한국통계학회 논문집 제1권1호
1994년 12월 pp. 20-26

Statistical Inferences on the Lognormal Hazard Function under Type I Censored Data[1)]

Kil-Ho Cho, In-Suk Lee, Jeen-Kap Choi[2)]

Abstract

The hazard function is a non-negative function that measures the propensity of failure in the immediate furture, and is frequently used as a decision criterion, especially in replacement decisions.

In this paper, we compute approximate confidence intervals for the lognormal hazard function under Type I censored data, and show how to choose the sample size needed to estimate a point on the hazard function with a specified degree of precision. Also we provide a table that can be used to compute the sample size.

1. Introduction

The lognormal distribution function has been widely used as a lifetime distribution model (Whittemore and Altschuler(1976) and Crow and Shimizu(1988)). If lifetime T has a lognormal distribution with the probability density function(p.d.f.)

$$f_T(t) = \frac{1}{(2\pi)^{1/2}\sigma t}\exp\left[-\frac{1}{2}\left(\frac{\log t-\mu}{\sigma}\right)^2\right], \quad t>0, \tag{1.1}$$

then Y=logT has a normal distribution with p.d.f.

$$f_Y(y) = \frac{1}{(2\pi)^{1/2}\sigma}\exp\left[-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2\right], \quad -\infty<y<\infty.$$

The corresponding lognormal hazard function h can be written as

$$h(t)=h(t\ ;\ \mu,\ \sigma)=\frac{\phi(z)}{\Phi(-z)\sigma t}, \quad t>0,$$

where ϕ and Φ are the p.d.f. and the cumulative distribution function(c.d.f.) of the standard normal distribution, respectively, and

$$z=\frac{\log t-\mu}{\sigma}.$$

The hazard function is a non-negative function that measures the propensity of failure in the immediate furture as a function of time, and is frequently used as a decision criterion, especially in replacement decisions.

Many authors have contributed to the statistical methodology of the lognormal distribution.

---

1) This paper was supported (in part) by Non Directed Research Fund, Korea Research Foundation, 1993.

2) Department of Statistics, Kyungpook National University, Taegu, 702-701, KOREA.

Laurent(1963), Land(1972), and Shimizu and Iwase(1981) considered the statistical inferences for the parameters of the lognormal distribution. Estimation for the quantile function of the distribution was discussed by Evans and Shaban(1974) and Kramer and Paik(1979). Zacks and Even(1966) and Nelson and Schmee(1979) studied the efficiencies for estimators of the reliability function. Nadas(1969) and Jones(1971) introduced methods for computing confidence intervals for the lognormal hazard function under complete samples.

In this paper, we compute approximate confidence intervals for the lognormal hazard function under Type I censored data, and show how to choose the sample size needed to estimate a point on the hazard function with a specified degree of precision. Also we provide a table that can be used to compute the sample size.

2. Statistical Inferences on the Hazard Function

We use the maximum likelihood(ML) method to estimate μ, σ, and the hazard function $h(t_e)$ for any specified time t_e. Consider Type I sampling scheme involving observation on the lifetime of n independent individuals with p.d.f. (1.1). Then the log likelihood function is given by

$$\log L(\mu,\ \sigma) = -r\log\sigma - \frac{1}{2\sigma^2}\sum_{i\in D}(y_i-\mu)^2 + (n-r)\log\Phi\left(-\frac{y_c-\mu}{\sigma}\right),$$

where y_i is log lifetimes for $i \in D$, y_c is a log censoring time, D is the set of individuals for which y_i is an observed log lifetime, and r is the number of observed lifetimes.

The first derivatives of logL are

$$\frac{\partial \log L}{\partial \mu} = \frac{1}{\sigma^2}\sum_{i\in D}(y_i-\mu) + \frac{(n-r)}{\sigma}\left[\frac{\phi(z_c)}{\Phi(-z_c)}\right]$$

$$\frac{\partial \log L}{\partial \sigma} = -\frac{r}{\sigma} + \frac{1}{\sigma^3}\sum_{i\in D}(y_i-\mu)^3 + \frac{(n-r)}{\sigma}\left[\frac{z_c\,\phi(z_c)}{\Phi(-z_c)}\right],$$

where

$$z_c = \frac{y_c-\mu}{\sigma}.$$

Some iterative procedures can be used to solve ML equation for the MLE's $\hat{\mu}$ and $\hat{\sigma}$. Wolynetz(1979) suggested more effective iterative procedure than Newton-Raphson method.

By the invariant property of MLE, the MLE of $h(t)$ is given by

$$\hat{h}(t) = h(t\ ;\ \hat{\mu},\ \hat{\sigma}),\quad t>0\ .$$

To obtain an approximate confidence interval for $h(t_e)$ at any specified time t_e, we consider the approximate large sample normal distribution for $\log[\hat{h}(t_e)]$. This procedure is generally preferred over the one based on the approximate large sample normal distribution of $\hat{h}(t_e)$.

The logarithm of the hazard function of T at any specified time t_e, l, is

$$l = l(t_e;\ \mu,\ \sigma) = \log[h(t_e)]$$
$$= -\frac{1}{2}\log(2\pi) - \frac{1}{2}\left(\frac{\log t_e - \mu}{\sigma}\right)^2 - \log\left[\Phi\left(-\frac{\log t_e - \mu}{\sigma}\right)\right] - \log\sigma - \log t_e.$$

Therefore, the MLE of l is

$$\hat{l} = l(t_e;\ \hat{\mu},\ \hat{\sigma}).$$

Theorem 2.1. The asymptotic variance of $\hat{l}$ is

$$AVar(\hat{l}) = \frac{1}{\sigma^2}\Big[\left(-z_e + \frac{\phi(z_e)}{\Phi(-z_e)}\right)^2 AVar(\hat{\mu})$$
$$+2\left(-z_e + \frac{\phi(z_e)}{\Phi(-z_e)}\right)\left(-z_e^2 + z_e\frac{\phi(z_e)}{\Phi(-z_e)} + 1\right) ACov(\hat{\mu},\ \hat{\sigma})$$
$$+\left(-z_e^2 + z_e\frac{\phi(z_e)}{\Phi(-z_e)} + 1\right)^2 AVar(\hat{\sigma})\ \Big].$$

where

$$z_e = \frac{\log t_e - \mu}{\sigma}.$$

Proof. By Taylor series approximation,

$$AVar(\hat{l}) = \left(\frac{\partial l}{\partial \mu}\right)^2 Avar(\hat{\mu}) + 2\left(\frac{\partial l}{\partial \mu}\right)\left(\frac{\partial l}{\partial \sigma}\right) Acov(\hat{\mu},\ \hat{\sigma}) + \left(\frac{\partial l}{\partial \sigma}\right)^2 AVar(\hat{\sigma}).$$

Since

$$l = l(t_e;\ \mu,\ \sigma) = \log\left[\frac{\phi(z_e)}{\Phi(-z_e)\sigma t}\right],$$

the first derivatives of l are given by

$$\frac{\partial l}{\partial \mu} = -\frac{1}{\sigma}\left[-z_e + \frac{\phi(z_e)}{\Phi(-z_e)}\right]$$

$$\frac{\partial l}{\partial \sigma} = -\frac{1}{\sigma}\left[-z_e^2 + z_e\frac{\phi(z_e)}{\Phi(-z_e)} + 1\right].$$

Hence this completes the proof.

The $AVar(\hat{\mu})$, $AVar(\hat{\sigma})$, and $ACov(\hat{\mu},\ \hat{\sigma})$ depend only on z_c(Cohen(1961)).

If we standardize $AVar(\hat{l})$ with respect to sample size n, then

$$V(z_c, z_e) = n\, AVar(\hat{l}).$$

$V(z_c, z_e)$ depends only on the standardized logtimes z_c and z_e, which are functions of the times t_c, t_e and the unknown parameters μ and σ.

Planning values of μ and σ are obtained from experience. It is also possible to obtain z_c and z_e directly from planning values for $p_c = F_T(t_c)$ and $p_e = F_T(t_e)$, the proportions of units that are anticipated to fail by times t_c and t_e, where F_T is c.d.f. of lifetime T. Straightforward computation gives;

$$z_\psi = \Phi^{-1}(p_\psi), \quad \psi = c \text{ or } e.$$

Alternatively, one can obtain z_c and z_e by having planning values for σ and $p_i = F_T(t_i)$ for any t_i.

Let

$$z_i = \Phi^{-1}(p_i).$$

Then

$$z_\psi = z_i + \frac{1}{\sigma} \log\left(\frac{t_\psi}{t_i}\right), \quad \psi = c \text{ or } e.$$

Table 1 provides values of $V(z_c, z_e)$, as a function of z_e for several values of z_c. These values include most practical applications of interest. The table of $V(z_c, z_e)$ shows the followings;

1. For any fixed z_e, $V(z_c, z_e)$ decreases as z_c increases because less censoring provides greater precision.

2. For a fixed standardized log censoring time z_c, $V(z_c, z_e)$ decreases rapidly and then increases slowly as z_e increases.

Given $\hat{\mu}$ and $\hat{\sigma}$, one can estimate $AVar\{\log[\hat{h}(t_e)]\}$ for any specified time t_e by using

$$\widehat{AVar}\{\log[\hat{h}(t_e)]\} = V(\hat{z}_c, \hat{z}_e)/n.$$

Theorem 2.2. A $100(1-\gamma)\%$ approximate confidence interval for the hazard function at any specified time t_e, $h(t_e)$, is

$$[\hat{h}(t_e)/\hat{q},\ \hat{h}(t_e)\hat{q}]$$

where

$$\hat{q} = \exp[Z_{(1-\gamma/2)}\, \widehat{SD}\{\log[\hat{h}(t_e)]\}],$$

$Z_{(1-\gamma/2)}$ is the $1-\gamma/2$ quantile of the standard normal distribution, and $\widehat{SD}\{\log[\hat{h}(t_e)]\}$ is an estimate of the standard deviation of $\log[\hat{h}(t_e)]$.

Proof. Since

$$\frac{\log[\hat{h}(t_e)] - \log[h(t_e)]}{\widehat{SD}\{\log[\hat{h}(t_e)]\}} \approx N(0,1) \qquad as\ n \to \infty,$$

the result can be proved easily.

Now, we show how to choose n large enough to estimate $h(t_e)$ to within a specified factor q (q>1) with a specified probability $1-\gamma$ ($0<\gamma<1$); that is, choose n so that

$$P[\hat{h}(t_e)/q < h(t_e) < \hat{h}(t_e)q] = 1-\gamma. \qquad (2.1)$$

Thoremm 2.3. The approximate sample size required to meet the criterion of equation (2.1) is

$$n = [Z_{(1-\gamma/2)}/\log(q)]^2\ V(z_c,\ z_e).$$

Proof. Since, in large samples,

$$\frac{\log[\hat{h}(t_e)] - \log[h(t_e)]}{\left[\frac{V(z_c,\ z_e)}{n}\right]^{1/2}} \approx N(0,1),$$

$$\log(q) = Z_{(1-\gamma/2)}\ [V(z_c,\ z_e)/n]^{1/2}.$$

Therefore, this completes the proof.

The sample size depends on the model parameters through z_c and z_e.

3. Example

The data below have been discussed by Nelson and Schmee(1979) and Lawless(1982), and show the number of thousand miles at which different locomotive controls failed in a life test involving 96 controls. The test was terminated after 135.0 thousand miles. The failure times for the 37 failed items are 22.5, 37.5, 46.0, 48.5, 51.5, 53.0, 54.5, 57.7, 66.5, 68.0, 69.5, 76.5, 77.0, 78.5, 80.0, 81.5, 82.0, 83.0, 84.0, 91.5, 93.5, 102.5, 107.0, 108.5, 112.5, 113.5, 116.0, 117.0, 118.5, 119.0, 120.0, 122.5, 123.0, 127.5, 131.0, 132.5, and 134.0. In addition, there are 59 censoring times, all equal to 135.0. We shall assume that the failure times are lognormally distributed.

The MLE's for μ and σ are $\hat{\mu}=4.12$ and $\hat{\sigma}=0.71$. An estimate of the hazard function at $t_e=80$ thousand miles is needed, since a warranty on the locomotive controls covers the first 80000 miles.

Then

$$z_c = \frac{\log t_c - \mu}{\sigma} = \frac{\log 135 - 4.12}{0.71} = 1.12$$

$$z_e = \frac{\log t_e - \mu}{\sigma} = \frac{\log 80 - 4.12}{0.71} = 0.37$$

$$V(1.12,\ 0.37) \doteq 0.29$$

$$\hat{h}(80) = \frac{\phi(0.37)}{\Phi(-0.37) * 0.71 * 80} = 0.0184$$

$$\hat{q} = 1.1137.$$

If this estimate is to be within a factor 1.05 of the true value with probability 0.95 (q=1.05 and $1-\gamma$=0.95), the desired sample size is

$$n = [1.96/\log 1.05]^2 * (0.29) \doteq 468.$$

Also, an approximate 95% confidence interval for $h(80)$ is (0.0165, 0.0205).

Table 1. values of $V(z_c,\ z_e)$

Z_e \ Z_c	-2.0	-1.5	-1.0	-0.5	0	0.5	1.0	1.5	2.0	3.0	4.0	5.0
-4.0	5742	2015	904	519	376	332	312	295	257	186	109	58.8
-3.5	3374	1184	531	306	223	199	175	151	139	115	72.8	38.7
-3.0	1834	643	288	166	122	111	107	96.5	81.7	65.2	41.3	27.5
-2.5	905	316	141	81.6	60.5	56.1	53.2	50.9	42.7	30.6	20.9	13.7
-2.0	401	139	61.5	35.1	26.2	24.8	23.2	21.2	19.7	16.1	14.3	11.4
-1.5	164	56.1	24.0	13.2	3.61	9.24	89.2	7.23	5.49	4.63	2.97	1.16
-1.0	73.5	24.3	9.74	4.78	3.07	2.79	2.47	2.04	1.76	0.83	0.43	0.19
-0.5	49.0	15.9	6.06	2.59	1.24	0.78	0.61	0.47	0.41	0.17	0.06	0.04
0	49.2	16.2	6.28	2.73	1.24	0.56	0.26	0.23	0.18	0.06	0.03	0.02
0.5	56.1	18.8	7.27	3.49	1.76	0.91	0.44	0.18	0.10	0.05	0.02	0.01
1.0	63.2	21.4	8.81	4.27	2.34	1.37	0.82	0.46	0.21	0.09	0.01	0.01
1.5	69.1	23.6	9.81	4.92	2.82	1.79	1.20	0.80	0.51	0.11	0.02	0.01
2.0	73.6	25.2	10.6	5.42	3.21	2.13	1.53	1.12	0.82	0.34	0.06	0.02
3.0	79.4	27.2	11.7	6.08	3.73	2.65	2.01	1.63	1.35	0.87	0.46	0.18
4.0	82.6	28.5	12.3	6.47	4.05	2.92	2.32	1.97	1.73	1.31	0.91	0.56
5.0	84.5	29.3	12.6	6.71	4.25	3.11	2.54	2.21	2.00	1.64	1.28	0.94

References

[1] Cohen, A. C. (1961), Tables for maximum likelihood estimates: singly truncated and singly censored samples, *Technometrics*, Vol. 3, 535-541.

[2] Crow, E. L. and Shimizu, K. (1988), *Lognormal Distributions*, Marcel Dekker, Inc.

[3] Evans, I. G. and Shaban, S. A. (1974), A note on estimation in lognormal models, *Journal of the American Statistical Association*, Vol. 69, 779-781.

[4] Jones, C. F. (1971), A confidence interval for the lognormal hazard, *Technometrics*, Vol. 13, 885-888.

[5] Kramer, H. C. and Paik, M. (1979), A central t approximation to the noncentral t distribution, *Technometrics*, Vol. 21, 357-360.

[6] Land, C. E. (1972), An evaluation of approximate confidence interval methods for lognormal means, *Technometrics*, Vol. 14, 145-158.

[7] Laurent, A. G. (1963), The lognormal distribution and the translation method: description and estimation problems, *Journal of the American Statistical Association*, Vol. 58, 231-235.

[8] Lawless, J. F. (1982), *Statistical Models* and *Methods* for *Lifetime Data* New York: Wiley.

[9] Nadas, A. (1969), A confidence region for the lognormal hazard function, *Technometrics*, Vol. 11, 387-388.

[10] Nelson, W. B. and Schmee, J. (1979), Inference for (log) normal life distributions from small singly censored samples and blue's, *Technometrics*, Vol. 21, 43-54.

[11] Shimizu, K. and Iwase, K. (1981), Uniformly minimum variance unbiased estimation in lognormal and related distribution, *Communications in Statistics*, *Theory* and *Methods*, Vol. 10, 1127-1147.

[12] Whittemore, A. and Altschuler, B. (1976), Lung cancer incidence in cigarette smokers: further analysis of Doll and Hill's data for Britsh physicians, *Biometrics*, Vol. 32, 805-816.

[13] Wolynetz, M. S. (1979), Statistical algorithms AS138 and AS139, *Applied*, *Statistics*, Vol. 28, 185-206.

[14] Zacks, S. and Even, M. (1966), The efficiencies in small samples of the maximum likelihood and best unbiased estimators of reliability functions, *Journal of the American Statistical Association*, Vol. 61, 1033-1051.