KYBERNETIKA — VOLUME 47 (2011), NUMBER 1, PAGES 15–37

ON THE COMPOUND POISSON-GAMMA DISTRIBUTION

Christopher S. Withers and Saralees Nadarajah

The compound Poisson-gamma variable is the sum of a random sample from a gamma distribution with sample size an independent Poisson random variable. It has received wide ranging applications. In this note, we give an account of its mathematical properties including estimation procedures by the methods of moments and maximum likelihood. Most of the properties given are hitherto unknown.

Keywords: compound Poisson-gamma, estimation, expansions, moments

Classification: 62E15, 62E17, 62E20

1. INTRODUCTION

Suppose that the number of times it rains in a given time period, say N, has a Poisson distribution with mean λ, so that

$$P(N = i) = \exp(-\lambda)\ \lambda^i / i! = p_i$$

say. Suppose also that when it rains the amount of rain falling has a gamma distribution,

$$R \sim \alpha g(\alpha x, \rho) = \alpha^\rho x^{\rho-1} \exp\left(-\alpha x\right) / \Gamma(\rho)$$

for $x > 0$ with known shape parameter ρ. Suppose too that the rain falling with the time period are independent of each other and of N. Then the total rainfall in the time period is:

$$S = \sum_{i=1}^{N} R_i, \tag{1.1}$$

where $\{R_i\}$ are independent random variables with distribution that of R. Suppose now we observe the rainfall for n such periods, say $S_1, \ldots, S_n$.

The model given by (1.1) is known as the *Poisson-gamma model.* It was proposed on page 223 of Fisher and Cornish [5] for rainfall. Many authors have studied the Poisson-gamma model since then. For the case $\rho = 1$, (that is exponential rainfall) the maximum likelihood estimates are studied in Buishand [1], Ozturk [14] and

Revfeim [15]. A moments estimate and allowance for seasonality are given in Revfeim [15]. Hadjicostas and Berry [9] consider estimation based on Markov Chain Monte Carlo. Xia et al. [19] consider estimation based on a combination of maximum hierarchical-likelihood and quasi-likelihood.

Several generalizations of the Poisson-gamma model have also been proposed. Nahmias and Demmy [13] propose a logarithmic version with applications to model leadtime demand. Fukasawa and Basawa [6] propose a state-space version. Christensen et al. [3] propose a hierarchical version with applications to model environmental monitoring. Henderson and Shimakura [10] propose a version to account for between-subjects heterogeneity and within-subjects serial correlation. Galue [7] proposes a generalization involving the intractable H function. Most recently, Choo and Walker [2] have proposed a multivariate version with applications to model spatial variations of disease.

Applications of the Poisson-gamma model have been wide ranging. Among others, it has been used to model radiocarbon-dated depth chronologies, gravel bedload velocity data, catch and effort data, distribution of micro-organisms in a food matrix, numbers of ticks on red grouse chicks, regional organ blood flow, pluviometric irregularity for the Spanish Mediterranean coast, identification of crash hot spots, multiple lesions per patient, recruitment in multicentre trials, BSE in western France, human capital distribution, mortality data, insurance, mall visit frequency, pump-failure data, mine equipment injury rates, the influence of gamete concentration on sperm-oocyte fusion and two-stage cluster sampling.

It appears however that many mathematical properties of (1.1) have not been known. The purpose of this note is to provide an account of mathematical properties of the Poisson-gamma distribution including estimation issues. Except possibly for the cumulants and some of the estimation procedures, the results given are new and original. It is expected that this note could serve as a source of reference and encourage further research with respect to the Poisson-gamma model.

Section 2 gives various representations for the moment generating function, moments and cumulants of S. The representations for the moment generating function and the moments involve the Bell polynomial.

Note that the probability of no rain in any such period is $P(S=0)=p_0 = \exp(-\lambda) = 1-q_0$ say, and the amount of rain in any such period given that it does rain is: $S_+ = S|(S>0)$ with probability density function

$$f_{\boldsymbol{\theta}}^{+}(x) = q_0^{-1} \sum_{i=1}^{\infty} p_i \ \alpha g\left(\alpha x, i\rho\right) = \left\{\exp\left(\lambda\right)-1\right\}^{-1} \exp\left(-\alpha x\right) \ x^{-1} \ r_\rho\left(\nu x^\rho\right)$$

for $\boldsymbol{\theta} = (\lambda, \alpha, \rho)$ and $x > 0$, where ρ may or may not be known, $\nu = \lambda \alpha^\rho$ and

$$r_\rho(y) = \sum_{i=1}^{\infty} y^i / \left\{i!\Gamma(i\rho)\right\}. \tag{1.2}$$

In terms of the Dirac δ-function, S has probability density function

$$f_{\boldsymbol{\theta}}(x) = p_0\delta(x) + q_0 f_{\boldsymbol{\theta}}^{+}(x) = \exp\left(-\lambda\right)\delta(x) + \exp\left(-\lambda-\alpha x\right) \ x^{-1} r_\rho\left(\nu \ x^\rho\right). \tag{1.3}$$

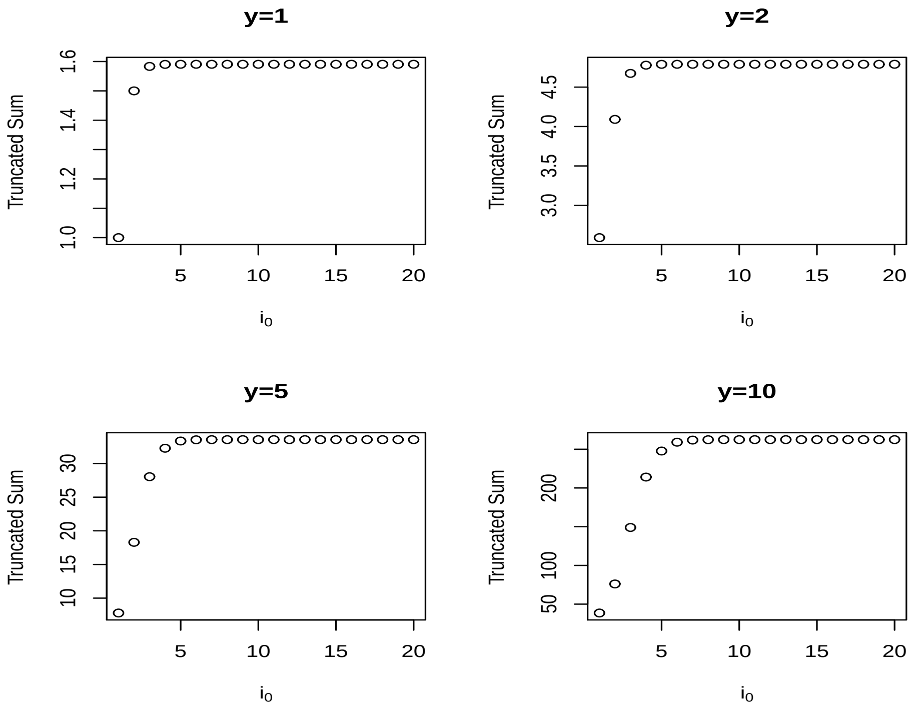

Fig. 1.1. The truncated sum in (1.4) versus i_0 for $y = 1, 2, 5, 10$ and $\rho = 1$.

For $\rho = 1$ this distribution was given by Le Cam [12] and equations (A2)-(A4) of Buishand [1]. See also Revfeim [16]. The function $r_\rho(y)$ converges quickly. Figure 1.1 shows how the truncated sum

$$\sum_{i=1}^{i_0} y^i / \{i!\Gamma(i\rho)\} \tag{1.4}$$

increases with respect to i_0 for $y = 1, 2, 5, 10$ and $\rho = 1$. We can see that only a few terms are needed to reach convergence even for y as large as 10.

Some exact expressions and expansions for $r_\rho(y)$ are given in Section 3. Some expansions for the probability density, cumulative distribution and the quantile functions of S are also given in Section 3. These extend expansions for the percentiles of S for the case $\rho = 1$ given by Fisher and Cornish [5].

In Appendix A, we show in a general setting that the unconditional maximum likelihood estimates $\widehat{\boldsymbol{\theta}}$ are more efficient than the maximum likelihood estimates $\widehat{\boldsymbol{\theta}}_c$ that condition on the number of zeros in the sample, say $M = n - m$. For our problem this inefficiency of conditioning is particularly poor when λ is small, that is when it seldom rains.

Sections 4 and 5 compare the maximum likelihood estimates with the maximum likelihood estimates conditioning on wet periods, and the maximum likelihood estimates with the moments estimates. Both sections assume that ρ is known, so that the unknowns are $\boldsymbol{\theta} = (\lambda, \alpha)'$. The asymptotic relative efficiency of $\widehat{\boldsymbol{\theta}}_c$ to $\widehat{\boldsymbol{\theta}}$ is plotted in Section 4. An exact expression and an expansion for the associated Fisher information are also given in Section 4.

Section 6 extends most of the results of Sections 4 and 5 to the case, where ρ is unknown. When considering moments estimates in Sections 5 and 6, we also give their asymptotic distributions. Finally, Section 7 illustrates the results of Sections 4 and 5 using two real rainfall data sets.

The advantage of choosing R to be gamma is that an explicit form is available for $f_{\boldsymbol{\theta}}$, the probability density function of S, and so the maximum likelihood estimates may be more easily computed. This is not the case for R lognormal or Weibull (both positive random variables) since their convolutions do not have closed forms, but it is the case for $R \sim \mathcal{N}(\mu, \sigma^2)$, giving

$$\begin{aligned} f_{\boldsymbol{\theta}}^{+}(x) &= (1-p_0)^{-1} \sum_{i=1}^{\infty} p_i \phi \left((x - i\mu)\, i^{-1/2} \sigma^{-1} \right) i^{-1/2} \sigma^{-1} \\ &= \beta \sum_{i=1}^{\infty} b^i \exp\left(-c/i\right) i^{-1/2} / i! \end{aligned}$$

for $\beta = \{\exp(\lambda) - 1\}^{-1} \exp(\mu x) \sigma^{-1} (2\pi)^{-1/2}$, $b = \lambda \exp(-\mu^2 \sigma^{-2}/2)$, $c = x^2 \sigma^{-2}/2$ and $\phi(\cdot)$ the standard normal probability density function. However, this rules out applications such as rainfall, where R may not be negative.

Throughout, we shall write $C_n \approx \sum_{r=0}^{\infty} c_{rn}$ to mean that that for $i \geq 1$ under suitable regularity conditions $C_n - \sum_{r=0}^{i-1} c_{rn}$ converges to zero as $n \to \infty$. We shall also write $\dot{\omega}(\cdot)$ to denote the first derivative of $\omega(\cdot)$.

2. MOMENT PROPERTIES

Theorem 2.1 gives $\mathrm{E} \exp(tS)$, $\mathrm{E} S^r$ and $\mu_r(S) = \mathrm{E}(S - \mathrm{E} S)^r$ in terms of the rth Bell polynomial

$$B_r(\mathbf{y}) = \sum_{k=1}^{r} B_{rk}(\mathbf{y})$$

for $r \geq 1$. The partial exponential Bell polynomials $\{B_{rk}\}$ are tabled on page 307 of Comtet [4]. For $\mathbf{y} = (y_1, y_2, \ldots)$, $B_r(\mathbf{y})$ may be defined by

$$\exp\left\{ \sum_{i=1}^{\infty} y_i t^i / i! \right\} = 1 + \sum_{r=1}^{\infty} B_r(\mathbf{y}) t^r / r!. \tag{2.1}$$

Theorem 2.2 gives the cumulants of S.

Theorem 2.1. The moment generating function, raw moments and the central moments of S are given by

$$\mathrm{E}\exp(tS) = \exp\left\{\sum_{i=1}^{\infty} y_i (t/\alpha)^i / i!\right\}, \tag{2.2}$$

$$\mathrm{E}S^r = \alpha^{-r} B_r\left(\lambda\rho_1,\ \lambda\rho_2, \ldots\right), \tag{2.3}$$

$$\mu_r(S) = \alpha^{-r} B_r\left(0, \lambda\rho_2,\ \lambda\rho_3, \ldots\right) \tag{2.4}$$

for $y_i = \lambda\rho_i$ and $\rho_i = [\rho]_i = \rho(\rho+1)\cdots(\rho+i-1)$. In particular,

$$\begin{aligned}
\mu_2\alpha^2 &= y_2 = \lambda\rho_2,\\
\mu_3\alpha^3 &= y_3 = \lambda\rho_3,\\
\mu_4\alpha^4 &= y_4 + 3y_2^2 = \lambda\rho_4 + 3\lambda^2\ \rho_2^2,\\
\mu_5\alpha^5 &= y_5 + 10y_2y_3 = \lambda\rho_5 + 10\lambda^2\rho_2\rho_3,\\
\mu_6\alpha^6 &= y_6 + 15y_2y_4 + 10y_3^2 + 15y_2^3 = \lambda\rho_6 + 5\lambda^2\left(3\rho_2\rho_4 + 2\rho_3^2\right) + 15\lambda^3\rho_2^3,\\
\mu_7\alpha^7 &= y_7 + 21y_2y_5 + 35y_3y_4 + 105y_2^2\ y_3 = \lambda\rho_7 + 7\lambda^2\left(3\rho_2\rho_5 + 5\rho_3\rho_4\right)\\
&\quad +105\lambda^3\rho_2^2\rho_3,\\
\mathrm{E}S\alpha &= y_1 = \lambda\rho,\\
\mathrm{E}S^2\alpha^2 &= y_2 + y_1^2 = \lambda\rho_2 + \lambda^2\rho^2,\\
\mathrm{E}S^3\alpha^3 &= \lambda\rho_3 + 3\lambda^2\rho_1\rho_2 + \lambda^3\rho_1^3,\\
\mathrm{E}S^4\alpha^4 &= \lambda\rho_4 + \lambda^2\left(4\rho_1\rho_3 + 3\rho_2^2\right) + 6\lambda^3\rho_1^2\rho_2 + \lambda^4\rho_1^4,\\
\mathrm{E}S^5\alpha^5 &= \lambda\rho_5 + 5\lambda^2\left(\rho_1\rho_4 + 2\rho_2\rho_3\right) + 5\lambda^3\left(2\rho_1\rho_3 + 3\rho_1\rho_2^2\right) + 10\lambda^4\rho_1^3\rho_2 + \lambda^5\rho_1^5,\\
\mathrm{E}S^6\alpha^6 &= \lambda\rho_6 + \lambda^2\left(6\rho_1\rho_5 + 15\rho_2\rho_4 + 10\rho_3^2\right) + 15\lambda^3\left(\rho_1^2\rho_4 + 4\rho_1\rho_2\rho_3 + \rho_2^3\right)\\
&\quad +5\lambda^4\left(4\rho_1^3\rho_3 + 9\rho_1^2\rho_4\right) + 15\lambda^5\rho_1^4\rho_2 + \lambda^6\rho_1^6,\\
\mathrm{E}S^7\alpha^7 &= \lambda\rho_7 + 7\lambda^2\left(\rho_1\rho_6 + 3\rho_2\rho_5 + 5\rho_3\rho_4\right)\\
&\quad +7\lambda^3\left(3\rho_1^2\rho_5 + 15\rho_1\rho_2\rho_4 + 10\rho_1\rho_3^2 + 15\rho_2^2\rho_3\right)\\
&\quad +35\lambda^4\left(\rho_1^3\rho_4 + 6\rho_1^2\rho_2\rho_3 + 3\rho_1\rho_2^3\right)\\
&\quad +35\lambda^5\left(\rho_1^4\rho_3 + 3\rho_1^3\rho_2^2\right) + 21\lambda^6\rho_1^5\rho_2 + \lambda^7\rho_1^7.
\end{aligned}$$

Proof. Note $\mathrm{E}\exp(tS)|N = u^N$ for $u = (1-t/\alpha)^{-\rho} = \mathrm{E}\exp(tG/\alpha)$ and $G \sim$ gamma(ρ), a gamma random variable with unit scale parameter. So,

$$\begin{aligned}
\mathrm{E}\exp(tS) &= \sum_{N=0}^{\infty}(1-t/\alpha)^{-N\rho}\frac{\lambda^N\exp(-\lambda)}{N!}\\
&= \exp\left\{\lambda\left[(1-t/\alpha)^{-\rho} - 1\right]\right\}\\
&= \exp\left\{\lambda\sum_{i=1}^{\infty}\binom{-\rho}{i}(-1)^i(t/\alpha)^i\right\},
\end{aligned}$$

so (2.2) follows. Next (2.3) follows by a direct application of the partial exponential Bell polynomials defined by (2.1). Finally,

$$\mathrm{E} \exp \{t (S - \mathrm{E}S)\} = \exp \left\{ \sum_{i=2}^{\infty} y_i (t/\alpha)^i / i! \right\},$$

so (2.4) follows. □

Theorem 2.2. The cumulants of S are given by

$$\kappa_r = \kappa_r(S) = \lambda \alpha^{-r} [\rho]_r. \tag{2.5}$$

P r o o f. Follows from (2.2). □

3. EXPANSIONS

In this section, we provide various expansions. We begin with $r_\rho(y)$ of (1.2). Theorem 3.1 derives exact expressions for $r_\rho(y)$. Some expansions for $r_\rho(y)$ for large y are given by Theorem 3.2. Finally, Theorem 3.3 provides expansions for the probability density, cumulative distribution and the quantile functions of S.

Theorem 3.1. We have

$$r_k(y) = k \mathcal{R}_k \left(y/k^k \right) \tag{3.1}$$

for $k = 1, 2, \ldots$, where

$$\mathcal{R}_k(z) = z \left(\frac{\partial}{\partial z} \right) {}_0F_k \left(; \frac{1}{k}, \frac{2}{k}, \ldots, \frac{k-1}{k}, 1; z \right)$$

and ${}_0F_q$ is the hypergeometric function defined by

$${}_0F_q \left(; \tau_1, \tau_2, \ldots, \tau_q; z \right) = \sum_{i=0}^{\infty} \frac{1}{[\tau_1]_i [\tau_2]_i \cdots [\tau_q]_i} \frac{z^i}{i!},$$

see Section 9.14 of Gradshteyn and Ryzhik [8]. In particular,

$$r_1(y) = y \sum_{i=0}^{\infty} y^i / \{ i!(i+1)! \} = z I_1(z)/2$$

at $z = 2y^{1/2}$, where $I_\nu(\cdot)$ is the modified Bessel function. So, if $\rho = 1$, S_+ has probability density function

$$f_{\boldsymbol{\theta}}^{+}(x) = \{ \exp(\lambda) - 1 \}^{-1} \exp(-\alpha x) \, x^{-1} z I_1(z)/2 \tag{3.2}$$

at $z = 2(\lambda \alpha x)^{1/2}$ and S has (unconditional) probability density function $\exp(-\lambda)$ $\delta(x)$ + $\exp(-\lambda - \alpha x)$ $x^{-1} z I_1(z)/2$, where δ is the Dirac function.

Proof. Note that

$$
\begin{aligned}
{}_0F_k\left(;\frac{1}{k},\frac{2}{k},\ldots,\frac{k-1}{k},1;z\right) &= \sum_{i=0}^{\infty}\frac{1}{[1/k]_i[2/k]_i\cdots[(k-1)/k]_i[1]_i}\frac{z^i}{i!}\\
&= \sum_{i=0}^{\infty}\frac{\left(k^i\right)^k}{k^i[1/k]_ik^i[2/k]_i\cdots k^i[(k-1)/k]_ik^i[1]_i}\frac{z^i}{i!}\\
&= \sum_{i=0}^{\infty}\frac{\left(k^k\right)^i}{(ki)!}\frac{z^i}{i!},
\end{aligned}
$$

so

$$\mathcal{R}_k(z)=\sum_{i=0}^{\infty}\frac{\left(k^k\right)^i}{(ki)!}\frac{z^i}{(i-1)!}.$$

The result in (3.1) follows. □

Equation (3.2) of Theorem 3.1 was given by Revfeim [15] (with $\mu=\alpha^{-1}$, $\lambda=\rho T$) but the first term omitted. Consequently, the likelihood derivatives are wrong by $O(\exp(-\lambda))$ which is negligible for λ large.

Theorem 3.2. For large y,

$$r_\rho(y)\approx\exp\left\{(\rho+1)\xi\right\}\left\{\rho\xi(2\pi)^{-1}(\rho+1)^{-1}\right\}^{1/2}\sum_{i=0}^{\infty}e_i\xi^{-i},$$

where $\xi=(y\rho^{-\rho})^{1/(\rho+1)}$, and e_i is given by equation (3.3) in Withers and Nadarajah [18, available on-line]. In particular, $e_0=1$, $e_1=-(\rho+1)^{-1}/24$, and $e_2=\{(1+\rho^{-1})^2/9-23(\rho+1)^{-2}/28\}/32$. So,

$$\log r_\rho(y)=(\rho+1)\xi+O(\log\xi)=ay^b+O(\log y) \tag{3.3}$$

for $a=(\rho+1)\rho^{-\rho/(\rho+1)}$ and $b=1/(\rho+1)$.

Proof. Follows by Example 4.2 of Withers and Nadarajah [18], available on-line. □

Revfeim [15] conjectures that the function

$$\Lambda_\rho(y)=\sum_{i=0}^{\infty}y^{\rho(i+1)}/\left\{i!\Gamma[\rho(i+1)+1]\right\}$$

satisfies $\partial_y\log\Lambda_\rho(y)\approx(\rho/y)^b$ and so $\log\Lambda_\rho(y)\approx(\rho+1)(y/\rho)^{\rho b}$, obtained by integration, as $y\to\infty$ for $b=(\rho+1)^{-1}$. Since

$$r_\rho\ (y^\rho)=\sum_{i=0}^{\infty}y^{\rho(i+1)}/\left\{(i+1)!\Gamma\left[\rho(i+1)\right]\right\},$$

we have $r_\rho(y) = \rho\Lambda_\rho(y^{1/\rho})$, so this suggests that

$$\log r_\rho(y) = \log\rho + \log\Lambda_\rho\left(y^{1/\rho}\right) \approx \log\rho + ay^b \tag{3.4}$$

as $y \to \infty$, where $a = (\rho+1)\rho^{-\rho b}$. Equation (3.3) in Theorem 3.2 confirms (3.4).

Theorem 3.3. Let $X = \kappa_2^{-1/2}(S - \kappa_1)$, where $\kappa_r = \kappa_r(S)$. Then, in terms of Φ, the standard normal cumulative distribution function, and ϕ, we have

$$P(X \le x) = P_\lambda(x) \approx \Phi(x) - \phi(x)\sum_{r=1}^\infty \lambda^{-r/2}\ h_r(x), \tag{3.5}$$

$$\Phi^{-1}\left(P_\lambda(x)\right) \approx x - \sum_{r=1}^\infty \lambda^{-r/2} f_r(x), \tag{3.6}$$

$$P_\lambda^{-1}\left(\Phi(x)\right) \approx x + \sum_{r=1}^\infty \lambda^{-r/2} g_r(x), \tag{3.7}$$

$$\dot{P}_\lambda(x) = p_\lambda(x) \approx \phi(x)\left\{1 + \sum_{r=1}^\infty \overline{h}_r(x)\right\}, \tag{3.8}$$

where h_r, $\overline{h}_r$, f_r and g_r are polynomials in x and $\{l_r = (\rho^2+\rho)^{-r/2}[\rho]_r\}$ given by Section 3 of Withers [17] with $l_1 = l_2 = 0$. In particular, in terms of the Hermite polynomials $He_r(x) = \phi(x)^{-1}\ (-d/dx)^r\ \phi(x)$,

$$\begin{aligned} h_1 &= f_1 = g_1 = l_3 He_2/6, \\ h_2 &= l_4 He_3/24 + l_3^2 He_5/72, \\ h_r(x) &= \sum_{j=1}^r He_{r+2j-1}(x) c_{rj}/j!, \end{aligned} \tag{3.9}$$

and

$$\overline{h}_r(x) = \sum_{j=1}^r He_{r+2j}(x) c_{rj}/j!, \tag{3.10}$$

where

$$\begin{aligned} c_{rj} &= \sum\left\{l_{r_1}\cdots l_{r_j}/\left(r_1!\cdots r_j!\right) : r_1 \ge 3, \ldots, r_j \ge 3, r_1 + \cdots + r_j = r + 2j\right\} \\ &= \left(\rho^2+\rho\right)^{-r/2-j}\sum\left\{\binom{\rho+1+s_1}{s_1+2}\cdots\binom{\rho+1+s_j}{s_j+2}\right. \\ &\qquad\qquad \left. : s_1 \ge 1, \ldots, s_j \ge 1,\ s_1 + \cdots + s_j = r\right\}. \end{aligned}$$

The number of terms in this last sum is the number of partitions of r into j parts allowing for permutations, that is $N_{rj} = \binom{r-1}{j-1}$.

Proof. By (2.5), $\kappa_r = \kappa_r(S) = \lambda\alpha^{-r}[\rho]_r$, so $X = \kappa_2^{-1/2}(S - \kappa_1)$ satisfies $\kappa_r(X) = \lambda^{1-r/2} l_r$ for $r \geq 3$. So, the expansions of Fisher and Cornish [5] apply for large λ to $P_\lambda(x) = P(X \leq x)$ and its probability density function $p_\lambda(x)$. □

Suppose $\rho = 1$. Then $c_{rj} = 2^{-r/2-j}\binom{r-1}{j-1}$ so (3.9), (3.10) give h_r and $\overline{h}_r$ explicitly. Also $\{g_r(x), 1 \leq r \leq 6\}$ are given by $m^{1/2}I$, mII, ... on pages 223-224 of Fisher and Cornish [5] and tabled for various levels on page 223.

Since S has a discrete component, it would seem preferable to apply (3.8) to the probability density function of its continuous component, S_+, rather than for S itself, that is to $X_+ = \kappa_2^{-1/2}(S_+ - \kappa_1)$. This can be justified since it is easy to show that $\kappa_r(S_+) = \kappa_r + O(\exp(-\lambda))$ as $\lambda \to \infty$, so for P_λ, p_λ the cumulative distribution and probability density functions of X^+, $O(\exp(-\lambda))$ should be added to the right hand sides of (3.5) – (3.8).

The moments of S_+ are related to those of S by $\mathrm{E}(S^+)^r = \mathrm{E}S^r/q_0$. Also

$$
\begin{aligned}
\mathrm{E}\left(S_+ - \mathrm{E}S_+\right)^r &= \sum_{i=0}^{r} \binom{r}{i} \left(\mathrm{E}\left(S_+\right)\right)^{r-i} \mathrm{E}\left(S_+^i\right) \\
&= \sum_{i=0}^{r} \binom{r}{i} \left(\lambda\rho\right)^{r-i} \alpha^{i-r} q_0^{i-r-1} \mathrm{E}\left(S^i\right) \\
&= \sum_{i=0}^{r} \binom{r}{i} \left(\lambda\rho\right)^{r-i} \alpha^{i-r} q_0^{i-r-1} \mathrm{E}\left(\left(S - \mathrm{E}(S) + \mathrm{E}(S)\right)^i\right) \\
&= \sum_{i=0}^{r}\sum_{j=0}^{i} \binom{r}{i}\binom{i}{j} \left(\lambda\rho\right)^{r-j} \alpha^{j-r} q_0^{i-r-1} \mu_j\left(S\right)
\end{aligned}
$$

for $r \geq 1$.

4. THE MAXIMUM LIKELIHOOD ESTIMATE

Consider a random sample of size n from S with probability density function (1.3). Let the positive values be $S_1, \ldots, S_m$, so there are $M = n - m$ zeros. We assume $m > 0$. We also assume that ρ is known - an assumption that is maintained in Section 5.

Here, we consider unconditional maximum likelihood estimates, $\widehat{\boldsymbol{\theta}}$, and conditional maximum likelihood estimates, $\widehat{\boldsymbol{\theta}}_c$. From standard maximum likelihood estimation theory,

$$n^{1/2}\left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}\right) \xrightarrow{L} \mathcal{N}\left(\mathbf{0}, \mathbf{I}(\boldsymbol{\theta})^{-1}\right) \tag{4.1}$$

and

$$m^{1/2}\left(\widehat{\boldsymbol{\theta}}_c - \boldsymbol{\theta}\right) \xrightarrow{\mathcal{L}} \mathcal{N}\left(\mathbf{0}, \mathbf{I}^+(\boldsymbol{\theta})^{-1}\right) \tag{4.2}$$

as $m, n \to \infty$, where, setting $\partial_{\boldsymbol{\theta}} = \partial/\partial\boldsymbol{\theta}$,

$$
\begin{aligned}
\mathbf{I}(\boldsymbol{\theta}) &= \mathrm{E}\partial_{\boldsymbol{\theta}} \log f_{\boldsymbol{\theta}}(S) \partial'_{\boldsymbol{\theta}} \log f_{\boldsymbol{\theta}}(S), \\
\mathbf{I}^+(\boldsymbol{\theta}) &= \mathrm{E}\partial_{\boldsymbol{\theta}} \log f^+_{\boldsymbol{\theta}}\left(S_+\right) \partial'_{\boldsymbol{\theta}} \log f^+_{\boldsymbol{\theta}}\left(S_+\right).
\end{aligned}
$$

Then by Appendix A,

$$\mathbf{I}(\boldsymbol{\theta}) = p_0 q_0^{-1} \begin{pmatrix} 10 \\ 00 \end{pmatrix} + q_0 \mathbf{I}^+(\boldsymbol{\theta}). \tag{4.3}$$

Since $q\mathbf{I}^+(\boldsymbol{\theta})^{-1} > \mathbf{I}(\boldsymbol{\theta})^{-1}$, $\widehat{\boldsymbol{\theta}}$ is more efficient $\widehat{\boldsymbol{\theta}}_c$.

Theorem 4.1 provides the unconditional maximum likelihood estimates for $\boldsymbol{\theta}$.

Theorem 4.1. Set $\nu = \lambda \alpha^\rho$, $\Delta(y) = \partial_y \log r_\rho(y)$, $u(x) = x^\rho \Delta(\nu\, x^\rho)$, $U(\nu) = u(S)$, $U_i(\nu) = u(S_i)$, $\overline{S}_+$ the sample mean of the $\{S_i\}$, $\overline{U}$ is the sample mean of the $\{U_i\}$, and $W(\nu) = \nu \overline{U}(\nu)$. Then, the maximum likelihood estimates, $\widehat{\boldsymbol{\theta}}$, are given by

$$\widehat{\alpha} = \left(\rho / \overline{S}_+\right) W\left(\widehat{\nu}\right), \ \widehat{\lambda} = Q^{-1}\left(2m\right), \tag{4.4}$$

where $Q^{-1}(\cdot)$ is the inverse function of $Q(\lambda) = (n-2m)/\{\exp(\lambda)-1\} + (m/\lambda)W(\widehat{\nu})$. Furthermore, $\widehat{\nu}$ satisfies

$$F\left(\widehat{\nu}\right) = 0, \tag{4.5}$$

where $F(\widehat{\nu}) = \widehat{\nu} - \{Q^{-1}(2m)\}(\rho/\overline{S}_+)^\rho W^\rho(\widehat{\nu})$.

P r o o f. Note that $m \sim \mathrm{Bi}(n, 1-\exp(-\lambda))$ and $P(m=0) = \exp(-n\lambda)$. The likelihood is

$$L = \binom{n}{m} p_0^{n-m} q_0^m \prod_{i=1}^m f_{\boldsymbol{\theta}}^+\left(S_i\right).$$

Recall that $p_0 = 1 - \exp(-\lambda) = 1 - q_0$. So, the maximum likelihood estimates, $\widehat{\boldsymbol{\theta}} = (\widehat{\lambda}, \widehat{\alpha})$, satisfy

$$\alpha^\rho \sum_{i=1}^m S_i^\rho \frac{\dot{r}_\rho\left(\lambda \alpha^\rho S_i^\rho\right)}{r_\rho\left(\lambda \alpha^\rho S_i^\rho\right)} = 2m - \frac{n-2m}{\exp(\lambda)-1} \tag{4.6}$$

and

$$\lambda \rho \alpha^{\rho-1} \sum_{i=1}^m S_i^\rho \frac{\dot{r}_\rho\left(\lambda \alpha^\rho S_i^\rho\right)}{r_\rho\left(\lambda \alpha^\rho S_i^\rho\right)} = \sum_{i=1}^m S_i. \tag{4.7}$$

Equations (4.6) and (4.7) reduce to

$$(m/\lambda) W\left(\nu\right) = 2m - \frac{n-2m}{\exp(\lambda)-1}$$

and

$$(m\rho/\alpha) W\left(\nu\right) = m\overline{S}_+,$$

respectively, so (4.4) follows. Finally, (4.5) follows by using $\widehat{\nu} = \widehat{\lambda}\widehat{\alpha}^\rho$ and (4.4). □

Theorem 4.2 checks that $\boldsymbol{\theta} = p\lim_{n\to\infty} \widehat{\boldsymbol{\theta}}$ is in fact a solution of (4.4).

Theorem 4.2. We have $\widehat{\boldsymbol{\theta}} = \boldsymbol{\theta}$ satisfying (4.4) in the limit as $n \to \infty$.

Proof. For a general function g, $\mathrm{E}g(S) = p_0 g(0) + q_0 \mathrm{E}g(S_+)$, so $\mathrm{E}S_+ = \lambda q_0^{-1} \rho \alpha^{-1}$; also $\mathrm{E}U(\nu) = \{\exp(\lambda) - 1\}^{-1} \partial_\nu L(\nu, \alpha)$ for

$$L(\nu, \alpha) = \int_0^\infty x^{-1} \exp(-\alpha x)\, r_\rho\left(\nu x^\rho\right) \mathrm{d}x = \exp(\lambda) - 1 \text{ at } \lambda = \nu \alpha^{-\rho}.$$

So, $\partial_\nu L(\nu, \alpha) = q_0^{-1}\, \lambda \nu^{-1}$, and the result follows. □

The information matrix corresponding to $\widehat{\boldsymbol{\theta}}$ is given by Theorem 4.3 in terms of the function:

$$K(\lambda, \rho) = \int_0^\infty \exp\left\{-(y/\lambda)^{1/\rho}\right\} \dot{r}_\rho(y)^2 r_\rho(y)^{-1} y\, \mathrm{d}y. \tag{4.8}$$

Theorem 4.4 provides an expansion for $K(\lambda, \rho)$ for small λ.

Theorem 4.3. The information matrix for $f_+(\boldsymbol{\theta})$ is given in terms of $L_{cd} = \mathrm{E}\, S_+^c\, U(\nu)^d$ by $\mathbf{I}_{f_+}(\boldsymbol{\theta}) = \mathbf{B} + \mathbf{b} L_{02}$, where $\mathbf{B}$ and $\mathbf{b}$ are 2×2 matrices given by:

$$\begin{aligned} B_{11} &= q_0^{-2} - 2 q_0^{-1} \lambda^{-1} \nu L_{01}, \\ B_{22} &= L_{20} - 2 \rho \alpha^{-1} \nu L_{11}, \\ B_{12} &= q_0^{-1}\left(L_{10} - \rho \alpha^{-1} \nu L_{01}\right) - \lambda^{-1} \nu L_{11}, \end{aligned}$$

and

$$\begin{aligned} b_{11} &= \lambda^{-2} \nu^2, \\ b_{22} &= \rho^2 \alpha^{-2}\, \nu^2, \\ b_{12} &= \rho \alpha^{-1} \lambda^{-1} \nu^2. \end{aligned}$$

The required $\{L_{cd}\}$ are:

$$L_{10} = q_0^{-1} \alpha^{-1} \lambda \rho, \tag{4.9}$$
$$L_{20} = q_0^{-1}\, \alpha^{-2} \lambda \rho \left\{1 + \rho(\lambda + 1)\right\}, \tag{4.10}$$
$$L_{01} = \mathrm{E}U(\nu) = q_0^{-1} \alpha^{-\rho} / \left\{\exp(\lambda) - 1\right\}, \tag{4.11}$$
$$L_{11} = -\left\{\exp(\lambda) - 1\right\}^{-1} \partial_\nu \partial_\alpha\, L(\nu, \alpha) = q_0^{-1} \left\{\exp(\lambda) - 1\right\}^{-1} \rho \alpha^{-\rho-1}, \tag{4.12}$$
$$L_{02} = \left\{\exp(\lambda) - 1\right\}^{-1} \rho^{-1} \nu^{-2} K(\lambda, \rho) \tag{4.13}$$

for $K(\lambda, \rho)$ of (4.8).

Theorem 4.4. We have

$$K(\lambda, \rho) \approx \rho \sum_{j=0}^\infty a_j^{-1} g_j^* \varepsilon^{j+1},$$

where

$$\begin{aligned}
&\varepsilon = \lambda^{\rho^2},\ a_j = \Gamma(\rho)^{-1}\Gamma\left((j+1)\rho\right),\ b_j = a_j/(j+1),\\
&\Omega_{\mathbf{a}} = \sum_{j=1}^{\infty} a_j x^j/j!,\ \Omega_{\mathbf{b}} = \sum_{j=1}^{\infty} b_j x^j/j!,\\
&G(x) = \dot{r}_\rho(x)^2 r_\rho^{-1}(x)x = \Gamma(\rho)^{-1}\left(1+\Omega_{\mathbf{a}}\right)^2\left(1+\Omega_{\mathbf{b}}\right)^{-1},\\
&g_j = G^{(j)}(0)/j!,\ g_j^* = g_j/g_0.
\end{aligned}$$

So,

$$g_j^* = (j!)^{-1}\sum_{r=0}^{j}\binom{j}{r} d_r e_{j-r},$$

where $d_r = 2a_r + C_r(2,\mathbf{a})$, $e_r = C_r(-1,\mathbf{b})$ and

$$C_r(\lambda,\mathbf{a}) = \sum_{i=0}^{r} B_{ri}(\mathbf{a})\langle\lambda\rangle_i$$

for $\langle\lambda\rangle_i = \Gamma(\lambda+1)/\Gamma(\lambda+1-i) = \lambda(\lambda-1)\cdots(\lambda-i+1)$ and $B_{ri}(\mathbf{a})$ the partial exponential Bell polynomial tabled on pages 307–308 of Comtet [4].

Proof. Follows by Theorem 2.1 of Withers and Nadarajah [18], available online. □

A plot of $K(\lambda,\rho)$ is given by Figure 4.1 for $\rho = 0.5, 1, 2, 4, 8$. Note that K was computed for $\rho = 1, 2, 4, 8$ using Gauss–Laguerre quadrature formula. The transformation $x = (y/\lambda)^{-1/\rho}$ gives K in the required form $K(\lambda,\rho) = \int_0^\infty \exp(-x) f(x)\,\mathrm{d}x$; here, $f(x) = \dot{r}_\rho(y)^2 r_\rho(y)^{-1} y\mathrm{d}y/\mathrm{d}x$. However, f was found to increase at a faster than polynomial rate, necessitating the introduction of an exponential scaling factor: a second transformation $z = (1-c)x$ gives the more useful form $K(\lambda,\rho) = \int_0^\infty \exp(-z) g(z)\,\mathrm{d}z \approx \sum_{i=1}^{J} w_i g(z_i)$, where $g(z) = \exp(-cx) f(x)\mathrm{d}x/\mathrm{d}z$. The weights and abscissae were provided by the NAG FORTRAN Library Routine D01BBF. The constant, c, was set to 0.01. For $\rho = 0.5$, K was found to be better approximated by automatic integration with the NAG FORTRAN Library Routine D01AMF. Note that K for $\rho = 8$ could not be estimated accurately for $\lambda > 5$, so the graph of K for $\rho = 8$ has been truncated at $\lambda = 5$.

Solution of $\widehat{\nu}$ of (4.5) can be done by Newton's method:

$$\widehat{\nu} = \nu_\infty,\ \text{where } \nu_{i+1} = \nu_i - \dot{F}\left(\nu_i\right)^{-1} F\left(\nu_i\right) \tag{4.14}$$

and ν_0 is an initial estimate, obtained possibly by some prior knowledge. If $\nu_0 = \nu + O_p(n^{-1/2})$, for example, if ν_0 is the moments estimate of Section 5, then $\nu^* = \nu_1$ has the same asymptotic properties as $\widehat{\nu}$ so no further iterations are necessary: let $\boldsymbol{\theta}^\star = P(\nu^*)$, where $\widehat{\boldsymbol{\theta}} = P(\widehat{\nu})$ is given by (4.4), then

$$n^{1/2}\left(\boldsymbol{\theta}^* - \boldsymbol{\theta}\right) \xrightarrow{\mathcal{L}} \mathcal{N}\left(\mathbf{0}, \mathbf{I}(\boldsymbol{\theta})^{-1}\right)$$

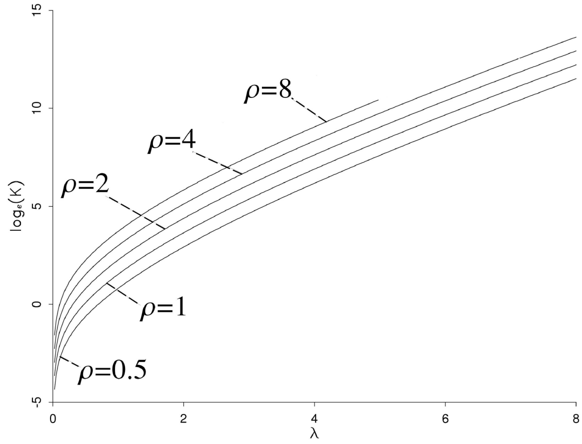

Fig. 4.1. $K(\lambda, \rho)$ of (4.8).

as $n \to \infty$, so for $t(\boldsymbol{\theta})$ any function in $\mathbb{R}$, a confidence interval for $t(\boldsymbol{\theta})$ of level $2\Phi(x) - 1 + O(n^{-1})$ is given by

$$|t(\boldsymbol{\theta}) - t(\boldsymbol{\theta}^*)| \leq n^{-1/2} x v(\boldsymbol{\theta}^*)^{1/2},$$

where $v(\boldsymbol{\theta}) = \dot{\mathbf{t}}(\boldsymbol{\theta})' \mathbf{I}(\boldsymbol{\theta})^{-1} \dot{\mathbf{t}}(\boldsymbol{\theta})$, $\dot{\mathbf{t}}(\boldsymbol{\theta}) = \partial_{\boldsymbol{\theta}} t(\boldsymbol{\theta})$ and $\mathbf{I}(\boldsymbol{\theta})$ is given by (4.3), (4.9)-(4.13) and (4.14). A test of $H_0 : t(\boldsymbol{\theta}) = t_0$ of the same level is given by accepting H_0 if

$$|t(\boldsymbol{\theta}) - t_0| \leq n^{-1/2} x v(\boldsymbol{\theta}^*)^{1/2}.$$

Theorem 4.5 is the analogue of Theorem 4.1 for maximum likelihood estimation conditional on m. Theorem 4.6 compares the unconditional and conditional maximum likelihood estimates.

Theorem 4.5. The maximum likelihood estimates, $\widehat{\boldsymbol{\theta}}_c$, conditional on m, are given by

$$\widehat{\alpha} = \left(\rho/\overline{S}_+\right) W(\widehat{\nu}), \ \widehat{\lambda} = Q_0^{-1}(W(\widehat{\nu})),$$

where $Q_0^{-1}(\cdot)$ is the inverse function of $Q_0(\lambda) = \lambda/\{1 - \exp(-\lambda)\}$.

P r o o f. The likelihood is

$$L \propto \prod_{i=1}^{m} f_{\boldsymbol{\theta}}^{+}(S_i).$$

So, the maximum likelihood estimates, $\widehat{\boldsymbol{\theta}} = (\widehat{\lambda}, \widehat{\alpha})$, satisfy

$$\alpha^{\rho} \sum_{i=1}^{m} S_i^{\rho} \frac{\dot{r}_{\rho}\left(\lambda \alpha^{\rho} S_i^{\rho}\right)}{r_{\rho}\left(\lambda \alpha^{\rho} S_i^{\rho}\right)} = \frac{m}{1-\exp(-\lambda)} \tag{4.15}$$

and

$$\lambda \rho \alpha^{\rho-1} \sum_{i=1}^{m} S_i^{\rho} \frac{\dot{r}_{\rho}\left(\lambda \alpha^{\rho} S_i^{\rho}\right)}{r_{\rho}\left(\lambda \alpha^{\rho} S_i^{\rho}\right)} = \sum_{i=1}^{m} S_i. \tag{4.16}$$

Equations (4.15) and (4.16) reduce to

$$(m/\lambda) W(\nu) = \frac{m}{1-\exp(-\lambda)}$$

and

$$(m\rho/\alpha) W(\nu) = m\overline{S}_{+},$$

respectively, so the result follows. □

Theorem 4.6. By (A.3) of Appendix A, the asymptotic relative efficiency of the conditional maximum likelihood estimates $\widehat{\boldsymbol{\theta}}_c$ to the unconditional maximum likelihood estimates $\widehat{\boldsymbol{\theta}}$ is given by:

$$ARE\left(\widehat{\lambda}_c \text{ to } \widehat{\lambda}\right) = q_0^2 \left\{p_0 I_{22}^{+} / \delta_{+} + q_0^2\right\}^{-1} = e_{\lambda}, \text{ say,} \tag{4.17}$$

$$ARE\left(\widehat{\alpha}_c \text{ to } \widehat{\alpha}\right) = \left\{p_0\, q_0^{-2} / I_{11}^{+} + 1\right\}\ e_{\lambda} = e_{\alpha}, \text{ say,} \tag{4.18}$$

where $\mathbf{I}^{+} = \mathbf{I}_{f_{+}}(\boldsymbol{\theta})$ of Theorem 4.3 and $\delta_{+} = \det(\mathbf{I}^{+})$. So, $e_{\alpha} > e_{\lambda}$ and neither depend on α.

The e_{α} and e_{λ} are plotted against λ in Figures 4.2 and 4.3. Clearly, the conditional maximum likelihood estimate for λ is very poor if λ is small, that is if it seldom rains in each period. The same is true of the conditional maximum likelihood estimate for the scale parameter α near $\rho = 1$, that is when the rainfall amounts $\{R_i\}$ are nearly exponential.

For exponential rainfall (that is $\rho = 1$) Buishand [1] gives the correct maximum likelihood estimate, but Ozturk [14] and Revfeim [15] do not: in effect they take the probability density function of S as $q_0 f_{\boldsymbol{\theta}}^{+}(x)$ and ignore the zeros in the data; see, for example, equation (3) of Revfeim [15]. The error will be small if λ is large.

5. THE MOMENTS ESTIMATE

Theorem 5.1 provides the moments estimates of $\boldsymbol{\theta}$ and its asymptotic distribution.

Theorem 5.1. If $\overline{S}$ and $\widehat{\vartheta}$ are the sample mean and variance for a random sample of size n from S, the moments estimators $\boldsymbol{\theta}$ are given by

$$\widetilde{\lambda} = \left(1+\rho^{-1}\right) \overline{S}^2 / \widehat{\vartheta},\ \widetilde{\alpha} = (\rho+1)\ \overline{S}/\widehat{\vartheta} \tag{5.1}$$

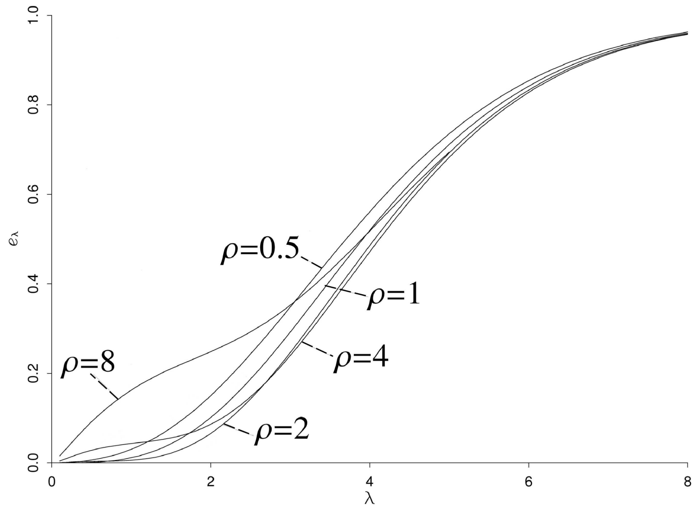

Fig. 4.2. e_λ of (4.17).

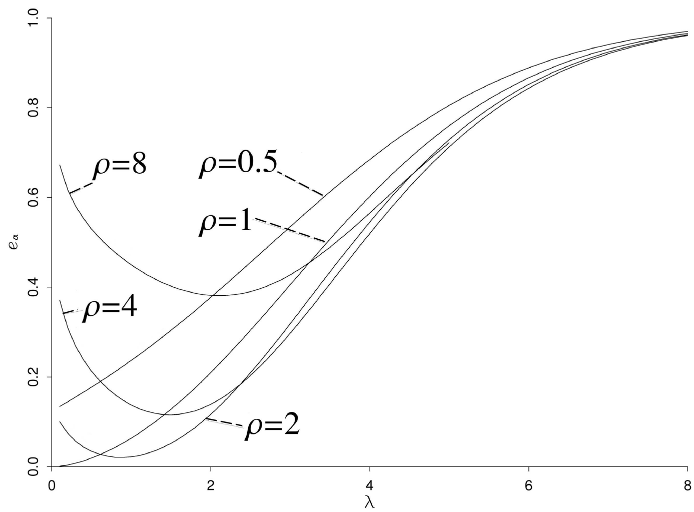

Fig. 4.3. e_α of (4.18).

with

$$n^{1/2}\left(\widetilde{\boldsymbol{\theta}}-\boldsymbol{\theta}\right) \stackrel{\mathcal{L}}{\to} \mathcal{N}\left(\mathbf{0}, \mathbf{V}(\boldsymbol{\theta})\right)$$

as $n \to \infty$, where $\mathbf{V} = \mathbf{V}(\boldsymbol{\theta})$ is given by

$$\begin{aligned} V_{11} &= 2\lambda^2 + \lambda\left(\rho^2+\rho+2\right)\rho^{-1}(\rho+1)^{-1}, \\ V_{12} &= -(2/\alpha)\left\{\lambda + \rho^{-1}(\rho+1)^{-1}\right\}, \\ V_{22} &= (1/\alpha)^2\left\{2 + \lambda^{-1}(\rho+3)\rho^{-1}(\rho+1)^{-1}\right\}. \end{aligned}$$

P r o o f. It follows from Theorem 2.2 that $\mathrm{E}S = \lambda\rho/\alpha$, var $S = \lambda\rho(\rho+1)/\alpha^2$. So, (5.1) follows. For a general function $f : \mathbb{R}^2 \to \mathbb{R}$, $\widetilde{w} = f(\overline{S}, \widehat{\vartheta})$ satisfies

$$n^{1/2}\left(\widetilde{w} - w\right) \stackrel{\mathcal{L}}{\to} \mathcal{N}(0, V) \qquad (5.2)$$

as $n \to \infty$, where

$$\begin{aligned} & w = f\left(\mathrm{E}S, \mathrm{var}S\right), \\ & V = f_{\cdot 1}^2 v_{11} + 2 f_{\cdot 1} f_{\cdot 2} v_{12} + f_{\cdot 2}^2 v_{22}, \\ & f_{\cdot i} = f_{\cdot i}\left(\mathrm{E}S, \mathrm{var}S\right),\ f_{\cdot i}\left(x_1, x_2\right) = \partial_{x_i} f\left(x_1, x_2\right), \\ & v_{11} = \mu_2 = \kappa_2,\ v_{12} = \mu_3 = \kappa_3,\ v_{22} = \mu_4 - \mu_2^2 = \kappa_4 + 2\kappa_2^2, \\ & \mu_r = \mathrm{E}\left(S - \kappa_1\right)^r. \end{aligned}$$

For $\mathbf{f}$ a vector, (5.2) also holds with

$$V_{12} = f_{1\cdot 1} f_{2\cdot 1} v_{11} + \left(f_{1\cdot 2} f_{2\cdot 1} + f_{1\cdot 1} f_{2\cdot 2}\right) v_{12} + f_{1\cdot 2} f_{2\cdot 2} v_{22},$$

where $f_{i\cdot j} = f_{\cdot j}$ for f_i the ith component of $\mathbf{f}$. Applying this to (5.1), we obtain the remainder the theorem. □

The asymptotic covariances of $\widehat{\boldsymbol{\theta}}$, $\widehat{\boldsymbol{\theta}}_c$ and $\widetilde{\boldsymbol{\theta}}$ all have the form $\left(\begin{smallmatrix} U_{11} & U_{12}/\alpha \\ U_{12}/\alpha & U_{22}/\alpha^2 \end{smallmatrix}\right)$, where $\mathbf{U}$ depends on λ but not α.

Plots of the asymptotic relative efficiency of $\widetilde{\lambda}$ to $\widehat{\lambda}$, and $\widetilde{a} = 1/\widetilde{\alpha}$ to $\widehat{a} = 1/\widehat{\alpha}$ against λ are given in Figures 5.1 and 5.2.

6. THE THREE PARAMETER PROBLEM

So far we have assumed ρ known. Typically ρ is taken as one. We now remove this assumption and set $\boldsymbol{\theta} = (\lambda, \alpha, \rho)'$. This augmented model can be used to test $\rho = 1$ say before using the first. The resulting maximum likelihood estimates, $\widehat{\boldsymbol{\theta}}$, given by Theorem 6.1 are awkward, but the moments estimates, $\widetilde{\boldsymbol{\theta}}$, given by Theorem 6.2 are manageable.

Theorem 6.1. We have $\widehat{\boldsymbol{\theta}}$ satisfying (4.4), (4.5) at $\widehat{\rho}$ and $\overline{h}_{\cdot\rho}(S_+) = 0$, where $\overline{h}_{\cdot\rho}(S_+)$ is the sample mean of $\{h_{\cdot\rho}(S_i)\}$, $h_{\cdot\rho}(x) = \nu(\log x)x^\rho\Delta(\nu x^\rho, \rho) + \mathcal{Q}(\nu x^\rho, \rho)$, $\Delta(y, \rho) = \Delta(y)$ and $\mathcal{Q}(y, \rho) = \partial_\rho \log r_\rho(y)$.

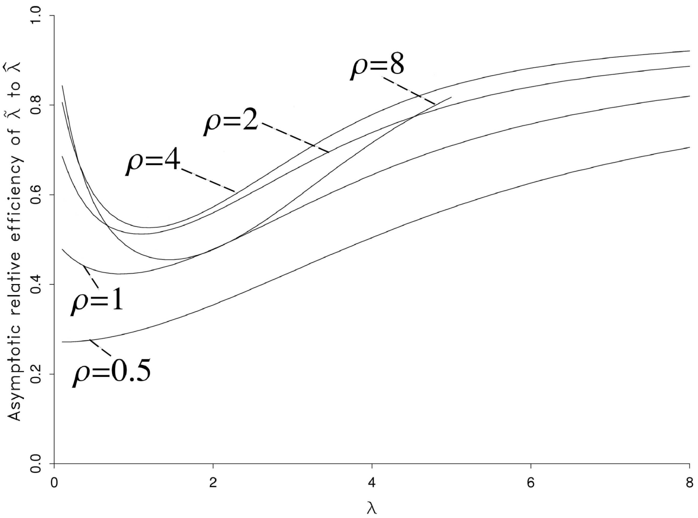

Fig. 5.1. Asymptotic relative efficiency of $\widetilde{\lambda}$ to $\widehat{\lambda}$.

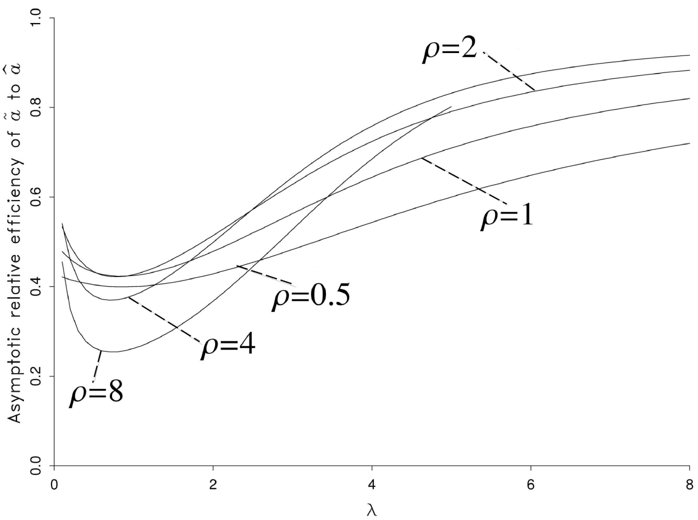

Fig. 5.2. Asymptotic relative efficiency of $\widetilde{a} = 1/\widetilde{\alpha}$ to $\widehat{a} = 1/\widehat{\alpha}$.

P r o o f. Follows since (4.1) – (4.3) hold with 2 replaced by 3, and $\binom{10}{00}$ by $\begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$. □

Again one can use Newton's equation to solve the equations (4.4), (4.5) and $\overline{h}_{\cdot\rho}(S_+) = 0$ for $(\widehat{\nu},\ \widehat{\rho}) = (\nu_\infty,\ \rho_\infty)$ or use $\widehat{\boldsymbol{\theta}}$ obtained from the first iteration of:

$$\begin{pmatrix} \nu_{i+1} \\ \rho_{i+1} \end{pmatrix} = \begin{pmatrix} \nu_i \\ \rho_i \end{pmatrix} - \dot{\mathbf{F}}(\nu,\rho)^{-1}\,\mathbf{F}(\nu,\rho), \tag{6.1}$$

where $\mathbf{F}(\nu,\rho)' = (F(\nu), \overline{h}_{\cdot\rho}(S_+))$ and $\dot{\mathbf{F}}(\nu,\rho) = \partial\mathbf{F}(\nu,\rho)/\partial(\nu,\rho)$, or one can use $\boldsymbol{\theta}^*$ given by (4.4) at (ν_1,ρ_1) of (6.1), where $(\nu_0,\ \rho_0)$ is a moments estimate.

In practice, one may be prepared to sacrifice efficiency and just use the moments estimates from $\kappa_r(S), 1 \le r \le 3$ given by Theorem 6.2.

Theorem 6.2. The moments estimates of $\boldsymbol{\theta}$ are given by

$$\widetilde{\boldsymbol{\theta}} = \left(\widetilde{\lambda}, \widetilde{\alpha}^{-1}, \widetilde{\rho}\right)' = \left(k_1^2\ k_2^{-1}(2-l)^{-1},\ k_2^{-1}k_3 - k_2k_1^{-1},\ (l-1)^{-1} - 1\right)',$$

where k_i is the ith sample cumulant (biased or unbiased) and $l = k_1k_2^{-2}k_3$. Also

$$n^{1/2}\left(\widetilde{\boldsymbol{\theta}} - \boldsymbol{\theta}\right) \xrightarrow{\mathcal{L}} \mathcal{N}(\mathbf{0}, \mathbf{V})$$

as $n \to \infty$, where

$$V_{11} = \sum_{k=1}^{3} \lambda^k\ V_{11\cdot k},$$

where

$$\begin{aligned} V_{11\cdot 1} &= -\rho^{-3}(\rho+1)^{-1}\left(11\rho^4 + 28\rho^3 - 2\rho^2 + 36\rho + 127\right), \\ V_{11\cdot 2} &= \rho^{-2}\left(18\rho^2 + 69\rho + 74\right), \\ V_{11\cdot 3} &= 6\rho^{-1}(\rho+1) \end{aligned}$$

and

$$V_{ij} = \sum_{k=-1}^{2} \lambda^k V_{ij\cdot k}$$

for $(i,\ j) \neq (1,1)$, where

$$
\begin{aligned}
& V_{12\cdot -1} = 0,\ V_{12\cdot 0} = \rho^{-2}(\rho+1)^{-1}\left(3\rho^4 - 12\rho^3 - 110\rho^2 - 206\rho - 112\right), \\
& V_{12\cdot 1} = -\rho^{-1}\left(24\rho^2 + 55\rho + 18\right),\ V_{12\cdot 2} = \rho + 1, \\
& V_{13\cdot -1} = 0,\ V_{13\cdot 0} = \rho^{-2}\left(3\rho^4 + 11\rho^3 + 15\rho^2 - \rho - 18\right), \\
& V_{13\cdot 1} = -\rho^{-1}(\rho+1)(\rho+2)\left(12\rho^2 + 28\rho + 3\right),\ V_{13\cdot 2} = -6(\rho+1)^2, \\
& V_{22\cdot -1} = \rho^{-1}(\rho+1)^{-1}\left(4\rho^6 + 40\rho^5 + 156\rho^4 + 289\rho^3 + 219\rho^2 - 27\rho - 89\right), \\
& V_{22\cdot 0} = 2(2\rho+3)\left(2\rho^3 + 7\rho^2 + 2\rho - 9\right), \\
& V_{22\cdot 1} = \rho(\rho+2)(2\rho+5),\ V_{22\cdot 2} = 6\rho(\rho+1), \\
& V_{23\cdot -1} = (\rho+2)\left(5\rho+14\right),\ V_{23\cdot 0} = -(\rho+1)(\rho+2)/(26\rho+57), \\
& V_{23\cdot 1} = -6\rho(\rho+1)^2,\ V_{23\cdot 2} = 0, \\
& V_{33\cdot -1} = 2\rho^{-1}(\rho+5),\ V_{33\cdot 0} = -(\rho+1)^2(\rho+2)(2\rho-5), \\
& V_{33\cdot 1} = 6\rho(\rho+1)^3,\ V_{33\cdot 2} = 0.
\end{aligned}
$$

Proof. The covariance follows by Rule 10 of Section 12.14, equations (10.10) and (3.38) of Kendall and Stuart [11], $v_{1i} = \kappa_{i+1}$, $v_{22} = \kappa_4 + 2\kappa_2^2$, $v_{23} = \kappa_5 + 6\kappa_3\kappa_2$ and $v_{33} = \kappa_6 + 9\kappa_4\kappa_2 + 9\kappa_3^2 + 6\kappa_2^3$. □

To test say $H : \rho = 1$, one accepts H at level $\Phi(x) - 1 + O(n^{-1})$ if and only if $m^{1/2}|\widetilde{\rho} - 1| \leq xV_{33}(\widetilde{\lambda},\ 1)^{1/2}$, where $V_{33}(\lambda,\ 1) = 12(\lambda^{-1} + 3 + 4\lambda)$.

7. APPLICATION

Here, we illustrate the results of Sections 4 and 5 using real data. We use the annual maximum daily rainfall data for the years from 1907 to 2000 for fourteen locations in west central Florida: Clermont, Brooksville, Orlando, Bartow, Avon Park, Arcadia, Kissimmee, Inverness, Plant City, Tarpon Springs, Tampa International Airport, St Leo, Gainesville, and Ocala. The data were obtained from the Department of Meteorology in Tallahassee, Florida.

Consider the distribution of S for $\rho = 1$, so the unknown parameters are λ and α. We fitted this distribution by the three methods: unconditional maximum likelihood estimation (Theorem 4.1), conditional maximum likelihood estimation (Theorem 4.5), and moments estimation (Theorem 5.1). The computer code used for implementing these estimation procedures can be obtained from the corresponding author.

Remarkably, unconditional maximum likelihood estimation provided the best fit for each location. The details for two of the locations, Orlando and Bartow, are given.

For Orlando, the estimates are $\widehat{\lambda} = 9.565$, $\widehat{\alpha} = 2.373$, $\widehat{\lambda}_c = 9.115$, $\widehat{\alpha}_c = 1.895$, $\widetilde{\lambda} = 7.394 \times 10^{-7}$ and $\widetilde{\alpha} = 0.183$. For Bartow, the estimates are $\widehat{\lambda} = 8.447$, $\widehat{\alpha} = 2.055$, $\widehat{\lambda}_c = 8.988$, $\widehat{\alpha}_c = 1.530$, $\widetilde{\lambda} = 3.404 \times 10^{-6}$ and $\widetilde{\alpha} = 0.184$.

The corresponding fitted probability density functions superimposed with the observed histograms are shown in Figures 7.1. and 7.2. It is clear that the general

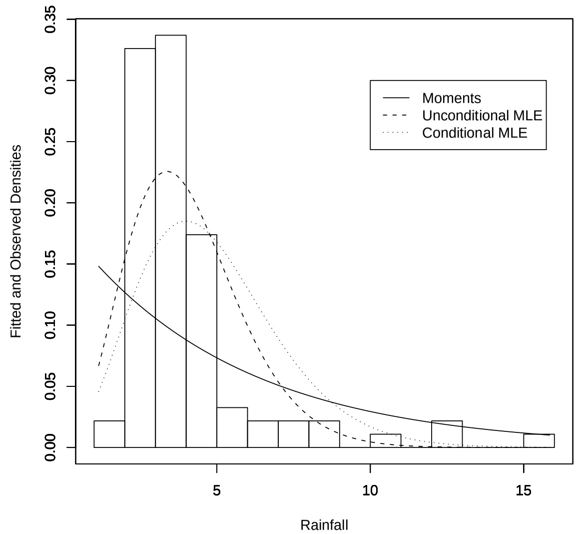

Fig. 7.1. Fitted probability density functions of S for rainfall data from Orlando.

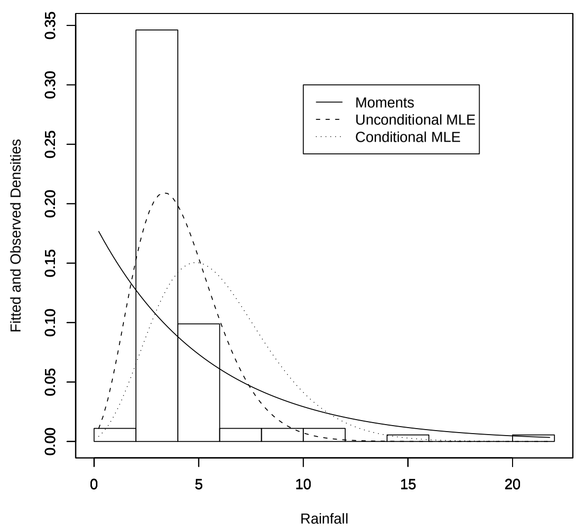

Fig. 7.2. Fitted probability density functions of S for rainfall data from Bartow.

pattern of the data is best captured by unconditional maximum likelihood estimation.

APPENDIX A

Theorem A.1 shows that it is less efficient to condition on M when finding the maximum likelihood estimate. Corollary A.1 derives the relative asymptotic efficiency of the maximum likelihood estimate versus the moments estimate of Section 4.

Theorem A.1. Consider a random variable

$$X = \begin{cases} 0, & \text{with probability } p = p(\boldsymbol{\theta}), \\ X_+, & \text{with probability } q = 1 - p, \end{cases}$$

where X_+ has probability density function $f_{\boldsymbol{\theta}}^+(x)$ with respect to Lebesgue measure and $\boldsymbol{\theta}$ in $\mathbb{R}^s$. Suppose we have a random sample of size n from X. Let M be the number of zeros, and $X_1, \ldots, X_m$ the non-zero values, where $m = n - M$. Let $\mathbf{I}_f^+(\boldsymbol{\theta})$ and $\mathbf{I}_f(\boldsymbol{\theta})$ denote the Fisher information matrix for X_+ and X, that is for $f_{\boldsymbol{\theta}}^+(X)$ and $f_{\boldsymbol{\theta}}(x) = p\delta(x) + qf_{\boldsymbol{\theta}}^+(x)$. Then $\widehat{\boldsymbol{\theta}}$ is more efficient asymptotically with equality if and only if p does not depend on $\boldsymbol{\theta}$. The relative asymptotic efficiency is $qI^{ii}/I^{+ii} = e_i$ say, where I^{ij} and I^{+ij} are the (i,j)th elements of $\mathbf{I}_f(\boldsymbol{\theta})^{-1}$ and $\mathbf{I}_f^+(\boldsymbol{\theta})^{-1}$, respectively.

Proof. Note that

$$\mathbf{I}_f(\boldsymbol{\theta}) = \mathrm{E}\partial_{\boldsymbol{\theta}} \log f_{\boldsymbol{\theta}}(X)\partial'_{\boldsymbol{\theta}} \log f_{\boldsymbol{\theta}}(X),$$

so

$$\mathbf{I}_f(\boldsymbol{\theta}) = (pq)^{-1}\mathbf{p}_{\cdot\boldsymbol{\theta}}\mathbf{p}'_{\cdot\boldsymbol{\theta}} + q\mathbf{I}_f^+(\boldsymbol{\theta}),$$

where $\mathbf{p}_{\cdot\boldsymbol{\theta}} = \partial p(\boldsymbol{\theta})/\partial\boldsymbol{\theta}$. So, if $\widehat{\boldsymbol{\theta}}_c$ and $\widehat{\boldsymbol{\theta}}$ are the conditional (on M) and unconditional maximum likelihood estimates,

$$m^{1/2}\left(\widehat{\boldsymbol{\theta}}_c - \boldsymbol{\theta}\right) \xrightarrow{\mathcal{L}} \mathcal{N}\left(\mathbf{0}, \mathbf{I}_f^+(\boldsymbol{\theta})^{-1}\right), \; n^{1/2}\left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}\right) \xrightarrow{\mathcal{L}} \mathcal{N}\left(\mathbf{0}, \mathbf{I}_f(\boldsymbol{\theta})^{-1}\right) \tag{A.1}$$

as $n \to \infty$, so

$$m^{1/2}\left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}\right) \xrightarrow{\mathcal{L}} \mathcal{N}\left(\mathbf{0}, q\mathbf{I}_f(\boldsymbol{\theta})^{-1}\right) \tag{A.2}$$

as $m \to \infty$. Comparing (A.1) and (A.2), $\widehat{\boldsymbol{\theta}}$ is more efficient asymptotically since $q\mathbf{I}_f(\boldsymbol{\theta})^{-1} = q(I^{ij})$ is less than or equal to $\mathbf{I}_f^+(\boldsymbol{\theta})^{-1} = (I^{+ij})$ elementwise. □

Corollary A.1. Suppose $\boldsymbol{\theta} \in \mathbb{R}^2$ and p depends on θ_1 but not θ_2. Set $p_1 = \partial p/\partial\theta_1$, $\mathbf{I} = \mathbf{I}_f(\boldsymbol{\theta})$ and $\mathbf{I}^+ = \mathbf{I}_f^+(\boldsymbol{\theta})$. Then

$$\mathbf{I} = p_1^2(pq)^{-1}\begin{pmatrix}10\\00\end{pmatrix} + q\mathbf{I}^+,$$

$$\det\mathbf{I} = p_1^2 p^{-1} I_{22}^+ + q^2 \det\mathbf{I}_+.$$

Also,

$$e_1 = q^2 \left(p_1^2 p^{-1} I_{22}^{+} / \det \mathbf{I}_{+} + q^2\right)^{-1}, \; e_2 = e_1 \left(1 + p_1^2 p^{-1} q^{-2} / I_{11}^{+}\right). \tag{A.3}$$

Here, e_1 and e_2 are the relative asymptotic efficiencies defined in the statement of Theorem A.1. They are compared in Theorem 4.6.

ACKNOWLEDGMENTS

The authors would like to thank the Editor and the two referees for careful reading and for their comments which greatly improved the paper.

(Received April 1, 2010)

REFERENCES

[1] T. A. Buishand: Stochastic Modelling of Daily Rainfall Sequences. Wageningen, Netherlands, Mededelingen Landbouwhogeschool 1977.

[2] L. Choo and S. G. Walker: A new approach to investigating spatial variations of disease. J. Roy. Statist. Soc. A *171* (2008), 395–405.

[3] A. Christensen, H. Melgaard, and J. Iwersen: Environmental monitoring based on a hierarchical Poisson-gamma model. J. Quality Technology *35* (2003), 275–285.

[4] L. Comtet: Advanced Combinatorics. Reidel Publishing Company, Dordrecht 1974.

[5] R. A. Fisher and E. A. Cornish: The percentile points of distributions having known cumulants. Technometrics *2* (1960), 209–225.

[6] T. Fukasawa and I. V. Basawa: Estimation for a class of generalized state-space time series models. Statist. Probab. Lett. *60* (2002), 459–473.

[7] L. Galue: A generalized hyper Poisson-gamma distribution associated with the H-function. Hadronic J. *30* (2007), 63–79.

[8] I. S. Gradshteyn and I. M. Ryzhik: Tables of Integrals, Series and Products. Fourth edition. Academic Press, New York 1965.

[9] P. Hadjicostas and S. M. Berry: Improper and proper posteriors with improper priors in a Poisson-gamma hierarchical model. Test *8* (1999), 147–166.

[10] R. Henderson and S. Shimakura: A serially correlated gamma frailty model for longitudinal count data. Biometrika *90* (2003), 355–366.

[11] M. Kendall and A. Stuart: The Advanced Theory of Statistics. Volume 1. MacMillan, New York 1977.

[12] L. Le Cam: A stochastic description of precipitation. In: Proc. Fourth Berkeley Symposium on Mathematical Statistics and Probability (J. Neyman, ed.), University of California Press, Berkeley 1961, volume 3, pp. 165–186.

[13] S. Nahmias and W. S. Demmy: The logarithmic Poisson gamma-distribution – a model for leadtime demand. Naval Research Logistics *29* (1982), 667–677.

[14] A. Ozturk: On the study of a probability distribution for precipitation totals. J. Appl. Meteorology *20* (1981), 1499–1505.

[15] K. J. A. Revfeim: Comments "On the study of a probability distribution for precipitation totals". J. Appl. Meteology *21* (1982), 97–100.

[16] K. J. A. Revfeim: A theoretically derived distribution for annual rainfall totals. Internat. J. Climatology *10* (1990), 647–650.

[17] C. S. Withers: Asymptotic expansions for distributions and quantiles with power series cumulants. J. Roy. Statist. Soc. B *46* (1984), 389–396.

[18] C. S. Withers and S. Nadarajah: Saddlepoint Expansions in Terms of Bell Polynomials. Technical Report, Applied Mathematics Group, Industrial Research Ltd., Lower Hutt, New Zealand 2010. Avaiable on-line at http://arxiv.org/.

[19] N. Xia, Z.-Z. Zhang, and Z.-L. Ying: Convergence rate of the L-N estimator in Poisson-gamma models. Acta Math. Appl. Sinica *22* (2006), 639–654.

Christopher Withers, Applied Mathematics Group, Industrial Research Limited, Lower Hutt. New Zealand.

e-mail: c.withers@irl.cri.nz

Saralees Nadarajah, School of Mathematics, University of Manchester, Manchester M13 9PL. United Kingdom.

e-mail: mbbsssn2@manchester.ac.uk