ESTIMATION PARAMETER OF SUSCEPTIBLE - INFECTED - SUSCEPTIBLE (SIS) MODEL WITH MAXIMUM LIKELIHOOD METHOD USING STEEPEST DESCENT APPROACH TO SPREAD OF TYPHOID DISEASE.

Aurelia Farrah Syafira[1], Respatiwulan[2], and Sri Subanti[3]
[1,2,3]Statistics, Sebelas Maret University
Ir. Sutami 36 Kentingan Street, Jebres, Surakarta, Central Java.
[1]aureliafara09@student.uns.ac.id , [2]respatiwulan@staff.uns.ac.id ,
[3]srisubanti@student.uns.ac.id

ABSTRACT

Typhoid is a disease that can be spread through physical contact. Individuals who contract the disease can be re-infected depending on their immune system. This research uses the Susceptible Infected Susceptible (SIS) model to study the spread of typhoid disease with infection data that follows the Poisson distribution, so the infection rate is assumed to follow the Poisson distribution (λ_t). The parameter (λ_t) is estimated with the Maximum Likelihood Estimator (MLE). Parameter estimation with MLE did not give exact results, so a numerical approach with the steepest descent method was used. After determining the parameter estimates through the steepest descent approach, the parameters are used to simulate the typhus disease spread model. The result of this research is the parameter estimation of SIS model with infection rate assumed to be Poisson distribution (λ_t). Parameter estimates are used to predict infection rates based on historical data to understand how past changes affect the current condition of typhoid spread through data plotting.

Keywords: Typhoid Disease, SIS Model, Poisson Distribution, MLE, Steepest Descent.

INTRODUCTION

Typhoid is a disease caused by the bacterium Salmonella typhi. According to Widoyono (2009) [10], Typhoid is a bacterial infection of the small intestine and occasionally occurs in the bloodstream, caused by the bacteria Salmonella typhi A, B, and C. Typhoid can also cause gastroenteritis (stomach inflammation) as a result of bacterial infection.

Typhoid is transmitted through direct contact from an infected individual to another. Typhoid is generally transmitted through food and drink contaminated with the bacteria and from infected individuals through their feces. The bacteria is usually in water contaminated with feces and sticks to the food or drink consumed. The mathematical models that are applied to recognize the spread of an infectious disease in a particular place are known as epidemic models. According to Darmawati and Nur (2019) [4], some researchers have used the Susceptible - Infected - Susceptible (SIS) epidemic model to model the spread of an infectious disease.

The SIS model is a disease spread model that has the characteristic that everyone is susceptible to certain diseases and recovery through medical treatment or natural processes. It does not guarantee that individuals are immune to the infection, meaning that individuals can be re-infected with the same disease at any time. Therefore, this SIS model can be used for mathematical modeling of the spread of typhoid disease.

Several studies related to infectious diseases have been conducted, ranging from researchers in the health sector to researchers in the field of mathematics. The mathematic model research conducted by Darmawati and Nur (2019) [4] use the SIS model to examine the malaria infection model, where the disease infection rate follows the Poisson distribution (λ_t). The characteristics of malaria are similar to typhoid disease which can be contagious, but individuals who have recovered do not have immunity to the disease. Therefore, this research discusses the spread of typhoid disease which is assumed to follow the Poisson distribution (λ_t).

The Poisson distribution can describe the rate of typhoid infection in a certain time interval. Therefore, the SIS model with disease infection rate can be assumed to follow the Poisson distribution (λ_t). According to Bolstad (1998) [2], The data follows a Poisson distribution with a disease infection rate requires parameter estimation using the Maximum Likelihood Estimator (MLE).

The Maximum Likelihood Estimator method is a maximizes value of the Likelihood function (Nurlaila, 2013) [7]. In some cases the Maximum Likelihood Estimator did not give an exact result, so a numerical approach is needed. One of the simplest numerical methods to maximize the function is the steepest descent method, which was first introduced by Cauchy in 1847 (Bazaraa, et al. 2005) [1]. Darmawati et al. (2019) [5] research, has the problem that the MLE method did not give an exact, because it will be approch by the steepest descent method. According to previous researchs, parameter estimation of SIS model with typhoid infection rate assumed to follow the Poisson distribution (λ_t) using Maximum Likelihood Estimator method with steepsest descent approach simulation.

METHOD

The model used in this study is the estimation of model parameters and is applied to the spread of typhus disease. Based on research conducted by Darmawati et al. (2019) [5], the following steps were taken to observe the objectives in this reserach.

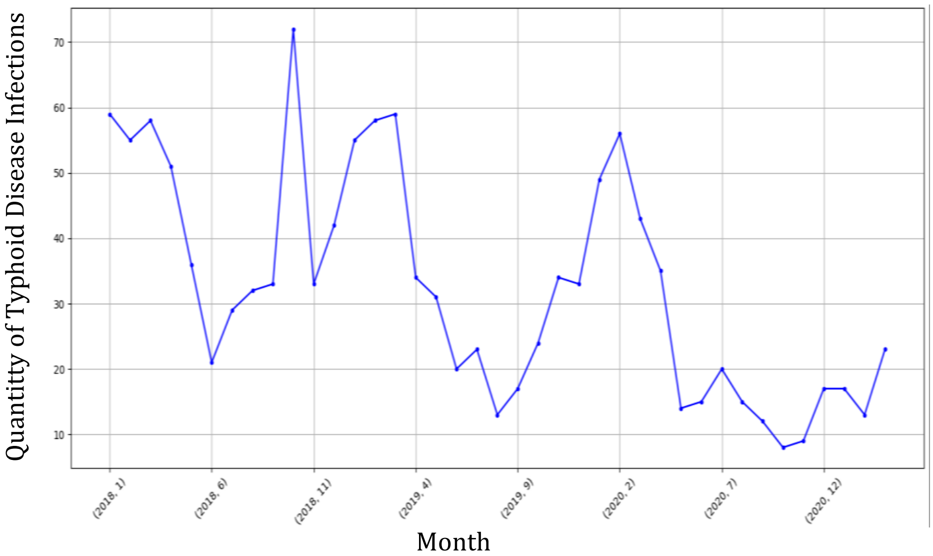
1. Collect and analyzing data
   The data used consists of monthly data on the number of typhoid cases from 2018 - 2020. The number of observation units consists of 36 months.
2. Construct the SIS model as follows:
   a. Determine the assumptions and parameters required in the SIS model of typhoid disease.
   b. Assume the infection rate is Poisson distributed (λ_t).
   c. Construct SIS model with infection rate in Poisson-distributed population (λ_t).
3. Determine the parameter estimation λ_t using the Maximum Likelihood Estimator as follows:
   a. Determine the likelihood function.
   b. Determine the log likelihood function (ln L(λ_t)).
4. Perform a numerical steepest descent simulation to obtain the following likelihood function parameter estimation values:
   a. Determine the initial value and ε (constant) to determine the tolerable error.
   b. Determine the gradient of the function at λ_1.
   c. Qualify the iteration, provided that $\|\partial Z(\lambda_{k+1})\| < \varepsilon$ the iteration can stop.
   d. Find λ_k by locating the extremum of $F(X_k + \lambda_k d_k)$, through differentiating the function $F(X_k + \lambda_k d_k)$ and equating it to zero, with the search direction $d_k = -\partial F(X_k)$.
   e. The value of λ_k is determine using the formula $\lambda_{k+1} = \lambda_k + \lambda_{k-1} d_k$.
   f. Simulate of steepest descent method.
   g. Interpretation of simulation results on the disease of interest.

FINDINGS AND DISCUSSION

1. Collect and Analyzing Data

The data analysis is conducted to determine the suitability of the distribution from the data. This research uses typhoid disease data from 2018-2020 from UNS Hospital. The data consists of the quantity of individuals infected each month. Below is the plot of typhoid disease data from 2018-2020 presented in Picture 1.

Figure 1:
Data Plot of Typhoid Disease in 2018 – 2020



Based on the figure 1, there are significant fluctuations each year with some prominent peaks, especially in the middle of the year. The data plot shows that a decrease in the number of infections occurred after the highest peak followed by a gradual increase before reaching the next peak. Based on the data, it is assumed that the infection rate follows a Poisson distribution λ_t. The SIS model that has been adjusted to the assumptions in Equation 1 is then constructed into a new SIS model.

2. Construct The SIS Model

The SIS models that consider the transmission rate and recovery rate and assume the death rate is equal to the birth rate, so the total population is constant and the recovered patient returns to being susceptible. The SIS model can be expressed by Equation 1 (Brauer et al, 1945) [3].

$$\begin{aligned} \frac{dS}{dt} &= -\frac{\eta}{N}SI + \gamma I \\ \frac{dI}{dt} &= \frac{\eta}{N}SI - \gamma I \end{aligned} \tag{1}$$

The SIS model has the following assumptions:

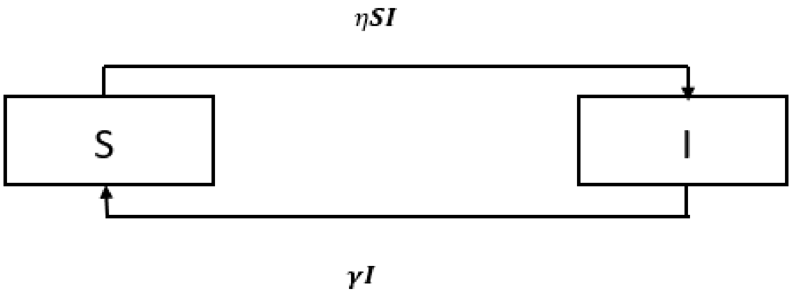
1. The spread of the disease occurs in a constant and closed population, so there are no individuals entering or leaving the population.
2. The population is homogeneous, so that each individual has the same characteristics and opportunities to contract the disease.
3. There is only one disease in the population.

4. Birth rate and death rate are ignored.
5. Recovered individuals can be re-infected and enter the susceptible population again.

Each individual is healthy, but vulnerable to disease. According to Hethcote (2000) [6], the SIS model of a population is categorized into two groups: the infected category (I) and the susceptible category (S). The infected category (I) comprises individuals who are infected with the disease, while the susceptible category (S) comprises healthy individuals who are at risk of contracting the disease.

Disease spread in the SIS model occurs because individuals from the susceptible (S) group move to the infected (I) group, because they are infected with typhoid disease due to physical contact. This movement is called the rate of infection in the population (ηSI). Individuals in category I recover from the disease with a recovery rate of γ. Furthermore, the number of individuals in category I who recover from the disease is γI and re-enter the S category, because individuals who have recovered do not have permanent immunity to the disease. The SIS model disease transmission scheme is shown in Figure 2.

Figure 2:
SIS Model Scheme



The values used for the rate of infection in the population (ηSI) parameter are approximate. In reality, the value of typhoid cases is derived from data on the number of cases that have occurred, with the current number of cases depending on the number of cases the previous time. Therefore, it is assumed that the population infection rate (ηSI) of typhoid data follows a Poisson distribution with Y_t being the number of new typhoid infections that occur each month. To replace the infection rate in the population (ηSI) in Equation 1, the infection rate in the population of Poisson-distributed data can be written as λ_t, wich is:

$$\lambda_t = \mu + \alpha\lambda_{t-1} \tag{2}$$

The parameter λ_t is the infection rate in the population at time t that follows a Poisson distribution, μ is the average value of λ_t and α is a coefficient that shows the effect of the previous infection rate λ_{t-1}on the current infection rate λ_t. This model states that the value of λ_t at time t is influenced by the previous value λ_{t-1}, by adding a constant μ and a factor α.

After determining the infection rate model in a Poisson distributed population (λ_t), the next step is to estimate the parameters in λ_t using Maximum Likelihood Estimator (MLE). The parameters are μ and α, with the following algorithm:

1. Initialize the initial values $S(0), I(0)$;
2. Input the initial values of N, η, γ, μ, and α arbitrarily;
3. Model construction;

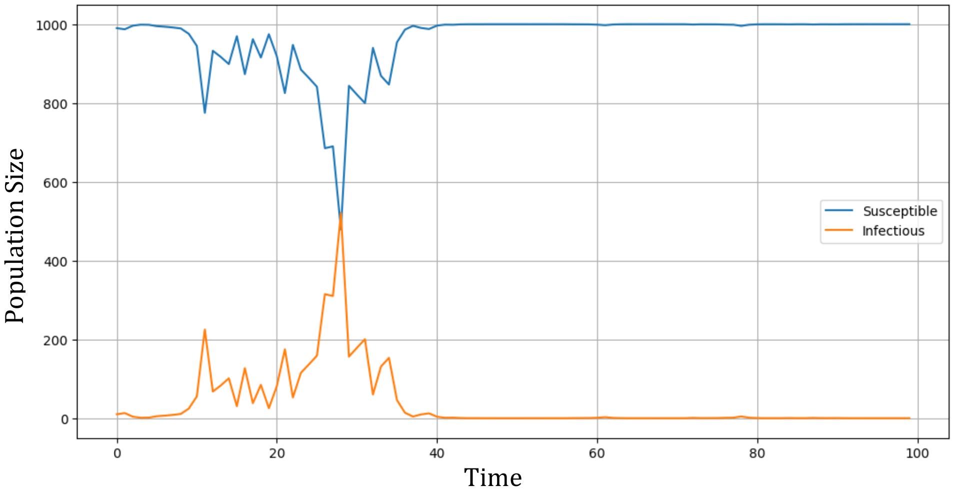
Based on the above algorithm, the parameter values used are available in table 1.

Table 1:
Parameters of Typhoid Disease SIS Model

Parameters	**Description**	**Value**	**Units**
N	total number of individuals in a population	1000	human
γ	Cure rate	0.7	per day
μ	Average of λ_t	0.9	per day
α	Coefficient describing the impact of the value of λ_{t-1} on λ_t	0.1	per day
$S(0)$	Initial number of susceptible humans	990	human
$I(0)$	Initial number of infected humans	10	human

The following is the pattern of typhoid disease spread using the above algorithm and parameter values. The results of the above algorithm will be shown in Figure 3.

Figure 3 :
The SIS Model Simulation



The peak of infections occurs approximately at time 20–25, with the number of infected individuals reaching more than 400. During the increase in infections, the number of susceptible individuals experiences a sharp decline. After the number of infected individuals reaches its peak, it begins to decline commencing approximately at time 25. During this decline the recovered individuals become susceptible again, causing the number of susceptible individuals starts to increase again. After time 39, the number of infected individuals decreases drastically and approaches zero, indicating that the outbreak has subsided. The number of susceptible individuals stabilizes at approximately 990, almost the same as the initial number, with only minor fluctuations in the number of infections. This indicates that stability has been reached, with very few individuals remaining infected and the majority of the population becoming susceptible again. After looking at Picture 3 with the infection rate in a Poisson distributed population, we can construct the SIS model as follows:

$$\begin{aligned} \frac{dS}{dt} &= -\lambda_t + \gamma I \\ \frac{dI}{dt} &= \lambda_t - \gamma I \end{aligned} \tag{3}$$

with infection rate in a Poisson distributed data population (λ_t), wich is $\lambda_t = \mu + \alpha\lambda_{t-1}$.

3. Parameter Estimation λ_t Using The Maximum Likelihood Estimator

The estimation of parameters in λ_t, that is μ and α, uses the Maximum Likelihood Estimator method. The Poisson distribution with parameter λ and infection data $y_1, y_2, y_3, \dots, y_t$, the probability density function is as follows (Sugito and Mukid, 2011) [8]:

$$f(y,\lambda) = \frac{{\lambda_t}^{y_t} e^{-\lambda_t}}{y_t!} \tag{4}$$

Based on Equation 4, the Likelihood function for Poisson distribution is as follows (Nurlaila et al., 2013) [7]:

$$\begin{aligned} L(y,\lambda) &= \frac{e^{-\lambda_1}{\lambda_1}^{y_1}}{y_1!} \times \frac{e^{-\lambda_2}{\lambda_2}^{y_2}}{y_2!} \times \frac{e^{-\lambda_3}{\lambda_3}^{y_3}}{y_3!} \times \dots \times \frac{e^{-\lambda_t}{\lambda_t}^{y_t}}{y_t!} \\ &= \prod_{t=2}^{T} \frac{e^{-\lambda_t}{\lambda_t}^{y_t}}{y_t!} \end{aligned} \tag{5}$$

Accordingly, the log-likelihood function is

$$\begin{aligned} \mathrm{Ln}(L(y,\lambda)) &= \sum_{t=2}^{T} \log \frac{e^{-\lambda_t}{\lambda_t}^{y_t}}{y_t!} \\ &= \sum_{t=2}^{T} -\lambda_t + \sum_{t=2}^{T} y_t \log \lambda_t - \sum_{t=2}^{T} \log(\, y_t!) \end{aligned} \tag{6}$$

Then, substitute $\lambda_t = \mu + \alpha\lambda_{t-1}$ into Equation 6.

$$F(\lambda) = \sum_{t=2}^{T} -(\mu + \alpha\lambda_{t-1}) + \sum_{t=2}^{T} y_t \log \mu + \alpha\lambda_{t-1} - \sum_{t=2}^{T} \log(\, y_t!) \tag{7}$$

Equation 7 is the log-likelihood function. Because parameter estimation using MLE does not produce an exact result, it will be assisted by the steepest descent method to obtain the parameter value. Determining the parameter estimation value with the steepest descent method approach refers to the algorithm in Utomo's (2016) [9] research, the algorithm is as follows:

1) Take $\lambda_1 = \{\mu_1, \alpha_1\} \in R^2$ as the initial parameter value and error value $\varepsilon = 1e^{-10}$.
2) Determine $\partial F(\lambda_1) = \partial F(\mu_1, \alpha_1)$with,

$$\partial F(\mu,\alpha) = \left[\frac{\partial F(\mu,\alpha)}{\partial \mu_1}, \frac{\partial F(\mu,\alpha)}{\partial \mu_1}\right]$$

$\partial F(\lambda_1)$ is the gradient of function F at λ_1, then determine it for $\partial F(\lambda_1)$ and do it for $\partial F(\lambda_k)$). The calculation of the gradient of the log-likelihood function, is as follows:

i. Partial derivative of μ:

$$\frac{\partial F(\mu,\alpha)}{\partial \mu_1} = \sum_{t=2}^{T} -(\mu + \alpha\lambda_{t-1}) + \sum_{t=2}^{T} y_t \log \mu + \alpha\lambda_{t-1}$$

$$-\sum_{t=2}^{T} \log(y_t!)$$

$$\frac{\partial F(\mu,\alpha)}{\partial \mu_1} = -1 + y_t \cdot \frac{1}{\mu + \alpha\lambda_{t-1}}$$

Therefore, the partial derivative of $F(\mu,\alpha)$ with respect to μ is:

$$\frac{\partial F(\mu,\alpha)}{\partial \mu_1} = \sum_{t=2}^{T} \left(-1 + \frac{y_t}{\mu + \alpha\lambda_{t-1}}\right)$$

ii. Partial derivative of α:

$$\frac{\partial F(\mu,\alpha)}{\partial \alpha_1} = \sum_{t=2}^{T} -(\mu + \alpha\lambda_{t-1}) + \sum_{t=2}^{T} y_t \log \mu + \alpha\lambda_{t-1}$$

$$-\sum_{t=2}^{T} \log(y_t!)$$

$$\frac{\partial F(\mu,\alpha)}{\partial \alpha_1} = -\lambda_{t-1} + y_t \cdot \frac{\lambda_{t-1}}{\mu + \alpha\lambda_{t-1}}$$

Terefore, the partial derivative of $F(\mu,\alpha)$ with respect to α is:

$$\frac{\partial F(\mu,\alpha)}{\partial \alpha_1} = \sum_{t=2}^{T} \left(-\lambda_{t-1} + \frac{y_t\lambda_{t-1}}{\mu + \alpha\lambda_{t-1}}\right)$$

3) If $\|\partial F(\lambda_k)\| = \sqrt{\frac{\partial F}{\partial \mu_1}^2 + \frac{\partial F}{\partial \alpha_1}^2} < \varepsilon$, then iteration stops, otherwise iteration continues.
4) Find $\lambda_k$by finding the extreme point of $F(\lambda_k + \pi_k d_k)$, by derivating the function $F(\lambda_k + \lambda_k d_k))$ and equating to zero with the search direction $d_k = -\partial F(\lambda_k)$.
5) The value of λ_k is determined using the formula$\lambda_{k+1} = \lambda_k + \pi_k d_k$.

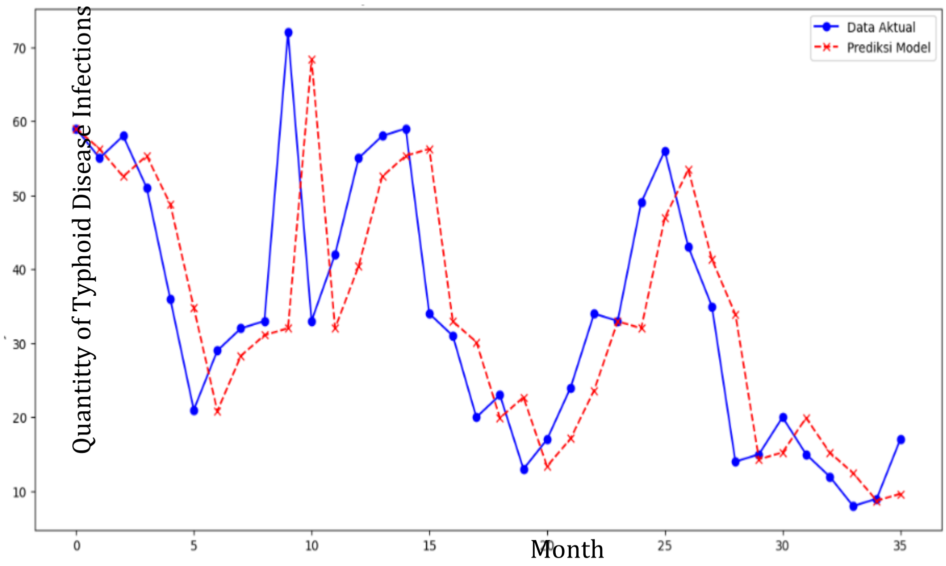
Based on the above algorithm, a numerical simulation is carried out with parameter values referring to the research of Darmawati et al. (2019) [5] to obtain the likelihood function value using Python software.

Table 2:
Numerical Simulation of Steepest Descent

Simulations	**Initial Parameters**	**Parameter Estimation**	**Log-Likelihood Function Value**
First	$\mu = 0.8$ $\alpha = 0.01$	$\mu = 1.1461$ $\alpha = 0.9350$	-200.2988
Second	$\mu = 0.8$ $\alpha = 0.02$	$\mu = 1.1658$ $\alpha = 0.9345$	-200.4041
Third	$\mu = 0.9$ $\alpha = 0.01$	$\mu = 1.2084$ $\alpha = 0.9333$	-200.6346
Fourth	$\mu = 0.9$ $\alpha = 0.2$	$\mu = 1.1818$ $\alpha = 0.9340$	-200.4911
Fifth	$\mu = 0.9$ $\alpha = 0.02$	$\mu = 1.2238$ $\alpha = 0.9329$	-200.7181
Sixth	$\mu = 0.9$ $\alpha = 0.3$	$\mu = 1.1415$ $\alpha = 0.9351$	-200.2737

Based on the results from Table 2, the fifth simulation is the simulation that produces the smallest log likelihood value. Therefore, the initial parameter value $\mu = 0.9$, $\alpha = 0.02$ produces a parameter estimate of $\mu = 1.2238, \alpha = 0.9329$ with and log likelihood function value of -200.7181. Furthermore, these parameters are used for model simulation as shown in Figure 4.

Figure 4:
Simulation Results of Steepest Descent Parameter Estimation



Based on Picture 4, it can be observed that the quantitiy of thypoid disease infections the actual data began to decline from approximately 60 cases to about 20 cases in the sixth month, then around month 10 experienced a sharp spike of about 70 cases, but the estimated did not reach the same high peak. This suggests that the infection rate model using likelihood estimation is less responsive to sudden or extreme changes in the infection data. After experiencing a peak in month 10, the number of infections dropped drastically to about 30 cases in month 15 and continued to decline until about month 19 with around 13 cases. The estimate followed the downward trend well, but did not match the exact low values of the actual data. After the decline, about month 21 quantitiy of thypoid disease infections gradually increased, reaching a peak about month 25 with 59 infections. Then it declined again until about month 28, the estimate captured this increasing and decreasing trend well, although there were small differences that prevented the estimate not exactly match the actual data values. After the decline, the quantity of infections remained relatively low with little fluctuation until month 35. The estimate followed the trend observed in the actual data, although there were small differences. This is due to the tolerance error (ε) value when estimating parameters of $1e^{-10}$ or equal to 0.0000000001.

CONCLUSION

Based on the results and discussions that have been carried out in this study, it can be concluded that:

1. The SIS epidemic model with the infection rate in the population is Poisson distributed (λ_t) as in Equation 3.
2. Parameter estimation log - likelihood ($F(\lambda)$) using steepest descent with initial parameter value $\mu = 0.9, \alpha = 0.02$ produces parameter estimates of $\mu = 1.2238, \alpha = 0.9329$and log likelihood function value -200.7181.
3. The numerical simulation results of the infection rate model using likelihood estimation are good at capturing the general patterns and basic trends in the monthly typhoid infection data. Model estimation are often close to actual values, especially in periods that do not have sudden or extreme changes.
4. The model does not capture sharp increases and decreases accurately, such as in month 10 and around month 20. The model estimation provide smoother values and do not always reflect the sharp fluctuations seen in the actual data.

REFERENCES

[1] Bazaraa, M., Sherali, H., and Shetty, C. (2005). Nonlinear Programming: Theory and Algorithms. New Jersey: A John Wiley & Sons, Inc. Published.

[2] Bolstad, B. (1998). Comparing Some Iterative Methods of Parameter Estimation for Censored Gamma Data. Thesis. The University of Wakaito.

[3] Brauer, F., Driessche, P. v., and Wu, J. (1945). Lecture Notes in Mathematics : Mathematical Epidemiology. Springer.

[4] Darmawati, dan Nur, W. (2019). Model SIS Stokastik pada Penyakit Malaria Berdasarkan Data Pasien. *SAINTIFIK; 5(1)*, 53 - 57.

[5] Darmawati, Nur, W., dan Musafira. (2019). Estimasi Parameter Model SIS Stokastik Penyebaran Penyakit Malaria dengan Metode Steepest Descent. *Jurnal Matematika, Sains, dan Pembelajarannya*, 145 - 149.

[6] Hethcote, H. W. (2000). The Mathematics of Infectious Disease. *SIAM Review 42*, 599 - 653.

[7] Nurlaila, D., Kusnandar, D., dan Sulistianingsih, E. (2013). Perbandingan Metode Maximum Likelihood Estimation (MLE) dan Metode Bayes dalam Pendugaan Parameter Distribusi Eksponensial. *Buletin Ilmiah Mat. Stat. dan Terapannya (Bimaster)*, 51 - 56.

[8] Sugito, dan Mukid, M. A. (2011). Distribusi Poisson dan DIstribusi Eksponensial dalam Proses Stokastik. *Media Statistika*, *Vol. 4*, *No.2*, 113 -120.

[9] Utomo, R. (2016). Metode Numerik Stepest Descent Terinduksi Newton Pemecahan Masalah Optimasi Tanpa Kendala. *Mosharafa; 5(3)*, 187 - 194.

[10] Widoyono.(2009). Penyakit Tropis Epidemiologi, Penularan, Pencegahan, dan Pemberantasannya. Erlangga.