ANZIAM J. 46 (E) ppE85–E104, 2004

Simulation of monthly rainfall totals

K. Rosenberg* J. Boland* P. G. Howlett*

(Received 11 August 2003; revised 10 September 2004)

Abstract

The observed distribution of non-zero rainfall totals for a given month is often modelled using a maximum likelihood estimate for the Gamma distribution. In this paper we show that a Gamma distribution can be regarded as a zero order approximation to a density distribution constructed by a series of associated Laguerre polynomials. The coefficients of the series are easily calculated and used to improve the shape of the initial approximation by adjusting higher order moments. We show that this more general method models joint probability distributions for two or more months and in particular that the series model does not require an assumption of independence between months. Finally we explore how the series method generates simulated data that is statistically indistinguishable from the observed data. We illustrate our methods on a case study at Mawson Lakes and although monthly correlations may not be significant we note that the rainfall records at Koonamore Station show many significant correlations for successive months.

*Centre for Industrial and Applied Mathematics, University of South Australia, Mawson Lakes, Australia 5095. mailto:Kathrine.Rosenberg@unisa.edu.au

See http://anziamj.austms.org.au/V46/E/Rosenberg/home.html for this article, Published November 25, 2004. ISSN 1446-8735

Contents

1 Introduction

We previously modelled [8] the non-zero monthly rainfall at Mawson Lakes by a maximum likelihood Gamma distribution $\Gamma(\alpha, \beta)$. Researchers in other countries [1, 10, 11] have also used the Gamma distribution for fitting rainfall data.

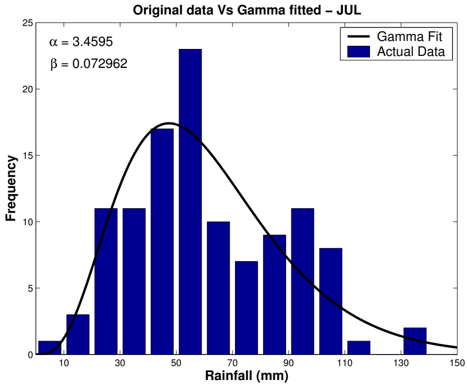
The maximum likelihood Gamma distribution matches the first order statistics but does not match higher order statistics and consequently may not match the observed characteristics. At Mawson Lakes there are some months where the Gamma distribution does not appear to have the correct shape. In this paper we propose a more general model in which the Gamma distribution is extended to a rapidly convergent series of associated Laguerre polynomials $L_m^\alpha(\beta x)$, where α and β are obtained from the parameters for the original Gamma distribution. The more general model allows us to match the observed higher order moments and hence allows us to match more of the observed characteristics. In Figure 1 we show how the shape of the fitted distribution converges to the shape of the observed distribution for the month of July as the number of terms increases.

If there are some observed zero values it may be best to assume that $\Pr[X = 0] > 0$. In such cases the data is modelled by a mixed distribution and the Gamma distribution or the more general series representation is used to model the non-zero part of the distribution.

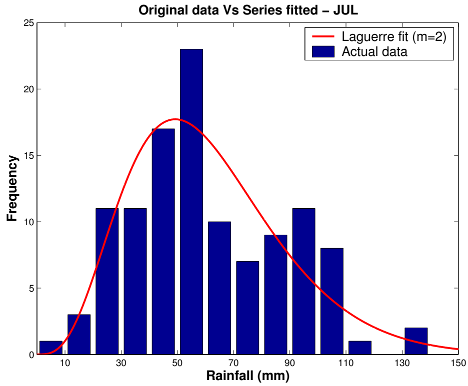
We also show that the series of associated Laguerre polynomials can be used to model the joint probability distribution for two or more months. Once again we show that the series method can be regarded as a generalisation of the elementary method in which the joint distribution is modelled by a product of independent Gamma distributions. In this case observe that the series method does not require an assumption of independence. In order to obtain analytic expressions for the various marginal distributions we will show that it is convenient to use new variables representing weighted totals and weighted proportions. For two variables x_1 and x_2 we define

$$t = \beta_1 x_1 + \beta_2 x_2 \quad \text{and} \quad s = \frac{\beta_1 x_1}{\beta_1 x_1 + \beta_2 x_2},$$

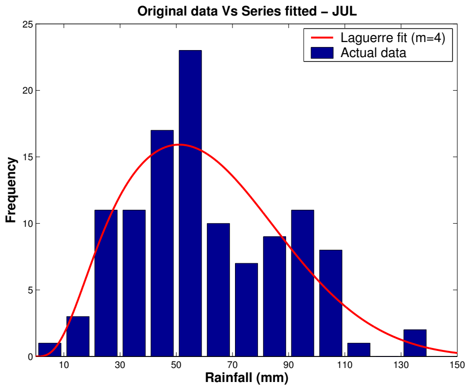
for the weighted total and weighted proportion of the first month respectively. Our methods make extensive use of several well-known formulae from the theory of special functions. In particular we use power series expansions



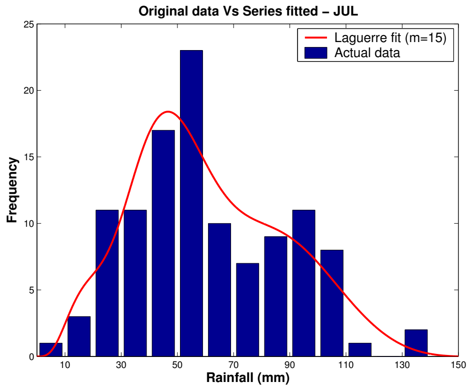
(a) Gamma (Series $M = 0$)



(b) Series ($M = 2$)



(c) Series ($M = 4$)



(d) Series ($M = 15$)

Figure 1: Compares the fitted densities for the Gamma and Laguerre series with the original data for July. There are 114 years of rainfall records at Mawson Lakes and all monthly recorded totals were non-zero. The parameters $\alpha = 3.4595$ and $\beta = 0.0730$ were determined by the maximum likelihood method

and standard integrals involving the Gamma and Beta functions. We emphasise that the calculations are performed easily in MATLAB on standard desktop computers.

The cumulative marginal probability for the weighted total and the cumulative conditional probability for the weighted proportion contributed by the first month can now be used to generate simulated rainfall totals for each month in a two month period. We use standard statistical tests to show that simulated data cannot be distinguished from the observed data and we extend our methods to the more general case of n months. We are also able to simulate certain special cases such as unusually dry spells or prolonged periods of high rainfall. Such unusual events are of great interest in catchment planning and management.

2 The associated Laguerre polynomials

For each non-negative integer m the associated Laguerre polynomial $L_m^\alpha(x)$ of order α is the unique polynomial solution of degree m to the differential equation

$$xy'' + (\alpha + 1 - x)y' + my = 0\,.$$

This equation is known as the associated Laguerre differential equation. The associated Laguerre polynomial

$$L_m^\alpha(x) = \sum_{p=0}^{m} \frac{\Gamma(\alpha+m+1)(-1)^p x^p}{\Gamma(\alpha+p+1)p!(m-p)!}\,. \tag{1}$$

The associated Laguerre polynomials satisfy an orthogonality relationship on the interval $[0, \infty)$:

$$\frac{m!}{\Gamma(\alpha+m+1)} \int_0^\infty L_m^\alpha(x) L_n^\alpha(x) x^\alpha e^{-x}\,dx = \begin{cases} 0\,, & m \neq n\,, \\ 1\,, & m = n\,, \end{cases} \tag{2}$$

with weight function $x^\alpha e^{-x}$. For further details see Bell [2] or Lebedev [4].

3 The probability distribution for one variable

Rosenberg et al. [8] modelled the distribution of non-zero monthly rainfall totals at Mawson Lakes as a random variable X_i for each $i = 1, 2, \ldots, 12$ with probability density given in non-standard notation by a Gamma distribution

$$\varphi(x_i) = \frac{\beta_i}{\Gamma(\alpha_i + 1)} (\beta_i x_i)^{\alpha_i} e^{-\beta_i x_i} . \tag{3}$$

They used maximum likelihood estimation to find the parameters α and β, and a Chi-square test was used to check the goodness of fit. The initial model formulated by Rosenberg et al. is here refined to find a probability density function (PDF) with matching statistics using a series of associated Laguerre polynomials [9] in the form

$$\varphi(x_i) = \sum_{m=0}^{\infty} c_m L_m^{\alpha_i}(\beta_i x_i)(\beta_i x_i)^{\alpha_i} e^{-\beta_i x_i} . \tag{4}$$

We calculate the coefficients c_m, from the standard orthogonality relationships (2), which imply

$$c_m = \frac{\beta_i m!}{\Gamma(m + \alpha_i + 1)} E\left[L_m^{\alpha_i}(\beta_i X_i)\right] .$$

In practice the PDF is represented by a truncated associated Laguerre series with density

$$\varphi(x_i) \quad \sim \quad \sum_{m=0}^{M} c_m L_m^{\alpha_i}(\beta_i x)(\beta_i x_i)^{\alpha_i} e^{-\beta_i x_i} .$$

It is a matter of judgment as to how many terms are used. The expansion of the PDF in (4) gives

$$\begin{aligned} \varphi(x_i) &= c_0 (\beta_i x_i)^{\alpha_i} e^{-\beta_i x_i} + c_1 L_1^{\alpha_i}(\beta_i x_i)(\beta_i x_i)^{\alpha_i} e^{-\beta_i x_i} \\ &\quad + c_2 L_2^{\alpha_i}(\beta_i x_i)(\beta_i x_i)^{\alpha_i} e^{-\beta_i x_i} + \cdots \\ &\sim \frac{\beta^{\alpha_i+1}}{\Gamma(\alpha_i + 1)} x_i^{\alpha_i} e^{-\beta_i x_i} , \end{aligned}$$

since $c_0 = \beta_i/\Gamma(\alpha_i + 1)$. The above expansion shows that the first term of the PDF is the maximum likelihood Gamma distribution, which means that the Gamma distribution is equivalent to the associated Laguerre series distribution with $M = 0$. The other terms from the expansion successively improve the fit of the truncated series to the data. There is no mathematical guarantee that the truncated series will be non-negative for $M = 1, 2, \ldots$ but in practice this property is usually preserved.

4 Correlation of monthly rainfall totals

Many assume that monthly rainfall totals are independent. At Mawson Lakes the observed correlations suggest this assumption is unjustified. We use Spearmans correlation method to test the significance of the correlations for monthly totals, see Table 1. Details of the Spearmans correlation method can be found in books by Hoel [3] or Pollard [6].

Our methodology can be applied to any location and some locations will have more inter-monthly dependence than others. Koonamore Station in the North-East pastoral district of South Australia has several pairs of monthly rainfall totals which have highly significant correlations, see Table 2. Note, for example, the block of significant correlations in the months August to November. On the other hand the number of correlations at Mawson Lakes are far less significant. At both locations there is correlation between months which are not adjacent; this suggests that an autoregressive model of order 1, or a Markov Process, would be inappropriate.

The method proposed in the following sections does not require the variables to be independent and can therefore account for any inter-monthly dependence which may be present.

Table 1: P-values resulting from Spearmans correlation method on the monthly rainfall totals at Mawson Lakes. Significant results at the 0.05 level are shown in bold.

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep
Jan	-	0.170	0.632	0.294	0.820	0.667	0.236	0.303	0.109
Feb	0.170	-	0.586	0.199	0.735	0.152	**0.017**	0.348	0.976
Mar	0.632	0.586	-	0.709	0.287	0.693	0.655	0.483	0.149
Apr	0.294	0.199	0.709	-	0.696	0.071	0.729	0.521	0.076
May	0.820	0.735	0.287	0.696	-	0.251	0.430	**0.038**	**0.045**
Jun	0.667	0.152	0.693	0.071	0.251	-	0.115	0.618	0.389
Jul	0.236	**0.017**	0.655	0.729	0.430	0.115	-	0.139	0.225
Aug	0.303	0.348	0.483	0.521	**0.038**	0.618	0.139	-	**0.013**
Sep	0.109	0.976	0.149	0.076	**0.045**	0.389	0.225	**0.013**	-
Oct	0.882	0.818	0.899	0.781	0.070	0.424	0.398	0.052	0.310
Nov	0.326	0.825	0.076	0.826	0.583	0.645	0.376	**0.026**	0.118
Dec	0.255	0.724	0.500	**0.010**	0.363	0.724	0.589	0.544	0.646

	Oct	Nov	Dec
Jan	0.882	0.326	0.255
Feb	0.818	0.825	0.724
Mar	0.899	0.076	0.500
Apr	0.781	0.826	**0.010**
May	0.070	0.583	0.363
Jun	0.424	0.645	0.724
Jul	0.398	0.376	0.589
Aug	0.052	**0.026**	0.544
Sep	0.310	0.118	0.646
Oct	-	**0.022**	0.055
Nov	**0.022**	-	0.192
Dec	0.055	0.192	-

Table 2: P-values resulting from Spearmans correlation method on the monthly rainfall totals at Koonamore Station. Significant results at the 0.05 level are shown in bold.

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep
Jan	-	0.972	0.517	**0.000**	**0.001**	0.981	**0.014**	0.414	0.793
Feb	0.972	-	0.077	0.547	0.790	0.910	0.934	0.699	0.362
Mar	0.517	0.077	-	0.324	**0.046**	**0.047**	0.131	0.857	0.887
Apr	**0.000**	0.547	0.324	-	**0.000**	0.955	0.317	0.454	0.140
May	**0.001**	0.790	**0.046**	**0.000**	-	0.584	0.417	0.441	0.526
Jun	0.981	0.910	**0.047**	0.955	0.584	-	**0.043**	0.934	0.338
Jul	**0.014**	0.934	0.131	0.317	0.417	**0.043**	-	0.706	0.120
Aug	0.414	0.699	0.857	0.454	0.441	0.934	0.706	-	**0.005**
Sep	0.793	0.362	0.887	0.140	0.526	0.338	0.120	**0.005**	-
Oct	0.272	**0.021**	0.905	**0.033**	0.777	0.125	0.928	0.448	**0.000**
Nov	0.336	0.498	0.298	0.781	0.318	0.768	0.311	**0.015**	**0.005**
Dec	0.281	0.540	0.847	0.285	0.516	**0.044**	0.068	0.997	0.127

	Oct	Nov	Dec
Jan	0.272	0.336	0.281
Feb	**0.021**	0.498	0.540
Mar	0.905	0.298	0.847
Apr	**0.033**	0.781	0.285
May	0.777	0.318	0.516
Jun	0.125	0.768	**0.044**
Jul	0.928	0.311	0.068
Aug	0.448	**0.015**	0.997
Sep	**0.000**	**0.005**	0.129
Oct	-	0.253	0.027
Nov	0.253	-	0.665
Dec	0.027	0.665	-

5 The joint distribution for two variables

We represent the two dimensional joint density for monthly rainfall in the form of a series of associated Laguerre polynomials

$$\varphi(x) = \sum_{m=0}^{\infty} c_m \left[\prod_{i=1}^{2} L_{m_i}^{\alpha_i}(\beta_i x_i) w_{\alpha_i}(\beta_i x_i) \right], \tag{5}$$

where $w_{\alpha_i}(\beta_i x_i) = (\beta_i x_i)^{\alpha_i} e^{-\beta_i x_i}$ is a weight function, and we write $x = (x_1, x_2)$ and $m = (m_1, m_2)$. The orthogonality of the associated Laguerre polynomials implies that the coefficients formula

$$c_m = \prod_{j=1}^{2} \frac{\beta_j m_j!}{\Gamma(m_j + \alpha_j + 1)} E\left[\prod_{j=1}^{2} L_{m_j}^{\alpha_j}(\beta_j X_j) \right],$$

for each m. We calculate the expected values from the observed data. We consider only the non-zero data points. If zero data points occur then the model can be extended to include all data using a mixed distribution.

6 Transformations and corresponding marginal distributions

To calculate the associated cumulative probabilities we change variables to

$$t = \beta_1 x_1 + \beta_2 x_2 \quad \text{and} \quad s = \frac{\beta_1 x_1}{\beta_1 x_1 + \beta_2 x_2},$$

where t is the weighted total and s is the weighted proportion for the first month. Equivalently, we can write

$$\beta_1 x_1 = st \quad \text{and} \quad \beta_2 x_2 = (1 - s)t\,.$$

The Jacobian determinant for the transformation is calculated as

$$|\det J| = \left|\det\begin{bmatrix} \frac{\partial x_1}{\partial s} & \frac{\partial x_1}{\partial t} \\ \frac{\partial x_2}{\partial s} & \frac{\partial x_2}{\partial t} \end{bmatrix}\right| = \left|\det\begin{bmatrix} \frac{t}{\beta_1} & \frac{s}{\beta_1} \\ \frac{-t}{\beta_2} & \frac{(1-s)}{\beta_2} \end{bmatrix}\right| = \frac{t}{\beta_1\beta_2}$$

and the joint density function $\varphi(x)$, for the random variables X_1 and X_2, in (5) is transformed into the new joint density function

$$\psi(s,t) = \sum_{m=0}^{\infty} c_m L_{m_1}^{\alpha_1}(st) L_{m_2}^{\alpha_2}((1-s)t)(st)^{\alpha_1}(1-s)^{\alpha_2} t^{\alpha_2} e^{-t} \frac{t}{\beta_1\beta_2} \tag{6}$$

for the random variables

$$T = \beta_1 X_1 + \beta_2 X_2 \quad \text{and} \quad S = \frac{\beta_1 X_1}{\beta_1 X_1 + \beta_2 X_2} .$$

Using the formula (1) for the associated Laguerre polynomials, equation (6) becomes

$$\psi(s,t) = \sum_{m,p} c_m k(m,p) s^{\alpha_1+p_1}(1-s)^{\alpha_2+p_2} t^{\alpha_1+\alpha_2+p_1+p_2+1} e^{-t} \tag{7}$$

where

$$k(m,p) = \prod_{i=1}^{2} \frac{\Gamma(\alpha_i+m_i+1)(-1)^{p_i}}{\Gamma(\alpha_i+p_i+1)(m_i-p_i)!p_i!} \frac{1}{\beta_i} .$$

It follows that the marginal density $h(t) = \int_0^1 \psi(s,t)ds$ for the weighted total T is given by

$$\begin{aligned} h(t) &= \sum_{m,p} c_m k(m,p) t^{\alpha_1+\alpha_2+p_1+p_2+1} e^{-t} \int_0^1 s^{\alpha_1+p_1}(1-s)^{\alpha_2+p_2} ds \\ &= \sum_{m,p} c_m k(m,p) t^{\alpha_1+\alpha_2+p_1+p_2+1} e^{-t} B(\alpha_1+p_1+1, \alpha_2+p_2+1) , \end{aligned}$$

where $B(\cdot,\cdot)$ denotes the standard Beta function[1]. The cumulative marginal probability for T is given by

$$H(\tau) \;=\; \int_0^\tau h(t)\,dt$$

[1] $B(x,y) = \int_0^1 s^{x-1}(1-s)^{y-1} ds$

$$= \sum_{m,p} c_m k(m,p)\Gamma(\alpha_1+\alpha_1+p_1+p_2+2;\tau)\Gamma(\alpha_1+\alpha_1+p_1+p_2+2) \times B(\alpha_1+p_1+1,\alpha_2+p_2+1)\,, \tag{8}$$

where $\tau \in [0,\infty)$, $\Gamma(\,\cdot\,)$ is the standard Gamma function[2] and $\Gamma(\,\cdot\,;\tau)$ is the incomplete Gamma function[3].

7 Simulation of rainfall for two months

Suppose we wish to simulate rainfall for a period of two months; we need to select a random weighted total and then, for the given weighted total, we also need to find a random weighted proportion. The weighted two monthly total $t = \beta_1 x_1 + \beta_2 x_2$, where x_1 and x_2 denote rainfall totals in the first and second months respectively. Since $H(\tau) \in [0,1]$ we choose a randomly generated weighted total τ by generating a random number $r \in [0,1]$ and then solving the equation $H(\tau) = r$ to find τ. This is depicted in Figure 2 for the months July and August. In this case $r = 0.376$ and $\tau = 7.8$.

Given a weighted total $\tau \in [0,\infty)$ we wish to find the weighted proportion for the first month. To do this we must find the conditional CDF for S given that $T = \tau$. By applying Bayes formula we have

$$\begin{aligned}
\Pr[S<\sigma \mid \tau \le T < \tau+\Delta\tau] &= \frac{\Pr[S<\sigma \;\;\&\;\; \tau \le T < \tau+\Delta\tau]}{\Pr[\tau \le T < \tau+\Delta\tau]} \\
&= \frac{\int_\tau^{\tau+\Delta\tau}\left[\int_0^\sigma \psi(s,t)\,ds\right]\,dt}{\int_\tau^{\tau+\Delta\tau}\left[\int_0^1 \psi(s,t)\,ds\right]\,dt} \\
&= \frac{\int_\tau^{\tau+\Delta\tau} h(\sigma;t)\,dt}{\int_\tau^{\tau+\Delta\tau} h(t)\,dt}
\end{aligned}$$

---

[2] $\Gamma(a) = \int_0^\infty x^{a-1}e^{-x}dx$

[3] $\Gamma(a;\tau) = \frac{1}{\Gamma(a)}\int_0^\tau t^{a-1}e^{-t}dt$

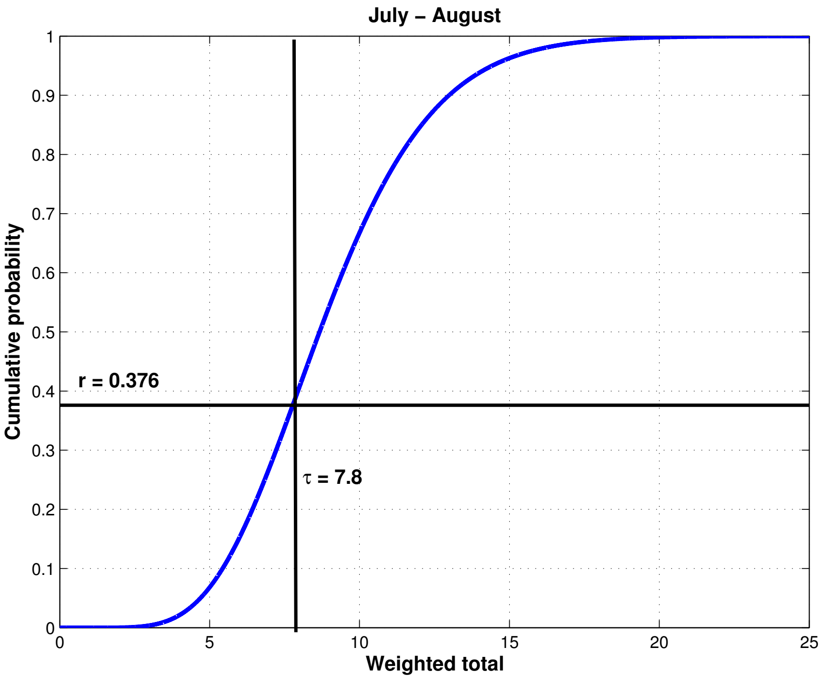


Figure 2: Using the CDF to generate a weighted total

$$\to \frac{h(\sigma;\tau)}{h(\tau)} \qquad \text{as} \quad \Delta\tau \to 0 \tag{9}$$

where $h(\sigma;t) = \int_0^\sigma \psi(s,t)\,ds$ and is given by

$$\begin{aligned} h(\sigma;\tau) &= \sum_{m,p} c_{[+,m]}\, k(m,p)\, \tau^{\alpha_1+\alpha_2+p_1+p_2+1} e^{-\tau} B(\alpha_1+p_1+1, \alpha_2+p_2+1) \\ &\qquad \times B(\alpha_1+p_1+1, \alpha_2+p_2+1; \sigma)\,. \end{aligned}$$

Also note that $h(1;\tau) = h(\tau)$. The above limit (9) exists for all τ and since the function $G(\sigma;\tau) = h(\sigma;\tau)/h(\tau)$ is non-negative and monotone increasing with $G(0;\tau)$ and $G(1;\tau) = 1$ we define the conditional probability

$$\begin{aligned} \Pr[S < \sigma \mid T = \tau] &= \lim_{\Delta\tau\to 0} \Pr[S < \sigma \mid \tau \le T < \tau + \Delta\tau] \\ &= \frac{h(\sigma;\tau)}{h(\tau)} \\ &= G(\sigma;\tau)\,. \end{aligned} \tag{10}$$

Find the weighted proportion σ by generating another random number $r_1 \in [0,1]$ and solving the equation $G(\sigma;\tau) = r_1$. Finally, the simulated rainfall total is (x_1^*, x_2^*), where

$$x_1^* = \frac{\tau\sigma}{\beta_1} \quad \text{and} \quad x_2^* = \frac{\tau(1-\sigma)}{\beta_2}\,.$$

8 Results for the two dimensional case

Incorporating the ideas and formulas above we generated (using MATLAB) some synthetic data for a two monthly period of rainfall at Mawson Lakes. Table 3 compares the statistics for the generated weighted totals and the actual weighted totals. Rainfall data was supplied from the Bureau of Meteorology. There are 100 generated weighted totals for each two monthly

Table 3: Statistics of the weighted totals for each two monthly period

	Actual Median	Generated Median	Actual SD	Generated SD	P-value
December–January	2.51	2.20	1.48	1.43	0.216
March–April	1.81	1.86	1.40	1.16	0.848
September–October	5.39	4.61	2.23	2.54	0.400

combination. The joint probability density for the two months is modelled by using a truncated series of associated Laguerre polynomials

$$\varphi(x) = \sum_{m=0}^{M} c_m \left[\prod_{i=1}^{2} L_{m_i}^{\alpha_i}(\beta_i x_i) w_{\alpha_i}(\beta_i x_i) \right] , \qquad (11)$$

with $M = 2$. Applying a Mann–Whitney test (details of which can be found in [5]), in each of the above examples, shows there is not enough evidence to reject the null hypothesis that the actual and generated totals come from the same population. This is because each P-value is greater than the significance level ($\alpha = 0.05$).

9 A simulated dry or wet event

If we can simulate a dry or wet season, then we can simulate the performance of a water cycle management system when water supply is low or high. It is important to understand how the system will behave during extreme periods. The occurrence of droughts and prolonged periods of high rainfall are of great interest in catchment planning and management.

To generate a random total, given that we wish to simulate a *one in* q dry event, we solve the equation $H(\tau) = r_1/q$, where r_1 is an random number in $[0, 1]$ and r_1/q will be a random number in $[0, 1/q]$.

To generate a random total, given that we wish to simulate a *one in* q wet event, we solve the equation $H(\tau) = 1 - r_1/q$, where $H(\tau)$ will be in the interval $[1-1/q, 1]$. Once τ has been generated a weighted proportion can be generated from $G(\sigma;\tau) = r_2$, where r_2 is a random number in $[0,1]$.

10 Extension to n months

The method in Section 6 is extended to n months with the following joint probability density function,

$$\varphi(x) = \sum_{m=0}^{\infty} c_m \left[\prod_{i=1}^{n} L_{m_i}^{\alpha_i}(\beta_i x_i) w_{\alpha_i}(\beta_i x_i)\right] \tag{12}$$

where $w_{\alpha_i}(\beta_i x_i) = (\beta_i x_i)^{\alpha_i} e^{-\beta_i x_i}$ is a weight function, and $m = (m_1, \ldots, m_n)$ and $x = (x_1, \ldots, x_n)$.

We have to make the change of variables such that

$$\begin{aligned} t &= \beta_1 x_1 + \cdots + \beta_n x_n\,, \quad s_1 = \frac{\beta_1 x_1}{\beta_1 x_1 + \beta_2 x_2}\,, \\ s_2 &= \frac{\beta_1 x_1 + \beta_2 x_2}{\beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3}\,, \quad \ldots, \quad s_{n-1} = \frac{\beta_1 x_1 + \cdots + \beta_{n-1} x_{n-1}}{\beta_1 x_1 + \cdots + \beta_n x_n}\,, \end{aligned} \tag{13}$$

where n is the number of months. Equivalently the above equations are

$$\beta_i x_i = \begin{cases} t \prod_{j=1}^{n-1} s_j\,, & \text{for } i = 1\,, \\ t(1 - s_{i-1}) \prod_{j=1}^{n-1} s_j\,, & \text{for } i = 2, \ldots, n\,. \end{cases}$$

For this change of variables the absolute value of the Jacobian determinant

$$|\det J| = t^{n-1} \prod_{i=1}^{n} \frac{1}{\beta_i} \prod_{i=2}^{n-1} s_i^{i-1}. \tag{14}$$

Consequently the joint density function $\varphi(x)$ in (12), for the random variables $X_1, \ldots, X_n$, transforms into the new joint density $\psi(s_1, \ldots, s_n, t)$,

$$\psi(s_1, \ldots, s_{n-1}, t) = \sum_{m=0}^{\infty} c_m \prod_{i=1}^{n} L_{m_i}^{\alpha_i}(\pi_i) w_{\alpha_i}(\pi_i) \prod_{j=1}^{n-1} t \frac{s_{j+1}^{j}}{\beta_j}\,. \tag{15}$$

for the random variables $S_1, \ldots, S_{n-1}$ and T. We define

$$\begin{aligned}
\pi_1 &= s_1 \cdots s_{n-1} t\,, \\
\pi_2 &= (1-s_1) s_2 \cdots s_{n-1} t\,, \\
\pi_3 &= (1-s_2) s_3 \cdots s_{n-1} t\,, \\
&\vdots \\
\pi_n &= (1-s_{n-1}) t\,.
\end{aligned}$$

Given the joint density function in (15) we find, using the method described in Section 6, the cumulative marginal probability distribution for the random variable T and the conditional cumulative marginal probability distribution for S_{n-1} given $T=\tau$, S_{n-2} given $T=\tau$ and $S_{n-1}=\sigma_{n-1}$ and so on. These are

$$\begin{aligned}
H(\tau) &= \sum_{m,p} c_m k(m,p) \prod_{i=1}^{n-1} B\left(i + \sum_{j=1}^{i} (\alpha_j + p_j), \alpha_{i+1} + p_{i+1} + 1\right) \\
&\quad \times \Gamma\left(n + \sum_{j=1}^{n} (\alpha_j + p_j)\right) \Gamma\left(n + \sum_{j=1}^{n} (\alpha_j + p_j); \tau\right)
\end{aligned}$$

and

$$\begin{aligned}
G(\sigma_{n-1};\tau) &= \sum_{m,p} c_m k(m,p) b_{n-1}(\sigma_{n-1};p) \tau^{(n-1)+\sum_{i=1}^{n}(\alpha_i+p_i)} e^{-\tau} \\
G(\sigma_{n-2};\tau,\sigma_{n-1}) &= \sum_{m,p} c_m k(m,p) b_{n-2}(\sigma_{n-2};p) \tau^{(n-1)+\sum_{i=1}^{n}(\alpha_i+p_i)} e^{-\tau} \\
&\quad \times \sigma_{n-1}^{n-2+\sum_{j=1}^{n-1}(\alpha_j+p_j)} (1-\sigma_{n-1})^{\alpha_n+p_n} \\
&\vdots \\
G(\sigma_1;\tau,\sigma_{n-1},\ldots,\sigma_2) &= \sum_{m,p} c_m k(m,p) b_1(\sigma_1;p) \tau^{(n-1)+\sum_{i=1}^{n}(\alpha_i+p_i)} e^{-\tau} \\
&\quad \times \sigma_{n-1}^{n-2+\sum_{j=1}^{n-1}(\alpha_j+p_j)} (1-\sigma_{n-1})^{\alpha_n+p_n} \\
&\quad \times \cdots \times \sigma_2^{\alpha_1+\alpha_2+p_1+p_2+1} (1-\sigma_2)^{\alpha_3+p_3}\,,
\end{aligned}$$

where

$$k(m,p) = \prod_{i=1}^{n} \frac{\Gamma(\alpha_i + m_i + 1)(-1)^{p_i}}{\Gamma(\alpha_i + p_i + 1)(m_i - p_i)!p_i!} \frac{1}{\beta_i},$$

and

$$\begin{aligned} b_h(\sigma_h; p) &= B\left(h + \sum_{j=1}^{h}(\alpha_j + p_j), \alpha_{h+1} + p_{h+1} + 1; \sigma_h\right) \\ &\quad \times \prod_{i=1}^{h} B\left(i + \sum_{j=1}^{i}(\alpha_j + p_j), \alpha_{i+1} + p_{i+1} + 1\right); \end{aligned}$$

$B(\cdot, \cdot; \sigma)$ is the incomplete Beta function.[4]

10.1 Simulation of rainfall for n months

In order to simulate rainfall we must firstly generate a random number $r \in [0,1]$ and solve $N_T(\tau) = r$. Given τ, our random weighted total, we then generate a succession of random numbers $r_1, r_2, \ldots, r_{n-1} \in [0,1]$ and solve equations using the conditional cumulative densities in sequence, that is

$$\begin{aligned} &N_{S_{n-1}}(\sigma_{n-1}; \tau) = r_1, \quad N_{S_{n-2}}(\sigma_{n-2}; \tau, \sigma_{n-1}) = r_2, \\ &\ldots, \quad N_{S_1}(\sigma_1; \tau, \sigma_{n-1}, \ldots, \sigma_2) = r_{n-1}. \end{aligned}$$

This paper has shown theoretically how to generate any number of monthly rainfall totals for a particular location. In practice this method is difficult to implement for more than three months because the numerical calculations become large. A multistage procedure overcomes these limitations and still retains the characteristics of the original probability distribution. Details can be found in Rosenberg [7].

[4] $B(x,y;a) = \frac{1}{B(x,y)} \int_0^a u^{x-1}(1-u)^{y-1} du$.

11 Conclusions and further work

We have applied our work to generate synthetic rainfall data for Mawson Lakes in South Australia. The aim of the Mawson Lakes water management project is to build a model of the water supply system in the Mawson Lakes catchment area. Wastewater and stormwater can be treated to provide a valuable resource in an urban environment with a large population and limited water supplies. Management of the resource is essential for efficient reuse.

Our methods generates simulated synthetic rainfall data which can be used as an input into the water cycle management system to simulate the consequent behaviour of the system for a given season. Our method allows us to match the statistics of the original data.

Our initial simulations are quite promising. The ideas can be extended to more than two variables and we can generate synthetic rainfall data for a range of different situations. In particular the engineers at Mawson Lakes are interested in simulating the effects of a *one in ten* dry season. They will find only 11 different realisations by looking through 114 years of records but we can generate a large number of *equally likely* events that have not yet necessarily occurred.

References

[1] M. E. Baldwin and S. Lakshmivarahan. Rainfall classification using histogram analysis: An example of data mining in meteorology. *Intelligent Engineering Systems Through Artificial Neural Networks*, 2002. E86

[2] W. W Bell. *Special Functions for Scientists and Engineers*. D. Van Nostrand Company Ltd., 1972. E89

[3] P. G. Hoel. *Elementary Statistics.* John Wiley and Sons, Inc., 2nd edition edition, 1966. E91

[4] N. N. Lebedev. *Special Functions and Their Applications.* Dover Publications, Inc., 1972. E89

[5] R. Meddis. *Statistical Handbook for Non-Statiticians.* McGraw Hill Book Company (UK) Limited, 1975. E99

[6] J. H. Pollard. *A Handbook of Numerical and Statistical Techniques: with examples mainly from the life sciences.* Cambridge University Press, 1977. E91

[7] K. Rosenberg. *Stochastic modelling of rainfall and generation of synthetic rainfall data at Mawson Lakes.* PhD thesis, University of South Australia, 2004. E102

[8] K. Rosenberg, J. Boland, and P. G. Howlett. Application of Laguerre polynomials to the simulation of rainfall at Mawson Lakes. In *Ninth International Conference on Urban Drainage, 9ICUD.* Portland, Oregon USA, September 2002. E86, E90

[9] K. Rosenberg, J. Boland, and P. G. Howlett. Modelling and simulation of rainfall data at Mawson Lakes. In *Ecosystems and Sustainable Development 2003 , ECOSUD.* Siena, Italy, June 2003. E90

[10] Z. Sen and A. G. Eljadid. Rainfall distribution function for Libya and rainfall prediction. *Hydrological Sciences*, 44(5):665–680, 1999. E86

[11] D. S. Wilks. Maximum likelihood estimation for the Gamma distribution using data containing zeros. *Journal of Climate*, 3:1495–1501, 1990. E86