Bayesian estimation in Kibble's bivariate gamma distribution

George Iliopoulos[1], Dimitris Karlis[2] and Ioannis Ntzoufras[2]

[1] Department of Statistics and Insurance Science, University of Pireaus,
80 Karaoli and Dimitriou str., 18534 Pireaus, GREECE

[2] Department of Statistics, Athens University of Economics,
76 Patission str. 10434 Athens, GREECE

Abstract

The paper describes Bayesian estimation for the parameters of Kibble's (1941) bivariate gamma distribution. The density of this distribution can be written as a mixture, allowing for a simple data augmentation scheme. An MCMC algorithm is constructed to facilitate Bayesian estimation. We show that the resulting chain is geometrically ergodic and thus a regenerative sampling procedure is applicable allowing for estimation of ergodic means' standard errors. Bayesian hypothesis testing procedures are developed to test both the dependence hypothesis of the two variables as well as the hypothesis that their means are equal. A reversible jump MCMC algorithm is proposed to carry out this model selection problem. Real and simulated datasets are used to illustrate the proposed methodology.

Key words and phrases: Downton's bivariate exponential distribution; Kibble's bivariate gamma distribution; Markov chain Monte Carlo; regenerative simulation; reversible jump.

1 Introduction

In several circumstances the researcher has to deal with paired positive data. Such examples usually arise in clinical trials involving survival analysis (paired lifetimes) and hydrology (flows in two distinct districts of the same river) among others. Currently, several models based on bivariate exponential distributions have been proposed in the statistical literature; for example Lagakos (1978), Wang and Ghosh (2000), Achcar and Leandro (1998), Damien and Muller (1998) used bivariate exponential models as survival models. For applications in hydrology, see Yue *et al.* (2001).

Consider the case of bivariate continuous positive data with positive skewness to each direction. Bivariate extensions of gamma densities can be used to construct plausible models for such data. They can take a variety of forms; for a review see Kotz *et al.* (2000, chapter 48). Multivariate or even bivariate gamma distributions have not been extensively used at the past, mainly due to their complicated nature.

Among the families of bivariate gamma models, the present paper focuses on the distribution introduced by Kibble (1941). Two random variables X, Y are said to follow Kibble's bivariate gamma distribution (hereafter KBGD) if they have joint density function

$$f(x,y|v,\lambda_1,\lambda_2,\rho) = \frac{(\lambda_1\lambda_2)^v}{(1-\rho)\Gamma(v)}\left(\frac{xy}{\rho\lambda_1\lambda_2}\right)^{\frac{v-1}{2}}\exp\left(-\frac{\lambda_1 x+\lambda_2 y}{1-\rho}\right)I_{v-1}\left(\frac{2\sqrt{\rho\lambda_1\lambda_2 xy}}{1-\rho}\right), \qquad (1)$$

where $x, y, \lambda_1, \lambda_2 > 0$, $0 \leq \rho < 1$, $v > 0$ and $I_\nu(\cdot)$ is the modified Bessel function of the first kind of order ν defined as

$$I_\nu(\alpha) = \sum_{\kappa=0}^{\infty} \frac{(\alpha/2)^{2\kappa+\nu}}{\Gamma(\kappa+\nu+1)\kappa!}, \quad \alpha > 0.$$

In (1) the marginal distributions of X, Y are $\mathcal{G}(v, \lambda_1)$, $\mathcal{G}(v, \lambda_2)$, respectively, that is, gamma with shape v and means v/λ_1, v/λ_2. The conditional expectation and variance of X given $Y = y$ are

$$\begin{aligned} \mathsf{E}(X|Y=y) &= v\lambda_1^{-1}(1-\rho) + \rho\lambda_1^{-1}\lambda_2 y, \\ \mathsf{Var}(X|Y=y) &= v\lambda_1^{-2}(1-\rho)^2 + 2\rho(1-\rho)\lambda_1^{-2}\lambda_2 y, \end{aligned}$$

i.e. they are linear in y. Moreover, $\mathsf{Corr}(X, Y) = \rho$ and X, Y are independent iff $\rho = 0$. In the special case $v = 1$, (1) is Downton's (1970) bivariate exponential distribution.

In what follows, f denotes probability densities. Using the series representation of the Bessel function (1) takes the form

$$f(x, y|v, \lambda_1, \lambda_2, \rho) = \sum_{\kappa=0}^{\infty} f(\kappa|\rho) f\left(x \left| v+\kappa, \frac{\lambda_1}{1-\rho}\right.\right) f\left(y \left| v+\kappa, \frac{\lambda_2}{1-\rho}\right.\right)$$

where $f(\kappa|\rho) = \frac{\Gamma(v+\kappa)}{\Gamma(v)\kappa!}(1-\rho)^v\rho^\kappa$, $\kappa \in \mathbf{Z}_+ := \{0, 1, 2, \ldots\}$, is the probability mass function of the negative binomial distribution $\mathcal{NB}(v, 1-\rho)$ and $f(x|v+\kappa, \lambda_1/(1-\rho))$, $f(y|v+\kappa, \lambda_2/(1-\rho))$ are the probability density functions of gamma distributions $\mathcal{G}(v+\kappa, \lambda_1/(1-\rho))$, $\mathcal{G}(v+\kappa, \lambda_2/(1-\rho))$ respectively. Clearly, X and Y are conditionally (given κ) independent gamma variables.

KBGD allows for two interesting interpretations which can be useful for modelling survival or reliability data. It can be considered either as a competing risk model (where two different causes of failure can occur) or as a frailty model. A frailty model assumes that the event times are conditionally independent given the frailty, which is an unobserved heterogeneity introducing dependence between the two variables (see for example Hougaard, 1984). Thus, KBGD is in fact a frailty model with the negative binomial as frailty distribution.

The present paper aims to introduce a Bayesian procedure for estimating the parameters of KBGD. We propose an MCMC approach using the data augmentation offered by the mixture representation of the density. Furthermore, since typical Bayesian approach treats ρ as a continuous random variable, it does not allow for testing the hypothesis of independence ($\rho = 0$). For this reason, we have constructed a reversible jump MCMC (RJMCMC) algorithm in order to estimate posterior model probabilities and evaluate the support on the corresponding hypotheses of independence or dependence between the two variables of interest. In addition, in paired data interest may also lie on evaluating the difference between the means of the two measurements. For example, if the data refer to clinical measurements of a patient before and after a treatment, we would like to examine the treatment effect via testing the hypothesis of equal means before and after treatment. Implementation and results for Bayesian model comparison (using RJMCMC) which correspond and evaluate the above hypotheses are provided in detail.

The remaining of the paper proceeds as follows: Section 2 provides Bayesian estimation of the parameters through an MCMC scheme. Properties of the produced chain are also examined. Section 3 describes Bayesian hypothesis testing through RJMCMC. Reduced models with either zero correlation ($\rho = 0$) and/or equal marginal means ($\lambda_1 = \lambda_2$) are compared to the KBGD. Furthermore, model diagnostics are discussed. Simulated examples and a real data application can be found in Section 4, while concluding remarks in Section 5.

2 Bayesian estimation of the parameters

Let $(x_1, y_1), \ldots, (x_n, y_n)$ be a random sample from (1) and set $\mathbf{x} = (x_1, \ldots, x_n)$, $\mathbf{y} = (y_1, \ldots, y_n)$, $s_1 = \sum x_i$, $s_2 = \sum y_i$. Consider Bayesian estimation of the parameters $\lambda_1, \lambda_2, \rho$ to which is assigned the prior distribution

$$\lambda_1, \lambda_2, \rho \sim f(\lambda_1|\rho) f(\lambda_2|\rho) f(\rho),$$

where

$$\lambda_j|\rho \sim \mathcal{G}(c_j, d_j/(1-\rho)) \text{ for } j = 1, 2 \text{ and } \rho \sim \mathcal{B}eta(c_3, d_3).$$

The discussion will be simplified if we reparametrize (1) by setting $\mu_j = \lambda_j/(1-\rho)$, $j = 1, 2$. Then, μ_1, μ_2, ρ are mutually independent with prior distributions

$$\mu_j \sim \mathcal{G}(c_j, d_j) \text{ for } j = 1, 2 \text{ and } \rho \sim \mathcal{B}eta(c_3, d_3).$$

The mixture representation of the data distribution leads to a complicated posterior. In such cases it is convenient to proceed by data augmentation treating the mixture parameters as unobserved random variables.

So, let $\boldsymbol{\kappa} = (\kappa_1, \ldots, \kappa_n)$ be the unobserved sample from the negative binomial distribution $\mathcal{NB}(v, 1-\rho)$ such that x_i, y_i are independent conditional on κ_i. It can be seen that conditional on $\mu_1, \mu_2, \rho, \mathbf{x}, \mathbf{y}$, the κ_i's are independent with probability mass function

$$f(\kappa_i|\mu_1, \mu_2, \rho, \mathbf{x}, \mathbf{y}) \propto \frac{(\rho\mu_1\mu_2 x_i y_i)^{\kappa_i}}{\Gamma(\kappa_i + v)\kappa_i!}\ ,\ \kappa_i = 0, 1, 2, \ldots.$$

This is the Bessel distribution, $\mathcal{B}es\,(v-1, 2\sqrt{\rho\mu_1\mu_2 x_i y_i}\,)$, recently introduced by Yuan and Kalbfleisch (2000). For later use we note that $\mathsf{E}\{\kappa_i|\mu_1, \mu_2, \rho, \mathbf{x}, \mathbf{y}\} = \sqrt{\rho\mu_1\mu_2 x_i y_i} R_{v-1}(2\sqrt{\rho\mu_1\mu_2 x_i y_i})$, where $R_{v-1}(z) = I_v(z)/I_{v-1}(z)$ is the Bessel quotient. Notice also that for $v \geq 1/2$ it holds $0 < R_{v-1}(z) < 1$, $\forall\, z > 0$ (see Amos, 1974). Now, conditional on $\boldsymbol{\kappa}, \mathbf{x}, \mathbf{y}$, the parameters μ_1, μ_2, ρ are independent with distributions

$$\mu_j|\boldsymbol{\kappa}, \mathbf{x}, \mathbf{y} \sim \mathcal{G}(c_j + nv + \textstyle\sum \kappa_i, d_j + s_j),\ j = 1, 2,\ \rho|\boldsymbol{\kappa}, \mathbf{x}, \mathbf{y} \sim \mathcal{B}eta(c_3 + \textstyle\sum \kappa_i, d_3 + nv).$$

The Gibbs sampler sequentially simulates from

$$\begin{array}{rcl} \mu_1|\mathbf{x}, \mathbf{y}, \boldsymbol{\kappa} & \sim & \mathcal{G}\,(c_1 + nv + \sum \kappa_i, d_1 + s_1) \\ \mu_2|\mathbf{x}, \mathbf{y}, \boldsymbol{\kappa} & \sim & \mathcal{G}\,(c_2 + nv + \sum \kappa_i, d_2 + s_2) \\ \rho|\mathbf{x}, \mathbf{y}, \boldsymbol{\kappa} & \sim & \mathcal{B}eta\,(c_3 + \sum \kappa_i, d_3 + nv) \\ \kappa_i|\mathbf{x}, \mathbf{y}, \mu_1, \mu_2, \rho & \sim & \mathcal{B}es\left(v-1, 2\sqrt{\rho\mu_1\mu_2 x_i y_i}\right),\ i = 1, 2, \ldots, n. \end{array} \tag{2}$$

Simulation from gamma and beta distributions is a well–known task. On the other hand, algorithms for simulating Bessel variates have been provided by Devroye (2002) and Iliopoulos and Karlis (2003).

2.1 Geometric ergodicity

Denote by $\Theta = (0, \infty)^2 \times [0, 1) \times \mathbf{Z}_+^n$ the state space of the Markov chain arising from the Gibbs sampling scheme (2), $\mathcal{B}(\Theta)$ the associated σ–algebra, $P(\boldsymbol{\theta}, \cdot)$ the transition kernel and π its stationary distribution. Let $\boldsymbol{\theta} = (\mu_1, \mu_2, \rho, \boldsymbol{\kappa})$, $\boldsymbol{\theta}' = (\mu_1', \mu_2', \rho', \boldsymbol{\kappa}')$ be two consecutive states of the chain. It is easily seen that the transition kernel has density

$$p(\boldsymbol{\theta}, \boldsymbol{\theta}') = f(\mu_1'|\kappa) f(\mu_2'|\kappa) f(\rho'|\kappa) \prod_{i=1}^{n} f(\kappa_i'|\mu_1', \mu_2', \rho'), \tag{3}$$

where $\kappa = \sum_{i=1}^{n} \kappa_i$.

We will show that (at least when $v \geq 1/2$) the chain is geometrically ergodic. This means that the chain converges in total variation to π at an exponential rate implying further that under some additional conditions the central limit theorem holds for certain functions of interest.

It is straightforward to see that the chain is π–irreducible, aperiodic and Harris recurrent. Hence, we need to establish a drift condition and an associated minorization condition (see for example Rosenthal, 1995 and Jones and Hobert, 2001). Recall that a drift condition holds if for some function $V : \Theta \to [0, \infty)$, some $a \in (0, 1)$ and some $b \in (0, \infty)$,

$$\mathsf{E}\{V(\boldsymbol{\theta}')|\, \boldsymbol{\theta}\} \leq aV(\boldsymbol{\theta}) + b, \ \forall\, \boldsymbol{\theta} \in \Theta.$$

On the other hand, a minorization condition holds if for some probability measure Q, some π–positive set $C \in \mathcal{B}(\Theta)$ and some $\varepsilon > 0$,

$$P(\boldsymbol{\theta}, A) \geq \varepsilon Q(A), \ \forall\, \boldsymbol{\theta} \in C, \ A \in \mathcal{B}(\Theta). \tag{4}$$

In fact this is a special minorization condition holding for the first step of the chain. Then, the chain is geometrically ergodic if (4) holds with $C = \{\boldsymbol{\theta} \in \Theta : V(\boldsymbol{\theta}) \leq d\}$ for any $d > 2b/(1 - a)$.

A drift condition. Let $V(\boldsymbol{\theta}) = \kappa = \sum \kappa_i$. Then,

$$\begin{aligned}
\mathsf{E}\{V(\boldsymbol{\theta}')|\boldsymbol{\theta}\} &= \sum_{i=1}^{n} \mathsf{E}\left\{\mathsf{E}[\kappa_i'|\mu_1', \mu_2', \rho']|\, \boldsymbol{\kappa}\right\} \\
&= \sum_{i=1}^{n} \mathsf{E}\left\{ \sqrt{\rho'\mu_1'\mu_2' x_i y_i}\, R_{v-1}\left(2\sqrt{\rho'\mu_1'\mu_2' x_i y_i}\right) \Big| \boldsymbol{\kappa}\right\} \\
&\leq R_{v-1}^* \sum_{i=1}^{n} \mathsf{E}\left\{ \sqrt{\mu_1'\mu_2' x_i y_i} \,\Big|\, \boldsymbol{\kappa}\right\} \qquad [\text{where } R_{v-1}^* = \sup_z R_{v-1}(z)] \\
&= R_{v-1}^* \sum_{i=1}^{n} \sqrt{\frac{x_i y_i}{(d_1 + s_1)(d_2 + s_2)}} \prod_{j=1}^{2} \frac{\Gamma(c_j + nv + \kappa + 1/2)}{\Gamma(c_j + nv + \kappa)} \\
&\leq R_{v-1}^* \sum_{i=1}^{n} \sqrt{\frac{x_i y_i}{(d_1 + s_1)(d_2 + s_2)}} \prod_{j=1}^{2} \sqrt{c_j + nv + \kappa} \qquad (5) \\
&\leq R_{v-1}^* \sum_{i=1}^{n} \sqrt{\frac{x_i y_i}{(d_1 + s_1)(d_2 + s_2)}}\, (\max\{c_1, c_2\} + nv + \kappa) \\
&\leq aV(\boldsymbol{\theta}) + b,
\end{aligned}$$

where

$$a = R_{v-1}^* \sum_{i=1}^{n} \sqrt{\frac{x_i y_i}{(d_1 + s_1)(d_2 + s_2)}} \ \text{ and } \ b = (\max\{c_1, c_2\} + nv)a. \tag{6}$$

For the inequality in (5) we have used the fact that $\Gamma(x + 1/2)/\Gamma(x) \leq x^{1/2}$ (see Kershaw, 1983). When $v \geq 1/2$, $R_{v-1}^* = 1$ and thus, by the Cauchy–Schwarz inequality, $0 < a < 1$. Hence the required drift condition has been established, at least for $v \geq 1/2$. The above inequality, may hold also for smaller v's depending on the data and the prior hyperparameters d_1, d_2.

A minorization condition. We will prove that for any $\kappa^* > 0$, the set $C = \{\boldsymbol{\theta} \in \Theta : V(\boldsymbol{\theta}) \leq \kappa^*\}$ satisfies (4). The technique is based on Rosenthal (1995) (see also Jones and Hobert, 2001).

From (3) it is obvious that

$$p(\boldsymbol{\theta},\boldsymbol{\theta}') \geq f(\boldsymbol{\kappa}'|\mu_1',\mu_2',\rho') \inf_{\kappa\in[0,\kappa^*]} \{f(\mu_1'|\kappa)f(\mu_2'|\kappa)f(\rho'|\kappa)\} = f(\boldsymbol{\kappa}'|\mu_1',\mu_2',\rho')f^*(\mu_1')f^*(\mu_2')f^*(\rho'),$$

where the superscript "$*$" denotes infimum with respect to $\kappa \in [0,\kappa^*]$. In order to proceed we will need the following lemmas.

LEMMA 1. *Let $f(x;a,b)$ denote the $\mathcal{G}(a,b)$ density, i.e., $f(x;a,b) = \Gamma(a)^{-1}b^a x^{a-1}e^{-bx}$, $x>0$. Then, for $0<a_1<a_2$,*

$$\inf_{a\in[a_1,a_2]} f(x;a,b) = \begin{cases} f(x;a_2,b), & x \leq x^* \\ f(x;a_1,b), & x > x^*, \end{cases}$$

with $x^* = b^{-1}\left\{\frac{\Gamma(a_2)}{\Gamma(a_1)}\right\}^{\frac{1}{a_2-a_1}}$.

LEMMA 2. *Let $f(x;a,b)$ denote the $\mathcal{B}eta(a,b)$ density, i.e., $f(x;a,b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}x^{a-1}(1-x)^{b-1}$, $0<x<1$. Then, for $0<a_1<a_2$,*

$$\inf_{a\in[a_1,a_2]} f(x;a,b) = \begin{cases} f(x;a_2,b), & x \leq x^* \\ f(x;a_1,b), & x > x^*, \end{cases}$$

with $x^* = \left\{\frac{\Gamma(a_1+b)\Gamma(a_2)}{\Gamma(a_2+b)\Gamma(a_1)}\right\}^{\frac{1}{a_2-a_1}}$.

Set now

$$\begin{aligned} \varepsilon &= \sum_{\boldsymbol{\kappa}} \int_\rho \int_{\mu_2} \int_{\mu_1} f(\boldsymbol{\kappa}|\mu_1,\mu_2,\rho) f^*(\mu_1)f^*(\mu_2)f^*(\rho) d\mu_1 d\mu_2 d\rho \\ &= \left(\int_{\mu_1} f^*(\mu_1)d\mu_1\right)\left(\int_{\mu_2} f^*(\mu_2)d\mu_2\right)\left(\int_\rho f^*(\rho)d\rho\right) \end{aligned} \tag{7}$$

and

$$q(\boldsymbol{\theta}') = \varepsilon^{-1} f(\boldsymbol{\kappa}'|\mu_1',\mu_2',\rho')f^*(\mu_1')f^*(\mu_2')f^*(\rho'), \tag{8}$$

with f^*'s being the corresponding infimums arising from Lemmas 1 and 2. It can be easily seen that (4) holds with $Q(A) = \int_A q(d\boldsymbol{\theta})$. Note also that the computation of ε can be done by evaluating four incomplete gamma and two incomplete beta functions. Moreover, it does not depend on the observed data.

By proving the drift and minorization conditions we have established the geometric ergodicity of the chain. Based on this property, one could try using Rosenthal's (1995) Theorem 12 in order to obtain crude quantitative bounds on the convergence to stationarity after a finite number of iterations. However, in cases where x's and y's are strongly correlated, ε is too close to zero making this approach of limited use (see example in Subsection 2.3).

2.2 Estimation of the parameters

Ergodic means and Rao–Blackwellized estimators. Let $\boldsymbol{\theta}_j = (\mu_1^{(j)},\mu_2^{(j)},\rho^{(j)},\boldsymbol{\kappa}^{(j)})$, $j=1,2,\ldots,M$ be the simulated output and set $\kappa^{(j)} = \sum \kappa_i^{(j)}$, $j=1,2,\ldots,M$. Then, under squared error loss, the Bayes estimator of every (integrable) function $h:\Theta\to\mathbf{R}$ can be approximated by the ergodic mean

$M^{-1}\sum_{j=1}^{M} h(\boldsymbol{\theta}_j)$ which by the ergodic theorem converges almost surely to $\mathsf{E}_\pi[h(\boldsymbol{\theta})] = \int h(\boldsymbol{\theta})\pi(d\boldsymbol{\theta})$. In particular, the Bayes estimators of λ_1, λ_2, ρ, $\phi = \lambda_1/\lambda_2$ can be approximated by

$$\hat{\lambda}_i = \frac{1}{M}\sum_{j=1}^{M} \mu_i^{(j)}(1-\rho^{(j)}),\ i=1,2,\ \hat{\rho} = \frac{1}{M}\sum_{j=1}^{M}\rho^{(j)},\ \hat{\phi} = \frac{1}{M}\sum_{j=1}^{M}\mu_1^{(j)}/\mu_2^{(j)},$$

respectively. Moreover, one can also consider the Rao–Blackwellized estimators

$$\begin{aligned}
\tilde{\lambda}_i &= \frac{1}{M}\sum_{j=1}^{M} \mathsf{E}\left\{ \mu_i^{(j)}(1-\rho^{(j)}) \middle| \kappa^{(j)} \right\} \frac{d_3+nv}{d_i+s_i} = \frac{1}{M}\sum_{j=1}^{M} \frac{c_i+nv+\kappa^{(j)}}{c_3+d_3+nv+\kappa^{(j)}},\ i=1,2,\\
\tilde{\rho} &= \frac{1}{M}\sum_{j=1}^{M} \mathsf{E}\left\{ \rho^{(j)} \middle| \kappa^{(j)} \right\} = \frac{1}{M}\sum_{j=1}^{M} \frac{c_3+\kappa^{(j)}}{c_3+d_3+nv+\kappa^{(j)}},\\
\tilde{\phi} &= \frac{1}{M}\sum_{j=1}^{M} \mathsf{E}\left\{ \mu_1^{(j)}/\mu_2^{(j)} \middle| \kappa^{(j)} \right\} \frac{d_2+s_2}{d_1+s_1} = \frac{1}{M}\sum_{j=1}^{M} \frac{c_1+nv+\kappa^{(j)}}{c_2+nv+\kappa^{(j)}-1}.
\end{aligned}$$

The results of Liu *et al.* (1994) ensure that in stationary Markov chains constructed by data augmentation, as is (2), estimators arising by Rao–Blackwellization reduce the asymptotic variances.

REMARK. Note that the forms of Rao–Blackwellized estimators together with their almost sure convergence to the true values give explicit bounds for the Bayes estimators. Since $0 \le \kappa^{(j)} < \infty$, $\forall\, j = 1,2,\ldots,M$, we have that

$$\begin{aligned}
\frac{d_3+nv}{d_i+s_i}\min\left\{1, \frac{c_i+nv}{c_3+d_3+nv}\right\} \le\ & \hat{\lambda}_i\ \le \frac{d_3+nv}{d_i+s_i}\max\left\{1, \frac{c_i+nv}{c_3+d_3+nv}\right\} \text{ (a.s.)},\ i=1,2,\\
\frac{c_3}{c_3+d_3+nv} \le\ & \hat{\rho}\ < 1 \text{ (a.s.)},\\
\frac{d_2+s_2}{d_1+s_1}\min\left\{1, \frac{c_1+nv}{c_2+nv-1}\right\} \le\ & \hat{\phi}\ \le \frac{d_2+s_2}{d_1+s_1}\max\left\{1, \frac{c_1+nv}{c_2+nv-1}\right\} \text{ (a.s.)}.
\end{aligned}$$

In the same way, it can be seen that the above parametic functions have finite higher moments. We actually need moments of order $2+\epsilon$ for some $\epsilon > 0$ for the central limit theorem to hold for the above ergodic means. Thus, if we could estimate consistently the asymptotic variances we would also obtain corresponding confidence intervals for the MCMC approximations.

Regenerative sampling. In order to obtain consistent estimators of the variances of various ergodic means we will use the method of Mykland *et al.* (1995). It is applicable since the chain is geometrically ergodic (see Hobert *et al.*, 2002). Recall that the method proceeds as follows. Assume the existence of a measurable function $s(\cdot)$ and a probability measure Q on $\mathcal{B}(\Theta)$ such that

$$P(\boldsymbol{\theta}, A) \ge s(\boldsymbol{\theta})Q(A),\ \forall\, \boldsymbol{\theta} \in \Theta,\ A \in \mathcal{B}(\Theta). \tag{9}$$

Note that the existence of such an $s(\cdot)$ and Q has been already shown in Subsection 2.1 it can be taken $s(\boldsymbol{\theta}) = \varepsilon I_{[0,\kappa^*]}(\kappa)$ and Q the corresponding "infimum" probability distribution. Consider the chain $(\boldsymbol{\theta}_1,\delta_1),(\boldsymbol{\theta}_2,\delta_2),\ldots$, where each δ_j is a Bernoulli random variate with probability of success $s(\boldsymbol{\theta}_j)$, and $\boldsymbol{\theta}_1 \sim Q(\cdot)$. Setting $A = \Theta$ in (9) it is seen that $s(\cdot) \in [0,1]$, hence the δ's are well defined. Now, if $\delta_j = 1$ then draw $\boldsymbol{\theta}_{j+1}$ from $Q(\cdot)$ *independently* of the past whereas if $\delta_j = 0$ then draw $\boldsymbol{\theta}_{j+1}$ from the residual distribution $R(\boldsymbol{\theta}_j,\cdot) = \{P(\boldsymbol{\theta}_j,\cdot) - s(\boldsymbol{\theta}_j)Q(\cdot)\}\{1-s(\boldsymbol{\theta}_j)\}^{-1}$. Notice that the transition kernel of the marginal chain $\boldsymbol{\theta}_1,\boldsymbol{\theta}_2,\ldots$ remains $P(\boldsymbol{\theta},\cdot)$. Clearly, the (random) times for which $\boldsymbol{\theta}_{j+1} \sim Q(\cdot)$

constitute regeneration times. Let $1 = \tau_1 < \ldots < \tau_R < \tau_{R+1}$ be the first $R+1$ regeneration times and set $N_t = \tau_{t+1} - \tau_t$, $S_t = \sum_{j=\tau_t}^{\tau_{t+1}-1} h(\boldsymbol{\theta}_j)$, $t = 1, 2, \ldots, R$. The pairs $(N_1, S_1), (N_2, S_2), \ldots, (N_R, S_R)$ are iid and as $R \to \infty$,

$$\bar{S}/\bar{N} = \frac{1}{\tau_{R+1}} \sum_{j=1}^{\tau_{R+1}-1} h(\boldsymbol{\theta}_j) \xrightarrow{\text{a.s}} \mathsf{E}_\pi(h) \quad \text{and} \quad R^{1/2}[\bar{S}/\bar{N} - \mathsf{E}_\pi(h)] \xrightarrow{\text{w}} \mathcal{N}(0, \sigma_h^2),$$

where $\sigma_h^2 = \mathsf{E}_Q\{[S_1 - N_1 \mathsf{E}_\pi(h)]^2\}/\{\mathsf{E}_Q(N_1)\}^2$. The asymptotic variance σ_h^2 can be consistently estimated by

$$\hat{\sigma}_h^2 = \frac{\sum_{t=1}^{R}(S_t - N_t\,\bar{S}/\bar{N})^2}{R\bar{N}^2}. \tag{10}$$

Note finally, that assuming $P(\boldsymbol{\theta}, \cdot)$ and $Q(\cdot)$ have densities $p(\boldsymbol{\theta}, \cdot)$ and $q(\cdot)$, the Q–draws can be avoided using the following procedure. After drawing $\boldsymbol{\theta}_j$ generate $\boldsymbol{\theta}_{j+1}$ rather than δ_j from the original kernel $P(\boldsymbol{\theta}_j, \cdot)$. Then, draw δ_j conditional on $\boldsymbol{\theta}_j$, $\boldsymbol{\theta}_{j+1}$, that is, generate a Bernoulli random variate with probability of success $s(\boldsymbol{\theta}_j)q(\boldsymbol{\theta}_{j+1})/p(\boldsymbol{\theta}_j, \boldsymbol{\theta}_{j+1})$. If it takes the value one a regeneration at time $j+1$ has been occured. The same method can be used to avoid drawing $\boldsymbol{\theta}_1$ from Q. Just start from an arbitrary distribution and proceed as before until the first regeneration. Then set $\boldsymbol{\theta}_1$ equal to this draw and discard all previous values.

Returning back to our context consider $\kappa_1, \kappa_2 \in \mathbf{Z}_+$ with $\kappa_1 < \kappa_2$ and set $C = (0, \infty)^2 \times (0, 1) \times \{\boldsymbol{\kappa} \in \mathbf{Z}_+^n : \kappa_1 \leq \sum \kappa_i \leq \kappa_2\} \in \mathcal{B}(\Theta)$. Let ε, $q(\cdot)$ be as in (7), (8) respectively but with f^*'s now denoting the infimums over $[\kappa_1, \kappa_2]$ rather than $[0, \kappa^*]$. Let also $s(\boldsymbol{\theta}) = \varepsilon I_C(\boldsymbol{\theta}) = \varepsilon I_{[\kappa_1, \kappa_2]}(\sum \kappa_i)$. Then it can be seen that (9) holds. Note that ε does not need to be calculated at all. Set

$$\mu_i^* = \frac{1}{d_i + s_i} \left\{ \frac{\Gamma(c_i + nv + \kappa_2)}{\Gamma(c_i + nv + \kappa_1)} \right\}^{\frac{1}{\kappa_2 - \kappa_1}}, \; i = 1, 2, \; \rho^* = \left\{ \frac{\Gamma(c_3 + \kappa_1 + d_3 + nv)\Gamma(c_3 + \kappa_2)}{\Gamma(c_3 + \kappa_2 + d_3 + nv)\Gamma(c_3 + \kappa_1)} \right\}^{\frac{1}{\kappa_2 - \kappa_1}}$$

and starting with an arbitrary $\boldsymbol{\kappa} \in \mathbf{Z}_+^n$ with $\kappa = \sum \kappa_i \notin [\kappa_1, \kappa_2]$ run sequentially (2). When κ hits $[\kappa_1, \kappa_2]$ draw the next μ_1, μ_2, ρ in the usual way followed by a Bernoulli random variate δ with success probability

$$\frac{\Gamma(c_1 + nv + \kappa)\Gamma(c_2 + nv + \kappa)\Gamma(c_3 + \tilde{\kappa}(\rho) + d_3 + nv)\Gamma(c_3 + \kappa)}{\Gamma(c_1 + nv + \tilde{\kappa}(\mu_1))\Gamma(c_2 + nv + \tilde{\kappa}(\mu_2))\Gamma(c_3 + \kappa + d_3 + nv)\Gamma(c_3 + \tilde{\kappa}(\rho))} \times \mu_1^{\tilde{\kappa}(\mu_1) - \kappa} \mu_2^{\tilde{\kappa}(\mu_2) - \kappa} \rho^{\tilde{\kappa}(\rho) - \kappa},$$

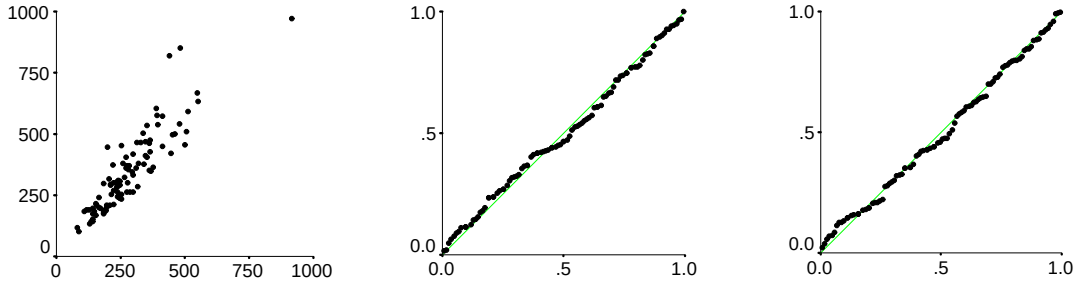
where $\tilde{\kappa}(\mu_1) = \kappa_2 I_{(0, \mu_1^*]}(\mu_1) + \kappa_1 I_{(\mu_1^*, \infty)}(\mu_1)$ and similarly for $\tilde{\kappa}(\mu_2)$ and $\tilde{\kappa}(\rho)$. If $\delta = 1$ then set $\boldsymbol{\theta}_1 = (\mu_1, \mu_2, \rho, \boldsymbol{\kappa}')$, $\boldsymbol{\kappa}'$ being the next $\boldsymbol{\kappa}$ value generated else continue until $\boldsymbol{\theta}_1$ is obtained.

Due to the ergodicity of the chain κ_1, κ_2 may be arbitrary since the set $[\kappa_1, \kappa_2]$ will be hit infinitely often. However, in practice the set should be chosen to have small waiting times but on the other hand not being too wide, or else ε will be negligible. So, it is suggested to inspect the $\sum \kappa_i$ values from a preliminary run of (2) in order to choose an appropriate set.

2.3 Example: The Missouri river flow data

We consider for illustration purposes the Missouri river flow data analyzed also in a different context by Nagaraja *et al.* (2002). The data represent annual instantaneous peak flow measurements (in 10^{-3} cubic feets per second) at Booneville, MO and Hermann, MO, during 1898–1997. Despite their time series nature, Nagaraja *et al.* (2002) justify why the data can be considered as a random sample from some bivariate distribution.

In Figure 1 one can see the scatterplot and gamma P-P plots of the data. The gamma distribution seems to fit well to their marginal distributions and the scatterplot has a typical form of KBGD with



(a) Booneville vs. Hermann (b) Booneville (c) Hermann

Figure 1: Figures of Missouri river flow data; (a) Scatterplot of Booneville (horizontal axis) vs. Hermann (vertical axis) peak flow measurements; (b) Gamma P-P plot of Booneville measurements; (c) Gamma P-P plot of Hermann measurements.

Parameter	Bayes estimate	Std. error	99% c.i.
λ_1	0.019234	10^{-8}	[0.019224, 0.019245]
λ_2	0.015764	10^{-8}	[0.015755, 0.015733]
ϕ	1.220355	5.5×10^{-6}	[1.220086, 1.220625]
ρ	0.896377	1.1×10^{-3}	[0.892554, 0.900200]

Table 1: Bayes estimates for the Missouri river flow data

large correlation. We chose $v = 5.4$ for the latter distribution which is the mean of the maximum likelihood estimators of the two marginal gamma shape parameters and $c_i = d_i = 0.001$, $i = 1, 2$, $c_3 = d_3 = 0.5$ for the priors' hyperparameters. Then, a and b in (6) are approximately 0.9955 and 537.6 and thus $2b/(1-a) \approx 238933$. Unfortunately, for any $\kappa^* > 238933$, ε is negligible making Rosenthal's (1995) bound on total variation practically uncalculatable. On the other hand, the regenerative sampling algorithm was run for $R = 500$ regenerations with $[\kappa_1, \kappa_2] = [4750, 4780]$. The mean number of iterations per regeneration was 96.64 and its coefficient of variation 0.8% suggesting that the variance estimates in (10) are adequate (see Mykland *et al.*, 1995). Detailed results are provided in Table 1.

3 Bayesian hypothesis tests using MCMC

Since KBGD degenerates to the product of two independent gamma distributions for $\rho = 0$, testing the hypothesis $H_0 : \rho = 0$ versus $H_1 : \rho \neq 0$ is of major interest. A test of secondary interest may be based on assessing whether the parameters λ_1 and λ_2 (or equivalently μ_1 and μ_2) of KBGD are equal.

According to the Bayesian paradigm, hypothesis testing concerning two competing models m_0 and m_1 are based on posterior models odds

$$PO_{01} = \frac{f(m_0|\text{data})}{f(m_1|\text{data})} \frac{f(\text{data}|m_0)}{f(\text{data}|m_1)} \times \frac{f(m_0)}{f(m_1)},$$

where $f(m)$ and $f(m|\text{data})$ are the prior and posterior probabilities of model m and $f(\text{data}|m)$ is the marginal likelihood of model. When no information is available, a default choice for prior model probabilities is $f(m_0) = f(m_1) = 1/2$. In this case the posterior model odds degenerates to the Bayes factor.

If we are interested in the comparison of more than two competing models then inference is based on the posterior distribution $f(m|\text{data})$ itself.

The integrals involved in the calculation of posterior model odds are analytically tractable only in certain restricted cases, therefore we may calculate them by asymptotic approximations or Monte Carlo methods. Here we facilitate the reversible jump Markov chain Monte Carlo (RJMCMC) algorithm of Green (1995) in order to evaluate various models based on KBGD.

3.1 Using reversible jump for KBGD

RJMCMC is a generalization of simple Metropolis algorithms sampling between models of different dimensions. Assume that model m has associated parameter vector $\boldsymbol{\theta}_{(m)}$ of dimension $d(\boldsymbol{\theta}_{(m)})$ and denote the current state of the Markov chain by $(m, \boldsymbol{\theta}_{(m)})$. Then the algorithm is described by the following steps:

- Generate the model parameters $\boldsymbol{\theta}_{(m)}$ of model m from $f(\boldsymbol{\theta}_{(m)}|\text{data}, m)$.
- Propose a new model m' with probability $j(m, m')$.
- Generate $\mathbf{u}$ (which can be of lower dimension than $\boldsymbol{\theta}'_{(m')}$) from a specified proposal density $q(\mathbf{u}|\boldsymbol{\theta}_{(m)}, m, m')$.
- Set $(\boldsymbol{\theta}'_{(m')}, \mathbf{u}') = g_{m,m'}(\boldsymbol{\theta}_{(m)}, \mathbf{u})$, where $g_{m,m'}$ is a specified invertible function. Hence $d(\boldsymbol{\theta}_{(m)}) + d(\mathbf{u}) = d(\boldsymbol{\theta}'_{(m')}) + d(\mathbf{u}')$.
- Accept the proposed move to model m' with probability

$$\alpha = \min\left\{1, \frac{f(\text{data}|m', \boldsymbol{\theta}'_{(m')}) f(\boldsymbol{\theta}'_{(m')}|m') f(m') j(m', m) q(\mathbf{u}'|\boldsymbol{\theta}_{(m)}, m', m)}{f(\text{data}|m, \boldsymbol{\theta}_{(m)}) f(\boldsymbol{\theta}_{(m)}|m) f(m) j(m, m') q(\mathbf{u}|\boldsymbol{\theta}_{(m')}, m, m')} \left|\frac{\partial g(\boldsymbol{\theta}_{(m)}, \mathbf{u})}{\partial(\boldsymbol{\theta}_{(m)}, \mathbf{u})}\right|\right\}.$$

In this section we focus on the comparison of four models: the independent gamma model (m_1) with common $\lambda_1 = \lambda_2 = \lambda$ and (parameter vector) $\boldsymbol{\theta}_{(m_1)} = (\lambda)$, KBGD ($m_2$) with common $\lambda_1 = \lambda_2 = \lambda$ and $\boldsymbol{\theta}_{(m_2)} = (\lambda, \rho)$, the independent gamma model (m_3) with $\lambda_1 \neq \lambda_2$ and $\boldsymbol{\theta}_{(m_3)} = (\lambda_1, \lambda_2)$ and the unrestricted KBGD (m_4) with $\boldsymbol{\theta}_{(m_4)} = (\lambda_1, \lambda_2, \rho)$.

We consider the following prior structure for the parameters of each model:

$$\begin{aligned}
f(\boldsymbol{\theta}_{(m_1)}|m_1) &= f(\lambda|m_1), \\
f(\boldsymbol{\theta}_{(m_2)}|m_2) &= f(\lambda, \rho|m_2) = f(\lambda|\rho, m_2) f(\rho|m_2), \\
f(\boldsymbol{\theta}_{(m_3)}|m_3) &= f(\lambda_1, \lambda_2|m_3) = f(\lambda_1|m_3) f(\lambda_2|m_3), \\
f(\boldsymbol{\theta}_{(m_4)}|m_4) &= f(\lambda_1, \lambda_2, \rho|m_4) = f(\lambda_1|\rho, m_4) f(\lambda_2|\rho, m_4) f(\rho|m_4),
\end{aligned}$$

where

$$f(\lambda|m_j) \sim \mathcal{G}(c_{0j}, d_{0j}) \text{ for } j = 1, 2, \tag{11}$$

$$f(\lambda_i|\rho, m_j) \sim \mathcal{G}(c_{ij}, d_{ij}/(1-\rho)) \text{ for } i = 1, 2,\ j = 3, 4, \tag{12}$$

$$f(\rho|m_j) \sim \mathcal{B}eta(c_{3j}, d_{3j}) \text{ for } j = 2, 4. \tag{13}$$

In order to make the MCMC more flexible we split the model selection procedure in two steps by introducing two binary latent indicators γ_1 and γ_2 taking values one if $\rho \neq 0$ and $\lambda_1 \neq \lambda_2$, respectively, and zero otherwise. The model indicator m is simply a function of the two binary indicators with $m = 1 + \gamma_1 + 2\gamma_2$. We construct a RJMCMC algorithm which can be summarized as follows.

1. (a) Propose $\gamma_1' = 1 - \gamma_1$ with probability one.

   (b) If $\gamma_1 = 0$ then generate ρ from $q_\rho(\rho) = \mathcal{B}eta(\tilde{a}_\rho, \tilde{b}_\rho)$, where $\tilde{a}_\rho$ and $\tilde{b}_\rho$ are tuning parameters of the RJMCMC algorithm and are specified appropriately to achieve high acceptance rate. For details on their specification see Subsection 3.3.

   (c) Accept the proposed move with probability $\alpha = \min\left(1, A^{1-2\gamma_1}\right)$ where

$$\begin{aligned} A &= \left[\frac{1}{(\rho\lambda_1\lambda_2)^n \prod x_i y_i}\right]^{\frac{v-1}{2}} \exp\left\{-\frac{\rho(\lambda_1 s_1 + \lambda_2 s_2)}{1-\rho}\right\}\left[\prod_{i=1}^{n} I_{v-1}\left(\frac{2\sqrt{\rho\lambda_1\lambda_2 x_i y_i}}{1-\rho}\right)\right]\left(\frac{\Gamma(v)}{1-\rho}\right)^n \\ &\times \frac{f(\lambda_1|\rho,\gamma_1=1,\gamma_2)}{f(\lambda_1|\gamma_1=0,\gamma_2)}\left(\frac{f(\lambda_2|\rho,\gamma_1=1,\gamma_2)}{f(\lambda_2|\gamma_1=0,\gamma_2)}\right)^{\gamma_2}\frac{f(\rho|\gamma_1=1,\gamma_2)}{q_\rho(\rho)}\frac{f(\gamma_1=1,\gamma_2)}{f(\gamma_1=0,\gamma_2)} \end{aligned}$$

   (d) If the proposed move is accepted then:

      i. If $\gamma_1' = 1$ then generate $\kappa_i \sim \mathcal{B}es(v-1, 2\sqrt{\rho\lambda_1\lambda_2 x_i y_i}/(1-\rho))$, $i = 1, \ldots, n$.

      ii. If $\gamma_1' = 0$ then set $\kappa_i = 0$, $i = 1, \ldots, n$.

2. (a) Propose $\gamma_2' = 1 - \gamma_2$ with probability one.

   (b) i. If $\gamma_2 = 0$ then

      - Generate u from $q_u(u) \sim \mathcal{G}(\tilde{a}_u, \tilde{b}_u)$, where $\tilde{a}_u$ and $\tilde{b}_u$ are RJMCMC tuning parameters and are specified in Subsection 3.3.
      - Set $(\lambda_1, \lambda_2) = g(\lambda, u)$. We facilitate the transformation based on the geometric mean resulting to the RJMCMC setup $(\lambda_1, \lambda_2) = (\lambda\sqrt{u}, \lambda/\sqrt{u})$

      ii. If $\gamma_2 = 1$ then set $(\lambda, u) = g^{-1}(\lambda_1, \lambda_2)$. For the geometric mean transformation we have $(\lambda, u) = (\sqrt{\lambda_1\lambda_2}, \lambda_1/\lambda_2)$.

   (c) Accept the proposed move with probability $\alpha = \min\left(1, A^{1-2\gamma_2}\right)$ where A is given by

$$\begin{aligned} A &= \left(\frac{\lambda_1\lambda_2}{\lambda^2}\right)^{nv+\sum\kappa_i}\exp\left\{\frac{(\lambda-\lambda_1)s_1 + (\lambda-\lambda_2)s_2}{1-\rho}\right\}\frac{f(\lambda_1|\rho,\gamma_1,\gamma_2=1)}{f(\lambda|\rho,\gamma_1,\gamma_2=0)} \\ &\times \frac{f(\lambda_2|\rho,\gamma_1,\gamma_2=1)}{q_u(u)}\left(\frac{f(\rho|\gamma_1,\gamma_2=1)}{f(\rho|\gamma_1,\gamma_2=0)}\right)^{\gamma_1}\frac{f(\gamma_1,\gamma_2=1)}{f(\gamma_1,\gamma_2=0)}\left|\frac{\partial g(\lambda,u)}{\partial(\lambda,u)}\right|. \end{aligned}$$

3. Set $m = 1 + \gamma_1 + 2\gamma_2$.

4. Generate model parameters from the Gibbs steps in (2) with $\lambda_j = \mu_j(1-\rho)$ for $j = 1, 2$. When $\lambda_1 = \lambda_2$ the common λ is generated from a gamma distribution with parameters $c_{0m} + 2nv + 2\gamma_1\sum\kappa_i$ and $(d_{0m} + s_1 + s_2)/(1-\rho)$.

In Step 1 we consider directly the marginal distribution $f(\lambda_1, \lambda_2, \rho|\gamma_1, \gamma_2)$ eliminating the latent variables κ_i. This simplifies the reversible jump step presented in Step 1 considerably, since the two compared models differ only by one parameter (ρ). The main drawback of this approach is the calculation of the Bessel function for each data point which is computationally inefficient. Alternatively, we may construct a RJMCMC keeping the latent data κ_i. Using such an approach in Step 1 of the above algorithm we will have to additionally generate κ_i from $q_i(\kappa_i)$, $i = 1, \ldots, n$, and then accept the proposed move with probability $\alpha = \min\left(1, A^{1-2\gamma_1}\right)$ with A given by

$$\begin{aligned} A &= \frac{f(\mathbf{x}, \mathbf{y}, \boldsymbol{\kappa}|\lambda_1, \lambda_2, \rho, \gamma_1 = 1, \gamma_2)}{f(\mathbf{x}, \mathbf{y}|\lambda_1, \lambda_2, \rho = 0, \gamma_1 = 0, \gamma_2)}\frac{f(\lambda_1|\rho,\gamma_1=1,\gamma_2)}{f(\lambda_1|\gamma_1=0,\gamma_2)}\left(\frac{f(\lambda_2|\rho,\gamma_1=1,\gamma_2)}{f(\lambda_2|\gamma_1=0,\gamma_2)}\right)^{\gamma_2} \\ &\times \frac{f(\rho|\gamma_1=1,\gamma_2)}{q_\rho(\rho)\prod_{i=1}^{n} q_i(\kappa_i)}\frac{f(\gamma_1=1,\gamma_2)}{f(\gamma_1=0,\gamma_2)}. \end{aligned}$$

Natural choices for the proposal $q_i(\kappa_i)$ are the Poisson and negative binomial distributions. On the other hand, the approach of generating $n+1$ additional parameters in each model comparison may lead to a low acceptance rate and hence to slow down the convergence of the algorithm.

Concerning the proposed algorithm in Step 2, it works efficiently when the posterior distributions of ρ under models m_2 and m_4 are close. This may not happen when the sample means $\bar{x} = s_1/n$ and $\bar{y} = s_2/n$ are far away. In such cases, we will also need to propose new values for ρ in order to jump directly to the correct posterior distribution. This can be achieved by either proposing the new ρ deterministically (by a suitable transformation) or probabilistically using proposal densities (see the Metropolized Carlin and Chib algorithm as described by Dellaportas *et al.*, 2002). Moreover, since we propose new values for ρ, we should use the marginal likelihood of the KBGD in order to avoid generation of the latent variables κ_i. The algorithm (for $\gamma_1 = 1$) should be modified and therefore after Step 2(b) we generate new ρ using the following step:

2(c) Generate ρ' from a proposal $q_{\rho,\gamma_2}(\rho')$. Set $\rho_1 = \gamma_2\rho + (1-\gamma_2)\rho'$, $\rho_0 = (1-\gamma_2)\rho + \gamma_2\rho'$ and accept the proposed move with probability $\alpha = \min\left(1, A^{1-2\gamma_2}\right)$ where

$$\begin{aligned} A &= \left(\frac{1-\rho_0}{1-\rho_1}\right)^n \left(\frac{\rho_0}{\rho_1}\right)^{\frac{n(v-1)}{2}} \exp\left\{\frac{\lambda(s_1+s_2)}{1-\rho_0} - \frac{\lambda_1 s_1 + \lambda_2 s_2}{1-\rho_1}\right\} \\ &\times \left[\prod_{i=1}^{n}\left\{I_{v-1}\left(\frac{2\sqrt{\rho_1\lambda_1\lambda_2 x_i y_i}}{1-\rho_1}\right)\left[I_{v-1}\left(\frac{2\lambda\sqrt{\rho_0 x_i y_i}}{1-\rho_0}\right)\right]^{-1}\right\}\right] \frac{f(\lambda_1|\rho_1,\gamma_1,\gamma_2=1)}{f(\lambda|\rho_0,\gamma_1,\gamma_2=0)} \\ &\times \frac{f(\lambda_2|\rho_1,\gamma_1,\gamma_2=1)}{q_u(u)} \left(\frac{q_{\rho,\gamma_2=1}(\rho_0)}{q_{\rho,\gamma_2=0}(\rho_1)} \frac{f(\rho_1|\gamma_1,\gamma_2=1)}{f(\rho_0|\gamma_1,\gamma_2=0)}\right)^{\gamma_1} \frac{f(\gamma_1,\gamma_2=1)}{f(\gamma_1,\gamma_2=0)}. \end{aligned} \tag{14}$$

If the proposed move is accepted, then we should also generate the new latent variables κ_i similarly to Step 1(d).

When ρ is proposed deterministically (for example $\rho_1 = C\rho_0$ or $\rho_1 = \rho_0 + C$; where C is a constant) then the ratio of proposals $q_{\rho,\gamma_2=1}(\rho_0)/q_{\rho,\gamma_2=0}(\rho_1)$ should be substituted by the corresponding derivative $\partial\rho_1/\partial\rho_0$. This RJMCMC variation is more general than the approach presented in Step 2 but involves more laborious computation since in (14) we need to compute the Bessel functions and additionally generate candidate values for ρ. In the majority of examples we have implemented our algorithm, the simpler version presented in Step 2 worked efficiently. In general, we suggest to estimate ρ assuming common and non-common means before running the RJMCMC. If their ratio is close to one then use the simpler approach of Step 2.

3.2 Specification of prior parameters for model comparison

A challenging and difficult task in Bayesian model comparison and hypothesis testing is the specification of prior distributions. Difficulties mainly arise due to the behavior of posterior model odds as noted by Lindley (1957) and Bartlett (1957). Essentially, we cannot use priors with large variance (which are thought to be noninformative) because in such case we will a posteriori fully support the simplest model.

Since ρ lies in the interval $[0,1)$ we may use standard noninformative prior distributions without any major problem. We propose to use a beta prior distribution (13) with $c_{3j} = d_{3j} = 1$ for $j = 2, 4$, i.e. the uniform distribution. Hence focus is given in the specification of prior distributions for λ, λ_1 and λ_2 which take positive values and consequently any prior expressing low information via an extremely large variance will activate Lindley–Bartlett paradox as discussed above. In order to specify plausible prior

distributions for the model comparison of interest we will use two approaches. The first approach uses "power prior" distributions (see Chen *et al.*, 2000) while the second tries to match the prior distributions used for different models.

Let us first consider for the independent gamma models m_1 and m_3 with parameter vectors $\boldsymbol{\theta}_{m_1} = (\lambda)$ and $\boldsymbol{\theta}_{m_3} = (\lambda_1, \lambda_2)$ that we have a priori imaginary data $(\mathbf{x}^*, \mathbf{y}^*) = ((x_i^*, y_i^*); i = 1, \ldots, n^*)$. We wish to use these a priori data for defining weak prior distribution which will be used for our model selection procedure with our actual data. For this reason, we consider as prior the posterior distribution resulting by a usual low information prior distribution (denoted by f_0) and the a priori "available" data $(\mathbf{x}^*, \mathbf{y}^*)$ with likelihood weighted to account for a fraction w of their sample size n^*. Under this approach, the prior distribution of the parameters λ_1 and λ_2 for model m_3 will be defined by

$$f(\lambda_1, \lambda_2 | \mathbf{x}^*, \mathbf{y}^*, \gamma_1 = 1, \gamma_2 = 0) \propto f(\mathbf{x}^*, \mathbf{y}^* | \lambda_1, \lambda_2, \gamma_1 = 1, \gamma_2 = 0)^w f_0(\lambda_1, \lambda_2 | \gamma_1 = 1, \gamma_2 = 0),$$

where $0 \leq w \leq 1$ is a parameter controlling the weight of belief on the prior imaginary data. Usually this fraction is taken equal to one data point, hence $w = 1/(2n^*)$. More generally, we may set $w = \xi/(2n^*)$, where ξ is the number of data points that the prior accounts for. When ξ tends to zero, the prior accounts for a negligible amount of information in the posterior distribution. We choose $\xi = 1$ as a baseline and use different values of ξ for sensitivity analysis.

Considering pre–prior distributions $\lambda_i \sim \mathcal{G}(c_i^*, d_i^*)$, $i = 1, 2$, and $w = \xi/(2n^*)$, the prior distribution based on these arguements should be given by (12) with

$$c_{i3} = c_i^* + v\xi/2, \;\; i = 1, 2, \quad d_{13} = d_1^* + \bar{x}^*\xi/2 \;\; \text{and} \;\; d_{23} = d_2^* + \bar{y}^*\xi/2.$$

Using the same arguments for model m_1 and gamma pre–prior distribution with parameters c^* and d^*, we conclude to a gamma prior distribution of type (11) with

$$c_{01} = c^* + v\xi \;\; \text{and} \;\; d_{01} = d^* + (\bar{x}^* + \bar{y}^*)\xi/2.$$

The construction of prior distributions for model m_2 and m_3 is not as straightforward as for the simpler case of the independent gamma distributions. Moreover, since for the parameter ρ we wish to implement a uniform prior, when no information is available, we consider the same procedure for $f(\mathbf{x}, \mathbf{y} | \boldsymbol{\kappa}, \lambda_1, \lambda_2, \rho)$. Hence, assuming the same pre–prior distributions as above, the one point information power prior is given by

$$f(\lambda | \boldsymbol{\kappa}, \rho, m_2) \;\sim\; \mathcal{G}\left(c^* + (v + \bar{\kappa}^*)\xi, \; d^* + \tfrac{1}{2}(\bar{x}^* + \bar{y}^*)\xi/(1-\rho)\right), \tag{15}$$

$$f(\lambda_1 | \boldsymbol{\kappa}, \rho, m_4) \;\sim\; \mathcal{G}\left(c_1^* + \tfrac{1}{2}(v + \bar{\kappa}^*)\xi, \; d_1^* + \tfrac{1}{2}\bar{x}^*\xi/(1-\rho)\right) \tag{16}$$

$$f(\lambda_2 | \boldsymbol{\kappa}, \rho, m_4) \;\sim\; \mathcal{G}\left(c_2^* + \tfrac{1}{2}(v + \bar{\kappa}^*)\xi, \; d_2^* + \tfrac{1}{2}\bar{y}^*\xi/(1-\rho)\right). \tag{17}$$

Since $\boldsymbol{\kappa}^*$ are only latent a priori data we may substitute $\bar{\kappa}^*$ by an arbitrary a priori supported value. We choose $\bar{\kappa}^* = 0$ in order to center our prior round the simpler model without correlation. In order the pre–prior to be noninformative, we take $c_i^* = d_i^* = c^* = d^*$ equal to a very low positive number. Moreover, we set $\bar{x}^* = \bar{y}^* = \omega$ in order to center our low information prior round the hypothesis of equal means and let ω varying to examine the robustness of the procedure on different choices of imaginary prior data. Note that for small values of ω we can claim that we use a low information prior since the variance will be large.

3.3 Specification of proposal densities.

Another important topic in constructing an efficient RJMCMC algorithm is the correct specification of the proposal densities $q_\rho(\rho)$, $q_u(u)$ and $q_i(\kappa_i)$, $i = 1, \ldots, n$, used in Steps 1 and 2 of the RJMCMC algorithm described in Subsection 3.1. The most usual way is to estimate the parameters of the proposal densities using the moments from pilot MCMC runs (see Dellaportas and Forster, 1999 and Dellaportas *et al.*, 2002). In our case, we run the MCMC for the unrestricted KBGD and set

$$\tilde{a}_\rho \bar{\rho}\{\bar{\rho}(1-\bar{\rho})/s_\rho^2 - 1\}, \quad \tilde{b}_\rho = (1-\bar{\rho})\{\bar{\rho}(1-\bar{\rho})/s_\rho^2 - 1\} \text{ and } \tilde{a}_u = \bar{u}^2/s_u^2, \quad \tilde{b}_u = \bar{u}/s_u^2,$$

where $\bar{\rho}$, s_ρ^2 are the mean and the variance of ρ and $\bar{u}$, s_u^2 are the mean and the variance of λ_1/λ_2 estimated from the MCMC pilot run.

3.4 Alternative model adequacy and diagnostic measures.

In order to check the fit of the models one may use a variety of Bayesian p-values given by $p_i = \mathsf{P}\left\{D_i(\mathbf{x}^{rep}, \mathbf{y}^{rep}, \boldsymbol{\theta}_{(m)}, m) > D_i(\mathbf{x}, \mathbf{y}, \boldsymbol{\theta}_{(m)}, m)\right\}$, where D_i is an appropriate statistic. In order to check the fit of the marginal distributions of x's and y's we have adopted χ^2 discrepancy quantities, as proposed by Gelman *et al.* (1995, p.172):

$$D_1(\mathbf{x}, \mathbf{y}, \boldsymbol{\theta}_{(m)}, m) = \sum_{i=1}^{n} \frac{(x_i - v/\lambda_1)^2}{v/\lambda_1^2} \quad \text{and} \quad D_2(\mathbf{x}, \mathbf{y}, \boldsymbol{\theta}_{(m)}, m) = \sum_{i=1}^{n} \frac{(y_i - v/\lambda_2)^2}{v/\lambda_2^2}.$$

Let us now define $x_i' = x_i / \sum_{k=1}^n x_k$, $y_i' = y_i / \sum_{k=1}^n y_k$, $i = 1, \ldots, n$. If $x_i \sim \mathcal{G}(v, \lambda_1)$ and $y_i \sim \mathcal{G}(v, \lambda_2)$, then $(x_1', \ldots, x_n')$, $(y_1', \ldots, y_n')$ will follow the same (Dirichlet) distribution. Hence the differences of their values should be minor. Therefore the statistics

$$D_3(\mathbf{x}, \mathbf{y}, \boldsymbol{\theta}_{(m)}, m) = \sum_{i=1}^{n} |x_i' - y_i'| \quad \text{and} \quad D_4(\mathbf{x}, \mathbf{y}, \boldsymbol{\theta}_{(m)}, m) = \max_{i,j} \left|x_i' - y_j'\right|$$

may indicate cases where x_i' and y_i' do not come from the same distribution which means that the initial assumption is violated. Statistic D_3 can also be used to identify whether the correlation implied by the fitted model is plausible since it compares the mean absolute paired distances. Statistic D_4 gives more weight to the tails of the distributions since maximum distances are compared rather than the whole sum of the distances. If our model does not hold, then the corresponding p-values should either low or high.

Additionally, one can calculate specific quantiles of the predictive distribution and their corresponding p-values. These will check how close is the predictive distribution to the observed data. Finally, probabilities of more extreme observations (see Gelfand *et al.*, 1992) may be used to trace possible outliers or surprising observations under the selected model.

4 Illustrative examples

4.1 Simulated Datasets

Six simulated datasets with $v = 1$ have been used to evaluate the efficiency of our proposed RJMCMC algorithms. Results based on runs of 60000 iterations, after discarding the first 10000 iterations as burn–in period, have been obtained for each dataset. For all MCMC runs we have used the prior distributions

	λ_1		λ_2		$\phi = \lambda_1/\lambda_2$		ρ	
1	2	(2.12±0.21)	2.0	(1.96±0.19)	1.00	(1.09±0.15)	0.1	(0.11±0.07)
2	2	(1.96±0.20)	1.8	(1.67±0.17)	1.11	(1.18±0.16)	0.1	(0.11±0.07)
3	1	(0.91±0.09)	1.5	(1.39±0.14)	0.67	(0.66±0.08)	0.1	(0.23±0.10)
4	2	(1.81±0.18)	2.0	(2.12±0.21)	1.00	(0.86±0.07)	0.6	(0.65±0.06)
5	2	(2.10±0.21)	1.8	(1.66±0.17)	1.11	(1.27±0.12)	0.6	(0.54±0.07)
6	1	(0.99±0.10)	1.5	(1.53±0.15)	0.67	(0.65±0.06)	0.6	(0.62±0.06)

Table 2: Actual parameter values (Posterior mean ± standard deviation) for simulated datasets; for all datasets $v = 1$.

in (15)–(17) with $\omega = 1$, $\xi = 1$, $c^* = d^* = 0$ and uniform prior for ρ. We have also generated posterior summaries using the same prior setup but with ξ small. However, the posterior distributions of λ_1, λ_2 and ρ did not change. Proposal distributions were specified using pilot runs of model m_4 of length of 2000 iterations and 1000 burn–in. Concerning the "jump" between models m_2 and m_4, the simpler approach of Step 2 was used and worked efficiently in all simulated datasets. Additionally, prior model probabilities were tuned to achieve posterior model probabilities in the interval $[0.2, 0.3]$ in order to be able to estimate log–Bayes factors with sufficient precision. Using this setup the percentage of γ_1 changes (or jumps) was higher than 68% for all simulated datasets while for γ_2 was higher than 70% for datasets 1–5 and about 40% for dataset 6. Table 2 summarizes the results for each dataset.

Results of the logarithm of the Bayes factors and the posterior probabilities for all models are given in Table 3. It is clear that correct models are highly supported using the default prior proposed in Subsection 3.2. In more details, model 1 is highly supported in datasets 1 and 2 with posterior probabilities 0.738 and 0.746, respectively. Both of these datasets were generated using $\rho = 0.10$ and ratio $\phi = \lambda_1/\lambda_2$ equal to 1 and 1.11 (posterior means 1.09 and 1.18), respectively. The hypothesis of $\rho = 0$ is supported with posterior probabilities of 0.79 and 0.83, respectively, while the hypothesis of $\lambda_1 = \lambda_2$ is supported with posterior probabilities of 0.94 and 0.90, respectively.

For the third dataset with $\rho = 0.1$, $\phi = 0.67$ (posterior means: 0.23 and 0.66), model 4 is highly supported with posterior probability of 0.65. The hypotheses of $\rho \neq 0$ and $\lambda_1 \neq \lambda_2$ are supported with posterior probabilities of 0.69 and 0.92, respectively. The support of $\rho \neq 0$ here sources from the relatively high sample correlation. Note the property of Bayesian paradigm to attribute uncertainty on both hypotheses (0.31 for $\rho = 0$ vs. 0.69 for $\rho \neq 0$).

Regarding datasets 4–6, the hypothesis of $\rho \neq 0$ is strongly supported in all datasets since $\rho = 0.6$ (posterior means $r \geq 0.54$). In more details, model 2 is supported for datasets 4 and 5, while model 4 is supported for dataset 6. For dataset 4, the hypothesis of $\lambda_1 = \lambda_2$ is highly supported (posterior probability 0.84) as expected since $\phi = 1$ (posterior mean of $\phi = 0.86$). In dataset 5 the hypothesis of $\lambda_1 = \lambda_2$ is marginally supported with posterior probability 0.57. This result seems to be plausible since $\phi = 1.11$ and its posterior mean was found to be equal to 1.27. Finally, in dataset 6 the hypothesis of $\lambda_1 \neq \lambda_2$ is strongly supported (posterior probability 0.999) as expected, since $\phi = 0.67$ and the corresponding posterior mean is equal to 0.62.

In all the above simulated examples, we have also calculated the model diagnostics of Subsection 3.4. As expected, the results do not indicate any violation of the assumptions. Tables and Figures can be found at the web address: `www.stat-athens.aueb.gr/~jbn/papers/paper12.htm`.

i		m_i	Simulated Examples 1	2	3	4	5	6
1	$\log B_{41}$	$\lambda_1 = \lambda_2,\ \rho = 0$	-4.09	-3.75	2.67	19.35	13.19	22.10
2	$\log B_{42}$	$\lambda_1 = \lambda_2,\ \rho \neq 0$	-2.78	-2.19	2.82	-1.64	-0.29	7.93
3	$\log B_{43}$	$\lambda_1 \neq \lambda_2,\ \rho = 0$	-1.40	-1.51	0.88	21.59	14.68	20.31
1	$f(m_1\|\text{data})$	$\lambda_1 = \lambda_2,\ \rho = 0$	0.738	0.746	0.045	0.000	0.000	0.000
2	$f(m_2\|\text{data})$	$\lambda_1 = \lambda_2,\ \rho \neq 0$	0.199	0.157	0.039	0.838	0.573	0.001
3	$f(m_3\|\text{data})$	$\lambda_1 \neq \lambda_2,\ \rho = 0$	0.050	0.079	0.268	0.000	0.000	0.000
4	$f(m_4\|\text{data})$	$\lambda_1 \neq \lambda_2,\ \rho \neq 0$	0.012	0.018	0.648	0.162	0.427	0.999

Table 3: Estimates of posterior model probabilities and log-Bayes Factors of model m_4 versus m_i, $i = 1, 2, 3$.

$f(\rho)$	ξ	λ_1	λ_2	ρ
$\mathcal{B}eta(1,1)$	1	0.01926 (0.00083)	0.01578 (0.00068)	0.910 (0.012)
	10^{-9}	0.01925 (0.00083)	0.01577 (0.00068)	0.899 (0.016)
$\mathcal{B}eta(1/2,1/2)$	1	0.01924 (0.00083)	0.01577 (0.00068)	0.911 (0.012)
	10^{-9}	0.01923 (0.00083)	0.01576 (0.00068)	0.900 (0.016)

Table 4: Posterior means (standard deviations) for the Missouri river flow data (10000 iterations; 1000 burn-in).

4.2 The Missouri river flow data (continued)

Consider the Missouri river flow data used in Subsection 2.3. In this dataset, both hypotheses ($\rho = 0$ and $\lambda_1 = \lambda_2$) are useful since we wish to quantify evidence against or in favor of the dependence between the two measurements and/or equality of the river peak flow rates. Following Subsection 2.3, v is set equal to 5.4. Initially, we have used two sets of prior distributions to assess the sensitivity of the posterior distribution on different prior values of ξ. Both prior setups are given by equations (15)–(17) with $\omega = 1$, $c^* = d^* = 0$ but for the first one we consider $\xi = 1$ while for the second one $\xi = 10^{-9}$. In both prior distributions the uniform prior for ρ was used. The posterior distributions of all parameters were found to be close for both choices; see Table 4 for posterior summaries from MCMC runs of length 10000 with additional 1000 iterations as burn–in. Since the posterior distribution is robust when ξ tends to zero we have decided to use $\xi = 1$ for the model comparison. Furthermore, we have also considered the above prior setups using a $\mathcal{B}eta(1/2, 1/2)$, instead of uniform, prior distribution for ρ; differences are again minor (see Table 4).

Concerning the "jump" between models m_2 and m_4, we have used the more general algorithm that additionally proposes new values for ρ. This was decided because the sample correlations were found 0.80 and 0.89 under models m_2 and m_4 respectively. We have tried two approaches: The first one generates ρ from beta proposals with means 0.80 and 0.89 respectively, while the second simply sets $\rho_1 = \rho_0 + 0.09$. A total of 60000 iterations were considered for all RJMCMC runs discarding the first 10000 as burn-in period. The chains were tuned to visit all models with probabilities in $[0.2, 0.3]$ and hence being able to estimate efficiently all Bayes factors. Both approaches were efficient with percentage of accepted jumps approximately equal to 50% and give equivalent results.

In Table 5 we report results using for $\xi = 1$ and uniform or $\mathcal{B}eta(1/2, 1/2)$ prior for ρ. Results using both priors are identical, hence we may assume that the effect of using either of the two prior setups is minimal for this model comparison. From the Bayes factors' values we clearly see that model m_4 with

$f(\rho)$	$\log B_{41}$	$\log B_{42}$	$\log B_{43}$
$\mathcal{B}eta(1,1)$	93.9	36.4	91.3
$\mathcal{B}eta(1/2,1/2)$	94.0	36.6	91.4

Table 5: Logarithm of Bayes factors for the Missouri river flow data (50000 iterations and additional 10000 iterations as burn-in).

$\lambda_1 \neq \lambda_2$ and $\rho \neq 0$ is strongly supported. Following the interpretation of Kass and Raftery (1995), the evidence in favor of both hypotheses is very strong.

Concerning the fit indices of Subsection 3.4, all four p-values do not indicate any violation of the KBGD model assumption. Looking at the quantiles, only the p-value for the first quartile of y (out of the fourteen calculated) was found to be low indicating a possible misfit of this quantity on the marginal distribution of y. Finally, only three datapoints were found to have low probability of more extreme observation. Since the number of observations flagged as possible outliers is low there is no strong evidence against our fitted KBGD model. Note that the number of potential outliers is much higher for the two independent gamma models (models m_1 and m_3). More details on the above quantities can be fount at `www.stat-athens.aueb.gr/~jbn/papers/paper12.htm` .

From a practical point of view, an advantage of our approach is that quantities of interest can be calculated from the MCMC output. The above results show that the mean annual instantaneous peak flow measurements at Booneville, MO, and Hermann, MO, cannot be considered as equal and the corresponding measurements are highly correlated. Also note that the posterior values are robust on different choices of prior parameters (see Table 4) which indicate that the data provide strong information about the model parameters. Moreover, for water resources management and flood dynamics, it is of interest to report that, according to our selected model, the probability that the peak flow in Hermann will be larger than that in Booneville is estimated as 0.85.

5 Discussion

The present paper described Bayesian estimation and hypothesis testing for KBGD and its special case, Downton's bivariate exponential distribution. The data augmentation offered by the mixture representation of the density is used to construct the MCMC scheme. The major advantages of the proposed Gibbs sampler and augmentation scheme are: Firstly, in all steps we have simple known distributions which makes the sampler easily implemented. Secondly, the sampling scheme we propose is automatic in the sense that there is no need for searching a suitable proposal distribution or calibrating the variance of a simple random walk Metropolis step. Finally, using this augmentation scheme we avoid evaluating the Bessel function n times in each Metropolis comparison which may cause numerical problems (unexpected overflows) and considerably slow down the algorithm. This problem will be more intense in the case of numerical integration. So, our approach, although considerably increases the parameter space, ensures that there will not be any numerical problems and makes the computation automatic.

Extensions of the model to cover the case when covariates are available, as well as when some observations are censored can be also developed along the lines described in the paper. Covariates may be added but in this case Metropolis–Hastings steps are needed in order to generate from the conditional distributions. Censoring can be treated as usual within the MCMC setting by generating "missing" data at each iteration. Note also, that we have used the mixture representation of the density in order to

perform efficient data augmentation.

The proposed methodology can be generalized to more than two dimensions rather easily. For example, we may define a q–dimensional multivariate gamma distribution with density

$$f(\boldsymbol{x}|v, \boldsymbol{\lambda}, \rho) = \sum_{\kappa=0}^{\infty} g(\kappa|\rho) \prod_{i=1}^{q} f(x_i|v+\kappa, \lambda_i/(1-\rho)).$$

The case of $q = 2$ and $g(\kappa|\rho) \sim \mathcal{NB}(v, 1-\rho)$ is examined in the present paper. For $q = 1$ we obtain a density similar to that of the noncentral gamma. The extension of our proposed methodology to cover such multivariate models is straightforward. Moreover, we may change or relax the parametric assumptions on $g(\kappa|\rho)$ to construct a more general and flexible multivariate gamma distribution.

Finally, we have constructed a trans–dimensional MCMC algorithm (RJMCMC) for evaluating the posterior model odds and implementing Bayesian hypothesis testing and model selection between competing models. Starting from KBGD one may check whether the two variables are independent, corresponding to the case when $\rho = 0$, or whether the marginal means are equal corresponding to the case when $\lambda_1 = \lambda_2$. The latter is of special importance when the paired observations belong to before and after observations and we like to examine whether the treatment had an effect. An interesting topic for further research is the convergence properties of that trans–dimensional Markov chain, which is however a difficult task. One can argue that since all within–model moves are produced by Gibbs sampling and hence are always accepted, the whole chain is Harris recurrent (see Roberts and Rosenthal, 2004). On the other hand, although the simulated Markov chain within each model is geometrically ergodic (for models 1 and 3 we sample directly from the posterior distribution whereas for model 2 the geometric ergodicity can be proved with similar arguments as in Subsection 2.1), it is not clear whether such a property is retained in the whole chain due to the complicated expressions of between–model jump probabilities.

APPENDIX

Proof of Lemma 1. Observe first that $f(x; a_1, b) < f(x; a_2, b)$ if and only if $x > x^*$. Now, the partial derivative of $f(x; a, b)$ with respect to a has the sign of $\log(bx) - \psi(a)$, where $\psi(a) = \Gamma'(a)/\Gamma(a)$ is the digamma function. Since ψ is strictly increasing, $f(x; a, b)$ is increasing if $\psi(a) < \log(bx)$ and decreasing if $\psi(a) > \log(bx)$. Hence, for $x \leq b^{-1}\exp[\psi(a_1)]$, the infimum of $f(x; a, b)$ equals $f(x; a_2, b)$, whereas for $x \geq b^{-1}\exp[\psi(a_2)]$, the infimum of $f(x; a, b)$ equals $f(x; a_1, b)$. For $b^{-1}\exp[\psi(a_1)] \leq x \leq b^{-1}\exp[\psi(a_2)]$ the infimum is the smallest among $f(x; a_1, b)$, $f(x; a_2, b)$ and this depends on whether x is smaller or greater than x^*. This completes the proof.

Proof of Lemma 2. Similarly to the proof of Lemma 1, $f(x; a_1, b) < f(x; a_2, b)$ if and only if $x > x^*$. The partial derivative of $f(x; a, b)$ with respect to a has the sign of $\psi(a+b) - \psi(a) + \log x$. After observing that $\psi(a+b) - \psi(a)$ is strictly decreasing in a for any $b > 0$, the proof proceeds as before.

REFERENCES

Achcar, J.A. and Leandro, R.A. (1998). Use of Markov chain Monte Carlo in a Bayesian analysis of the Block and Basu bivariate exponential distribution. *Annals of the Institute of Statistical Mathematics*, **50**, 403-416.

Amos, D.E. (1974). Computation of modified Bessel functions and their ratios. *Mathematics of Computation*, **28**, 239–251.

Bartlett, M.S. (1957). Comment on D.V. Lindley's statistical paradox. *Biometrika*, **44**, 533–534.

Chen, M.H., Ibrahim, J.G. and Shao, Q.M. (2000). Power prior distributions for generalized linear models. *Journal of Statistical Planning and Inference*, **84**, 121–137.

Damien, P. and Muller, P. (1998). A Bayesian bivariate failure time regression model. *Computational Statistics and Data Analysis*, **28**, 77-85.

Dellaportas P. and Forster J J (1999). Markov chain Monte Carlo model determination for hierarchical and graphical log-linear models. *Biometrika*, **86**, 615–633.

Dellaportas, P., Forster, J.J. and Ntzoufras, I. (2002). On Bayesian model and variable selection using MCMC. *Statistics and Computing*, **12**, 27–36.

Devroye, L. (2002). Simulating Bessel random variables. *Statistics and Probability Letters*, **57**, 249–257.

Downton, F. (1970). Bivariate exponential distributions in reliability theory. *Journal of the Royal Statistical Society* B, **32**, 408–417.

Gelfand, A.E., Dey, D.K. and Chang, H. (1992). Model determination using predictive distributions with implementation via sampling-based methods. *Bayesian Statistics*, **4**, (Bernardo, J.M., Berger, J.O., Dawid, A.P. and Smith, A.F.M., eds.), pp. 147 – 168. Oxford University Press.

Gelman, A., Carlin, J.B., Stern, H.S. and Rubin, D.B. (1995). *Bayesian Data Analysis*. First edition. Chapman and Hall.

Green, P. (1995). Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika*, **82**, 711–732

Iliopoulos, G. and Karlis, D. (2003). Simulation from the Bessel distribution with applications. *Journal of Statistical Computation and Simulation*, **73**, 491–506.

Jones, G.L. and Hobert, J.R. (2001). Honest exploration of intractable probability distributions via Markov chain Monte Carlo. *Statistical Science*, **16**, 312–334.

Hobert, J.P., Jones, G.L., Presnell, B. and Rosenthal, J.S. (2002). On the applicability of regenerative simulation in Markov chain Monte Carlo. *Biometrika*, **89**, 731–743.

Hougaard, P. (1984). Life tables methods for heterogeneous populations: distributions describing the heterogeneity. *Biometrika*, **71**, 75–83.

Kass, R.E. and Raftery, A.E. (1995). Bayes Factors. *Journal of the American Statistical Association*, **90**, 773–795.

Kershaw, D. (1983). Some extensions of W. Gautschi's inequalities for the gamma function. *Mathematics of Computation*, **41**, 607–611.

Kibble, W.F. (1941). A two-variate gamma type distribution. *Sankhyã*, **5**, 137–150.

Kotz, S., Balakrishnan, N. and Johnson, N.L. (2000). *Continuous Multivariate Distributions*, **1**, Second edition. New York, Wiley.

Lagakos, S.W. (1978). A covariate model for partially censored data subject to competing causes of death. *Applied Statistics*, **27** , 235–241

Lindley, D.V. (1957). A statistical paradox. *Biometrika*, **44**, 187–192.

Liu, J.S., Wong, W.H. and Kong, A. (1994). Covariance structure of the Gibbs sampler with applications to the comparisons of estimators and augmentation schemes. *Biometrika*, **81**, 27–40.

Mykland, P., Tierney, L. and Yu, B. (1995). Regeneration in Markov chain samplers. *Journal of the American Statistical Association*, **90**, 233–241.

Nagaraja, H.N., Choudhary, P.K. and Matalas, N. (2002). Number of records in a bivariate sample with appliacation to Missouri river flood data. *Methodology and Computing in Applied Probability*, **4**, 377–391.

Roberts, G.O. and Rosenthal, J.S. (2004). Harris recurrence of Metropolis–within–Gibbs and trans–dimensional Markov chains. Preprint, `http://probability.ca/jeff/ftpdir/harris.ps.Z`

Rosenthal, J.S. (1995). Minorization conditions and convergence rates for Markov chain Monte Carlo. *Journal of the American Statistical Association*, **90**, 558–566.

Wang, C.P. and Ghosh, M. (2000). Bayesian analysis of bivariate competing risks models. *Sankhyã*, **B**, **62**, 388–401

Yuan, L. and Kalbfleisch, J.D. (2000). On the Bessel distribution and related problems. *Annals of the Institute of Statistical Mathematics*, **52**, 438–447.

Yue, S., Quarda, T.B.M.J and Bobee, B. (2001). A review of bivariate gamma distributions for hydrological application. *Journal of Hydrology*, **246**, 1-18.