On Type-II Hybrid Censored Lindley Distribution

Bhupendra Singh[1], Puneet Kumar Gupta[2], Vikas Kumar Sharma[*3]

[1,2]Department of Statistics, Chaudhary Charan Singh University, Meerut-250005, India
[*3]Department of Statistics and DST-CIMS, Banaras Hindu University, Varanasi-221005, India

[*3]vikasstats@rediffmail.com

Received 4 October, 2013; Revised 10 January, 2014; Accepted 20 January, 2014; Published 18 May, 2014

Abstract

The present paper deals with the inferential study of Lindley distribution when the data are type-II hybrid censored. In classical set up, the maximum likelihood estimate of the parameter with its standard error are computed. Further, by assuming Jeffrey's invariant and gamma priors for the unknown parameter, Bayes estimate along with its posterior standard error and highest posterior density credible interval of the parameter are obtained. Markov Chain Monte Carlo technique such as Metropolis-Hastings algorithm has been utilized to simulate draws from the posterior density of the parameter. Finally, a real data study is conducted for illustrative purpose.

Keywords

Lindley Distribution; Type-II Hybrid Censoring; Maximum Likelihood Estimate; Bayes Estimate; Metropolis-Hastings Algorithm; Highest Posterior Density Credible Interval

Introduction

Lindley distribution was introduced by Lindley (1958) in the context of Bayesian statistics, as a counter example of fudicial statistics. Ghitany et al. (2008) observed that this distribution can be quite effectively used in lifetime experiments, particularly as an alternative of exponential distribution, as it also has only scale parameter. More so, in real world, we rarely encounter the engineering systems which have constant hazard rate through their life span. Therefore, it seems practical to assume hazard rate as a function of time. Lindley distribution is one of the distributions, having time-dependent hazard rate. Recently, several authors like Ghitany et al. (2008), Krishna and Kumar (2011), Mazucheli and Achcar (2011), Singh and Gupta (2012) and Al-Mutairi et al. (2013) has investigated various inference problems using Lindley distribution as a lifetime model. On the other hand, the some extension or modification of the Lindley distribution like, Ghitany et. al. (2011), Gomoz-Deniz and Calderin-Ojeda (2011), Ghitany et al. (2013) are also available in the reliability/survival literature.

In reliability literature, Type-I and Type-II censoring are the most regularly used censoring schemes. The mixture of Type-I and Type-II censoring scheme is known as hybrid censoring scheme. In this censoring scheme, n items are put on test and the test is terminated when the pre-chosen number R out of n items are failed or when a pre-decided time T on the test has been reached. In other words, we can say that the termination point of the test is $T^{\dagger} = \min\{X_{R:n}, T\}$. From now on, we call this the Type-I censoring scheme. Epstein (1954) was the first to introduce this Type-I hybrid censoring and it is quite applicable in reliability acceptance test in MIL-STD-781C (1977). After words, Type-I hybrid censoring scheme is used by many authors like Draper and Guttmen (1987), Chen and Bhattacharya (1988), Ebrhimi (1998), Childs et al. (2003), Kundu (2007) and Gupta and Singh (2013).

In Type-I hybrid censoring scheme, we assume that the number of observed failures is at least one. Also, there may be very few failures occurring up to the pre-decided time T. In view of this, Childs et al. (2003) proposed a new hybrid censoring scheme called Type-II hybrid censoring scheme described as follows: let n identical items are put on test then terminate the experiment at the random time $T^{*} = \max\{X_{R:n}, T\}$. So, we have ensured that at least R failure is observed. For more details on estimation under Type-II hybrid censoring scheme refer to Child et al. (2003) , Banarjee and Kundu (2008) and Singh et al. (2013). For a comprehensive review of various hybrid censoring scheme, see Balakrishnan an Kundu (2013).

In lieu of above considerations, the paper is organized as follows. In section 2, we describe the model by assuming Type-II hybrid censored data from Lindley distribution. Section 3 devoted to the maximum likelihood estimators (MLE) of the unknown parameter. It is observed that the MLE is not obtained in closed form, so it is not possible to derive the exact distribution of the MLE. Therefore, we propose to use the asymptotic distribution of the MLE to construct the approximate confidence interval. Further, Bayes estimate along with its posterior standard error and highest posterior density credible (HPD) interval of the parameter are obtained in section 4. Markov Chain Monte Carlo (MCMC) technique such as Metropolis-Hastings algorithm has been utilized to generate simulated draws from the posterior density of the parameter. In Section 5, a real data set has been analyzed for illustration purpose .

Model Description

Suppose n identical units are put to test under the same environmental conditions. It is assumed that R and T are known in advance and the item once failed will not be replaced. Therefore, under type-II hybrid censoring scheme, we have following types of observations:

Case I: $\{x_{1:n} < .... < x_{R:n}\}$ if $x_{R:n} > T$

CaseII: $\{x_{1:n} < .... < x_{R:n} < x_{R+1:n} < ... < x_{m:n} < T < x_{m+1:n}\}$ if $R \le m < n$ and $x_{m:n} < T < x_{m+1:n}$

Case III: $\{x_{1:n} < .... < x_{n:n} < T\}$ if $x_{R:n} > T$

Here, $x_{1:n} < x_{2:n} < ...$ denote the observed failure times of the experimental units. For schematic representation of the Type-II hybrid censoring scheme refer to Banerrjee and Kundu (2007). It is to be noted that although we do not observe $x_{m+1:n}$, but $x_{m:n} < T < x_{m+1:n}$ means that the m^th^ failure took place before T. Let the life time random variable X has a Lindley distribution with parameter θ i.e. the probability density function (PDF) of x is given by;

$$f(x) = \frac{\theta^2}{(1+\theta)}(1+x)e^{-\theta x}; \quad x,\theta > 0$$

Based on the observed data, the likelihood function is given by

Case I:

$$L(\underset{\sim}{x}|\theta) = \frac{\theta^{2R}}{(1+\theta)^n}\prod_{i=1}^{R}(1+x_{i:n})\left[1+\theta(1+x_{R:n})\right]^{n-R} \times e^{-\theta\left\{\sum_{i=1}^{R}x_{i:n}+(n-R)x_{R:n}\right\}} \tag{1}$$

Case II:

$$L(\underset{\sim}{x}|\theta) = \frac{\theta^{2m}}{(1+\theta)^n}\prod_{i=1}^{m}(1+x_{i:n})\left[1+\theta(1+T)\right]^{n-m} \times e^{-\theta\left\{\sum_{i=1}^{m}x_{i:n}+(n-m)T\right\}} \tag{2}$$

Case III:

$$L(\underset{\sim}{x}|\theta) = \frac{\theta^{2n}}{(1+\theta)^n}\prod_{i=1}^{n}(1+x_{i:n})e^{-\theta\sum_{i=1}^{n}x_{i:n}} \tag{3}$$

Then the combined likelihood for above 3 cases can be written as

$$L = L(\underset{\sim}{x}|\theta) = \frac{\theta^{2D}}{(1+\theta)^n}\prod_{i=1}^{D}(1+x_{i:n})\left[1+\theta(1+Z)\right]^{n-D} e^{-\theta\left\{\sum_{i=1}^{D}x_{i:n}+(n-D)Z\right\}} \tag{4}$$

Now, if we assume that D is the observed number of failures and

$$Z = \begin{cases} x_{R:n} & if \;\; D = R \\ T & if \;\; D > R \end{cases}$$

Then the log likelihood function for equation in (4) can be written

$$\log L = 2D\log(\theta) - n\log(1+\theta) + (n-D)\log\left[1+\theta(1+Z)\right] + \sum_{i=1}^{D}\log(1+x_{i:n}) - \theta\left[\sum_{i=1}^{D}x_{i:n} + (n-D)Z\right] \tag{5}$$

Maximum Likelihood Estimates

The first derivative of equation in (5) with respect to θ is given by

$$\frac{\partial \log L}{\partial\theta} = \frac{2D}{\theta} - \frac{n}{1+\theta} + \frac{(n-D)(1+Z)}{1+\theta(1+Z)} - \sum_{i=1}^{D}x_{i:n} + (n-D)Z \tag{6}$$

The second derivative of equation in (5) with respect to θ is given by

$$\frac{\partial^2 logL}{\partial\theta^2} = -\frac{2D}{\theta^2} + \frac{n}{(1+\theta)^2} + (n-D)\left[\frac{(1+Z)}{1+\theta(1+Z)}\right]^2 \tag{7}$$

The MLE of θ will be the solution of the following non-linear equation

$$\frac{2D}{\theta} - \frac{n}{1+\theta} + \frac{(n-D)(1+Z)}{1+\theta(1+Z)} - \sum_{i=1}^{D}x_{i:n} + (n-D)Z = 0 \tag{8}$$

From equnation (8), we see that $\hat{\theta}$ can't be obtained in closed form. However it can be solved for $\hat{\theta}$ by using some suitable numerical iterative procedure such as Newton-Raphson method. The observed Fisher's information is given by

$$I(\hat{\theta}) = -\frac{\partial^2 logL}{\partial\theta^2}\Big|_{\theta=\hat{\theta}} \tag{9}$$

Also, the asymptotic variance of $\hat{\theta}$ is given by

$$Var\left(\hat{\theta}\right)=\frac{1}{I\left(\hat{\theta}\right)} \tag{10}$$

The sampling distribution of $\frac{\left(\hat{\theta}-\theta\right)}{\sqrt{Var\left(\hat{\theta}\right)}}$ can be approximated by a standard normal distribution. The large-sample $(1-\gamma)100\%$ confidence interval for θ is given by $\left[\hat{\theta}_L,\hat{\theta}_U\right]=\hat{\theta}\pm z_{\frac{\gamma}{2}}\sqrt{Var\left(\hat{\theta}\right)}$.

Bayes Estimation

Here, we have also conducted a Bayesian study by assuming the following gamma prior for θ;

$$g(\theta)\propto\theta^{\beta-1}e^{-\alpha\theta};\quad \theta,\,\alpha,\,\beta>0$$

Here, the hyper parameters α and β are assumed to be known real numbers. Based on the above prior information, the joint density function of the sample observations and θ becomes

$$L\left(\underset{\sim}{x},\theta\right)\alpha\frac{\theta^{2D+\beta-1}}{\left(1+\theta\right)^n}\left[1+\theta\left(1+Z\right)\right]^{n-D}e^{-\theta\left\{\sum_{i=1}^{D}x_{i:n}+(n-D)Z+\alpha\right\}} \tag{11}$$

Based on the $L\left(\underset{\sim}{x},\theta\right)$, the posterior density function of θ, given the data is given by

$$\pi\left(\theta|\underset{\sim}{x}\right)=\frac{L\left(\underset{\sim}{x}|\theta\right)g_1\left(\theta|\alpha,\beta\right)}{\int_0^\infty L\left(\underset{\sim}{x}|\theta\right)g_1\left(\theta|\alpha,\beta\right)d\theta} \tag{12}$$

Therefore, if $h(\theta)$ is any function of θ, then its Bayes estimate under the squared error loss function is given by

$$\hat{h}(\theta)=E_{\theta|data}\left[h(\theta)\right]=\frac{\int_0^\infty h(\theta)L\left(\underset{\sim}{x}|\theta\right)g_1\left(\theta|\alpha,\beta\right)d\theta}{\int_0^\infty L\left(\underset{\sim}{x}|\theta\right)g_1\left(\theta|\alpha,\beta\right)d\theta} \tag{13}$$

Since, it is not possible to compute (12) and therefore (13) analytically. Therefore, we propose the MCMC method to draw samples from the posterior density function and then to compute the Bayes estimate and HPD credible interval.

Metropolis-Hastings algorithm:

Step-1: Start with any value satisfying target density $f\left(\theta^{(0)}\right)>0$.

Step-2: Using current $\theta^{(0)}$ value, generate a proposal point $\left(\theta\_prop\right)$ from the proposal density $q\left(\theta^{(1)},\theta^{(2)}\right)$ = $P\left(\theta^{(1)}\to\theta^{(2)}\right)$ i.e., the probability of returning a value of $\theta^{(2)}$ given a previous value of $\theta^{(1)}$.

Step-3: Calculate the ratio at the proposal point $\left(\theta\_prop\right)$ and current $\theta^{(i-1)}$ as:

$$\rho=\log\left[\frac{f\left(\theta\_prop\right)q\left(\theta\_prop,\theta^{(i-1)}\right)}{f\left(\theta^{(i-1)}\right)q\left(\theta^{(i-1)},\theta\_prop\right)}\right]$$

Step-4: Generate U from uniform on (0, 1) and take Z=log U.

Step-5: If $Z<\rho$, accept the move i.e., θ_prop and set $\theta^{(0)}=\theta\_prop$ and return to step 1. Otherwise reject it and return to step- 2.

Step-6: Repeat the above procedure N times and record the sequence of the parameter θ as $\theta_1,\theta_2,......,\theta_N$.

Step-7: The Bayes estimate of θ and corresponding posterior variance is taken as the mean and variance of the generated values of θ respectively.

Stpe-8: Let $\theta_{(1)}\le\theta_{(2)}\le......\le\theta_{(M)}$ denote the ordered value of $\theta_{(1)},\theta_{(2)},......\le\theta_{(M)}$. Then, following Chen and Shao (1999), the 100$(1-\gamma)$% HPD interval for θ is $\left(\theta_{\left(M+i^*\right)},\theta_{\left(M+i^*+[(1-\gamma)(M-N)]\right)}\right)$ where, i^* is so chosen that

$$\theta_{\left(M+i^*+[(1-\gamma)(M-N)]\right)}-\theta_{\left(M+i^*\right)}$$
$$=\min_{N\le i\le(M-N)-[(1-\gamma)(M-N)}\left(\theta_{\left(N+i+[(1-\gamma)(M-N)]\right)}-\theta_{\left(N+i\right)}\right)$$

Real Data Analysis

This section performs a real data analysis for illustrative purpose. We use the data set of waiting times (in minutes) before service of 100 bank customers as discussed by Ghitany et al. (2008). The waiting times in minutes are as follows:

0.8, 0.8, 1.3, 1.5, 1.8, 1.9, 1.9, 2.1, 2.6, 2.7, 2.9, 3.1, 3.2, 3.3, 3.5, 3.6, 4.0, 4.1, 4.2, 4.2, 4.3, 4.3, 4.4, 4.4, 4.6, 4.7, 4.7, 4.8, 4.9, 4.9, 5.0, 5.3, 5.5, 5.7, 5.7, 6.1, 6.2, 6.2, 6.2, 6.3, 6.7, 6.9, 7.1, 7.1, 7.1, 7.1, 7.4, 7.6, 7.7, 8.0, 8.2, 8.6, 8.6, 8.6, 8.8, 8.8, 8.9, 8.9, 9.5, 9.6, 9.7, 9.8, 10.7, 10.9, 11.0, 11.0, 11.1, 11.2, 11.2, 11.5, 11.9, 12.4, 12.5, 12.9, 13.0, 13.1, 13.3, 13.6, 13.7, 13.9, 14.1, 15.4, 15.4, 17.3, 17.3, 18.1, 18.2, 18.4, 18.9, 19.0, 19.9, 20.6, 21.3, 21.4, 21.9, 23.0, 27.0, 31.6, 33.1, 38.5.

Ghitany et al. (2008) already observed that the Lindley distribution can be quite effectively used to analyze this data set. Therefore, no need to perform goodness of fit test for the above data set.

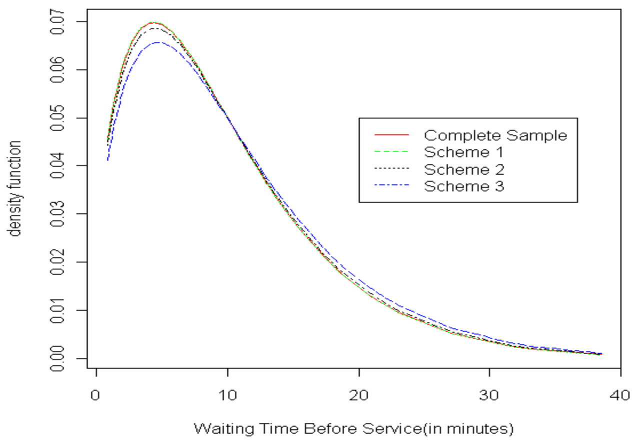

FIG. 1 : DIFFERENT ESTIMATED DESITY FUNCTION FOR COMPLETE, SCHEME 1, SCHEME 2, AND SCHEME 3 DATA SET

TABLE 1: CLASSICAL AND BAYES ESTIMATES WITH THEIR STANDARD ERRORS (S.E.)/ POSTERIOR STANDARD ERRORS (PSE) FOR COMPLETE AND HYBRID CENSORED DATA:

Scheme	ML Estimate (SE)	Bayes estimates (PSE)	
		Jeffrey's Prior	Gamma Prior
Complete	0.1865 (0.0132)	0.1866 (0.0130)	0.1866 (0.0129)
Scheme 1	0.1869 (0.0147)	0.1869 (0.0144)	0.1869 (0.0143)
Scheme 2	0.1834 (0.0169)	0.1833 (0.0166)	0.1834 (0.0164)
Scheme 3	0.1758 (0.0188)	0.1759 (0.0182)	0.1763 (0.0179)

TABLE 2: 95% CONFIDENCE/ HPD INTERVALS WITH THEIR WIDTHS FOR COMPLETE AND HYBRID CENSORED DATA:

Scheme	{Confidence Interval} [Width]	{HPD Intervals} [Width]	
		Jeffrey's Prior	Gamma Prior
Complete	{0.1605, 0.2126} [0.0520]	{0.1612, 0.2120} [0.0507]	{0.1620, 0.2121} [0.0500]
Scheme 1	{0.1580, 0.2159} [0.0578]	{0.1601, 0.2159} [0.0558]	{0.1592, 0.2147} [0.0555]
Scheme 2	{0.1501, 0.2166} [0.0665]	{0.1515, 0.2165} [0.0649]	{0.1522, 0.2163} [0.0640]
Scheme 3	{0.1389, 0.2126} [0.0737]	{0.1412, 0.2121} [0.0708]	{0.1421, 0.2120} [0.0698]

For analyzing the above data set with Type-II hybrid censoring, we have created three artificially Type-II hybrid censored data sets from the above complete (uncensored) data under the following censoring schemes:

Scheme 1: R = 75, T=12 (25 % Censored data)
Scheme 2: R = 50, T=8 (50 % Censored data)
Scheme 3: R = 35, T=6 (65 % Censored data)

In all the cases, we have estimated the unknown parameter using the classical and Bayes methods of estimation.

As the MLE of θ can't be obtained in closed form, so, in this case, we have used nlm() function of R environment. Bayes estimates of θ and HPD interval are obtained using gamma and Jeffrey priors. Using Metropolis-Hastings algorithms; we generated 5,000 realizations of the parameter θ from the posterior density in (11). The convergence of the sequences of parameter for their stationary distributions has been checked through different starting values.

Bayes estimates of the parameter with gamma priors have been obtained by setting the values of prior's parameter as $\theta = E(\theta) = \beta/\alpha$ and put the value of prior's parameters as zero to obtain Bayes estimate with Jeffrey's prior. The results of the above three schemes have been summarized in Table 1-2. Note that, in the Tables 1-2, the entries in the brackets "()" represents SEs/PSEs and that in the brackets {} and [] respectively represent confidence /HPD interval and the widths of the interval. To see the consequence of censoring on the estimation of the unknown parameters, we have also plotted the four density functions based on MLE for complete, Scheme 1, Scheme 2 and Scheme 3 data sets in Fig. 1.

For all the numerical computations, the programs are developed in R-environment. From the Table 1-2 and Fig. 1, we observed the following:

- Bayes estimation with gamma prior provides more precise estimates as compared to the MLEs (in terms of SE/PSE). Although Jeffrey priors perform similar to MLE even with the Type-II hybrid censored data.
- It is observed that the length of the HPD credible intervals based on informative priors are slightly shorter than the corresponding length of the HPD credible intervals based on non-informative priors, as expected.
- From Fig. 1, it is observed that the goodness of fit of the Lindley distribution is quite acceptable even with the type-II hybrid censored data based on scheme 3. Although, estimated plot under all the censoring scheme could not estimate the upper tail properly because of the absence of information in that region. The loss of information increases respectively according to Scheme 1, Scheme 2 and Scheme 3, which is an obvious fact.

REFERENCES

Al-Mutairi D.K., Ghitany M.E., and Kundu D, Inference on stress-strength reliability from Lindley distribution, to appear in Comm. Statist. Theory Methods, 42(8) (2013), 1443-1463.

Balakrishnan N., Aggarwala R., Progressive censoring: theory, methods and applications, Boston: Birkhauser, 2000.

Balakrishnan N., Kundu D., Hybrid censoring: Models, inferential results and applications. Computational Statistics and Data Analysis, 57(1) (2013), 166-209.

Banerjee A. and Kundu D., Inference based on Type-II Hybrid Censored Data from a Weibull, IEEE Trans. Rel., 57(2)(2008), 369-379.

Chen S. M. and G.K. Bhattacharya, Exact confidence bounds for an exponential parameter Hybrid censoring, Communications in Statistics, Theory and Methods, 17(6) (1988), 1858-1870.

Chen M. H. and Shao Q. M., Monte Carlo Estimation of Bayesian credible and HPD intervals, J. Comput. Graph Statistics, 6 (1999), 69-92.

Childs A., Chandrasekhar B., Balakrishnan N. and Kundu D., 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 (2003), 319-330.

Draper N. and Guttman I., Bayesian analysis of hybrid life tests with exponential failure times, Annals of the Institute of Statistical Mathematics, 39 (1987), 219-225.

Dube S., Pradhan B. and Kundu D., parameter estimation for the hybrid censored log-normal distribution, Journal of Statistical Computation and Simulation, 81 (3) (2011), 275-282.

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

Ganguly A., Mitra S., Samanta D. and Kundu D., Exact inference for the two-parameter exponential distribution under type-II hybrid censoring scheme, Journal Stat. Plan. Inf., 142 (3) (2012) 613:625.

Ghitany M.E., Al-Mutairi D. K., Balakrishnan N. and Al-Enezi L.J., Power Lindley distribution and associated inference. Computatinal Statistics and Data Analysis 64 (2013), 20-33.

Ghitany M.E., Atieh B. and Nadarajah S., Lindley distribution and its application, Math. Comput. Simul., 78 (4) (2008), 493-506.

Gomoz-Deniz E. and Calderin-Ojeda E., The discrete Lindley distribution: properties and applications, Journal of Statistical Computation and Simulation, 81 (2011), 1405-1416.

Gupta P. K. and Singh B., Parameter estimation of hybrid censored data with Lindley distribution, Int J Sys Assur Eng Manag 4 (4) (2013) 378-385.

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

Krishna H. and Kumar K., Reliability estimation in Lindley Distribution with progressively type II censored sample, math. Comput. Simul., 82(2) (2011), 281-294.

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

Kundu D. and Pradhan B., Estimating the Parameters of the Generalized Exponential Distribution in Presence of the Hybrid Censoring, Communications in Statistics – Theory and Methods, 38(12) (2009), 2030-2041.

Lindley D.V., Fiducial Distribution and Bayes' theorem, J. Roy. Stat. Soc., Series B 20 (1958), 102-107.

Mazucheli J. and Achcar J.A., The Lindley Distribution applied to competing risks lifetime data, Comput. Methods Programs Biomed., 104 (2) (2011), 241-288.

MIL-STD-781-C, Reliability Design Qualifications and Production Acceptance Test, Exponential Distribution, U.S. Government Printing Office, Washington, D.C., 1977.

Singh B. and Gupta, P. K., Load-sharing system model and its application to the real data set, Mathematics and Computers in Simulation, 82 (2012), 1615-1629.

Singh S.K., Singh U. and Vikas Kumar Sharma, Bayesian Analysis for Type-II Hybrid Censored Sample from Inverse Weibull Distribution, Int J Syst Assur Eng Manag, 4(3)(2013), 241–248.