

SIMULATION FOR PARAMETERS ESTIMATION OF TYPE II RIGHT CENSORED DATA FROM WEIBULL DISTRIBUTION BASED ON SEVERAL INTENSITIES

DIAN KURNIASARI*, ANDRO PRANAJAYA, WARSONO, YEFTANUS ANTONIO,
MUSTOFA USMAN, AND SUTOPO HADI

ABSTRACT. Survival analysis is aimed to suspect the probability of survival, recurrence, death, and other events until a certain time period. Type II right censored data is a combination of survival analysis involving right censored data and type II censoring. When an experiment stops until the r -th failure and all n samples are tested at the same time, then the data used is the right censored data. Censoring type II occurs when the number of failures is fixed and the observation time is random. The survival function of the Weibull distribution is used to obtain the distributed function of the type II right censored data. To estimate the parameters of the distribution function, the maximum likelihood method is used. This study will show the effect of the intensity censored (the proportion of the censored data) on parameter estimation of the type II right censored data. The simulation also considers the effect of parameter values and sample size. The result is intensity censored greatly influence the bias value of parameter estimation, but bias is not affected by sample size.

1. Introduction

As a branch of mathematics, statistics are growing very rapidly due to the discovery of various tools to solve real-life problems. Survival analysis [1] as a statistical tool aims to estimate the probability of survival, relapse, death, and other events until a certain time period. This analysis usually used in the fields of engineering, biology, medicine [2] and others. The difference between survival analysis and other statistical analysis is the existence of censored data [3]. In survival analysis, there are four types of censors, that are left censored [4], right censored, interval censored, and random censored. Types of censoring also can be divided into three types, namely type I censoring [5], type II censoring, and progressive censoring. Right censored is a type of censoring with r -th sample as the smallest observation on a random sample of size n ($1 \leq r \leq n$). In other words, if the total sample of size n , then the trial will be stopped until r failure. All n units test at the same time. In type II censoring, observations terminated after several number of failures obtained, or it can be said that number of failures is fixed and time of observation is random. On the right censor type II, number

2000 *Mathematics Subject Classification.* Primary ?????; Secondary ?????

Key words and phrases. Survival Analysis, Right Censored Data, Type II Censoring, Weibull Distribution, Maximum Likelihood Estimation.

of individual at the beginning of the study are determined and the time until the occurrence of death is determined by a certain amount. On the left censor types II, the starting point of the research done at the sequences of failure time r ($r \geq n$). In this study, survival function of Weibull distribution [6] is used (life time of data is assumed to follow Weibull distribution). Not only used for damage phenomenon with rate of damage depends on the lifetime of the component, Weibull distribution is often used to solve engineering problems, such as damage period of capacitors, transistors, photo cell conductive, corrosion metal fatigue and etc.

2. Type II Right Right-Censored Data from Weibull Distribution

Let x_1, x_2, \dots, x_n random variable that states the length of life assumed to be independent identically distribution and continuous distribution with probability density function $f(x)$ and survival function $S(x)$ [6]. In type II censored data, data consists of the smallest life time r in a random sample of n life time, that are $x_{(1)}, x_{(2)}, \dots, x_{(r)}$ where $r < n$. So, the first density function of $x_{(1)}, x_{(2)}, \dots, x_{(r)}$ [6] is defined as follows:

$$f(x_{(1)}, x_{(2)}, \dots, x_{(r)}) = \frac{n!}{(n-r)!} \prod_{i=1}^r [f(x_{(i)})] [S(x_{(r)})]^{(n-r)}. \quad (2.1)$$

A random variable X is said to have Weibull distribution with shape parameter β and scale parameter μ if and only if the probability density function of X is:

$$f(x_{(i)}) = \left(\frac{\beta}{\mu}\right) \left(\frac{x_i}{\mu}\right)^{\beta-1} \exp\left[-\left(\frac{x_i}{\mu}\right)^\beta\right] \quad (2.2)$$

and the survival function of Weibull distribution defined below:

$$S(x_{(r)}) = \exp\left[-\left(\frac{x_{(r)}}{\mu}\right)^\beta\right] \quad (2.3)$$

then probability density function for Weibull distribution of type II right censored data obtained by equation (2.1) is:

$$\begin{aligned} f(x_{(1)}, x_{(2)}, \dots, x_{(r)}) &= \frac{n!}{(n-r)!} \prod_{i=1}^r [f(x_{(i)})] [S(x_{(r)})]^{(n-r)}. \\ &= \frac{n!}{(n-r)!} \left[\left(\frac{\beta}{\mu}\right)^r \prod_{i=1}^r \left(\frac{x_i}{\mu}\right)^{\beta-1} \exp\left[-\left(\frac{x_i}{\mu}\right)^\beta\right] \right] \\ &\quad \left[\exp\left[-\left(\frac{x_{(r)}}{\mu}\right)^\beta\right] \right]^{(n-r)}. \end{aligned} \quad (2.4)$$

3. Maximum Likelihood Estimation

3.1. Estimation of Full Data. In this section, maximum likelihood method used to estimate the Weibull distribution parameters of full data [7]. Weibull

distribution density function with parameters β and μ are

$$f(x; \beta, \mu) = \left(\frac{\beta}{\mu}\right) \left(\frac{x}{\mu}\right)^{\beta-1} \exp\left[-\left(\frac{x}{\mu}\right)^\beta\right]; x > 0; \beta, \mu > 0 \quad (3.1)$$

and the likelihood function of the Weibull distribution is

$$L(x; \beta, \mu) = \prod_{i=1}^n \frac{\beta}{\mu^\beta} x_i^{(\beta-1)} e^{(-\frac{x_i}{\mu})^\beta} = \left(\frac{\beta}{\mu^\beta}\right)^n \prod_{i=1}^n x_i^{(\beta-1)} e^{(-\frac{x_i}{\mu})^\beta} \quad (3.2)$$

To maximize the likelihood function, likelihood form of the Weibull distribution as follows

$$\ln L(x; \beta, \mu) = n(\ln(\beta) - \beta \ln(\mu)) + (\beta - 1) \sum_{i=1}^n \ln(x_i) - \sum_{i=1}^n \left(\frac{x_i}{\mu}\right)^\beta \quad (3.3)$$

If the likelihood function in equation (3.2) derived the parameter μ then equated to zero, then the estimator for μ is obtained for the estimator of the Weibull distribution for the full data

$$\frac{\partial \ln L}{\partial \mu} = \frac{\partial \left(n(\ln \beta - \beta \ln \mu) + (\beta - 1) \sum_{i=1}^n \ln x_i - \sum_{i=1}^n \left(\frac{x_i}{\mu}\right)^\beta \right)}{\partial \mu} \quad (3.4)$$

$$\frac{\partial \ln L}{\partial \mu} = \frac{n\beta}{\mu} + \frac{\beta}{\mu^{\beta+1}} \sum_{i=1}^n x_i^\beta = 0 \quad (3.5)$$

$$\hat{\mu} = \left[\frac{\sum_{i=1}^n x_i^{\hat{\beta}}}{n} \right]^{\frac{1}{\hat{\beta}}} \quad (3.6)$$

by derived $\ln L(\mu, \beta)$ to the β then equated to zero, the obtained β estimate for Weibull distribution for the full data

$$\frac{\partial \ln L}{\partial \beta} = \frac{\partial \left(n(\ln \beta - \beta \ln \mu) + (\beta - 1) \sum_{i=1}^n \ln x_i - \sum_{i=1}^n \left(\frac{x_i}{\mu}\right)^\beta \right)}{\partial \beta} \quad (3.7)$$

$$\frac{n}{\hat{\beta}} - n \ln \hat{\mu} + \sum_{i=1}^n \ln x_i - \sum_{i=1}^n \left(\frac{x_i}{\hat{\mu}}\right)^{\hat{\beta}} \ln \left(\frac{x_i}{\hat{\mu}}\right) = 0 \quad (3.8)$$

$$\frac{\sum_{i=1}^n x_i^{\hat{\beta}} \ln x_i}{\sum_{i=1}^n x_i^{\hat{\beta}}} - \frac{1}{\hat{\beta}} = \frac{\sum_{i=1}^n \ln x_i}{n} \quad (3.9)$$

β estimator in (3.9) has complex form and can't be solved exactly, so it takes a certain algorithm that can be used to perform the survival analysis of Weibull distribution. In this case is done using Newton Raphson [8] iteration method.

3.2. Estimation of Type II Right-Censored Data. To estimate parameters of the Weibull distribution models for right-censored data, the maximum likelihood

method is used follows

$$L(\mu, \beta) = \frac{n!}{(n-r)!} \left[\left(\frac{\beta}{\mu} \right)^r \prod_{i=1}^r \left(\frac{x_i}{\mu} \right)^{\beta-1} \exp \left[- \left(\frac{x_i}{\mu} \right)^\beta \right] \right] \left[\exp \left[- \left(\frac{x_r}{\mu} \right)^\beta \right] \right]^{n-r} \quad (3.10)$$

To maximize the likelihood function, \ln function is used. So, Weibull likelihood form of Weibull distribution follows:

$$\begin{aligned} \ln(L(\mu, \beta)) &= \ln \left(\frac{n!}{(n-r)!} \right) \ln \left[\prod_{i=1}^r \left(\frac{\beta}{\mu} \right) \left(\frac{x_i}{\mu} \right)^{\beta-1} \exp \left[- \left(\frac{x_i}{\mu} \right)^\beta \right] \right] \\ &\quad \ln \left[\exp \left[- \left(\frac{x_r}{\mu} \right)^\beta \right] \right]^{n-r} \\ &= \ln \frac{n!}{(n-r)!} + r (\ln \beta - \ln \mu^\beta) + (\beta - 1) \sum_{i=1}^r \ln x_i - \sum_{i=1}^r \left(\frac{x_i}{\mu} \right)^\beta \\ &\quad - (n-r) \left(\frac{x_r}{\mu} \right)^\beta \end{aligned} \quad (3.11)$$

If the likelihood function (3.11) derived the parameter μ and equated to zero, then we obtain μ estimate for Weibull distribution right censored data

$$\begin{aligned} \frac{\partial \ln L(\mu, \beta)}{\partial \mu} &= \\ \frac{\partial \left(\ln \frac{n!}{(n-r)!} + r (\ln \beta - \ln \mu^\beta) + (\beta - 1) \sum_{i=1}^r \ln x_i - \sum_{i=1}^r \left(\frac{x_i}{\mu} \right)^\beta + (n-r) \left(\frac{x_r}{\mu} \right)^\beta \right)}{\partial \mu} &= \\ -r \hat{\beta} \hat{\mu}^{\hat{\beta}} + \hat{\beta} \sum_{i=1}^r x_i^{\hat{\beta}} + (n-r) \hat{\beta} x_r^{\hat{\beta}} &= 0 \end{aligned} \quad (3.12)$$

$$\hat{\mu} = \left[\frac{(n-r) x_r^{\hat{\beta}} + \sum_{i=1}^r x_i^{\hat{\beta}}}{r} \right]^{\frac{1}{\hat{\beta}}} \quad (3.13)$$

The next step is to find estimators for the parameters β , by deriving $\ln L(\mu, \beta)$ against β then equated to zero, the obtained β estimate for Weibull distribution right censored data

$$\begin{aligned} \frac{\partial \ln L(\mu, \beta)}{\partial \beta} &= \\ \frac{\partial \left(\ln \frac{n!}{(n-r)!} + r (\ln \beta - \ln \mu^\beta) + (\beta - 1) \sum_{i=1}^r \ln x_i - \sum_{i=1}^r \left(\frac{x_i}{\mu} \right)^\beta + (n-r) \left(\frac{x_r}{\mu} \right)^\beta \right)}{\partial \beta} &= \\ \frac{r}{\hat{\beta}} - r \ln \hat{\mu} + \sum_{i=1}^r \ln x_i - \sum_{i=1}^r \left(\frac{x_i}{\hat{\mu}} \right)^{\hat{\beta}} \ln \left(\frac{x_i}{\hat{\mu}} \right) + (n-r) \left(\frac{x_r}{\hat{\mu}} \right)^{\hat{\beta}} \ln \frac{x_r}{\hat{\mu}} &= 0 \end{aligned} \quad (3.14)$$

$$\begin{aligned}
\frac{r}{\widehat{\beta}} = & r \ln \left[\frac{(n-r) x_r^{\widehat{\beta}} + \sum_{i=1}^r x_i^{\widehat{\beta}}}{r} \right]^{\frac{1}{\widehat{\beta}}} \\
& + \sum_{i=1}^r \ln x_i - \sum_{i=1}^r \left[\frac{x_i^{\widehat{\beta}}}{(n-r) x_r^{\widehat{\beta}} + \sum_{i=1}^r x_i^{\widehat{\beta}}} \right] \ln \left(\frac{x_i}{\left[\frac{(n-r) x_r^{\widehat{\beta}} + \sum_{i=1}^r x_i^{\widehat{\beta}}}{r} \right]^{\frac{1}{\widehat{\beta}}}} \right) \\
& + (n-r) \left(\frac{x_r^{\beta}}{\left[\frac{(n-r) x_r^{\widehat{\beta}} + \sum_{i=1}^r x_i^{\widehat{\beta}}}{r} \right]} \right) \ln \left(\frac{x_r}{\left[\frac{(n-r) x_r^{\widehat{\beta}} + \sum_{i=1}^r x_i^{\widehat{\beta}}}{r} \right]^{\frac{1}{\widehat{\beta}}}} \right)
\end{aligned} \tag{3.15}$$

the estimators for β for the model (3.16) can't solved exactly, so it requires numerical algorithm that can be used to perform the survival analysis of Weibull distribution. Newton Raphson algorithm [8] selected as an iteration method.

$$\widehat{\beta}_{n+1} = \widehat{\beta}_n - \frac{f(\widehat{\beta}_0)}{f'(\widehat{\beta}_0)} \tag{3.16}$$

4. Result For Nurmerical Simulation Process

This chapter will discuss the results of simulations that have been carried out. Simulation is done by generating data with different sample sizes, the values of the different parameters and different proportions censored [9]. In this study, it will be seen how the parameter estimators of full data on Weibull distribution and parameter estimators for right censored data of Weibull distribution. Parameter estimation performed using Maximum Likelihood method [6].

Numerical results in this study were obtained from iteration process using R software that used to estimate the value of the parameter β on the parameter estimators for the full data and parameter estimator for right censored data.

Determination of the value of initial estimate in μ and β parameters in this study are resembling to Guure [5]. That study started by generating Weibull distribution data with different sample sizes, this study used a sample size is $n = 100$, $n = 200$, $n = 400$ are repeated 1000 times for each n . While the parameter values used for μ is 1 and 0.5, while for β is 1 and 1.5. And the proportion censored is 1%, 5%, 10%, 15%. Simulation results with different sample sizes, the values of the different parameters and different proportions of censored will be presented in Table 1.

This result done by performing iterative process to estimate the parameter β (beta) to get estimated value of μ with $n=100$, $n=200$ and $n=400$ are repeated 1000 times for each n . The estimated values of β and the estimated value of μ for full data, as well as the estimated value for the parameter β and the estimated value for the parameter μ for censored data with censored intensity (1%, 5%, 10%, and 15%).

TABLE 1. Estimation value and bias on parameters estimation of full and type II right censored data from weibull distribution with 1000 repetitions.

n	Censored Intensity	μ	β	Type II Right Censored		Bias of Estimator	
				$\hat{\mu}$	$\hat{\beta}$	μ	β
100	0% (full data)	1	1	1.001023	1.01173	0.001023	0.01173
	1%			0.937112	1.152813	0.062888	0.152813
	5%			0.849441	1.518728	0.150559	0.518728
	10%			0.79798	1.76773	0.20202	0.76773
	15%			0.752705	1.819821	0.247295	0.819821
	0% (full data)	0.5	1.5	0.500232	1.521579	0.000232	0.021579
	1%			0.191433	1.741347	0.308567	0.241347
	5%			0.110567	2.105245	0.389433	0.605245
	10%			0.086929	2.6662	0.413071	1.1662
	15%			0.080102	2.732074	0.419898	1.232074
200	0% (full data)	1	1	0.996243	1.004957	0.003757	0.004957
	1%			0.92746	1.156748	0.07254	0.156748
	5%			0.832284	1.517197	0.167716	0.517197
	10%			0.768674	1.742913	0.231326	0.742913
	15%			0.712449	1.766025	0.287551	0.766025
	0% (full data)	0.5	1.5	0.50188	1.51415	0.00188	0.01415
	1%			0.187923	1.745869	0.312077	0.245869
	5%			0.114978	2.305152	0.385022	0.805152
	10%			0.083418	2.661998	0.416582	1.161998
	15%			0.075901	2.710141	0.424099	1.210141
400	0% (full data)	1	1	0.998652	1.0029	0.001348	0.0029
	1%			0.925841	1.161956	0.074159	0.161956
	5%			0.829013	1.524025	0.170987	0.524025
	10%			0.76099	1.748575	0.23901	0.748575
	15%			0.703095	1.76185	0.296905	0.76185
	0% (full data)	0.5	1.5	0.50025	1.50107	0.00025	0.00107
	1%			0.18743	1.737327	0.31257	0.237327
	5%			0.114973	2.285217	0.385027	0.785217
	10%			0.083567	2.626875	0.416433	1.126875
	15%			0.076754	2.644991	0.423246	1.144991

Simulation results, value of μ and β estimators of Weibull distribution on the full data get near to the initial value of the parameter μ and β , but not for the value of the parameter μ and β censored data. For example, it can be seen in Table 1 to the intensity of censored data at 1%, 5%, 10%, 15% for $n = 100$, $n = 200$, $n = 400$, the initial value of the parameter $\mu = 1$ and $\beta = 1$, the value produced by the estimator μ is always less than 1 and result value of the difference estimator will be smaller and be smaller as the magnitude of the intensity of censored, while the

β value for the resulting estimator is always more than one and will always be in line with the magnitude of the intensity censored, and if we look again for any n with beginning value of the parameter $\mu = 1$ and $\beta = 1$ with the intensity of the sensor by 1%, value of μ and β estimators produced is not very different even with no significant increase in the n number.

To find out how influential the amount of data (n) of the resulting bias in the estimation of parameters μ and β can be done by adding the bias generated each censored at each intensity parameter is then divided by the number of intensity censored. For example, to calculate the average bias generated by the parameters μ and β with $n = 100$ and the initial value of $\mu = 1$ and $\beta = 1$ is

$$\bar{\mu} = \frac{\sum_{i=1}^5 \hat{\mu}}{5} = \frac{0.663785}{5} = 0.132757,$$

$$\bar{\beta} = \frac{\sum_{i=1}^5 \hat{\beta}}{5} = \frac{2.270822}{5} = 0.454164.$$

Average bias generated by the parameters μ and β with $n = 200$ and the initial value of $\mu = 1$ and $\beta = 1$ is

$$\bar{\mu} = \frac{\sum_{i=1}^5 \hat{\mu}}{5} = \frac{0.762890}{5} = 0.152578$$

$\bar{\beta} = \frac{\sum_{i=1}^5 \hat{\beta}}{5} = \frac{2.187840}{5} = 0.437568$. Average bias generated by the parameters μ and β with $n = 400$ and the initial value of $\mu = 1$ and $\beta = 1$ is

$$\bar{\mu} = \frac{\sum_{i=1}^5 \hat{\mu}}{5} = \frac{0.782409}{5} = 0.156481,$$

$$\bar{\beta} = \frac{\sum_{i=1}^5 \hat{\beta}}{5} = \frac{2.199306}{5} = 0.439861.$$

From the above calculation, it can be concluded that the bias produced by the censored data parameters μ and β are not affected by the number of data n due to the bias calculation is done, because the more the number of data n then the resulting bias parameter estimators μ and β is not significantly different.

5. Graph Of Weibull Distribution (Full Data)

The following graphs are based on density function of opportunities Full Weibull distribution using R Software. By following a pre-determined parameters that has been previously reported by Guure and Ibrahim [5].

Figure 1 are Weibull distribution graph in various parameter. Every graph is representation of different cases. In Figure A-1, shape parameter β has two different value, (i.e. $\beta_1 = 1$ and $\beta_2 = 1.5$), and scale parameter μ is fixed ($\mu_1 = \mu_2 = \mu = 1$). From this graph, the shape of distribution function is different.

Next case is in Figure A-2. In this case, shape parameter is same as before, but scale parameter $\mu = 0.5$. Shape of distribution function in this graph is same as before, but the smaller parameter values make the graph more pointed. Figure B-1. Illustrates the shape of the distribution function with different location parameters, then one is more pointed than the other. Finally, Figure B-2 have two similar shape (i.e. shape parameter is fixed $\beta = 1$).

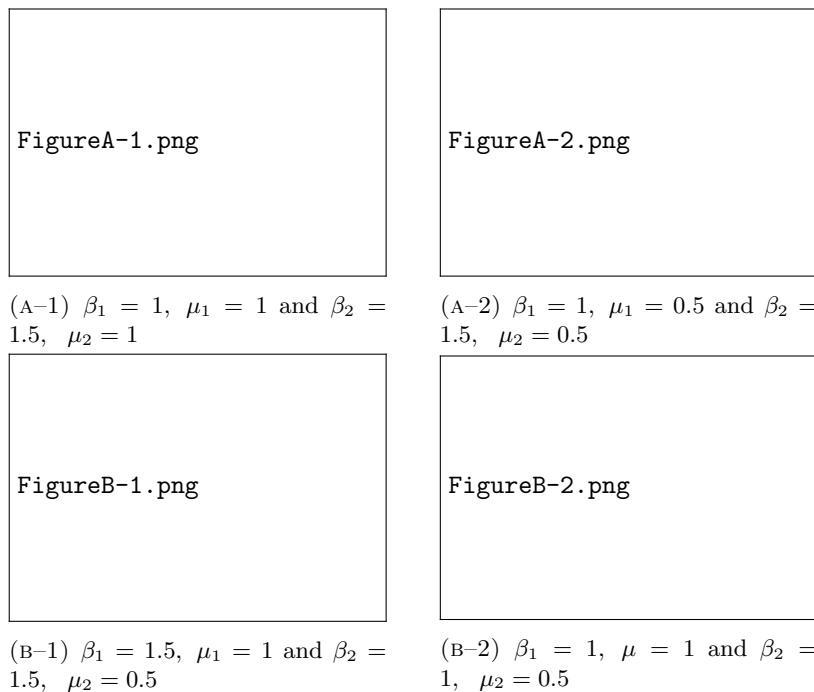


FIGURE 1. Weibull distribution graph.

6. Conclusion

The censored intensity is greatly influence the value of the parameter estimators $\hat{\mu}$ and $\hat{\beta}$, this can be seen in Table 4.1 due to the greater intensity of censored, then the value of estimator $\hat{\mu}$ will decreasing if the value of the censored intensity is increasing, as well as estimator for $\hat{\beta}$ will be increasing if the value of censored intensity is increasing. Bias produced by the estimators μ and β censored data will be worth more in line with the magnitude of the intensity censored.

Bias produced by the censored data parameters μ and β are not affected by the number of data n due to the bias calculation is done, because the more the number of data n then the resulting bias parameter estimators μ and β are not significantly different.

References

1. Lee, E. T., Wang, J. W.: *Statistical Methods for Survival Data Analysis*, John Wiley & Sons Inc, Oklahoma, 2003.
2. Collett, D.: *Modelling Survival Data in Medical Research*, Chapman & Hall, London, 2003.
3. Engelhardt, M., Bain, L. J., Blumenthal, S.: Statistical Analysis of a Weibull Process with Left-Censored Data, in (eds) *Survival Analysis: State of the Art. Nato Science (Series E: Applied Sciences)* **211** (1992) 173–195.
4. Lawless, J. F.: *Statistical Models and Methods for Lifetime Data*, John Wiley & Sons Inc, New York, 2003.

5. Guure, B.C. and Ibrahim , A. N: Methods for Estimating the 2-Parameter Weibull Distribution with Type-I Censored Data, *Res. J. Appl. Sci. Eng. Technol* **5**(3) (2013) 689-694.
6. Hogg, R. V., Craig, A. T.: *Introduction to Mathematical Statistics*, Pearson, Buston, 2013.
7. Al-Kanani , H. I., Shaima, A. J.: Estimate Survival Function for The Brain Cancer Disease by Using Three Parameters Weibull Distribution, *J. Basrah Res. (Sci.)* **37**(1817–2695) (2011) 109–121.
8. Atkinson, K. E.: *An Introduction to Numerical Analysis*, John Wiley & Sons Inc, New York, 1989.
9. Yan, J.: Survival Analysis: Techniques for Censored and Truncated Data, *J. Am. Stat. Assoc.* **99**(467) (2004) 900–901.

DIAN KURNIASARI: DEPARTMENT OF MATHEMATICS, FACULTY OF MATHEMATICS AND NATURAL SCIENCE, UNIVERSITAS LAMPUNG, INDONESIA
E-mail address: dian.kurniasari@fmipa.unila.ac.id

ANDRO PRANAJAYA: DEPARTMENT OF MATHEMATICS, FACULTY OF MATHEMATICS AND NATURAL SCIENCE, UNIVERSITAS LAMPUNG, INDONESIA
E-mail address: andro.pranajayaramadhan@gmail.com

WARSONO: DEPARTMENT OF MATHEMATICS, FACULTY OF MATHEMATICS AND NATURAL SCIENCE, UNIVERSITAS LAMPUNG, INDONESIA
E-mail address: warsono.1963@fmipa.unila.ac.id

YEFTANUS ANTONIO: DEPARTMENT OF MATHEMATICS, FACULTY OF MATHEMATICS AND NATURAL SCIENCE, UNIVERSITAS LAMPUNG, INDONESIA
E-mail address: yeftanus@gmail.com

MUSTOFA USMAN: DEPARTMENT OF MATHEMATICS, FACULTY OF MATHEMATICS AND NATURAL SCIENCE, UNIVERSITAS LAMPUNG, INDONESIA
E-mail address: mustofa.usman@fmipa.unila.ac.id

SUTOPO HADI: DEPARTMENT OF CHEMISTRY, FACULTY OF MATHEMATICS AND NATURAL SCIENCE, UNIVERSITAS LAMPUNG, INDONESIA
E-mail address: sutopo.hadi@fmipa.unila.ac.id