Journal of the Korean Data & Information Science Society
2012, **23**(6), 1271–1277

http://dx.doi.org/10.7465/jkdi.2012.23.6.1271
한국데이터정보과학회지

Estimation in the exponential distribution under progressive Type I interval censoring with semi-missing data

Hyejung Shin[1] · Kwangho Lee[2]

[12]Department of Statistics, Yeungnam University

Received 10 October 2012, revised 9 November 2012, accepted 14 November 2012

Abstract

In this paper, we propose an estimation method of the parameter in an exponential distribution based on a progressive Type I interval censored sample with semi-missing observation. The maximum likelihood estimator (MLE) of the parameter in the exponential distribution cannot be obtained explicitly because the intervals are not equal in length under the progressive Type I interval censored sample with semi-missing data. To obtain the MLE of the parameter for the sampling scheme, we propose a method by which progressive Type I interval censored sample with semi-missing data is converted to the progressive Type II interval censored sample. Consequently, the estimation procedures in the progressive Type II interval censored sample can be applied and we obtain the MLE of the parameter and survival function. It will be shown that the obtained estimators have good performance in terms of the mean square error (MSE) and mean integrated square error (MISE).

Keywords: Exponential distribution, maximum likelihood estimator, progressive Type I interval censored sample, progressive Type II interval censored sample, semi-missing data.

1. Introduction

Interval censored data are widely occurred in industrial, economic, medical, social science research, and lifetime studies. Of course, the progressive Type I and progressive Type II interval censored sample schemes have been applied in those areas. A number of studies have employed the progressive censored samples. The progressive Type II censored sample scheme was detailedly introduced by Balakrishnan and Aggarwala (2000). Aggarwala (2001) proposed progressive interval-censoring and Ashour and Afify (2007) studied the exponentiated Weibull family under progressive Type I interval censoring with random removal. Ng and Wang (2009) suggested some methods for estimating parameters of a Weibull distribution. Further Shin *et al.* (2010) studied the estimation problems in the exponential distribution under progressive Type I interval censored sample.

[1] Instructor, Department of Statistics, Yeungnam University, Gyeongsan 712-749, Korea.
[2] Corresponding author: Professor, Department of Statistics, Yeungnam University, Gyeongsan 712-749, Korea. E-mail: khlee@yu.ac.kr

We are interested in the estimation of the parameter and survival function for the exponential distribution under progressive Type I interval censored sample with missing data. Under the progressive Type I interval censored sample scheme, it is not permitted to observe the exact failure time, but the number of failure units within the preset time interval can only be recorded. In this paper, we consider the case that some data are missing in an interval but some of those can appear the next any interval. We will say this type of missing data as semi-missing one. General estimation method could not be applied under the progressive Type I interval censored sample with semi-missing data. Thus, we convert this kind of sample data to the progressive type II interval censored data, and obtain the MLE of the parameter and the estimator of the survival function. This paper is organized as follows. In section 2, we propose an estimation method under the progressive Type I interval censored sample with semi-missing data. In section 3, simulation studies and results are discussed. Conclusions and suggestions are given in section 4.

2. Proposed estimation method

Let T be the failure time under exponential distribution with a scale parameter σ. The probability density function (pdf) and the cumulative distribution function (cdf) as follow:

$$f(t;\sigma) = \frac{1}{\sigma} e^{-t/\sigma}, \tag{2.1}$$

$$F(t;\sigma) = 1 - e^{-t/\sigma}, \qquad t > 0,\ \sigma > 0. \tag{2.2}$$

In progressive type I interval censored sample, we can observe the number of failure units $X_1,\ X_2,\ \cdots,\ X_m$ in prefixed m intervals $(0, T_1]$, $(T_1, T_2]$, $\cdots$, $(T_{m-1}, T_m]$ for n experiment units, and remove $R_1,\ R_2,\ \cdots,\ R_m$ for remaining live units in each time $T_1,\ T_2,\ \cdots,\ T_m$, respectively. In the special case in which the interval lengths are the same, say $T_i = i \cdot t$, $i = 1, 2, \cdots, m$. An explicit expression for the MLE of σ obtained by (Aggarwala, 2001), that is

$$\hat{\sigma}_I = \frac{t}{\ln[1 + \sum_{i=1}^{m} X_i / (\sum_{i=2}^{m} (i-1) X_i + \sum_{i=1}^{m} i R_i)]}. \tag{2.3}$$

Also, Shin *et al.* (2010) generated the failure times $Y_{j:IM,n}$ based on $X_1,\ X_2,\ \cdots,\ X_m$ in progressive Type I interval censored sample. If $Y_{j:IM,n}$ of IM units are generated in each time intervals and the number of removing units IR_j are generated in $Y_{j:IM,n}$ then $Y_{j:IM,n}$ can be applied to the estimation of parameter in progressive type II censored scheme, where $j = 1, 2, \cdots, IM$, $IM = X_1 + X_2 + \cdots + X_m$, $n = IM + IR_1 + \cdots + IR_{IM}$.

Let

$$Y_{1:IM,n} \leq Y_{2:IM,n} \leq \cdots \leq Y_{IM:IM,n}. \tag{2.4}$$

Then it is simple to derive the MLE of σ (Balakrishnan and Sandu, 1996) as

$$\hat{\sigma} = \frac{1}{IM} \sum_{j=1}^{IM} (IR_j + 1) Y_{j:IM,n}. \tag{2.5}$$

In clinical or longitudinal studies involving periodic follow-ups. For example, the experimental units exist from the first time interval $(T_0, T_1]$ to the i-th time interval $(T_{i-1}, T_i]$, but they may fail to appear in subsequent intervals. If some units fail to appear in the $(i+k)$-th interval, then we cannot obtain the exact failure times because the units may have been missing in a certain intervals. In such a case, the length of the interval for the semi-missing units are not same, we cannot estimate the MLE of σ by (2.3) when some intervals have semi-missing data. To overcome the problem, we suggest another method for making failure times for progressive Type I interval censored sample with semi-missing data.

Denote $X_1, X_2, \cdots, X_m$ indicate the number of failure units; S is the number of semi-missing units, where $S < n$; R_i is the number of removing units from remaining units at times $T_1, T_2, \cdots, T_m$(m is fixed value); IR_j is the number of removing units at failure times $Y_{j:IM,n}$ for $j = 1, 2, \cdots, IM$; the total number of units in the experiment is $n = IM + IR_1 + IR_2 + \cdots + IR_{IM}$ where $IM = X_1 + X_2 + \cdots + X_m + S$. And Y_j is generated failure times in each interval. We use $X_1, X_2, \cdots, X_m$ and S to generate Y_j and IR_j units in each interval.

Now, we assume that there are semi-missing units in some intervals. For example, if a semi-missing unit occurs in the first interval $(T_0, T_1]$, then the unit is observed in the second interval $(T_1, T_2]$. In this case, we obtain the failure times Y_j as follows;

$$
\begin{aligned}
Y_1 &= T_0 + \frac{T_1 - T_0}{X_1 + 1},\ IR_1 = 0, \\
&\vdots \\
Y_{X_1 - 1} &= T_0 + (X_1 - 1)\frac{T_1 - T_0}{X_1 + 1},\ IR_{X_1 - 1} = 0, \\
Y_{X_1} &= T_0 + X_1 \frac{T_1 - T_0}{X_1 + 1},\ IR_{X_1} = R_1.
\end{aligned}
$$

And if the missing unit Y_{j+S} reappear in the second interval $(T_1, T_2]$, then

$$
\begin{aligned}
Y_{X_1 + S} &= \frac{T_2 - T_0}{2} = T_1, IR_{X_1 + S} = 0, \\
Y_{X_1 + 1 + S} &= T_1 + \frac{T_2 - T_1}{X_2 + 1},\ IR_{X_1 + 1 + S} = 0, \\
&\vdots \\
Y_{X_1 + X_2 - 1 + S} &= T_1 + (X_2 - 1)\frac{T_2 - T_1}{X_2 + 1},\ IR_{X_1 + X_2 - 1 + S} = 0, \\
Y_{X_1 + X_2 + S} &= T_1 + X_2 \frac{T_2 - T_1}{X_2 + 1},\ IR_{X_1 + X_2 + S} = R_2.
\end{aligned}
$$

Similarly, Y_j are obtained for the remaining intervals, and the final interval $(T_{m-1}, T_m]$

$$
\begin{aligned}
Y_{X_1 + X_2 + \cdots + X_{m-1} + 1 + S} &= T_{m-1} + \frac{T_m - T_{m-1}}{X_m + 1}, \\
&\quad IR_{X_1 + X_2 + \cdots + X_{m-1} + 1 + S} = 0,
\end{aligned}
$$

$$\begin{aligned}
&\vdots \\
Y_{X_1+X_2+\cdots+X_{m-1}+X_m-1+S} &= T_{m-1} + (X_m - 1)\frac{T_m - T_{m-1}}{X_m + 1}, \\
&\quad IR_{X_1+X_2+\cdots+X_{m-1}+X_m-1+S} = 0, \\
Y_{X_1+X_2+\cdots+X_{m-1}+X_m+S} &= T_{m-1} + X_m\frac{T_m - T_{m-1}}{X_m + 1}, \\
&\quad IR_{X_1+X_2+\cdots+X_{m-1}+X_m+S} = R_m,
\end{aligned}$$

where $S = 1$.

In general, if there are semi-missing units between the i-th and the $(i+k-1)$-th intervals, then those units are observed in the subsequent interval. Therefore, in the first interval $(T_0, T_1]$, Y_j are obtained by

$$\begin{aligned}
Y_1 &= T_0 + \frac{T_1 - T_0}{X_1 + 1},\ IR_1 = 0, \\
&\vdots \\
Y_{X_1-1} &= T_0 + (X_1 - 1)\frac{T_1 - T_0}{X_1 + 1},\ IR_{X_1-1} = 0, \\
Y_{X_1} &= T_0 + X_1\frac{T_1 - T_0}{X_1 + 1},\ IR_{X_1} = R_1.
\end{aligned}$$

In the second interval $(T_1, T_2]$,

$$\begin{aligned}
Y_{X_1+1} &= T_1 + \frac{T_2 - T_1}{X_2 + 1},\ IR_{X_1+1} = 0, \\
&\vdots \\
Y_{X_1+X_2-1} &= T_1 + (X_2 - 1)\frac{T_2 - T_1}{X_2 + 1},\ IR_{X_1+X_2-1} = 0, \\
Y_{X_1+X_2} &= T_1 + X_2\frac{T_2 - T_1}{X_2 + 1},\ IR_{X_1+X_2} = R_2.
\end{aligned}$$

In the same way, Y_j are obtained for the remaining intervals. But if the semi-missing units S show up in $(i+k)$-th interval $(T_{i+k-1}, T_{i+k}]$, Y_j and Y_{j+S} are

$$\begin{aligned}
Y_{X_1+X_2+\cdots+X_{i+k-1}+1} &= \frac{T_{i+k} - T_i}{2}, \\
&\quad IR_{X_1+X_2+\cdots+X_{i+k-1}+1} = 0, \\
&\vdots \\
Y_{X_1+X_2+\cdots+X_{i+k-1}+S} &= \frac{T_{i+k} - T_i}{2}, \\
&\quad IR_{X_1+X_2+\cdots+X_{i+k-1}+S} = 0,
\end{aligned}$$

$$
\begin{aligned}
Y_{X_1+X_2+\cdots+X_{i+k-1}+1+S} &= T_{i+k-1} + \frac{T_{i+k}-T_{i+k-1}}{X_{i+k}+1}, \\
&\quad IR_{X_1+X_2+\cdots+X_{i+k-1}+1+S} = 0, \\
&\vdots \\
Y_{X_1+X_2+\cdots+X_{i+k-1}+X_{i+k}-1+S} &= T_{i+k-1} + (X_{i+k}-1)\frac{T_{i+k}-T_{i+k-1}}{X_{i+k}+1}, \\
&\quad IR_{X_1+X_2+\cdots+X_{i+k-1}+X_{i+k}-1+S} = 0, \\
Y_{X_1+X_2+\cdots+X_{i+k-1}+X_{i+k}+S} &= T_{i+k-1} + X_{i+k}\frac{T_{i+k}-T_{i+k-1}}{X_{i+k}+1}, \\
&\quad IR_{X_1+X_2+\cdots+X_{i+k-1}+X_{i+k}+S} = R_{i+k}.
\end{aligned}
$$

In the final interval $(T_{m-1}, T_m]$, Y_j are

$$
\begin{aligned}
Y_{X_1+X_2+\cdots+X_{m-1}+1+S} &= T_{m-1} + \frac{T_m-T_{m-1}}{X_m+1}, \\
&\quad IR_{X_1+X_2+\cdots+X_{m-1}+1+S} = 0, \\
&\vdots \\
Y_{X_1+X_2+\cdots+X_{m-1}+X_m-1+S} &= T_{m-1} + (X_m-1)\frac{T_m-T_{m-1}}{X_m+1}, \\
&\quad IR_{X_1+X_2+\cdots+X_{m-1}+X_m-1+S} = 0, \\
Y_{X_1+X_2+\cdots+X_{m-1}+X_m+S} &= T_{m-1} + X_m\frac{T_m-T_{m-1}}{X_m+1}, \\
&\quad IR_{X_1+X_2+\cdots+X_{m-1}+X_m+S} = R_m,
\end{aligned}
$$

for $i = 1, 2, \ldots, m-1$, $k \leq i$, $S < n$.

As above, all the failure times Y_j are obtained and let $Y_{j:IM,n}$ be ordered failure time Y_j. Using the progressive Type II censored scheme, we propose an estimator $\hat{\sigma}_{II_M}$ of the scale parameter σ under the exponential distribution by (2.5).

$$
\hat{\sigma}_{II_M} = \frac{1}{IM}\sum_{j=1}^{IM}(IR_j+1)Y_{j:IM,n}, \tag{2.6}
$$

where $IM = X_1 + X_2 + \cdots + X_m + S$.

Also we are interested in a survival function of time t. Consider a survival function of time t,

$$
S(t) = 1 - F(t) = P(T_i > t), \qquad 0 < t < \infty, \tag{2.7}
$$

where T_i is a nonnegative random variable. Using (2.2) and (2.7), the survival function of t under exponential distribution is derive to

$$
S(t) = e^{-t/\sigma}. \tag{2.8}
$$

Thus we obtain $\hat{S}_{II_M}(t)$ by using $\hat{\sigma}_{II_M}$ from (2.8) as

$$
\hat{S}_{II_M}(t) = e^{-t/\hat{\sigma}_{II_M}}. \tag{2.9}
$$

3. Simulation studies

3.1. Generating the failure times

We perform simulation studies about the numerical performance of the estimator and the survival function by using the proposed estimator. For estimation, first, we obtain the number of failure times for progressive Type I interval censored sample by the simulation algorithm (Aggarwala, 2001). Second, we determine the failure times, estimate the parameter and the survival function based on the progressive Type II censored scheme.

The starting point is $T_0 = 0$, and the end point is T_m. The number of failure times are obtained at preset times $T_1, T_2, \ldots, T_m$ (m is fixed value) in each time intervals, an exponential distribution with given fixed percentages $p_1, p_2, \cdots, p_{m-1}$ ($p_m = 1$) of units to be removed. Failure times $X_1, X_2, \cdots, X_m$ for the progressive Type I interval censored sample are generated by the algorithm (Aggarwala, 2001).

In this paper, we calculate the proposed estimator $\hat{\sigma}_{II_M}$ and $\hat{S}_{II_M}(t)$. To compare the numerical performance of the each estimators, we compute the mean square error (MSE) for the estimator of a scale parameter and the mean integrated square error (MISE) for the estimator of the survival function.

3.2. Numerical results

We perform the simulation algorithm under the exponential distribution with the scale parameter $\sigma = 3$. We consider the sample sizes $n = 20, 30$, and 50, the number of intervals $m = 10$, and the time interval $t = 1$ and let the percentage of removed units be

$$P_A = (0, 0, 0, 0, 0, 0, 0, 0, 0, 1),$$
$$P_B = (0.25, 0.25, 0.25, 0.25, 0, 0, 0, 0, 0, 1),$$
$$P_C = (0.5, 0.5, 0.5, 0.5, 0, 0, 0, 0, 0, 1).$$

To evaluate the performance of the proposed estimators, we simulate the MSE and the MISE for the proposed estimator from the Monte Carlo method, the simulation procedure was repeated 1,000 times.

We estimate the bias and MSE of $\hat{\sigma}_{II_M}$. In addition to, we examine in the MISE of $\hat{S}_{II_M}(t)$. As expected, the MSE and MISE for the proposed estimator decrease as the sample size n increase. For example, two missing units in $(T_0, T_1]$ are observed in $(T_1, T_2]$. According to Table 3.1, which shows the bias and MSE for $\hat{\sigma}_{II_M}$, the MISE of $\hat{\sigma}_{II_M}$. Although the two units are not observed in the first interval, the simulation results are efficient.

Table 3.1 The bias and MSE of $\hat{\sigma}_{II_M}$, and the MISE of $\hat{S}_{II_M}(t)$

		$\hat{\sigma}_{II_M}$		$\hat{S}_{II_M}(t)$
n	p	Bias	MSE	MISE
20	p_A	-0.016572	0.446666	0.003434686
	p_B	-0.104048	0.535993	0.004299477
	p_C	-0.263467	0.641896	0.005777040
30	p_A	0.002090	0.291217	0.002293241
	p_B	-0.074248	0.373977	0.002974430
	p_C	-0.201310	0.500820	0.004270334
50	p_A	-0.007521	0.177879	0.001415436
	p_B	-0.064594	0.240818	0.001957395
	p_C	-0.123487	0.317347	0.002609882

4. Conclusions

We consider that some intervals have semi-missing data in progressive Type I interval censored sample, and propose the estimation method of those sample. To evaluate the performance of the proposed method, we obtain the number of failure times for the progressive Type I interval censored sample and determine failure times. Thus, the obtained sample is the same as the progressive Type II censored sample. We estimate the parameter and the survival function. For this, we use the exponential distribution with an unknown scale parameter, and the estimator of parameter is based on that method. The results of the simulation studies indicate as follows: The proposed estimator $\hat{\sigma}_{II_M}$ is shown good performance and those have a consistency in terms of the MSE and bias. Also, the proposed estimator $\hat{S}_{II_M}(t)$ is shown good performance and those have a consistency in terms of the MISE.

A few things add to the conclusion, we will consider to shift removal points of units in next study. Also, we will give variety to the time-interval-lengths and distribution.

References

Aggarwala, R. (2001). Progressive interval censoring: Some mathematical results with applications to inference. *Communications in Statistics-Theory and Methods,* **30**, 1921–1935.

Amin, Z. H. (2008). A note on the parameter estimation for the log-normal distribution based on progressively Type I interval censored samples. *Model Assisted Statistics Application,* **3**, 169–176.

Ashour, S. K. and Afify, W. M. (2007). Statistical analysis exponentiated Weibull family under Type I progressive interval censoring with random removal. *Journal of Applied Sciences Research,* **3**, 1851-1863.

Balakrishnan, N. and Aggarwala, R. (2000). *Progressive censoring: Theory, methods and applications,* Birkhauser, Boston.

Balakrishnan, N. and Sandu, R. A. (1996). Best linear unbiased and maximum likelihood estimation for exponential distributions under general progressive Type II censored samples. *Sankhya: The Indian Journal of Statistics*, **58**, 1–9.

Ng, T. H. K. and Wang, Z. (2009). Statistical estimation for the parameters of Weibull distribution based on progressively Type-I interval censored sample. *Journal of Statistical Computation and Simulation,* **79**, 145–159.

Shin, H. J., Lee, K. H. and Cho, Y. S. (2010). Parameter estimation for exponential distribution under progressive type I interval censoring. *Journal of the Korean Data & Information Science Society,* **21**, 927-934.