Probability Distribution of Rainfall in Medan

Elly Rosmaini[1] and Yoni Yolanda Saphira[1]*

[1]Department of Mathematics, Universitas Sumatera Utara, Medan, 20155, Indonesia

Abstract. In this paper we chose three stations in Medan City, Indonesia to estimate Monthly Rainfall Data i.e. Tuntungan, Tanjung Selamat, and Medan Selayang Stations. We took the data from 2007 to 2016. In this case fitted with Normal, Gamma, and Lognormal Distributions. To estimate parameters, we used this method. Furthermore, Kolmogorov-Smirnov and Anderson Darling tests were used the goodness-of-fit test. The Gamma and Normal Distributions is suitable for Tuntungan and Medan Selayang Stations were stated by Kolmogorov-Smirnov's test. Anderson Darling's test stated that Gamma Distribution was suitable for all stations.

Keyword: Probability Distribution, Test of Goodness-Of-Fit, The Method of Moments.

Abstrak. *Pada makalah ini kami memilih tiga stasiun pada kota Medan, Indonesia untuk memperkirakan Data Curah Hujan Bulanan yaitu Stasiun Tuntungan, Tanjung Selamat dan Medan Selayang. Kami mengambil data dari 2007 sampai 2016. Pada kasus ini dicocokkan dengan distribusi Normal, Gamma dan Lognormal. Untuk memperkirakan parameter, kami menggunakan metode ini. Lebih lanjut, uji Kolmogorov-Smirnov dan Anderson Darling digunakan untuk uji goodness-of-fit. Distribusi Gamma dan Normal cocok untuk stasiun Tuntungan dan Medan Selayang ditetapkan oleh uji Kolmogorov-Smirnov. Uji Anderson Darling menyatakan bahwa distribusi Gamma cocok untuk semua stasiun.*

Kata Kunci: *Distribusi Probabilitas, Uji Goodness-Of-Fit, Metode dari Momen*

1. Introduction

The distribution function of rainfall and rainfall prediction in Libya has been studied by Sen and Eljadid (1999). Distribution of monthly rainfall in Libya has met the Gamma probability distribution functions are tested using Chi-Square test. The series of all rainfall recorded at least 20 years in Libya investigated statistically and Gamma distribution parameters are calculated on existing stations. Prediction rainfall amounts around 10, 25, 50, and 100 mm is achieved make use of the probability function.

Suhaila Abdul Aziz Jamaludin and (2007) classified daily rainfall data according to four types of rain (types 1, 2, 3, and 4). Gamma, Weibull, Kappa, and mixed Exponential are four distributions which are tested to conform to the number of daily rainfall in Malaysia peninsula. Parameter estimation for each distribution uses the maximum likelihood method. Based on tests of goodness-

*Corresponding author at: Universitas Sumatera Utara, Medan, 20155, Indonesia

E-mail address: elly1@usu.ac.id

of-fit, Mixed Exponential distribution declared as the most appropriate distribution to describe the amount of daily rainfall in Malaysia peninsula.

Based on the book "Statistical Methods in Hydrology" (Hann CT, 1977), it has summarized that the Gamma distribution is the most appropriate for monthly and annual precipitation. In the book "Arid Lands and Water Evaluation and Management" (Maliva and Missimer, 2012), has shown that one of the distribution Normal, Log-Normal, Gamma, Weibull and Gumbel in accordance with the rainfall data from the arid and semi-arid.

We take three stations as samples for the area of Medan and use the Normal distribution, Gamma and Log-Normal. The goodness-of-fit test is used by Kolmogorov-Smirnov dan Anderson Darling. The rainfall pattern at the three stations in Medan can be used as consideration in predicting rainfall so that it can anticipate problems such as problems in the development of infrastructure and flood prevention in Medan.

2. Materials and Methods

a) Normal distribution
A random variable X is said to be normally distributed, if its density function has the form (Montgomery and Runger, 2014):

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} . e^{\left[\frac{-(x-\mu)^2}{2\sigma^2}\right]} ; -\infty < x < \infty \tag{1}$$

with parameter μ where $-\infty < \mu < \infty$ and $\sigma > 0$.

Mean of normal distribution: $E(X) = \mu$

Variance of the normal distribution: $Var(X) = \sigma^2$

b) Gamma distribution
A random variable X is said to be Gamma distributed if and only if its density function has the form (Montgomery and Runger, 2014):

$$f(x) = \frac{\lambda^r x^{r-1} e^{-\lambda x}}{\Gamma(r)} ; x > 0, \lambda > 0, r > 0 \tag{2}$$

Mean of gamma distribution: $E(X) = \frac{r}{\lambda}$

Variance gamma distribution: $Var(X) = \frac{r}{\lambda^2}$

c) Lognormal Distribution
Let W has a normal distribution with mean θ and variance ω^2, then $= \exp(W)$ is a random variable which is lognormal distributed with the following probability density function (Montgomery and Runger, 2014):

$$f(x) = \frac{1}{x\omega\sqrt{2\pi}} e^{\left[-\frac{(\ln(x)-\theta)^2}{2\omega^2}\right]} ; 0 < x < \infty \tag{3}$$

Mean of Lognormal distribution: $E(X) = e^{\theta+\frac{\omega^2}{2}}$

Variance of Lognormal distribution: $Var(X) = e^{2\theta+\omega^2}(e^{\omega^2} - 1)$

d)   Moment Method

Let $X_1, X_2, \ldots, X_n$ are random samples of probability distribution $f(x)$ where $f(x)$ can be discrete probability mass function or continuous probability density function. The k-th pulation moment is (Wackerly, Mendenhall, and Scheaffer, 2008):

$$\mu_k' = E(X^k) \tag{4}$$

The appropriate k-th sample moment is:

$$m_k' = \frac{1}{n} \sum_{i=1}^{n} X_i^k \tag{5}$$

Moment method is based on the intuitive idea that states that the moment of the sample should provide a good estimate from appropriate population moment. That is, m_k' has to be a good estimation of μ_k', where $k = 1, 2, \ldots$. Next, since population moment $\mu_1', \mu_2', \ldots, \mu_k'$ is a function of population parameter, then moment of population can be equated to the moment of sample and completing the desired estimator. Therefore, the method of momentscan be expressed as follows:

Choose as estimation of parameter values which is a solution of the equation, $\mu_k' = m_k'$, where $k = 1, 2, \ldots, t$, and t is the number of parameters to be estimated.

e)   Test goodness-of-fit

The procedure steps to execute the Kolmogorov-Smirnov test to give a collection of sample values $x_1, x_2, \ldots, x_i$ examined from the population X by article Soong (2004) is as follows:

a. Arranging the sample values from the smallest to biggest, denoted by x_i.

b. Determining the observed distribution function $F^0[x_i]$ at each x_i using $F^0[x_i] = \frac{i}{N}$.

c. Determining theoretical distribution function $F_x[x_i]$ at each x_i using distribution hypothesis which has been obtained and the deviation is determined from the equation:
$$d_2 = \max\{|F^0[x_i] - F_x[x_i]|\} \tag{6}$$

d. The maximum absolute value d_2 obtained from the equation $d_2 = \max\{|F^0[x_i] - F_x[x_i]|\}$, compared to the critical value shown in the statistics table. If the critical value is greater than equal d_2, then d tested distribution is suitable to describe the observed data, if it is not holds, that is the critical value greater than equal d_2, we get the tested distribution is not suitable for Kolmogorov-Smirnov, which is shown as follows (Lothar, 1984):

Table 1 The critical value of Kolmogorov-Smirnov test

limit to d_2	α
$\dfrac{1{,}073}{\sqrt{n}}$	0,20
$\dfrac{1{,}138}{\sqrt{n}}$	0,15
$\dfrac{1{,}224}{\sqrt{n}}$	0,10
$\dfrac{1{,}358}{\sqrt{n}}$	0,05
$\dfrac{1{,}628}{\sqrt{n}}$	0,01

Anderson Darling test is used to determine the distribution of the sample data. This test is a modification of the Kolmogorov-Smirnov test. Anderson Darling test uses a specific data distribution in calculating critical value.

The formula used to test Anderson Darling is:

$$AD = \left[\sum_{i=1}^{n} \frac{1-2i}{n} \{\ln(F_0[Z_i]) + \ln(1 - F_0[Z_{n-i+1}])\} \right] - n \tag{7}$$

where AD = Anderson Darling Test

   n = Amount of data

   F = Cumulative distribution function

Anderson Darling test uses the critical value according to the tested probability distribution as follows (Ang and Tang, 2007):

$$A^* = AD + \frac{0.2 + \frac{0.3}{\alpha}}{n} \; ; for \; gamma \; distribution \tag{8}$$

$$A^* = \frac{\alpha}{1 + \frac{0.75}{n} + \frac{2.25}{n^2}} \; ; for \; normal \; and \; lognormal \; distribution \tag{9}$$

where A^* = Critical value of Anderson Darling

n = Amount of data

α = The level of significance

Table 2 α significance level of Anderson Darling test

	α Level of Significance					
K	0.25	0.10	0.05	0.025	0.01	0.005
1	0.486	0.657	0.786	0.917	1.092	1.227
2	0.477	0.643	0.768	0.894	1.062	1.190
3	0.475	0.639	0.762	0.886	1.052	1.178
4	0.473	0.637	0.759	0.883	1.048	1.173
5	0.472	0.635	0.758	0.881	1.045	1.170
6	0.472	0.635	0.757	0.880	1.043	1.168
8	0.471	0.634	0.755	0.878	1.041	1.165
10	0.471	0.633	0.754	0.877	1.040	1.164
12	0.471	0.633	0.754	0.876	1.038	1.162
15	0.470	0.632	0.754	0.876	1.038	1.162
20	0.470	0.632	0.753	0.875	1.037	1.161
∞	0.470	0.631	0.752	0.873	1.035	1.159

3. Results and Discussion

The method of moment is used to estimate the parameters in the Normal, Gamma and Lognormal distribution. The result of parameter estimation by moment method using equation (4) and (5) shown in the following table:

Table 3 The result of parameter estimation using Method of Moments

Station	Normal		Gamma		Lognormal	
	$\hat{\mu}$	$\hat{\sigma}^2$	$\hat{\lambda}$	$\hat{r}$	$\hat{\theta}$	$\hat{\omega}^2$
Medan Selayang	229,742	14.885,5416	0,015433903	3,545815743	5,312735695	0,24843959
Tanjung Selamat	182	17.288	0,010527534	1,916011188	4,994033351	0,419946678
Tuntungan	241,925	20.636,786	0,011722998	2,836086291	5,33761395	0,30202762

Parameter estimation results are used to the test goodness-of-fit, in this case, Kolmogorov-Smirnov and Anderson Darling test. This goodness-of-fit test is used to examine the appropriate distribution of the monthly rainfall in Medan Selayang, Tanjung Selamat, and Tuntungan stations. The results of the goodness of-fit test for Normal Distribution, Gamma and Lognormal are as follows:

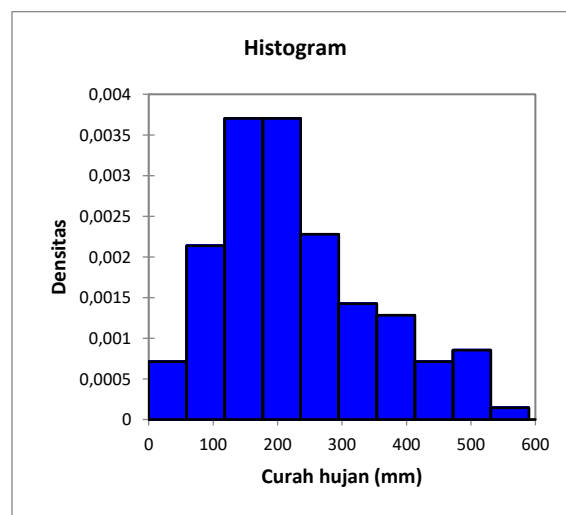**Table 4** The results of Goodness-of-Fit test using the Kolmogorov-Smirnov

Station	n	Critical value	Normal	Gamma	Lognormal
Medan Selayang	120	0,123237575	0,102710069	0,102710011	0,424515726
Tanjung Selamat	120	0,123237575	0,125003484	0,125003474	0,39424123
Tuntungan	120	0,123237575	0,086193286	0,086193283	0,407912624

Based on the results of the Kolmogorov-Smirnov test using equation (6) and Table 1, it is clear that the Normal and Gamma Distribution is apropriate with the monthly rainfall data in Medan Selayang and Tuntungan stations. While the monthly rainfall data at Tanjung Selamat station is not appropriate with the three distributions.
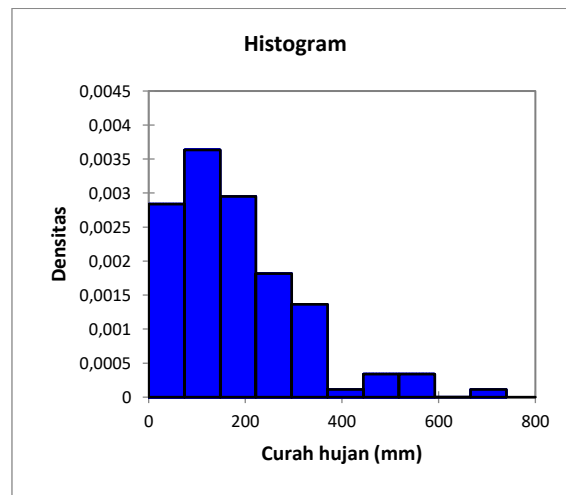
Table 5 The results of Goodness-of-Fit test using Anderson Darling

| Station	Normal		Gamma		Lognormal	
	A^*	AD	A^*	AD	A^*	AD
Medan Selayang	0,747213166	1,71859701	1,723586083	1,718594948	0,747213166	1,529518631
Tanjung Selamat	0,747213166	3,277427696	3,282419502	3,277428367	0,747213166	1,486938223
Tuntungan	0,747213166	1,791307447	1,796298604	1,791307469	0,747213166	1,518192187

Based on the result of Anderson Darling test using equation (7), (8), (9), and Table 2, it is easy to see that the Gamma distribution is appropriate with all the stations in this study.

Figure 1 Histogram of Medan Selayang station

Figure 2 Histogram of Tanjung Selamat station
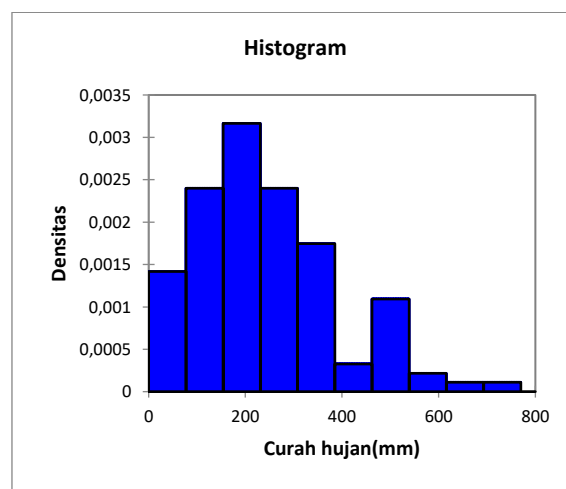
Figure 3 Histogram of Tuntungan station

4. Conclusion

The results of Kolmogorov-Smirnov goodness-of-fit test states that the Normal and Gamma Distribution is appropriate with the monthly rainfall data in Medan Selayang and Tuntungan stations. The results of Anderson Darling goodness-of-fit stated that the Gamma distribution is appropriate with all of the stations in this study.

REFERENCES

[1]   A. HS. Ang and W. H. Tang, *Probability Concepts in Engineering: Emphasis on Applications to Civil and Environmental Engineering*, 2nd Ed., John Willey and Sons, Hoboken, New Jersey, 2007.

[2]   *C. T. Haan, Statistical Methods in Hydrology, The Iowa State University Press, Ames, 1977.*

[3]   *R. Maliva and T. Missimer, Arid Lands and Water Evaluation and Management, Springer, Berlin, 2012.*

[4]   D. Montgomery and G. C. Runger, *Applied Statistics and Probability for Engineers*, John Wiley and Sons, USA, 2014.

[5]   L. Sachs, *Applied Statistics: A Handbook of Techniques*, Springer, Verlag, 1984.

[6] *Z. Sen and A. G. Eljadid, "Distribution Function for Libya and Rainfall Prediction", Hydrological Sciences Journal, Istanbul, Turkey, vol. 44, no. 5, pp. 665-680, 1999.*

[7] *S. Jamaludin and A. A. Jemain, "The Statistical Distributions to The Daily Rainfall Amount in Peninsular Malaysia", Technology Journal, vol. 46C, pp. 33-48, 2007.*

[8] *T. T. Soong, Fundamentals of Probability and Statistics for Engineers, John Wiley and Sons, New York, USA, 2004.*

[9] D. Wackerly, W. Mendenhall, and R. L. Scheaffer, *Mathematical Statistics with Applications*, Thomson Brooks / Cole, USA, 2008.