INTERVAL ESTIMATION FOR SURVIVOR FUNCTION ON TWO PARAMETERS EXPONENTIAL DISTRIBUTION UNDER TYPE II CENSORING WITH BOOTSTRAP PERCENTILE

Akhmad Fauzy, Noor Akma Ibrahim, Isa Daud, Mohd. Rizam Abu Bakar
Institute for Mathematical Research, Universiti Putra Malaysia, Serdang, Selangor Malaysia

Abstract
In this article, two methods are proposed to find the interval estimation for survivor function of two parameters exponential distribution under type II censoring. Survivor function is a probability of an individual surviving till time t. Lawless (1982), Bain and Engelhardt (1992) and Bury (1999) usually use traditional method to construct interval estimation. This interval needs an assumption that sample is exponentially distributed. We will use another method, known as the Bootstrap percentile. This method gives shorter interval than the traditional method and this method does not need an assumption that the sample is distributed exponentially.

Introduction

The statistical analysis of what is variously referred to as lifetime, survival time or failure time data has developed into an important topic for researchers in many areas, especially in the engineering and biomedical sciences. Applications of lifetime distribution methodology range from investigations into the endurance of manufactured items so research involving human diseases.

The exponential distribution is used widely in the modelling of lifetime data. Historically the exponential distribution was the first lifetime model in which statistical methods in survival analysis were extensively developed. What distinguishes survival analysis from other fields of statistics is censoring. Vaguely speaking, a censored observation contains only partial information about the random variable of interest. The three types of censoring are the type I censoring, the type II censoring and random censoring (Miller, 1981).

Failure times that resulted from an experiment that is run over a fixed time period in such a way that an individual's lifetime is known exactly only if it is less than some predetermined value are said to be type I censored or time censoring. A type II censored data involves observations in which only the r smallest observations in a random sample of n items are observed. Censoring times are often effectively random. For example, in medical trial patients may enter the study in a more or less random fashion, according to their time of diagnosis. If the study is terminated at some prearranged date, then censoring time, that is the lengths of time from an individual's entry into the study until the termination of the study, are random (Lawless, 1982).

Lawless (1982), Bain and Engelhardt (1992) and Bury (1999) use traditional method to construct interval for survivor function on two parameters exponential distribution under type II censoring. This interval needs an assumption that sample is exponentially distributed. Bootstrap method is a computer-based method for assigning measures of accuracy to statistical estimates, especially to calculate the confidence interval. The aim of using bootstrap method is to gain the best estimation from minimal data (Efron and Thibshirani, 1993).

Fauzy and Ibrahim (2002a) use bootstrap method to construct interval estimation for one parameter exponential distribution under type II censoring. In Fauzy and Ibrahim (2002b) bootstrap method was utilised to construct the interval estimation for survivor function on one parameter exponential distribution under type II censoring. And again bootstrap method was used to construct the interval for two parameters exponential distribution under type II censoring (Fauzy and Ibrahim, (2002c)). In this paper comparison study will be made on the interval estimation of the survivor function using both the conventional method and bootstrap percentile method.

Methodology

The data used for illustration is taken from the text *Statistical Models and Methods for Lifetime Data*, by Lawless (1982 page 130). We begin with the interval estimation for survivor function on two parameters exponential distribution under type II censoring using the traditional method. This will be followed with the searching of convergence condition from bootstrap's repeated samples. Next, after knowing the convergence condition, confidence interval for survivor function on two parameters exponential distribution under type II censoring with bootstrap percentile method can be computed.

Theory

The statistical analysis of what is variously referred to as lifetime, survival time, or failure time data has developed into an important topic for researchers in many areas, especially in the engineering, medical research, insurance and biological sciences. The data in these various fields can be described as type I censoring or time censoring, type II censoring or lastly complete censoring. In type II censoring only the r smallest observations in a random sample of n items are observed.

The actual survival time of an individual, t, can be regarded as the value of a variable T, which can take any non-negative value. The different values that T can take have a probability distribution, and we call T the random variable associated with the survival time. Now suppose that the random variable T has a probability distribution with underlying probability density function $f(t)$. The distribution function of T is then given by:

$$F(t) = P(T < t) = \int_0^t f(u)\,\mathrm{d}u \qquad (1)$$

which represents the probability that the survival time is less than some value t. The survivor function, $S(t)$, is defined to be the probability that the survival time is greater than or equal to t, and so:

$$S(t) = P(T \geq t) = 1 - F(t) \qquad (2)$$

The survivor function can therefore be used to represent the probability that an individual survives from the time origin to some time beyond t (Collett, 1996).

Type II Censoring

The two parameters exponential distribution has probability density function (Lawless, 1982):

$$f(t;\mu,\theta) = \frac{1}{\theta}\exp\left(-\frac{t-\mu}{\theta}\right) \;;\; t \geq \mu,\ \mu \geq 0,\ \theta > 0 \qquad (3)$$

The joint probability density function (pdf) of the r smallest observations $t_{(1)} < t_{(2)} < \ldots < t_{(r)}$ in a random sample of size n from (1) is:

$$\frac{n!}{(n-r)!}\frac{1}{\theta^{r}}\exp\left(-\frac{1}{\theta}\sum_{i=1}^{r}(t_{(i)}-\mu)-\frac{n-r}{\theta}(t_{(i)}-\mu)\right), \quad t_{(i)} \geq \mu \tag{4}$$

To obtain the maximum likelihood estimation (m.l.e.) of $\hat{\mu}$ and $\hat{\theta}$ note that for any $\theta > 0$ (4) decreases as μ decreases. Also note that, regardless of the value of θ, $t_{(1)}$ is the largest value μ can take on in (4), since $\mu \leq t_{(1)} < t_{(2)} < \ldots < t_{(r)}$. Hence:

$$\hat{\mu} = t_{(1)} \quad \text{and} \quad \hat{\theta} = \frac{\left(\sum_{i=1}^{r} t_{(i)} + (n-r)\,t_{(r)} - n\,t_{(1)}\right)}{r} \tag{5}$$

Confidence intervals for θ and μ are respectively:

$$\frac{2r\hat{\theta}}{\chi^{2}_{(1-\alpha/2,2r)}} = \hat{\theta}_{\min} < \theta < \frac{2r\hat{\theta}}{\chi^{2}_{(\alpha/2,2r)}} = \hat{\theta}_{\max} \text{ and}$$

$$\left(\hat{\mu} - \frac{r\,\hat{\theta}\,F_{(2;2r-2)}}{n(r-1)}\right) = \hat{\mu}_{\min} < \mu < \hat{\mu} = \hat{\mu}_{\max} \tag{6}$$

Interval estimation for survivor function is:

$$\exp\left(-\frac{(t-\hat{\mu}_{\min})}{\hat{\theta}_{\min}}\right) \langle \; \mathrm{S}(t) \; \langle \; \exp\left(-\frac{(t-\hat{\mu}_{\max})}{\hat{\theta}_{\max}}\right) \tag{7}$$

Bootstrap Percentile Method

In setting up of the bootstrap method to find the confidence intervals and estimating significance levels, this method consists of approximating the distribution of a function of the observations and the underlying distribution, such as a pivot, denoted by Efron as the bootstrap distribution of this quantity. This distribution is obtained by replacing the unknown distribution by the empirical distribution of the data in the definition of the statistical function, and then resampling the data to obtain a Monte Carlo distribution for the resulting random variable (Bickel and Freedman, 1981).

Bootstrap method is a computer-based method for assigning measures of accuracy to statistical estimates, especially to calculate the confidence interval. Bootstrap itself comes from "*pull oneself up by one's Bootstrap*" which means to stand up by own feet, do with minimal resources. The minimal resource is a minimum data, data which away from certain assumption or data with no assumption at all about the population distribution. The aim of using bootstrap method is to gain the best estimation from minimal observation.

The Bootstrap's percentile procedures for the interval estimation of survivor function on two parameters exponential distribution under type II censoring are as follows:

1. give an equal opportunity 1/*r* to every observation of *r* sample data of type II censoring,

2. take *r* sample with replication,
3. do step 2 until *B* times in order to get an "*independent bootstrap replications*", $\hat{\beta}_r^{*1}, \hat{\beta}_r^{*2}, \ldots, \hat{\beta}_r^{*B}$, and search for convergence condition. Calculate S(*t*) as follows:

$$\mathrm{S}(t)_r^{*i} = \exp\left(- \frac{\left(t_r^{*i} - \mu_r^{*i}\right)}{\theta_r^{*i}} \right), \text{ with } \mu_r^{*i} = t_{(1)}^{*i} \text{ and } \theta_r^{*i} = \frac{\mathrm{T}_r^{*i}}{r},$$

$$\mathrm{T}_r^{*i} = \sum_{i=1}^{r} t_{(i)}^{*i} + (n-r)\, t_{(r)}^{*i} \tag{8}$$

4. define the confidence interval at the level (1-α) of the bootstrap percentile for survivor function of two parameters exponential distribution under type II censoring as

$$\left[S(t)_r^{*(\alpha/2)}, S(t)_r^{*(1-\alpha/2)} \right] \tag{9}$$

Expressions (9) refer to the ideal bootstrap situation we must the number of bootstrap replications is infinite. So if B = 2000 and α = 0.05, $S(t)_r^{*(\alpha/2)}$ is the 50*th* and $S(t)_r^{*(1-\alpha/2)}$ is the 1950*th* ordered value of the replications.

Results And Discussion

This data was taken from the text *Statistical Models and Methods for Lifetime Data*, by Lawless (1982 page 130). The data are mileages for 19 military personnel carriers that failed in service and can be described by using exponential model. There is no censoring, and the mileages are:

162, 200, 271, 320, 393, 508, 539, 629, 706, 777,
884, 1008, 1101, 1182, 1463, 1603, 1984, 2355, 2880

This data is of type II censored and exponentially distributed with two parameters. We will construct interval estimation survivor function at 700 and 1250 hours or S(700) and S(1250).

Traditional Method

By equations (4)-(7), the survivor functions at 700 and 1250 hours are

S(700) = 0.525493 and S(1250) = 0.272208

Based on the above two parameters of exponential distribution under type II censoring, the interval estimation for survivor function at S(700) and S(1250), are tabulated in Table 1 and Table 2.

Table 1. The floor (F), ceiling (C) and interval widths (IW) for survivor function at t = 700 at the level of (L) 99 % and 95 %

L	F	C	IW
99 %	0.219771	0.738703	0.518932
95 %	0.306927	0.696797	0.389870

Table 2. The floor (F), ceiling (C) and interval widths (IW) for survivor function at t = 1250 at the level of (L) 99 % and 95 %

L	F	C	IW
99 %	0.075689	0.542008	0.466319
95 %	0.119619	0.481630	0.362011

Bootstrap Percentile

Bootstrap's repeated result gives a convergence condition that begins at B = 4850. The plot between bias and replication is shown in Figure 1.

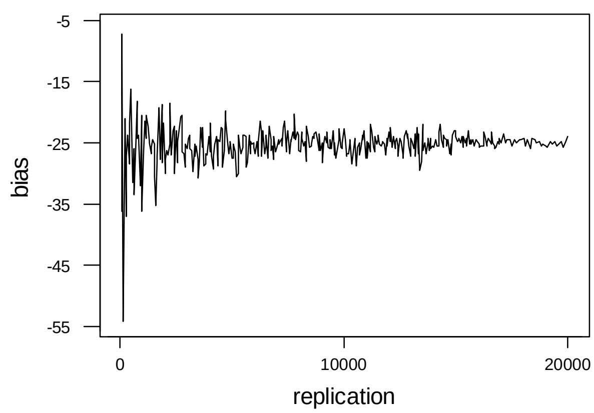

Figure 1. Plot between bias and replication

Estimations θ, μ and S(*t*) at this replication are

$$\hat{\theta} = 812.265 \ ; \ \hat{\mu} = 187.398 \ ; \ S(700) = 0.524033 \ ; \ S(1250) = 0.266005$$

After doing the bootstrap process, the floor (F), ceiling (C) and interval widths (IW) at t = 700 and 1250 at the level of (L) 99 % and 95 % are depicted in Table 3 and Table 4.

Table 3. The floor (F), ceiling (C) and interval widths (IW) for survivor function at t = 700 at the level of (L) 99 % and 95 %

L	F	C	IW
99 %	0.298157	0.736704	0.438547
95 %	0.362916	0.664835	0.301919

Table 4. The floor (F), ceiling (C) and interval widths (IW) for survivor function at t = 1250 at the level of (L) 99 % and 95 %

L	F	C	IW
99 %	0.083981	0.441740	0.357759
95 %	0.127638	0.399036	0.271398

Comparison of Interval Widths

Table 5 gives the interval widths of the survivor function on two parameters exponential distribution under type II censoring using the traditional method and the bootstrap percentile method.

Table 5. Comparison interval widths of survivor function at t =700 and t =1250 at level of 99 % and 95 %

Method	S(700)		S(1250)	
	99 %	95 %	99 %	95 %
Traditional	0.518932	0.389870	0.466319	0.362011
Bootstrap percentile	0.438547	0.301919	0.357759	0.271398
Widths	0.080385	0.087951	0.108560	0.090613

Conclusion

Bootstrap percentile method has more potential in constructing interval estimation for survivor function on two parameters exponential distribution under type II censoring than the traditional method. And this method does not need assumption that the sample has to have an exponential distribution.

References

Bain, Lee J. and Max Engelhardt. 1992. *Introduction to Probability and Mathematical Statistics*. Second edition. Boston: PSW-KENT Publishing Company.

Bickel, P. J and Freedman, D. A. 1981. Some Asymptotic Theory for the Bootstrap. *The Annals of Statistics*. 9 1196-1217.

Bury, Karl. 1999. *Statistical Distributions in Engineering*. Cambridge: Cambridge University Press.

Collett, D. 1996. *Modelling Data In Medical Research*. London: Chapman & Hall.

Efron, B. and Tibshirani R. 1993. *An Introduction to the Bootstrap*. New York: Chapman & Hall.

Fauzy, Akhmad and N. Akma, I. 2002a. Interval Estimation For One Parameter Exponential Distribution Under Type II Censoring With Bootstrap Percentile. *Proceeding National Conference on Mathematic*, Department of Mathematic, Diponegoro University. 9 March 2002.

Fauzy, Akhmad and N. Akma, I. 2002b. Interval Estimation For Survivor Function On One Parameter Exponential Distribution Under Type II Censoring With Bootstrap Percentile.

Proceeding National Conference on Mathematic 11th, Department of Mathematic, Malang University. 22-25 July 2002.

Fauzy, Akhmad and N. Akma, I. 2002c. Interval Estimation For Two Parameters Exponential Distribution Under Type II Censoring With Bootstrap Percentile. *Proceeding National Conference on Mathematic* 11th, Department of Mathematic, Malang University. 22-25 July 2002.

Lawless. 1982. *Statistical Models and Methods for Lifetime Data*. New York: John Wiley & Sons.

Miller, Rupert G. 1981. *Survival Analysis*. New York: John Wiley & Sons.

Appendix

```
"bootstrap" <- function(x,B)
{
data1 <- matrix(sample(x, size=r*B, replace=T),nrow=B)
data1
}

x <- c(162, 200, 271, 320, 393, 508, 539, 629, 706, 777,
       884, 1008, 1101, 1182, 1463, 1603, 1984, 2355, 2880)
r <- length(x)
n <- 19

t <- sum(x)+(n-r)*x[19]-n*x[1]
teta <- t/r
mu <- x[1]

alpha <- 0.01
tatas1 <- (2*r*teta) / (qchisq(alpha/2,2*r-2))
tbawah1 <- (2*r*teta) / (qchisq(1-alpha/2,2*r-2))

muatas1 <- mu
mubawah1 <- mu - (r*teta*qf(1-alpha,2,2*r-2)/(n*(r-1)))

alpa <- 0.05
tatas2 <- (2*r*teta) / (qchisq(alpa/2,2*r-2))
tbawah2 <- (2*r*teta) / (qchisq(1-alpa/2,2*r-2))

muatas2 <- mu
mubawah2 <- mu - (r*teta*qf(1-alpa,2,2*r-2)/(n*(r-1)))

t1 <- 700
surt1  <- exp(-(t1-mu)/teta)
surb1a <- exp(-(t1-mubawah1)/tbawah1)
sura1a <- exp(-(t1-muatas1)/tatas1)
int1a <- sura1a-surb1a

surb1b <- exp(-(t1-mubawah2)/tbawah2)
sura1b <- exp(-(t1-muatas2)/tatas2)
int1b <- sura1b-surb1b

t2 <- 1250
surt2 <- exp(-(t2-mu)/teta)
surb2a <- exp(-(t2-mubawah1)/tbawah1)
```

```
sura2a <- exp(-(t2-muatas1)/tatas1)
int2a <- sura2a-surb2a

surt2b <- exp(-(t2-mu)/teta)
surb2b <- exp(-(t2-mubawah2)/tbawah2)
sura2b <- exp(-(t2-muatas2)/tatas2)
int2b <- sura2b-surb2b

B <- 4850
xstar <- bootstrap(x,B)

data2 <- apply(xstar,1,sort)
data3 <- apply(xstar,1,sum)
theta <- (data3+(n-r)*data2[r,]-n*data2[1,])/r
thetabar <- mean(theta)
bias <- thetabar-teta

miu1 <- data2[1,]
miu <- mean(miu1)

surb1 <- exp(-(t1-miu1)/theta)
surbar1 <- mean(surb1)

urut1 <- sort(surb1)
l1 <- urut1[(alpha/2)*B]
u1 <- urut1[(1-alpha/2)*B]
sel1 <- u1-l1

l2 <- urut1[(alpa/2)*B]
u2 <- urut1[(1-alpa/2)*B]
sel2 <- u2-l2

surb2 <- exp(-(t2-miu1)/theta)
surbar2 <- mean(surb2)

urut2 <- sort(surb2)
b1 <- urut2[(alpha/2)*B]
a1 <- urut2[(1-alpha/2)*B]
sela1 <- a1-b1

b2 <- urut2[(alpa/2)*B]
a2 <- urut2[(1-alpa/2)*B]
sela2 <- a2-b2
```