THE EGYPTIAN STATISTICAL JOURNAL
ISSR., Cairo Univ., Vol.37, No.2, 1993.

PREDICTION OF STH ORDERED OBSERVATION IN DOUBLY TYPE-II CENSORED SAMPLE FROM ONE-PARAMETER EXPONENTIAL DISTRIBUTION

AHMED A. SOLIMAN AND A. H. ABD ELLAH

Dept. of Mathematics, Assiut University, Sohag 82516, Egypt.

Abstract This paper deals with the problem of predicting the sth ordered observation X_s in a sample of size n from one-parameter exponential distribution, based on the first $(r-k+1)$ ordered observations (Doubly Type-II censoring $(0 \leq k < r < s \leq n)$)

It is shown how to find an interval estimte for $X_{(s)}$ and how the result can be used to predict the remaining elapsed time in certain life test experiments involving items whose life times follow an exponential distribution.

key words and phrases: Exponential distribution, prediction problems, doubly type-II censoring, life test.

1-Introduction

A prediction interval is an interval which uses the results of past sample to obtain the results of a future sample from the same population with a specified probability. One of the earliest papers on this subject was that by Baker [1]. Review papers on prediction interval are given by Hahn and Nelson [4] and by Patel [12].

Kaminsky and Nelson [6] gives the prediction interval for $X_{(s)}$ based on the observed values of subsets of order statistics from a sample. They use the factor C_α where C_α is 100α lower percentage point of the distribution of the random variable $R = (X_{(s)} - X_{(r)})/\theta^*$ and θ^* is the best linear unbiased estimator of θ, but the authers have approximated the distribution of R by an F-distribution. This approximation would be reasonable, especially when the number of degees of freedom in numerator and denominator are not too different. J.F. Lawless[7] solves the problem of predicting $X_{(s)}$ in the case of one-parameter exponential distribution and assumes that each of first r values were observed.Let $X_1 < X_2 < \cdots < X_n$ be a random ordered sample of size n from the one-parameter exponential distribution .

$$f(x,\theta) = \theta \exp(-\theta x) \quad x, \theta > 0$$

In this paper we shall obtain a prediction interval for the sth ordered observation $X_{(s)}$ based on a sample of doubly type-II censored so that one observes only $X_{(k)}$, $X_{(k+1)}$, $\cdots$,$X_{(r)}$. The results obtained can be considered as extensions of [7]. This prediction interval would be useful in life testing problems where some of the data are missing or where optimal selection of the ordered statistics is desired.

2-Prediction Intervals for X_s

Consider an ordered random sample $X_1 < X_2 < \cdots < X_n$ of size n from the exponential distribution with mean θ^{-1}, having denisty

$$f(x,\theta) = \theta\ exp(-\theta x), \quad x,\theta > 0 \tag{1}$$

Suppose that only $X_{(k)}, X_{(k+1)}, \cdots, X_{(r)}$ of the same sample are available (doubly type-II censoring sample) $(0 \leq k < r \leq n)$.

Let

$$T = kX_{(k)} + \sum_{i=k+1}^{r} X_{(i)} + (n-r)X_{(r)}$$

In which for k=0 and $X_{(0)} \equiv 0$ gives the case of singly type-II censoring sample.

For given $0 \leq k < r < s \leq n$ (see Likeš [11]), we consider the variate

$$U = U(k,r,s,n)$$

defined by

$$U = (X_{(s)} - X_{(r)})/T \tag{2}$$

We note that (see Epstein and Sobel [3],Tanis [13]) U has a distribution not involving θ.

The probabilty denisty function of U (see for example, Kendall [5] and Lawless [10]) is found to be

$$dF(X_{(k)}, X_{(r)}, X_{(s)}) = \frac{(F(X_{(k)})^{k-1}(F(X_{(r)}) - F(X_{(k)}))^{r-k-1}(F(X_{(s)}) - F(X_{(r)})^{s-r-1}(1-F(X_{(s)}))^{n-s}}{\beta(k,r-k)\beta(r,s-r)\beta(s,n-s+1)} .dF(X_{(k)})dF(X_{(r)})dF(X_{(s)}) \tag{3}$$

Where $F(t) = 1 - exp(-\theta t)$, $\beta(a,b)$ is the beta function ratio.

$$dF(X_{(k)}, X_{(r)}, X_{(s)}) = \frac{(1-exp(-\theta X_{(k)}))^{k-1}}{\beta(k,r-k)} . \frac{(exp(-\theta X_{(k)}) - exp(-\theta X_{(r)}))^{r-k-1}}{\beta(r,s-r)} . \frac{(exp(-X_{(r)}\theta) - exp(-\theta X_{(s)}))^{s-r-1}}{\beta(s,n-s+1)} .(exp(-\theta X(s)))^{n-s}\theta^3 exp(\ \theta X_{(k)})exp(-\theta X_{(r)}))exp(-\theta X_{(s)})) .dF(X_{(k)})dF(X_{(r)})dF(X_{(s)}) \tag{4}$$

Making the transformations $Y = X_{(r)} - X_{(k)}$ $\quad Z = X_{(s)} - X_{(r)}$, $\quad$ and $X_{(k)} = X_{(k)}$.

Then equation (4) becomes

$$dF(X_{(k)},y,z)=\frac{(1-exp(-\theta X_{(k)}))^{k-1}}{\beta(k,r-k)}$$
$$\cdot\frac{exp(-\theta X_{(k)})(n-r+1)(1-exp(-\theta y)^{r-k-1}exp(-r\theta y)}{\beta(r,s-r)}$$
$$\cdot\frac{(1-exp(-\theta z))^{s-r-1}exp(-\theta sz)exp(-\theta(y+z)(n+1))}{\beta(s,n-s+1)}$$
$$.\theta^3 dF(X_{(k)})dF(y)dF(z) \quad (5)$$

Integrating out of $X_{(k)}$ and Y, we find the density of Z is given by

$$f(z;\theta)=\frac{exp(-\theta z(n-s+1))(1-exp(-\theta z))^{s-r-1}}{\beta(s-r,n-s+1)}\theta \quad (6)$$

The density of T is given by

$$f(t;\theta)=\frac{(\theta t)^{r-k-1}exp(-\theta t)}{\Gamma(r-k)}\theta \quad (7)$$

Since Z, T are independent, we have the joint density of Z and T as

$$f(z,t;\theta)=\frac{(exp(-\theta z))^{n-s+1}(1-exp(-\theta z))^{s-r-1}(\theta t)^{r-k-1}}{\Gamma(r-k)\beta(s-r,n-s+1)}exp(-\theta t)\theta^2 \quad (8)$$

Making the transformation $U=\frac{Z}{T}$, $T=T$ the joint density of U and T is

$$f(u,t;\theta)=\frac{exp(-\theta t)}{\Gamma(r-k)}\frac{(exp(-\theta tu))^{n-s+1}}{\beta(s-r,n-s+1)}$$
$$.(1-exp(-\theta tu))^{s-r-1}(\theta t)^{r-k-1}\theta^2 t=\frac{exp(-\theta t)}{\Gamma(r-k)}$$
$$\cdot\frac{(exp(-\theta tu))^{n-s+1}}{\beta(s-r,n-s+1)}(\theta t)^{r-k}\theta$$
$$\cdot\sum_{i=0}^{s-r-1}(-1)^i\binom{s-r-1}{i}exp(-\theta tui)$$
$$=\frac{\theta(\theta t)^{r-k}}{\Gamma(r-k)\beta(s-r,n-s+1)}$$
$$\cdot\sum_{i=0}^{s-r-1}(-1)^i\binom{s-r-1}{i}exp(-\theta t(1+(n-s+i+1)u) \quad (9)$$

Integrating out T, we obtain the denisty of U as shown:

$$f(u) = \frac{1}{\Gamma(r-k)\beta(s-r, n-s+1)} \cdot \sum_{i=0}^{s-r-1} (-1)^i \binom{s-r-1}{i} \frac{\Gamma(r-k+1)}{(1+(n-s+i+1)u)^{r-k+1}}$$
$$= \frac{r-k}{\beta(s-r, n-s+1)} \cdot \sum_{i=0}^{s-r-1} (-1)^i \binom{s-r-1}{i} \frac{1}{(1+(n-s+i+1)u)^{r-k+1}} \qquad ,u > 0 \quad (10)$$

Integration yields

$$Pr(U \geq t) = \frac{1}{\beta(s-r, n-s+1)} \cdot \sum_{i=0}^{s-r-1} (-1)^i \binom{s-r-1}{i} \frac{1}{(n-s+i+1)(1+(n-s+i+1)t)^{r-k}}$$
$$= P(t, k, r, s, n) \quad (11)$$

Whence the distribution function of U is given by:

$$F(t) = 1 - Pr(U \geq t)$$

Probability statements about U give prediction statements about X_s, on the basis of observed $X_{(k)}, X_{(k+1)}, \cdots, X_{(r)}, \quad T$

For example, the statement $Pr(U \leq t_0) = \alpha$ yields the prediction statement

$$Pr(X_{(s)} \leq X_{(r)} + t_0 T) = \alpha \quad (12)$$

giving a (one-sided) $100\alpha\%$ prediction interval for $X_{(s)}$.

For given values of n, k, r, s, t the probabilities $Pr(t, k, r, s, n)$ can be evaluated on a computer.

Special Cases

(i) When $s = n$ (that is we wish to predict the largest observation on the basis of the $(r-k+1)$ smallest, we obtain

$$P(t, k, r, n, n) = Pr(U_1 \geq t) = \frac{1}{\beta(n-r, n-n+1)} \sum_{i=0}^{n-r-1} (-1)^i \binom{n-r-1}{i} \frac{1}{(i+1)(1+(i+1)t)^{r-k}}$$

where $U_1 = \frac{X_{(n)} - X_{(r)}}{T}$. This can be conveniently rewritten as

$$P(t,k,r,n,n) = 1 - \sum_{i=0}^{n-r} (-1)^i \binom{n-r}{i} \frac{1}{(1+it)^{r-k}}$$

Where the distribution function of $U_1 = U(t,k,r,n,n)$ is given by

$$Pr(U_1 \le t) = \sum_{i=0}^{n-r} (-1)^i \binom{n-r}{i} \frac{1}{(1+it)^{r-k}} \tag{13}$$

(ii) In the special case where r=s-1, it is known that the statistic

$$(n-s+1)(s-k-1)\frac{(X_{(s)} - X_{(s-1)})}{T}$$

has an F distribution with (2, 2(s-k-1)) degrees of freedom (see for example [8]) i.e

$$F = (n-s+1)(s-k-1)\left(\frac{X_{(s)} - X_{(s-1)}}{T}\right) = (n-s+1)(s-k-1)U_2$$

where $U_2 = \frac{X_{(s)} - X_{(s-1)}}{T}$ since

$$\begin{aligned} \alpha &= Pr\left[F \le f_{(2,2(s-k-1);\alpha)}\right] \\ &= Pr\left[U_2 \le \frac{f_{(2,2(s-k-1);\alpha)}}{(n-s+1)(s-k-1)}\right] \end{aligned}$$

However if r=s-1 we obtain the factor

$$U_2(k, s-1, s, n; \alpha) = \frac{f_{(2,2(s-k-1);\alpha)}}{(n-s+1)(s-k-1)} \tag{14}$$

where $f_{(a,b;\alpha)}$ is the 100α lower percentile of an F distribution with (a,b) degrees of freedom. Table 1 gives $U_2(k, s-1, s, n; \alpha)$ for n=10, s=7, r=6, k=0(1)5 using α= 0.95, 0.99, 0.995, 0.05, 0.01, and 0.005 values with the exact values in parentheses. We checked the accuracy of these factor values by comparing them with the exact values using the relation in (14). We note that k=0 gives the same results of Bain [2] (single type II censoring).

Table 1

				α					
n	s	r	k	.95	.99	.995	.05	.01	.005
10	7	6	0	.972	1.7316	2.1274	.012878	.00251	.00125
				(.97132)	(1.73165)	(2.1274)	(.012878)	(.002514)	(.00125)
			1	1.025	1.89	2.356	.012889	.002516	.001253
				(1.0257)	(1.8898)	(2.3567)	(.012889)	(.002515)	(.001253)
			2	1.115	2.1625	2.7606	.012906	.002516	.001253
				(1.1147)	(2.16228)	(2.7606)	(.01231)	(.002515)	(.001253)
			3	1.285	2.73	3.66	.012933	.002516	.001254
				(1.2858)	(2.7311)	(3.6360)	(.012933)	(.002516)	(.001254)
			4	1.735	4.5	6.57107	.012987	.002518	.001254
				(1.7360)	(4.5)	(6.57107)	(.012989)	(.002518)	(.001254)
			5	4.75	24.75	49.75	.0131578	.002525	.001256
				(4.75)	(24.75)	(49.75)	(.013157)	(.002525)	(.001256)

3- Applications of The Prediction Result

3.1-A life Test Where All Units Are Observed Until Failure

Consider a life test where n units, whose lifetimes follow the same exponential distribution, are put on test simultaneously, and where some units are observed until failure. We can, using the results derived above, given a prediction interval for the largest life time X_n on the basis of the $(r-k+1)$ smallest life times $X_k \leq X_{k+1} \leq \cdots \leq X_r$;$Xn$ is in this case the total elapsed time required to complete the test.

Numerical example

We use Lawless's example of Section 3.1 [7]. Ten items, whose failure distribution is the same one-parameter exponential, are put on test simultaneously. The test is terminated after the first 4 failures are observed, giving failure times: 30, 90, 120, 170. For this example n=10, s=10, r=4, k=0, and then T=1430. We can find from (13) that $Pr(U_1 \leq 2.098)$ is very nearly 95%, this yields the prediction statement:

$$Pr(X_{(10)} \leq 170 + (2.098)(1430)) = Pr(X_{(10)} \leq 3173) = 0.95$$

That is, we can be (approximately) 95% confident that the total elapsed test time will not exceed 3173 hours. (the same result of Lawless [7]).

Now, we consider some cases in which some of the failure times are missing.

Case (i) The failure #1 is unobserved, for this case, $k = 2$, $T = 1490$, $t = 2.12$, we can be 95% confident that $X_{(10)}$ will not exceed 3329.

Case (ii) the failures #1 and #2 went unobserved, for this case, $k = 3$, $T = 1550$, $t = 2.13$, we can be 95% confident that $X_{(10)}$ will not exceed 3472.

3.1-B A Life Test Where testing is terminated After The rth failure

Consider a life testing situation similar to that of 3.1-A but suppose that it had been decided beforhand to terminate the test after fifth failure on the basis of the first four failure times, we can compute say, an upper 95% prediction limit for $X_{(5)}$ with $s = 5$, $r = 4$, $k = 0$, $n = 10$. We consider $U_2 = \frac{(X_{(5)} - X_{(4)})}{T}$ using $Pr(U_2 \leq t)$, we note that in this case $Pr(U_2 \leq .185833) = 0.95$ given the observed values $X_{(4)} = 170, T = 1430$ this yields the prediction statement $Pr(X_{(5)} \leq 170 + (.185833)(1430)) = Pr(X_{(5)} \leq 436) = .95$. We can be 95% confident that the fifth failure will occur before 436 hours.

If $k = 1$, $T = 1430$, $t = .285555$, $X_{(5)} = 578$, we can be 95% confident that $X_{(5)}$ will not exceed 578.

References

[1]. Baker, G.A., *The probability that the mean of a second sample will differ from the mean of a first sample by less than a certain multiple of the standard deviation of the first sample.*, Annals of Mathematical Statistics **6**, **pp. 197-201. (1935).**

[2]. Bain,T.P. & Patel,K.J., *Factors for Calculating Prediction Intervals for Samples from a One-parameter Exponential distribution.*, Journal of Quality Technology **23,No.1 pp 48-52 (1991).**

[3]. Epstein,B. & Sobel,M, *Life testing*, J. Amer. Statist. Assoc. **48, pp 486-502 (1953).**

[4]. Hahn, G.J. and Nelson, W., *A Survey of prediction intervals and their applications.*, Journal of Quality Technology **5,No.4 pp 178-188 (1973).**

[5]. Kendall,M. G. and Stuart,A. S., *The Advanced Theory of Statistics.*, vol. 1 (3rd edition), Charles Griffin and Co. Ltd., Landon, 1969.

[6]. Kenneth,S.K.& Nelson,I.P., *Prediction intervals for the exponential distribution using subset data.*, Technometrics **16,No.1 pp 57-59 (1974).**

[7]. Lawless,J.F., *A Prediction Problem Concerning Samples From the Exponential Distribution, With Application in Life Testing.*, Technometrics **13 pp 725-730 (1971).**

[8]. Lawless,J.F., *On Prediction Intervals for Samples from the exponential distribution and prediction limits for system survival.*, Sankhya.B, **34,pp 1-14 (1972).**

[9]. Lawless,J.F., *Prediction intervals for the two-parameter exponential distribution.*, Technometrics **19, pp 469-472 (1977).**

[10]. Lawless,J.F., *Statistical Modles and Methods for Life Times Data.*, John Wiley& Sons,, New York, 1982.

[11]. Likeš ,J., *Prediction of sth ordered observation for the two-parameter exponential distribution.*, Technometrics **16,No.2 pp 241-244 (1974).**

[12]. Patel,J.K., *Prediction Intervals - A Review.*, Commun. Statist. Theory Meth. **18(7) pp 2393-2465 (1989).**

[13]. Tanis,E. A., *Linear Forms in the Order Statistics from an Exponential Distribution.*, Annals Math. Statist. **vol.35 pp 270-276 (1964).**

MATHEMATICS DEPT., FACULTY OF SCIENCE AT SOHAG, SOHAG 82516, EGYPT.