ESTIMATION OF GENERALIZED GAMMA DISTRIBUTION PARAMETER WITH PROBABILITY WEIGHTED MOMENT METHOD

Dian Kurniasari[1*], Warsono[1], Widiarti[1], Yeftanus Antonio[1], M. Azram[2] and Mustofa Usman[1]

[1]Department of Mathematics, Faculty of Mathematics and Natural Sciences, Universitas Lampung, Indonesia

[2]Department of Mathematics and Statistics (DOMAS), Sultan Qaboos University, Muscat, Oman

*Email: dian.kurniasari@fmipa.unila.ac.id

ABSTRACT: *The classical Gamma distribution with two parameters is the distribution most commonly used in modeling the distribution of environmental quality data. However, this distribution is less precise for environmental quality data fittings. One way to overcome this is by making generalization of Gamma distribution. In this study Generalized Gamma distribution with parameters **b**, **d**, and **k** will be used as a model of water quality data. The parameters of the Generalized Gamma distribution probability model are to be estimated with the Probability Weighted Moment Method. To see the accuracy of estimations in various sample sizes, the Monte Carlo simulation or experiment is used to generate data. Using the data, the result indicates initial guess value b = 0.14, d = 3.28 and k = 0.001.Simulation with Monte Carlo experiments for parameter estimation with Probability Weighted Moment Method on parameters **b**, **d** and **k** shows better results if the sample sizes used are larger. For modeling environment quality data that has Generalized Gamma distribution, regular (routine) sampling is necessary. The Probability Weighted Moment method can be an alternative method of estimation used in Generalized Gamma distribution.*

Keywords:Generalized Gamma distribution, Probability Weighted Moment, Monte Carlo Simulation, Water Quality.

1. INTRODUCTION

Gamma distribution is one of the most commonly used distributions in environmental quality data modeling. The gamma distribution is one of a continuous probability distribution family with two parameters. However, this distribution does not necessarily fit the data well for all types of environmental quality data. Berger, et al.,[1] noted, for the sulphur dioxide data in the Gent region of Belgium, the gamma distribution gives a more precise picture. One way to overcome this is to develop the distribution of gamma into a more generalized distribution. In this case, the generalized model must contain a model commonly used in environmental pollution data modeling. For modeling to be generally applicable to each state of data, the gamma distribution is generalized by three parameters also called as Generalized Gamma (GG) distribution

Gamma distribution is very useful in modeling data distribution, among others: survival. Because of its importance, the parameter estimation for the distribution of data that has Gamma probability model should be done precisely, accurately, and efficiently. One of the most popular methods of estimating the parameters of a distribution is the maximum likelihood method.

Unfortunately, the maximum likelihood method is based on large sample theory, so this method often works poorly for data with "small" sample sizes or even for data with "medium" sample sizes. It is therefore very interesting to look for alternative methods to estimate the parameters of a distribution, and in this case it is proposed to use the probability-weighted moment method.

Marani, et al.,[2] used a Generalized Gamma model for modeling air quality data distribution with satisfactory results. Thus, it is very interesting to apply the same model to other environmental pollution data, such as water quality data.To find out if a distribution model works well in modeling a data

set, the parameters of the model should be estimated first. One of the most well known methods of estimation is the maximum likelihood method. Greenwood et al. [3] and Holland and Fitz-Simons [4] stated, since the maximum likelihood method is based on large sample theory, this method often works less satisfactorily for "small" sample sizes data or even for "medium" sample sizes. To address this problem, [3] proposed the use of the Probability Weighted Moment method, as an alternative to the maximum likelihood method, to estimate the parameters of some distributions, such as the Gumbel distribution and the generalized lambda. Meanwhile, Shoukri, et al.,[5] applied the alternative method to log-logistics distribution. Therefore, to estimate the parameters of the Generalized Gamma distribution, this study uses Probability Weighted Moment method as an alternative to the maximum likelihood method. This study also examines the effect of sample size by Monte Carlo simulation or experiment on the appearance of Probability Weighted Moment method in estimating the parameters of environmental quality data distribution following the *Generalized Gamma* distribution.

2. LITERATURE REVIEW

2.1 Generalized Gamma Distribution

According to Diciccio [6], a random variable X is said to have a generalized distribution of gamma probabilities with parameters α, β and θ if and only if the probability function of X is:

$$f(x) = \frac{\beta}{\Gamma(\alpha)} \frac{x^{\beta\alpha-1}}{\theta^{\beta\alpha}} e^{-\left(\frac{x}{\theta}\right)^{\beta}} ; x > 0; \alpha, \beta, \theta > 0$$

Parameters α and β are known as shape parameters and θ parameter is known as scale parameter.

2.2 The Probability Weighted Moment Method

In relation to the weakness of the maximum likelihood method described earlier for a "small" sample,[3] introduce the use of the Probability Weighted Moment method as an alternative to the maximum likelihood method. The Probability Weighted Moment method of the random variable X with the function of the cumulative distribution is defined as follows:

$$M_{r,s,t} = E[X^r(F(X)^s) \; (1 - F(X)^t)]$$

In this case r, s, and t are the real numbers. If $s = t = 0$ and r is a non-negative integer, then $M_{r,0,0}$ is a conventionally known moment of probability. Let $X(F)$ be the inverse of cumulative distribution, then the Probability Weighted Moment can be written in the form:

$$M_{r,s,t} = \int_0^1 (X(F)^r)F^s(1 - F(X))^t dF$$

Methods of Probability Weighted Moment was introduced by Landwehr, et al.,[7] in Gumbel distribution. More specifically Hosking, et al.,[8] discussed the characteristics of probability distribution parameters of generalized extreme-value distribution generated by the Probability Weighted Moment method. Shoukri, et al.,[5] concluded that based on the simulation results the instant-Probability Weighted Moment method can be calculated, without having to go through the iteration process and always produce a visible value, and the bias and variance are smaller on the "small" sample size, equal to 15 and 25.

3. RESEARCH METHOD

3.1 Method of Estimating Parameters

To compare the appearance of the Probability Weighted Moment method in estimating the Generalized Gamma distribution parameters, the estimation of the environmental data distribution parameters model is performed by this method. Before the method is applied, it is necessary to derive and develop the predictions procedure for the Generalized Gamma distribution.

In Probability Weighted Moment method, the procedure of estimation begins with finding the inverse function of the cumulative distribution of the Generalized Gamma distribution, i.e. $X(F)$. Then the estimation parameter is calculated by solving the following equation:

$$M_{r,s,t} = \int_0^1 (X(F)^r)F^s(1 - F(X))^t dF$$

in this case $r = 1$, $s = s$, and $t = 0$. Having obtained the parameter estimation, which is still expressed in the form M_r, The unbiased estimators for M_r are obtained based on the samples of $X_{(1)} < X_{(2)} < \ldots < X_{(n)}$ from the random sample of size n and by solving the equation:

$$\widehat{M}_r = \frac{1}{n}\sum_{j=1}^{n} \frac{(j-1)(j-2)\ldots(j-r)}{(n-1)(n-2)\ldots(n-r)} X_{(j)}$$

A review of the performance of the Probability Weighted Moment method in estimating the Generalized Gamma distribution parameters for the various sample sizes is done by Monte Carlo simulation or experiment. The assessment of the two methods performance is based on the features of unbiasedness and variances, i.e. by looking at the biased values and variances of the probability parameters produced for different sample sizes.

3.2 Data Usage

The purpose of this study is to obtain the results from the Probability Weighted Moment method on Generalized Gamma Distribution on seawater quality data. To illustrate the estimation results, the data will be used as the basis for determining the parameters. This research uses sea water quality data obtained from University of Lampung. The location of data collection is on several rivers in the Coastal Area of Bandar Lampung City (Lampung Bay) which include Way Sukamaju, Way Keteguhan, Way Kuripan, Way Kunyit, Way Kuala, Way Lunik, and Way Galih. The seawater quality are assessed in through physical parameter including TDS and TSS as well as chemical parameters including DO, COD, BOD, Hardness, Alkalinity, PO4, SO4, Nitrite, Nitrate, Iron (Fe), Sulfide, Pb, Hg, Cu, and Cd with the same unit in mg/l.

3.3 Monte Carlo Simulation Scenario

Monte Carlo simulation scenarios to be conducted in this study are as follows:

a. Generating a random sample of size n = 5.
b. Generating a random sample of size n = 10.
c. Generating a random sample of size n = 25.
d. Generating a random sample of size n = 50
e. Generating a random sample of size n = 100.
f. Generating a random sample of size n = 500.
g. Generating a random sample of size n = 1000.

for parameter estimation and simulation used in this research is program facilities provided in program R version 3.Generation of samples for all sample sizes above is conducted by simulation each of N = 500 times.

The sample sizes n = 5, n = 10 and n = 25 are considered to represent "small" samples, n = 50 is considered to represent a "medium" sample, and the sample sizes n = 100, n = 500, and n = 1000 are considered to represent " large" samples. The simulated data of Generalized Gamma distribution with predetermined parameters is done by utilizing uniform distribution.

4. RESULTS AND DISCUSSION

4.1 Estimation of Generalized Gamma Distribution Parameters Using Probability Weighted Moment Method (PWM)

To estimate the parameters of the Generalized Gamma distribution using the Probability Weighted Moment method, the first step is to determine the cumulative distribution function (CDF) of the Generalized Gamma distribution for subsequent use in searching M_(r,s,t) as the basis for applying Probability Weighted Moment method and then determines the parameter estimator.

4.2 Cumulative Distribution Function of Generalized Gamma Distribution

The cumulative function of Generalized Gamma distribution can be obtained by the following steps:

$$F(x) = \int_{-\infty}^{x} f(t)dt$$

$$F(x) = \int_{0}^{x} f(t)dt$$

$$= \int_{0}^{x} \frac{\beta}{\Gamma(\alpha)} \frac{t^{\beta\alpha-1}}{\theta^{\beta\alpha}} e^{-\left(\frac{t}{\theta}\right)^{\beta}} dt = \frac{\beta}{\Gamma(\alpha)\theta^{\beta\alpha}} \int_{0}^{x} t^{\beta\alpha-1} e^{-\left(\frac{t}{\theta}\right)^{\beta}} dt$$

Due to the many parameters in the integral function, it is necessary to simplify as follows:

$$u = \left(\frac{t}{\theta}\right)^{\beta}$$

$$t = u^{\frac{1}{\beta}}\theta$$

$$dt = \frac{1}{\beta} u^{\frac{1}{\beta}-1} \theta du$$

Limit:

$$t = 0 \rightarrow u = 0$$

$$t = x \rightarrow u = \left(\frac{x}{\theta}\right)^{\beta}$$

So the above equation can be written as:

$$F(x) = \frac{\beta}{\Gamma(\alpha)\theta^{\beta\alpha}} \int_{0}^{\left(\frac{x}{\theta}\right)^{\beta}} \left(u^{\frac{1}{\beta}}\theta\right)^{\beta\alpha-1} e^{-u} \frac{1}{\beta} u^{\frac{1}{\beta}-1} \theta du$$

$$= \frac{\beta}{\Gamma(\alpha)\theta^{\beta\alpha}} \frac{\theta^{\beta\alpha}}{\beta} \int_{0}^{\left(\frac{x}{\theta}\right)^{\beta}} u^{\alpha-\frac{1}{\beta}} e^{-u} u^{\frac{1}{\beta}-1} du$$

$$= \frac{1}{\Gamma(\alpha)} \int_{0}^{\left(\frac{x}{\theta}\right)^{\beta}} u^{\alpha-1} e^{-u} du$$

$$= \frac{1}{\Gamma(\alpha)} \int_{0}^{z} u^{\alpha-1} e^{-u} du \; ; \; z > 0$$

Thus, the cumulative function (CDF) obtained from the Generalized Gamma distribution is:

$$F(x) = \frac{1}{\Gamma(\alpha)} \int_{0}^{z} u^{\alpha-1} e^{-u} du \; ; \; z > 0 \quad and \quad z = \left(\frac{x}{\theta}\right)^{\beta}$$

The cumulative function of the GB2 distribution is an incomplete gamma function, so the inverse of the cumulative function cannot be resolved analytically but numerically. Therefore, to find the probability weighted form of the GB2 distribution used to estimate the parameter values should also be resolved numerically.

4.3 Numerical Richardson Integral Method
Richardson extrapolation is a method that uses two estimations of an integral to compute a more accurate third estimator. The estimations and errors associated with the multi-application trapezoidal rule can be described in general as

$$I = I(h) + E(h) \tag{1}$$

where I is the true value of the integral, $I(h)$ is an estimation of a trapezoidal rule with a segmented application n with the width of the step $h = \frac{b-a}{n}$ and $E(h)$ is the truncation error. If we make two different estimations using step width h_1 and h_2 then we get the following equation:

$$I(h_1) + E(h_1) = I(h_2) + E(h_2) \tag{2}$$

errors from multi-application trapezoidal rules can be estimated as:

$$E \cong -\frac{(b-a)^3}{12 \, n^2} f''$$

Since the value $n = \frac{b-a}{h}$ so the above equation can be changed into

$$E \cong -\frac{(b-a)}{12} h^2 f'' \tag{3}$$

If it is assumed that f'' is a constant which means it is not affected by step width then the value of E can be used to determine the ratio of both errors:

$$\frac{E(h_1)}{E(h_2)} \cong \frac{h_1^2}{h_2^2}$$

This calculation has an important effect on removing f'' from the calculation. Furthermore, the above ratio equation can be changed to:

$$E(h_1) \cong E(h_2) \left(\frac{h_1}{h_2}\right)^2 \tag{4}$$

Then substitute equation (4) into equation (2), and the following results are obtained:

$$I(h_1) + E(h_2) \left(\frac{h_1}{h_2}\right)^2 = I(h_2) + E(h_2)$$

$$E(h_2) \cong \frac{I(h_1)-I(h_2)}{1-\left(\frac{h_1}{h_2}\right)^2} \tag{5}$$

As a result we have developed an estimation of the deduction error in terms of the integral estimation and the width of the step. This estimation can be substituted into: $I = I(h_2) + E(h_2)$.

To produce an improved integral estimation:

$$I \cong I(h_2) + \frac{I(h_2) - I(h_1)}{\left(\frac{h_1}{h_2}\right)^2 - 1}$$

This shows that the error of this estimation is h^4. For a special case where the interval is divided into two $\frac{h_2-h_1}{2}$, the equation will be:

$$I \cong I(h_2) + \frac{I(h_2) - I(h_1)}{2^2 - 1}$$

or formed into:

$$I \cong \frac{4}{3} I(h_2) - \frac{1}{3} I(h_1)$$

This approach is a subset of a more general method for combining integral to produce an error that is $O(h^4)$ which can then be used to look for smaller errors i.e. $O(h^6)$, $O(h^8)$, and so on.

4.4 Determination of Water Quality Data Parameters
The parameter values to be estimated are obtained from the water quality parameter in some rivers in coastal area of

Bandar Lampung city (Source: University of Lampung). Using the data, the parameter values are obtained $b = 0.14$, $d = 3.28$ and $k = 0.001$. Data on water quality parameters in some rivers in the coastal area of Bandar Lampung city are believed to have distributed Generalized Gamma from Kolmogov Smirnov test. In Kolmogorov Smirnov rank test of Generalized Gamma distribution is rank 2 with p-value= 0.07232.

4.5 Simulation and Evaluation of Generalized Gamma Distribution Parameter Estimation

The Generalized Gamma distribution has three parameters where d and k are the shape parameters, and b is the scale parameter. The parameter estimation is performed by using software R. The parameter values to be estimated are parameters of seawater quality data obtained before, they are $b = 0.14$, $d = 3.28$ and $k = 0.001$.

After obtaining the parameter values to be estimated, then simulation is performed to evaluate the characteristics of each estimator. A good estimator is an unbiased, efficient, and consistent predictor. In this simulation the bias value, variance, and mean square error (MSE) of the estimations obtained by using sample sizes 5, 10, 25, 50, 100, 500, and 1000 are compared. Next, we will discuss the results of the estimation evaluation on parameter b on the Generalized Gamma distribution with the Probability Weighted Method in Table 1. From the results of parameter estimation b in Table 1, it shows that the larger the samples used, the closer the estimator value of the actual parameter. This result can be supported with an increasingly small bias value as the sample size increases. The larger the sample size used, the closer to zero the bias value. Variance of estimator for parameter b shows that if the size of the sample used is larger then the value of the variance is also getting closer to zero. Likewise for the MSE value estimator for b, the larger the sample used, the smaller the error value. Based on the results of the estimator evaluation shown in Table 1, the estimator for b with the Probability Weighted Moment Method on the Generalized Gamma distribution, the larger the sample size used in the estimation by the Probability Weighted Moment Method, the better the estimation will be with the bias, variance, and MSE approaching zero.

Table 1: The predicted, biased, variance, and MSE values for parameter b with initial value b = 0.14 from Generalized Gamma distribution

Sample Size	$\hat{b}$	Bias $\hat{b}$	Variance $\hat{b}$	MSE $\hat{b}$
5	11.4004	11.2604	10.79596	137.5926
10	10.83954	10.69954	10.35214	124.7433
25	10.15808	10.01808	9.98652	110.3484
50	9.15263	9.01263	8.61373	89.84123
100	7.889112	7.749112	7.02612	67.07486
500	5.31125	5.17125	4.98748	31.72931
1000	3.14271	3.00271	2.87251	11.88878

Table 2: The predicted, biased, variance, and MSE, values for parameter d with initial value d =3.28 of the Generalized Gamma distribution

Sample Size	$\hat{d}$	Bias $\hat{d}$	Variance $\hat{d}$	MSE $\hat{d}$
5	21.79714	18.51714	11.0779	353.9624
10	21.52569	18.24569	10.90974	343.8148
25	20.5857	17.3057	10.17834	309.6656
50	19.2222	15.9422	9.03112	263.1849
100	17.56315	14.28315	7.96581	211.9742
500	15.25442	11.97442	6.01637	149.4031
1000	12.42064	9.14064	3.25024	86.80154

The estimation value of parameter d gets closer to the actual parameter as the number of samples increases. Table 2 summarizes estimator evaluation by bias, variance, and MSE. For the bias of estimated parameter d in the Generalized Gamma distribution, bias will be smaller or closer to zero as the number of samples gets larger. Likewise on the variance and the MSE, the value of the variance and the MSE is getting closer to zero when the sample size used is larger.

Thus, for estimating the parameter d by the Least Square Method indicates better estimation results when the sample size used is larger indicated by the bias, the variance, and the MSE approaching zero.

Table 3 presents the estimator evaluation results with the bias, variances, and MSEs by the estimation parameter k with the Probability Weighted Moment method on Generalized Gamma distribution. Based on the results in Table 3, the estimation of parameter k with the Probability Weighted Moment Method produces bias, variance, and MSE approaching zero as the sample size increases. The estimated value of the parameter k is also closer to the actual value for the larger sample size. As for estimation parameters b and d, estimation parameter k gets better as the sample size gets larger based on the evaluation of estimation bias, variance, and MSE parameters. This suggests that the Probability Weighted Moment Method will be more accurate to estimate the parameters of water quality data in Generalized Gamma distribution when the sample size used is larger.

Table 3: The predicted, biased, variance, and MSE values for parameter k with initial value k = 0.001 of Generalized Gamma distribution

Sample Size	$\hat{k}$	Bias $\hat{k}$	Variance $\hat{k}$	MSE $\hat{k}$
5	9.99128	9.98128	8.54105	108.167
10	9.62705	9.61705	8.02417	100.5118
25	9.1823	9.1723	7.89802	92.02911
50	8.06921	8.05921	6.81571	71.76658
100	6.27371	6.26371	5.23102	44.46508
500	4.20023	4.19023	4.00163	21.55966
1000	2.14125	2.13125	1.97423	6.516457

4.6 Fitting Parameter Estimation on Water Quality Data
Fitting on water quality data in several rivers in Coastal Area of Bandar Lampung City is known to have Generalized Gamma distribution with parameter b = 0.14, d = 3.28 and k = 0.001. This simulation result of parameter estimation with Least Squares method indicates that the larger the sample size used, the better the parameter estimation is.

These results provide the foundation that the environmental quality data of several rivers in the Coastal Area of Bandar Lampung City follow the more Generalized Gamma or Generalized Gamma distribution. To obtain information for both modeling and distribution, parameter estimation with Probability Weighted Moment Methods gives better results for larger samples. Therefore, in using Generalized Gamma distribution as distribution in modeling of water quality data especially in Coastal Area of Bandar Lampung City it is required to do continuous or routine sampling so that the number of samples used is larger and gives better estimation results with Probability Weighted Moment Method.

5. CONCLUSION
The water quality data on several rivers in the coastal area of Bandar Lampung follow the Generalized Gamma distribution with parameter value $b = 0.14$, $d = 3.28$ and $k = 0.001$. The simulation result for parameter estimation with Probability Weighted Moment method for parameters b, d and k is better if the sample size used is larger based on the evaluation value of bias estimation, the variance and the MSE approaching zero.

For the modeling of environmental quality data that has Generalized Gamma distribution, regular (routine) sampling is necessary because the parameter estimation on the distribution with Probability Weighted Moment method will be better if the number of samples used is larger. The Probability Weighted Moment method can be an alternative method of estimation used in Generalized Gamma distribution.

6. REFERENCES
[1] Berger, A., Melice, J.L., Demuth, C. (1982). Statistical distributions of daily and high atmospheric SO2-concentrations. Atmospheric Environment 16(12):2863-2877.

[2] Marani, A., Lavagni, I., and Buttazzoni, C. (1986). Statistical study of air pollutant concentrations via Generalized Gamma distributions. Journal of the Air Pollution Control Association 36(11):1250-1254.

[3] Greenwood, J.A., J.M. Landweehr, N.C. Matalas, and J.R. Wallis. 1979. Probability weighted moments: Definition and relation to parameters of several distributions expressible in inverse form. Water Resources Res. 15: 1049–1054.

[4] Holland, D.M. and T. Fitz-Simons. 1982. Fitting statistical distributions to air quality data by the maximum likelihood method. Atmospheric Environment 16 (5): 1071-1076.

[5] Shoukri, M.M., I.U.H. Mian, and D.S. Tracy. 1988. Sampling properties of estimators of the log-logistic distribution with application to Canadian precipitation data. The Canadian Journal of Statistics 16(3): 223-236.

[6] DiCiccio, T.J. 1987. Approximate Inference for The Generalized Gamma Distribution. Techno metrics 29(1): 33-40.

[7] Landwehr, J.M., N.C. Matalas, and J.R. Wallis. 1979. Probability weighted moments compared with some traditional techniques in estimating Gumbel parameters and quantiles. Water Resources Res. 15: 1055-1064.

[8] Hosking, J.R.M., J.R. Wallis, and E.F. Wood. 1985. Estimation of the generalized extreme-value distribution by the method of probability-weighted moments. Technometrics 27: 251-261.