new models are named as exponentiated models by some authors and generalized models by some authors. For instance if the $F(x)$ is exponential, $[F(x)]^k$ is named as generalised exponential [2], if $F(x)$ is Weibull, $[F(x)]^k$ is named as exponentiated Weibull [5]. Banking on this notion, generalized Rayleigh distribution was studied by [6], as a process of revisit to Burr type X distribution. Its cumulative distribution function (CDF), probability density function (PDF) and hazard function are given by

$$F(x; k) = (1 - e^{-x^2})^k; x > 0, k > 0 \dots\dots\dots (1.4)$$

$$f(x; k) = 2kxe^{-x^2} (1 - e^{-x^2})^{k-1}; \quad x > 0, k > 0 \quad \dots\dots\dots (1.5)$$

$$h(x; k) = \frac{2kxe^{-x^2}(1-e^{-x^2})^{(k-1)}}{1-(1-e^{-x^2})^k} \quad \dots\dots\dots (1.6)$$

Burr [1] has suggested a number of forms of cumulative distribution functions that might be useful in modeling various practical situations. In all, he suggested twelve models as listed below.

- (I) $F(x) = x; \quad 0 < x < 1,$
- (II) $F(x) = (e^{-x} + 1)^{-k},$
- (III) $F(x) = (x^{-c} + 1)^{-k}; \quad 0 < x < \infty,$
- (IV) $F(x) = \left[\left(\frac{c-x}{x} \right)^{1/c} + 1 \right]^{-k}; \quad 0 < x < c,$
- (V) $F(x) = (ce^{-\tan x} + 1)^{-k}; \quad -\frac{\pi}{2} < x < \frac{\pi}{2},$
- (VI) $F(x) = (ce^{-k \sinh x} + 1)^{-k},$
- (VII) $F(x) = 2^{-k}(1 + \tanh x)^k,$
- (VIII) $F(x) = \left(\frac{2}{\pi} \arctan e^x \right)^k,$
- (IX) $F(x) = 1 - \frac{2}{c[(1+e^x)^k - 1] + 2},$
- (X) $F(x) = (1 - e^{-x^2})^k; \quad 0 < x < \infty,$
- (XI) $F(x) = \left(x - \frac{1}{2\pi} \sin 2\pi x \right)^k; \quad 0 < x < 1,$
- (XII) $F(x) = 1 - (1 + x^c)^{-k}; \quad 0 < x < \infty.$

Thus generalized Rayleigh distribution and Burr type X distributions are one and the same.

In the above models k and c are the positive parameters involved in the respective models. The first model is the well-known uniform distribution also included by Burr [1]. Among these twelve forms, the type X and type XII models are most frequently applied by many researchers. Our focus is Burr Type X model, whose expressions are given in equations (1.4), (1.5), and (1.6). If a scale parameter say λ is introduced, the cumulative distribution function, probability density function and hazard function are respectively given by

$$F(x; k, \lambda) = (1 - e^{-(\lambda x)^2})^k; \quad x > 0, k > 0, \lambda > 0, \quad \dots\dots\dots (1.8)$$

$$f(x; k, \lambda) = 2k\lambda^2 x e^{-(\lambda x)^2} (1 - e^{-(\lambda x)^2})^{(k-1)}; \quad x > 0, k > 0, \lambda > 0, \quad \dots\dots\dots (1.9)$$

$$h(x; k, \lambda) = \frac{2k\lambda^2 x e^{-(\lambda x)^2} (1 - e^{-(\lambda x)^2})^{(k-1)}}{1 - (1 - e^{-(\lambda x)^2})^k} \quad \dots\dots\dots (1.10)$$

Expressions in (1.4), (1.5) and (1.6) are called standard Burr type X model, those in equations (1.8), (1.9) and (1.10) are called scaled Burr type X or Two parameter Burr type X model.

In this paper we attempt to present how a failure censored sample can be converted into a complete sample with the help of estimates of the missing observations using the uncensored observations. Thus we have a complete sample wherein the missing observations of a censored sample are filled with their estimates using uncensored observations. Such a sample is called a Pseudo Complete Sample (PCS) [7] and the references therein are some of the works in this direction. In Section – II we study the need for modifications to log likelihood equations to estimate the scale parameter for a known shape parameter and suggest two admissible methods that would lead to modified maximum likelihood estimates for complete samples. We describe a method of pseudo

completion of a given censored sample in Section – III and use such a Pseudo Complete Sample to get the modified maximum likelihood estimation.

II. MODIFIED MAXIMUM LIKELIHOOD ESTIMATION FROM COMPLETE SAMPLES

The probability density function of the two parameter Burr type X distribution is given by

$$f(x; k, \lambda) = 2k\lambda^2 x e^{-(\lambda x)^2} (1 - e^{-(\lambda x)^2})^{(k-1)}; x > 0, k > 0, \lambda > 0 \quad (2.1)$$

Its log likelihood equations to get the maximum likelihood estimates of λ and k from a complete sample are given by (after simplification).

$$\frac{\partial \log L}{\partial \lambda} = 0 \Rightarrow \frac{2n}{\lambda} - 2\lambda \sum_{i=1}^n x_i^2 + 2\lambda(k-1) \sum_{i=1}^n \frac{x_i^2 e^{-(\lambda x_i)^2}}{1 - e^{-(\lambda x_i)^2}} = 0 \quad (2.2)$$

$$\frac{\partial \log L}{\partial k} = 0 \Rightarrow \frac{n}{k} + \sum_{i=1}^n \ln(1 - e^{-(\lambda x_i)^2}) = 0 \quad (2.3)$$

The maximum likelihood estimator of k is a closed form expression involving λ given as

$$\hat{k} = \frac{-n}{\sum_{i=1}^n \ln(1 - e^{-(\lambda x_i)^2})} \quad (2.4)$$

The maximum likelihood estimator of λ is an iterative solution of the equation (2.2) involving k . In order to overcome the iterative nature of the solution we proceed as follows. Equation (2.2) can be rewritten as

$$2n - 2 \sum_{i=1}^n z_i^2 + 2(k-1) \sum_{i=1}^n \frac{z_i^2 e^{-z_i^2}}{1 - e^{-z_i^2}} = 0, \quad (2.5)$$

where $z_i = \lambda x_i$.

$$\text{Consider the expression } g(z_i) = \frac{z_i^2 e^{-z_i^2}}{1 - e^{-z_i^2}} \quad (2.6)$$

of equation (2.5).

We approximate this expression by a linear one in z_i in a small neighborhood of the i^{th} quantile of the population say

$$g(z_i) \cong \alpha_i + \beta_i z_i. \quad (2.7)$$

With this approximation equation (2.5) becomes a quadratic equation in λ given by

$$A\lambda^2 + B\lambda + C = 0. \quad (2.8)$$

Where, $A = \sum_{i=1}^n z_i^2$, $B = -(k-1) \sum_{i=1}^n \beta_i z_i$, $C = -n - (k-1) \sum_{i=1}^n \alpha_i$.

Positive root of this equation is an estimate of λ called the modified maximum likelihood estimate (MMLE) of λ . It can be seen that A,B,C depend on the ordered observations x_1, x_2, \dots, x_n , the shape parameter k and the slope, intercept of the linear approximation (2.7). In order to find α_i, β_i , we suggest two methods of [3].

Method-I:

Let $p_i = \frac{i}{n+1}, i = 1, 2, \dots, n$.

Let z_i^*, z_i^{**} be the solutions of the following equations

$$F(z_i^*) = p_i^*, F(z_i^{**}) = p_i^{**},$$

Where $p_i^* = p_i - \sqrt{\frac{p_i q_i}{n}}, p_i^{**} = p_i + \sqrt{\frac{p_i q_i}{n}}$, $F(\cdot)$ is the cdf of standard Burr type X distribution, and

$q_i = 1 - p_i$.

The expressions for z_i^*, z_i^{**} are

$$z_i^* = \sqrt{-Ln \left[1 - \left(p_i - \sqrt{\frac{p_i q_i}{n}} \right)^k \right]}, \quad (2.9)$$

$$z_i^{**} = \sqrt{-Ln \left[1 - \left(p_i + \sqrt{\frac{p_i q_i}{n}} \right)^k \right]} \quad (2.10)$$

The slope β_i and intercept α_i of the linear approximation in the equation (2.7) are given by

$$\beta_i = \frac{g(z_i^{**}) - g(z_i^*)}{z_i^{**} - z_i^*}, \quad (2.11)$$

$$\alpha_i = g(z_i^*) - \beta_i z_i^*. \quad (2.12)$$

Method-II:

Considering Taylor's expansion of $g(z_i)$ upto its first derivative w.r.t z_i in a neighborhood of population quantile corresponding to p_i , we get

$$\beta_i = g'(\xi_i), \quad (2.13)$$

$$\alpha_i = g(\xi_i) - \beta_i \xi_i, \quad (2.14)$$

Where ξ_i is the quantile of Burr type X distribution, given as the solution of the equation? $F(\xi_i) = p_i$.

$$\text{i.e., } \xi_i = \sqrt{-Ln \left[1 - \left(p_i - \sqrt{\frac{p_i q_i}{n}} \right)^k \right]}. \quad (2.15)$$

It can be seen from (2.6) that

$$g'(\xi_i) = \frac{2\xi_i e^{-\xi_i^2} (1 - \xi_i - e^{-\xi_i^2} + \xi_i e^{-\xi_i^2} - \xi_i^2 e^{-\xi_i^2})}{(1 - e^{-\xi_i^2})^2} \quad (2.16)$$

Substituting (2.15) in (2.16) we get

$$\beta_i = g'(\xi_i) = \frac{2\xi_i e^{-\xi_i^2} (1 - \xi_i - e^{-\xi_i^2} + \xi_i e^{-\xi_i^2} - \xi_i^2 e^{-\xi_i^2})}{(1 - e^{-\xi_i^2})^2} \quad (2.17)$$

Using β_i in (2.14) we get α_i .

In the above two modified methods, the basic principle is that certain expressions in the log likelihood equation are linearised in a neighborhood of the population quantile which depends on the size of the sample also. The larger the size, the narrower is the neighborhood and hence the closer is the approximation. That is, the exactness of the approximation becomes finer and finer for large values of 'n'. Hence the approximate log likelihood equation and the exact log likelihood equation tend to each other as $n \rightarrow \infty$. Hence the exact and modified MLEs are asymptotically identical (Tiku *et al.* [8]). The same may not be true in small samples and these are to be assessed with the help of small sample characteristics of the MMLEs.

III. PSEUDO COMPLETION OF A CENSORED SAMPLE

Let $x_1 < x_2 < \dots < x_{n-1}$ be a censored sample of a planned sample of size n in which the last observation is missing. If ξ_i is the i^{th} quantile in a standard population we make use of a pivotal equation as $\lambda x_i \cong \xi_i$. Let $\Delta_1, \Delta_2, \dots, \Delta_{n-2}$ are the successive differences obtained from x_1, x_2, \dots, x_{n-1} .

$$\text{i.e., } \Delta_i = x_i - x_{i-1}, \quad i=2, 3, \dots, (n-1).$$

$$\text{Let } \bar{\Delta} = \frac{\sum_{i=2}^{n-1} \Delta_i}{n-2} \quad \text{and} \quad \bar{\xi} = \frac{\sum_{i=1}^n \xi_i}{n}.$$

In the first iteration the missing observation x_n is estimated as

$$\hat{x}_{n(1)} = x_{n-1} + \bar{\Delta} \quad (3.1)$$

First Pseudo Complete sample is $x_1, x_2, x_3, \dots, x_{n-1}, \hat{x}_{n(1)}$.

$$\text{Let } \bar{x}_{(1)} = \frac{\sum_{i=1}^{n-1} x_i + \hat{x}_{n(1)}}{n} \quad (3.2)$$

$$\hat{\lambda}_{(1)} = \frac{\bar{\xi}}{\bar{x}_{(1)}} \quad (3.3)$$

$$\text{The second iterative estimate of } x_n \text{ is given by } \hat{x}_{n(2)} = \frac{\xi_n}{\hat{\lambda}_{(1)}} \quad (3.4)$$

Second Pseudo Complete Sample is $x_1, x_2, x_3, \dots, x_{n-1}, \hat{x}_{n(2)}$.

where $\hat{x}_{n(2)}$ is in (3.4).

$$\text{Let } \bar{x}_{(2)} = \frac{\sum_{i=1}^{n-1} x_i + \hat{x}_{n(2)}}{n} \quad (3.5)$$

$$\text{Let } \hat{\lambda}_{(2)} = \frac{\bar{\xi}}{\bar{x}_{(2)}} \quad (3.6)$$

$$\text{Third iterative value of } x_{(n)} \text{ is given by } \hat{x}_{n(3)} = \frac{\xi_n}{\hat{\lambda}_{(2)}} \quad (3.7)$$

This procedure is to be repeated till at any j^{th} iteration

$$|\hat{x}_{n(j)} - \hat{x}_{n(j+1)}| < 0.01 \quad (3.8)$$

As soon as this convergence is achieved, the ultimate Pseudo Complete Sample is considered for general inferential purposes.

As a matter of illustration we consider the following example of a live sample of size 23 from [4]: 17.88, 28.92, 33, 41.52, 42.12, 45.6, 48.8, 51.84, 51.96, 54.12, 55.56, 67.8, 68.64, 68.64, 68.88, 84.12, 93.12, 98.64, 105.12, 105.84, 127.92, 128.04, 174.4. These observations represent the number of million revolutions before failing of ball-bearings in a life test. From this we regard the following four sets of ordered observations as true complete samples of sizes 5,10,15,20 respectively.

Set-1: n=5; 17.88, 28.92, 33, 41.52, 42.12

Set-2: n=10; 17.88, 28.92, 33, 41.52, 42.12, 45.6, 48.8, 51.84, 51.96, 54.12

Set-3: n=15; 17.88, 28.92, 33, 41.52, 42.12, 45.6, 48.8, 51.84, 51.96, 54.12, 55.56, 67.8, 68.64, 68.64, 68.88

Set-4: n=20; 17.88, 28.92, 33, 41.52, 42.12, 45.6, 48.8, 51.84, 51.96, 54.12, 55.56, 67.8, 68.64, 68.64, 68.88, 84.12, 93.12, 98.64, 105.12, 105.84

Assuming that these observations follow Burr type X distribution, we treat the highest ordered observation of each of the above four samples to be missing and estimate the same as described earlier. The Pseudo Complete Sample so obtained is made use of to get the MMLE of λ from the Pseudo Complete Sample as well as the true complete sample by methods – I and – II of Section - II. The results of Pseudo Complete Sample are presented in Tables 1 to 4 for n=5(5)20 respectively along with the estimated standard errors shown in the parentheses.

Table 1: Pseudo Completion of a Sample with Estimate of Largest Missing Observation (n=5)

n	k	\hat{x}_5	Pseudo Complete Sample (PCS)	True Complete Sample (TCS)	Estimates of λ based on			
					PCS		TCS	
					MMLE-I	MMLE-II	MMLE-I	MMLE-II
5	1.50	49.51161	17.88,28.92,33,41.52,49.51161	17.88, 28.92,	0.03141 (0.006641)	0.030421 (0.006106)	0.033164 (0.007012)	0.031974 (0.006417)
	2.00	46.65493	17.88,28.92,33,41.52,46.65493	33, 41.52,	0.034713 (0.00654)	0.031423 (0.005303)	0.03585 (0.006754)	0.032296 (0.00545)
	2.50	44.88519	17.88,28.92,33,41.52,44.88519	42.12	0.037258 (0.006105)	0.031119 (0.004309)	0.037977 (0.006223)	0.031601 (0.004376)
	3.00	43.65793	17.88,28.92,33,41.52,43.65793		0.039324 (0.005991)	0.030265 (0.003542)	0.039734 (0.006053)	0.030509 (0.003571)

Table 2: Pseudo Completion of a Sample with Estimate of Largest Missing Observation (n=10)

n	k	\hat{x}_{10}	Pseudo Complete Sample (PCS)	True Complete Sample (TCS)	Estimates of λ based on			
					PCS		TCS	
					MMLE-I	MMLE-II	MMLE-I	MMLE-II
10	1.50	70.51067	17.88,28.92,33,41.52,42.12,45.6,48.8,51.84,51.96,70.51067	17.88, 28.92, 33, 41.52,	0.024504 (0.003507)	0.023634 (0.003198)	0.026103 (0.003736)	0.025042 (0.003389)
	2.00	66.22923	17.88,28.92,33,41.52,42.12,45.6,48.8,51.84,51.96,66.22923	42.12,45.6, 48.8, 51.84,	0.026864 (0.0034)	0.024231 (0.002686)	0.028189 (0.003568)	0.025247 (0.002799)
	2.50	63.54704	17.88,28.92,33,41.52,42.12,45.6,48.8,51.84,51.96,63.54704	51.96, 54.12	0.028694 (0.003184)	0.023944 (0.002268)	0.02984 (0.003311)	0.024718 (0.002342)
	3.00	61.67855	17.88,28.92,33,41.52,42.12,45.6,48.8,51.84,51.96, 61.67855		0.03019 (0.003085)	0.023293 (0.001874)	0.031208 (0.003189)	0.023912 (0.001923)

Table 3: Pseudo Completion of a Sample with Estimate of Largest Missing Observation (n=15)

n	k	\hat{x}_{15}	Pseudo Complete Sample (PCS)	True Complete Sample (TCS)	Estimates of λ based on			
					PCS		TCS	
					MMLE-I	MMLE-II	MMLE-I	MMLE-II
15	1.50	88.36869	17.88,28.92,33,41.52,42.12,45.6,48.8,5 1.84,51.96,54.12,55.56,67.8,68.64,68.6 4,88.36869	17.88, 28.92, 33, 41.52,	0.017347 (0.002003)	0.020191 (0.002202)	0.021681 (0.002504)	0.020837 (0.002272)
	2.00	82.85789	17.88,28.92,33,41.52,42.12,45.6,48.8,5 1.84,51.96,54.12,55.56,67.8,68.64,68.6 4,82.85789	42.12,45.6, 48.8, 51.84, 51.96,	0.022904 (0.002344)	0.020684 (0.001837)	0.023441 (0.002399)	0.021082 (0.001872)
	2.50	79.38872	17.88,28.92,33,41.52,42.12,45.6,48.8,5 1.84,51.96,54.12,55.56,67.8,68.64,68.6 4,79.38872	54.12, 55.56,67.8, 68.64,68.64, 68.88	0.024434 (0.002211)	0.020456 (0.001567)	0.024841 (0.002248)	0.020711 (0.001586)
	3.00	76.96624	17.88,28.92,33,41.52,42.12,45.6,48.8,5 1.84,51.96,54.12,55.56,67.8,68.64,68.6 4,76.96624		0.025692 (0.00213)	0.019926 (0.001319)	0.026006 (0.002156)	0.020095 (0.00133)

Table 4: Pseudo Completion of a Sample with Estimate of Largest Missing Observation (n=20)

n	k	\hat{x}_{20}	Pseudo Complete Sample (PCS)	True Complete Sample (TCS)	Estimates of λ based on			
					PCS		TCS	
					MMLE-I	MMLE-II	MMLE-I	MMLE-II
20	1.50	111.5974	17.88,28.92,33,41.52,42.12,45.6,48.8,51.84,51.96,5 4.12,55.56,67.8,68.64,68.64,68.88,84.12,93.12,98.6 4,105.12,111.5974	17.88,28.92, 33, 41.52, 42.12,45.6,	0.016915 (0.001665)	0.016408 (0.001523)	0.017028 (0.001676)	0.016509 (0.001532)
	2.00	82.85789	17.88,28.92,33,41.52,42.12,45.6,48.8,51.84,51.96,5 4.12,55.56,67.8,68.64,68.64,68.88,84.12,93.12,98.6 4,105.12,82.8579	48.8, 51.84, 51.96,54.12, 55.56,67.8, 68.64,68.64,68.88,84.12,93	0.018918 (0.001716)	0.017219 (0.001341)	0.018492 (0.001678)	0.016894 (0.001316)
	2.50	100.039	17.88,28.92,33,41.52,42.12,45.6,48.8,51.84,51.96,5 4.12,55.56,67.8,68.64,68.64,68.88,84.12,93.12,98.6 4,105.12,100.039	.12,98.64, 105.12,105.84	0.019787 (0.001521)	0.016824 (0.001098)	0.019671 (0.001512)	0.016748 (0.001093)
	3.00	96.90877	17.88,28.92,33,41.52,42.12,45.6,48.8,51.84,51.96,5 4.12,55.56,67.8,68.64,68.64,68.88,84.12,93.12,98.6 4,105.12,96.90877		0.020839 (0.001476)	0.01426 (0.000802)	0.02066 (0.001463)	0.014124 (0.000794)

We see from these tables that estimated standard error of an estimate from true complete sample and pseudo complete sample differ little indicating the possible admissibility of the pseudo completion procedure for censored sample, suggested in this paper. However, this is to be established in a general frame work as this remark is made out of a single live example only.

IV. REFERENCES

- [1] Burr, I.W. (1942). Cumulative Frequency Functions. *The Annals of Mathematical Statistics*, vol.13, 215-232.
- [2] Gupta, R.D., and Kundu D. (1999). Generalized Exponential Distributions. *Australian and New Zealand Journal of Statistics*, vol.41, No.2, 173-188.
- [3] Kantam, R.R.L., and Ravikumar, M.S. (2015). Modified Maximum Likelihood Estimation in Burr Type X Distribution. *Inter Stat*, June, pp: 1-14.
- [4] Lawless, J. F. (2003). *Statistical Models and Methods for Lifetime Date*. John Wiley and Sons, Second Edition, Hoboken, pp.228.
- [5] Mudholkar, Govind S., and Srivastava, Deo Kumar. (1993). Exponentiated Weibull family for analyzing bathtub failure-rate. *Reliability, IEEE Transaction*, vol.42, No.2, 299-302.
- [6] Raqab, M. Z., and Kundu, D. (2006). Burr type X distribution: revisited. *Journal of Probability and Statistical Science*, vol.8, 179-198.
- [7] Sriram, B. and Kantam, R.R.L. (2014). Estimation of Scale Parameter and Reliability in Gamma Distribution Pseudo Completion of Censored Sample. *Integrated Journal of Engineering and Technology*, vol.1 (15), pp: 142-151, 2014.
- [8] Tiku, M.L., Tan, W.Y., and Balakrishnan, N. (1986). *Robust Inference*. Marcel Dekker. New York.