Estimation of type I censored exponential distribution parameters using objective bayesian and bootstrap methods (case study of chronic kidney failure patients)

A Wiranto, A Kurniawan¹, D A Fitria, Suliyanto, N Chamidah
Department of Mathematics, Faculty of Sciences and Technology, Airlangga University, Surabaya, Indonesia

¹Corresponding Author : ardi-k@fst.unair.ac.id

Abstract. Bayesian point estimation is an estimation method based on prior selection and loss function. In Objective Bayesian estimation are chosen prior to Jeffrey and used intrinsic discrepancy loss functions based on the Kullback-Leibler divergence equation which will have a minimum effect of data on the posterior distribution. The objective Bayesian point estimator provides estimates of population parameters based solely on the assumed population distribution and data. The goal of this paper is to estimate parameters from the exponential distribution on type II censored data using the objective Bayesian and bootstrap methods. The bootstrap method is used to resampling and built a confidence intervals for parameters which will be estimated. The methods were applied on the life-time data of 63 patients of chronic renal failure and the initial diagnosis was non-diabetic disease with bootstrap methods using 10,100, and 1000 times used in this study. So that the bigger bootstrap samples rendered the estimated value $\hat{\theta}$ will be better and the result confidence interval ranges narrower.

1. Introduction
Survival data analysis is an analysis that discusses the survival of an object or individual in certain operational conditions. This analysis is usually applied to observe the survival of a product up to observations in the field of human disease that can be used to improve product quality and develop modern medical methods [1]. In observing survival data, life time data can be divided into censored and uncensored data. Censored observations occur if the survival data of the observed individuals is unknown [7].

One type of censorship is type I censorship, which occurs when all the objects studied are tested at the same time and the test is stopped after the specified observation time limit [6]. Observation with censorship is often have disadvantages, which one is the limited sample observed. So, in the analysis a method called bootstrap method used to solve the disadvantages. The bootstrap approach uses a recurring sampling method (resample). Based on the characteristics of the Bootstrap method which is a data-based simulation method for the certain simpler statistical inferential will require help from modern computers [4].

Exponential Distribution is a random variable distribution that is often used for modeling purposes. Exponential distribution is widely used as a lifetime model based on research that includes the durability of survival time. One of the inferences carried out in statistical modeling is parameter estimation [5]. To
do estimations, the objective of this research is to use the Bayesian method which is an estimation method using a reference prior so that the completion of the estimation depends only on the presumption model and observation data [2]. The objective Bayesian method uses the intrinsic discrepancy loss function, which is the minimum function adapted from the Kullback-Leibler discriminant function. The principle of this function is to represent the parameters of a distribution with other parameters which still have the same information as the initial parameters, the parameter parameters are called proxies [3].

Several researches related to the use of bootstrap method in the field of life time analysis have been carried out. Bootstrap method can be used as an alternative method for predicting confidence intervals for geometric Weibull process model parameters in life test analysis for type II censored data [8]. Then another research concerning the Objective Bayesian estimation in uniform distribution. The study was able to produce intrinsic estimator values for θ and the intrinsic estimator obtained was close to the minimum value of the expected losses [3]. Based on the description above, this paper contains research on the application of parameter estimation Exponential distribution for type I censored data with an Objective Bayesian and bootstrap method. The result of those method used to modelling using the survival time data of patients with chronic renal failure.

2. Data and Research Method

The data used in this research based on survival time of patients with chronic renal failure undergoing Hemodialysis therapy diagnosed cause is non-diabetic disease in the hospital Dr. Sosodoro Djatikoesoemo Bojonegoro in 2014 was 63 patients. This paper describes the steps in estimating parameters, the confidence interval with the bootstrap method, and the application of theory to data on chronic kidney failure.

2.1 Estimation of Type I Censored Exponential Distribution Parameters with Objective Bayesian Method

The methods that used to obtain the estimation of type I censored exponential distribution parameters with objective bayesian methods
a. Take random sample \( t_1, t_2, \ldots, t_n \) from Exponential distribution \( \theta \)
b. Determine the intrinsic discrepancy of Exponential distribution using the formula
\[
\delta(\theta_0, \theta) = n \min K(\theta_0 | \theta) K(\theta | \theta_0) \text{ dengen } K(\theta | \theta_0) = \frac{f(t | \theta)}{f(t | \theta_0)} \int t dt
\]
c. Determine the censored time limit as \( L_i \) and the lifetime \( T_i \) censored or uncensored with \( d_i = 1 \) if \( t_i = T_i \) or \( T_i < L_i \) and \( d_i = 0 \) if \( t_i = L_i \) or \( T_i \geq L_i \)
d. Determine the likelihood function \( L(\theta | t) \) from Exponential distribution of type I censored data
\[
L(\theta | t) = \prod_{i=1}^{n} \left( \frac{f(t_i)}{f(L_i)} \right)^{d_i} S(L_i)^{1-d_i}
\]
e. Define the prior distribution use Jeffrey’s method \( \pi(\theta) \propto \sqrt{I(\theta)} \)
   Where, \( I(\theta) \) is Information of Fisher’s from \( \theta \).
f. Define posterior distribution
\[
\pi(\theta | t) = \frac{L(\theta | t) \pi(\theta)}{\int L(\theta | t) \pi(\theta) d\theta}
\]
g. Define intrinsic statistics \( d(\theta_0 | t) = \int_\theta \delta(\theta | \theta_0) \pi(\theta | t) d\theta \)
h. Obtain objective Bayesian estimation by \( \arg \min_{\theta \neq \theta_0} d(\theta_0 | t) \)
Determination Of Confidence Interval With Bootstrap Method

In this section will be explained how to acquire Mean Square Error from the parameter, and also obtain the lower bound and upper bound according to bootstrap sample size.

a. Suppose $t_1, t_2, ..., t_n$ Exponential Distribution
b. Make a withdrawal by returning data at step a one by one 63 times
c. Perform parameter estimation as in step 1 and obtained $\hat{\theta}^*$ (1)
d. Repeat steps b and c as much as $B = 10$ times so that an estimator is obtained $\hat{\theta}^*$ (1), $\hat{\theta}^*$ (3), ..., $\hat{\theta}^*$ (10)
e. Calculate the average of $\hat{\theta} = \sum_{i=1}^{10} \hat{\theta}^*(i)$
f. Calculate the MSE by $\text{MSE}(\hat{\theta}) = \frac{\sum_{i=1}^{B} (\hat{\theta} - \hat{\theta}^*(i))^2}{B}$
g. Sort the value of $\hat{\theta}^*$ (i) from smaller to bigger
h. Determine confidence interval of 95% for $\hat{\theta}^*$ with 2.5% quantile for lower bound and 97.5% quantile for upper bound.

2.3 Apply the theory obtained from steps 2.1 and 2.2 to the survival data of patients with chronic renal failure.

3. Result and Analysis

3.1 Estimation of Exponential Distribution Parameters in Type I Censored Data with Objective Bayesian Methods

If $t_1, t_2, ..., t_n$ is a random sample with Exponential distribution with scale parameters is $\theta$, then it is known that the form of probability distribution function of Exponential distribution functions as follows:

$$ f(t) = \frac{1}{\theta} \exp\left(-\frac{t}{\theta}\right), \theta > 0, t > 0 \quad (1) $$

The first step to estimate using the Objective Bayesian method is determine intrinsic discrepancy as follows:

a. Intrinsic Discrepancy for Type I Censored Data of Exponential Distribution

In general Kullback-Leibler equation based on $f(t|\theta)$ are $K(\theta_0|\theta)$ and $K(\theta|\theta_0)$ as follows:

$$ K(\theta_0|\theta) = \frac{\theta}{\theta_0} - 1 - \ln\left(\frac{\theta}{\theta_0}\right) \quad \text{and} \quad K(\theta|\theta_0) = \frac{\theta_0}{\theta} - 1 - \ln\left(\frac{\theta_0}{\theta}\right) \quad (2) $$

Furthermore, the intrinsic discrepancy will be determined $\delta(\theta_0, \theta)$ as follows:

$$ \delta(\theta_0, \theta) = \min\left\{ \frac{\theta}{\theta_0} - 1 - \ln\left(\frac{\theta}{\theta_0}\right), \frac{\theta_0}{\theta} - 1 - \ln\left(\frac{\theta_0}{\theta}\right) \right\} \quad (3) $$

Mathematically, we can determine the minimum Kullback-Leibler equation with the conditions $\theta_0 > 0$ and $\theta > 0$. With the $\theta_0$ is a proxy from $\theta$ which the value cannot be determined, then Mathematica Software is used to be solved numerically. After determining the intrinsic discrepancy, the next step is to determine the survival function in type I censored samples.

b. Determine Jeffrey Prior’s for Type I Censored Exponential Distribution

The form of the survival function for exponential distribution is:

$$ S(t) = \exp\left(-\frac{t}{\theta}\right) \quad (4) $$
Type I censored sampling is by determining the number of observation samples as much as n and the censor time as \( L_i \). In taking type I censored samples, researchers can subject the censor deadline subjectively according to the research needs. The research stopped according to the predetermined time, if it exceeds the time limit, it is classified as censored observation and vice versa. Following is the form of the likelihood function of type I censored Exponential distribution:

\[
L(\theta | t) = \frac{1}{\theta} \exp \left\{ -\frac{\sum_{i=1}^{n} d_i}{\theta} \right\} \quad \text{with} \quad r = \sum_{i=1}^{n} d_i
\]  

(5)

Where, the \( d_i \) shows that the lifetime of \( T_i \). Censored or Not Censored

The form of prior distribution \( \pi(\theta) \) can be determined using the Jeffrey's method. In Jeffrey's method, it is known that the proportional to the square root of Fisher's information \( I(\theta) \) is obtained by looking for the second derivative of the function \( \ln(f(t | \theta)) \). The Fisher’s information formula is given below.

\[
I(\theta) = -E \left[ \frac{1}{\theta^2} - \frac{2t}{\theta^3} \right] = \frac{1}{\theta^2}
\]

(6)

Using the Jeffrey's method, the prior distribution is obtained as follows:

\[
\pi(\theta) \propto \frac{1}{\sqrt{\theta^2}} = \frac{1}{\theta}
\]

(7)

c. Determine Posterior Distribution for Type I Censored Exponential Distribution

Posterior distribution is obtained from the result of a combination of likelihood functions with prior distributions that have been previously known.

\[
\pi(\theta) = \frac{\left( \sum_{i=1}^{n} t_i \right)^r \exp \left\{ -\frac{\sum_{i=1}^{n} t_i}{\theta} \right\}}{\theta^r + 1 \Gamma(r)}
\]

(8)

d. Parameter Estimation Results \( \theta \) with Objective Bayesian Method

The last step to estimate with the Objective Bayesian method is to determine the intrinsic statistics and \( \hat{\theta}^* \) based on 2.1 as follows

\[
d(\theta_i | t) = \int_0^\infty \frac{\delta(\theta_i | t) \exp \left\{ -\frac{\sum_{i=1}^{n} t_i}{\theta} \right\}}{\theta^r + 1} \, d\theta
\]

(9)

\[
\hat{\theta}^* = \arg \min_{\theta \in \Theta} \int_0^\infty \frac{\delta(\theta | t) \exp \left\{ -\frac{\sum_{i=1}^{n} t_i}{\theta} \right\}}{\theta^r + 1} \, d\theta
\]

(10)

Equation (10) will determine the estimated value of the parameter for \( \theta \). The solution for Equation (10) is not closed form, so we use a Mathematica software to run the estimation.
3.2 Application Of Exponential Distribution Parameter Estimation Results On Type I Censored Data Of The Lifetime For Patients With Chronic Renal Failure

The data used in this research are data on patients with chronic renal failure undergoing Hemodialysis therapy with the initial cause is non-diabetic disease. In this research 63 samples were observed together, after that built the samples with bootstrap value consists of 10, 100, and 1000 times and the observation was stopped until the 100th day. The results of the parameter estimation are as follows:

| Bootstrap | Estimate | MSE     |
|-----------|----------|---------|
| 10        | 165.358  | 1841.303|
| 100       | 166.825  | 951.5675|
| 1000      | 162.280  | 922.2375|

Table 1 shows the results of parameter estimation when bootstrapping is 10, 100, and 1000 times. It can be said that estimated $\hat{\theta}$ values obtained for each bootstrap sample size are not much different, but for $MSE$ values can conclude that the more bootstrap simulation performed the smaller $MSE$ value will produced. Furthermore, the confidence interval for the parameter $\hat{\theta}^*$ is calculated using the bootstrap method. So that the lower and upper limits are obtained as in Table 2, and when presented in the form of a histogram can be seen in Figure 1.
Table 2. Confidence Intervals for Each Bootstrap

| Bootstrap Sample Size | Lower Bound | Upper Bound |
|-----------------------|-------------|-------------|
| 10                    | 98.8478     | 224.571     |
| 100                   | 117.318     | 235.893     |
| 1000                  | 112.857     | 225.69      |

Based on table 2 that obtain from smallest value and biggest value from estimated parameter value obtained. With a 95% confidence interval seen for the bootstrap, the bigger the confidence interval produced will be narrower. This is in accordance with the illustration in Figure 2, for bootstrap observations as much as 10 times shows that the beam tends to be wide. When bootstrap is done 100 times, the beam size will be narrower, and so on when bootstrap is done 1000 times, the beam size will narrow. This shows that the more bootstrap is carried out, the estimation results for each test will be conical. So that the confidence interval is getting shorter.

4. Conclusion

The application of estimation results to the data on chronic renal failure patients shows that the more bootstrap simulations performed can produce smaller \( MSE \) values. The bigger the bootstrap the better the estimation \( \hat{\theta} \) is generated. Thus, the best fit model is \[ f(t) = \frac{t}{112.857} \exp\left(-\frac{t}{112.857}\right), \] with the \( t \) are consists of lifetime. With a 95% confidence interval seen for the bootstrap where, the more bootstrap sample size performed then affect the confidence interval produced will be even shorter.

References

[1] Barlow R E, and Proschan F 1996 *Mathematical Theory of Reliability Society for Industrial and Applied Mathematics* (Philadelphia)
[2] Bernardo J M 2010 *Integrated Objective Bayesian Estimation and Hypothesis Testing* (Oxford: Oxford University Press)
[3] Bernardo J M, and Juarez M A 2010 *Intrinsic Estimation* (Bayesian Statistics vol 7) (Oxford: Oxford University Press) pp 465-476
[4] Cox D R 1953 *J.R. Stat. Soc. B* **21** 411-421
[5] Efron B, and Tibshirani R J 1993, *An Introduction to the Bootstrap*, Chapman and Hall (New York)
[6] Harinaldi 2006 *Statistics Principles for Engineering and Science* Erlangga (Jakarta)
[7] Lee E T 1992 *Statistical Methods for Survival Data Analysis* John Willey & Sons, Inc. (Canada).
[8] Lee E T, and Wang J W 2003 *Statistical Methods for Survival Data Analysis*, 3th Edition (Oklahoma: University of Oklahoma Health Sciences Centre)
[9] Shanker R, Fesshaye H, and Selvaraj S, On *Modelling of Lifetimes Data Using Exponential and Lindley Distribution, Biom Biostat Int, 2*(5).
[10] Solih A., Cahyanadari R, and Tarkin 2017 *Estimation of Confidence Intervals Parameters of Weibull Geometric Process Models in Life Test Analysis for Censored Type Data II* ISTEK 10(1).