Journal of Computer Science & Computational Mathematics, Volume 5, Issue 1, March 2015
DOI: 10.20967/jcscm.2015.01.002

Independent Gamma Variables of McKay Distribution for Rainfall Model

R. Zakaria*, N. H. Moslim

Faculty of Industrial Sciences & Technology, Universiti Malaysia Pahang, Lebuhraya Tun Razak, 26300 Gambang, Kuantan Pahang, Malaysia,
*Corresponding author email: roslinazairimah@ump.edu.my

***Abstract*:** The application of mathematical models of rainfall is crucial in order to have a better understanding in terms of rainfall characteristics. The rainfall models have been used widely to improve water management, to construct hydrological structures and as an input in climatological studies. In this study, the gamma distribution is used to fit the marginals of monthly rainfall data. The parameters of the gamma distribution, shape and scale, are estimated using the maximum likelihood estimation method. Then, the sum of monthly rainfall amounts is modelled using the McKay distribution for independent gamma variables. To illustrate the independent example, rainfall data is used between the months in monsoon season for Kuantan station. The fit between the observed and generated rainfall data is assessed using Kolmogorov-Smirnov goodness of fit test. The results show that the McKay distribution is suitable to model the sum of monthly rainfall amounts for two independent gamma variables. The model is also useful for generating synthetic rainfall amounts for independent months in simulation studies.

***Keywords*:** McKay distribution, rainfall model, independent gamma variables**.**

1. Introduction

The research on the rainfall models has been increasing rapidly to improve water management, to construct hydrological structures and to complement climatological studies. The application of mathematical models of rainfall is crucial in order to have a better understanding in terms of rainfall characteristics. In general, the rainfall model of the study will either be based on sequences of dry and wet days or the rainfall amount on the wet days. A Markov chain model has been used extensively to model rainfall occurrences [1, 2, 3, 4, 5], and a two-parameter gamma distribution is frequently applied to model rainfall amount [6, 7, 8]. The aggregation of rainfall amounts is also important for water management. Zakaria [9, 10] uses two types of distribution related to gamma distribution to form summation of rainfall amounts. The first type is the sum of *n* gamma variables which has been introduced by Alouini et al. [11] in the telecommunication field. The latter type is the sum of gamma variables based on the McKay distribution. In a different study, Moschopoulos [12] introduces the sum of independent gamma variables as a single gamma series with different parameters. In this study, the sum of two gamma variables based on the McKay distribution is applied to model the sum of rainfall amounts for independent gamma variables. This model is also able to generate synthetic rainfall amounts.

2. Area of study

The Peninsular Malaysia experienced two significant seasons, wet and dry, respectively. The wet season is experienced by the northeast monsoon (October-March) whereas the dry season by the southwest monsoon (May-August). Monthly rainfall data used in this study have been obtained from the Malaysian Meteorological Department for the period of 30 years (1971 – 2000). A meteorological station on the East Coast of Peninsular Malaysia is selected for the purpose of illustration of the methodology. The months within southwest monsoon are used for Kuantan station (Lat. 3°47 N, Lon. 103°13 E).

3. Modelling the sum of monthly rainfall data

This study shows how to fit the sum of monthly rainfall amounts by using a model based on McKay distribution. An application of the temporal case will be presented by fitting the sum of monthly rainfall amounts to independent gamma variables between two months. The model is used whenever the correlation between the months is not significant. The procedures for the data analysis are as follows: (1) select the positive pairs (wet,wet) (2) calculate the correlation coefficient (3) fit each data set to gamma distribution and estimate the parameters (4) find the average of the shape parameters and use it to re-estimate the value of the scale parameters (5) calculate the sum of the associated pairs and fit them to McKay distribution for independent gamma variables (6) compare the fit between the observed sum and the generated sum and show the graph of PDF and CDF for both.

3.1 Probability density function of gamma variables data

The probability density function (PDF) of gamma variables, X with shape and scale parameters which is written as $X \sim G(\alpha, \beta)$ is given by

$$f(x;\alpha,\beta) = \left(x^{\alpha-1} \exp\left(-x/\beta \right) \right) \Big/ \left(\Gamma(\alpha) \beta^{\alpha} \right) \tag{1}$$

for $x > 0$ and $\alpha, \beta > 0$. Note that, X represents the random variable of monthly rainfall amounts whereas the shape parameter controls the shape of the rainfall distribution and the scale parameter determines the variation of the rainfall amount data. For convenience, the maximum likelihood estimation method (MLE) is used for parameter estimation. The MLE method determines a set of parameters which maximise the likelihood function. Then, the parameters are obtained by differentiating the log likelihood function with respect to the parameters of the distribution. The logarithm of the likelihood function is as follows

$$\ln L = -N\ln\Gamma(\alpha) - N\alpha\ln\beta + (\alpha-1)\sum \ln x - \left(\sum x/\beta\right)$$

In many cases, the rainfall data has skewed distribution. Therefore, the Spearman's rank correlation is used to accommodate the non-normal data. The Spearman's rank correlation coefficient is denoted by ρ_{ij} of X_i and X_j with the formula of

$$\rho_{ij} = 1 - \left[\left(6\sum_{k=1}^{n} d_{ijk}^2\right) \Big/ \left(n\left(n^2-1\right)\right)\right]$$

where the difference between the ranked pairs $R(x_i, k)$ and $R(x_j, k)$ is denoted by $d_{ijk} = R(x_{i,k}) - R(x_{j,k})$ and n is the number of pairs. This study limits the sum to only two random variables since the model will become more complex for higher numbers of months, refer to [9]. The bivariate McKay distribution is considered tractable in terms of moment generating function and cumulants [13]. There are two forms for the McKay distribution. The focus will be on the form used for the bivariate gamma distribution. Holm and Alouini [13] derive general formulae for the probability density function (PDF) of the sum and the difference of two correlated gamma variables in the form of the bivariate McKay distribution. They define the sum form as type I McKay distribution and the difference form as type II McKay distribution. This study considers type I McKay distribution for the sum of gamma variables and it will be shown that the type I McKay distribution is suitable for modelling rainfall amounts.

3.2 Probability density function for the sum of independent gamma variables

Consider a set of N independent gamma variables, $\{X_i\}_{i=1}^{N}$ with parameters α and β_i written as $X_i \sim G(\alpha, \beta_i)$ where the PDF of $\{X_i\}_{i=1}^{N}$ is given by equation (1). The type I McKay distribution is first defined for the sum of independent gamma variables where $\Sigma = \sum_{i=1}^{n} X_i$ and the probability density function (PDF) is given by

$$f_\Sigma(\sigma) = \frac{\sqrt{\pi}}{\Gamma\left(a+\frac{1}{2}\right)} \frac{\left(c^2-1\right)^{a+\frac{1}{2}}}{2^a b^{a+1}} \sigma^a \exp\left(-\frac{\sigma c}{b}\right) I_a\left(\frac{\sigma}{b}\right) \quad (2)$$

where a, b, c are real parameters with $a > -1/2$, $b > 0$ and $c > 1$. The gamma function and modified Bessel function of the first kind of order a are denoted by $\Gamma(\cdot)$ and $I_a(\cdot)$, respectively. The parameters a, b and c can be expressed in terms of the parameters α and β for the sum of two independent gamma variables given in type I McKay distribution as

$a = \alpha - \frac{1}{2}$, $b = 2\beta_1\beta_2/|\beta_1-\beta_2|$ and $c = (\beta_1+\beta_2)/|\beta_1-\beta_2|$.

For further details, the reader is referred to [13] and [14].

3.3 Kolmogorov-Smirnov goodness of fit test

The Kolmogorov-Smirnov goodness of fit test is used to compare the cumulative distributions between the observed data, $F_1(x)$ and generated data $F_2(x)$. The cumulative distribution function of a random variable X is defined as $F(x) = P(X \le x)$. The hypotheses of the two-sample Kolmogorov–Smirnov test are

$H_0: F_1(x) = F_2(x)$ and $H_1: F_1(x) \ne F_2(x)$

and the test statistic is

$$D_{1,2} = \max_x \left|F_1(x) - F_2(x)\right| \quad (4)$$

where $F_1(x)$ and $F_2(x)$ are the cumulative distribution functions of the observed and generated data, respectively. The value of D determines the absolute maximum difference between the CDF of the observed and generated data. If the P-value is greater than the significance level of $\alpha = 0.05$, it indicates that the two distributions are not significantly different from each other.

4. Results and discussion

Two months within the southwest monsoon of Kuantan station are selected, namely May and June from 1971–2000. The Spearman's correlation between the months is calculated to be 0.12 and it is considered as independent. Then, the sum of the independent gamma variables is used to illustrate the temporal application. Firstly, the selected positive pairs (wet,wet) of the data will be fitted marginally to the gamma distribution. The estimated parameters are given in Table 1, in which the comparison between the mean and variance of the observed data and by using the formula are presented. It is observed that the means matched perfectly, however the variance demonstrates a discrepancy which may be due to the small correlation.

Table 1: Shape and scale parameters, means and variances of observed and estimated values

Data	Parameter		Mean (mm)		Variance	
	α	β	observed	$\alpha\beta$	observed	$\alpha\beta^2$
May	4.79	35.99	172.40	172.40	5434.57	6206.34
June	4.38	37.45	164.10	164.10	6580.34	6145.27

In the PDF of independent gamma variables (2), it requires a single value of shape parameter (α) with distinct values of β . Hence, the average of α , $\bar{\alpha} = (4.79+4.38)/2 = 4.59$ is used to re-estimate the new scale parameters denoted by $\beta_1^* = 37.60$ and $\beta_2^* = 35.79$. The synthetic data of the sum is generated using the PDF of independent gamma variables with parameters $\bar{\alpha}$, β_1^* and β_2^*. The statistics of the observed and generated sum of rainfall amounts are given in Table 2.

Table 2: Means and variances for the sum of two months of observed and generated rainfall amounts

Data	**Mean**	**Variance**
Observed	336.5	12615.8
Generated	333.4	11735.4

The Kolmogorov-Smirnov goodness of fit test is used to compare the distribution of the observed and generated sum. Based on the Kolmogorov-Smirnov goodness of fit test, the $(\text{P-value} = 0.8105) > (\alpha = 0.05)$ shows that the observed sum and the generated sum are not significantly different at the 5% significance level. Figure 1 and Figure 2 show plots of PDF and CDF of the observed and generated data from the sum of two independent gamma variables and the plots support the statistical results.

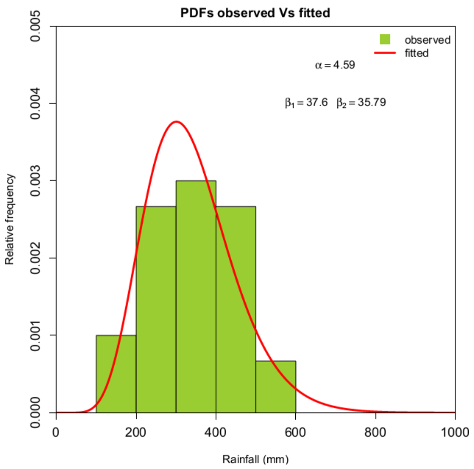

Figure 1: PDF plots of observed vs generated for the sum of independent gamma variables

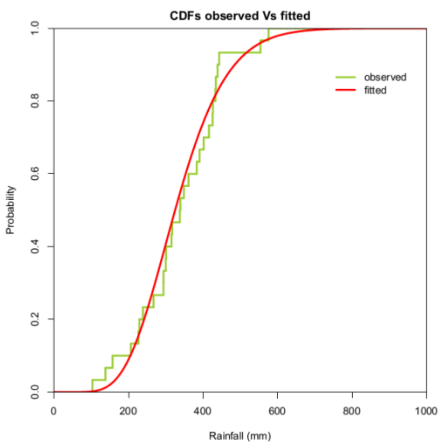

Figure 2: CDF plots of observed vs generated for the sum of independent gamma variables

5. Conclusion

The sum of two independent gamma variables based on the McKay distribution is applied to model the sum of monthly rainfall amounts for two consecutive months whenever the correlation between them is insignificant. It is also highlighted that the model is easy to implement for the two independent variables. The average α is used since the model required a single α and their values were not much different. In conclusion, based on McKay distribution the model of sum of two independent gamma variables is suitable in modelling rainfall amounts for random variables with insignificant correlation. The McKay distribution has also been extended to model correlated variables by including extra parameter, namely correlation parameter.

Acknowledgements

The authors are indebted to the Malaysian Meteorological Department for providing the rainfall data. This study was supported by the Universiti Malaysia Pahang under the RDU120101 grant. Cordial gratitude is also extended to an anonymous reviewer who considerably improved the paper.

References

[1] K. R. Gabriel, J. Neumann, “A Markov chain model for daily rainfall occurrence at Tel Aviv,” *Quarterly Journal of the Royal Meteorological Society*, vol. 88, no. 375, pp. 90–95, 1962.

[2] E. Chin, “Modelling daily precipitation occurrence process in Markov chain,” *Water Resources Research*, vol. 13, no. 6, pp. 949–956, 1977.

[3] R. Stern, R. Coe, “Fitting models to daily rainfall data,” *Journal of Applied Meteorology*, vol. 21, pp. 1024–1031, 1982.

[4] S. Deni, A. Jemain, K. Ibrahim, “Fitting optimum order of Markov chain models for daily rainfall occurrences in Peninsular Malaysia,” *Theoretical and Applied Climatology*, vol. 97, no. 1, pp. 109–121, 2009.

[5] P. Dash, “A Markov chain modelling of daily precipitation occurrences of Odisha,” *International Journal of Advanced Computer and Mathematical Sciences*, vol. 3, no. 4**,** pp. 482–486, 2012.

[6] D. Woolhiser, J. Roldan, “Seasonal and regional variability of parameters for stochastic dailly precipitation models: South Dakota, USA,” *Water Resources Research*, vol. 22, no. 6, pp. 965–978, 1982.

[7] S. Jamaludin, A. Jemain, “Fitting daily rainfall amount in Peninsular Malaysia using several types of exponential distributions,” *Journal of Applied Sciences Research*, vol. 3, no. 10**,** pp. 1027–1036, 2007.

[8] G. Husak, J. Michaelsen, C. Funk, “Use of the gamma distribution to represent monthly rainfall in Africa for drought monitoring applications,” *International Journal of Climatology*, vol. 27, no. 7, pp. 935–944, 2007.

[9] R. Zakaria, *Mathematical modelling of rainfall in the Murray-Darling Basin*. Ph.D. thesis, University of South Australia, 2011.

[10] R. Zakaria, P. G. Howlett, J. Piantadosi, J. W. Boland, "Rainfall modelling using the sum of independent gamma variables," *American Institute of Physics Conference Series*, eds. B. Ari-Wahjoedi, R. Razali & M. Narahari, volume 1482 of *American Institute of Physics Conference Series*, vol. 1482, no. 1, pp. 369–372, 2012.

[11] M. Alouini, A. Abdi, M. Kaveh, "Sum of gamma variates and performance of wireless communication systems over Nakagami-fading channels," *IEEE Trans Veh Technol*, vol. 50, no. 6, pp. 1471–1480, 2001.

[12] P. G. Moschopoulos, "The distribution of the sum of independent gamma random variables," *Ann. Inst. Statist. Math.–Part A*, vol. 37, no. 1, pp. 541–544, 1985.

[13] H. Holm, M. S. Alouini, "Sum and difference of two squared correlated Nakagami variates in connection with the McKay distribution," *IEEE Transactions On Communications*, vol. 52, no. 8**,** pp. 1367–1376, 2004.

[14] I. Gradshteyn, I. Ryzhik., *Table of Integrals, Series and Products.* Academic Press: San Diego, CA, 6th edition, 2000.