A proposed reparametrization of gamma distribution for the analysis of data of rainfall-runoff driven pollution

B. LAGOS ÁLVAREZ UNIVERSIDAD DE CONCEPCIÓN, CHILE

G. FERREIRA UNIVERSIDAD DE CONCEPCIÓN, CHILE and M. VALENZUELA HUBE UNIVERSIDAD DE CONCEPCIÓN, CHILE

Received : March 2011. Accepted : October 2011

Abstract

A generalized gamma (GG) distribution of four parameters was first introduced by Amoroso 1925, and since then, different distributions emerged as subclasses of this model. This model is commonly used to model lifetime data or data with a right skewed unimodal density function. In this article, we use a reparameterization of the GG distribution that is compared with other usual two-parameter distributions, Weibull, generalized exponential (Gupta and Kundu 1999), and gamma, using a real data set with a high coefficient of asymmetry and kurtosis (Valenzuela M. 2009). Akaike's information criterion and Bayesian information criterion indicates that our reparametrization of the gamma distribution is better. Besides a Monte Carlo simulation study, shows the behavior of five estimation methods: least squared, weighted least squared, moments, probability weighted moments and maximum likelihood methods.

Keywords : Gamma distribution, Maximum likelihood estimators, Moment estimators, Probability weighted moment estimators, Weighted least squares estimators, Water pollution and watershed.

1. Introduction

Environmental pollution, especially water pollution, in the last decades has become more serious all over the world. Non-point source pollution has become increasingly important for the study of water quality and pollutant transport processes within river watersheds. The dominant process leading to the spatial redistribution of pollutants and contamination of water bodies is surface flow. In data from the watersheds, one problem is to determine a probability distribution model of pollutants in water and soil to provide reliable features closer to the real data set. Few studies concerning with the distribution of pollution were carried out for the distribution of rainfallrunoff driven pollution. Hydrologic data are typically skewed, and so are soil contaminant concentration data from sites which behavior is rain-fall dependent. The presence of one or more outliers is the rule rather the exception in a data set which appears to come from this area. The impact of outliers values can be as powerful that some of them can completely dominate the estimate of the interesting quantities such as mean contamination levels at various areas of a polluted site. The modeling problem becomes more complex when a small number of high concentration of pollutants is observed.

In order to solve the above problem we propose a reparameterization of the GG distribution. This reparameterization leads to a gamma distribution with two parameters that is compared with the known gamma distribution, Weibull distribution, and generalized exponential (GE) distribution, our proposed appears to be the most appropriate for our data set.

We begin by describing the two-parameter gamma distribution or simply gamma distribution, it has been used as a universal statistical law for various complex stochastic events in various fields, such as, physical, chemical and biological systems, among others. This distributions can be written as

(1.1)
$$f(x;\alpha,\beta) = \frac{x^{\alpha-1}}{\beta^{\alpha} \Gamma(\alpha)} e^{-x/\beta}, \ x > 0$$

where $\Gamma(z) = \int_0^\infty y^{z-1} e^{-z} dy$ is the gamma function at z > 0, and α , β are positive parameters, the notation of this distribution will use $X \sim G(\alpha, \beta)$.

The gamma distribution, which covers a wide range of skewness, has been tried. It is seen that a gamma probability model shows a good fit to the rainfall-related variables, for example; in the article of Mooley and Appa Rao (1971), is examined the adjustment of gamma distribution to the annual precipitations in India covering several climatic regimes. They demonstrated that annual precipitation adjusted to a gamma distribution, in another example, mean and maximum rainfall associated to cyclonic storms of the Bay of Bengal striking the two sections of the east coast of India were studied by Mooley and Mohile (1986) examining the rainfall distributions along the coast on both sides of a point P at which the storms struck the coast at a right angle during the period 1877–1980. The distribution is highly skewed and Suhaila 1998 have used this distribution to fit rainfall intensity for Peninsular Malaysia using hourly rainfall data.

Recently, new releases have emerged as competitors to the gamma distribution, for example the GE distribution or exponentiated exponential distribution, has been introduced and studied quite and extensively by Gupta and Kundu (1999) and Gupta and Kundu (2001).We consider a subclass of the family of Suzuki (1964) distribution that is different to the two-parameter gamma distribution and different to the three-parameter gamma distribution proposed by Stacy (1962) and is not incorporated in the list of Crooks (2010).

The objective of the present article is to develop based-likelihood inference for the parameters of the reparametirized gamma distribution which yield reliable estimates of parameters from data sets in the presence of outliers, and also to compare the performances of various estimation procedures. The five different estimation methods for the distribution parameters considered, are maximum likelihood estimators, method of moment estimators, weighted least squares estimators, least squares and probability weighted moment estimators. These procedures have not been extensively studied for this reparameterized gamma distribution, to which we aim to assess its ability to explain the pollution data observed in response to rainfall-runoff phenomena from watersheds.

This paper is organized as follows. In Section 2, we define the proposed reparametrization for the distribution of Suzuki (1964) and its properties. In Section 3, we introduce five different estimation methods for the parameter of the distribution considered. Section 4 describes the asymptotic inference of parameters via likelihood function. In Section 5, we present a study via simulation comparing the performance of the different estimation methods. Section 6 illustrates the Akaike's information criterion (AIC) and Bayesian information criterion (BIC) for the Weibull distribution, GE distribution, gamma two parameter and the proposed distribution considering a real data set. Finally, we draw some conclusions in section 7. The ap-

pendix shows the tables of the statistics mean and variance coefficient for the variable following the proposed distribution.

2. The Model

A original form of probability density function (pdf) of the GG distribution of three positive parameters a, d and p, is given by

(2.1)
$$f(x; a, d, p) = \frac{p \ x^{d-1}}{a^d \Gamma(d/p)} e^{-(x/a)^p}, \quad x \ge 0$$

with their cumulative distribution function, which does not have a closed formula, is given by

(2.2)
$$F(x; a, d, p) = \int_0^x f(y; a, d, p) dy = \frac{\gamma((x/a)^p, d/p)}{\Gamma(d/p)}$$

where $\gamma(s, x) = \int_0^x t^{s-1} e^{-t} dt$ denotes the incomplete gamma function, see Stacy (1962) and Johnson et. al. (1994). If a random variable X has a pdf as in the equation (2.2), ... then will be denoted as $X \sim GG_1(a, d, p)$.

A particular parametrization the pdf in (2.2), leads to the well-known density function proposed by Suzuki (1964), this parametrization is done by the following transformation $\delta = p$, $\nu = d$ and $\eta = 1/a^p$, then we have

(2.3)
$$f(x;\eta,\nu,\delta) = \frac{\delta \eta^{\nu/\delta} x^{\nu-1}}{\Gamma(\nu/\delta)} e^{-\eta x^{\delta}}, \quad x > 0.$$

We use the following notation $GG_2(\eta, \nu, \delta)$ for the distribution given by (2.3), that satisfies $GG_2(\eta, \nu, \delta) := GG_1(1/\eta^{1/\delta}, \nu, \delta)$.

Stacy and Mihram (1965) manages the estimation of parameters through method moments of log X incorporating the transformation k = d/p. On the other hand we obtain another form commonly used in practice making the parametrization in (2.2); k = d/p, $\beta = a$ and $\psi = p$, so

(2.4)
$$f(x;k,\beta,\psi) = \frac{\psi \ x^{k\psi-1}}{\beta^{k\psi}\Gamma(k)}e^{-(x/\beta)^{\psi}}, \quad x > 0.$$

For the distribution given by the pdf in (2.4) we use the notation $GG_3(k,\beta,\psi)$, and therefore we have $GG_3(k,\beta,\psi) := GG_2(1/\beta^{\psi},k\psi,\psi) := GG_1(\beta,k\psi,\psi)$. Moreover using the notations above we can write GG_2 in terms of GG_3 as $GG_2(\eta,\nu,\delta) := GG_3(\nu/\delta,1/\eta^{1/\delta},\delta)$. The cumulative function distribution of $GG_2(\eta,\nu,\delta)$ evaluated at $x, F(x;\eta,\nu,\delta)$, can be written as $F(x;\eta,\nu,\delta) = \gamma(\eta x^{\delta},1/\delta)/\Gamma(\nu/\delta)$.

As discussed above, different parameterizations of the function (2.2) have been carried out, we will use a new parametrization given by $\eta = 1/\rho$, $\nu = 1$, and $\delta = 1/\alpha$. This parametrization allows us fix a parameter, so the pdf of Suzuki (1964) becomes a function of density of two parameters $GG_2(1/\rho, 1, 1/\alpha)$, where the pdf is given by

(2.5)
$$f(x;\eta,\nu,\delta) = \frac{1}{\alpha \Gamma(\alpha)} e^{-x^{1/\alpha}/\rho}, \quad x > 0.$$

The cumulative function distribution of $GG_2(1/\rho, 1, 1/\alpha)$ evaluated at x, can be written as $F(x; \rho, \alpha) = \gamma(x^{1/\alpha}/\rho, \alpha)/\Gamma(\alpha)$.

We will compare this distribution with distributions of two parameters such as Weibull, $G(\alpha, \beta)$ and GE distribution. The properties of the new density function generated by this parametrization will be discussed in Section 4. For purposes of simplicity, all properties will be described under the pdf of Suzuki i. e. we will use the density function $GG_2(\eta, \nu, \delta)$.

2.1. Properties

In this section we will study the particular case of the parametrization proposed by Suzuki, hence we will consider the form (2.3) to study the properties of GG_2 utilizing the relation $\int_0^\infty x^p e^{-\beta x^q} = \Gamma(m)/(q\beta^m)$, with m = (p+1)/q in Gradshteyn and Ryzhik (2007) and the power series of the function $g(z) = e^z$ we obtain the moment generating function

(2.6)
$$\mathcal{M}(t) = E\{e^{tX}\} = \frac{1}{\Gamma(\nu/\delta)} \sum_{k=0}^{\infty} \frac{t^k}{k!} \frac{\Gamma(\frac{\nu+k}{\delta})}{\alpha \eta^{k/\delta}}.$$

From (2.6) the not central moments, until order r, are given by

(2.7)
$$\mu_r := E\{X^r\} = \frac{1}{\eta^{r/\delta}} \frac{\Gamma((\nu+r)/\delta)}{\Gamma(\nu/\delta)},$$

with $r \in \mathbf{Z}^+$.

Therefore, the calculation of mean and variance follow immediately from the formula (2.7)

$$E\{X\} = \frac{1}{\eta^{1/\delta}} \frac{\Gamma((\nu+1)/\delta)}{\Gamma(\nu/\delta)},$$

 $\operatorname{Var}\{X\} = \frac{1}{\eta^{2/\delta} \Gamma(v/\delta)} \left[\Gamma\left(\left(v+2\right)/\delta \right) - \Gamma^2\left(\left(v+1\right)/\delta \right)/\Gamma\left(v/\delta \right) \right]$ (2.8)

We observed that, if $X = Y/\eta^{1/\delta}$, with $Y \sim GG_2(\nu, 1, \delta)$, then

(2.9)
$$\left[E\{Y\}/E\{X\}\right]^{\delta} = \eta$$

The coefficient of variation of X is given by

(2.10)
$$CV\{X\} = \sqrt{\Gamma(\nu/\delta)} \sqrt{\frac{\Gamma((\nu+2)/\delta)}{\Gamma^2((\nu+1)/\delta)} - \frac{1}{\Gamma(\nu/\delta)}}$$

with

$$\begin{split} &\Gamma((\nu+2)/\delta)>\Gamma^2((\nu+1)/\delta)/\Gamma(\nu/\delta), \, \text{for } \delta>0.\\ &\text{The skewness, } \beta_1\,=\,\mu_3^\bullet/\mu_2^{\bullet 3/2} \, \text{ and kurtosis } \beta_2\,=\,\mu_4^\bullet/\mu_2^{\bullet 2} \, \text{ coefficients,} \end{split}$$
where the symbol "•" denotes the moment centered around the mean, which can be calculated as

$$\beta_1 = \frac{\mu_3 - 2\mu_1\mu_2 + 2\mu_1^2}{(\mu_2 - \mu_1^2)^{3/2}}, \quad \beta_2 = \frac{(\mu_4 - \mu_1(4\mu_3 - 6\mu_1\mu_2 + 3\mu_1^3))}{(\mu_2 - \mu_1^2)^2},$$

where μ_i , i = 1, 2, 3, 4 are the first, second, third and fourth moments given in relation (2.9).

If $X \sim GG_3(k, \beta, \psi)$, the stochastic representation of $Y \sim GG_3(k, \beta^{\psi}, 1)$ in function of X can be obtained following the relation $Y = X^{\psi}$, and for $V \sim GG_3(k, 1, 1)$, the standard gamma distribution, in function of X, can be written as $V = (X/\beta)^{\psi}$. So, the generation of a simple sample random of finite size from $X = GG_3(k, \beta, \psi)$ may be obtained from $GG_3(k, \beta^{\psi}, 1)$, as $X = Y^{1/\psi}$ or from $GG_3(k, 1, 1)$ as $X = \beta V^{1/\psi}$. On the other hand, in the case of our parametrization, if we have $X \sim G(1/\eta, \nu/\delta)$ the stochastic representation of $Y \sim GG_2(\eta, \nu, \delta)$ in function of X can be obtained following the relation $Y = X^{1/\delta}$. So any feature of interest for Y can be obtained from X with a usual two-parameter gamma distribution, specifically $a = 1/\eta$ and $d = \nu/\delta$. The algorithm to compute pseudo random number from $GG_3(k, 1, 1)$, $GG_3(k, \beta^{\psi}, 1)$ or $G(1/\eta, \nu/\delta)$ is been developed on different statistics softwares.

3. Estimation methods

In this section we study the behavior of the five appropriate estimate methods that can be apply to us distributions with a highly skewed real data set. Such as; moments, least squares, weighted least squares, probability weighted moment and maximum likelihood, under the particular $GG_2(\eta, \nu, \delta)$ distributions. We expect that these robust procedures, as weighted least squares, and probability weighted moment, have a better performance that moments method, least square and maximum likelihood.

3.1. Moment estimation method

The moments estimation method can be obtained from the no-central moments of $X \sim GG_2(\eta, \nu, \delta)$ given in the equations (2.9), solving for ν , η and δ , besides we used the equations of Konn and Lomdahl (2004) to obtain,

$$\frac{\mu_1^2}{\mu_2} = \frac{\Gamma^2(\frac{\nu+1}{\delta})}{\Gamma(\frac{\nu}{\delta})\Gamma(\frac{\nu+2}{\delta})}, \qquad \frac{\mu_1^3}{\mu_3} = \frac{\Gamma^3(\frac{\nu+1}{\delta})}{\Gamma^2(\frac{\nu}{\delta})\Gamma(\frac{\nu+3}{\delta})}, \qquad \mu_1\eta^{1/\delta} = \frac{\Gamma(\frac{\nu+1}{\delta})}{\Gamma(\frac{\nu}{\delta})}.$$
(3.1)

So it could be used to estimate the initial point to obtain the least squares, weighted least squares, probability weighted moment method and estimators maximum likelihood estimation, which are described in the next subsection.

Or finding the root, say $\tilde{\nu}$ and $\tilde{\delta}$ for the nonlinear equation

(3.2)
$$CV\{X\} - \frac{S_X}{\overline{X}} = 0.$$

And then, replacing the $\tilde{\nu}$ and $\tilde{\delta}$ at the last expression (3.1), we have for $\tilde{\eta}$

(3.3)
$$\widetilde{\eta} = exp \bigg\{ \widetilde{\delta} \bigg[log(\Gamma((1+\widetilde{\nu})/\widetilde{\delta})) - log(\Gamma(\widetilde{\nu}/\widetilde{\delta})) - log(\overline{X}) \bigg] \bigg\},$$

being the moment estimator for η . Or using the relation (11), we have

(3.4)
$$\widetilde{\eta} = \left[\frac{E_{(\widetilde{\nu},\widetilde{\delta})}\{Y\}}{\overline{X_n}}\right]^{\widetilde{\delta}}$$

where $E_{(\tilde{\nu},\tilde{\delta})}$ is the expectation on $GG_2(1,\tilde{\nu},\tilde{\delta})$ distribution.

3.2. Least squares and weighted least squares estimators

In this subsection we present the regression-based method estimators for the unknown parameters of the GG_2 distribution. The method of ordinary least squares and the method of weighted least squares (Swain et. al., 1988) that is defined in term of the order statistics making of it procedure a robust alternative methods, useful for the estimation of parameters of a skewed distribution, that is our case. The methodology is as follows:

Let $X_{1:n} \leq X_{2:n} \leq \cdots \leq X_{n:n}$ be the order statistics of X_1, \ldots, X_n , with $X_i \sim GG_2(\eta, \nu, \delta)$. Observe that $F(X_{r:n})$ behaves like the *r*-th order statistic of a sample of size *n* from the uniform distribution (0; 1), so $U_{r:n} \sim Beta(n+2, r+1)$ and

(3.5)
$$E\{U_{r:n}\} = r/(n+1), \quad Var\{U_{r:n}\} = \frac{r(n-r+1)}{(n+1)^2(n+2)},$$

and

(3.6)
$$Cov\{U_{r:n}, U_{s:n}\} = \frac{r(n-s+1)}{(n+1)^2(n+2)}, \text{ for } r < s.$$

The weighted least squares estimators (WLSE) of the unknown parameters (η, ν, δ) can be obtained by

(3.7)
$$(\overline{\eta}, \overline{\nu}, \overline{\delta}) = \arg \min_{\eta, \nu, \delta} \sum_{r=1}^{n} w_r \Big[F(X_{r:n}; \eta, \nu, \delta) - E\{U_{r:n}\} \Big]^2.$$

If $w_r = 1$ this method is known as the least squares (LSE), in other case, we consider $0 < w_r < 1$, more specifically $w_r = 1/Var\{F(X_{r:n}; \eta, \nu, \delta)\}$.

In term of incomplete gamma function we have for (3.7)

(3.8)
$$(\overline{\eta}, \overline{\nu}, \overline{\delta}) = \arg \min_{\eta, \nu, \delta} \sum_{r=0}^{n} w_r \left[\frac{\gamma(\eta \, x_{r:n}^{\delta}, \nu/\delta)}{\Gamma(\nu/\delta)} - \frac{r}{(n+1)} \right]^2.$$

3.3. Probability weighted moment method

The Probability-weighted moments (PWM) method, which has been investigated by many researchers, was originally proposed by Greenwood et. al. (1979), where PWM is commonly used in theoretical studies and empirical purposes, it is an alternative estimation method for the probability distributions, and it has enabled the determination of parameters that are more stable against possible outliers in the dataset. The PWM method can be generally used to estimate parameters of a distribution which inverse form cannot be expressed explicitly. The (s, r)th PWM of X that follows a distribution with cdf $F(x; \beta)$, say $w_{s,r}$, is defined by

$$w_{s,p,r}(\beta) = E\{X^s S^p(X;\beta)F^r(X;\beta)\} = \int_0^\infty x^s S^p(x;\beta)F^r(x;\beta)dF(x;\beta),$$

for $s, p, r \in \mathbf{N}$ and $S(\cdot; \beta)$ is the survival function of X.

An usual practical application of PWM is when s = 1 and p = 0, we use, $w_{1,0,r} := w_r$. Similar notation will be use for the sample version, say $\hat{w}_{1,0,r} := \hat{w}_r$.

So using the practical formula we have the sample version of w_r as \hat{w}_r , know as Landwerh formula, defined as

 $\hat{w}_r = \frac{1}{n} \sum_{j=1}^n X_{j:n} E\{U_{j:n}^r\} \text{ with } E\{U_{j:n}^r\} = \frac{B(j+r,n+1-j)}{B(j,n+1-j)}, r = 1, 2, 3.$ Solving numerically the three non-linear equations $w_r(\beta) = \hat{w}_r, r = 1, 2, 3,$

we can calculate the PWM estimator for the elements of the parameter vector $\beta = (\nu, \eta, \delta)$. Moreover we make use of relation

(3.9)
$$w_r(\beta) = \int_0^1 Q(y;\beta) y^r dy,$$

where $Q(y;\beta)$ is the quantile function of the variable $X \sim GG_2(\nu,\eta,\delta)$, which can be calculated using numerical procedures.

4. Inference likelihood-based

The log-likelihood function for (η, ν, δ) based on a simple random sample $\{x_1, \ldots, x_n\}$ of size *n* from $X \sim GG_2(\eta, \nu, \delta)$ defined by (2.3), is given by

$$l(\eta,\nu,\delta) = n \left[\ln(\delta) + \frac{\nu}{\delta} \ln(\eta) - \ln(\Gamma(\nu/\delta)) \right] + (\nu-1) \sum_{i=1}^{n} x_i - \eta \sum_{i=1}^{n} x_i^{\delta}.$$
(4.1)

The general assumption considered for the GG_2 distribution are those that allow inference based on likelihood until first order, calling the conditions of the distribution proposal, as, conditions of regularity (Cox and Hinkley 1974).

4.1. Maximum likelihood method

The maximum likelihood estimator (MLE) of (η, ν, δ) can be obtained numerically solving the equation of likelihood given in Kohn and Lomdahl (2004). In the case $X \sim GG_3(k, \beta, \psi)$ the procedure of Lawless (1980) can be used, Cox et. al. (2007) and Noufaily and Jones (2009) used this procedure and concluded that they have obtained consistent estimators, even for small sample size. If we considered the form (2.3), with $\nu = 1$, $\delta = 1/\alpha$ and $\eta = 1/\rho$, that is the case of our parametrization $GG_2(1/\rho, 1, 1/\alpha)$, the likelihood equation become from the first derivatives in relation of α and ρ of log-likelihood, given by

(4.2)
$$\frac{\partial l}{\partial \alpha} = -\frac{n}{\alpha} - n \ln(\rho) - n\psi(\alpha) + \frac{1}{\alpha^2 \rho} \sum_{i=1}^n x_i^{1/\alpha} \ln(x_i)$$

(4.3)
$$\frac{\partial l}{\partial \rho} = -\frac{n\alpha}{\rho} + \frac{1}{\rho^2} \sum_{i=1}^n x_i^{1/\alpha}.$$

From equation (4.3), we obtained the MLE of ρ as function of α , given by

(4.4)
$$\widehat{\rho}(\alpha) = \frac{1}{\alpha} \sum_{i=1}^{n} x_i^{1/\alpha}.$$

So, if we know the α parameter, the MLE of ρ can be obtained from (4.4). Now if both parameter are unknown, we can proceed as follows: substitute $\hat{\rho}(\alpha)$ at (4.2) to obtained

$$g(\alpha) = l(\widehat{\rho}(\alpha), \rho) = -n\ln(\alpha) + n\alpha\ln(\alpha) - n\alpha\ln(\sum_{i=1}^{n} x_i^{1/\alpha}) - n\ln(\Gamma(\alpha)) - \alpha.$$
(4.5)

And using the procedure of the first derivative of $g(\alpha)$, the MLE of α can be obtained, solving the equation

$$g'(\alpha) = -\frac{n}{\alpha} + n\ln(\alpha) + n - n\ln\left(\sum_{i=1}^{n} x_i^{1/\alpha}\right) + \frac{n}{\alpha} \frac{\sum_{i=1}^{n} x_i^{1/\alpha} \ln(x_i)}{\sum_{i=1}^{n} x_i^{1/\alpha}} - n\Psi(\alpha) - 1 = 0.$$
(4.6)

We note that (4.6) can be written as

$$(4.7) h(\alpha) = \alpha$$

where

$$h(\alpha) = \left[1 - \frac{1}{n} + \ln(\alpha) - \ln\left(\sum_{i=1}^{n} x_i^{1/\alpha}\right) - \Psi(\alpha) + \frac{1}{\alpha} \frac{\sum_{i=1}^{n} x_i^{1/\alpha} \ln(x_i)}{\sum_{i=1}^{n} x_i^{1/\alpha}}\right]^{-1}$$
(4.8)

and α can be resolver from a fixed point equation $h(\alpha) = \alpha$. The Fisher's information matrix is given by

$$\mathbf{I}_n(\alpha,\rho) = \left[\begin{array}{cc} I_{\alpha\alpha} & I_{\alpha\rho} \\ I_{\rho\alpha} & I_{\rho\rho} \end{array} \right]$$

where

$$I_{\alpha\alpha} = n \left[\psi'(\alpha) - \frac{1}{\alpha^2} + \frac{2}{\alpha^3 \rho} a_1 + \frac{1}{\alpha^4 \rho} a_2 \right], \ I_{\alpha\rho} = n \left[\frac{1}{\rho} + \frac{1}{\alpha^2 \rho^2} a_1 \right], \ I_{\rho\rho} = n \frac{1}{\rho},$$

with

$$E\{x^{1/\alpha}\} = \alpha\rho, \ a_1 = E\{\ln(x)x^{1/\alpha}\} < \infty, \ a_2 = E\{\ln^2(x)x^{1/\alpha}\} < \infty,$$
(4.9)

and $\psi'(\cdot)$ is the trigamma function. Therefore, the MLE of (α, ρ) , when the size sample, n, is large, has a asymptotic distribution given by

(4.10)
$$\sqrt{n}(\widehat{\alpha} - \alpha, \widehat{\rho} - \rho) \sim N_2(\mathbf{0}, \mathbf{I}_n^{-1}(\widehat{\alpha}, \widehat{\rho})).$$

4.2. Confidence regions and testing of hypotheses

From (4.10) we can obtain approximate confidence intervals for α and ρ . Besides we will focus on testing hypotheses about the α and ρ parameters in conjoint form using approximate confidence region for (α, ρ) . For onedimensional inference for shape or scale, we do not believe, being a separate study to review in detail the possibility of UMP test.

The problem considered is to test a simple null hypothesis for the interest parameter vector (α, ρ) ,

(4.11)
$$H_0: \alpha = \alpha_0, \quad \rho = \rho_0 \quad \text{vs} \quad H_0: \alpha \neq \alpha_0 \text{ or } \rho \neq \rho_0$$

and for this, first we will use the likelihood ratio statistic test

 $R(\alpha_0, \rho_0) = L(\alpha_0, \rho_0) / L(\widehat{\alpha}, \widehat{\rho})$

or the log-likelihood ratio statistics test, given by

 $S_1(\alpha_0, \rho_0) = -2[l(\alpha_0, \rho_0) - l(\widehat{\alpha}, \widehat{\rho})]$

where $\hat{\alpha}$ and $\hat{\rho}$ are the MLE's of α and ρ respectively and for large size sample, n, the distribution of S_1 under H_0 is approximately chi-square with two degrees of freedom, making possible the use of $S_1(\alpha_0, \rho_0)$ to test the hypothesis (4.11).

An approximate $100(1 - \gamma)\%$ likelihood-ratio-based confidence region for (α, ρ) is the set of all values of (α, ρ) such that $S_1(\alpha, \rho) < \chi^2_{1-\gamma;2}$ or $R(\alpha, \rho) > \exp\{-\chi^2_{1-\gamma;2}/2\}.$

As second procedure to test (4.11) we will use the Wald's statistics, given by

(4.12)
$$S_2(\alpha, \rho) = (\widehat{\alpha} - \alpha, \widehat{\rho} - \rho) \mathbf{I}_n(\widehat{\alpha}, \widehat{\rho}) (\widehat{\alpha} - \alpha, \widehat{\rho} - \rho)^\top,$$

where the asymptotic distribution for the statistics $S_2(\alpha, \rho)$ under H_0 is chisquare with two degrees of freedom. Generating an approximate $100(1 - \gamma)\%$ likelihood-ratio-based confidence region for (α, ρ) as the set of all values of (α, ρ) such that $S_2(\alpha, \rho) < \chi^2_{1-\gamma;2}$. Therefore, it can be easily checked whether (α, ρ) is inside or outside a give contour and hence whether H_0 is accepted or rejected at a given significance level γ .

Graphically we can test the hypothesis (4.11) displaying a contour plot of $S_i(\alpha, \rho) - \chi^2_{1-\gamma;2} = 0$, i = 1 or 2.

If we choose S_i as statistics test, because to we want 5% significant level to test (4.11) we should reject the hypothesis if (α, ρ) falls outside the contour $S_i(\alpha, \rho) - 5.991 = 0$. We will make use of S_2 statistics test.

5. Simulation Monte Carlo Study

In this section we study empirical behavior of the five estimation methods, these methods have been described in Section 3 and the MLE described in the Section 4, for the parameter of the proposed distribution $GG_2(\eta, \nu, \delta)$ with $\nu = 1$, $\delta = 1/\alpha$ and $\eta = 1/\rho$.

We calculated the relative bias and the square root of the mean squared error which are denoted by (1) and (2) in the Tables respectively, the estimations are based in Monte Carlos study over 5,000 repetitions. For the MOM, PWM methods we must use a numerical procedure to solve the non-linear equation system and for the LS, WLS and ML we use numerical procedure to do maximization (minimization). All methods were repeated during the entire experiment for $(\alpha, \rho) \in \{(0.5, 2), (1, 2), (2, 0.5), (2, 2)\}$, the skewness (β_1) and kurtosis (β_2) coefficients for the vector parameter set considered are show in the Table 1.

Table 1 : The symmetric (β_1) and kurtosis (β_2) coefficients for the vector parameter set $(\alpha, \rho) \in \{(0.5, 2), (1, 2), (2, 0.5), (2, 2)\}.$

(α, ρ)	(0.5,2)	(1,2)	(2,0.5)	$(2,\!2)$
β_1	3.00	4.77	6.26	7.07
β_2	17.31	37.41	5.53	87.72

The sample size, n, considered are n = 50, 100, 200, 250, 300, 350. The initial values used for the iterative maximization procedure and solving no-linear equation were obtained from coefficient of variation given in the relation (2.10). Tables 2-3 display the results. Looking at these tables we observe that:

• In general when sample size increases, the relative biases and the MSE decrease for all the methods, excepting the ML method. As a consequence of that, all estimators are asymptotically unbiased and consistent of α when ρ is known.

- The behavior of the different method considered are not uniform in relation to sample size, for example, for $n = 50, 100, (\alpha, \rho) = (0.5, 2)$ the PWM is better for estimate α than ML method, while if n = 200 the opposite occurs. Fixing the mean squared error as criteria for comparison, we obtained that for n > 50 we observed that the WLM method is uniformly better than LM. Now if we compare the MOM with PWM methods, we have that for n > 100 the PWM is better that MOM in almost all parameter considered, the exception is for $(\alpha, \rho) = (1, 2)$.
- The behavior of all methods considered are strongly dependent of α , we have the same conclusion when analyzing a real data set: if we look at $\rho = 2$ and move α is observed that the behavior of relative biases to estimate ρ , by the MOM, PWM and MLE methods those are not uniform, but the LS and WLS methods have a behavior uniform. If we considered the behavior MSE the above conclusion is opposite. On the same scene, but now the interest is to analyze the behavior when estimating α , the methods that have a uniform behavior in the relative biases and MSE are the LS and WLS.
- Moreover, if n < 100 we can say that if the interest parameter is ρ and nuisance is α , the best method is WLS, and for n > 100 the MLE is better.

Table 2 : Bias (1) and square root of the mean squared error (2) The
first line displays the results for the location parameter, α , and the second
one for the scale parameter, ρ .

n		M	OM	I	'S	W	\mathbf{LS}	PV	VM	Ν	ſL
		(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)
50	$\begin{array}{c} \alpha = 0.5 \\ \rho = 2 \end{array}$	0.430 -0.449	0.058 0.889	1.057 -0.605	$0.333 \\ 1.494$	0.826 -0.567	$0.204 \\ 1.324$	0.295 -0.352	$0.036 \\ 0.686$	0.368 -0.306	$0.065 \\ 1.872$
	$\begin{array}{c} \alpha = 1 \\ \rho = 2 \end{array}$	$-0.002 \\ 0.171$	$\begin{array}{c} 0.039 \\ 1.416 \end{array}$	$0.451 \\ -0.363$	$0.335 \\ 1.024$	$0.372 \\ -0.352$	$\begin{array}{c} 0.225 \\ 0.815 \end{array}$	$-0.021 \\ 0.241$	$\begin{array}{c} 0.043 \\ 1.644 \end{array}$	$\begin{array}{c} 0.121 \\ 0.027 \end{array}$	$0.093 \\ 2.241$
	$\begin{array}{c} \alpha=2\\ \rho=0.5 \end{array}$	$-0.177 \\ 0.456$	$\begin{array}{c} 0.278 \\ 0.126 \end{array}$	$0.200 \\ -0.108$	$\begin{array}{c} 0.484 \\ 0.024 \end{array}$	$0.145 \\ -0.067$	$\begin{array}{c} 0.348 \\ 0.023 \end{array}$	$-0.155 \\ 0.448$	$\begin{array}{c} 0.291 \\ 0.152 \end{array}$	$\begin{array}{c} 0.013 \\ 0.085 \end{array}$	$\begin{array}{c} 0.230 \\ 0.045 \end{array}$
	$\begin{array}{c} \alpha=2\\ \rho=2 \end{array}$	$-0.174 \\ 1.690$	$\begin{array}{c} 0.274\\ 46.736\end{array}$	0.189 -0.010	$0.468 \\ 17.983$	$\begin{array}{c} 0.118\\ 0.143\end{array}$	$\begin{array}{c} 0.332 \\ 19.559 \end{array}$	$-0.155 \\ 1.647$	$\begin{array}{c} 0.285 \\ 46.715 \end{array}$	$\begin{array}{c} 0.016\\ 0.431 \end{array}$	$0.227 \\ 20.700$
100	$\begin{array}{c} \alpha = 0.5 \\ \rho = 2 \end{array}$	0.230 -0.264	$\begin{array}{c} 0.021 \\ 0.431 \end{array}$	0.597 -0.438	$0.107 \\ 0.844$	0.437 -0.376	$0.062 \\ 0.685$	0.167 -0.187	$\begin{array}{c} 0.016 \\ 0.386 \end{array}$	0.148 -0.087	$0.021 \\ 1.248$
	$\begin{array}{c} \alpha = 1 \\ \rho = 2 \end{array}$	$-0.044 \\ 0.238$	$\begin{array}{c} 0.026 \\ 1.319 \end{array}$	$0.219 \\ -0.212$	$\begin{array}{c} 0.099 \\ 0.574 \end{array}$	$0.158 \\ -0.146$	$\begin{array}{c} 0.066 \\ 0.549 \end{array}$	$-0.042 \\ 0.247$	$\begin{array}{c} 0.029 \\ 1.388 \end{array}$	$\begin{array}{c} 0.020\\ 0.130\end{array}$	$0.038 \\ 1.608$
	$\substack{\alpha=2\\\rho=0.5}$	$-0.138 \\ 0.298$	$\begin{array}{c} 0.173 \\ 0.051 \end{array}$	0.097 -0.053	$\begin{array}{c} 0.195 \\ 0.013 \end{array}$	$0.066 \\ -0.028$	$\begin{array}{c} 0.141 \\ 0.012 \end{array}$	$-0.099 \\ 0.228$	$\begin{array}{c} 0.156 \\ 0.043 \end{array}$	$-0.004 \\ 0.051$	$\begin{array}{c} 0.112 \\ 0.017 \end{array}$
	$\begin{array}{c} \alpha=2\\ \rho=2 \end{array}$	$-0.142 \\ 0.942$	$0.178 \\ 9.507$	0.091 -0.028	$0.196 \\ 1.634$	$\begin{array}{c} 0.052 \\ 0.060 \end{array}$	$0.152 \\ 1.400$	$-0.118 \\ 0.879$	$\begin{array}{c} 0.183 \\ 9.438 \end{array}$	-0.008 0.222	$\begin{array}{c} 0.116 \\ 2.354 \end{array}$
200	$\begin{array}{c} \alpha = 0.5 \\ \rho = 2 \end{array}$	$0.127 \\ -0.151$	$\begin{array}{c} 0.009 \\ 0.258 \end{array}$	0.327 -0.289	$\begin{array}{c} 0.036 \\ 0.448 \end{array}$	0.235 -0.232	$\begin{array}{c} 0.021 \\ 0.366 \end{array}$	0.101 -0.113	$\begin{array}{c} 0.008 \\ 0.245 \end{array}$	$0.070 \\ -0.042$	$0.008 \\ 0.458$
	$\begin{array}{c} \alpha = 1 \\ \rho = 2 \end{array}$	$-0.040 \\ 0.170$	$\begin{array}{c} 0.017 \\ 0.653 \end{array}$	$0.124 \\ -0.130$	$\begin{array}{c} 0.041 \\ 0.351 \end{array}$	0.083 -0.077	$\begin{array}{c} 0.028\\ 0.338\end{array}$	$-0.031 \\ 0.157$	$\begin{array}{c} 0.018\\ 0.666\end{array}$	$\begin{array}{c} 0.001 \\ 0.083 \end{array}$	$0.020 \\ 0.603$
	$\begin{array}{c} \alpha=2\\ \rho=0.5 \end{array}$	$-0.104 \\ 0.198$	$\begin{array}{c} 0.099 \\ 0.022 \end{array}$	$0.055 \\ -0.031$	$\begin{array}{c} 0.091 \\ 0.007 \end{array}$	$0.037 \\ -0.017$	$\begin{array}{c} 0.067 \\ 0.006 \end{array}$	$-0.061 \\ 0.125$	$\begin{array}{c} 0.079 \\ 0.015 \end{array}$	$-0.003 \\ 0.027$	$0.055 \\ 0.007$
	$\begin{array}{c} \alpha = 2 \\ \rho = 2 \end{array}$	-0.097 0.512	$\begin{array}{c} 0.098 \\ 2.896 \end{array}$	$0.054 \\ -0.024$	$0.099 \\ 0.601$	$0.037 \\ -0.002$	$\begin{array}{c} 0.074 \\ 0.514 \end{array}$	-0.070 0.430	$0.097 \\ 2.793$	$-0.001 \\ 0.099$	$0.062 \\ 0.759$

ne for t	he scale p	oaram	eter, ρ			1		, ,		
n	м	ЭM	L	s	W	LS	PW	VМ	М	L
	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)

Table 3 : Bias (1) and square root of the mean squared error (2) The d

250	$\alpha = 0.5 \\ \rho = 2$	0.109 -0.132	$0.007 \\ 0.223$	$0.276 \\ -0.254$	$0.027 \\ 0.375$	0.198 -0.204	$0.016 \\ 0.307$	0.090 -0.103	$0.007 \\ 0.211$	0.060 -0.043	$0.007 \\ 0.324$
	$\begin{array}{c} \alpha = 1 \\ \rho = 2 \end{array}$	$-0.037 \\ 0.142$	$\begin{array}{c} 0.013\\ 0.472\end{array}$	$0.105 \\ -0.116$	$\begin{array}{c} 0.031\\ 0.289 \end{array}$	0.071 -0.071	$\begin{array}{c} 0.021\\ 0.266\end{array}$	$-0.027 \\ 0.127$	$\begin{array}{c} 0.014\\ 0.481 \end{array}$	$0.000 \\ 0.063$	$\begin{array}{c} 0.015 \\ 0.413 \end{array}$
	$\begin{array}{c} \alpha=2\\ \rho=0.5 \end{array}$	$-0.095 \\ 0.175$	$\begin{array}{c} 0.082\\ 0.017\end{array}$	0.048 -0.028	$\begin{array}{c} 0.074 \\ 0.006 \end{array}$	0.032 -0.016	$\begin{array}{c} 0.054 \\ 0.005 \end{array}$	$-0.053 \\ 0.103$	$\begin{array}{c} 0.062\\ 0.011 \end{array}$	$-0.003 \\ 0.021$	$0.045 \\ 0.006$
	$\begin{array}{l} \alpha=2\\ \rho=2 \end{array}$	-0.083 0.407	$\begin{array}{c} 0.076 \\ 1.843 \end{array}$	0.050 -0.033	$\begin{array}{c} 0.080\\ 0.452 \end{array}$	0.036 -0.016	$\begin{array}{c} 0.060 \\ 0.399 \end{array}$	$-0.054 \\ 0.319$	$\begin{array}{c} 0.074 \\ 1.733 \end{array}$	$\begin{array}{c} 0.004 \\ 0.063 \end{array}$	$0.049 \\ 0.497$
300	$\begin{array}{c} \alpha = 0.5 \\ \rho = 2 \end{array}$	0.096 -0.118	$\begin{array}{c} 0.006 \\ 0.199 \end{array}$	0.241 -0.230	$\begin{array}{c} 0.021 \\ 0.323 \end{array}$	0.173 -0.183	$\begin{array}{c} 0.013 \\ 0.267 \end{array}$	0.081 -0.093	$\begin{array}{c} 0.005 \\ 0.190 \end{array}$	0.053 -0.040	0.005 0.279
	$\begin{array}{c} \alpha = 1 \\ \rho = 2 \end{array}$	$-0.035 \\ 0.130$	$\begin{array}{c} 0.011 \\ 0.388 \end{array}$	0.089 -0.098	$\begin{array}{c} 0.024\\ 0.246\end{array}$	0.059 -0.059	$\begin{array}{c} 0.017\\ 0.223\end{array}$	-0.024 0.113	$\begin{array}{c} 0.012\\ 0.392 \end{array}$	-0.002 0.058	$0.012 \\ 0.330$
	$\begin{array}{c} \alpha=2\\ \rho=0.5 \end{array}$	-0.086 0.153	$0.070 \\ 0.013$	$0.041 \\ -0.026$	$0.06 \\ 0.005$	$0.028 \\ -0.016$	$\begin{array}{c} 0.045 \\ 0.004 \end{array}$	$-0.045 \\ 0.085$	$\begin{array}{c} 0.053 \\ 0.009 \end{array}$	$-0.002 \\ 0.016$	$0.037 \\ 0.004$
	$\begin{array}{c} \alpha=2\\ \rho=2 \end{array}$	-0.079 0.365	$\begin{array}{c} 0.065 \\ 1.405 \end{array}$	$0.044 \\ -0.033$	$\begin{array}{c} 0.065 \\ 0.383 \end{array}$	$0.032 \\ -0.018$	$\begin{array}{c} 0.048\\ 0.326\end{array}$	$-0.051 \\ 0.275$	$\begin{array}{c} 0.061 \\ 1.280 \end{array}$	$\begin{array}{c} 0.003 \\ 0.051 \end{array}$	$0.039 \\ 0.387$
350	$\begin{array}{c} \alpha = 0.5 \\ \rho = 2 \end{array}$	0.081 -0.100	$\begin{array}{c} 0.005 \\ 0.171 \end{array}$	0.206 -0.203	$\begin{array}{c} 0.016 \\ 0.271 \end{array}$	0.147 -0.159	$0.01 \\ 0.225$	0.069 -0.080	$\begin{array}{c} 0.005 \\ 0.165 \end{array}$	0.042 -0.031	$0.004 \\ 0.229$
	$\begin{array}{c} \alpha = 1 \\ \rho = 2 \end{array}$	$-0.032 \\ 0.114$	$\begin{array}{c} 0.010\\ 0.310\end{array}$	0.08 -0.093	$\begin{array}{c} 0.021\\ 0.211\end{array}$	$0.054 \\ -0.057$	$\begin{array}{c} 0.014 \\ 0.189 \end{array}$	$-0.022 \\ 0.096$	$\begin{array}{c} 0.011\\ 0.311 \end{array}$	$-0.002 \\ 0.047$	$\begin{array}{c} 0.011 \\ 0.261 \end{array}$
	$\substack{\alpha=2\\\rho=0.5}$	$-0.082 \\ 0.144$	$\begin{array}{c} 0.061 \\ 0.012 \end{array}$	$0.037 \\ -0.023$	$\begin{array}{c} 0.051 \\ 0.004 \end{array}$	$0.025 \\ -0.014$	$\begin{array}{c} 0.038\\ 0.003 \end{array}$	$-0.041 \\ 0.077$	$\begin{array}{c} 0.044 \\ 0.007 \end{array}$	$-0.002 \\ 0.015$	$\begin{array}{c} 0.032 \\ 0.004 \end{array}$
	$\alpha = 2$ $\rho = 2$	-0.072 0.316	$0.055 \\ 1.078$	$0.044 \\ -0.045$	$0.056 \\ 0.318$	0.033 -0.031	$\begin{array}{c} 0.042 \\ 0.269 \end{array}$	$-0.042 \\ 0.221$	$\begin{array}{c} 0.050 \\ 0.944 \end{array}$	$0.006 \\ 0.030$	0.032 0.303

6. Application

6.1. Data set

One of the most frequent types of contamination in rural watersheds is fecal pollution from different sources. Microbiological contamination is dispersed, sporadic, and influenced by a range of interacting environmental factors such as the watersheds physical characteristics, climatic conditions, and agricultural management practices. We will consider an application to a data set, obtained from Valenzuela M. et. al. (2009), which corresponds to the presence of Colony Forming Units (CFU) of Fecal Coliform (FC) per 100 ml of groundwater samples. FC presence in a water sample is a strong evidence of fecal contamination, which is useful as an indicator of the probable presence of microorganisms capable of causing diseases in humans. The small rural watershed has an area of 10.8 km^2 , and is located in the Biobio Region, Chile. The catchment area is sparsely inhabited by families dedicated to traditional agriculture, and is characterized by a Mediterranean climate with a long dry season leading to water shortages and a short wet season. The watershed soils have low permeability and capacity to provide underground water. Moisture accumulation in the watershed takes place between April and June. The major runoff period of the year is from July to October when the ground is saturated and almost all the precipitation that falls in the watershed runs off. Precipitation is scarce between November and March, with practically no base flow in the watershed. Farmers obtain small amounts of water from private wells.

On the average, these are 7.0 m deep and yield a median of $1.1 Lmin^{-1}$. Groundwater is used as drinking water, for other domestic purposes, orchards, gardens, greenhouses, and livestock production. Forty-two wells were chosen with the Stratified Random Sample (Murray 2002) and sitelocation data were determined with global positioning system units (Garmin 12XL, Garmin International Inc., Kansas, USA). The sampling periods were defined in accordance with the precipitation regime and variations in the hydrologic levels in the wells. Based on these criteria, four sampling seasons were established (March, June, September, and December). Water samples were analyzed for fecal coliforms (FC). Aseptic sample collections were taken in sterilized flasks. Samples were held at 5 ^{o}C after being collected and for no more than 6 h until reaching the laboratory. Results were expressed in colony forming units (CFU) per 100 mL. FC concentrations were analyzed with a membrane filtration technique following standard methods (Clesceri and Eaton 1998). Aliquots (100, 10, and 1 mL) of each water sample were filtered through a 0.45μ Millipore membrane filter. All samples were tested in triplicate. Results were reported as CFU /100 ml.

Results for FC obtained in the four seasons were analyzed independently by looking for spatial correlations using the definitions of Gearys and Morans Index (Anselin 1988). The null hypothesis of no spatial correlation between data at each station was not rejected. A seasonal trend was identified. Concentrations of FC varied over time and showed a pattern similar to rainfall which appeared to exert a local influence on the indicator concentrations. Indicator counts turned out to be significantly related to certain watershed features during specific months. Inherent well site characteristics and its surroundings, as well as rainfall are the main factors that affect groundwater quality in the watershed.

Statistic	March	June	September	December
\overline{X}_n	203.4634	372.1707	28.7333	136.2821
S_X^2	108129.3049	797497.3951	2603.2984	59947.5236
S_X/\overline{X}_n	1.6162	2.3995	1.7757	1.7966
$\widetilde{\beta}_1$	1.7195	5.4733	3.7980	3.5197
$\tilde{\beta}_2$	4.7931	33.3043	19.5092	17.3986

Table 4: Summary of data set for each station.

Table 4 shows some statistical properties of data set, here β_1 and β_2 are the sample version of the skewness and kurtosis coefficients. From this Table we can observed that the data set are skewness and kurtosis coefficients more large that the case of a normal distribution.

6.2. Use of $GG_2(1/\rho, 1, 1/\alpha)$ distribution

Applying the procedure described in the Subsection 3.1 and from the Table 7 in the Appendix we obtained the initial point for the iterative maximization procedure to fitted the GG_2 and GE distribution, moreover we use the **fitdist** function of **R** to fit the Weibull and gamma distribution.

We have found that the $GG_2(1/\rho, 1, 1/\alpha)$ compared to the Weibull, gamma and GE (Gupta and Kundu, 1999) distributions, provides better fitting for a data set corresponding to the number of bacteria of fecal coliform class, measured per ml of water, which is useful as an indicator of the presence of microorganisms capable of causing disease in humans (Valenzuela, et al., 2009).

Model	Ma	rch	Ju	ine	Septe	ember	Dece	mber
	AIC	BIC	AIC	BIC	AIC	BIC	AIC	BIC
GG_2	478.04	481.47	560.32	563.75	359.62	363.09	447.43	450.76
Weibull	480.50	483.93	563.58	567.01	362.32	365.80	448.80	452.12
gamma	483.49	486.92	568.05	571.47	365.43	368.91	451.49	454.81
GE	483.57	486.99	569.12	572.59	366.10	369.57	452.02	455.35

Table 5: AIC, BIC for the GG_2 , Weibull, gamma and GE distribution

We can observe from Table 5 that the GG_2 improved the fit for the data set considered above.

The MLE and their standard errors (se) corresponding for α and ρ parameters of GG_2 are given in Table 6.

Table 6: MLE of Parameters of GG_2

Estimative	March	June	September	December
$\widehat{\alpha}$ (se)	4.475 (1.041)	1.780(0.359)	1.941(0.435)	2.695(0.674)
$\widehat{ ho}(se)$	0.557 (0.265)	12.544 (11.153)	2.341(1.465)	1.789(1.243)

Following the Section 4, we proceed to construct a confidence region to test the null hypothesis (4.11) for (α, ρ) , which arises from the asymptotic bi-variate normal distribution, given by

$$\sqrt{n}(\widehat{\alpha} - \alpha, \widehat{\rho} - \beta) \sim N_2 \left\{ \begin{bmatrix} 0\\0 \end{bmatrix}, \begin{bmatrix} 1.127 & -0.284\\-0.284 & 0.073 \end{bmatrix} \right\}$$

So the 95% region confidence for (α, ρ) is given by the contour plot

$$n(\hat{\alpha} - \alpha, \hat{\rho} - \rho) \begin{bmatrix} 38.361 & 148.646\\ 148.646 & 589.637 \end{bmatrix}^{-1} \begin{bmatrix} \hat{\alpha} - \alpha\\ \hat{\rho} - \rho \end{bmatrix} - \chi^2_{0.95;2} = 0$$

or

(6.1)
$$n(\widehat{\alpha} - \alpha, \widehat{\rho} - \beta) \begin{bmatrix} 1.127 & -0.284 \\ -0.284 & 0.073 \end{bmatrix} \begin{bmatrix} \widehat{\alpha} - \alpha \\ \widehat{\rho} - \rho \end{bmatrix} - 5.991 = 0$$



Figure 1 : An approximate 95% confidence ellipse of (α, ρ) , from GG_2 distribution, data set of March.

The ellipse (6.1) is provided in Figure 1, that can be used to test the hypothesis (4.11).

7. Conclusion

We fitted four model classes to a real data set, consisting in microbiological indicator pollution levels in groundwater of a rural watershed, driven by rainfall-runoff phenomena. We used the ML method to estimate the parameters corresponding to GG_2 , gamma, Weibull and generalized exponential distributions. The distribution chosen, using the AIC and BIC criteria, is proposed in this paper, i.e., $GG_2(1/\alpha, 1, 1/\rho)$ distribution, which is not incorporated in the list of (Crooks 2010). We can say that the $GG_2(1/\alpha, 1, 1/\rho)$ distribution allows data with kurtosis and high asymmetry, as was visible in the fit of our data.

From the simulation study performed, considering the square mean error, we obtained that if the interest parameter is the α and ρ the nuisance is preferable to use the ML method. Instead, if the ρ is the interest parameter and α is the nuisance, the preferable method is WLS. But if we make inferences to α and ρ together, we can use the results of Subsection 4.2 (Confidence regions and testing of hypotheses). From the semi-ellipse on the Figure 1, we can infer that the dependence between $\hat{\alpha}$ and $\hat{\rho}$ is strong, in term of the correlation, this being an ellipse with minor axis smaller compared to the principal axis. From above, we observe that it is useful to analyze the data using moment estimates when the coefficient kurtosis is not very large, what can potentially serve in data sets coming from pollution in watersheds, other than microorganisms, related to rainfall-runoff processes.

Acknowledgments

The research in this paper has been partially supported by grant DIUC 210.014.018-1.0 (Univesity of Concepción, Chile).

Table 7 : Table for $E\{Y\}$ and $CV\{Y\}$, $Y \sim GG_2(1/\alpha, 1, 1)$, for difference values of α .

B. Lagos Álvarez, G. Ferreira and M. Valenzuela Hube

α	$E\{Y\}$	$CV{Y}$
2.0	6.0000	1.5275
2.1	7.4121	1.5844
2.2	9.1996	1.6421
2.3	11.4692	1.7009
2.4	14.3602	1.7608
2.5	18.0541	1.8217
2.6	22.7879	1.8838
2.7	28.8724	1.9471
2.8	36.7159	2.0117
2.9	46.8556	2.0775
3.0	60.0000	2.1448
3.1	77.0862	2.2134
3.2	99.3553	2.2835
3.3	128.4554	2.3551
3.4	166.5789	2.4283
3.5	216.6488	2.5030

References

- Amoroso, L. : Ricerche intorno alla curva dei redditi (in Italian). Ann. Math. Pura Appl. 4 (21), pp. 123-159, (1925).
- [2] Anselin, L., Spatial Econometrics: Methods and Models. Kluwer Academic Publisher, The Nederlands, (1988).
- [3] Clesceri, L. S., G. A., and Eaton, A., Standard methods for the examination of water and wastewater, 20th Edition, Washington DC: American Public Health Association (APHA), (1998).
- [4] Cox, C., Chu, H., Schneider, M.F. and Munõz, A. Parametric survival analysis and taxonomy of hazard functions for the generalized gamma distribution. Statistics in Medicine, 26, pp. 4352-4374, (2007).
- [5] Cox, D.R. and Hinkley, D.V. (1974) Theoretical Statistics, London: Chapman and Hall.
- [6] Crooks G. E (2010) The Amoroso Distribution, Technical Notes, Physical Biosciences Division, Lawrence Berkeley National Lab, Berkeley.
- [7] Gradshteyn, I. S. and Ryzhik, I. M. (2007) Table of Integrals, Series and Products, Seventh Edition, Academic Press, San Diego, U.S.A., (2007).
- [8] Greenwood, J. A. Landwehr, J. M. Matalas, N. C. and Wallis J. R. Probability Weighted Moments: Definition and Relation to Parametes of several Distributions expressible in inverse Form. Water Resources Research, 15, pp. 1049-1054, (1979).
- [9] Gupta, R. D. and Kundu, D. Generalized Exponential Distributions, Australian and New Zealand Journal of Statistics, 41 (2), pp. 173-188, (1999).
- [10] Gupta, R. D. and Kundu, D. Exponentiated Exponential Family: an alternative to Weibull and Gamma Distribution. Biometrical Journal, 43, pp. 117-130, (2001).
- [11] Johnson, N. L., Kotz, S. and Balakrishnan, N. Continuous Univariate Distributions, Vol. 1, Wiley, New York, (1994).

- [12] Konn, H. and Lomdahl, P. S. (2004) Stochastic Processes Having Fractional Order Nonlinearity Associated with Hyper Gamma Distribution, Journal of the Physical Society of Japan, 73(3), pp. 573-579, (2004).
- [13] Lawless, J. F. Inference in the generalized gamma and log gamma distributions. Technometrics, 22, pp. 409-419, (1980).
- [14] Mooley, D. A., and Mohile, C. M., Some aspects of rainfall associated with cyclonic storms of the Bay of Bengal, International Journal of Climatology, 6, 149-160, (1986).
- [15] Mooley, D., and Appa Rao, G., Distribution function for seasonal and annual rainfall over India, Monthly Weather Review, 99, 796-799, (1971).
- [16] Murray, C., Sampling and data analysis for environmental microbiology, in Manual of Environmental Microbiology. 2nd ed, eds. K. M. Hurst, Crawford, and S. L.D., Washington D. C., USA: ASM Press, (2002).
- [17] Noufaily, A. and Jones, M. C. On Maximization of the Likelihood for the Generalized Gamma Distribution, in Technical Report in Statistics, Department of Mathematics and Statistics, The Open University, (2009).
- [18] Stacy, E. W. A generalization of the gamma distribution, Ann. Math. Stat. 33, pp. 1187-1192, (1962).
- [19] Stacy, E., and Mihram, G. (1965), Parameter estimation for a generalized gamma distribution, Technometrics, 7, 349-358.
- [20] Suzuki, E. Hyper gamma distribution and its fitting to rainfall data. Pap. Meteor. Geophys. 15, pp. 31-51, (1964).
- [21] Swain, J., Venkatraman, S. and Wilson, J. Least squares estimation of distribution function in Johnson's translation system, Journal of Statistical Computation and Simulation 29, pp. 271-297, (1988).
- [22] Valenzuela Mariella., Lagos B., Claret M., Mondaca M., Parra O. Ocurrence of faecal contamination in groundwater at a small rural watershed. *Chilean Journal of Agricultural Research*, 69 (2), pp. 235-243, (2009).

B. Lagos-Álvarez

Department of Statistic, University of Concepción Esteban Iturra, Barrio Universitario, Chile e-mail : bla@udec.cl

G. Ferreira

Department of Statistic, University of Concepción Esteban Iturra, Barrio Universitario, Chile e-mail : gpferrei@puc.cl

and

M. Valenzuela-Hube

Environmental Sciences Center EULA, University of Concepción, Chile e-mail : marvalenz@udec.cl