Independent Mixed-gamma Variables for Modelling Rainfall

ROSLINAZAIRIMAH ZAKARIA
Universiti Malaysia Pahang
Faculty of Industrial Sciences & Technology
Lebuhraya Tun Razak, 26300 Gambang
MALAYSIA
roslinazairimah@ump.edu.my

NOR HAFIZAH MOSLIM
Universiti Malaysia Pahang
Faculty of Industrial Sciences & Technology
Lebuhraya Tun Razak, 26300 Gambang
MALAYSIA
fizahm@ump.edu.my

Abstract: Understanding the rainfall process and characteristics are crucial to the efficient design of flood mitigation and construction of crop growth models. Modelling rainfall is not limited to fit the historical data to a suitable distribution but the model should be able to generate synthetic rainfall data. In this study, we derive sets of formulae of mean and variance for the sum of two and three independent mixed-gamma variables, respectively. Firstly, the positive data is fitted to gamma model marginally and the shape and scale parameters are estimated using the maximum likelihood estimation method. Then, the mixed-gamma model is defined to include zero and positive data. The formulae of mean and variance for the sum of two and three independent mixed-gamma variables are derived and tested using the daily rainfall totals from Pooraka station in South Australia for the period of 1901-1990. The results demonstrate that the values of generated mean and using formula are close to the observed mean. However, the values of the variance are sometimes over-estimated or under-estimated of the observed values. The observed variance is lower possibly due to correlation between the experimental data, that have not been included in the mixed-gamma models. The Kolmogorov–Smirnov and Anderson–Darling goodness of fit tests are used to assess the fit between the observed sum and the generated sum of independent mixed-gamma variables. In both cases, the observed sum is not significantly different from the generated sum of independent mixed-gamma model at 5% significance level. This methodology and formulae derived can be applied to find the sum of more than three independent mixed-gamma variables and the general form of the formulae can be derived.

Key–Words: Gamma variable, mixed-gamma variables, rainfall model, independent variables, synthetic data

1 Introduction

Extensive studies of many aspects of water related issues have been conducted throughout the world. Hydrology and climatology are particular areas of study that use rainfall as their input analysis. Hydrology concerns modelling water catchment, rainfall-runoff and storm water management. Climatology focuses on modelling for climate change and weather forecasting.

Various models have been used to model rainfall such as Markov model for modelling rainfall occurrence [2, 9, 4, 6, 13] and gamma distribution to model rainfall totals [1, 8, 10, 11, 12, 14, 15, 17, 18]. A gamma distribution has been used widely as the first choice to model rainfall totals due to its simplicity since it has only two parameters, shape and scale. Other statistical distributions are also employed to model rainfall totals such as exponential and mixed exponential distributions by Woolhiser and Roldan [5] and Weibull distribution by Sharda and Das [16]. However, the gamma distribution is only suitable for positive data.

A mixed-gamma distribution is an extended form of gamma distribution which is able to accommodate both zero and positive data. Rosenberg et al. [15] use a different type of mixed-gamma distribution associated with Laguerre polynomials whereas Piantadosi et al. [17] use the mixed-gamma distribution to generate synthetic rainfall totals on various timescale including daily, monthly and yearly. In this study, we apply the mixed-gamma distribution to derive the formulae of mean and variance for the sum of two and three independent mixed-gamma variables. We also generate the synthetic rainfall totals using the mixed-gamma distribution and assess the goodness of fit between the observed and generated rainfall totals.

2 Modelling the sum of independent monthly rainfall totals

Firstly, the rainfall totals is fitted marginally using gamma distribution and the parameters (shape and scale) of gamma distribution are estimated using maximum likelihood estimation (MLE) method. Then, the

ISBN: 978-1-61804-281-1

mixed-gamma distribution is used to model the sum of independent rainfall totals. The gamma and mixed-gamma distributions are presented as follows.

2.1 Gamma distribution

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

$$f(x;\alpha,\beta) = \frac{x^{\alpha-1}e^{-\frac{x}{\beta}}}{\Gamma(\alpha)\beta^{\alpha}} \tag{1}$$

for $x > 0$ and $\alpha, \beta > 0$. Note that, X represents the random variable of monthly rainfall totals whereas the shape parameter controls the shape of the rainfall distribution and the scale parameters determines the variation of the rainfall 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

$$\begin{aligned} \ln L &= -N\ln\Gamma(\alpha) - N\alpha\ln\beta \\ &\quad +(\alpha-1)\sum \ln x - \frac{\sum x}{\beta} \end{aligned} \tag{2}$$

2.2 Mixed-gamma distribution

The gamma probability density function (PDF) in equation (1) is defined for positive rainfall totals. A mixed-gamma model is defined for zero and positive rainfall totals [15]. The probability of zero rainfall totals is calculated as

$$p_0 = P(X=0) = \frac{n}{N}$$

where n is the count for zero rainfall totals and N is the total count of the rainfall data. The cumulative distribution function (CDF) for the mixed-gamma model is

$$\begin{aligned} F(x) &= P[X \le x] \\ &= P[X=0] + P[0 < X \le x] \\ &= p_0 + (1-p_0)F(x;\alpha,\beta) \end{aligned} \tag{3}$$

where we have written

$$F(x;\alpha,\beta) = \int_0^x f(\xi;\alpha,\beta)\, d\xi.$$

The mean is denoted by $E[X]$ where

$$\begin{aligned} E[X] &= p_0 \cdot 0 + (1-p_0)E[X] \\ &= (1-p_0)\alpha\beta \end{aligned} \tag{4}$$

for $X \sim G(\alpha,\beta)$ and the variance is $V[X] = E[X^2] - (E[X])^2$ where

$$\begin{aligned} E[X^2] &= p_0 \cdot 0^2 + (1-p_0)E[X^2] \\ &= (1-p_0)\alpha(\alpha+1)\beta^2. \end{aligned}$$

Therefore, the variance for a mixed-gamma random variable is

$$\begin{aligned} V[X] &= E[X^2] - (E[X])^2 \\ &= (1-p_0)\alpha(\alpha+1)\beta^2 - [(1-p_0)\alpha\beta]^2 \\ &= p_0(1-p_0)\alpha^2\beta^2 + (1-p_0)\alpha\beta^2. \end{aligned} \tag{5}$$

If X is a random variable with mixed-gamma distribution and parameters p_0 for the probability of zero, α for the shape and β for the scale then we will write $X \sim G(\alpha,\beta,p_0)$. If $p_0 = 0$ then there is no probability of a zero rainfall and we write $G(\alpha,\beta,0) = G(\alpha,\beta)$.

2.3 Formulae of sum of independent mixed-gamma variables

In this study, we also derive the formulae of mean and variance of the sum of two and three independent mixed-gamma variables. Consider two independent mixed-gamma variables X and Y both on $[0,\infty)$, with densities defined and denoted by $X \sim G(\alpha_1,\beta,p_{01})$ and $Y \sim G(\alpha_2,\beta,p_{02})$ where we assume $\beta_1 = \beta_2 = \beta$. To derive the density for the sum of two independent mixed-gamma variables X, Y, let $S = X + Y$ denote the sum where

$$\begin{aligned} F(x) &= p_{01} + (1-p_{01})F(x;\alpha_1,\beta) \\ G(y) &= p_{02} + (1-p_{02})F(y;\alpha_2,\beta). \end{aligned}$$

Hence the CDF for the sum of two independent mixed-gamma variables is

$$\begin{aligned} H(s) &= P[X+Y \le s] \\ &= p_{01}p_{02} + p_{01}(1-p_{02})F(s;\alpha_2,\beta) \\ &\quad +p_{02}(1-p_{01})F(s;\alpha_1,\beta) \\ &\quad +(1-p_{01})(1-p_{02})F(s;\alpha_1+\alpha_2,\beta). \end{aligned} \tag{6}$$

The expected value (mean) of the sum is calculated using the result obtained in equation (4), and is given by

$$\begin{aligned} E[S] &= p_{01}p_{02}\cdot 0 + p_{01}(1-p_{02})\alpha_2\beta \\ &\quad +p_{02}(1-p_{01})\alpha_1\beta \\ &\quad +(1-p_{01})(1-p_{02})(\alpha_1+\alpha_2)\beta \\ &= \beta[\alpha_1(1-p_{01}) + \alpha_2(1-p_{02})]. \end{aligned}$$

ISBN: 978-1-61804-281-1

Similarly, $E[S^2]$ is calculated using the result from equation (5);

$$\begin{aligned} E[S^2] &= \beta^2 \left[\alpha_1(\alpha_1+1)(1-p_{01}) \right. \\ &\quad +\alpha_2(\alpha_2+1)(1-p_{02}) \\ &\quad \left. +2\alpha_1\alpha_2(1-p_{01})(1-p_{02})\right]. \end{aligned}$$

Hence the variance $V[S]$ of the sum of two independent mixed-gamma variables is

$$\begin{aligned} V[S] &= E[S^2] - (E[S])^2 \\ &= \beta^2[\alpha_1(\alpha_1 p_{01}+1)(1-p_{01}) \\ &\quad +\alpha_2(\alpha_2 p_{02}+1)(1-p_{02})]. \end{aligned}$$

If $\alpha_1 = \alpha_2 = \alpha$, then the mean becomes

$$E[S] = \alpha\beta[2 - p_{01} - p_{02}] \tag{7}$$

and

$$\begin{aligned} E[S^2] &= \alpha^2\beta^2(4 - 3p_{01} - 3p_{02} + 2p_{01}p_{02}) \\ &\quad +\alpha\beta^2(2 - p_{01} - p_{02}). \end{aligned}$$

Hence, the variance is given by

$$\begin{aligned} V[S] &= \alpha^2\beta^2(p_{01} + p_{02} - p_{01}^2 - p_{02}^2) \\ &\quad +\alpha\beta^2(2 - p_{01} - p_{02}). \end{aligned} \tag{8}$$

The same procedure is applied for finding the formulae of mean and variance of the sum of three independent mixed-gamma variables. Consider three independent mixed-gamma variables defined on $(0, \infty)$ denoted by $X \sim G(\alpha_1, \beta)$, $Y \sim G(\alpha_2, \beta)$ and $Z \sim G(\alpha_3, \beta)$ with the same shape and scale parameters ($\alpha_1 = \alpha_2 = \alpha_3 = \alpha$ and $\beta_1 = \beta_2 = \beta_3 = \beta$), respectively. The mean and variance of the sum of three independent mixed-gamma variables are given by

$$E[S] = \alpha\beta[3 - p_{01} - p_{02} - p_{03}] \tag{9}$$

and

$$\begin{aligned} V[S] &= \alpha^2\beta^2(p_{01} + p_{02} + p_{03} - p_{01}^2 - p_{02}^2 - p_{03}^2) \\ &\quad +\alpha\beta^2(3 - p_{01} - p_{02} - p_{03}). \end{aligned} \tag{10}$$

The reader is referred to Zakaria [18] for further details of the derivation of mean and variance of sum of two and three independent mixed-gamma variables.

3 Case study

To test the formulae of mean and variance that we have derived in Section 2.3, we choose daily rainfall totals from Pooraka station in South Australia for the period 1901–1990. Before conducting further analysis, we do the marginal analysis for each selected month of December, January and February. The marginal analysis has three parts: 1. Calculate p_0 and fit observed positive rainfall totals to a gamma distribution (use MLE method, see Section 2.1), 2. fit generated data to gamma distribution and mixed-gamma distribution and 3. compare the mean and variance of observed and generated with the formula. The formulae of mean and variance for the mixed-gamma variables are given by (4), (5) and (7) – (10), respectively.

In the first part of marginal analysis, each data is fitted to the gamma distribution using formula given in (1). The shape and scale parameters ($\alpha_i, \beta_i;\ i = 1, 2, 3$) are estimated using the MLE method and the probability of zero rainfall totals ($p_{0i};\ i = 1, 2, 3$) is calculated. Then, we check the values of mean and variance using formulae of $\alpha\beta$ and $\alpha\beta^2$, respectively. Table 1 shows that the value of mean of the observed data matches the estimated mean but not the variance.

For part two, using the estimated parameters from part one, the synthetic rainfall totals are generated as follows:

- For the gamma model, generate three sets of data using the parameters ($\alpha_i, \beta_i;\ i = 1, 2, 3$).
- Similarly, for the mixed-gamma model, we generate three sets of data using the parameters ($\alpha_i, \beta_i, p_{0i};\ i = 1, 2, 3$).

Therefore, for each model we will have three sets of data, generated data for December, January and February from gamma model and mixed-gamma model.

4 Results and discussion

4.1 Marginal analysis of gamma and mixed-gamma models

The mean and variance are calculated for each data set and compared with the observed data and also with the formulae. For the gamma model, the formulae for the mean and variance are given by $\alpha\beta$ and $\alpha\beta^2$, respectively. For the mixed-gamma model, the formulae for the mean and variance are (4) and (5), respectively. Table 2 shows a comparison of the means and variances between the observed and generated data from the two models. The values of the mean from both models are reasonably close to the observed mean, however the values of the variance are sometimes over-estimated or under-estimated of the observed values.

ISBN: 978-1-61804-281-1

Table 1: Estimated parameters, mean (mm) and variance for December, January and February using gamma model

	Estimated parameters			Mean (mm)		Variance	
	p_0	α	β	$\alpha\beta$	observed	$\alpha\beta^2$	observed
Dec	0.0222	1.4954	17.9893	26.9011	26.9011	483.9318	401.8061
Jan	0.0667	1.2598	16.7341	21.0821	21.0821	352.7911	393.9547
Feb	0.0778	0.8829	24.3492	21.4988	21.4988	523.4788	556.5196

Table 2: Comparison of mean and variance of observed data with gamma and mixed-gamma models

	Observed	Gamma		Mixed-gamma	
		generated	formula	generated	formula
			Mean (mm)		
Dec	26.90	25.92	26.90	25.02	26.30
Jan	21.08	21.66	21.08	19.44	19.68
Feb	21.50	22.41	21.50	20.33	19.83
			Variance		
Dec	401.81	428.51	483.93	499.14	488.90
Jan	393.95	362.18	352.79	364.34	356.93
Feb	556.52	583.47	523.48	483.36	515.92

4.2 Mean and variance for the sum of two months

The application of the CDF of the sum of two independent mixed-gamma variables is tested using formula (6) on the months of December and January, which have similar scale values (β), using rainfall data from Pooraka station in South Australia for the period 1901–1990. Firstly, construct the sum of two rainfall totals by adding the corresponding rainfall totals for December and January. Then calculate the mean, variance and p_0 for the sum and also check the mean and variance using the formulae of mean and variance from the gamma model. The probability of zero is found to be $p_0 = 0.00$ and the estimated parameters are $\alpha = 2.90$ and $\beta = 15.87$. Table 3 shows the values of estimated and observed mean and variance, respectively.

Table 3: Mean and variance for the sum of two months rainfall using the gamma model

	Mean (mm)		Variance	
	$\alpha\beta$	observed	$\alpha\beta^2$	observed
Sum	45.98	45.98	667.36	729.56

The generated sum of rainfall data is formed from the sum of two individual generated set from the mixed-gamma model. In each generated set, we use the average of shape parameters ($\overline{\alpha} = (\alpha_1 + \alpha_2)/2 = 1.3776$), the average of scale parameters ($\overline{\beta} = (\beta_1 + \beta_2)/2 = 17.3617$) and the respective probability of zero rainfall ($p_{0i}, i = 1, 2$), refer Table 1. Then we take the corresponding sum of the two generated data.

In the next step, we calculate and compare the mean and variance of the sum of the observed data with the sum of the generated data of the mixed-gamma. Also, we compare the mean and variance of the generated sum obtained from the formulae given by (7) and (8), respectively. Table 4 shows a comparison of the mean and variance between the observed sum and the sum of two independent mixed-gamma variables. The values of generated mean (45.28 mm) and using formula (45.71 mm) are close to the observed mean (45.98 mm). The observed variance is lower possibly due to correlation between the experimental months data, that have not been included in the mixed-gamma models.

Table 4: Comparison of mean and variance for the sum of two months rainfall and the sum of two independent mixed-gamma variables

	Observed	Mixed-gamma	
		generated	formula
Mean (mm)	45.98	45.28	45.71
Variance	667.36	847.98	841.61

ISBN: 978-1-61804-281-1

4.3 Mean and variance for the sum of three months

Again, the same rainfall data from Pooraka, South Australia, for the period of 1901–1990 is used to test the formulae of mean and variance for the sum of three months. The months selected for the analysis are December, January and February. A similar procedure is followed as described in Section 4.2. Now the sum of rainfall data is constructed by adding three corresponding months. We use average α ($\overline{\alpha} = (\alpha_1 + \alpha_2 + \alpha_3)/3 = 1.2127$), average β ($\overline{\beta} = (\beta_1 + \beta_2 + \beta_3)/3 = 19.6909$) and the corresponding probability of zero rainfall ($p_{0i}; i = 1, 2, 3$) to generate three sets of synthetic rainfall totals, refer Table 1. Table 5 presents a comparison of the mean and variance for the observed data, generated data and the formula from sum of three independent mixed-gamma variables. We obtain a similar result as before that the values of generated mean (68.50 mm) and using formula (67.66 mm) are close to the observed mean (65.81 mm). Again, the variance of the data appears lower due to the same reason mentioned in the Section 4.2.

4.4 Goodness of fit test

The means and variances of the observed sum and the sum of independent mixed-gamma variables using generated data and formulae are compared. The Kolmogorov–Smirnov (KS) and Anderson–Darling (AD) goodness of fit tests are used to assess the fit between the observed sum and the sum of independent mixed-gamma variables. Table 6 gives the P-values for the sum of two and three independent mixed-gamma variables. In both cases, P-values > 0.05, hence the observed sum is not significantly different from the generated sum of independent mixed-gamma model at 5% significance level.

Table 5: Comparison of mean and variance for the sum of three months rainfall totals and the sum of three independent mixed-gamma variables

	Observed	Mixed-gamma	
		generated	formula
Mean (mm)	65.81	68.50	67.66
Variance	1237.53	1510.01	1421.03

5 Conclusion

The probability density function of the sum of two and three mixed-gamma variables are derived for the simplest cases when the scale parameters are common.

Table 6: P-values of goodness of fit tests for the sum of two and three independent mixed-gamma variables

	KS	AD
Sum of two	0.4649	0.4230
Sum of three	0.9750	0.6205

KS: Kolmogorov–Smirnov; AD: Anderson–Darling

We also obtained the formulae of the mean and variance for each of the case considered. Comparisons of the estimated values of the mean and variance from our simple model with both the estimated values from the fitted distribution and the observed values are presented. The results demonstrate that the formulae derived will give a good estimate for the mean and the goodness of fit test confirmed a good fit between the cumulative distributions of the observed sum and the generated sum. This methodology can be applied to find the sum of more than three independent mixed-gamma variables and the general form of the formulae can be derived.

Acknowledgements: The research was supported by the University Malaysia Pahang (Grant No. RDU120101).

References:

[1] N.T. Ison, A.M. Feyerherm and B.L. Dean, Wet period precipitation and the gamma distribution, *Journal of Applied Meteorology*, 10:658665, 1971.

[2] E.H. Chin, Modelling daily precipitation occurrence process in Markov chain, *Water Resources Research*, 13(6):949–956, 1977.

[3] A.M. Mathai, Storage capacity of a dam with gamma type inputs, *Ann. Inst. Statist. Math.–Part A*, 34(3), 1982, pp. 591-597.

[4] J. Roldan and D.A. Woolhiser, Stochastic daily precipitation models: A comparison of occurrence processes, *Water Resource Research*, 18(5):1451–1459, 1982.

[5] D.A. Woolhiser and J. Roldan, Seasonal and regional variability of parameters for stochastic dailly precipitation models: South Dakota, USA, *Water Resource Research*, 22(6):965–978, 1982.

[6] R.D. Stern and R. Coe, A model fitting analysis of daily rainfall data, *Journal of Royal Statistical Society Series A*, 147:134, 1984.

[7] P.G. Moschopoulos, The distribution of the sum of independent gamma random variables, *Ann. Inst. Statist. Math.–Part A*, 37, 1984, pp. 541-544.

[8] R.W. Katz and M.B. Parlange, Overdispersion phenomenon in stochastic modelling of precipitation, *J. Climate*, 11:591601, 1998.

[9] J. Martin-Vide and L. Gomez, Regionalization of Peninsular Spain based on the length of dry spells, *Int. J. Climatol.*, 1999, pp. 537-555.

[10] D.S. Wilks and R.L. Wilby, The weather generation game: A review of stochastic weather models, *Progress in Physical Geography*, 23(3):329357, 1999.

[11] H. Aksoy, Use of gamma distribution in hydrological analysis, *Turk J Engin Environ Sci.*, 24:419428, 2000.

[12] R. Srikanthan and T.A. McMahon, Stochastic generation of annual, monthly and daily climate data: A review, *Hydr. and Earth Sys. Sci.*, 5(4):633670, 2001.

[13] E. Bekele, Markov chain modelling and ENSO influences on the rainfall seasons of Ethiopia, National Meteorological Services Agency of Ethiopia, 2001, pp. 25–35.

[14] K. Rosenberg, J.W. Boland and P.G. Howlett, Simulation of monthly rainfall totals, *ANZIAM Journal* 46(E), 2004, pp. E85–E104.

[15] K. Rosenberg, Stochastic modelling of rainfall and generation of synthetic rainfall data at Mawson Lakes, *University of South Australia*, 2004.

[16] V.N. Sharda and P. K. Das, Modelling weekly rainfall data for crop planning in a sub-humid climate of India, *Agricultural Water Management*, 76:120–138, 2005.

[17] J. Piantadosi, J.W. Boland and P.G. Howlett, Simulation of rainfall totals on various time scales–Daily, Monthly and Yearly, *Environmental Modeling and Assessment* 14(4), 2009, pp. 431–438.

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

ISBN: 978-1-61804-281-1