IOSR Journal of Mathematics (IOSR-JM)
e-ISSN: 2278-5728, p-ISSN:2319-765X. Volume 9, Issue 1 (Nov. – Dec. 2013), PP 11-16
www.iosrjournals.org

Modeling the Distribution of Rainfall Intensity using Quarterly Data

H. G. Dikko[1], I. J. David[1], and H.R. Bakari[2]
[1] *(Department of Mathematics, Ahmadu Bello University, Zaria, Nigeria)*
[2] *(Department of Mathematics and Statistics, University of Maiduguri, Nigeria)*

Abstract: *Agricultural planning, drainage pattern and designs for erosion control all depends on how best the design of water management practices are been conducted. This can be, best achieved if there is a prior knowledge of the distribution pattern of rainfall characteristics. The effect of rainfall to man is diverse ranging from designs of agricultural systems to erosion control. This study is aimed at exploring the statistical distribution namely Gamma distribution on quarterly rainfall amount in Zaria. Quarterly rainfall data were been collected for a period of 38 years (1971 – 2008) from Nigeria Meteorological Agency (NMA) quoted in central bank of Nigeria (CBN) bulletin, the Gamma distribution was used to model the distribution of the quarterly rainfall amount. Kolmogorov – Smirnov, One Sample test was used to evaluate the model fit. The Gamma distribution adequately fit the quarterly rainfall data producing a suitable model base on the Kolmogorov – Smirnov One Sample test. The result could be very useful to agricultural planning, erosion control, etc.*

Keywords: *Chi-square test, Gamma distribution, Likelihood ratio test, quarterly, Rainfall intensity,*

I. Introduction

Water is one of the vital natural resources, which plays an important role in human life sustainability be it agriculture, industry and domestics. Shortage or excessive rainfall could be very harmful as there will be food scarcity and insecurity, water pollution, erosion, telecommunication problems, etc. All of this could lead to economic loss in a country. Therefore, prior knowledge of the distribution of rainfall intensity is important in drainage pattern and design in controlling erosion, agricultural planning, water pollution, etc. the aim of this research is to explore the statistical distribution of quarterly rainfall amount for Zaria – Nigeria. The Gamma distribution is been employed in fitting the model and the Kolmogorov – Smirnov One Sample test will be used to check the goodness of fit of the model. Below are a brief literature on the application of the Gamma distribution and other probability distribution on rainfall data by different scholars. The Gamma distribution is one of the probability distribution used in modeling rainfall data and its performance has been commented on. [1] stated that a study was conducted during 2000 – 2003 to analyze standard meteorological weekly rainfall data of Patna for 42 years (1960 – 2001) using Gamma probability distribution, which has been identified as the best-fit model out of seven competing distributions using Anderson – Darling goodness-of-fit test.

The spatial characteristic of rainfall rate was developed from a rainfall data in a tropical rainfall measuring mission. Gamma and Log-Normal distributions were used to model the rain rate and results indicate that both the models fit well with probability density function (PDF) of rainfall data [2]. Similarly, parametric family of PDFs were derived alongside Gamma and Log-normal for comparing the best fit in modeling rain rate in Darwin and Florida. Result indicates that the Log-Normal distribution out-performs both the gamma and the other family of PDFs [3].

[4] compared four probability distribution namely Generalized Pareto, Exponential, Beta and Gamma distribution in modeling the distribution of rainfall intensity using hourly data for Peninsular Malaysia. The result showed that based on the three goodness-of-fit test (Kolmogorov-Smirnov, Chi-squared and Anderson-Darling) all the models can be used to model the distribution of rainfall intensity. However, General Pareto was found to be the most suitable among the four probability distributions tested.

[5] applied the two parameter model alongside Markov model to describe the distribution of daily rainfall at two sites in Ghana and to predict the occurrence of the rainfall in order to develop a rainfall simulation model; using twenty years of rainfall data; the research confirms the applicability of the Gamma distribution.

[6] in his study on multisite daily rainfall data generation model for climate change conditions applied Gamma distribution on the rainfall amounts model he stated that the rainfall amounts on wet days are generated by using a Gamma distribution because it has been found to fit better than the routinely used distributions, Exponential and Weibull. [7] Presented a method to adapt stochastic daily weather generation models for generation of synthetic daily time series consistent with assumed future climates. The assumed climates were specified by the monthly means and variances of rainfall and temperature. He stated that for a two-part rainfall

model with Gamma distribution for rainfall amounts, there are four parameters (p_{11}, p_{01}, α, β). The shape and scale parameters of the Gamma distribution are denoted by α and β. The transition probabilities (p_{11}, p_{01}) are convenient for Monte Carlo simulation. However, he replaced the unconditional probability of a wet day (π) and a dependence parameter (d):

$\pi = p_{01}/ (1 + P_{01} - p_{11})$

$d = p_{11} - p_{01}$

[8] used eight probability distributions namely; Two parameter Gumbel and Gamma, Three parameter Generalized Normal, Generalized Pareto, Generalized Extreme Value, Pearson type 3, Log-Pearson type 3 and the five parameter Wakeby were tested. Using probability plot correlation coefficient test combined with root mean squared error, relative root mean squared error and maximum absolute deviation and it was concluded that the Generalized extreme value distribution was the most appropriate distribution for describing the annual maximum rainfall series in Malaysia.

II. Materials And Methods

2.1 Data collection and analysis

Zaria situated in Kaduna State, North Central Geo-political Zone/Region of Nigeria. Approximately, Zaria has an annual rainfall amount between 1000mm to 1200mm with annual number of wet days ranging from 60 to 95 days. Quarterly rainfall series data for this study was obtained from the Nigeria Meteorological Agency (NMA) published in Central Bank of Nigeria (CBN). The data cover a period of 38 years (1971 – 2008). There were occasion where rainfall was absent in some quarter of a particular year we discarded those quarters so as to avoid zero's in the data and we used a threshold value of 0.3999 because the lowest value in the data is 0.4 which is appropriate for the data since the lowest data recorded was 0.4. The SAS 9.0 was used for the modeling of the data.

2.2 Distribution of rainfall amount

The Gamma distribution was used to model the distribution of the rainfall amount. The Gamma distribution represents the sum of n exponentially distributed random variables. Both the shape and scale parameters (α, σ) can have non-integer values. Typically, the Gamma distribution is, defined in terms of a scale factor and a shape factor. When used to describe the sum of a series of exponentially distributed variables, the shape factor represents the number of variables and the scale factor is the mean of the exponential distribution $\left(\frac{1}{\sigma}\right)$. This is apparent when the profile of an exponential distribution with mean set to one is compared to a Gamma distribution with a shape factor of one and a mean of one.

2.3 Special cases of the Gamma distribution

When the shape parameter is set to one, and the scale parameter to the mean interval between events, the Gamma distribution simplifies to the exponential. A Chi-squared distribution is a gamma distribution in which the shape parameter set to the degrees of freedom divided by two and the scale parameter set to two [9]. Lastly, the Erlang distribution is been used to model the total interval associated with multiple Poisson events, the shape parameters represents the number of events and the scale parameter the average interval between events.

2.4 Applications of the Gamma distribution

Aside the application of the Gamma distribution in modeling rainfall data it is also applied based on intervals between events, which derive from it being the sum of one or more exponentially distributed variables. In this form, examples of its use include queuing models, the flow of items through manufacturing and distribution processes and the load on web servers and the many and varied forms of telecom exchange. However, due to its moderately skewed profile, it is been applied in modeling financial services where it has been used for modeling insurance claims and the size of loan defaults and as such has been used in probability of ruin and value at risk calculations.

2.5 The probability density function of Gamma distribution

The Gamma distribution with two parameters α and β denote the shape and scale parameters respectively. The Gamma model for daily rainfall amount is been tested with its probability density function given as follows:

$$P(x) = \begin{cases} \frac{1}{\alpha\Gamma(\alpha)} \exp\left(\frac{x-\theta}{\sigma}\right) & for\ all\ x > \theta \\ 0 & for\ all\ x \leq \theta \end{cases} \quad \text{..... (1)}$$

where; x is the random variable representing the daily rainfall amount, θ is the threshold parameter which is less than x, σ is the scale parameter ($\sigma > 0$) and α is the shape parameter ($\alpha > 0$). Since the Gamma distribution is a continuous distribution then the indicator function $I_{(0,\infty)}(x)$ implies that x takes value from zero to infinity and Γ is the Gamma function. The mean and variance of the Gamma distribution is given respectively as,

$$\frac{\alpha}{\sigma} \text{ and } \frac{\alpha}{\sigma^2} \quad \text{..... (2)}$$

Proposition I: If X_i, $i = 1, \ldots, n$ are independent gamma random variables with respective parameters (α_i, σ), then $\sum_{i=i}^{n} X_i$ is gamma with parameters $\sum_{i=1}^{n} \alpha_i, \sigma$. Since the gamma distribution with parameters (1, σ) reduces to the exponential with the rate σ [9]. Hence, proposition I leads to a corollary stated below.

Corollary I: If $X_1, \ldots, X_n$ are independent exponential random variables, each having rate σ, then $\sum_{i=i}^{n} X_i$ is a gamma random variable with parameters (n, σ) [9].

Based on proposition (I) above it worth stating a hypothesis for the study. Hence, the hypothesis to be tested at 5% significance level is as follows.

H_0: the estimated scale parameter value for Zaria Quarterly Rainfall indicates that the fitted gamma distribution model is not adequate.

H_1: the fitted gamma distribution model is an adequate model for Zaria Quarterly Rainfall Data.

2.6 Goodness-of-fit tests

The goodness of fit test employed in this research is the log-likelihood-ratio statistic known as the scaled deviance and the scaled Pearson's chi-square statistic. The log-likelihood can be expressed in terms of the mean parameter μ and the log-likelihood ratio which is the scaled deviance is expressed as:

$$\mathbf{D^*(y; \hat{\mu})} = -2\left(l(\hat{\mu}; y) - l(\hat{\mu}_{\max}; y)\right) \quad \text{..... 3}$$

where, $l(\hat{\mu}; y)$ is the log-likelihood under the model; $l(\hat{\mu}_{\max}; y)$ is the log-likelihood under the maximum achievable (saturated) model.

For generalized linear models, the scaled deviance can be expressed as:

$$D^*(y; \hat{\mu}) = \frac{1}{\phi} D(y; \hat{\mu}) \quad \text{..... 4}$$

where, $D(y; \hat{\mu})$ is the residual deviance for the model and it's the sum of individual deviance contributions and ϕ is the dispersion parameter.

The form of the individual deviance contributions, d_j, for gamma distribution model is:

$$-2 \log\left(\frac{y}{\hat{\mu}}\right) + 2\left(\frac{y - \hat{\mu}}{\hat{\mu}}\right) \quad \text{..... 5}$$

The dispersion parameter ϕ is

$$\phi = \frac{D(y; \hat{\mu})}{(n - p)} \quad \text{..... 6}$$

Over-dispersion is present in a data set if the estimated scaled deviance ($D^*(y; \hat{\mu}) / df$) > 1 and under-dispersed if ($D^*(y; \hat{\mu}) / df$) < 1. If ($D^*(y; \hat{\mu}) / df$) is slightly > 1 or < 1 then the data may contain outliers in the data set. The scaled Pearson's chi-square follows the same approach. The Pearson's chi-square statistic is defined as:

$$\chi^2 = \sum_{i=1}^{n} \frac{w_i (y_i - \mu_i)^2}{V(\mu_i)} \quad \text{..... 7}$$

where y_i is the response value; w_i is the weight which is set to one for all observations; μ_i is the predicted mean value and $V(\mu_i)$ is the variance function for the gamma distribution.

The scaled Pearson's chi-square is $\frac{\chi^2}{\phi}$ **…… 8**

The scaled versions of both statistics, under certain conditions, has a limiting chi-square distribution, with degrees of freedom equal to the number of observations minus the number of estimated parameters and are approximate guide to the goodness of fit of a given model [10].

III. Results And Discussion

Below are the results from the analysis for modeling Zaria quarterly rainfall data from 1971 to 2008. The results to be discussed are the parameter estimates, likelihood ratio (LR) statistics for testing the significance of the quarterly (QT) distributed effect and goodness of fit test.

Table I: Analysis of Parameter Estimates

Parameter	DF	Estimate	SE	Likelihood Ratio 95% Confidence Limits		Chi-Sq	Pr > Chi-Sq
Intercept	1	3.9993	0.128	3.7569	4.263	975.72	< .0001
QT1	1	-1.1978	0.2	-1.6634	-0.6978	24.21	< .0001
QT2	1	1.582	0.1762	1.2323	1.9295	80.58	< .0001
QT3	1	2.5279	0.1762	2.1781	2.8753	205.73	< .0001
QT4	0	0	0	0	0		
Scale	1	1.7942	0.211	1.4139	2.2423		

Note: Scale parameter was estimated by maximum likelihood

Table II: Type III LR Statistics

Source	DF	Chi-Sq	Pr > Chi-Sq
QT	3	152.73	< .0001

Table III: Criteria for Assessing Goodness of Fit

Criterion	DF	Value	Value/DF
Deviance	119	74.7472	0.6281
Scaled Deviance	119	134.1117	1.127
Pearson Chi-Sq	119	57.0986	0.4798
Scaled Chi-Sq	119	102.4465	0.8609
Log Likelihood		-744.942	

**Algorithm Converged*

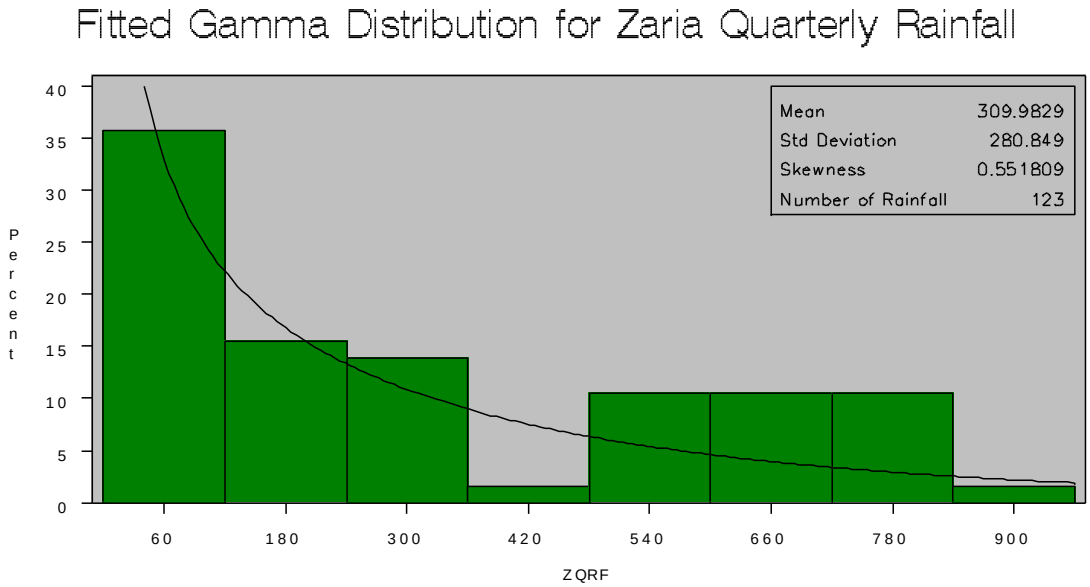

Figure I: Fitted Gamma Distribution For Zaria Quarterly Rainfall

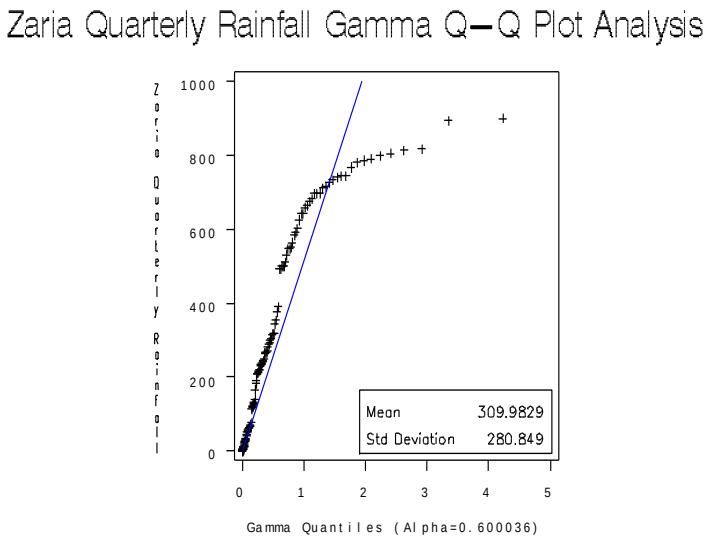

Figure Ia: Fitted Gamma Q-Q Plot for Zaria Quarterly Rainfall

From TABLE I above it can be observed that the intercept estimate of 3.9993 is highly significant base on the chi-square (χ^2) test p-value of 0.0001 which is less than α = 0.05 significant level. The intercept which is the estimated log mean of the fitted gamma distribution, so that the mean Zaria quarterly rainfall will be

μ = exp (Intercept)
= exp (3.9993)
= 54.56mm

A confidence interval for the mean Zaria quarterly rainfall is

(*exp* (3.757), *exp* (4.263)) = (42.82, 71.02)

The four quarters that is Q1, Q2, Q3 and Q4 rainfall are also highly significant to the model based on their χ^2 - test where their respective p-values of 0.0001 are less than α = 0.05 significant level. The zero value for the Q4 implies that it is redundant.

TABLE II presented the likelihood ratio test to determine if there is a significant difference between the four Zaria quarterly rainfalls that is Q1, Q2, Q3 and Q4. The p-value of the χ^2 - test (0.0001) is less than α = 0.05 significant level. Hence, it is concluded that there is a highly significant difference between the four quarters rainfall of Zaria.

The scale parameter which is the inverse of the gamma dispersion parameter and sometimes called the gamma index parameter have a value of 1.7942 ≠ 1 and the 95% profile likelihood confidence interval for the scale parameter is (1.4139, 2.2423) which none of the values are equals to 1. Therefore, our null hypothesis (H_0) of the estimated scale parameter value for Zaria Quarterly Rainfall indicates that the fitted gamma distribution model is not adequate is not accepted. Hence, we concluded that the fitted gamma distribution model is an adequate model for Zaria Quarterly Rainfall Data.

TABLE III of the presented result showed the goodness of fit criteria. The scaled deviance value divided by the associated *d.f* of 1.13 is not far from 1 meaning that the fitted model is adequate the difference in the value that is 0.13 could be due to outliers in the data set. Likewise, the scaled deviance value of 134.11 is also not too far from the associated *d.f* value of 119. The scaled deviance result for assessing the fitted model conforms to the result of the scaled Pearson's χ^2 value divided by the associated *d.f* of 0.8609 which is close to one, likewise, the scaled Pearson's χ^2 value of 102.45 is also not too far from the associated *d.f* value of 199. Therefore, the fitted gamma distribution model for Zaria quarterly rainfall is adequate.

While Fig. I showed the fitted quarterly rainfall data for Zaria and it can be observed from the plot that the distribution fitting on the histogram bars showed that the data modeled follows a gamma distribution. The skewness value of 0.55 is an evidence to say that Zaria quarterly rainfall is well modeled using a gamma distribution because the value is far less than 1 and not far from zero. Likewise, Fig. II which shows the gamma Quantile-Quantile plot shows that the data follows a gamma distribution.

IV. Conclusion

Results obtained from the analysis draws to the conclusion that the gamma distribution can be used to model the quarterly rainfall of Zaria adequately. The goodness of fit test showed the adequacy of the gamma

probability distribution in modeling the data where the scaled deviance and Pearson's chi-square values (1.13 and 0.8609 respectively) which are approximately one. Based on these findings, it is recommended that the gamma distribution could be used in describing the quarterly rainfall intensity.

References

[1] P. K.Das, N. Subash, , A. K. Sikka, V. N.Sharda, and N. K. Sharma, Modeling weekly rainfall using gamma probability distribution and Markov chain for crop planning in a sub-humid (dry) climate of central Biar. *Indian Journal of Agricultural Science*, 76, 2006, 358-361.

[2] H. K. Cho, K. P. Bowman and G. R. North, A comparison of Gamma and Log-normal distributions for characterizing satellite rain rates from the tropical rainfall measuring mission, *Journal of Applied meteorology*., 43(165) 2004,1586-1597

[3] B. Kedem, H .Pavlopoulos, X. Guan, and D. A. Short, Probability distribution model for Rain rate, *Journal of Applied Meteorology*, *33*, 1994, 486-1493, 1520-0450.

[4] D. Salisu, S. Supiah, and A. Azmi, Modeling the Distribution of Rainfall Intensity Using Hourly Data. *American Journal of Environmental Sciences*, 6(3), 2010, 238-243.

[5] S. G. K. Adiku,, P. W. A. Dayananda, C. W. Rose and G. N. N. Dowuona, Analysis of within-season rainfall characteristics and simulation of daily rainfall in two Savannah Zones in Ghana, *Agric. Forest meteorology*., *86(96)* 1997, 51-62, 168-1923,14-8.

[6] R. Srikanthan, A multisite daily rainfall data generation model for climate change conditions, *18th World IMACS/MODSIM Congress, Cairns, Australia 13-17 July*. 2009. *http://mssanz.org.au/modsim09*

[7] D. S.Wilks, Adopting stochastic weather generation algorithms for climate change studies, *Climate Change, 22*, 1992, 67-84.

[8] M. D Zalina, M. N. M. Desa, V. V. Nguyen and A. H. M. Kassim, Selecting a probability distribution for extreme rainfall series in Malaysia. *Water Sci. Techno., 45*, 63-68, 2002).

[9] S. M. Ross, *Introduction to Probability and Statistics for Engineers and Scientists*. Third Edition, (London: Elsevier Academic Press 2004).

[10] P. McCullagh, and J. A. Nelder, *Generalized Linear Models*, Second edition, (London: Chapman and Hall, 1989).