SIMULTANEOUS MODELING OF RAINFALL QUANTITY AND INCIDENCE IN PANGASINAN USING POISSON-GAMMA DISTRIBUTION

Jennifer R. Miguel and Marycel T. Sajise
Pangasinan State University, Philippines
Benguet State University, Philippines
jhenrah_06@yahoo.com
marycelsajise@gmail.com

This study aimed to fit a model based from the Poisson-Gamma Distribution for simultaneous data of rainfall quantity and incidence. The characteristics of the data was also investigated using descriptive statistics. The rainfall data from year 2000 to 2016 of Dagupan City was utilized. Dry months were observed from November to April while wet months from May to October of which August recorded the maximum rain. Sine and cosine terms were used as predictors. A single model was derived in predicting the monthly rainfall quantity. Reparametrization of the parameters μ, φ and β was adopted and found to predict the probability of zero rainfall well for all months except for the months of January to April. The model was considered valid though a bootstrapping procedure.

INTRODUCTION

Teaching statistics in a research based method is highly encouraged to enable student learning develop flexible understanding, including both basic factual and conceptual knowledge, and must know how to use that knowledge critically. Rainfall modeling is one of the fields where statistics can be applied. This study tried to explore a non-traditional statistical distribution from the Tweedie family of distributions, particularly the Poisson-Gamma Distribution, that can simultaneously model the probability of rainfall quantity which is continuous containing precise zero outcome (i.e., days or a month without a rainfall quantity is recorded) and rainfall incidence which is discrete because it has only two states: precipitation occurred and no precipitation occurred. It aimed to widen perspectives, insights, and knowledge in the field of applied statistics and its application to meteorology.

This study attempted to apply the method used by Dunn (2004) in modelling simultaneously the rainfall quantity and the probability of precipitation occurring. Dunn (2004) who effectively proved in his study that the rainfall incidence and quantity can be modeled simultaneously using Poisson-Gamma Distributions below the class distributions known as Tweedie family of distributions. This study served as a pioneer research in the Philippines about simultaneous modeling of rainfall incidence and quantity using non-traditional statistical method under the Tweedie Model. This study also served as a guide and blueprint in making the same research paper on other regions in the Philippines.

METHODS

The complete monthly rainfall data were gathered from Philippine Atmospheric Geophysical and Astronomical Services Administration (PAGASA) Dagupan City Synoptic Office Philippines. The scope of the data that was utilized is from the year 2000 to 2016. The modeling of incidence and quantity of rainfall is only limited in the use of Poisson-Gamma Distribution.

The predictors or covariates that would be used were the sine and cosine terms. Hasan & Dunn (2011) and Swan (2006) used sine and cosine terms and explained that using these two as covariates is simple yet performs well. The sine and cosine terms are in the forms of $\sin(\frac{2\pi m}{12})$ *and* $\cos(\frac{2\pi m}{12})$ where in m=1,2,3,4,…,12 are the months in a year.

Descriptive statistics like boxplot, summary statistics and mean-variance relationship were employed. The mean-variance relationship test was done by computing and taking the log of the mean and variance of rainfall amount each month and plotting the resulting values.

To determine the model based from Poisson-Gamma Distribution used in model fitting, parameter estimates were determined using the Tweedie distribution. Fitting the model to the data

In M. A. Sorto, A. White, & L. Guyot (Eds.), *Looking back, looking forward. Proceedings of the Tenth International Conference on Teaching Statistics (ICOTS10, July, 2018), Kyoto, Japan.* Voorburg, The Netherlands: International Statistical Institute. iase-web.org

required the estimation of many parameters such β, φ and p. The main key of this estimation process was to estimate the index parameter p by obtaining the maximum likelihood estimates of β and φ and compute the log-likelihood. This process was repeated for a range of p values. Forecasted values were generated using the derived models. Model validation was performed using Bootstrapping Technique.

RESULTS

The boxplot of the monthly rainfall distribution of Dagupan City from 2000-2016 (Figure 1) was employed to fully understand the data. The monthly rainfall data in Pangasinan from 2000-2016 revealed that it has wet months from May to October and dry months from November to April.

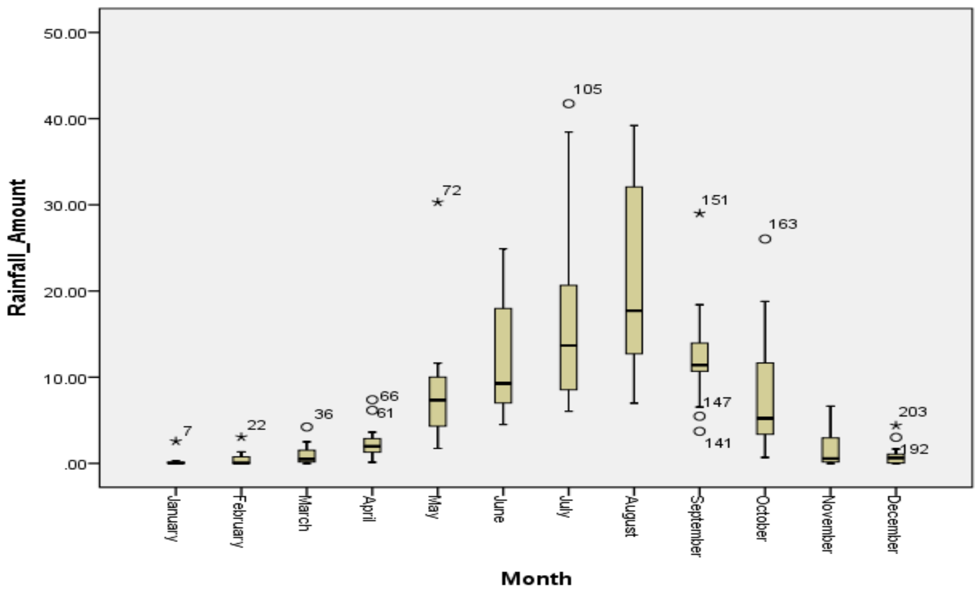

Figure 1. Boxplot of the monthly rainfall distribution of Pangasinan

The mean-variance relationship was computed and the result shows an approximate linear relationship between the group means and group variances. This means that the variance is approximately proportional to some power of p by mean.

The researcher derived the formula for the forecasted mean rainfall quantity for month m as: $\hat{\mu}_i = \exp\left\{\beta_0 + \beta_1 \sin\left(\frac{2\pi m}{12}\right) + \beta_2 \cos\left(\frac{2\pi m}{12}\right)\right\}$

Using the sine and cosine predictors, the $\hat{p} = 1.59$ was utilized. The significance of the computed coefficients were tested using the Wald Chi-Square test and it shows that both the sine and cosine terms have significant coefficients. They are all significant at α =0.05 based on the standard asymptotic Wald Z-test.The regression coefficients and the dispersion parameters is presented in Table 1. It was observed that there were little differences between the two datasets.

Table 1. Estimated values of parameters

Station	Dataset	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\varphi}$
Dagupan	Overall	1.257*	-1.319*	-1.352*	1.227
City	Estimation	1.282*	-1.229*	-1.284*	1.336

*significant at $\alpha = .05$

Table 1 includes the estimated values of the coefficients and the dispersion parameter φ for the predictors in the fitted rainfall model. Using the estimated model, the mean rainfall quantity for specific month was derived. Figure 2 shows the mean rainfall quantity between the observed (blue line) and predicted (red line). Based from the computed values, it was concluded that the performance of the model is good in predicting the future amount considering the differences in value is small and its corresponding confidence interval.

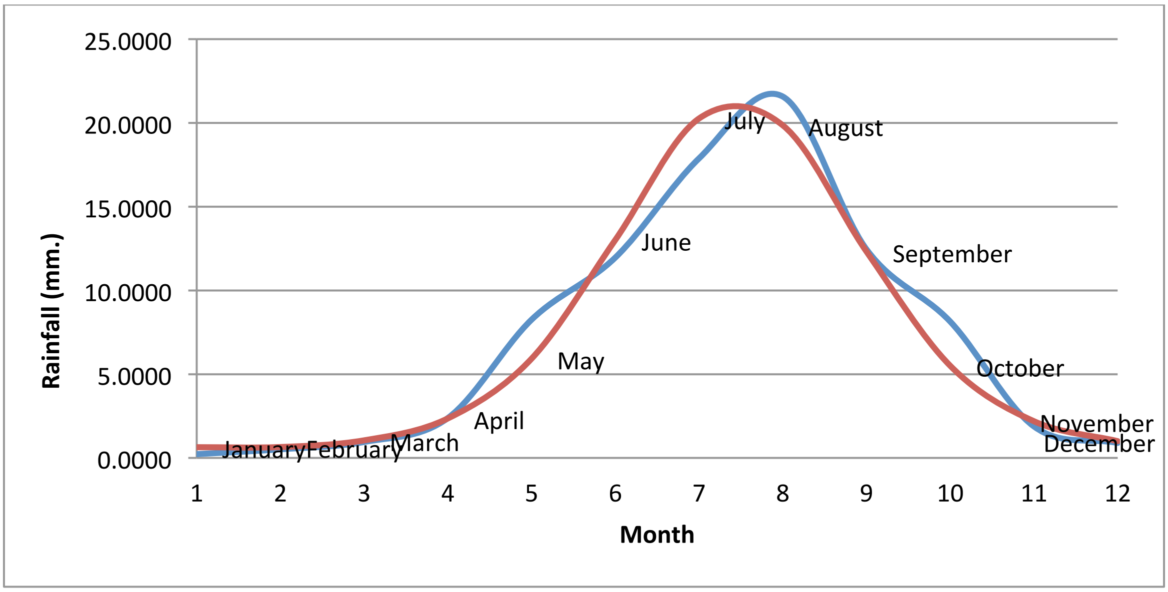

Figure 2. Predicted versus actual observed means

Figure 3 also shows the plotted forecasted and actual observed probability of no rain occurrence The computed value for the estimated dispersion parameter used is $\hat{\varphi} = 1.3360$. Using the reparameterized distribution parameters suggested by Dunn (2004), the mean number of rainfall events (λ), the shape of the rainfall Gamma distribution (γ), and quantity of rain per rainfall event $\alpha\gamma$, were derived. Based from (Dunn & Smyth, 2005), the probability of recording no rain is $\pi_0 = \Pr(Y = 0) = e^{-\lambda}$, where $\lambda = \frac{\mu^{2-p}}{\varphi\,(2-p)}$ was utilized in computing the predicted probability of zero rainfall occurrence. It is shown that these two follow the same path; thus making the model performs well in predicting zero rainfall occurrence (except for the months of January to April).

It was concluded that that the model adopted and used in predicting zero rainfall performs good for the rainfall data in Pangasinan except for the months of January to April (see Figure 3). The importance of this model in predicting the probability of zero rainfall occurrence per month can be reflected in the field of agriculture specifically in the shifting of cropping time or alternative cropping.

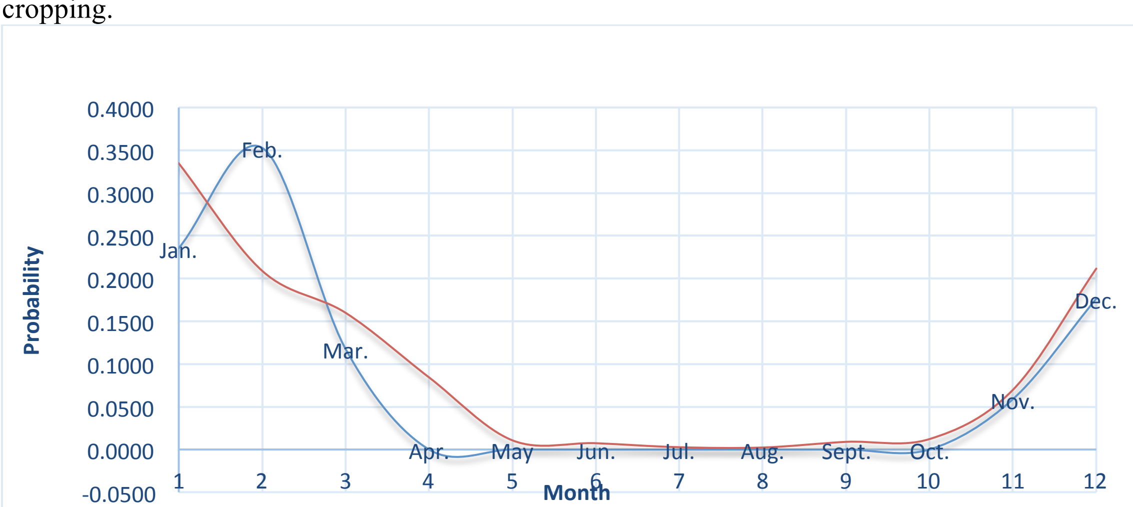

Figure 3. Predicted versus actual observed probability of no rain

To test the validity of the model, bootstrapping technique was conducted. It was found that the estimators produce negative bias except for the sine term. This indicates that the intercept, cosine term and scale estimators are slightly underestimated for they produced values greater than -0.025. The sine term is unbiased for it yielded zero value for bias. These estimators can still be considered as able and good in representing the population parameters.

DISCUSSIONS

Maximum amount of rain was recorded for the month of August as presented in Figure 1. Dry months are from November to April while wet months from May to October. From the data, Pangasinan is considered Type I Climate according to the Climate classification of PAGASA. The mean of the overall dataset was is computed to be 7.2609 mm. with median equal to 3.4710 mm. The percentage of months with no rainfall based from the overall dataset is 7.8%. The mean-variance results implies that variance of the rainfall quantity is dependent on the mean showing an approximate linear relationship of $y = 0.27 + 1.35^{x}$.

The formula for the forecasted mean rainfall quantity for month m is $\hat{\mu}_i = \exp\left\{\beta_0 + \beta_1 \sin\left(\frac{2\pi m}{12}\right) + \beta_2 \cos\left(\frac{2\pi m}{12}\right)\right\}$. Computed values are the following: $\widehat{\beta_0} = 1.282$; $\widehat{\beta_1} = -1.229$; and $\widehat{\beta_2} = -1.284$. The model used to determine the occurrence of zero rainfall per month is π_0 = Pr(Y = 0) =$e^{-\widehat{\lambda}}$, where λ is the mean number of rainfall event.

Further, based from the computed values after utilizing $\widehat{\mu}_i = \exp\left\{\beta_0 + \beta_1 \sin\left(\frac{2\pi m}{12}\right) + \beta_2 \cos\left(\frac{2\pi m}{12}\right)\right\}$ to predict mean rainfall quantity, it is concluded that the performance of the model is good because it resembles the actual observed data. After using π_0 = Pr(Y = 0) =$e^{-\widehat{\lambda}}$ to determine the occurrence of zero rainfall per month, it is concluded that the model adopted and used in predicting zero rainfall performs good for the rainfall data in Pangasinan except for the months of January to April. Both the standard error of the estimates for before and after bootstrapping process produce less than 0.2 standard errors which is pretty small. This means that the ability of the model to make rainfall forecasts is good. Lastly, the bootstrapping techniques further validate the capacity of the model in predicting the amount of rainfall in Pangasinan.

Based from these results that were observed and recorded, it is recommended that future studies can be done using daily data instead of monthly data though it will take so much time and effort or use different timescale aside from daily or monthly. Also, future researchers can consider fitting the Poisson-Gamma distribution in other rainfall data aside from Pangasinan (e.g. rainfall data from PAGASA Baguio and PAGASA Science City of Muñoz) and compare it with the current results. To further improve the model, other predictors in forecasting the mean rainfall amount can be used aside from sine and cosine terms.

LITERATURE CITED

Brown, R. B. (2006). *Doing your dissertation in business and management: the reality of research and writing*. Sage Publications.

Dunn, P.K. (2004). *Occurrence and quantity of precipitation can be modeled simultaneously*. Wiley Interscience: www.interscience.wiley.com. Accessed on July 7, 2017.

Hasan, M.M. and P.K. Dunn. (2010). *A simple poisson-gamma model for modeling rainfall occurrence and amount simultaneously*. Agricultural and Forest Meteorology,150:1319-1330.

Hasan, M.M. and P.K. Dunn. (2011). *Two tweedie that are near-optimal for modeling monthly rainfall in australia*. International Journal of Climatology, 31:1389-97.

Swan, T. (2006). *Generalized estimating equations when the response variable has a tweedie distribution: an application of multi-site rainfall modeling*. The University of Southern Queensland: Toowoomba, Queensland, Australia.