Journal of the Korean Statistical Society (2007), **36**: *1, pp 93–109*

SOME GENERALIZED GAMMA DISTRIBUTION

SARALEES NADARAJAH[1] AND ARJUN K. GUPTA[2]

ABSTRACT

Gamma distributions are some of the most popular models for hydrological processes. In this paper, a very flexible family which contains the gamma distribution as a particular case is introduced. Evidence of flexibility is shown by examining the shape of its pdf and the associated hazard rate function. A comprehensive treatment of the mathematical properties is provided by deriving expressions for the nth moment, moment generating function, characteristic function, Rényi entropy and the asymptotic distribution of the extreme order statistics. Estimation and simulation issues are also considered. Finally, a detailed application to drought data from the State of Nebraska is illustrated.

AMS 2000 subject classifications. Primary 33C90; Secondary 62E99.
Keywords. Drought modeling, gamma distribution, generalized gamma distribution.

1. INTRODUCTION

A random variable X is said to have the standard gamma distribution if its probability density function (*pdf*) is given by

$$f(x) = \frac{\lambda^\alpha x^{\alpha-1} \exp(-\lambda x)}{\Gamma(\alpha)} \tag{1.1}$$

for $x > 0$, $\alpha > 0$ and $\lambda > 0$. Gamma distributions are some of the most popular models for hydrological processes (Yue, 2001; Yue *et al.*, 2001; Shiau *et al.*, 2006; references therein). The aim of this paper is to introduce a generalization of (1.1) that could have much wider applicability in hydrology. The generalization is given by the *pdf*

$$f(x) = Cx^{\alpha-1}(x+z)^\rho \exp(-\lambda x) \tag{1.2}$$

Received May 2006; accepted September 2006.

[1]Corresponding author. School of Mathematics, University of Manchester, Manchester M60 1QD, U.K. (e-mail: saralees.nadarajah@manchester.ac.uk)

[2]Department of Mathematics and Statistics, Bowling Green State University, Bowling Green, OH 43403, U.S.A.

for $x > 0$, where $C = C(\alpha, \lambda, z, \rho)$ denotes a normalizing constant. We refer to this new distribution as the *generalized gamma* (GG) distribution. The parameter ranges are given by $\alpha > 0$, $\lambda > 0$, $z > 0$ and $-\infty < \rho < \infty$. By equation (2.3.6.9) in Prudnikov *et al.* (1986)

$$\int_0^\infty x^{\alpha-1}(x+z)^\rho \exp(-\lambda x)\, dx = z^{\alpha+\rho}\Gamma(\alpha)\Psi(\alpha, \alpha+1+\rho; \lambda z),$$

where $\Psi(a, b; u)$ denotes the confluent hypergeometric function defined by

$$\Psi(a, b; u) = \frac{1}{\Gamma(a)} \int_0^\infty t^{a-1}(1+t)^{b-a-1} \exp(-ut) dt. \tag{1.3}$$

Thus, the normalizing constant C is given by

$$\frac{1}{C(\alpha, \lambda, z, \rho)} = z^{\alpha+\rho}\Gamma(\alpha)\Psi(\alpha, \alpha+1+\rho; \lambda z). \tag{1.4}$$

The GG distribution given by (1.2) shares some of the attractive properties of the gamma distribution. For example, if a random variable X has the GG distribution with parameters $(\alpha, \lambda, a, \rho)$ then aZ also has the GG distribution with parameters $(\alpha, \lambda/a, az, \rho)$. The GG distribution is very flexible. Its particular cases include: the standard gamma distribution for $\rho = 0$ and the exponential distribution for $\rho = 0$ and $\alpha = 1$.

In the rest of this paper, we provide a comprehensive description of the mathematical properties of (1.2). We examine its shape and associated hazard rate function. We derive formulas for the nth moment, moment generating function (*mgf*), characteristic function (*cf*), Rényi entropy, and the asymptotic distribution of the extreme order statistics. We also consider estimation and simulation issues. Finally, an application to drought data from the State of Nebraska is illustrated to show (1.2) is a better model than (1.1).

2. SHAPE

The first and second derivatives of $\log f(x)$ for the GG distribution are:

$$\frac{d \log f(x)}{dx} = \frac{\alpha-1}{x} + \frac{\rho}{x+z} - \lambda$$

and

$$\frac{d^2 \log f(x)}{d^2 x} = \frac{1-\alpha}{x^2} - \frac{\rho}{(x+z)^2}. \tag{2.1}$$

Standard calculations based on these derivatives show that $f(x)$ can exhibit the following shapes:

- If $0 < \alpha < 1$ then f monotonically decreases with $f(0) = \infty$ and $f(\infty) = 0$.
- If $\alpha > 1$ then f has a single mode at $x = x_0$ with f increasing for all $x \leq x_0$ and decreasing for all $x > x_0$. Furthermore, $f(0) = 0$, $f(\infty) = 0$ and x_0 is the solution of

$$\frac{\alpha - 1}{x} + \frac{\rho}{x + z} = \lambda.$$

- If $\alpha = 1$ and either $\rho \leq 0$ or $-\lambda z \leq -\rho < 0$ then f monotonically decreases with $f(0) = 1/\{z\Psi(1, 2 + \rho; \lambda z)\}$ and $f(\infty) = 0$.
- If $\alpha = 1$ and $\rho > \lambda z$ then f has a single mode at $x = x_0 = (\rho - \lambda z)/\lambda$ with f increasing for all $x \leq x_0$ and decreasing for all $x > x_0$. Furthermore, $f(0) = 1/\{z\Psi(1, 2 + \rho; \lambda z)\}$ and $f(\infty) = 0$.

Note that unlike the standard gamma *pdf*, f can exhibit a unimodal shape even when $\alpha = 1$. In fact, if $\alpha = 1$ then $f(x) = C(x + z)^\rho \exp(-\lambda x)$ is a generalization of the exponential distribution and shares many of its attractive properties, including closure under scale transformations. Some of the possible shapes of f for selected values of (α, ρ) and $\lambda = z = 1$ are illustrated in Figure 2.1.

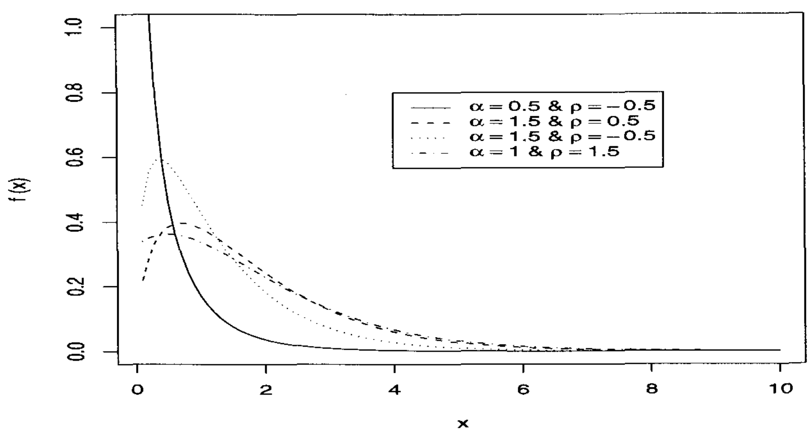

FIGURE 2.1 *Pdf of the generalized gamma distribution (1.2) for selected values of* (α, ρ) *and* $\lambda = z = 1$.

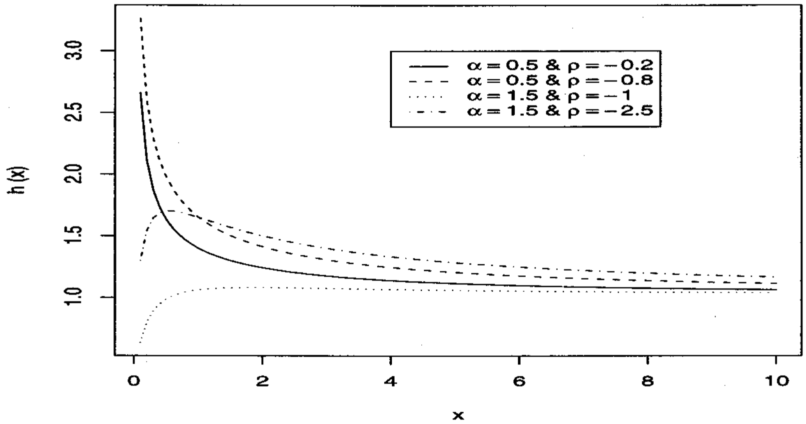

FIGURE 2.2 *Hazard rate function of the generalized gamma distribution (1.2) for selected values of* (α, ρ) *and* $\lambda = z = 1$.

The hazard rate function defined by $h(x) = f(x)/\{1 - F(x)\}$ is an important quantity characterizing life phenomena. A closed form expression for $h(x)$ is difficult to obtain for the GG distribution. However, we have illustrated, in Figure 2.2, the shape of the hazard rate function for selected values of (α, ρ) and $\lambda = z = 1$. Evidently, the GG distribution exhibits failure rates which are monotonically increasing and monotonically decreasing. More importantly, the GG distribution also exhibits a shape where the failure rate initially increases before decreasing for all x. Clearly, this is one feature that is not shared by the standard gamma distribution.

Moreover, since f is log-concave for $\rho > 0$ and $\alpha > 1$ (this follows from the second derivative $(\log f)''(x)$ in (2.1)), note that the generalized gamma must have increasing failure rate for $\rho > 0$ and $\alpha > 1$ (Gupta and Brown, 2001).

3. MOMENTS

The n^{th} moment of a random variable X having the GG distribution is

$$E\left(X^n\right) = C \int_0^\infty x^{n+\alpha-1}(x+z)^\rho \exp\left(-\lambda x\right) dx,$$

where C is the normalizing constant given by (1.4). By the definition, (1.3), of the confluent hypergeometric function,

$$\int_0^\infty x^{n+\alpha-1}(x+z)^\rho \exp(-\lambda x)\,dx$$
$$= z^{n+\alpha+\rho}\Gamma(n+\alpha)\Psi(n+\alpha, n+\alpha+1+\rho; \lambda z).$$

Thus, the n^{th} moment becomes

$$E(X^n) = \frac{z^n\Gamma(n+\alpha)\Psi(n+\alpha, n+\alpha+1+\rho; \lambda z)}{\Gamma(\alpha)\Psi(\alpha, \alpha+1+\rho; \lambda z)}. \tag{3.1}$$

In particular,

$$E(X) = \frac{z\alpha\Psi(1+\alpha, \alpha+2+\rho; \lambda z)}{\Psi(\alpha, \alpha+1+\rho; \lambda z)}$$

and

$$\mathrm{Var}(X) = \frac{z^2\alpha(\alpha+1)\alpha\Psi(2+\alpha, \alpha+3+\rho; \lambda z)}{\Psi(\alpha, \alpha+1+\rho; \lambda z)} - \frac{z^2\alpha^2\Psi^2(1+\alpha, \alpha+2+\rho; \lambda z)}{\Psi^2(\alpha, \alpha+1+\rho; \lambda z)}.$$

Figure 3.1 illustrates the flexibility of the GG distribution. We have plotted the skewness and kurtosis measures for the standard gamma ($\rho = 0$) and generalized gamma ($\rho = -0.5, 0.5$) distributions. Evidently the GG distribution shows greater degree of flexibility as compared to the standard gamma.

The *mgf* of X defined by $M(t) = E(\exp(tX))$ can be derived in a similar manner to (3.1) by application of the definition (1.3). It turns out that

$$M(t) = \frac{\Psi(\alpha, \alpha+1+\rho; (\lambda-t)z)}{\Psi(\alpha, \alpha+1+\rho; \lambda z)}.$$

Thus, the *cf* defined by $\psi(t) = E(\exp(itX))$ takes the form

$$\psi(t) = \frac{\Psi(\alpha, \alpha+1+\rho; (\lambda-it)z)}{\Psi(\alpha, \alpha+1+\rho; \lambda z)},$$

where $i = \sqrt{-1}$ is the complex number.

4. Rényi Entropy

An entropy of a random variable X is a measure of variation of the uncertainty. Rényi entropy is defined by

$$\mathcal{J}_R(\gamma) = \frac{1}{1-\gamma}\log\left\{\int f^\gamma(x)dx\right\},$$

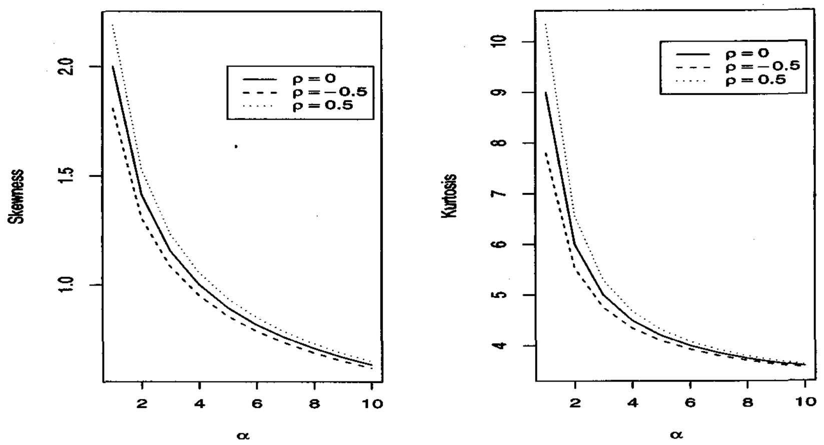

FIGURE 3.1 *Skewness and kurtosis measures versus* $\alpha = 1, 2, \ldots, 10$ *for the standard and generalized gamma distributions.*

where $\gamma > 0$ and $\gamma \neq 1$ (Rényi, 1961). For the *pdf* (1.2), note that one can write

$$
\begin{aligned}
f^{\gamma}(x) = & \frac{\Gamma(\alpha\gamma - \gamma + 1) z^{1-\gamma}}{\Gamma^{\gamma}(\alpha)} \frac{\Psi(\alpha\gamma - \gamma + 1, \alpha\gamma - \gamma + 2 + \gamma\rho; \lambda\gamma z)}{\Psi^{\gamma}(\alpha, \alpha + 1 + \rho; \lambda z)} \\
& \times \frac{x^{\alpha\gamma - \gamma}(x + z)^{\rho\gamma}}{z^{\alpha\gamma - \gamma + 1 + \gamma\rho} \Gamma(\alpha\gamma - \gamma + 1)} \\
& \times \frac{\exp(-\lambda\gamma x)}{\Psi(\alpha\gamma - \gamma + 1, \alpha\gamma - \gamma + 2 + \gamma\rho; \lambda\gamma z)}.
\end{aligned}
$$

Note that the last term integrates to 1 over $0 < x < \infty$. Hence, the Rényi entropy can be expressed as

$$
\begin{aligned}
\mathcal{J}_R(\gamma) = \log z + \frac{1}{1-\gamma} \log \Bigg\{ & \frac{\Gamma(\alpha\gamma - \gamma + 1)}{\Gamma^{\gamma}(\alpha)} \\
& \times \frac{\Psi(\alpha\gamma - \gamma + 1, \alpha\gamma - \gamma + 2 + \gamma\rho; \lambda\gamma z)}{\Psi^{\gamma}(\alpha, \alpha + 1 + \rho; \lambda z)} \Bigg\}. \qquad (4.1)
\end{aligned}
$$

Another well-known entropy measure is the Shannon entropy defined by $E[-\log f(X)]$. Its form can be determined by limiting $\gamma \uparrow 1$ in (4.1).

5. ASYMPTOTICS

If $X_1, \ldots, X_n$ is a random sample from (1.2) and if $\bar{X} = (X_1 + \cdots + X_n)/n$ denotes the sample mean then by the usual central limit theorem $\sqrt{n}(\bar{X} -$

$E(X))/\sqrt{\mathrm{Var}(X)}$ approaches the standard normal distribution as $n \to \infty$. Sometimes one would be interested in the asymptotics of the extreme values $M_n = \max(X_1, \ldots, X_n)$ and $m_n = \min(X_1, \ldots, X_n)$. Note from (1.2) that

$$1 - F(t) \sim \frac{z^{-\rho-\alpha} t^{\alpha+\rho-1} \exp(-\lambda t)}{\lambda \Gamma(\alpha) \Psi(\alpha, \alpha+1+\rho; \lambda z)} \tag{5.1}$$

as $t \to \infty$ and that

$$F(t) \sim \frac{t^\alpha}{\alpha z^\alpha \Gamma(\alpha) \Psi(\alpha, \alpha+1+\rho; \lambda z)}$$

as $t \to 0$. Thus, it follows that

$$\lim_{t\to\infty} \frac{1 - F(t + x/\lambda)}{1 - F(t)} = \exp(-x)$$

and

$$\lim_{t\to 0} \frac{F(tx)}{F(t)} = x^\alpha.$$

Hence, it follows from Theorem 1.6.2 in Leadbetter *et al.* (1987) that there must be norming constants $a_n > 0$, b_n, $c_n > 0$ and d_n such that

$$\Pr\{a_n (M_n - b_n) \le x\} \to \exp\{-\exp(-x)\}$$

and

$$\Pr\{c_n (m_n - d_n) \le x\} \to 1 - \exp(-x^\alpha)$$

as $n \to \infty$. The form of the norming constants can also be determined. For instance, using Corollary 1.6.3 in Leadbetter *et al.* (1987), one can see that $a_n = \lambda$ and that b_n satisfies $1 - F(b_n) \sim 1/n$ as $n \to \infty$. Using the fact (5.1), one can show that

$$b_n = -\frac{1}{\lambda} \log\left\{\frac{\Gamma(\alpha)\Psi(\alpha, \alpha+1+\rho; \lambda z)}{(\lambda z)^{-\rho-\alpha}}\right\} + \frac{1}{\lambda}\log n + \frac{\alpha+\rho-1}{\lambda}\log\log n$$

satisfies $1 - F(b_n) \sim 1/n$. The constants c_n and d_n can be determined by using the same corollary.

6. ESTIMATION

Here, we consider estimation of the parameters of (1.2) by the method of maximum likelihood and the method of moments. The log-likelihood for a random sample $x_1, \ldots, x_n$ from (1.2) is:

$$\begin{aligned}
&\log L(\alpha, \lambda, z, \rho) \\
&= -n \log K + (\alpha - 1) \sum_{i=1}^{n} \log x_i + \rho \sum_{i=1}^{n} \log (x_i + z) - \lambda \sum_{i=1}^{n} x_i, \qquad (6.1)
\end{aligned}$$

where $K = K(\alpha, \lambda, z, \rho) = z^{\alpha+\rho}\Gamma(\alpha)\Psi(\alpha, \alpha+1+\rho; \lambda z)$. The first order derivatives of (6.1) with respect to the four parameters are:

$$\begin{aligned}
\frac{\partial \log L}{\partial \alpha} &= -\frac{n}{K}\frac{\partial K}{\partial \alpha} + \sum_{i=1}^{n} \log x_i, \\
\frac{\partial \log L}{\partial \lambda} &= -\frac{n}{K}\frac{\partial K}{\partial \lambda} - \sum_{i=1}^{n} x_i, \\
\frac{\partial \log L}{\partial z} &= -\frac{n}{K}\frac{\partial K}{\partial z} + \rho \sum_{i=1}^{n} \frac{1}{x_i + z}, \\
\frac{\partial \log L}{\partial \rho} &= -\frac{n}{K}\frac{\partial K}{\partial \rho} + \sum_{i=1}^{n} \log (x_i + z).
\end{aligned}$$

Thus, the maximum likelihood estimators of the four parameters are the simultaneous solutions of the equations:

$$\frac{n}{K}\frac{\partial K}{\partial \alpha} = \sum_{i=1}^{n} \log x_i, \qquad (6.2)$$

$$\frac{n}{K}\frac{\partial K}{\partial \lambda} = -\sum_{i=1}^{n} x_i, \qquad (6.3)$$

$$\frac{n}{K}\frac{\partial K}{\partial z} = \rho \sum_{i=1}^{n} \frac{1}{x_i + z}, \qquad (6.4)$$

$$\frac{n}{K}\frac{\partial K}{\partial \rho} = \sum_{i=1}^{n} \log (x_i + z). \qquad (6.5)$$

We know from (1.4) that the constant K involves the confluent hypergeometric function and numerical routines are widely available for evaluating this function.

The method of moments estimators of the four parameters are the simultaneous solutions of the equations:

$$\frac{z\Gamma(1+\alpha)\Psi\left(1+\alpha, 2+\alpha+\rho; \lambda z\right)}{\Gamma(\alpha)\Psi\left(\alpha, \alpha+1+\rho; \lambda z\right)} = s_1,$$

$$\frac{z^2\Gamma(2+\alpha)\Psi\left(2+\alpha, 3+\alpha+\rho; \lambda z\right)}{\Gamma(\alpha)\Psi\left(\alpha, \alpha+1+\rho; \lambda z\right)} = s_2,$$

$$\frac{z^3\Gamma(3+\alpha)\Psi\left(3+\alpha, 4+\alpha+\rho; \lambda z\right)}{\Gamma(\alpha)\Psi\left(\alpha, \alpha+1+\rho; \lambda z\right)} = s_3,$$

$$\frac{z^4\Gamma(4+\alpha)\Psi\left(4+\alpha, 5+\alpha+\rho; \lambda z\right)}{\Gamma(\alpha)\Psi\left(\alpha, \alpha+1+\rho; \lambda z\right)} = s_4,$$

where $s_k = (1/n)\sum_{i=1}^{n} x_i^k$ is the k^{th} sample moment, $k = 1, 2, 3, 4$.

For interval estimation of $(\alpha, \lambda, z, \rho)$ and tests of hypothesis, one requires the Fisher information matrix. The second order derivatives of (6.1) with respect to the four parameters are:

$$\frac{\partial^2 \log L}{\partial \alpha^2} = -\frac{n}{K}\frac{\partial^2 K}{\partial \alpha^2} + \frac{n}{K^2}\left(\frac{\partial K}{\partial \alpha}\right)^2,$$

$$\frac{\partial^2 \log L}{\partial \alpha \partial \lambda} = -\frac{n}{K}\frac{\partial^2 K}{\partial \alpha \partial \lambda} + \frac{n}{K^2}\frac{\partial K}{\partial \alpha}\frac{\partial K}{\partial \lambda},$$

$$\frac{\partial^2 \log L}{\partial \alpha \partial z} = -\frac{n}{K}\frac{\partial^2 K}{\partial \alpha \partial z} + \frac{n}{K^2}\frac{\partial K}{\partial \alpha}\frac{\partial K}{\partial z},$$

$$\frac{\partial^2 \log L}{\partial \alpha \partial \rho} = -\frac{n}{K}\frac{\partial^2 K}{\partial \alpha \partial \rho} + \frac{n}{K^2}\frac{\partial K}{\partial \alpha}\frac{\partial K}{\partial \rho},$$

$$\frac{\partial^2 \log L}{\partial \lambda^2} = -\frac{n}{K}\frac{\partial^2 K}{\partial \lambda^2} + \frac{n}{K^2}\left(\frac{\partial K}{\partial \lambda}\right)^2,$$

$$\frac{\partial^2 \log L}{\partial \lambda \partial z} = -\frac{n}{K}\frac{\partial^2 K}{\partial \lambda \partial z} + \frac{n}{K^2}\frac{\partial K}{\partial \lambda}\frac{\partial K}{\partial z},$$

$$\frac{\partial^2 \log L}{\partial \lambda \partial \rho} = -\frac{n}{K}\frac{\partial^2 K}{\partial \lambda \partial \rho} + \frac{n}{K^2}\frac{\partial K}{\partial \lambda}\frac{\partial K}{\partial \rho},$$

$$\frac{\partial^2 \log L}{\partial z^2} = -\frac{n}{K}\frac{\partial^2 K}{\partial z^2} + \frac{n}{K^2}\left(\frac{\partial K}{\partial z}\right)^2 - \rho\sum_{i=1}^{n}\frac{1}{(x_i+z)^2},$$

$$\frac{\partial^2 \log L}{\partial z \partial \rho} = -\frac{n}{K}\frac{\partial^2 K}{\partial z \partial \rho} + \frac{n}{K^2}\frac{\partial K}{\partial z}\frac{\partial K}{\partial \rho} + \sum_{i=1}^{n}\frac{1}{x_i+z},$$

$$\frac{\partial^2 \log L}{\partial \rho^2} = -\frac{n}{K}\frac{\partial^2 K}{\partial \rho^2} + \frac{n}{K^2}\left(\frac{\partial K}{\partial \rho}\right)^2.$$

Thus, using the fact

$$E\left[\frac{1}{(X+z)^m}\right] = \frac{1}{z^m}\frac{\Psi(\alpha, \alpha+1+\rho-m; \lambda z)}{\Psi(\alpha, \alpha+1+\rho; \lambda z)},$$

the elements of the Fisher information matrix can be calculated as

$$E\left(-\frac{\partial^2 \log L}{\partial \alpha^2}\right) = \frac{n}{K}\frac{\partial^2 K}{\partial \alpha^2} - \frac{n}{K^2}\left(\frac{\partial K}{\partial \alpha}\right)^2,$$

$$E\left(-\frac{\partial^2 \log L}{\partial \alpha \partial \lambda}\right) = \frac{n}{K}\frac{\partial^2 K}{\partial \alpha \partial \lambda} - \frac{n}{K^2}\frac{\partial K}{\partial \alpha}\frac{\partial K}{\partial \lambda},$$

$$E\left(-\frac{\partial^2 \log L}{\partial \alpha \partial z}\right) = \frac{n}{K}\frac{\partial^2 K}{\partial \alpha \partial z} - \frac{n}{K^2}\frac{\partial K}{\partial \alpha}\frac{\partial K}{\partial z},$$

$$E\left(-\frac{\partial^2 \log L}{\partial \alpha \partial \rho}\right) = \frac{n}{K}\frac{\partial^2 K}{\partial \alpha \partial \rho} - \frac{n}{K^2}\frac{\partial K}{\partial \alpha}\frac{\partial K}{\partial \rho},$$

$$E\left(-\frac{\partial^2 \log L}{\partial \lambda^2}\right) = \frac{n}{K}\frac{\partial^2 K}{\partial \lambda^2} - \frac{n}{K^2}\left(\frac{\partial K}{\partial \lambda}\right)^2,$$

$$E\left(-\frac{\partial^2 \log L}{\partial \lambda \partial z}\right) = \frac{n}{K}\frac{\partial^2 K}{\partial \lambda \partial z} - \frac{n}{K^2}\frac{\partial K}{\partial \lambda}\frac{\partial K}{\partial z},$$

$$E\left(-\frac{\partial^2 \log L}{\partial \lambda \partial \rho}\right) = \frac{n}{K}\frac{\partial^2 K}{\partial \lambda \partial \rho} - \frac{n}{K^2}\frac{\partial K}{\partial \lambda}\frac{\partial K}{\partial \rho},$$

$$E\left(-\frac{\partial^2 \log L}{\partial z^2}\right) = \frac{n}{K}\frac{\partial^2 K}{\partial z^2} - \frac{n}{K^2}\left(\frac{\partial K}{\partial z}\right)^2 + \frac{n\rho\Psi(\alpha, \alpha+\rho-1; \lambda z)}{z^2\Psi(\alpha, \alpha+1+\rho; \lambda z)},$$

$$E\left(-\frac{\partial^2 \log L}{\partial z \partial \rho}\right) = \frac{n}{K}\frac{\partial^2 K}{\partial z \partial \rho} - \frac{n}{K^2}\frac{\partial K}{\partial z}\frac{\partial K}{\partial \rho} - \frac{n\Psi(\alpha, \alpha+\rho; \lambda z)}{z\Psi(\alpha, \alpha+1+\rho; \lambda z)},$$

$$E\left(-\frac{\partial^2 \log L}{\partial \rho^2}\right) = \frac{n}{K}\frac{\partial^2 K}{\partial \rho^2} - \frac{n}{K^2}\left(\frac{\partial K}{\partial \rho}\right)^2.$$

7. SIMULATION

Here, we consider simulating from (1.2) by the rejection method with envelope g (say). It is well known that the scheme for simulating is given by:

Step 1. Simulate $X = x$ from the *pdf* g.

Step 2. Simulate $Y = UMg(x)$, where U is an independent uniform random variable on $(0, 1)$ and M is taken such that $f(x)/g(x) < M$ for all x.

Step 3. Accept $X = x$ as a realization of a random variable with the *pdf* f if $Y < f(x)$. If $Y \geq f(x)$ return to Step 2.

For the *pdf* (1.2), one can choose the envelope to be

$$g(x) = \begin{cases} F(z)g_1(x), & \text{if } 0 < x \leq z, \\ (1 - F(z))\, g_2(x), & \text{if } z < x < \infty, \end{cases}$$

where

$$g_1(x) = \frac{\alpha x^{\alpha-1}}{z^\alpha}$$

and

$$g_2(x) = \frac{\lambda^{\alpha+\rho} x^{\alpha+\rho-1} \exp(-\lambda x)}{\Gamma(\alpha+\rho, z)},$$

where $\Gamma(\cdot, \cdot)$ denotes the complementary incomplete gamma function defined by

$$\Gamma(\alpha, z) = \int_z^\infty \exp(-t) t^{\alpha-1} dt.$$

Note that g_1 is a power-function *pdf* while g_2 is the *pdf* of a truncated gamma distribution. Standard calculations show that the constant M can be chosen to be

$$M = \frac{\max(1, 2^\rho)}{\Gamma(\alpha)\Psi(\alpha, \alpha+1+\rho; \lambda z)} \max\left[\frac{1}{\alpha F(z)}, \frac{\Gamma(\alpha+\rho, z)}{(\lambda z)^{\alpha+\rho}\{1 - F(z)\}}\right].$$

Hence, simulating from (1.2) amounts to simulating from the power-function and truncated gamma distributions for which methods are widely available.

8. APPLICATION

Here, we return to the drought problem discussed in Section 1 and provide an application of the model given by (1.2). The drought data from the State of Nebraska is used, freely downloadable from the web–site:

http://lwf.ncdc.noaa.gov/oa/climate/onlineprod/drought/xmgrg3.html.

TABLE 8.1 *Basic drought statistics for Nebraska PDSI data*

Climate division	*Number of droughts*	*Drought frequency (number/year)*	*Mean drought duration (months)*	*Standard deviation of drought duration (months)*
1	83	0.75	6.0	8.0
2	66	0.60	8.6	12.0
3	89	0.81	6.3	9.7
5	81	0.74	6.3	10.5
6	90	0.82	6.3	10.1
7	81	0.74	6.1	9.7
8	76	0.69	6.5	13.4
9	74	0.67	7.5	10.9

TABLE 8.2 *Parameter estimates of (1.1) for drought data*

Climate division	$\hat{\lambda}$	$\hat{\alpha}$
1	0.038	0.419
2	0.022	0.381
3	0.030	0.369
5	0.027	0.372
6	0.031	0.390
7	0.034	0.407
8	0.028	0.394
9	0.026	0.409

The data comprises of the monthly modified Palmer Drought Severity Index (PDSI) from the period from January 1895 to December 2004. A drought is said to have happened when PDSI is below 0 and is defined by the theory of runs (Yevjevich, 1967). The State of Nebraska is divided into eight climate divisions numbered 1, 2, 3, 5, 6, 7, 8 and 9—there is no climate division 4 for Nebraska. Some statistics of the observed drought for the eight climatic divisions are summarized in Table 8.1.

Using the PDSI data, data on drought intensity were obtained for each climate division. The gamma distribution given by (1.1) has been the traditional model for drought intensity data (Shiau *et al.*, 2006). Here, we show that the generalization given by (1.2) provides a significant improvement. We fitted the models given by (1.1) and (1.2) to the observed drought intensity data from each climate division. The fitting of the models was performed by the method of maximum likelihood by solving the equations (6.2)–(6.5). The quasi-Newton algorithm nlm in the R software package (Dennis and Schnabel, 1983; Schnabel *et al.*, 1985;

TABLE 8.3 *Parameter estimates of (1.2) for drought data*

Climate division	$\hat{\lambda}$	$\hat{\alpha}$	$\hat{z}$	$\hat{\rho}$
1	0.011	3.473	0.071	-3.654
2	0.010	59.316	0.001	-59.296
3	0.005	1.161	0.555	-1.598
5	0.000	0.789	3.738	-1.816
6	0.002	1.155	1.249	-1.890
7	0.003	1.396	0.858	-2.074
8	0.000	1.106	2.548	-2.184
9	0.004	1.019	1.522	-1.550

TABLE 8.4 *p values of the Likelihood Ratio Test*

Climate division	*NLLH for (1.1)*	*NLLH for (1.2)*	*p-value*
1	254.656	239.693	0.000000
2	224.629	215.403	0.000098
3	266.869	249.179	0.000000
5	254.966	237.529	0.000000
6	277.862	256.113	0.000000
7	251.247	231.099	0.000000
8	243.325	220.443	0.000000
9	248.608	236.953	0.000009

Ihaka and Gentleman, 1996) was used to solve the equations. The possibility of detecting local maxima was avoided by the execution of the algorithm with different starting values.

The parameter estimates for (1.1) and (1.2) are given in Tables 8.2 and 8.3, respectively. The negative logarithm of the maximized likelihoods (NLLH) are shown in Table 8.4. We performed a standard likelihood ratio test of $z = 0$ and $\rho = 0$ by comparing twice the difference of the two negative logarithms with a chi square distribution with two degrees of freedom. The p values of this test are also shown in Table 8.4. The p values are much much smaller than the nominal 0.05. It follows that the fit of (1.2) is significantly better than that of (1.1) for each of the climate divisions.

The goodness of fits of (1.1) and (1.2) can be examined by probability plots. A probability plot is where the observed probability is plotted against the probability predicted by the fitted model. To check the goodness of fit given by (1.1), one would plot $F(x_{(i)})$ versus $(i-0.375)/(n+0.25)$, where $F(\cdot)$ is the cdf corresponding to (1.1) and $x_{(i)}$ are the sorted values, in the ascending order, of the observed

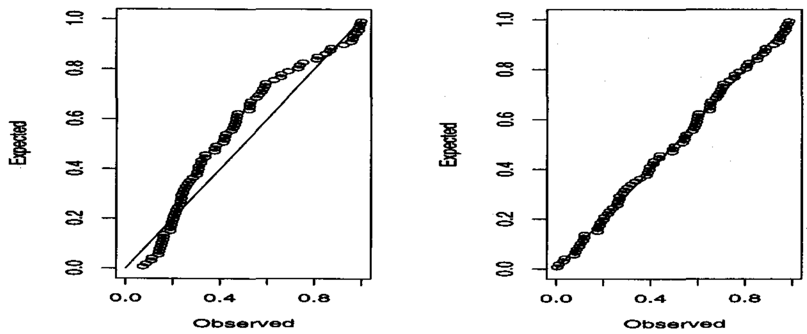

FIGURE 8.1 *Probability plots on the fits of (1.1) and (1.2) for drought intensity data from climate division 1 of Nebraska.*

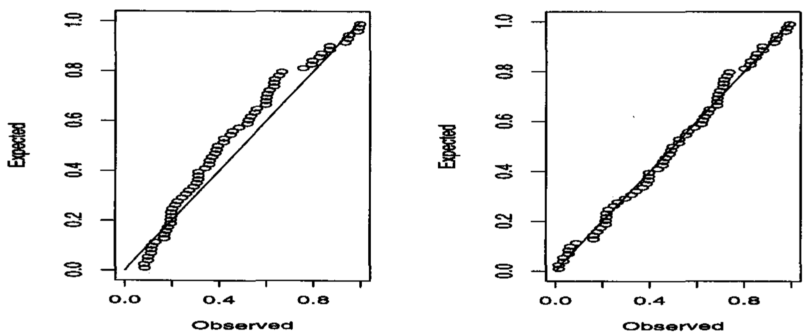

FIGURE 8.2 *Probability plots on the fits of (1.1) and (1.2) for drought intensity data from climate division 2 of Nebraska.*

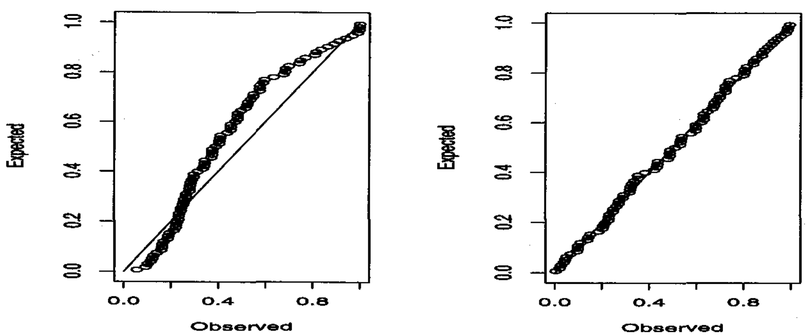

FIGURE 8.3 *Probability plots on the fits of (1.1) and (1.2) for drought intensity data from climate division 3 of Nebraska.*

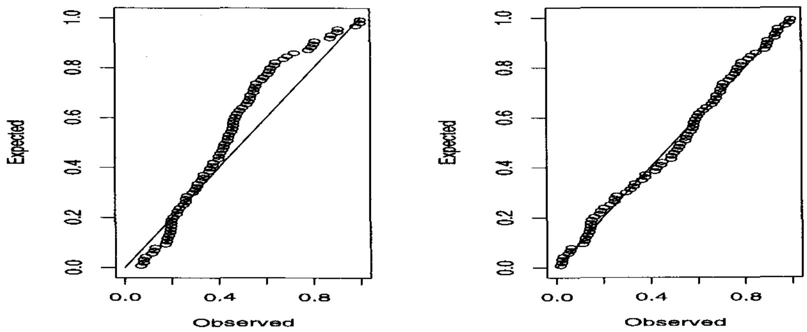

FIGURE 8.4 *Probability plots on the fits of (1.1) and (1.2) for drought intensity data from climate division 5 of Nebraska.*

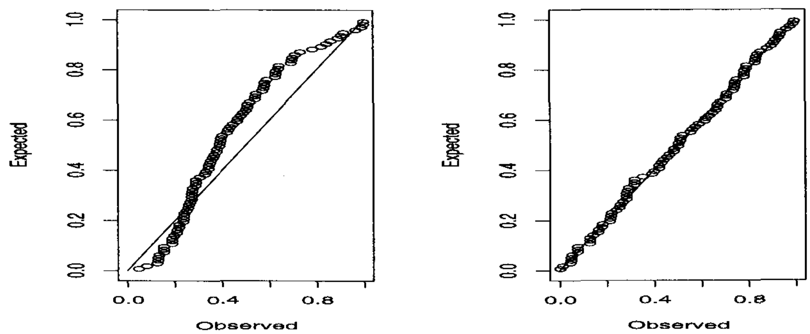

FIGURE 8.5 *Probability plots on the fits of (1.1) and (1.2) for drought intensity data from climate division 6 of Nebraska.*

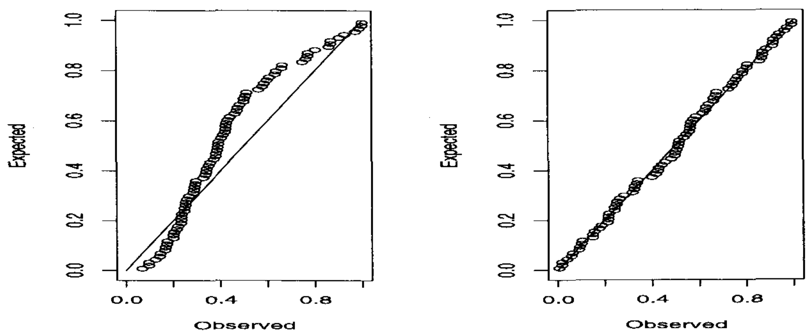

FIGURE 8.6 *Probability plots on the fits of (1.1) and (1.2) for drought intensity data from climate division 7 of Nebraska.*

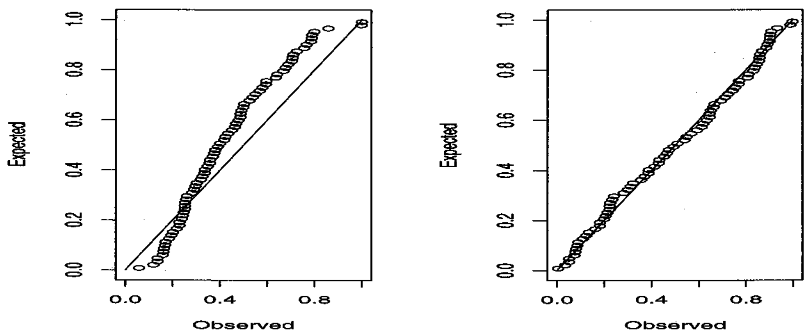

FIGURE 8.7 *Probability plots on the fits of (1.1) and (1.2) for drought intensity data from climate division 8 of Nebraska.*

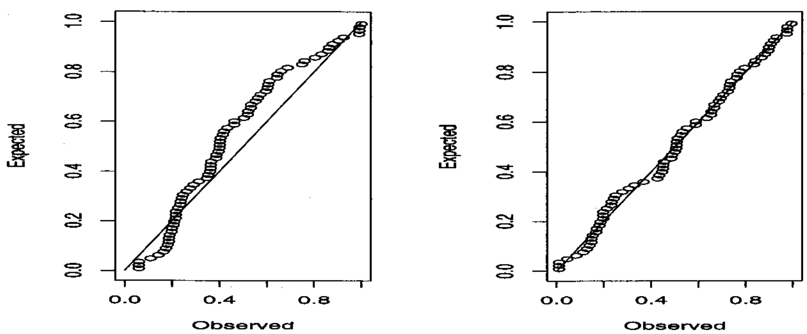

FIGURE 8.8 *Probability plots on the fits of (1.1) and (1.2) for drought intensity data from climate division 9 of Nebraska.*

drought intensity data. To check the goodness of fit given by (1.2), one would plot $F(x_{(i)})$ versus $(i-0.375)/(n+0.25)$, where $F(\cdot)$ is the cdf corresponding to (1.2). These plots for the eight climate divisions are shown in Figures 8.1 to 8.8 The plots on the left and right correspond to the gamma and generalized gamma distributions, respectively. It is evident that the generalized gamma distribution provides an excellent fit for each climate division. Furthermore, its fit appears significantly better than that of the gamma distribution at least visually. This finding is consistent with the conclusions from the likelihood ratio test.

ACKNOWLEDGMENTS

The authors would like to thank the Editor–in–Chief and the two referees for carefully reading the paper and for their great help in improving the paper.

REFERENCES

DENNIS, J. E. AND SCHNABEL, R. B. (1983). *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*, Prentice-Hall, Englewood Cliffs, New Jersey.

GUPTA, R. C. AND BROWN, N. (2001). "Reliability studies of the skew-normal distribution and its application to a strength-stress model", *Communications in Statistics. Theory and Methods*, **30**, 2427–2445.

IHAKA, R. AND GENTLEMAN, R. (1996). "R: A language for data analysis and graphics", *Journal of Computational and Graphical Statistics*, **5**, 299–314.

LEADBETTER, M. R., LINDGREN, G. AND ROOTZÉN, H. (1987). *Extremes and Related Properties of Random Sequences and Processes*, Springer-Verlag, New York/ Berlin.

PRUDNIKOV, A. P., BRYCHKOV, Y. A. AND MARICHEV, O. I. (1986). *Integrals and Series*, Gordon and Breach Science Publishers, Amsterdam.

RÉNYI, A. (1961). "On measures of entropy and information", *Proceedings of the 4th Berkeley Symposium on Mathematical Statistics and Probability*, **1**, 547–561, University of California Press, Califonia.

SCHNABEL, R. B., KOONTZ, J. E. AND WEISS, B. E. (1985). "A modular system of algorithms for unconstrained minimization", *Association for Computing Machinery. Transactions on Mathematical Software*, **11**, 419–440.

SHIAU, J.-T., FENG, S. AND NADARAJAH, S. (2006). "Assessment of hydrological droughts for the Yellow River, China using copulas", In *Hydrological Processes.*

YEVJEVICH, V. (1967). "An objective approach to definitions and investigations of continental hydrologic droughts", *Hydrologic Paper*, **23**, Colorado State University, Fort Collins.

YUE, S. (2001). "A bivariate gamma distribution for use in multivariate flood frequency analysis", *Hydrological Processes*, **15**, 1033–1045.

YUE, S., OUARDA, T. B. M. J. AND BOBEE, B. (2001). "A review of bivariate gamma distributions for hydrological application", *Journal of Hydrology*, **246**, 1–18.