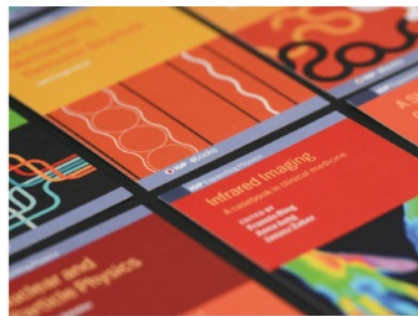# C/Triana_2019_J._Phys

by Matematika Pak Joko

PAPER · OPEN ACCESS

Exponential Distribution Parameter Estimation with Bayesian SELF Method in Survival Analysis

View the article online for updates and enhancements.

Exponential Distribution Parameter Estimation with Bayesian SELF Method in Survival Analysis

Yuli Triana[1], Joko Purwadi[1,*]

[1]Mathematics Study Program, Faculty of Applied Science and Technology, Ahmad Dahlan University Yogyakarta

* Email: joko@math.uad.ac.id

Abstract. This Paper discussed the Exponensial distribution parameter estimation using Bayesian SELF method in survival analysis with $\hat{\theta}$ as SELF Bayesian estimator for the θ parameter. Survival analysis corresponds to a method that relates to time starting from the start point to the occurrence of a particular event or an endpoint. The event referred in this research is the event of death. The distribution used in this research is an exponential distribution. Bayesian SELF estimation of $\hat{\theta}$ on exponential distribution for the censored data was obtained by minimizing expectations of loss function. Application of Bayesian SELF method on acute coronary syndrome patient's data, it was obtained $\hat{\theta} = 0.18583$ that indicates the patients' probability to survive is high and the probability for the patients to fail is low.

1. Introduction

Survival analysis or life-resistant analysis is a method related to the time of origin or start point until the occurrence of a special event or end point. The Data obtained in the field of health is an observation of the observed patient and the time of the failure of each individual [1]. The difference between survival analysis and other statistical analyses is the presence of censored observation concepts and uncensored observations. In analyzing the survival data, there are two models that can be used that are parametric models and nonparametric models. Some parametric models consist of the Weibull distribution, Log-Normal distribution, LOG-logistics distribution and exponential distribution. In the field of health, exponential distribution can be used to examine the survival data, where the research is derived from the survival data in acute coronary syndrome cases by estimating the parameters of the distribution. In this study used the Bayesian *SELF* method obtained by minimizing the expectation of loss function.

The purpose of this study is to determine the estimation of parameters of the exponential distribution and apply them to the case data of acute coronary syndrome sufferers. The time distribution of survival is usually described in three functions, namely survival function, probability density function and hazard function.

2. Material and Methods

Probability density function is the probability an individual dies or experiences instantaneous events in a time interval t until $t + \Delta t$ which is equivalent to $f(t)$. Survival function defined as an probbility stating that an individual can survive until the time t [1], which is defined as the following:

$$S(t) = 1 - F(t). \tag{1}$$

If T random variables are not negative at intervals $[0, \infty)$ That indicates the time of the individual experiencing events in a population, the chances that individuals experience events at intervals $(t + \Delta t)$ is defined as hazard function [1] $h(t)$:

$$h(t) = \frac{f(t)}{S(t)}. \tag{2}$$

Exponential distribution is the most commonly used distribution in the modeling of reusability and analysis of life tests. If T is the survival time of the T continuous random variable that follows the exponential distribution with θ parameters, then, probability density function of exponential distribution is the probability density function of exponential distribution [2] is:

$$f(t) = \theta e^{-\theta t}, t > 0, \theta > 0 \tag{3}$$

the cumulative distribution function for exponential distribution is:

$$F(t, \theta) = 1 - e^{-\theta t} \tag{4}$$

Bayesian is a method that views parameters as variables describing the initial information about the parameters before the observation is performed and expressed in a distribution called the prior distribution [3]. Furthermore, the posterior (posterior distribution) information is a combination of two sources of information regarding the parameters of the statistical model, i.e., likelihood from the distribution of samples and preliminary information (prior distribution). The result is expressed in a posterior distribution form which then becomes the basis in the Bayesian method.

Bayesian SELF method Defined as follow :

$$L(\theta, \hat{\theta}) = (\hat{\theta} - \theta)^2$$

Where $\hat{\theta}$ is the Bayesian SELF estimator for the θ parameter [4]. The Bayesian estimate of θ on the exponential distribution for censored data is achieved by minimizing the expectations of the loss function obtained as follows:

$$\frac{\partial \left[E\left(L(\theta, \hat{\theta}) \right) \right]}{\partial \theta} = 0$$

$$\hat{\theta} = E(\theta) \tag{5}$$

Likelihood function is defined as join density function of n random variabel ([5],[6]). join density function of $f_{t_1, t_2, \cdots, t_n}(t_1, t_2, \cdots, t_n; \theta)$ That considers the function of θ. if $t_1, t_2, \cdots, t_n$ is a random sample of the density function $f(t; \theta)$, join density function is:

$$L(\theta) = \prod_{i=1}^{n} (t_i; \theta).$$

On the Bayesian approach selects the prior distribution which shows uncertainty about the unknown θ parameter being a major problem in this approach. Based on the likelihood function the prior distribution is divided into two groups [7]: 1) Based on the distribution of data pattern identification result there are prior conjugates and prior non-conjugates, 2) based on the

determination of each of the parameters of the prior distribution pattern there are prior informative and prior non-informative.

Posterior distribution is a conditional density function of θ if known as the value of T observation, on the Bayesian method, inferences are based on the posterior distribution. So the posterior distribution can be defined as [4]:

$$f(\theta|t) \propto L(t;\theta)f(\theta)$$

The symbol of $\propto$ states that the posterior distribution is propotional or comparable to the likelihood function if multiplied by the prior distribution. When it is continuous θ, the prior and posterior distribution of θ can be presented with density function ([8[[9]). Conditional density function of one random variable if known value of the second random variable is just a join density function with two random variables it is divided by a marginal density function of second random variables. Join density function of $f(\theta, t)$ and marginal density function $g(t)$ is generally unknown, Only the prior distribution and the likelihood function are usually expressed. So the posterior density function for continuous random variables can be written as [10]:

$$f(\theta|t) = \frac{f(t;\theta)f(\theta)}{\int_{-\infty}^{\infty} f(\theta) f(t;\theta)d\theta} \tag{6}$$

The stages of research on the exponential distribution using the SELF Bayesian method is as follows:

1. Determine the survival function, and the hazard function.
2. Determine the likelihood function.
3. Formulating the prior and posterior distribution.
4. Estimate exponential distribution parameters with Bayesian SELF method.
5. Apply the estimation parameters of the Bayesian SELF method in acute coronary syndrome patient data.

3. Results and Discussion

3.1 Survival function and Hazard function of exponential distribution

Based on the equation (1), the survival function of t is:

$$S(t;\theta) = 1 - F(t;\theta)$$
$$= e^{-\theta t} \tag{7}$$

Based on the equation (2), the hazard function of t is:

$$h(t) = \frac{f(t)}{S(t)}$$
$$= \theta \tag{8}$$

3.2 Likelihood function on the Bayesian SELF method

The likelihood function is a function that is considered a function of θ, for $t_1, t_2, \cdots, t_n$ that is fixed. In the censored data of observation (t_i, δ_i) the likelihood function is:

$$L(\theta) = \prod_{i=1}^{n} [f(t_i)]^{\delta_i}[S(t_i)]^{1-\delta_i}$$

with δ_i is a censorship indicator, $\delta = 1$ if the data is not censored and $\delta = 0$ if the data is censored. The value of t_i is the survival time of the patient with $i = 1,2,3,\cdots,n$. So the likelihood function of the exponential distribution for the censored data has the following forms:

$$L(t_i;\theta,\delta_i) = \prod_{i=1}^{n}\left[\theta e^{-\theta t_i}\right]^{\delta_i}\left[e^{-\theta t_i}\right]^{1-\delta_i}$$
$$= \theta^{\sum_{i=1}^{n}\delta_i}\, e^{-\theta\sum_{i=1}^{n}t_i}$$

3.3 Forming the Prior distribution

In the Bayesian method, when a particular distribution with a parameter in it is described with θ, it is likely that the θ parameter follows a particular distribution of probabilities called the prior distribution. In this research, Gamma distribution was set as a comrade distribution for exponential distribution with the θ parameter where $0 < \theta < \infty$. The prior distribution for θ-on exponential distribution with α and β parameters, the probability density function as follows

$$f(\theta) = \frac{1}{\beta^\alpha \Gamma(\alpha)}\theta^{\alpha-1}e^{-\frac{1}{\beta}\theta} \qquad \alpha > 0, \beta > 0, \theta > 0$$

Determination of α and β parameters for gamma distribution (α, β) which is used as a prior distribution can be solved by matching between the mean and variance of gamma distribution with the mean and exponential distribution variance. Thus obtained $\alpha = 1$ and $\beta = \frac{1}{\theta}$. With θ is the average of failure time.

3.4 Determining Posterior Distribution

The posterior distribution can be expressed by comparison between common density functions and marginal functions. So based on equations (6) Posterior distribution can be expressed as follows:

$$f(\theta,t) = \frac{\theta^{\sum_{i=1}^{n}\delta_i}\, e^{-\theta\sum_{i=1}^{n}t_i}.\frac{1}{\beta^\alpha\Gamma(\alpha)}\theta^{\alpha-1}e^{-\frac{1}{\beta}\theta}}{\int_0^\infty \theta^{\sum_{i=1}^{n}\delta_i}\, e^{-\theta\sum_{i=1}^{n}t_i}.\frac{1}{\beta^\alpha\Gamma(\alpha)}\theta^{\alpha-1}e^{-\frac{1}{\beta}\theta}\,d\theta}$$
$$\propto \theta^{\sum_{i=1}^{n}\delta_i}\, e^{-\theta\sum_{i=1}^{n}t_i}\theta^{\alpha-1}e^{-\frac{1}{\beta}\theta}$$
$$\propto \theta^{\sum_{i=1}^{n}\delta_i+\alpha-1}e^{-\theta(\sum_{i=1}^{n}t_i+\frac{1}{\beta})}$$

Thus, the posterior distribution of the exponential distribution is Gamma distribution $\left(\alpha = \sum_{i=1}^{n}\delta_i + \alpha, \beta = \left(\sum_{i=1}^{n}t_i + \frac{1}{\beta}\right)^{-1}\right)$.

3.5 Estimation Parameter of Bayesian SELF Method

Because in this research using the estimation of the Square Error Loss Function (SELF) parameter, the Bayesian estimator used is the mean of the posterior distribution. So based on equations (5) obtained:

$$\hat{\theta} = E(\theta)$$
$$\hat{\theta}_{BS} = \frac{(\sum_{i=1}^{n}\delta_i + \alpha)}{\left(\sum_{i=1}^{n}t_i + \frac{1}{\beta}\right)} \qquad (9)$$

from the equation (9) can be obtained estimation of the parameters of survival function and hazard function with the Bayesian SELF of the exponential distribution method is:

$$\hat{S}_{BS}(t_i; \hat{\theta}_{BS}) = e^{-\frac{(\sum_{i=1}^n \delta_i + \alpha)}{(\sum_{i=1}^n t_i + \frac{1}{\beta})} t_i} \tag{10}$$

$$\hat{h}_{BS}(t_i; \hat{\theta}_{BS}) = \frac{(\sum_{i=1}^n \delta_i + \alpha)}{(\sum_{i=1}^n t_i + \frac{1}{\beta})} \tag{11}$$

3.6 Case Study In Acute Coronary Syndrome Patient Data

The data conformance test used is the Anderson-Darling test, by using the R software obtained output from the Anderson-Darling test of *p-value* = 0.08529. Because the value of *p-value* > α is 0.08529 > 0.05 then H_0 is accepted which means the data is exponential distribution.

3.7 Estimation parameters of Survival function and Hazard function

From acute coronary syndrome sufferers, it is known that $\sum_{i=1}^n \delta_i = 63$, $\sum_{i=1}^n t_i = 340$ dan $\beta = 0.227272$ So that the estimation of parameters with the Bayesian SELF method is:

$$\hat{\theta}_{BS} = \frac{(\sum_{i=1}^n \delta_i + \alpha)}{(\sum_{i=1}^n t_i + \frac{1}{\beta})}$$

$$\hat{\theta}_{BS} = 0.18583$$

So that it can be known the Bayesian SELF estimation for the θ parameter of the exponential distribution is 0.18583 which indicates the patient's probablity to endure is high and the probability of failure are low. And can also be known estimation of Bayesian SELF for survival function and hazard function as follows:

$$\hat{S}_{BS}(t_i; \hat{\theta}_{BS}) = e^{-\frac{64}{344,400141} t_i}$$

and if from the data is taken one sample that is in the individual to 63 with a long time treated is 12 days then the estimate obtained the Survival function as follows:

$$\hat{S}_{BS}(12; \hat{\theta}_{BS}) = e^{-\frac{64}{344,400141} 12}$$
$$= 0.10753$$

So if the patient's time treated is 12 days then the patient's probability to survive is 0.10753. and the hazard function estimation as follows

$$\hat{h}_{BS}(t_i; \hat{\theta}_{BS}) = \frac{(\sum_{i=1}^n \delta_i + \alpha)}{(\sum_{i=1}^n t_i + \frac{1}{\beta})}$$

$$\hat{h}_{BS}(t_i; \hat{\theta}_{BS}) = 0.18583$$

Survival Function dan Hazard Function Graph
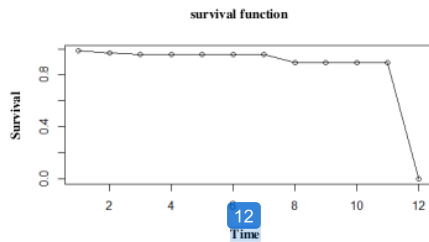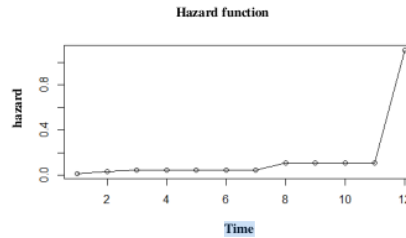
survival function

Hazard function

Figure 1

Figure 2

from Figure 1, it can be known that the graph of survival function is getting down closer to 0, which means the individual probability to survive the longer the more the decline. While from the figure 2, It can be noted that the graphic hazard is inversely proportional to the survival function graph. The graph of hazard function is getting up close to 1, That means a person's probability to experience failure or death is increasingly rising or higher.

4. Conslusion

Based on the analysis had been done, the research conclusion is as follows :

1. Estimation of exponential distribution parameters by Bayesian SELF method on survival analysis is obtained as follows:

$$\hat{\theta}_{BS} = \frac{\left(\sum_{i=1}^{n} \delta_i + \alpha\right)}{\left(\sum_{i=1}^{n} t_i + \frac{1}{\beta}\right)}$$

2. Application of estimation of exponential distribution parameters by Bayesian SELF method in survival analysis of acute coronary syndrome is $\hat{\theta}_{BS} = 0.18583$ That indicates the patient's probability to endure is high and the probability of the patient failing are low.

Reference

[1]   Ibrahim J G, Ming-Hui C, Sinha, Debajyot, 2001. *Bayesian Survival Analysis*. Springer-Verlag.

[2]   Ahsanullah, M. dan Hamedani G.G. 2010. *Exponentiated Distribution: Theory And Methods*. Book Series: Mathematics Research Developments New York: Nova Science Publishers, Inc.

[3]   Amelia S R 2018 *Estimasi Parameter Distribusi Weibull Dua Parameter Menggunakan Metode Bayes*. Makassar: Universitas Islam Negeri Alauddin Indonesia.

[4]   Luoma A 2014 *Introduction To Bayesian* Finland: University of Tampere.

[5]   Bain L J and Engelhardt M 1992 *Introduction to Probability and Mathematical Statistic*. *California*: Duxbury of Wathfor, Inc.

[6]   Rarasati I P 2012 *Estimasi Parameter Distribusi Gamma Dengan Metode Bayes* Malang: Universitas Isam Negeri Maulana Malik Ibrahim Indonesia.

[7]   Box, G E P and Tiao G C 1973 *Bayesian Inference In Statistical Analysis* Addison-Wesley: Publishing Company,Inc.

[8]   Collect D, 1993 *Modelling Survival Data in Medical Research* New York: Chapman and Hall.

[9]   Sagita I, Satyahadewi N, Rizki S W 2018 *Bimaster* **7** 353-362

[10] Aulia R, Fajriah N, Salam N, 2011 *Epsilon* **5** 40-52

C/Triana_2019_J._Phys