Analysis of Hybrid Censored Data from the Lognormal Distribution

A. Habibi Rad[†,*] and F. Yousefzadeh[‡]

[†] Ferdowsi University of Mashhad

[‡] Birjand University

Abstract. The mixture of Type I and Type II censoring schemes, called the hybrid censoring. This article presents the statistical inferences on lognormal parameters when the data are hybrid censored. We obtain the maximum likelihood estimators (MLEs) and the approximate maximum likelihood estimators (AMLEs) of the unknown parameters. Asymptotic distributions of the maximum likelihood estimators are used to construct approximate confidence intervals. Monte Carlo simulations are performed to compare the performances of the different methods and one data set is analyzed for illustrative purposes.

Keywords. Approximate maximum likelihood estimate; asymptotic distribution; hybrid censoring; maximum likelihood estimate; Monte Carlo simulation.

1 Introduction

The two most common censoring schemes are termed as Type I and Type II censoring schemes. Briefly, they can be described as follows. Consider n items under observations in a particular experiment. In the conventional Type I censoring scheme, the experiment continues up to a pre-specified time T. On the other hand, the conventional Type II censoring scheme requires the experiment to continue until a pre-specified number of failures $R(\leqslant n)$ occur. The mixture of Type I and Type II censoring schemes is known as the

---

[*] Corresponding author

hybrid censoring scheme. Suppose n identical units are put to test under the same environmental conditions and the lifetime of each unit is independent and identically distributed ($i.i.d.$) random variables. The test is terminated when a pre-chosen number R, out of n items have failed or a pre-determined time T, on test has been reached.

Therefore, in the presence of hybrid censoring schemes, we have one of the following types of observations:

Case I: $\{x_{1:n} < \cdots < x_{R:n} < T\}$

Case II: $\{x_{1:n} < \cdots < x_{d:n} < T\}$,   if $0 \leqslant d < R$ and $x_{d:n} < T < x_{d+1:n}$,

here $X_{1:n} < X_{2:n} < ...$ denote the observed failure times of the experimental units.

The hybrid censoring scheme was first introduced by Epstein (1954), but recently it becomes quite popular in the reliability and life-testing experiments. Chen and Bhattacharya (1998) obtained the exact distribution of the conditional maximum likelihood estimator of and proposed a one-sided confidence interval. Draper and Guttman (1987) considered this problem from the Bayesian point of view and obtained the two sided credible intervals of the mean lifetime using inverted Gamma prior. Comparison of the different methods can be found in Gupta and Kundu (1998). For some related work, one may refer to Ebrahimi (1990, 1992) and Jeong et al. (1996). Some simplifications of the exact distribution have been suggested by Childs et al. (2003). Recently, Kundu (2007) and Banerjee and Kundu (2008) both investigated a hybrid censoring scheme, in which the life-distribution is Weibull, also Kundu and Pradhan (2009) worked on hybrid censoring data based on the generelized exponential distribution.

In this paper we consider the analysis of the hybrid censored lifetime data when the lifetime of each experimental unit follows a lognormal distribution because of its flexibility and wide scale applicability.

The rest of the article is organized as follows. In Section 2, we describe the model and the maximum likelihood estimators. Approximate maximum likelihood estimators of the unknown parameters are provided in Sections 3. Simulation results are presented in Section 4 and one real data set has been analyzed in Section 5. Finally conclusions appear in Section 6.

2   Maximum Likelihood Estimator

Suppose the lifetime random variable Y has a lognormal distribution with μ and σ parameters with probability density function (pdf) as

$$f(x; \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma x} e^{-\frac{(\ln x - \mu)^2}{2\sigma^2}} \qquad x > 0, \ -\infty < \mu < \infty, \ \sigma > 0. \qquad (1)$$

In this section we provide the maximum likelihood estimators (MLE) of the unknown parameters. The likelihood function based on the observed data, for Case I is

$$l(\mu, \sigma) = \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^R \prod_{i=1}^{R} \left(\frac{1}{x_{i:n}}\right) e^{-\sum_{i=1}^{R} \frac{(\ln x_{i:n} - \mu)^2}{2\sigma^2}} \left\{\bar{F}(x_{R:n})\right\}^{n-R}, \qquad (2)$$

and for Case II is

$$l(\mu, \sigma) = \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^d \prod_{i=1}^{d} \left(\frac{1}{x_{i:n}}\right) e^{-\sum_{i=1}^{d} \frac{(\ln x_{i:n} - \mu)^2}{2\sigma^2}} \left\{\bar{F}(T)\right\}^{n-d}, \qquad (3)$$

where F is the cumulative distribution function (cdf) of lognormal distribution. The logarithm of (2) and (3) can be written as

$$L(\mu, \sigma) = R \ln(\sqrt{2\pi}\sigma) - \sum_{i=1}^{R} \ln x_{i:n} - \sum_{i=1}^{R} \frac{(\ln x_{i:n} - \mu)^2}{2\sigma^2}$$
$$+ (n - R) \ln \left\{ \bar{\Phi} \left(\frac{\ln x_{R:n} - \mu}{\sigma} \right) \right\}, \qquad (4)$$

$$L(\mu, \sigma) = d \ln(\sqrt{2\pi}\sigma) - \sum_{i=1}^{d} \ln x_{i:n} - \sum_{i=1}^{d} \frac{(\ln x_{i:n} - \mu)^2}{2\sigma^2}$$
$$+ (n - d) \ln \left\{ \bar{\Phi} \left(\frac{\ln T - \mu}{\sigma} \right) \right\}, \qquad (5)$$

where Φ is the cumulative distribution function (cdf) of the standard Normal distribution.

Taking derivatives with respect to μ and σ of (4) and putting then equal to zero we obtain

$$\frac{\partial L}{\partial \mu} = \sum_{i=1}^{R} \frac{\ln x_{i:n} - \mu}{\sigma^2} + \frac{n - R}{\sigma} \cdot \frac{\phi\left(\frac{\ln x_{R:n} - \mu}{\sigma}\right)}{\bar{\Phi}\left(\frac{\ln x_{R:n} - \mu}{\sigma}\right)} = 0, \qquad (6)$$

www.SID.ir

$$\frac{\partial L}{\partial \sigma} = \frac{R}{\sigma} + \sum_{i=1}^{R} \frac{(\ln x_{i:n} - \mu)^2}{\sigma^3} + \frac{(n-R)(\ln x_{R:n} - \mu)}{\sigma^2} \cdot \frac{\phi\left(\frac{\ln x_{R:n}-\mu}{\sigma}\right)}{\bar{\Phi}\left(\frac{\ln x_{R:n}-\mu}{\sigma}\right)} = 0, \quad (7)$$

where ϕ is the pdf of the standard normal distribution.

From (6), we have

$$-\sum_{i=1}^{R} \frac{\ln x_{i:n} - \mu}{\sigma} = (n-R)\frac{\phi\left(\frac{\ln x_{R:n}-\mu}{\sigma}\right)}{\bar{\Phi}\left(\frac{\ln x_{R:n}-\mu}{\sigma}\right)}. \tag{8}$$

Using (8) in (7) we obtain

$$R + \sum_{i=1}^{R} \frac{(\ln x_{i:n} - \mu)^2}{\sigma^2} + \frac{(\ln x_{R:n} - \mu)}{\sigma}\left(-\sum_{i=1}^{R} \frac{\ln x_{i:n} - \mu}{\sigma}\right) = 0. \tag{9}$$

So, we have

$$\sigma = \sqrt{\frac{(\ln x_{R:n} - \mu)\left(\sum\limits_{i=1}^{R} \ln x_{i:n} - \mu\right) - \sum\limits_{i=1}^{R}(\ln x_{i:n} - \mu)^2}{R}} = u(\mu), \tag{10}$$

using (10) in (8) we obtain

$$\mu = h(\mu), \tag{11}$$

where

$$h(\mu) = \frac{1}{R}\left\{(n-R)u(\mu)\frac{\phi\left(\frac{\ln x_{R:n}-\mu}{u(\mu)}\right)}{\bar{\Phi}\left(\frac{\ln x_{R:n}-\mu}{u(\mu)}\right)} + \sum_{i=1}^{R} \ln x_{i:n}\right\}.$$

We propose a simple iterative scheme to solve (11) for μ. Start with an initial guess of μ, say $\mu^{(0)}$, obtain $\mu^{(1)} = h(\mu^{(0)})$ and proceeding in this way obtain $\mu^{(n+1)} = h(\mu^{(n)})$. Stop the iterative procedure, when $|\mu^{(n+1)} - \mu^{(n)}| < \epsilon$, some pre-assigned tolerance limit.

The likelihood equation in Case II (5) can be re-written as

$$\sigma = \sqrt{\frac{(\ln T - \mu)\left(\sum\limits_{i=1}^{d} \ln x_{i:n} - \mu\right) - \sum\limits_{i=1}^{d}(\ln x_{i:n} - \mu)^2}{d}} = u(\mu). \tag{12}$$

Using (10) in (8) we obtain

$$\mu = h(\mu) \tag{13}$$

where

$$h(\mu) = \frac{1}{d} \left\{ (n-d)u(\mu) \frac{\phi\left(\frac{\ln T - \mu}{u(\mu)}\right)}{\bar{\Phi}\left(\frac{\ln T - \mu}{u(\mu)}\right)} + \sum_{i=1}^{d} \ln x_{i:n} \right\}.$$

Similar procedure as above can be used to solve for μ.

Since the MLEs when they exist, are not in compact forms, so the next section we propose the approximate maximum likelihood estimates which have explicit forms.

3 Approximate Maximum Likelihood Estimation

In this section, we use the Approximate Maximum Likelihood Estimation method (AMLE) to estimate the scale and location parameters μ and σ. We first consider AMLEs for Case I:

We expand the function $\frac{\phi(z_{i:n})}{\bar{\Phi}(z_{i:n})}$, where $z_{i:n} = \frac{\ln x_{i:n} - \mu}{\sigma}$, $i = 1, \ldots, R$, in Taylor series around the point $\Phi^{-1}(p_R) = \xi_R$ where $p_R = \frac{R}{n+1}$.

Note that

$$\frac{\phi(z_{R:n})}{\bar{\Phi}(z_{R:n})} \approx \alpha + \beta z_{R:n}, \tag{14}$$

where

$$\alpha = \frac{\phi(\xi_R)}{\bar{\Phi}(\xi_R)} - \xi_R \left[-\xi_R \frac{\phi(\xi_R)}{\bar{\Phi}(\xi_R)} + \left\{ \frac{\phi(\xi_R)}{\bar{\Phi}(\xi_R)} \right\}^2 \right],$$

$$\beta = -\xi_R \frac{\phi(\xi_R)}{\bar{\Phi}(\xi_R)} + \left\{ \frac{\phi(\xi_R)}{\bar{\Phi}(\xi_R)} \right\}^2.$$

Using the approximation (14) in (6) and (7), we have

$$\frac{\partial L}{\partial \mu} \approx \sum_{i=1}^{R} z_{i:n} + (n-R)(\alpha + \beta z_{R:n}) = 0, \tag{15}$$

$$\frac{\partial L}{\partial \sigma} \approx -R + \sum_{i=1}^{R} z_{i:n}^2 + (n-R)z_{R:n}(\alpha + \beta z_{R:n}) = 0. \tag{16}$$

From (15), we obtain solution of $\hat{\mu}$ as

$$\hat{\mu} = A_1 + \hat{\sigma} A_2, \tag{17}$$

where

$$A_1 = \frac{\sum\limits_{i=1}^{R} y_{i:n} + \beta y_{R:n}}{R + (n-R)\beta},$$

$$A_2 = \frac{(n-R)\alpha}{R + (n-R)\beta},$$

where $y_{i:n} = \ln x_{i:n}, \ i = 1, \ldots, R$.

From (16), we obtain $\hat{\sigma}$ as a solution of the quadratic equation

$$B_1\sigma^2 + B_2\sigma + B_3 = 0, \tag{18}$$

where

$$B_1 = -R,$$

$$B_2 = 2A_1A_2R - 2A_2\sum_{i=1}^{R} y_{i:n} + \alpha(n-R)(y_{R:n} - A_1) - 2(n-R)\beta(y_{R:n} - A_1)A_2,$$

$$B_3 = \sum_{i=1}^{R}(y_{i:n} - A_1)^2 + (n-R)\beta(y_{R:n} - A_1)^2 > 0.$$

Therefore

$$\hat{\sigma} = \frac{B_2 + \sqrt{B_2^2 + 4RB_3}}{2R}, \tag{19}$$

is the only positive root.

Similarly for Case II: We expand the function $\frac{\phi(V)}{\Phi(V)}$, where $V = \frac{\ln T - \mu}{\sigma}$, in Taylor series around the point $\Phi^{-1}(p_d) = \xi_d$ where $p_d = \frac{d}{n+1}$.

Therefore, in this case we obtain the solutions of $\hat{\mu}$ and $\hat{\sigma}$ as the same steps as before

$$\hat{\mu} = A_1 + \hat{\sigma}A_2, \tag{20}$$

where

$$A_1 = \frac{\sum\limits_{i=1}^{d} y_{i:n} + \beta \ln T}{d + (n-d)\beta},$$

$$A_2 = \frac{(n-d)\alpha}{d + (n-d)\beta},$$

and

$$\hat{\sigma} = \frac{B_2 + \sqrt{B_2^2 + 4dB_3}}{2d}, \tag{21}$$

where

$$B_1 = -d,$$

$$B_2 = 2A_1A_2d - 2A_2\sum_{i=1}^{d}y_{i:n} + \alpha(n-d)(\ln T - A_1) - 2(n-d)\beta(\ln T - A_1)A_2,$$

$$B_3 = \sum_{i=1}^{d}(y_{i:n} - A_1)^2 + (n-d)\beta(\ln T - A_1)^2 > 0.$$

4 Numerical Results

In this section we present some simulation results to compare the performances of the different methods proposed in the previous sections. We take $\mu = 0$ and $\sigma = 1$ in all the ceases and we mainly compare the performances of the MLEs and AMLEs estimators of the unknown parameters, in terms of their average bias, the mean squared error, average confidence width and coverage percentage for different choices of $n(30, 40)$ and R when $T = 1$. All the computations are performed by using R software. The results are reported in Tables 1 and 2.

From Tables 1 and 2, we observed that the average biases for $(\hat{\mu}, \hat{\sigma})$ based on the MLE give smaller to those based on the AMLE but they are very close to each other, especially based on their mean squared error.

Now we compare average confidence widths and coverage percentages. In general it is observed that the two methods work well unless R values be very small respect to n.

5 Data Analysis

For illustrative purposes, we present here a data analysis using the proposed methods. The data set is taken from Lawless (1982, p. 491) and consists of failure times for 36 appliances subject to an automatic life test.

The ordered data are as follows: 11, 35, 49, 170, 329, 381, 708, 958, 1062, 1167, 1594, 1925, 1990, 2223, 2327, 2400, 2451, 2471, 2551, 2565, 2568, 2694, 2702, 2761, 2831, 3034, 3059, 3112, 3214, 3478, 3504, 4329, 6367, 6976, 7846, 13403.

www.SID.ir

Table 1. The average biases, the mean squared errors, average confidence widths and coverage percentages for $n = 30$ and $T = 1$

R		MLE				AMLE			
20	μ	0.011	0.056	0.092	0.936	0.0126	0.055	0.091	0.926
	σ	-0.018	0.044	0.081	0.830	-0.019	0.044	0.081	0.830
25	μ	0.010	0.055	0.091	0.943	0.014	0.056	0.092	0.915
	σ	-0.017	0.043	0.081	0.846	-0.016	0.044	0.081	0.862
30	μ	0.012	0.056	0.092	0.918	0.011	0.055	0.092	0.934
	σ	-0.017	0.043	0.081	0.855	-0.017	0.043	0.081	0.856

Table 2. The average biases, the mean squared errors, average confidence widths and coverage percentages for $n = 40$ and $T = 1$

R		MLE				AMLE			
25	μ	0.008	0.041	0.079	0.933	0.009	0.040	0.078	0.908
	σ	-0.013	0.033	0.070	0.862	-0.013	0.032	0.070	0.872
30	μ	0.008	0.040	0.078	0.931	0.009	0.040	0.078	0.937
	σ	-0.012	0.032	0.070	0.883	-0.013	0.032	0.069	0.871
40	μ	0.009	0.040	0.078	0.923	0.008	0.040	0.078	0.942
	σ	-0.012	0.032	0.071	0.879	-0.013	0.032	0.070	0.883

We have created an artificial hybrid censored data set from the above uncensored data set. In this case we let $R = 20$, $T = 2600$. Then we have estimated the unknown parameters using the MLEs and AMLEs. The MLEs of μ and σ are 7.8630 and 2.0472, respectively. The confidence widthes of μ and σ based on the MLEs are 2.087759 and 0.8967767, respectively. Similarly, The AMLEs of μ and σ are 5.760958 and 3.095316, respectively. The confidence widthes of μ and σ based on the AMLEs are 11.62383 and 1.391108, respectively.

6   Conclusions

In this paper we have considered the classical inference for the unknown parameters of the lognormal distribution when the data are hybrid censored. It is observed that the MLE method can not be obtained in closed form. But it is shown that the MLE estimators can be obtained by using a simple iterative procedure and the proposed AMLE estimators can be obtained in

explicit forms. From the simulation study it is observed that the AMLE performs well as the MLE and they are very closed to each other. Also we found for fixed n and T as R increases, no specific pattern observed in MSE, because of in this case no additional information are gathered, finally one data set analyzed for illustrative purposes.

Acknowledgment

References

Banerjee, A. and Kundu, D. (2008). Inference based on Type-II hybrid censored data from Weibull distribution, *IEEE Transactions on Reliability*, **57**, 369-378.

Chen, S. and Bhattacharya, G.K. (1998). Exact confidence bounds for an exponential parameter under hybrid censoring, *Communications in Statistics-Theory and Methods*, **17**, 1857-1870.

Childs, A., Chandrasekhar, B., Balakrishnan, N. and Kundu, D. (2003). Exact likelihood inference based on Type-I and Type-II hybrid censored samples from the exponential distribution, *Annals of the Institute of Statistical Mathematics*, **55**, 319-330.

Draper, N. and Guttman, I. (1987), Bayesian analysis of hybrid life-test with exponential failure times, *Annals of the Institute of Statistical Mathematics*, **39**, 219-255.

Epstein, B. (1954). Truncated life tests in the exponential case, *Annals of Mathematical Statistics*, **25**, 555-564.

Ebrahimi, N. (1990). Estimating the parameter of an exponential distribution from hybrid life test, *Journal of Statistical Planning and Inference,* **23**, 255-261.

Ebrahimi, N. (1992). Prediction intervals for further failures in exponential distribution under hybrid censoring, *IEEE Transactions on Reliability*, **41**, 127-132.

Gupta, R.D. and Kundu, D. (1998). Hybrid censoring schemes with exponential failure distribution, *Communications in Statistics-Theory and Methods*, **27**, 3065-3083.

Jeong, H.S., Park, J.I. and Yum, B.J. (1996). Development of (r,T) hybrid sampling plans for exponential lifetime distributions, *Journal of Applied Statistics*, **23**, 601-607.

www.SID.ir

Kundu, D. (2007). On hybrid censored Weibull distribution, *Journal of Statistical Planning and Inference*, **137**, 2127-2142.

Kundu, D. and Pradhan, B. (2009). Estimating the parameters of the generalized exponential distribution in presence of hybrid censoring, *Communications in Statistics, Theory and Methods*, **38**, 2030-2041.

Lawless, J.F. (1982), *Statistical Models and Methods for Lifetime Data*, John Wiley & sons, New York.

A. Habibi Rad
Department of Statistics,
School of Mathematical Science,
Ferdowsi University of Mashhad,
Mashhad, Iran.
P.O. Box 91775-1159
email: *ahabibi@um.ac.ir*

F. Yousefzadeh
Department of Statistics,
School of Science,
Birjand University,
Birjand, Iran.
P.O. Box 97175-615