Comparison of the Bayesian and Maximum Likelihood Estimation of Rayleigh Distribution for Right Censored Survival Data Type II

Rahajeng Ika D. P\textsuperscript{1,a}, Ida Fithriani\textsuperscript{2,b} and Siti Nurrohmah\textsuperscript{3,c}

\textsuperscript{1, 2, 3}Department of Mathematics, Universitas Indonesia, Depok, 16424, Indonesia.
E-mail: aajeng.desya@gmail.com, bida.fithriani@gmail.com, csnurrohmah@gmail.com

Abstract. In this paper, we used maximum likelihood and Bayes method to estimate parameter of Rayleigh distribution for right censored survival data type II to know the best method. The prior knowledge which used in Bayes method is Invers Gamma prior. Maximum likelihood and Bayes method under Square Error Loss Function will be compared. We compare these methods by Mean Squared Error (MSE) value using R program. After that, the result will be displayed in tables to facilitate the comparisons.

1. Introduction

The Rayleigh distribution is found to be useful in analyzing life time data in the fields of medicine, engineering, sciences and others. Life time or survival data cases usually used censored data, because to get all the life time data from a set of object observations it takes a lot of time and money. Censored data is data that cannot be observed in complete because there are some unobservable events. Abd Elfattah, et al [1] studied the efficiency of the maximum likelihood estimates of the parameter under three cases namely Type I, Type II and progressive Type II censored sampling. Howlader and Hossain [2] obtained Bayes estimators for the scale parameter and the reliability function in the case of Type II censored sampling. Dey [3] obtained Bayes estimators for the parameter and reliability function of the Rayleigh distribution under different loss function based on complete as well as Type II censored samples, also compared relative risk functions. Now in this study the two methods will be compared to see the estimates produced. Both methods will be compared based on the MSE value obtained. The smaller of MSE value the better estimated parameters will be.

The discussions of the paper are arranged as follows; likelihood function of Rayleigh distribution for right censored survival data type II, parameter estimation of Rayleigh distribution for right censored survival data type II with Maximum Likelihood and Bayes method, simulation study is discussed and the results are presented and followed by the conclusion.

2. Rayleigh distribution

The Rayleigh distribution has a wide range of applications including life testing experiments and clinical studies. This distribution has probability density function (PDF):

\[
f(x) = \frac{2x}{\theta} \exp \left( -\frac{x^2}{\theta} \right); \quad \theta > 0
\]  

And survival function
\[ S(x) = \exp\left(-\frac{x^2}{\theta}\right); \quad \theta > 0 \]  

(2)

### 3. Likelihood function of Rayleigh distribution for Right Censored Survival Data Type II

If an observation with the time of occurrence of the desired event is known for sure, this provides information about the probability that the event occurred at that time. For that reason, the probability density function (PDF) of \( X_i \) at a certain time \( x_i \) with \( i = 1, 2, ..., r \) contributes to the likelihood function for survival data. Mathematically it can be written as follows

\[ f(x_1) f(x_2) ... f(x_r) = \prod_{i=1}^{r} f(x_i) \]

For right censored observation, it is known that the time of occurrence of an event is greater than a certain time \( x_r \) (\( X_i > x_r \)). This means that the survival time is greater than \( x_r \). Survival function will contribute to the likelihood function for right censored data. Mathematically, the information provided by the observation is

\[ P(X_i > x_r) = S(x_r) \]

For right censored type II observations, it is known that the type II censored principle is that there are \( r \) observations of the observed \( n \) units or objects, and the experiment will be stopped after the \( r \)-event occurs. Units or objects that experience an event are the 1st to the \( r \)th units, each one component. Whereas units or objects that have not experienced an event as much as \( (n-r) \). The random sample size \( n \) with failure \( r \) objects experiencing the occurrence so that there is \( \frac{n!}{(n-r)!} \) possible sequence.

Assuming that the survival time is independent with censored time, the construction of the likelihood function for right type II censored data in the survival data is:

\[ f(x_i | \theta) = \frac{n!}{(n-r)!} f(x_1) f(x_2) ... f(x_r) \left[ P(X_{r+1} \geq x_r) ... P(X_n \geq x_r) \right] \]

\[ = \frac{n!}{(n-r)!} \left[ \prod_{i=1}^{r} f(x_i) \right] \left[ (1-P(X_{r+1} < x_r)) ... (1-P(X_n < x_r)) \right] \]

\[ = \frac{n!}{(n-r)!} \left[ \prod_{i=1}^{r} f(x_i) \right] \left[ (1-F(x_r)) ... (1-F(x_r)) \right] \]

\[ = \frac{n!}{(n-r)!} \left[ \prod_{i=1}^{r} f(x_i) \right] \left[ S(x_r) \right]^{n-r} \]

\[ = \frac{n!}{(n-r)!} \left[ \prod_{i=1}^{r} f(x_i) \right] \left[ S(x_r) \right]^{n-r} \]

Then substituting PDF and survival function of Rayleigh distribution from equation (1) and (2) obtained

\[ f(x_1, ..., x_n; \theta) = \frac{n!}{(n-r)!} \left( \frac{2}{\theta} \right)^r x_1^{r-1} \exp\left(-\sum_{i=1}^{n} \frac{x_i^2}{\theta}\right) \left[ \exp\left(-\frac{x_1^2}{\theta}\right) \right]^{(n-r)} \]

(3)
4. Parameter estimation of Rayleigh distribution for Right Censored Survival Data Type II with Maximum Likelihood

The maximum likelihood method uses the likelihood function which is defined as a joint pdf of random samples \( X_1, X_2, \ldots, X_n \) which has a probability function \( f(x; \theta): \theta \in \Omega \), where \( \theta \) is an unknown parameter and \( \Omega \) is the parameter space which is function of \( \theta \). But in survival analysis, the construction of the likelihood function is different from that described. The likelihood function depends on what information is obtained from the study of observations of the survival of each individual within a certain period of time.

The maximum likelihood method uses values in the parameter space \( \Omega \) that correspond to the maximum possible price of the observation data as an estimate of unknown parameters. In the application \( f(x_1, \ldots, x_n; \theta) \) shows the joint probability density function of the random sample. If parameter space is an open interval and \( f(x_1, \ldots, x_n; \theta) \) is a function that can be derived and is assumed to be maximum at \( \Omega \) then the maximum likelihood equation is:

\[
\frac{d}{d\theta} f(x_1, \ldots, x_n; \theta) = 0
\]

If the solution of the equation exists then the maximum of \( f(x_1, \ldots, x_n; \theta) \) can be fulfilled. If the completion of the equation is difficult to solve, then the \( f(x_1, \ldots, x_n; \theta) \) function can be logically algorithmic, with the maximum \( \ln f(x_1, \ldots, x_n; \theta) \), so that the natural logarithm equation maximum likelihood is:

\[
\frac{d}{d\theta} \ln f(x_1, \ldots, x_n; \theta) = 0
\]

The likelihood maximum estimator from \( \theta \) is obtained by solving the equation \( \frac{d}{d\theta} \ln f(x_1, \ldots, x_n; \theta) = 0 \). Based on equation (3) the natural logarithm (\( \ln \)) is drawn from the above likelihood function so that the log-likelihood function is obtained as follows:

\[
\ln f(x_1, \ldots, x_n; \theta) = \ln \frac{n!}{(n-r)!} + r \ln \left( \frac{2}{\theta} \right) + \left[ - \sum_{i=1}^{r} \left( \frac{x_i^2}{\theta} \right) + (n-r) \left( \frac{x_r^2}{\theta} \right) \right] + \sum_{i=1}^{r} \ln x_i
\]

By decreasing \( \ln f(x_1, \ldots, x_n; \theta) \) to the parameter \( \theta \) obtained:

\[
\frac{d}{d\theta} \ln f(x_1, \ldots, x_n; \theta) = -\frac{r}{\theta} + \frac{1}{\theta^2} \left( \sum_{i=1}^{r} x_i^2 + \frac{1}{\theta} x_r^2 (n-r) \right)
\]

Parameter estimation with maximum likelihood method is obtained by solving the equation \( \frac{d}{d\theta} \ln f(x_1, \ldots, x_n; \theta) = 0 \) so that it is obtained:
Parameter estimation of Rayleigh distribution for Right Censored Survival Data Type II with Bayes method

5.1 Prior Conjugate distribution for Rayleigh distribution

The initial step that needs to be done is to determine the prior distribution that will be used as the initial information (prior) of the parameters to be estimated. The use of different prior distributions will greatly affect the posterior distribution produced. Determination of conjugate prior on type II censored survival data will be the same as uncensored data. This is because the censored data will only affect the construction or determination of the likelihood function, while the prior selection in prior conjugate does not see the likelihood form but looks at the functional form of the distribution that the data has. In data with Rayleigh distribution with sample space from zero to infinite, the presumed priors that might be from the Gamma family are assumed. It can be seen that the functional form of \( f(x) \) from (1) associated with \( \theta \) is \( \frac{1}{\theta} \exp\left(-\frac{x^2}{\theta}\right) \). This is the functional form of the probability density function (PDF) Inverse Gamma so that the possible conjugate prior is Invers Gamma. PDF Invers Gamma is

\[
f(x) = \frac{1}{\Gamma(\alpha)\beta^\alpha} \left(\frac{1}{x}\right)^{\alpha+1} \exp\left(-\frac{1}{x\beta}\right); x \geq 0, \alpha > 0, \beta > 0
\]

So that the known priors is

\[
f(\theta) \propto \left(\frac{1}{\theta}\right)^{\alpha+1} \exp\left(-\frac{1}{\theta\beta}\right)
\]  

(4)

Note that \( \theta \sim \text{Invers Gamma} \left(\alpha, \beta\right) \) or \( \theta \) inverse Gamma with the parameters \( \alpha \) and \( \beta \).

5.2 Posterior Distribution

The Invers Gamma prior (4) and likelihood function (3) that have been previously obtained will be combined to produce a posterior distribution based on the equation

\[
f(\theta \mid x_i) = \frac{f(\theta) f(x_i \mid \theta)}{\int f(\theta) f(x_i \mid \theta) \, d\theta} = \frac{\left(\frac{1}{\theta}\right)^{\alpha+r+1} \exp\left(-\frac{\beta T + 1}{\theta\beta}\right)}{\Gamma(\alpha + r) \left(\frac{\beta}{\beta T + 1}\right)^{\alpha+r}}
\]

(5)

The equation (5) with \( T = \sum_{i=1}^r x_i^2 + (n - r) x_i^2 \) is a PDF posterior that distributes Inverse Gamma with parameters \( \left(\alpha + r, \frac{\beta}{\beta T + 1}\right) \). It is evident that the prior Invers Gamma is a conjugate prior for \( \theta \) because the posterior is also distributed Inverse Gamma.
5.3 **Point Estimation with Square Error Loss Function (SELF)**

We will find the point estimator for the parameter θ with Square Error Loss Function. The point estimator can be searched by minimizing expectations of loss function or minimizing posterior risk. Therefore, we will look for θ values that minimize 

\[ \text{RP}(\hat{\theta}_{\text{BAYES}}) = E \left[ L \left( \theta, \hat{\theta}_{\text{BAYES}} \right) \right] = E \left[ \left( \theta - \hat{\theta}_{\text{BAYES}} \right)^2 \right] \]

\[ = \int_0^\infty \theta^2 f(\theta | x_i) d\theta - 2 \hat{\theta}_{\text{BAYES}} \int_0^\infty \theta f(\theta | x_i) d\theta + \hat{\theta}_{\text{BAYES}}^2 \]

\[ \frac{\partial \text{RP}(\hat{\theta}_{\text{BAYES}})}{\partial \hat{\theta}_{\text{BAYES}}} = 0 \Rightarrow \hat{\theta}_{\text{BAYES}} = \int_0^\infty \theta f(\theta | x_i) d\theta \Rightarrow \text{RP}(\theta_{\text{BAYES}}) = E \left[ \theta | x_i \right] = \sum_{i=1}^r x_i^2 + (n-r)x_r^2 + \frac{1}{\beta} \]

\[ = \frac{\sum_{i=1}^r x_i^2 + (n-r)x_r^2 + 1}{(\alpha + r - 1)} \]

6. **Bias estimator of Maximum Likelihood and Bayes method with Square Error Loss Function (SELF)**

Let \( Z = \sum_{i=1}^r x_i^2 + (n-r)x_r^2 + \frac{1}{\beta} = T + \frac{1}{\beta} \) be proved that Z is a sufficient statistic of \( \theta \). Based on the statistical theorem it is sufficient to find that \( u_i(x_1, x_2, ..., x_n) = T + \frac{1}{\beta} = Z \) is sufficient statistics for the parameter \( \theta \). So that 

\[ E \left[ \hat{\theta}_{\text{BAYES}} \right] = \frac{\sum_{i=1}^r x_i^2 + (n-r)x_r^2 + 1}{(\alpha + r - 1)^2} \neq \theta \] and 

\[ E \left[ \hat{\theta}_{\text{MLE}} \right] = \frac{(r-1)}{r \sum_{i=1}^r x_i^2 + (n-r)x_r^2} \neq \theta . \] This means to compare the two estimators using Mean Squared Error (MSE).

7. **Comparison of Bayes and Maximum Likelihood estimation of Rayleigh distribution for Right Censored Survival Data Type II**

The estimation of the Rayleigh distribution parameter for right censored survival data type II with the maximum likelihood method is obtained \( \hat{\theta}_{\text{MLE}} = \frac{\sum_{i=1}^r x_i^2 + x_r^2 (n-r)}{r} \) only depending on the observation values and the number of censored observations \( r \). Whereas with Bayes method, prior distribution is needed and merging with the likelihood function will produce posterior distribution. This posterior distribution then used to estimate distribution parameters. The point estimator with Bayes is 

\[ \hat{\theta}_{\text{BAYES}} = \frac{\sum_{i=1}^r x_i^2 + (n-r)x_r^2 + \frac{1}{\beta}}{(\alpha + r - 1)} \]

depending on the parameter value of the prior, namely \( \alpha \) and \( \beta \).

8. **Data simulation**

In this simulation study, we generated data using software R. We have chosen \( n = 50, 100 \) and \( 200 \) with intensity censored by \( 10\%, 20\%, 30\% \) and \( 40\% \) which distributed Rayleigh (\( \theta \)) with \( \theta = 0.25, 0.75, 1, 1.25, 1.5, 2 \). For Bayes method, we have prior parameter \( \alpha \) and \( \beta \), the values of prior parameter are \( \alpha = \ ...
0.5 and $\beta = 1.5$. Since the right censored survival data type II is determined a number of $r$ events before the observation begins that’s mean for $n = 50$ with censored 10% the value of $r = 90% \times 50 = 45$ and so on.

**Table 1.** $MSE(\hat{\theta}_{MLE})$ and $MSE(\hat{\theta}_{BAYES})$ of Rayleigh Distribution for Right Censored Survival Data Type II

| $\theta$ | $n$ | Intensity Censored | $r$ | $MSE(\hat{\theta}_{MLE})$ | $MSE(\hat{\theta}_{BAYES})$ |
|----------|-----|--------------------|-----|---------------------------|-----------------------------|
| 0.25     | 50  | 10%                | 45  | 4.168136e-05              | 3.559403e-05               |
|          |     | 20%                | 40  | 4.293842e-05              | 3.602277e-05               |
|          |     | 30%                | 35  | 4.503905e-05              | 3.66718e-05                |
|          |     | 40%                | 30  | 4.465284e-05              | 3.536566e-05               |
|          |     | 10%                | 90  | 3.817222e-05              | 3.520246e-05               |
|          |     | 20%                | 80  | 3.535033e-05              | 3.212743e-05               |
|          | 100 | 30%                | 70  | 3.75998e-05               | 3.376595e-05               |
|          |     | 40%                | 60  | 3.81235e-05               | 3.591647e-05               |
|          |     | 10%                | 180 | 3.665041e-05              | 3.518094e-05               |
|          |     | 20%                | 160 | 3.706997e-05              | 3.541012e-05               |
|          |     | 30%                | 140 | 3.721033e-05              | 3.531248e-05               |
|          |     | 40%                | 120 | 3.93864e-05               | 3.712004e-05               |
| 0.75     | 50  | 10%                | 45  | 4.043043e-05              | 3.237272e-05               |
|          |     | 20%                | 40  | 4.360808e-05              | 3.426109e-05               |
|          |     | 30%                | 35  | 4.549689e-05              | 3.46972e-05                |
|          |     | 40%                | 30  | 4.619239e-05              | 3.358283e-05               |
|          |     | 10%                | 90  | 3.912299e-05              | 3.506691e-05               |
|          |     | 20%                | 80  | 4.182964e-05              | 3.713581e-05               |
|          |     | 30%                | 70  | 4.266257e-05              | 3.72672e-05                |
|          |     | 40%                | 60  | 4.639075e-05              | 3.987533e-05               |
|          |     | 10%                | 180 | 3.812645e-05              | 3.610027e-05               |
|          |     | 20%                | 160 | 3.870463e-05              | 3.641258e-05               |
|          |     | 30%                | 140 | 3.917013e-05              | 3.654087e-05               |
|          |     | 40%                | 120 | 3.993308e-05              | 3.684552e-05               |
| 1        | 50  | 10%                | 45  | 2.301654e-05              | 1.61797e-05                |
|          |     | 20%                | 40  | 2.858527e-05              | 1.829968e-05               |
|          |     | 30%                | 35  | 2.952622e-05              | 1.974223e-05               |
|          |     | 40%                | 30  | 3.157513e-05              | 1.992548e-05               |
|          |     | 10%                | 90  | 2.215875e-05              | 1.867442e-05               |
|          |     | 20%                | 80  | 2.798178e-05              | 2.360973e-05               |
|          |     | 30%                | 70  | 3.125132e-05              | 2.600661e-05               |
|          |     | 40%                | 60  | 3.144288e-05              | 2.53458e-05                |
|          |     | 10%                | 180 | 1.609579e-05              | 1.457405e-05               |
|          |     | 20%                | 160 | 2.729082e-05              | 1.993997e-05               |
|          |     | 30%                | 140 | 2.763881e-05              | 2.511652e-05               |
|          |     | 40%                | 120 | 2.995348e-05              | 2.690649e-05               |
| 1.25     | 50  | 10%                | 45  | 2.191551e-05              | 3.194386e-05               |
|          |     | 20%                | 40  | 2.232876e-05              | 3.386336e-05               |
|          |     | 30%                | 35  | 2.845831e-05              | 4.350578e-05               |
|          |     | 40%                | 30  | 2.060253e-04              | 2.554076e-04               |
|          |     | 10%                | 90  | 1.618725e-05              | 2.026977e-05               |
|          |     | 20%                | 80  | 1.62967e-05               | 2.710499e-05               |
|          |     | 30%                | 70  | 2.17927e-05               | 2.797948e-05               |
|          |     | 40%                | 60  | 2.469491e-05              | 3.246317e-05               |
| 100      |     |                    |     |                           |                            |
Table 1 show us that MSE can be seen from the estimated parameters θ using the Maximum Likelihood method and Bayes method by using Square Error Loss Function (SELF) on the right censored data type II. Note that the censored intensity greatly influences the estimated parameter value θ, because the MSE value for estimation θ in the right censored data of type II will be greater as the intensity of the censor increases. This means that the smaller the intensity of the censored MSE value is also getting smaller. This indicates that along with the intensity of the censored or $r/n$ is smaller than 1, all estimates obtained are generally closer to the real value for various parameter values θ.

For all parameters θ actually that will be estimated, namely θ = 0.25, 0.75 and 1, the result is that the MSE value with Bayes method always gives the smallest MSE value compared to the MSE value with the maximum likelihood method. For all parameters θ actually that will be estimated, namely θ = 1.25, 1.5 and 2, it is obtained that the MSE value with the maximum likelihood method always gives the smallest MSE value compared to the MSE value with the Bayes method. Or with another sentence, for all the parameters θ actually that will be estimated, namely θ > 1, the result is that the MSE value with the maximum likelihood method always gives the smallest MSE value compared to the MSE value with the Bayes method.

The conclusions about the best method can be summarized in Table 2.

### Table 2. The Best Method

| θ     | n   | Best Method   | θ     | n   | Best Method   |
|-------|-----|---------------|-------|-----|---------------|
| 0.25  | 50  | Bayes         | 1.25  | 100 | Maximum Likelihood |
| 0.75  | 50  | Bayes         | 1.5   | 50  | Maximum Likelihood |
Bayes   100   Maximum Likelihood  100
Bayes   200   Maximum Likelihood  200

Based on Table 2 for all parameters \( \theta \) actually to be estimated, namely \( \theta = 0.25, 0.75, \) and 1 or for \( \theta \leq 1 \), it can be concluded that the estimation \( \theta \) with Bayes method is better to use because it gives the best estimate compared to estimation \( \theta \) with the Maximum Likelihood method. For all parameters \( \theta \) actually to be estimated, namely \( \theta = 1.25, 1.5 \) and 2 or for \( \theta > 1 \), it can be concluded that the estimation \( \theta \) with the Maximum Likelihood method is better to use because it gives the best estimate compared to the estimation \( \theta \) with the Bayes method.

9. Conclusion
The point estimator by Maximum Likelihood method for right censored survival data type II is

\[
\hat{\theta}_{\text{MLE}} = \frac{\sum_{i=1}^{r} x_i^2 + x_r^2(1 - r)}{r}
\]

and Bayes method using Square Error Loss Function (SELF) for right censored survival data type II is

\[
\hat{\theta}_{\text{BAYES}} = \frac{\left(\sum_{i=1}^{r} x_i^2 + x_r^2(n - r)\right) + \frac{1}{\beta}}{\alpha + r - 1}
\]

Based on the data simulation, for \( \theta \leq 1 \) it is concluded that the estimator \( \theta \) with Bayes method is better used because it gives smaller MSE values compared to the estimator \( \theta \) with maximum likelihood method. For \( \theta > 1 \), it is concluded that the estimator \( \theta \) with maximum likelihood method is better used because it provides smaller MSE values than the estimated \( \theta \) with Bayes method.

Acknowledgement
This work is supported by Hibah PITTA 2018 funded by DRPM Universitas Indonesia No.2308/UN2.R3.1/HKP.05.00/2018

References
[1] Abd Elfattah, A. M., Amal S. Hassan, dan Ziedan, D. M. (2006). Efficiency of Maximum Likelihood Estimators under Different Censored Sampling Schemes for Rayleigh Distribution. Institute of Statistical Studies & Research, Cairo University.
[2] Howlader H. A and Hossain A. (1995). On Bayesian Estimation and Prediction from Rayleigh Distribution Based on Type II Censored Data. 24(9), 2249 – 2259.
[3] Dey, S. (2009). Comparison of Bayes Estimators of the Parameter and Reliability Function for Rayleigh Distribution under Different Loss Function. Malaysian Journal of Mathematical Science, 3, 217 – 261.
[4] Collett, D. (1997). Modeling Survival Data in Medical Research. London: Chapman & Hall.
[5] Klein, J. P. and Moeschberger, M. L. (1997). Survival Analysis : Techniques for Censored and Truncated Data. Springer-Verlag, New York.
[6] Smith, J. P. (2002). Analysis of Failure and Survival Data. Chapman and Hall.
[7] Walpole, R. E. (1995). Pengantar Statistika, Edisi ketiga. Jakarta: Penerbit PT Gramedia Pustaka Utama.