Revista Colombiana de Estadística
Diciembre 2013, volumen 36, no. 2, pp. 319 a 336

Bayesian Inference for Two-Parameter Gamma Distribution Assuming Different Noninformative Priors

Inferencia Bayesiana para la distribución Gamma de dos parámetros asumiendo diferentes a prioris no informativas

Fernando Antonio Moala[1,a], Pedro Luiz Ramos[1,b], Jorge Alberto Achcar[2,c]

[1]Departamento de Estadística, Facultad de Ciencia y Tecnología, Universidade Estadual Paulista, Presidente Prudente, Brasil

[2]Departamento de Medicina Social, Facultad de Medicina de Ribeirão Preto, Universidade de São Paulo, Ribeirão Preto, Brasil

Abstract

In this paper distinct prior distributions are derived in a Bayesian inference of the two-parameters Gamma distribution. Noniformative priors, such as Jeffreys, reference, MDIP, Tibshirani and an innovative prior based on the copula approach are investigated. We show that the maximal data information prior provides in an improper posterior density and that the different choices of the parameter of interest lead to different reference priors in this case. Based on the simulated data sets, the Bayesian estimates and credible intervals for the unknown parameters are computed and the performance of the prior distributions are evaluated. The Bayesian analysis is conducted using the Markov Chain Monte Carlo (MCMC) methods to generate samples from the posterior distributions under the above priors.

***Key words*:** Gamma distribution, noninformative prior, copula, conjugate, Jeffreys prior, reference, MDIP, orthogonal, MCMC.

Resumen

En este artículo diferentes distribuciones a priori son derivadas en una inferencia Bayesiana de la distribución Gamma de dos parámetros. A prioris no informativas tales como las de Jeffrey, de referencia, MDIP, Tibshirani y una priori innovativa basada en la alternativa por cópulas son investigadas. Se muestra que una a priori de información de datos maximales conlleva a una a

[a]Professor. E-mail: femoala@fct.unesp.br
[b]Student. E-mail: pedrolramos@hotmail.com
[c]Professor. E-mail: achcar@fmrp.usp.br

posteriori impropia y que las diferentes escogencias del parámetro de interés permiten diferentes a prioris de referencia en este caso. Datos simulados permiten calcular las estimaciones Bayesianas e intervalos de credibilidad para los parámetros desconocidos así como la evaluación del desempeño de las distribuciones a priori evaluadas. El análisis Bayesiano se desarrolla usando métodos MCMC (Markov Chain Monte Carlo) para generar las muestras de la distribución a posteriori bajo las a priori consideradas.

***Palabras clave*:** a prioris de Jeffrey, a prioris no informativas, conjugada, cópulas, distribución Gamma, MCMC, MDIP, ortogonal, referencia.

1. Introduction

The Gamma distribution is widely used in reliability analysis and life testing (see for example, Lawless 1982) and it is a good alternative to the popular Weibull distribution. It is a flexible distribution that commonly offers a good fit to any variable such as in environmental, meteorology, climatology, and other physical situations.

Let X be representing the lifetime of a component with a Gamma distribution, denoted by $\Gamma(\alpha, \beta)$ and given by

$$f(x \mid \alpha, \beta) = \frac{\beta^{\alpha}}{\Gamma(\alpha)} x^{\alpha-1} \exp\{-\beta x\}, \text{ for all } x > 0 \tag{1}$$

where $\alpha > 0$ and $\beta > 0$ are unknown shape and scale parameters, respectively.

There are many papers considering Bayesian inference for the estimation of the Gamma parameters. Son & Oh (2006) assume vague priors for the parameters to the estimation of parameters using Gibbs sampling. Apolloni & Bassis (2099) compute the joint probability distribution of the parameters without assuming any prior. They propose a numerical algorithm based on an approximate analytical expression of the probability distribution. Pradhan & Kundu (2011) assume that the scale parameter has a Gamma prior and the shape parameter has any log-concave prior and they are independently distributed. However, most of these papers have in common the use of proper priors and the assumption of independence a priori of the parameters. Although this is not a problem and have been much used in the literature we, would like to propose a noninformative prior for the Gamma parameters which incorporates the dependence structure of parameters. Some of priors proposed in the literature are Jeffrey$s$ (1967), MDIP (Zellner 1977, Zellner 1984, Zellner 1990, Tibshirani 1989), and reference prior (Bernardo 1979). Moala (2010) provides a comparison of these priors to estimate the Weibull parameters.

Therefore, the main aim of this paper is to present different noninformative priors for a Bayesian estimation of the two-parameter Gamma distribution. We also propose a bivariate prior distribution derived from copula functions (see for example, Nelsen 1999, Trivedi & Zimmer 2005a, Trivedi & Zimmer 2005b) in order to construct a prior distribution to capture the dependence structure between the parameters α and β.

We investigate the performance of the prior distributions through a simulation study using a small data set. Accurate inference for the parameters of the Gamma is obtained using MCMC (Markov Chain Monte Carlo) methods.

2. Maximum Likelihood Estimation

Let $X_1, \ldots, X_n$ be a complete sample from (1) then the likelihood function is

$$L(\alpha, \beta \mid \boldsymbol{x}) = \frac{\beta^{n\alpha}}{[\Gamma(\alpha)]^n}\Big(\prod_{i=1}^{n} x_i^{\alpha-1}\Big)\exp\Big\{-\beta\sum_{i=1}^{n} x_i\Big\} \tag{2}$$

for $\alpha > 0$ and $\beta > 0$.

Considering $\frac{\partial}{\partial\alpha}\log L$ and $\frac{\partial}{\partial\beta}\log L$ equal to 0 and after some algebric manipulations we get the likelihood equations given by

$$\widehat{\beta} = \frac{\widehat{\alpha}}{\overline{X}} \quad \text{and} \quad \log\widehat{\alpha} - \psi(\widehat{\alpha}) = log\left(\frac{\overline{X}}{\tilde{X}}\right) \tag{3}$$

where $\psi(k) = \frac{\partial}{\partial k} log\,\Gamma(k) = \frac{\Gamma'(k)}{\Gamma(k)}$ (see Lawless 1982) is the diGamma function, $\overline{X} = \frac{\sum_{i=1}^{n} x_i}{n}$ and $X = \left(\prod_{i=1}^{n} x_i\right)^{1/n}$. The solutions for these equations provide the maximum likelihood estimators $\widehat{\alpha}$ and $\widehat{\beta}$ for the parameters of the Gamma distribution (1). As closed form solution is not possible to evaluate (3), numerical techniques must used. The Fisher information matrix is given by

$$I(\alpha, \beta) = \begin{bmatrix} \psi'(\alpha) & -\frac{1}{\beta} \\ -\frac{1}{\beta} & \frac{\alpha}{\beta^2} \end{bmatrix} \tag{4}$$

where $\psi'(\alpha)$ is the derivative of $\psi(\alpha)$ called as triGamma function.

For large samples, approximated confidence intervals can be constructed for the parameters α and β through normal marginal distributions given by

$$\widehat{\alpha} \sim N(\alpha, \sigma_1^2) \text{ and } \widehat{\beta} \sim N(0, \sigma_2^2), \text{ for } n \to \infty \tag{5}$$

where $\sigma_1^2 = v\widehat{a}r(\widehat{\alpha}) = \frac{\widehat{\alpha}}{\widehat{\alpha}\psi'(\widehat{\alpha})-1}$ and $\sigma_2^2 = v\widehat{a}r(\widehat{\beta}) = \frac{\widehat{\beta}^2\psi'(\widehat{\alpha})}{\widehat{\alpha}\psi'(\widehat{\alpha})-1}$. In this case, the approximated $100(1-\Gamma)\%$ confidence intervals for each parameter α and β are given by

$$\widehat{\alpha} - z_{\frac{\Gamma}{2}}\sigma_1 < \alpha < \widehat{\alpha} + z_{\frac{\Gamma}{2}}\sigma_1 \text{ and } \widehat{\beta} - z_{\frac{\Gamma}{2}}\sigma_2 < \beta < \widehat{\beta} + z_{\frac{\Gamma}{2}}\sigma_2 \tag{6}$$

respectively.

3. Jeffrey's Prior

A well-known weak prior to represent a situation with little information about the parameters was proposed by Jeffreys (1967). This prior denoted by $\pi_J(\alpha,\, \beta)$ is derived from the Fisher information matrix $I(\alpha,\, \lambda)$ given in (4) as

$$\pi_J(\alpha, \beta) \propto \sqrt{det\, I(\alpha,\, \beta)} \tag{7}$$

Jeffrey's prior is widely used due to its invariance property under one-to-one transformations of parameters although there has been an ongoing discussion about whether the multivariate form prior is appropriate.

Thus, from (4) and (7) the Jeffreys prior for $(\alpha,\, \beta)$ parameters is given by:

$$\pi_J(\alpha, \beta) \propto \frac{\sqrt{\alpha\psi\,'(\alpha) - 1}}{\beta} \tag{8}$$

4. Maximal Data Information Prior (MDIP)

It is of interest that the data gives more information about the parameter than the information on the prior density; otherwise, there would not be justification for the realization of the experiment. Thus, we wish a prior distribution $\pi(\phi)$ that provides a gain in the information supplied by data in which the largest possible relative to the prior information of the parameter, that is, which maximize the information on the data. With this idea Zellner (1977), Zellner (1984), Zellner (1990) and Min & Zellner (1993) derived a prior which maximize the average information in the data density relative to that one in the prior. Let

$$H(\phi) = \int_{R_x} f(x \mid \phi) ln f(x \mid \phi) dx, x \in R_x \tag{9}$$

be the negative entropy of $f(x \mid \phi)$, the measure of the information in $f(x \mid \phi)$ and R_x the range of density $f(x \mid \phi)$. Thus, the following functional criterion is employed in the MDIP approach:

$$G[\pi(\phi)] = \int_{\mathrm{a}}^{\mathrm{b}} H(\phi)\pi(\phi)d\phi - \int_{\mathrm{a}}^{\mathrm{b}} \pi(\phi) \ln\, \pi(\phi) d\phi \tag{10}$$

which is the prior average information in the data density minus the information in the prior density. $G[\pi(\phi)]$ is maximized by selection of $\pi(\phi)$ subject to $\int_{\mathrm{a}}^{\mathrm{b}} \pi(\phi)d\phi = 1$. The solution is then a proper prior given by

$$\pi(\phi) = k \exp\Big\{H(\phi)\Big\} \quad a \leq \phi \leq b \tag{11}$$

where $k^{-1} = \int_{\mathrm{a}}^{\mathrm{b}} \exp\Big\{H(\phi)\Big\}d\phi$ is the normalizing constant.

Therefore, the MDIP is a prior that leads to an emphasis on the information in the data density or likelihood function. That is, its information is weak in comparison with data information.

Zellner (1977), Zellner (1984), Zellner (1990) shows several interesting properties of MDIP and additional conditions that can also be imposed to the approach reflection given initial information. However, the MDIP has restrictive invariance properties.

Theorem 1. *Suppose that we do not have much prior information available about α and β. Under this condition, the prior distribution MDIP, denoted by $\pi_Z(\alpha,\ \beta)$, for the parameters $(\alpha,\ \beta)$ of the Gamma density (1) is given by:*

$$\pi_Z(\alpha,\beta) \propto \frac{\beta}{\Gamma(\alpha)} \exp\Big\{(\alpha-1)\psi(\alpha)-\alpha\Big\} \tag{12}$$

Proof. Firstly, we have to evaluate the measure information $H(\alpha,\ \beta)$ for the Gamma density which is given by

$$H(\alpha,\beta) = \int_0^\infty \ln\Big(\frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1}\exp\{-\beta x\}\Big) f(x \mid \alpha,\beta)\, dx \tag{13}$$

and after some algebra, the result is

$$H(\alpha,\beta) = \alpha \ln\, \beta - \ln\, \Gamma(\alpha) + (\alpha-1)\int_0^\infty \ln(x) f(x \mid \alpha,\beta)\, dx - \beta E(X) \tag{14}$$

with $E(X) = \frac{\alpha}{\beta}$. □

Since the integral functions $\int_0^\infty u^{\alpha-1}e^{-u}\, du = \Gamma(\alpha)$ and $\int_0^\infty u^{\alpha-1}\log(u)e^{-u}\, du = \Gamma'(\alpha)$, the function (14) envolving these integrals can be expressed as

$$H(\alpha,\beta) = -\ln\, \Gamma(\alpha) + \ln\, \beta + (\alpha-1)\psi(\alpha) - \alpha \tag{15}$$

Therefore, the MDIP prior for the parameters α and β is given by

$$\pi_Z(\alpha,\beta) \propto \frac{\beta}{\Gamma(\alpha)} \exp\Big\{(\alpha-1)\psi(\alpha)-\alpha\Big\} \tag{16}$$

However, the corresponding joint posterior density is not proper, but surprisingly, the prior density given by

$$\pi_Z(\alpha,\beta) \propto \frac{\beta}{\Gamma(\alpha)} \exp\Big\{(\alpha-1)\frac{\psi(\alpha)}{\Gamma(\alpha)}-\alpha\Big\} \tag{17}$$

yields a proper posteriory density. Thus, we will use (17) as MDIP prior in the numerical illustration in Section 8.

5. Reference Prior

Another well-known class of noninformative priors is the reference prior, first described by Bernardo (1979) and further developed by Berger & Bernardo (1992).

The idea is to derive a prior $\pi(\phi)$ that maximizes the expected posterior information about the parameters provided by independent replications of an experiment relative to the information in the prior. A natural measure of the expected information about ϕ provided by data $\boldsymbol{x}$ is given by

$$I(\phi) = E_{\boldsymbol{x}}[K(p(\phi \mid \boldsymbol{x}), \pi(\phi))] \tag{18}$$

where

$$K(p(\phi \mid \boldsymbol{x}), \pi(\phi)) = \int_{\Phi} p(\phi \mid \boldsymbol{x}) \log \frac{p(\phi \mid \boldsymbol{x})}{\pi(\phi)} d\phi \tag{19}$$

is the Kullback-Leibler distance. So, the reference prior is defined as the prior $\pi(\phi)$ that maximizes the expected Kullback-Leibler distance between the posterior distribution $p(\phi \mid \boldsymbol{x})$ and the prior distribution $\pi(\phi)$, taken over the experimental data.

The prior density $\pi(\phi)$ which maximizes the functional (19) is found through calculus of variation and, the solution is not explicit. However, when the posterior $p(\phi \mid \boldsymbol{x})$ is asymptotically normal, this approach leads to Jeffreys prior for a single parameter situation. If on the other hand, we are interested in one of the parameters, being the remaining parameters nuisances, the situation is quite different, and the appropriated reference prior is not a multivariate Jeffrey prior. Bernardo (1979) argues that when nuisance parameters are present the reference prior should depend on which parameter(s) are considered to be of primary interest. The reference prior in this case is derived as follows. We will present here the two-parameters case in details. For the multiparameter case, see Berger & Bernardo (1992).

Let $\boldsymbol{\theta} = (\theta_1, \theta_2)$ be the whole parameter, θ_1 being the parameter of interest and θ_2 the nuisance parameter. The algorithm is as follows:

Step 1: Determine $\pi_2(\theta_2 \mid \theta_1)$, the conditional reference prior for θ_2 assuming that θ_1 is known, is given by,

$$\pi_2(\theta_2 \mid \theta_1) = \sqrt{\mathrm{I}_{22}(\theta_1, \theta_2)} \tag{20}$$

where $\mathrm{I}_{22}(\theta_1, \theta_2)$ is the (2,2)-entry of the Fisher Information Matrix.

Step 2: Normalize $\pi_2(\theta_2 \mid \theta_1)$.

If $\pi_2(\theta_2 \mid \theta_1)$ is improper, choose a sequence of subsets $\Omega_1 \subseteq \Omega_2 \subseteq \ldots \to \Omega$ on which $\pi_2(\theta_2 \mid \theta_1)$ is proper. Define the normalizing constant and the proper prior $p_m(\theta_2 \mid \theta_1)$ respectively as

$$c_{\mathrm{m}}(\theta_1) = \frac{1}{\int_{\Omega_m} \pi_2(\theta_2 \mid \theta_1) d\theta_2} \tag{21}$$

and

$$p_m(\theta_2 \mid \theta_1) = c_m(\theta_1) \pi_2(\theta_2 \mid \theta_1) 1_{\Omega_m}(\theta_2), \tag{22}$$

Step 3: Find the marginal reference prior $\pi_m(\theta_1)$ for $\theta_1$the reference prior for the experiment found by marginalizing out with respect to $p_m(\theta_2 \mid \theta_1)$. We obtain

$$\pi_m(\theta_1) \propto exp\Big\{ \frac{1}{2} \int_{\Omega_m} p_m(\theta_2 \mid \theta_1) log \left\| \frac{det\ I(\theta_1, \theta_2)}{I_{22}(\theta_1, \theta_2)} \right\| d\theta_2 \Big\} \tag{23}$$

Step 4: Compute the reference prior $\pi_{\theta_1}(\theta_1,\ \theta_2)$ when θ_1 is the parameter of interest

$$\pi_{\theta_1}(\theta_1,\theta_2)=\lim_{m\to\infty}\left(\frac{c_m(\theta_1)\pi_m(\theta_1)}{c_m(\theta_1^*)\ \pi_m(\theta_1^*)}\right)\pi(\theta_2\mid\theta_1) \tag{24}$$

where θ_1^* is any fixed point with positive density for all π_m.

We will derive the reference prior for the parameters of the Gamma distribution given in (1), where α will be considered as the parameter of interest and β the nuisance parameter.

Theorem 2. *The reference prior for the parameters of the Gamma distribution given in (1), where α will be considered as the parameter of interest and β the nuisance parameter, is given by:*

$$\pi_\alpha(\alpha,\beta)=\frac{1}{\beta}\sqrt{\frac{\alpha\psi'(\alpha)-1}{\alpha}} \tag{25}$$

If β is the parameter of interest and α the nuisance, thus the prior is

$$\pi_\beta(\alpha,\beta)\propto\frac{\sqrt{\psi'(\alpha)}}{\beta} \tag{26}$$

Proof. By the approach proposed by Berger & Bernardo (1992), we find the reference prior for the nuisance parameter β, conditionally on the parameter of interest α, given by

$$\pi(\beta\mid\alpha)=\sqrt{\mathrm{I}_{\beta\beta}(\alpha,\ \beta)}\propto\frac{1}{\beta} \tag{27}$$

where $\mathrm{I}_{\beta\beta}(\alpha,\ \beta)$ is the (2,2)-entry of the Fisher Information Matrix given in (4). □

As in Moala (2010), a natural sequence of compact sets for $(\alpha,\ \beta)$ is $(l_{1n},l_{2n})\times(q_{1n},q_{2n})$, so that $l_{1i},\ q_{1i}\to 0$ and $l_{2i},\ q_{2i}\to\infty$ when $i\to\infty$. Therefore, the normalizing constant is given by,

$$c_i(\alpha)=\frac{1}{\int_{q_{1i}}^{q_{2i}}\frac{1}{\beta}d\beta}=\frac{1}{\log\ q_{2i}-\log\ q_{1i}}. \tag{28}$$

Now from (23), the marginal reference prior for α is given by

$$\pi_i(\alpha)=\exp\left\{\frac{1}{2}\int_{q_{1i}}^{q_{2i}}c_i(\alpha)\frac{1}{\beta}\log\left\|\frac{\frac{\alpha\psi'(\alpha)-1}{\beta^2}}{\frac{\alpha}{\beta^2}}\right\|d\beta\right\} \tag{29}$$

which after some mathematical arrangement, we have

$$\pi_i(\alpha)=\sqrt{\frac{\alpha\psi'(\alpha)-1}{\alpha}}\exp\left\{\frac{1}{2}c_i(\alpha)\int_{q_{1i}}^{q_{2i}}\frac{1}{\beta}d\beta\right\} \tag{30}$$

Therefore, the resulting marginal reference prior for α is given by

$$\pi_i(\alpha)\propto\sqrt{\frac{\alpha\psi'(\alpha)-1}{\alpha}} \tag{31}$$

and the global reference prior for $(\alpha,\, \beta)$ with parameter of interest α is given by,

$$\pi_\alpha(\alpha,\beta) = \lim_{i\to\infty}\left(\frac{c_i(\alpha)\pi_i(\alpha)}{c_i(\alpha^*)\ \pi_i(\alpha^*)}\right)\ \pi(\beta\,|\,\alpha) \propto \frac{1}{\beta}\sqrt{\frac{\alpha\psi'(\alpha)-1}{\alpha}} \tag{32}$$

considering $\alpha^* = 1$

Similarly we obtain the reference prior considering β as the parameter of interest and α as nuisance. In this case, the prior is

$$\pi_\beta(\alpha,\beta) \propto \frac{\sqrt{\psi'(\alpha)}}{\beta} \tag{33}$$

6. Tibishirani's Prior

Given a vector parameter $\boldsymbol{\phi}$, Tibshirani (1989) developed an alternative method to derive a noninformative prior $\pi(\delta)$ for the parameter of interest $\delta = t(\boldsymbol{\phi})$ so that the credible interval for δ has coverage error $O(n^{-1})$ in the frequentist sense. This means that the difference between the posterior and frequentist confidence interval should be small. To achieve that, Tibshirani (1989) proposed to reparametrize the model in terms of the orthogonal parameters $(\delta,\, \lambda)$ (see Cox & Reid 1987) where δ is the parameter of interest and λ is the orthogonal nuisance parameter. In this way, the approach specifies the weak prior to be any prior of the form

$$\pi(\delta,\lambda) = g(\lambda)\sqrt{I_{\delta\delta}(\delta,\, \lambda)} \tag{34}$$

where $g(\lambda)> 0$ is an arbitrary function and $I_{\delta\delta}(\delta,\, \lambda)$ is the "delta" entry of the Fisher Information Matrix.

Theorem 3. *The Tibishirani's prior distribution $\pi_T(\alpha,\, \beta)$ for the parameters $(\alpha,\, \beta)$ of the Gamma distribution given in (1) by considering α as the parameter of interest and β the nuisance parameter is given by:*

$$\pi_T(\alpha,\beta) \propto \frac{1}{\beta}\sqrt{\frac{\alpha\psi'(\alpha)-1}{\alpha}} \tag{35}$$

Proof . For the Gamma model (1), we will propose an orthogonal reparametrization $(\delta,\, \lambda)$ where $\delta = \alpha$ is the parameter of interest and λ is the nuisance parameter to be evaluated. The orthogonal parameter λ is obtained by solving the differential equation:

$$I_{\beta\beta}\frac{\partial\beta}{\partial\alpha} = -I_{\alpha\beta} \tag{36}$$

□

From (4) and (36) we have

$$\frac{\alpha}{\beta^2}\frac{\partial\beta}{\partial\alpha} = \frac{1}{\beta} \tag{37}$$

Separating the variables, (37) becomes the following,

$$\frac{1}{\beta}\partial\beta = \frac{1}{\alpha}\partial\alpha \tag{38}$$

Integrating both sides we get,

$$\log\beta = log\,\alpha + h(\lambda) \tag{39}$$

where $h(\lambda)$ is an arbitrary function of λ.

By chosing $h(\lambda) = \log\lambda$, we obtained the solution to (36), the nuisance parameter λ orthogonal to δ,

$$\lambda = \frac{\beta}{\alpha} \tag{40}$$

Thus, the information matrix for the orthogonal parameters is given by

$$I(\boldsymbol{\delta}, \boldsymbol{\lambda}) = \begin{bmatrix} \psi'(\delta) - \frac{1}{\delta} & 0 \\ 0 & \frac{\delta}{\lambda^2} \end{bmatrix} \tag{41}$$

From (34) and (41), the corresponding prior for $(\delta,\, \lambda)$ is given by

$$\pi_{\boldsymbol{\delta}}(\delta, \lambda) \propto g(\lambda)\sqrt{\frac{\delta\psi'(\delta) - 1}{\delta}} \tag{42}$$

where $g(\lambda)$ is an arbitrary function.

Due to a lack of uniqueness in the choice of the orthogonal parametrization, then the class of orthogonal parameters is of the form $g(\lambda)$, where $g(\cdot)$ is any reparametrization. This non-uniqueness is reflected by the function $g(\cdot)$ corresponding to (26). One possibility, in the single nuisance parameter case, is to require that $(\delta,\, \lambda)$ also satisfies Stein's condition (see Tibshirani 1989) for λ with p taken as the nuisance parameter. Under this condition we obtain

$$\pi_{\lambda}(\delta, \lambda) \propto g^*(\delta)\frac{\sqrt{\delta}}{\lambda} \tag{43}$$

Now, requiring $g(\lambda)\sqrt{\frac{\delta\psi'(\delta)-1}{\delta}} = g^*(\delta)\frac{\sqrt{\delta}}{\lambda}$ we have that

$$\pi_T(\delta, \lambda) \propto \frac{1}{\lambda}\sqrt{\frac{\delta\psi'(\delta) - 1}{\delta}} \tag{44}$$

Thus, from (40), the prior expressed in terms of the $(\alpha,\, \beta)$ parametrization is given by

$$\pi_T(\alpha, \beta) \propto \frac{1}{\beta}\sqrt{\frac{\alpha\psi'(\alpha) - 1}{\alpha}} \tag{45}$$

Note that this prior coincides with reference prior (25) considering α as the parameter of interest.

7. Copula Prior

In this section we derive a bivariate prior distribution from copula functions (see for example, Nelsen 1999, Trivedi & Zimmer 2005*a*, Trivedi & Zimmer 2005*b*) in order to construct a prior distribution to capture the dependence structure between the parameters α and β. Copulas can be used to correlate two or more random variables and they provide great flexibility to fit known marginal densities.

A special case is given by the Farlie-Gumbel-Morgenstern copula which is suitable to model weak dependences (see Morgenstern 1956) with corresponding bivariate prior distribution for α and β given ρ,

$$\pi(\alpha, \beta \mid \rho) = f_1(\alpha) f_2(\beta) + \rho f_1(\alpha) f_2(\beta)[1 - 2F_1(\alpha)][1 - 2F_2(\beta)], \tag{46}$$

where $f_1(\alpha)$ and $f_2(\beta)$ are the marginal densities for the random quantities α and β; $F_1(\alpha)$ and $F_2(\beta)$ are the corresponding marginal distribution functions for α and β, and $-1 \leq \rho \leq 1$.

Observe that if $\rho = 0$, we have independence between α and β.

Different choices could be considered as marginal distributions for α and β as Gamma, exponential, Weibull or uniform distributions.

In this paper, we will assume Gamma marginal distribution $\Gamma(\mathrm{a}_1, b_1)$ and $\Gamma(\mathrm{a}_2, b_2)$ for α and β, respectively, with known hyperparameters a_1, a_2, b_1 and b_2. Thus,

$$\begin{aligned}\pi(\alpha, \beta \mid a_1, a_2, b_1, b_2, \rho) \propto \alpha^{\mathrm{a}_1 - 1} \beta^{\mathrm{a}_2 - 1} \exp\Big\{-b_1\alpha - b_2\beta\Big\} \times \\ \Big[1 + \rho\Big(1 - 2I(a_1, b_1\alpha)\Big)\Big(1 - 2I(a_1, b_1\alpha)\Big)\Big]\end{aligned} \tag{47}$$

where $I(k, x) = \frac{1}{\Gamma(\mathrm{k})} \int_0^x u^{k-1} e^{-u} du$ is the incomplete Gamma function.

Assuming the prior (47), the joint posterior distribution for α, β and ρ is given by,

$$\begin{aligned}p(\alpha, \beta, \rho \mid x) \propto \frac{\beta^{n\alpha}}{[\Gamma(\alpha)]^n} \Big(\prod_{i=1}^{n} x_i^{\alpha-1}\Big) \\ \exp\Big\{-\beta \sum_{i=1}^{n} x_i\Big\} \pi(\alpha, \beta \mid a_1, a_2, b_1, b_2, \rho) \pi(\rho)\end{aligned} \tag{48}$$

where $\pi(\rho)$ is a prior distribution for ρ.

In general, many different priors can be used for ρ; one possibility is to consider an uniform prior distribution for ρ over the interval $[-1, 1]$.

8. Numerical Illustration

8.1. Simulation Study

In this section, we investigate the performance of the proposed prior distributions through a simulation study with samples of size $n = 5$, $n = 10$ and $n = 30$ generated from the Gamma distribution with parameters $\alpha = 2$ and $\beta = 3$.

As we do not have an analytic form for marginal posterior distributions we need to appeal to the MCMC algorithm to obtain the marginal posterior distributions and hence to extract characteristics of parameters such as Bayes estimators and credible intervals. The chain is run for 10,000 iterations with a burn-in period of 1,000. Details of the implementation of the MCMC algorithm used in this paper are given below.

i) choose starting values α_0 and β_0.

ii) at step $i+1$, we draw a new value α_{i+1} conditional on the current α_i from the Gamma distribution $\Gamma(\alpha_i/\mathrm{c}, \mathrm{c})$;

iii) the candidate α_{i+1} will be accepted with a probability given by the Metropolis ratio

$$u(\alpha_i\,,\,\alpha_{i+1}) = \min\left\{1\,,\,\frac{\Gamma(\alpha_i/c, c)p\big(\alpha_{i+1},\,\beta_i \mid \boldsymbol{x}\big)}{\Gamma(\alpha_{i+1}/c, c)p\big(\alpha_i, \beta_i \mid \boldsymbol{x}\big)}\right\}$$

iv) sample the new value $\boldsymbol{\beta}_{i+1}$ from the Gamma distribution $\Gamma(\beta_i/d, d)$;

v) the candidate β_{i+1} will be accepted with a probability given by the Metropolis ratio

$$u(\beta_i\,,\,\beta_{i+1}) = \min\left\{1\,,\,\frac{\Gamma(\beta_i/d, d)p\big(\alpha_{i+1},\,\beta_{i+1} \mid \boldsymbol{x}\big)}{\Gamma(\beta_{i+1}/d, d)p\big(\alpha_{i+1}, \beta_i \mid \boldsymbol{x}\big)}\right\}$$

The proposal distribution parameters c and d were chosen to obtain a good mixing of the chains and the convergence of the MCMC samples of parameters are assessed using the criteria given by Raftery and Lewis (1992). More details of MCMC in oder to construct these chains see, for example, Smith & Roberts (1993), Gelfand & Smith (1990), Gilks, Clayton, Spiegelhalter, Best, McNiel, Sharples & Kirby (1993).

We examine the performance of the priors by computing point estimates for parameters α and β based on 1,000 simulated samples and then we averaged the estimates of the parameters, obtain the variances and the coverage probability of 95% confidence intervals. Table 1 shows the point estimates for α and its respective variances given between parenthesis. Table 2 shows the same summaries for β.

The results of our numerical studies show that there is little difference between the point estimates for both parameters α and β. However, the MDIP prior produces a much smaller variance than using the other assumed priors. The uniform prior and MLE estimate produce bad estimations with large variances showing

Table 1: Summaries for parameter α.

$\boldsymbol{\alpha = 2}$	Jeffreys	MDIP	Tibshirani	Reference	Copula	Uniform	MLE
$n = 5$	2.2529	2.1894	2.3666	2.3602	2.0909	3.3191	3.2850
	(2.1640)	(0.5112)	(3.0297)	(2.4756)	(2.1987)	(3.1577)	(7.5372)
$n = 10$	2.5227	2.2253	2.4138	2.4761	2.3068	2.9769	2.7013
	(1.3638)	(0.3855)	(1.3849)	(1.3301)	(1.2052)	(1.4658)	(1.8308)
$n = 30$	2.1259	2.1744	2.0606	2.1079	2.0369	2.2504	2.2253
	(0.2712)	(0.1910)	(0.2651)	(0.2728)	(0.2571)	(0.2829)	(0.3138)

Table 2: Summaries for parameter β.

$\boldsymbol{\beta = 3}$	Jeffreys	MDIP	Tibshirani	Reference	Copula	Uniform	MLE
$n = 5$	2.9136	3.2680	3.2161	3.1058	2.7486	4.3419	5.2673
	(4.2292)	(1.6890)	(6.1384)	(4.7787)	(4.2447)	(5.6351)	(23.1881)
$n = 10$	3.8577	3.7727	3.6705	3.7649	3.6112	4.5186	4.2960
	(3.8086)	(1.5872)	(3.8110)	(3.6667)	(3.5599)	(3.7803)	(5.4296)
$n = 30$	3.2328	3.4255	3.1475	3.1798	3.0805	3.4633	3.4005
	(0.7950)	(0.6292)	(0.7861)	(0.7856)	(0.7453)	(0.8335)	(0.9163)

that, despite being widely used in the literature, they are not suitable for the Gamma distribution. As expected, the performance of these priors improves when the sample size increases.

Frequentist property of coverage probabilities for the parameters α and β have also been studied to compare the priors and MLE. Table 3 summarizes the simulated coverage probabilities of 95% confidence intervals. For the three sample sizes considered here, the intervals of MDIP prior produce an over-coverage for small sample sizes while, the intervals of uniform prior and MLE seem to have an under-coverage for some cases. Coverage probabilities are very close to the nominal value when n increases.

Table 3: Frequentist coverage probability of the 95% confidence intervals for α and β.

$\boldsymbol{\alpha = 2}$	Jeffreys	MDIP	Tibshirani	Reference	Copula	Uniform	MLE
$n = 5$	96.30%	99.60%	97.20%	96.10%	95.30%	95.70%	95.60%
$n = 10$	96.40%	99.50%	94.90%	95.00%	95.80%	90.60%	95.30%
$n = 30$	96.10%	96.20%	98.10%	95.80%	96.80%	95.50%	95.00%
$\boldsymbol{\beta = 3}$	Jeffreys	MDIP	Tibshirani	Reference	Copula	Uniform	MLE
$n = 5$	96.60%	99.60%	97.50%	97.00%	96.30%	99.90%	94.30%
$n = 10$	98.10%	98.00%	96.00%	96.70%	97.80%	93.90%	95.70%
$n = 30$	97.30%	95.80%	96.90%	96.10%	97.40%	94.90%	96.80%

8.2. Rainfall Data Example

Data in Table 4 represent the average monthly rainfall obtained from the Information System for Management of Water Resources of the State of São Paulo, including a period of 56 years from 1947 to 2003, by considering the month of November.

Let us assume a Gamma distribution with density (1) to analyse the data.

TABLE 4: Historical rainfall averages over last 56 years in State of São Paulo.

0.2,3.5,2.8,3.7,8.7,6.9,7.4,0.8,4.8,2.5,2.9,3.1,4.0,5.0,3.8,3.5,5.4,3.3,2.9,
1.7,7.3,2.9,4.6,1.1,1.4,3.9,6.2,4.1,10.8,3.8,7.3,1.8,6.7,3.5,3.2,5.2,2.8,5.2,
5.4,2.2,9.9,2.1,4.7,5.5,2.6,4.1,5.4,5.5,2.1,1.9,8.8,1.3,24.1,5.4,6.2,2.9

Table 5 presents the posterior means assuming the different prior distributions and maximum likelihood estimates (MLE) for the parameters α and β.

TABLE 5: Posterior means for parameters α and β of rainfall data.

	Uniform	Jeffreys	Ref-$\boldsymbol{\beta}$	MDIP	Tibshirani	Copula	MLE
α	2.493	2.387	2.393	2.659	2.357	2.380	2.395
β	0.543	0.516	0.517	0.641	0.510	0.515	0.518

From Table 5, we observe similar inference results assuming the different prior distributions for α and β, except for MDIP prior as observed in the simulation study introduced in the example presented in section 8.1.

The 95% posterior credible intervals obtained using the different priors for the parameters are displayed in Table 6. The MLE intervals for the parameters α and β are given respectively by (1.56; 3.22) and (0.31; 0.72).

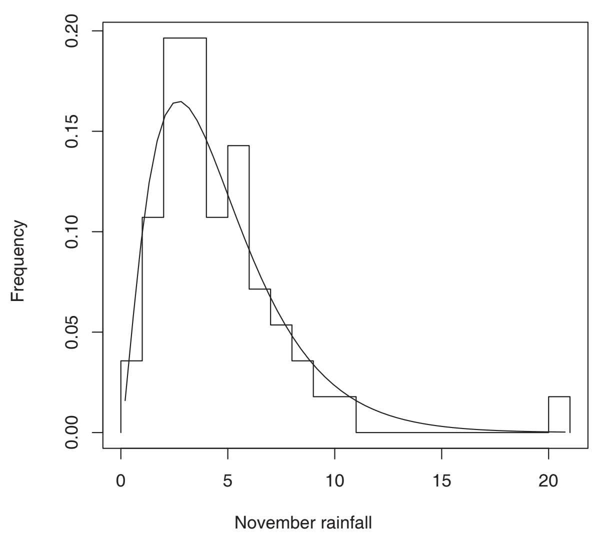

FIGURE 1: Histogram and fitted Gamma distribution for rainfall data.

TABLE 6: 95% posterior intervals for the parameters α and β of rainfall data.

	Uniform	Jeffreys	Ref-$\boldsymbol{\beta}$	MDIP	Tibshirani	Copula
α	(1.71; 3.43)	(1.63; 3.29)	(1.64; 3.28)	(1.91; 3.52)	(1.60; 3.25)	(1.60; 3.34)
β	(0.35; 0.76)	(0.33; 0.73)	(0.34; 0.73)	(0.44; 0.87)	(0.33; 0.73)	(0.32; 0.75)

Figure 2 shows the marginal posterior densities for both parameters α and β. We can see that the MDIP prior leads to a posterior slightly more sharply peaked for both parameters, while the other priors are quite similar, agreeing with simulated data with sample size $n = 30$.

To determine the appropriate prior distribution to be used with the rainfall data fitted by the Gamma distribution, some selection criteria can be examined. These include information-based criteria (AIC, BIC and DIC) given in the Table 7 for each prior distribution.

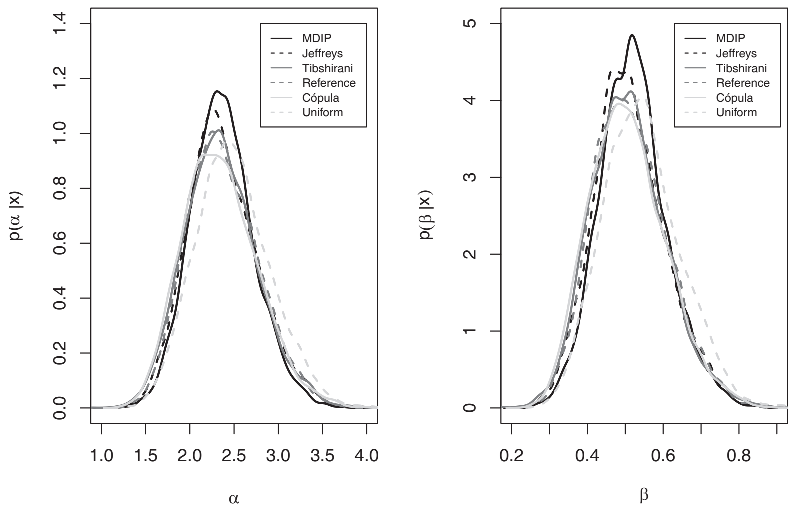

FIGURE 2: Plot of marginal posterior densities for the parameters α and β of rainfall data.

TABLE 7: Information-based model selection criteria (AIC, BIC and DIC) for rainfall data.

Prior	AIC	BIC	DIC
Jeffreys	272.213	268.162	267.827
MDIP	272.247	268.196	267.502
Ref-β	272.212	268.162	267.922
Tibshirani	272.219	268.169	268.068
Copula	272.222	268.171	268.197
Uniform	272.266	268.215	267.935

From the results of Table 7 and Figure 2 we observe that the choice of the prior distributions for parameters α and β has a negligible effect on the posterior distribution, surely due to the large amount of data in this study.

8.3. Reliability Data Example

In this example, we consider a lifetime data set related to an electrical insulator subjected to constant stress and strain introduced by Lawless (1982). The dataset does not have censored values and represent the lifetime (in minutes) to failure: 0.96, 4.15, 0.19, 0.78, 8.01, 31.75, 7.35, 6.50, 8.27, 33.91, 32.52, 16.03, 4.85, 2.78, 4.67, 1.31, 12.06, 36.71 and 72.89. Let us denote this data as "Lawless data". We assume a Gamma distribution with density (1) to analyse the data.

The maximum likelihood estimators and the Bayesian summaries for α and β, considering the different prior distributions are given in Table 8. Table 9 shows the 95% posterior intervals for α and β. The estimated marginal posterior distributions for the parameters are shown in Figure 3.

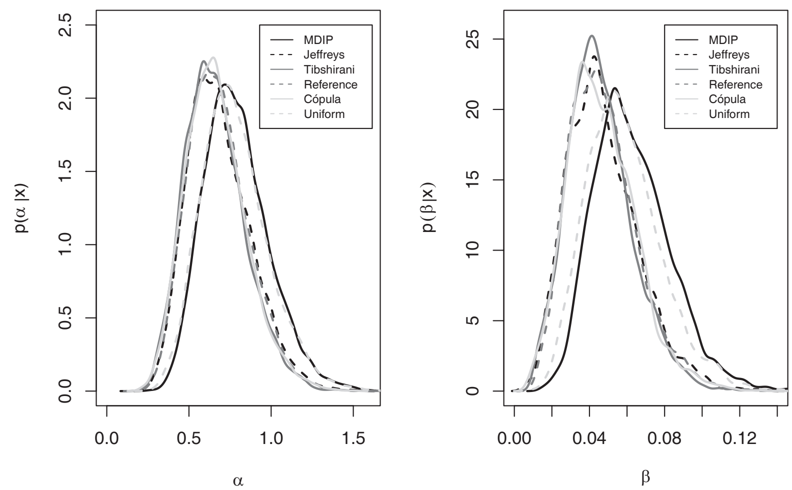

FIGURE 3: Plot of marginal posterior densities for the parameters α and β for Lawless data.

Tables 8 and 9 present the posterior statistics and 95% confidence intervals for both parameters resulting from the proposed priors. Again the performance of the MDIP prior clashes from the others.

TABLE 8: Posterior means for parameters α and β (Lawless data).

	Uniform	Jeffreys	Ref-$\boldsymbol{\beta}$	MDIP	Tibshirani	Copula	MLE
α	0.779	0.686	0.681	0.789	0.660	0.666	0.690
β	0.058	0.047	0.047	0.063	0.046	0.047	0.048

TABLE 9: 95% posterior intervals for the parameters α and β (Lawless data).

	Uniform	Jeffreys	Ref-$\boldsymbol{\beta}$	MDIP	Tibshirani	Copula
α	(0.433, 1.229)	(0.371, 1.111)	(0.374, 1.087)	(0.459, 1.232)	(0.350, 1.061)	(0.353, 1.080)
β	(0.025, 0.105)	(0.017, 0.091)	(0.018, 0.090)	(0.031, 0.108)	(0.017, 0.085)	(0.017, 0.087)

Table 10 shows the AIC, BIC and DIC values for all priors under investigation, with similar results as presented in Table 7 are obtained in this comparison which shows no differences using the different assumed priors.

Table 10: Information-based model selection criteria (AIC, BIC and DIC) for(Lawless data.

Prior	AIC	BIC	DIC
Jeffreys	143.125	141.236	141.409
MDIP	143.642	141.753	141.082
Ref-β	143.126	141.237	141.168
Tibshirani	143.148	141.259	141.247
Copula	143.138	141.249	141.471
Uniform	143.401	141.512	141.119

9. Conclusion and Discussion

The large number of noninformative priors can cause difficulties in the choosing one, especially when these priors does not produce similar results. Thus, in this paper, we presented a Bayesian analysis using a variety of prior distributions for the estimation of the parameters of the Gamma distribution.

We have shown that the use of the maximal data information process proposed by Zellner (1977), Zellner (1984), Zellner (1990) yields an improper posterior distribution for the parameters α and β. In this way, we proposed a "modified" MDIP prior analytically similar to the original one but with proper posterior. We also shown that the reference prior provides nonuniqueness of prior due to the choice of the parameter of interest, although the simulation shows the same performance. We have shown that the Tibshirani prior applied to the parameters of the Gamma distribution is equal to the reference prior when α is the parameter of interest.

Besides, a simulation study to check the impact of the use of different noninformative priors in the posterior distributions was also carried out. From this study we can conclude that it is necessary to carefully choose a prior for the parameters of the Gamma distribution when there is not enough data.

As expected, a moderated large sample size is need to achieve the desirable accuracy. In this case, the choice of the priors become irrelevant. However, the disagreement is substantial for small sample sizes.

Our simulation study indicates that the class of priors: Jeffreys, Reference, Tibshirani and Copula, had the same performance while the Uniform prior had worse performance. On the other hand , the "modified" MDIP prior produced the best estimations for α and β. Thus, the simulation study showed that the effect of the prior distributions can be substantial in the estimation of parameters and therefore the modified MDIP prior should be the recommended noninformative prior for the estimation of parameters of the Gamma distribution.

[Recibido: enero de 2013 — Aceptado: septiembre de 2013]

References

Apolloni, B. & Bassis, S. (2099), 'Algorithmic inference of two-parameter gamma distribution', *Communications in Statistics - Simulation and Computation* **38**(9), 1950–1968.

Berger, J. & Bernardo, J. M. (1992), On the development of the reference prior method, Fourth Valencia International Meeting on Bayesian Statistics, Spain.

Bernardo, J. M. (1979), 'Reference posterior distributions for Bayesian inference', *Journal of the Royal Statistical Society* **41**(2), 113–147.

Cox, D. R. & Reid, N. (1987), 'Parameter orthogonality and approximate conditional inference (with discussion)', *Journal of the Royal Statistical Society, Series B* **49**, 1–39.

Gelfand, A. E. & Smith, F. M. (1990), 'Sampling-based approaches to calculating marginal densities', *Journal of the American Statistical Association* **85**, 398–409.

Gilks, W., Clayton, D., Spiegelhalter, D., Best, N., McNiel, A., Sharples, L. & Kirby, A. (1993), 'Modeling complexity: Applications of Gibbs sampling in medicine', *Journal of the Royal Statistical Society, Series B* **55**(1), 39–52.

Jeffreys, S. H. (1967), *Theory of Probability*, 3 edn, Oxford University Press, London.

Lawless, J. (1982), *Statistical Models and Methods for Lifetime Data*, John Wiley, New York.

Min, C.-k. & Zellner, A. (1993), Bayesian Analysis, Model Selection and Prediction, *in* 'Physics and Probability: Essays in honor of Edwin T Jaynes', Cambridge University Press, pp. 195–206.

Moala, F. (2010), 'Bayesian analysis for the Weibull parameters by using noninformative prior distributions', *Advances and Applications in Statistics* (14), 117–143.

Morgenstern, D. (1956), 'Einfache beispiele sw edimensionaler vertielung', *Mit Mathematics Statistics* **8**, 234–235.

Nelsen, R. B. (1999), *An Introduction to Copulas*, Springer Verlag, New York.

Pradhan, B. & Kundu, D. (2011), 'Bayes estimation and prediction of the two-parameter Gamma distribution', *Journal of Statistical Computation and Simulation* **81**(9), 1187–1198.

Smith, A. & Roberts, G. (1993), 'Bayesian computation via the Gibbs sampler and related Markov Chain Monte Carlo Methods', *Journal of the Royal Statistical Society: Series B* **55**, 3–24.

Son, Y. & Oh, M. (2006), 'Bayesian estimation of the two-parameter Gamma distribution', *Communications in Statistics – Simulation and Computation* **35**, 285–293.

Tibshirani, R. (1989), 'Noninformative prioris for one parameters of many', *Biometrika* **76**, 604–608.

Trivedi, P. K. & Zimmer, D. M. (2005*a*), *Copula Modelling*, New Publishers, New York.

Trivedi, P. K. & Zimmer, D. M. (2005*b*), 'Copula modelling: An introduction to practicioners', *Foundations and Trends in Econometrics* (1), 1–111.

Zellner, A. (1977), Maximal data information prior distributions, *in* A. Aykac & C. Brumat, eds, 'In New Methods in the Applications of Bayesian Methods', North-Holland, Amsterdam.

Zellner, A. (1984), *Maximal Data Information Prior Distributions, Basic Issues in Econometrics*, The University of Chicago Press, Chicago, USA.

Zellner, A. (1990), Bayesian methods and entropy in economics and econometrics, *in* W. J. Grandy & L. Schick, eds, 'Maximum Entropy and Bayesian Methods', Dordrecht, Netherlands: Kluwer Academic Publishers, pp. 17–31.