Monthly rainfall distributions in Pematangsiantar

P Ismawati and E Rosmaini*
Department of Mathematics, Faculty of Mathematics and Natural Sciences, Universitas Sumatera Utara, Medan, Indonesia

*E-mail: ellyl@usu.ac.id

Abstract. Monthly rainfall data for three rainfall stations, Marihat, Bangun and Bah Jambi in Pematangsiantar are investigated during 2007 to 2016 to choose the most appropriate probability distribution in every station. Rainfall data obtained from the Meteorology Climatology and Geophysics Council (BMKG). In this paper Maximum Likelihood Method is used for estimation parameters of distributions Normal, Log Normal and Gamma. After the data is estimated then data is tested using Kolmogorov-Smirnov test. Based on the Kolmogorov-Smirnov test, Normal distribution has been discovered as the most appropriate distribution to represent monthly rainfall in Pematangsiantar.

1. Introduction
The amount of water availability in an area can be assessed several factors such as the rainfall of the area. Therefore, it is a substantial thing to manage the water supply optimally because it has a major impact on the economy and life of a region. The amount of water supply that is too little will have a bad impact for an area such as drought. However, the amount of water supply that is too much can also cause natural disasters that interfere with daily activities such as flooding. To manage water supplies, an understanding of the characteristics of rainfall in the form of rainfall distribution studies is needed.

Sen et. al (1999) studied that the distribution of monthly rainfall data in Libya is found as Gamma distribution that was confirmed based on the Chi-Squared test. Husak et. al (2006) studied the monthly distribution function of rainfall in Africa for the use of drought monitoring. It is found that the Gamma distribution fits approximately 98% for the monthly rainfall in Africa. Alghazali et. al (2014) fitted thirteen rainfall stations in Iraq with three probability distributions namely Normal, Gamma and Weibull with a goodness of fit test obtained from five of thirteen rainfall stations were Normal distributed. Mohamed et. al (2016) identified an appropriate probability distribution for the period 1971 to 2010 from fourteen rain stations in Sudan. The five tested distributions are: Normal, Log-Normal, Gamma, Weibull, and Exponential. It is found that the distribution of Normal and Gamma is the most suitable distribution of the fourteen rain stations in Sudan.

Based on the research of Sen et. al, Husak et. al, Mohamed et. al, and Alghazali et. al each using a commonly used distribution on arid and semi arid regions. As far as the authors know has not been found any research on the tropics are experiencing drought due to low rainfall. For that the authors try to examine areas with these characteristics, namely Pematangsiantar by taking samples of three rain stations using probability distribution Normal, Log Normal and Gamma.
2. Material and Methods

2.1 Data
Rainfall data is obtained from the Meteorology Climatology and Geophysics Agency (BMKG) during 2007 to 2016. The station is selected based on the pathway through each connected station. Picture 1 shows the histogram of Marihat Station, Picture 2 shows the histogram of Bangun Station, Picture 3 shows the histogram of Bah Jambi Station.

![Picture 1. Histogram of Marihat Station](image1)

![Picture 2. Histogram of Bangun Station](image2)

![Picture 3. Histogram of Bah Jambi Station](image3)
2.2 Maximum Likelihood
Redefined about Maximum Estimation Likelihood. Maximum Estimation Likelihood is a method that maximizes likelihood function. The principle used for the maximum estimate of likelihood is to choose \( \hat{\theta} \) as the point estimate for \( \theta \) which maximizes \( L(X; \theta) \). The maximum likelihood estimation method makes the likelihood function \( L(X; \theta) \) to be maximum by using logarithmic functions. So the likelihood logarithm function is denoted by \( \ln L(X; \theta) = l(X; \theta) \), where \( l(X; \hat{\theta}) > l(X; \theta) \). By using logarithms \( l(X; \theta) \), the likelihood estimator is obtained from the likelihood function derivative of its parameters, that is \( \frac{d(l(X; \theta))}{d\theta} = 0 \).

2.3 Normal Distribution
A random variable \( X \) is said to be normally distributed \( X \sim N(\mu, \sigma^2) \), if and only if the density function is in the form of:

\[
f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left[-\frac{1}{2} \left(x-\mu\right)^2 \right]; \quad -\infty < x < \infty, -\infty < \mu < \infty, \sigma^2 > 0
\]  

where:
- \( \mu \): expectation
- \( \sigma^2 \): variance

2.4 Gamma Distribution
Suppose \( X \) is a continuous random variable distributes gamma with parameters \( \alpha \) and \( \beta \), if and only if the density function is in the form of:

\[
f(x) = \begin{cases} 
\frac{1}{\Gamma(\alpha)} \beta^\alpha \theta^{\alpha-1} e^{-\frac{x}{\theta}}, & x > 0 \\
0, & \text{x lainnya}
\end{cases}
\]

If a random variable \( X \) is said to be gamma distributionn \( X \sim G(\alpha, \beta, 0) \) then the mean and variance of the Gamma distribution are:

\[
E(X) = \alpha\beta \quad \text{dan} \quad \text{Var}(X) = \alpha\beta^2
\]

2.5 Log Normal Distribution
Suppose a random variable $X$ is a positive real number ($0 < x < \infty$) such that $Y = \ln x$ is said to be a Log Normal distribution with the expectation $\mu$ and variance $\sigma^2$. $X = e^Y$ is a Log Normal distribution or can be written as $X \sim \mathcal{N}(\mu, \sigma^2)$. Since $X$ and $Y$ are connected with the relation $Y = \ln x$, then the density function of Log Normal distribution is:

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(\ln x-\mu)^2}{2\sigma^2}}, x > 0$$

(4)

If a random variable $X$ is a Log normal distribution $X \sim \mathcal{N}(\mu, \sigma^2)$ then the mean and variance of the Log Normal distribution are:

$$E(X) = e^{\mu + \frac{\sigma^2}{2}}$$

and

$$\text{Var}(X) = \sigma^2 e^{2\mu + \sigma^2} (e^{\sigma^2} - 1)$$

(5)

2.6 Goodness of Fit Test

Goodness of fit test is used based on the complete cumulative distribution function with predetermined parameters. In this paper, an appropriate distribution model for the data will be determined using the Kolmogorov-Smirnov test.

1. Kolmogorov-Smirnov Test

Kolmogorov-Smirnov principle is comparing the function of empirical distribution with a special distribution function contained in the Kolmogorov-Smirnov test table. Formulation of testing hypothesis Kolmogorov Smirnov:

$$H_0: D_{\text{max}} < D$$

$$H_1: D_{\text{max}} > D$$

Kolmogorov-Smirnov statistical test is shown in the following equation:

$$D_{\text{max}} = \text{Max} |P_e - P_t|$$

With critical areas refused $H_0$ if $D > D_\alpha$

where:

$H_0$ = the data is distributed according to the selected distribution

$H_1$ = the data is not distributed according to the selected distribution

$D$ = the critical value of form Kolmogorov-Smirnov table

$D_{\text{max}}$ = the largest absolute difference between empirical probability and theoretical probability

$P_e$ = empirical probability

$P_t$ = theoretical probability

3. Results and Discussions

3.1. Parameter Estimation with Maximum Likelihood Method

Three probability distributions are Normal, Log Normal and Gamma, each estimation of the parameters using Likelihood Estimation Method, the result obtained as the table below:
3.2. Kolmogorov-Smirnov Test

Kolmogorov-Smirnov's testing principle is to compare the value of $D$ obtained from the Kolmogorov-Smirnov critical table with the value of $D_{max}$ obtained from the observation calculation, with 95% confidence interval, then:

$$D = \frac{1.36}{\sqrt{n}}$$

Table 2: Critical values $D$ and observations $D_{max}$ Kolmogorov-Smirnov test for Rain Station in Pematangsiantar

| Stations     | Normal $\mu$, $\sigma$ | Log Normal $\hat{\mu}$, $\hat{\sigma}$ | Gamma $\hat{\alpha}$, $\hat{\beta}$ |
|--------------|-------------------------|-------------------------------------------|----------------------------------------|
| Marihat      | 237,883, 115,56         | 0,197216, 0,063164                        | 4,51, 52,69                           |
| Bangun       | 202,23, 100,79          | 2,231786, 0,085307                        | 4,02, 50,23                           |
| Bah Jambi    | 191,583, 106,76         | 2,186791, 0,135273                        | 3,21, 59,50                           |

Based on Table 2 it can be seen that the values of $D_{max}$ in the Normal distribution for each station are less than the values of $D$.

4. Conclusions

Normal distribution has been discovered as the most appropriate distribution to represent monthly rainfall in Pematangsiantar. Based on the Kolmogorov-Smirnov test results with $\alpha = 0.05$ the three rain stations are Normally distributed. If considered from a critical value, then the Log distribution of Normal is rated second and the Gamma distribution is rated to last.

References

[1] Mood A M, Graybill F A and Boes D C 1974 *Introduction to the Theory of Statistics* (New York: McGraw-Hill)
[2] Aitchison J and Brown J A C 1963 *The Log Normal Distribution With Special Reference To Its Uses In Economics* (Great Britain: Cambridge University Press)
[3] Hann C T 1977 *Statistical Methods in Hydrology* (Ames: The Iowa State University Press)
[4] Husak G J, Michaelsen J and Funk C 2006 *International Journal of Climatology* 27 935
[5] Jamaluddin S and Jemain A A 2007 *Journal of Applied Sciences* 7 1880
[6] Mohamed T M and Ibrahim A A A 2016 *SUST Journal of Engineering and Computer Sciences* 17 34
[7] Alghazali N O S and Alawadi D A H 2014 *Civil and Environmental Research* 6 40
[9] Ozturk A 1981 *Journal of Applied Meteorology* **20** 1499
[10] Roldan J and Woolhiser W A 1982 *Water Resources Research* **18** 1451
[11] Sen Z and Eljadid A G 1999 *Hydrological Sciences Journal* **44** 665
[12] Sharda V N and Das P K 2005 *Agricultural Water Management* **76** 120
[13] Thom H C 1958 *Monthly Weather Review* **86** 117
[14] Walpole R E, Myres R H, Myres S L and Ye K 2007 *Probability & Statistics for Engineers and Scientists* (United States: Prentice Hall)
[15] Wilks D 1990 *Journal of Climate* **3** 1495
[16] Wong R K W 1977 *Journal of Applied Meteorology* **16** 1360