Aligarh Journal of Statistics
Vol. 26 (2006), 81-86

A NOTE ON THE LOSS OF INFORMATION DUE TO CENSORING – THE PARETO CASE

A.B. Aich

ABSTRACT

This article is concerned with the loss of information that arises due to Type I censoring while discriminating between two Pareto populations. The discrepancy criterion used is due to Kullback and Leibler. It is shown that the loss of information depends on the truncation time as specified by the censoring procedure. An expression for the residual duration of the experiment is also obtained. Finally, a numerical illustration with simulated data is given.

1. INTRODUCTION

The problem of selection of suitable life distribution arises frequently in reliability study. For example, one might be interested in discriminating between two exponentials, Weibull or Pareto populations. There are methods to deal with these problems based on uncensored or complete data, thereby utilizing the total information. However, the reliability of most electronic parts produced these days has reached a high level, whereas the actual observation time within a laboratory will only exceptionally be larger than some thousand hours. This necessitates censoring of observations, which is naturally accompanied by a loss of information.

In this article, we consider the problem of discrimination between two Pareto populations based on data generated by Type I censoring. In this type of censoring the experiment is conducted with a fixed number of items and continued up to a specified time called the *truncation time*, while in Type II censoring the experiment is continued till a *fixed number of failures* is obtained. Thus, in Type I censoring, information regarding the life of the items which do not fail within the truncation time will be lost. Also, for an item of high reliability this information loss increases with decreasing truncation time, as fewer failures will occur within the truncation time. It is clear that the information loss due to censoring will be nil if truncation time is set at infinity as this situation corresponds to the case of no censoring.

Pareto distribution is widely used in life study, especially in situations where there is a *guarantee time* during which no failures occur. In a recent work, Tiwari *et al*. (1996) have considered this distribution under Bayesian perspective with type-I and type-II censored samples. Upadhyay and Shastri (1997) have also studied the same distribution via Gibbs sampler with doubly type-II censored data. Park (2005) has used the Kullback-Leibler (KL) information measure in testing for exponentiality under type-II censoring. Using the same information measure, Aich (2005) has considered the problem of selecting the truncation point under type-I censoring for the Weibull failure model. In this context, one can also see Tikhov (1991).

An explicit expression for the loss of information due to censoring is obtained here. This loss is seen to depend on the truncation time, in addition to other parameters. The choice of truncation time in Type I censoring is at present arbitrary and subjective. However, the result of this article can be used in this regard and the truncation time can be set ensuring a given level of the loss of information. An expression for the residual duration of the experiment is also obtained. This is defined to be the expected additional time up to which the experiment would continue, if allowed, beyond the truncation time. This quantity can be used to measure the *saving in time* due to censoring which is the primary concern of any censored experiment. Thus, the results derived in this article appear to be of considerable significance to the reliability engineers.

2. AMOUNT OF INFORMATION IN THE CENSORED DATA

Let a component have the life X with distribution given by

$$f(x) = \theta a^{\theta} x^{-(\theta+1)}, \qquad x \geq a > 0, \ \theta > 0 \tag{2.1}$$

which is Pareto with parameters θ and a where a is interpreted as the guarantee time during which failure cannot occur. We shall assume here a to be known.

We consider the problem of discrimination of the hypothesis $H_1 : \theta = \theta_1$ against $H_2 : \theta = \theta_2$, $\theta_1 \neq \theta_2$, on the basis of censored data.

Let $E^r = [(x_1, x_2, \cdots, x_r) : a \leq x_j \leq t, 1 \leq j \leq r]$ be the subspace (of the complete sample space R^n) which is induced by the Type I censoring with truncation time t, where n and r are respectively the number of identical units (i.e., the sample size) put to test and number of failures up to the time t. Then the mean amount of information in favour of H_1 against H_2 is [Kullback, (1958). p-5]

$$\Delta(H_1 : H_2 / E^r) = E_{H_1}[\ln(l_1 / l_2)] \tag{2.2}$$

where the l_1 and l_2 are the likelihood function under H_1 and H_2, respectively.

Also, the expectation on the right of (2.2) is performed under H_1.

Here, l_1 and l_2 are obtained from $l(\theta)$ on putting $\theta = \theta_1$ and θ_2, respectively, where

$$l(\theta) = \vartheta \theta^r a^{r\theta} \left(\prod_{j=1}^{r} x_j \right)^{-(\theta+1)} (a/t)^{\theta(n-r)}, \quad \vartheta = \frac{n!}{r!(n-r)!} \tag{2.3}$$

Thus, we have the log-likelihood ratio $\ln(l_1 / l_2)$ as

$$\ln(l_1 / l_2) = r \ln(\theta_1 / \theta_2) + r(\theta_1 - \theta_2) \ln a + (\theta_2 - \theta_1) \sum_{j=1}^{r} \ln x_j$$

$$+ (n-r)(\theta_2 - \theta_1) \ln(a/t) \tag{2.4}$$

We then obtain the expected value of (2.4) under H_1 and get $\Delta(H_1 : H_2 | E^r)$. This naturally depends on r, the random number of failures within the truncation time. We replace r by $E(r) = n \Pr(X \le t) = n(1 - (a/t)^{\theta_1})$ and obtain what might be called *the expected mean amount of information* and is given by

$$\Delta(H_1 : H_2) = n(1 - (a/t)^{\theta_1})[(\theta_2 / \theta_1) - 1 + \ln(\theta_1 / \theta_2)] \tag{2.5}$$

where all the 'expectations' were performed under H_1.

The quantity $\Delta(H_1 : H_2)$ is seen to depend, among other parameters, on the truncation time t. It is a measure of the information that we utilize to discriminate between two Pareto distributions under censoring.

3. RELATIVE LOSS OF INFORMATION DUE TO CENSORING

To obtain the loss of information due to censoring, we need an expression for the total information that would have been obtained had there been no censoring. Following the approach of the previous Section, the mean amount of information for discriminating H_1 against H_2 in the absence of censoring can be seen to be

$$\Delta^*(H_1 : H_2) = n[(\theta_2 / \theta_1) - 1 + \ln(\theta_1 / \theta_2)] = \Delta_{\max} \tag{3.1}$$

This can also be obtained from (2.5) by letting $t \to \infty$, as infinite truncation time is equivalent to the case of no-censoring. Thus, the relative amount of information provided by the censored data is $\Delta(H_1 : H_2) / \Delta_{\max}$ and hence the percentage loss of information due to censoring is given by

$$\partial = 100[1 - \Delta(H_1 : H_2) / \Delta_{\max}] = 100(a/t)^{\theta_1}, \ t \ge a \tag{3.2}$$

The equation (3.2) can be used to know the loss of information corresponding to a given truncation time. Conversely, this equation can also be used to decide about the truncation time for a Type I censoring when the loss of information is pre-fixed. We thus have a rule for the selection of truncation time which at present is rather arbitrary.

4. RESIDUAL DURATION OF THE EXPERIMENT UNDER TYPE I CENSORING

A quantity which is of immense importance in the context of Type I censoring is the *residual duration of the experiment*. This may be defined as the expected additional time up to which the experiment would continue if it is allowed to do so beyond the truncation time. This quantity would actually measure the *gain in time* due to censoring. Thus, a larger value of the residual duration of the experiment would imply that a substantial gain in time has been achieved by censoring while a smaller value of the same would mean that censoring has failed to be effective in saving time. To obtain an expression for the residual duration of the experiment, we note that the overall duration of the experiment depends on the life of the most reliable item, i.e., on $X_{(n)}$ where $X_{(i)}$ is the $i-$th ordered observation in a sample of size n. Keeping this in mind and also the fact that there has been r failures up to the time t, i.e., $X_{(1)} < t, X_{(2)} < t, \cdots X_{(r-1)} < t$, $X_{(r)} \le t$, we have the residual duration of the experiment as the conditional expectation

$$E[X_{(n)} - t / X_{(r+1)} > t, X_{(r+2)} > t, \cdots, X_{(n)} > t] = E(X_{(n)} - t / X_{(r+1)} > t)$$

$$= t\left(n(a/t)^{n\theta} \sum_{j=0}^{r} \binom{n-1}{j} [(t/a)^{\theta} - 1]^{j} \tau_j - 1 \right) \tag{4.1}$$

where

$$\tau_j = B(1 - 1/\theta, n - j) / I_{1-F(t)}(n - r, r + 1)$$

and

$$I_x(p,q) = \int_0^x v^{p-1}(1-v)^{q-1} dv / B(p,q)\,,\ 0 \le x \le 1,\ \ p,q > 0\,.$$

It is assumed that $\theta > 1$. The above expression of the residual duration of the experiment can be compared with the expected duration of the experiment without censoring which is $E(X_{(n)}) = n\,a\,B(1-(1/\theta), n)$. The numerical calculation of the expression (4.1) will pose no problem as the incomplete beta function ratio has been extensively tabulated (Pearson, 1948). On noting that both t/a and θ are assumed to be greater than unity and also $I_x(p,q)$ is a proper fraction, we have the following interesting inequality:

$$E(X_{(n)} / X_{(r+1)} > t) \geq (a/t)^{n\theta-1} E(X_{(n)}) \quad (4.2)$$

As the truncation time t increases, the loss of information due to censoring decreases while the gain in time also decreases. We thus need an optimal value of t which will achieve a compromise between these two conflicting goals. One can thus minimize the loss of information subject to a given residual duration of the experiment or maximize the residual duration of the experiment for given loss. Either of these can be used as criterion for choosing t. It is intuitively clear that such choices are cost effective.

5. NUMERICAL ILLUSTRATION

The following table, reproduced from Upadhyay and Shastri (1997), gives 30 simulated observations from (2.1) with $a = 2.0$ and $\theta_1 = 2.0$. The data are written in an ascending order.

Table 1: Ordered simulated data from Pareto distribution. $(n = 30)$

2.05	2.23	2.26	2.26	2.27	2.27	2.27	2.28	2.33	2.35
2.35	2.42	2.46	2.57	2.58	2.59	2.63	2.81	3.01	3.35
3.43	3.77	4.95	5.07	5.10	5.36	6.09	6.47	8.33	14.33

(a) Let ∂ be specified as 10.0. Then we have $t = 6.32$ (in some suitable units). With this truncation point, the simulated data show that $r = 27$ and $n - r = 3$. Thus, the percentage of lost observation is $100(3/30) = 10$, which exactly coincides with specified ∂.

(b) Let ∂ be specified as 15.0. Then t becomes 5.16. For this truncation point, we have $r = 25$ and $n - r = 5$. The percentage of lost observation is $100(5/30) = 16.7$. This compares favourably with the specified ∂.

Thus, the above simulation study confirms the efficacy of formula (3.2) for selecting the truncation point.

Acknowledgement

This work is supported by a grant No. PSW –030/ 02 ERO from the University Grants Commission, India. The author is also thankful to the referee for his valuable comments which have greatly improved the presentation of the paper.

REFERENCES

Aich, A.B. (2005): A note on the loss of information due to censoring – The Weibull case. Submitted.

Kullback, Soloman (1958): *Information Theory and Statistics.* John Wiley, New York.

Park, Sangun (2005): Testing exponentiality based on the Kullback-Leibler information with type-II censored data. *IEEE Trans. Reliab.*, **54,** 22-26.

Pearson, Karl (1948): *Tables of the Incomplete Beta-Function*, Cambridge University Press.

Tikhov, M.S. (1991): On reduction of test duration in the case of censoring. *Theory Probab. Appl.*, **36**, 629-633.

Tiwari, R.C., Yang, Y. and Zalkikar, J.N. (1996): Bayes estimation of the Pareto failure-model using Gibbs sampling. *IEEE Trans. Reliab*., **45,** 471-476.

Upadhyay, S. and Shastri, V. (1997): Bayes results for classical Pareto distribution via Gibbs sampler, with doubly-censored observations. *IEEE Trans.Reliab.*, **46,** 56-59.

Received : 17-03-2005

Revised : 14-01-2006

Director, Study Centers
Netaji Subhas Open University
Kolkata-700 020, India
e-mail: abaich_nsou123@rediffmail.com