

# C/Elviana\_2020\_J.\_Phys.

*by* Matematika Pak Joko

---

**Submission date:** 03-Feb-2021 01:44PM (UTC+0700)

**Submission ID:** 1500545060

**File name:** Elviana\_2020\_J.\_Phys.\_\_Conf.\_Ser.\_1503\_012004.pdf (867.36K)

**Word count:** 2385

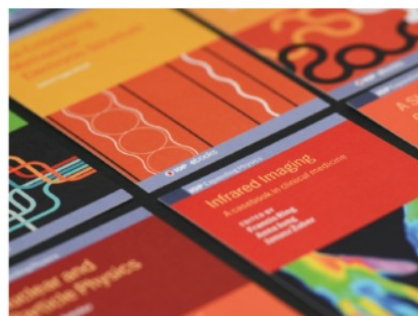
**Character count:** 12795

PAPER • OPEN ACCESS

## Parameters Estimation Of Rayleigh Distribution In Survival Analysis On Type Ii Censored Data Using The Bayesian Method

To cite this article: Erma Elviana and Joko Purwadi 2020 *J. Phys.: Conf. Ser.* **1503** 012004

View the [article online](#) for updates and enhancements.



**IOP ebooks™**

Bringing together innovative digital publishing with leading authors from the global scientific community.

Start exploring the collection—download the first chapter of every title for free.

## Parameters Estimation Of Rayleigh Distribution In Survival Analysis On Type Ii Censored Data Using The Bayesian Method

Erma Elviana<sup>1</sup>, Joko Purwadi<sup>2</sup>

<sup>1,2</sup>Mathematics Study Program, Faculty of Applied Science and Technology,  
Ahmad Dahlan University Yogyakarta

Email: [erma1500015027@webmail.uad.ac.id](mailto:erma1500015027@webmail.uad.ac.id),

**Abstract.** This research aims to determine the estimation of Rayleigh's distribution parameters on the data survival analysis of type II of tuberculosis patients. The method used in this research is the Bayesian method to perform estimate parameters. The process estimation parameter using the Bayesian method requires information from the likelihood function and the prior distribution. The prior distribution used was the prior distribution of Jeffreys. In this study, estimates of Rayleigh's distribution parameters were applied in the survival analysis data of tuberculosis sufferers. And the value of estimated parameters in patients with tuberculosis acquired  $\hat{\theta} = 0.001097324$  and the changes of the survival value decreased near to zero.

### 1. Introduction

The development of science and technology is very fast, especially in the field of mathematical statistics that have developed so far by the discovery of analytical tools that can be used to analyze a problem. The statistical analysis used to analyze the lifetime data is called Life-resistant analysis (Survival analysis).

Survival analysis is a collection of statistical procedures for the analysis of data where the variables studied are the time until the occurrence of the end-point [1]. The method used in survival analysis relies heavily on the assumption of the distribution of data used. The complete data form if all objects or individuals are tested to the occurrence of death or failure. Type I censored Data is the entire research object conducted for a certain time already set by the researcher. Type II censored data is the data generated when the test is terminated after death or failure [2]. Censored type II data is a lifetime data that there is an observation in a sample of  $n$ -size random with  $1 \leq r \leq n$ . In a research, type II censorship is more commonly used, because in this life test there is an observation as much as  $n$ , but the research is discontinued when the observation has failed to  $r$ , so that researchers can save time and money. To analyze live test data, a distribution is required. There are two models used in live test analysis, namely non-parametric models and parametric models. Non-parametric models are used if data does not follow certain assumptions. Parametric models are used when data follows



a specific distribution assumption. The research uses parametric models, the distribution used is Rayleigh's distribution.

The distribution of Rayleigh was first introduced in 1880 by an English physicist, Baron John William Strutt Rayleigh. Rayleigh's distribution is a special form of two-parameter Weibull distributions [4]. To find out if the assumed distribution of lifetime data illustrates the real state, an analysis of life-time data is required. The step to analyze the distribution function of lifetime data is to estimate the value of its distribution parameters. In statistics, the main branch of inference statistics is an estimation. There are two methods to estimate the parameters namely, the classic method and Bayes method. In the classic method The parameter is a fixed magnitude, while the Bayes method is used to find the posterior distribution [5].

## 2. Material and Methods

Rayleigh's distribution is a special form of two-parameter Weibull distributions, in some special cases the form ( $\beta$ ) parameters, from the Weibull distribution are value  $\beta = 2$ , known as Rayleigh's distribution [5]. The characteristic of Rayleigh's distribution is the hazard function that increase linearly against time [6]. The probability density function of Rayleigh's distribution is expressed as follows:

$$f(t) = 2\theta t e^{-\theta t^2} \quad (1)$$

There are two approaches to solve the  $\theta$  parameter, first is ia using the classical statistical approaches and the second is the Bayesian statistical approaches. The classic statistics approaches are fully relies on the inference process on sample data taken from the population. Whereas Bayesian statistics approaches, besides utilizing sample data obtained from the population also take into account an initial distribution. Statistical inference with Bayesian statistics approach differs from the classic statistical approach. The classic statistics approach looks at the  $\theta$  parameter as a fixed value parameter. While the Bayesian statistical approach views the parameter as a random variable that has a distribution, it is called the prior distribution. Subsequent distributions can be specified posterior distribution so that the acquired Bayesian estimator which is the average or mode of the posterior distribution [8].

### 2.1 Likelihood Function

Let  $X_1, X_2, \dots, X_n$  is a random variables with probability density function  $f(x_1, x_2, \dots, x_n; \theta)$  where  $\theta$  is the parameter that will be estimated. If  $L$  is a joint density function of  $X_1, X_2, \dots, X_n$  which is in view as a function of  $\theta$  then Likelihoodnya function is:

$$L(\theta; t_i) = f(x_1, x_2, \dots, x_n; \theta) = \prod_{i=1}^n f(x_i; \theta) \quad (2)$$

### 2.2 Prior and Posterior Distribution

The Prior distribution is grouped into two groups based on the form of the Likelihood function The first is related to the distribution form of data pattern identification of the prior conjugates (conjugate) and prior distribution non-conjugate. Secondly, related to the determination of each of the parameters in the Prior distribution pattern i.e. the prior distribution of informative and distribution of prior non-informative.

The posterior distribution is a conditional density function if known to the value of  $T$  observation, on the Bayesian method, inferences are based on the posterior distribution. So the posterior distribution can be expressed as follows [7]

$$f(\theta|t_i) \propto L(t_i; \theta)f(\theta) \quad (3)$$

The posterior density function for continuous random variable is as follows:

$$f(\theta|t_i) = \frac{f(\theta)f(t_i; \theta)}{\int_0^\infty f(\theta)f(t_i; \theta) d\theta} \quad (4)$$

The posterior distribution can be used to determine an estimated intervals of unknown parameter.

### 2.3 Survival Analysis and Censored Data

Survival analysis is a collection of statistical procedures for data analysis for which the outcome variable of interest is time until an event occurs [1].

there are three factors needed to determine the time of survival [3], namely

1. The time origin (starting point) of an event.
2. Time of final occurrence ((endpoints)) of an event.
3. The scale of measurement as part of time should be the event.

There are three types of censorship that is type I-censored data that is all observation objects conducted over a certain time set by the observation, type II sensor data occurs when the observation is terminated after  $n$  observations, there is  $r$  failure, damage or death with  $r < n$  and  $r$  is a random sample of  $n$  and the type III sensor data that occurs if the incoming object does not jointly, so that the sensory time is different[3].

Cumulative distribution function  $F(t)$  for continuous distribution with the probability density function  $f(t)$  is stated as follows [7]

$$F(t) = P(T \leq t) \text{ atau } F(x) = \int_0^t f(x) dx \quad (5)$$

### 2.4 Survival function and Hazard Function

Survival function denoted by  $S(t)$ , defined as the probability that the individual lasts longer than  $T$  with:

$$S(t) = P(T > t) = 1 - F(t) \quad (6)$$

Hazard function  $H(t)$  is the probability of a person failing after a specified unit of time, such as the opposite of the survival function  $S(t)$  [9],

$$h(t) = \frac{f(t)}{S(t)} \quad (7)$$

The goodness of fit distribution on data is an early stage of data analysis. To perform the conformity test the distribution was used in a Anderson-Darling test. The Anderson-Darling test is used to test whether the data follows a particular distribution.

Statistical test are :

$$A^2 = -n - \frac{1}{n} \sum_{i=1}^n (2i - 1)[\ln F(X_i) + \ln (1 - F(X_{n+1-i}))] \quad (8)$$

## 3. Results And Discussion

The cumulative distribution function for the Rayleigh distribution with the  $\theta$  parameter on the survival data is written as follows:

$$F(t) = 1 - e^{-\theta t^2}, \quad (9)$$

the survival function of Rayleigh's distribution is as follows:

$$S(t) = e^{-\theta t^2}, \quad (10)$$

and the hazard function of Rayleigh's distribution is as follows:

$$h(t) = 2\theta t. \quad (11)$$

### 3.1 Likelihood function on type II censored data

Type II censored data is derived from research or observation of live test on a research object which is product, individual, system, unit or component. Unknown  $n$  is the number of research objects and  $r$  (object of research damaged, failed or dead) which is a random sample of  $n$  where  $r < n$  and  $r$  is the sequential survival time  $(t_1, t_2, t_3, \dots, t_r)$  in a study.

Likelihood function for type II sensor data with  $\theta$  parameter in general is as follows:

$$L(\theta; t_i) = f(t_i; \theta) = \frac{n!}{(n-r)!} \left[ \prod_{i=1}^r f(t_i) \right] [S(t_r)]^{n-r} \quad (12)$$

next to the likelihood function of type II sensor data with the  $\theta$  parameter on the Rayleigh distribution then the equation substitution (1) and (10) to equation (12) so that it is obtained:

$$L(\theta; t_i) = \frac{n!}{(n-r)!} (2\theta)^r \left( \prod_{i=1}^r t_i \right) \exp \left( - \left( \sum_{i=1}^r \theta t_i^2 + \theta t_r^2 (n-r) \right) \right) \quad (13)$$

### 3.2 Prior and Posterior Rayleigh Distribution

In this study, to determine the prior distribution for  $\theta$  is used the Jeffreys method. The Jeffreys approach was used to obtain the prior distribution of the prior non informative distributions [10]. The process for determining prior Jeffreys is as follows:

$$\ln L(\theta; t_i) \quad (14)$$

$$= \ln \frac{n!}{(n-r)!} + r \ln(2\theta) + \sum_{i=1}^r \ln t_i - \left( \sum_{i=1}^r \theta t_i^2 + \theta t_r^2 (n-r) \right)$$

$$\frac{\partial \ln L(\theta; t_i)}{\partial \theta} = \frac{r}{\theta} - \left( \sum_{i=1}^r \theta t_i^2 + \theta t_r^2 (n-r) \right) \quad (15)$$

$$\frac{\partial^2 \ln L(\theta; t_i)}{\partial \theta^2} = -\frac{r}{\theta^2} \quad (16)$$

Functions can be obtained by taking a square root of the Fisher information ie:

$$I(\theta) = -E \left( \frac{\partial^2 \ln L(\theta; t_i)}{\partial \theta^2} \right) = \frac{r}{\theta^2} \quad (17)$$

therefore, the results of the prior distribution for  $\theta$  with Jeffreys rules are

$$g(\theta) \propto \sqrt{\theta} = \sqrt{\frac{r}{\theta^2}} = \frac{1}{\theta} \sqrt{r} \quad (18)$$

The posterior distribution for the  $\theta$  parameter is,

$$f(\theta|t_i) = \frac{f(\theta)f(t_i; \theta)}{\int_0^\infty f(\theta)f(t_i; \theta) d\theta}$$

$$= \frac{\frac{1}{\theta} \sqrt{r} \frac{n!}{(n-r)!} (2\theta)^r (\prod_{i=1}^r t_i) \exp(-(\sum_{i=1}^r \theta t_i^2 + \theta t_r^2 (n-r)))}{\int_0^\infty \frac{1}{\theta} \sqrt{r} \frac{n!}{(n-r)!} (2\theta)^r (\prod_{i=1}^r t_i) \exp(-(\sum_{i=1}^r \theta t_i^2 + \theta t_r^2 (n-r))) d\theta}$$

with  $T = (\sum_{i=1}^r t_i^2 + t_r^2 (n-r))$  then obtained,

$$f(\theta|t_i) = \frac{T^r \theta^{r-1} \exp(-\theta T)}{\Gamma(r)} \tag{19}$$

The estimate of the  $\theta$  parameter is

$$\begin{aligned} \hat{\theta} &= E(\theta) \\ &= \int_0^\infty \theta f(\theta|t_i) d\theta \\ &= \int_0^\infty \theta \frac{T^r \theta^{r-1} \exp(-\theta T)}{\Gamma(r)} d\theta \\ &= \frac{r}{T} \end{aligned} \tag{20}$$

### 3.3 Data Analysis

The data used in this research is data from tuberculosis patients in the hospital of PKU Muhammadiyah Yogyakarta. The period of termination of data taken from 2016 until 2018. There were 51 patients with tuberculosis and 13 patients who had failed (died) or  $R = 13$ .

The Goodness of fit test test is carried out using the R software obtained by the following results:

Table 1: Output results of Anderson Darling

A	p-value
0.54933	0.1497

The value of P-value = 0.1497 > 0.05 which means  $H_0$  is accepted, meaning that data follows Rayleigh's distribution. The value of  $\theta$  parameter of the Rayleigh distribution can be calculated as follow,

$$\hat{\theta} = \frac{r}{\sum_{i=1}^r \theta t_i^2 + \theta t_r^2 (n-r)} = 0.001097324$$

By using the value of the estimated Bayes of  $\theta$  parameter, then the probability of survival and failure rate of tuberculosis patients can be known.

Table 2. probability of survival and failure rate

Time	S(t)	h(t)
1	0.998903278	0.002194648
2	0.995620322	0.004389297
3	0.990172689	0.006583945
4	0.982596042	0.008778594
5	0.972939765	0.010973242
6	0.961266424	0.013167891
7	0.947651096	0.015362539
8	0.932180559	0.017557187
9	0.914952377	0.019751836
10	0.896073873	0.021946484



11	0.875661017	0.024141133
12	0.853837241	0.026335781
13	0.830732205	0.028530430
14	0.806480511	0.030725078
15	0.781220414	0.032919727
17	0.728238482	0.037309023

12

Based on table 2, it can be seen that the survival rate is continuing to be worth 0.728238482 on the 17th day, while the hazard value is getting bigger on the 17th day of 0.037309023.

The survival rate of failure odds and hazard value failure rate in table 2 can be formed chart for the survival and hazard values as follows:

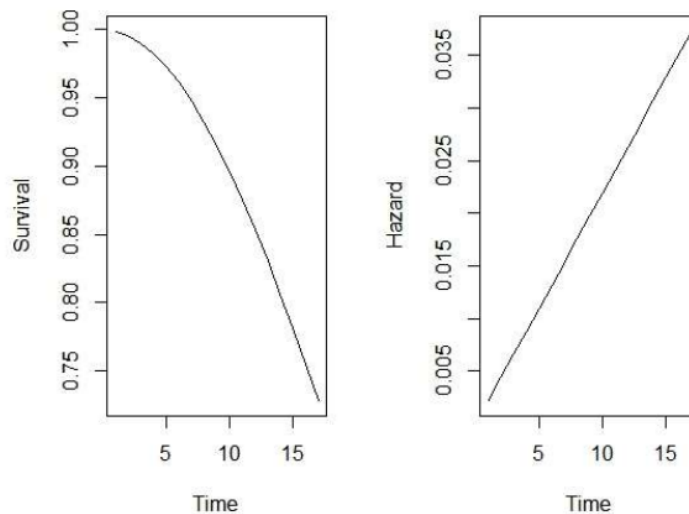


Figure 1. Survival Probability and failure rate graphs

Based on Figure 1 it can be concluded that the graphs on the odds of survival or survival value with the estimation of parameters using the Bayesian method can produce a moving survival value decreased close to zero, according to the characteristics of Survival function. In the graph of failure rate or hazard value indicates that of hazard function with parameter estimation using Bayesian method moves up linear time. The result of the hazard value corresponds to the characteristics of Rayleigh distribution.

#### 4. Conclusion

Conclusions that can be taken from the results of analysis and discussion on this research are as follows:

- i. The  $\theta$  parameter estimated for Rayleigh's distribution analysis on type II censored data using the Jeffreys prior are :

$$\hat{\theta} = \frac{r}{\sum_{i=1}^r t_i^2 + t_r^2(n-r)}$$



- ii. The result of parameter estimation is applied to the survival analysis of type II of patients with tuberculosis (TB), with  $n = 51$  and  $r = 13$  gives the value of  $\hat{\theta} = 0.001097324$ .
- iii. Survival probability or survival values with parameter estimates using the Bayes method can result in a moving survival value decreased close to zero, according to the characteristics of the survival function. Failure rate or hazard value indicates that of hazard function with parameter estimation using Bayes method Move up linearly to time. The result of the hazard value in this research is in accordance with the characteristics of Rayleigh distribution.

## 5. Reference

- [1] Kleinbaum, D. G., dan Klein, M. 2005. *Survival analysis: A Self-learning Text*. Springer.
- [2] Bain, L. J., dan Engelhardt, M. 1992. *Introduction to Probability and Mathematical Statistics*. California: Duxbury of Wathfor, Inc.
- [3] Lee, E. T. dan Wang, J. W. 2003. *Statistical Methods for Survival Data Analysis Third Edition*. Canada: Wiley
- [4] Dey, S., dan Dey, T. 2011.. *Statistical Journal*. **9** 213-226.
- [5] Lukitasari, A. D., Setiaji, A., Sasongko, L. R. 2015. *Dcartesian Jurnal* **4** 26-33.
- [6] Dey S, Maiti SS. 2012. *Electronic Journal of Applied Statistical Analysis*. **5(1)** 44-55
- [7] Lomax, Arto. 2014. *Introduction To Bayesian*. Finland: University of Tampere.
- [8] Hidayah, Entin. 2010. *Disagregasi Data Hujan Temporal Dengan Pendekatan Bayesian Sebagai Input Pemodelan Banjir*. Surabaya:ITS.
- [9] Lawless, J. F. 1982. *Statistical Models and Methods for Lifetime Data*. New York: John Wiley and Sons, inc.
- [10] Thamrin, S. A., Azhar, dan Jaya, A. K. 2018. *JUTEKS* **1**, 22-27.

ORIGINALITY REPORT

---

9%

SIMILARITY INDEX

7%

INTERNET SOURCES

6%

PUBLICATIONS

6%

STUDENT PAPERS

---

PRIMARY SOURCES

---

1

[jbe.tums.ac.ir](http://jbe.tums.ac.ir)

Internet Source

1%

2

[digitalcommons.wayne.edu](http://digitalcommons.wayne.edu)

Internet Source

1%

3

[repository.its.ac.id](http://repository.its.ac.id)

Internet Source

1%

4

[jurnal.uns.ac.id](http://jurnal.uns.ac.id)

Internet Source

1%

5

[epdf.pub](http://epdf.pub)

Internet Source

1%

6

[repository.unhas.ac.id](http://repository.unhas.ac.id)

Internet Source

1%

7

[stat.ethz.ch](http://stat.ethz.ch)

Internet Source

1%

8

Vikas Khare, Cheshta Khare, Savita Nema, Prashant Baredar. "Reliability Assessment Model", Elsevier BV, 2019

Publication

1%

---

9	Submitted to Higher Education Commission Pakistan Student Paper	1%
10	Patterson, R.. "Estimation of the variance of centring errors", Australian Surveyor, 1979. Publication	<1%
11	Submitted to Amity University Student Paper	<1%
12	<a href="http://isclo.telkomuniversity.ac.id">isclo.telkomuniversity.ac.id</a> Internet Source	<1%
13	<a href="http://silo.pub">silo.pub</a> Internet Source	<1%
14	<a href="http://www.tandfonline.com">www.tandfonline.com</a> Internet Source	<1%
15	<a href="http://research-information.bris.ac.uk">research-information.bris.ac.uk</a> Internet Source	<1%
16	Kusiak, A.. "On-line monitoring of power curves", Renewable Energy, 200906 Publication	<1%
17	Anwar Hossain, William Zimmer. "Comparison of estimation methods for weibull parameters: Complete and censored samples", Journal of Statistical Computation and Simulation, 2003 Publication	<1%

---

Exclude quotes      On

Exclude matches      Off

Exclude bibliography      On