Brazilian Journal of Probability and Statistics
2013, Vol. 27, No. 1, 1–19
DOI: 10.1214/10-BJPS133

Beta generalized distributions and related exponentiated models: A Bayesian approach

Jorge A. Achcar[a], Emílio A. Coelho-Barros[b] and Gauss M. Cordeiro[c]

[a]*Universidade de São Paulo, FMRP/USP*
[b]*Universidade Estadual de Maringá, DES/UEM*
[c]*Universidade Federal Rural de Pernambuco, DEINFO/UFRPE*

Abstract. We introduce a Bayesian analysis for beta generalized distributions and related exponentiated models. We review the exponentiated exponential, exponentiated Weibull and beta generalized exponential distributions. These distributions have been proposed as alternative extensions of the gamma and Weibull distributions in the analysis of lifetime data. Some posterior summaries of interest are obtained using Monte Carlo Markov chain (MCMC) methods. An application to a real data set is given to illustrate the potentiality of the Bayesian analysis.

1 Introduction

In recent years, several common distributions have been generalized via exponentiation. Let $G(y)$ be the cumulative distribution function (c.d.f.) of any continuous baseline distribution. The c.d.f. of the exponentiated G distribution is defined by elevating $G(x)$ to the power λ, say $F(y) = G(y)^{\lambda}$, where $\lambda > 0$ denotes an extra shape parameter. The baseline distribution is obtained as a special case when $\lambda = 1$. The advantage of this approach for modeling failure time data lies in its flexibility to model both monotonic as well as non-monotonic failure rates even though the baseline failure rate may be monotonic. Following this idea, Gupta, Gupta and Gupta (1998) introduced the exponentiated exponential (EE) distribution as a generalization of the exponential distribution. In the same way, Nadarajah and Kotz (2006) proposed four more exponentiated distributions that generalize the gamma, Weibull, Gumbel and Fréchet distributions and provided some mathematical properties for each distribution. Several other authors have considered exponentiated distributions, for example, Mudholkar and Hutson (1996), Gupta and Kundu (2001), Surles and Padgett (2001) and Kundu and Gupta (2007, 2008).

We consider a further extension of the exponentiated distributions starting from the baseline c.d.f. $G(y)$. Eugene, Lee and Famoye (2002) proposed a class of beta generalized distributions defined by

$$F(y) = I_{G(y)}(a,b) = \frac{1}{B(a,b)} \int_0^{G(y)} \omega^{a-1}(1-\omega)^{b-1}\, d\omega, \tag{1.1}$$

Key words and phrases. Beta generalized exponential distribution, exponentiated exponential distribution, exponentiated Weibull distribution, information matrix, maximum likelihood estimation.
Received February 2010; accepted November 2010.

where $B(a,b) = \int_0^1 \omega^{a-1}(1-\omega)^{b-1}\,d\omega$ is the beta function,

$$I_y(a,b) = B(a,b)^{-1}\int_0^y \omega^{a-1}(1-\omega)^{b-1}\,d\omega$$

denotes the incomplete beta function ratio, that is, the c.d.f. of the beta distribution with parameters $a > 0$ and $b > 0$. The role of these extra parameters is to introduce skewness and to very tail weight and the c.d.f. $G(y)$ could be quite arbitrary. The distribution F is so-called the beta G distribution. Evidently, the exponentiated G distribution is a special case of the beta G distribution when $b = 1$. Application of $Y = G^{-1}(V)$ to a beta random variable V with parameters a and b yields Y with c.d.f. (1.1).

The class of beta generalized distributions has been received special attention over the last years, in particular after recent works of Eugene, Lee and Famoye (2002) and Jones (2004). Eugene, Lee and Famoye (2002) defined the beta normal (BN) distribution by taking $G(y)$ in (1.1) to be the c.d.f. of the normal distribution and derived some of its first moments. General expressions for the moments of the BN distribution were obtained by Gupta and Nadarajah (2004). Nadarajah and Kotz (2004) proposed the beta Gumbel (BG) distribution by taking $G(y)$ as the c.d.f. of the Gumbel distribution and obtained closed-form expressions for the moments, the asymptotic distribution of the extreme order statistics and discussed maximum likelihood estimation. Further, Nadarajah and Kotz (2005) worked with the beta exponential (BE) distribution, derived the moment generating function, the first four cumulants, the asymptotic distribution of the extreme order statistics and also discussed maximum likelihood estimation. Some of the their results were generalized by Cordeiro, Simas and Stosic (2011) who investigated several mathematical properties for the beta Weibull (BW) distribution and obtained the maximum likelihood estimates (MLEs) of the model parameters.

The probability density function (p.d.f.) corresponding to (1.1) can be written as

$$f(y) = \frac{1}{B(a,b)} G(y)^{a-1}[1-G(y)]^{b-1} g(y), \tag{1.2}$$

where $g(y) = dG(y)/dy$. The p.d.f. $f(y)$ of the beta G will be most tractable when both functions $G(y)$ and $g(y)$ of the baseline distribution have simple analytic expressions. Except for some special choices for $G(y)$ in (1.1), it would appear that the p.d.f. $f(y)$ will be difficult to deal with in generality.

The rest of the paper is organized as follows. In Section 2, we review some exponentiated and beta generalized distributions. Section 3 proposes a Bayesian analysis for the EE, exponentiated Weibull (EW) and beta generalized exponential (BGE) distributions. Model selection is presented in Section 4. In Section 5, we give an application to a real data set to illustrate the Bayesian methods developed here. Finally, concluding remarks are given in Section 6.

2 Some exponentiated and beta generalized distributions

2.1 Exponentiated exponential distribution

The two parameter EE distribution (Gupta and Kundu (1999)) has been used as an alternative for the usual gamma and Weibull distributions in the analysis of lifetime data (Kundu and Gupta (2007, 2008), Raqab (2002), Raqab and Ahsanullah (2001), Zheng (2002)). This distribution is a special sub-model of the EW distribution (see Section 2.2) and can be used quite effectively in analyzing several lifetime data, particularly in place of the widely known gamma and Weibull distributions. If the shape parameter is one, then all the three distributions reduce to the one parameter exponential distribution. The three distributions represent generalizations of the exponential distribution in different ways. Unlike the gamma model, the density, distribution and survival functions of the EE distribution have convenient representations. In addition to the EW distribution, Nadarajah and Kotz (2006) studied the exponentiated gamma, exponentiated Gumbel and exponentiated Fréchet distributions by extending the gamma, Gumbel and Fréchet distributions in the same way that the EE distribution generalizes the exponential distribution. They also provide some mathematical properties for each exponentiated distribution.

The two parameter EE distribution has density function given by

$$f(y;\alpha,\lambda)=\alpha\lambda[1-\exp(-\lambda y)]^{\alpha-1}\exp(-\lambda y), \qquad y>0, \tag{2.1}$$

where $\alpha>0$ and $\lambda>0$ are the shape and scale parameters, respectively, like in the gamma and Weibull distributions. The EE density function is always right skewed and can be used quite effectively to analyze skewed data sets as an alternative to the more popular log-normal distribution. The p.d.f. (2.1) is concave and varies significantly depending on the shape parameter. For $\alpha<1$, it is a decreasing function and for $\alpha>1$, it is unimodal with mode at $\lambda^{-1}\log(\alpha)$, skewed, right tailed similar to the Weibull and gamma density functions. Both EE and gamma distributions can be considered as generalizations of the exponential distribution in different directions. In many situations, the EE distribution provides better fit than a gamma distribution and we can choose one of the two models to analyze a skewed data set. Even when α is very large, it is not symmetric. The mean, median and mode are nonlinear functions of the shape parameter and as α goes to infinity all of them tend to infinity. For large values of this parameter, the mean, median and mode are, approximately, equal to $\log(\alpha)$ but they converge at different rates. The EE distribution has several mathematical properties that are very similar to those of the gamma distribution but it has closed-form expressions for the distribution and survival functions like the Weibull distribution. The mean of both distributions diverges to infinity as the shape parameter goes to infinity. The corresponding survival and hazard functions are given by

$$S(y;\alpha,\lambda) = 1 - [1-\exp(-\lambda y)]^{\alpha}$$

and

$$h(y;\alpha,\lambda) = \frac{f(y;\alpha,\lambda)}{S(y;\alpha,\lambda)} = \frac{\alpha\lambda[1-\exp(-\lambda y)]^{\alpha-1}\exp(-\lambda y)}{1-[1-\exp(-\lambda y)]^{\alpha}},$$

respectively.

The hazard function of the EE distribution can be increasing from 0 to λ if $\alpha > 1$, decreasing from ∞ to 1 if $\alpha < 1$ and constant if $\alpha = 1$, similarly as the gamma distribution. Most properties of the EE distribution are identical in nature to those of the gamma distribution but computationally it is quite similar to the Weibull distribution. Hence, it can be used as an alternative distribution to the gamma and Weibull models and in some situations it might work better in terms of fitting than both models although it can not be guaranteed. In summary, the EE distribution is a good alternative for the gamma and Weibull models to be fitted to lifetime data.

We assume a random sample $\mathbf{y} = (y_1, \ldots, y_n)$ following the EE distribution (2.1). The log-likelihood function $\ell = \ell(\alpha,\lambda)$ for the model parameters reduces to

$$\ell(\alpha,\lambda) = n\log(\alpha) + n\log(\lambda) + (\alpha-1)\sum_{i=1}^{n}\log[1-\exp(-\lambda y_i)] - \lambda n\bar{y}, \qquad (2.2)$$

where $\bar{y}$ is the sample mean. The MLEs $\hat{\alpha}$ and $\hat{\lambda}$ of α and λ can be obtained from the score equations $\partial\ell/\partial\alpha = 0$ and $\partial\ell/\partial\lambda = 0$. The estimate $\hat{\alpha}$ is given by

$$\hat{\alpha} = \frac{-n}{\sum_{i=1}^{n}\log[1-\exp(\hat{\lambda}y_i)]}$$

and $\hat{\lambda}$ is the solution of the nonlinear equation

$$\frac{n}{\hat{\lambda}} - n\bar{y} + (\hat{\alpha}-1)\sum_{i=1}^{n}\frac{y_i\exp(-\hat{\lambda}y_i)}{[1-\exp(-\hat{\lambda}y_i)]} = 0.$$

The MLEs $\hat{\alpha}$ and $\hat{\lambda}$ of α and λ are consistent estimates with an asymptotic bivariate normal distribution Gupta and Kundu (1999), namely

$$(\hat{\alpha},\hat{\lambda})^T \sim N[(\alpha,\lambda)^T, I^{-1}(\alpha,\lambda)],$$

where $I(\alpha,\lambda)$ is the joint information matrix obtained from (2.2)

$$I(\alpha,\lambda) = \begin{pmatrix} E\Big(-\dfrac{\partial^2\ell}{\partial\alpha^2}\Big) & E\Big(-\dfrac{\partial^2\ell}{\partial\alpha\,\partial\lambda}\Big) \\ E\Big(-\dfrac{\partial^2\ell}{\partial\lambda\,\partial\alpha}\Big) & E\Big(-\dfrac{\partial^2\ell}{\partial\lambda^2}\Big) \end{pmatrix}, \qquad (2.3)$$

whose elements are given by

$$
\begin{aligned}
E\Big(-\frac{\partial^2\ell}{\partial\alpha^2}\Big) &= \frac{n}{\alpha^2},\\
E\Big(-\frac{\partial^2\ell}{\partial\alpha\,\partial\lambda}\Big) &= \frac{n}{\lambda}\Big\{[\psi(\alpha+1)-\psi(1)] - \frac{\alpha}{\alpha-1}[\psi(\alpha)-\psi(1)]\Big\},\\
E\Big(-\frac{\partial^2\ell}{\partial\lambda^2}\Big) &= \frac{n}{\lambda^2}\Big\{1+\frac{\alpha(\alpha-1)}{\alpha-2}\big[\psi'(1)-\psi'(\alpha-1)+[\psi(\alpha-1)-\psi(1)]^2\big]\\
&\qquad - \alpha\big[\psi'(1)-\psi'(\alpha)+[\psi(\alpha)-\psi(1)]^2\big]\Big\}.
\end{aligned} \tag{2.4}
$$

Here, $\Gamma(\cdot)$ and $\psi(\alpha) = d\log\Gamma(\alpha)/d\alpha = \Gamma'(\alpha)/\Gamma(\alpha)$ are the gamma and digamma functions, respectively.

2.2 Exponentiated Weibull distribution

The EW distribution was originally proposed by Mudholkar and Srivastava (1993) and later studied by Mudholkar, Srivastava and Freimer (1995). Its properties have been studied in more detail by Mudholkar and Hutson (1996) and Nassar and Eissa (2003). The EW family is an extension of the Weibull distribution obtained by adding an extra shape parameter. The beauty and importance of this distribution lies in its ability to model monotone as well as nonmonotone failure rates which are quite common in reliability and biological studies. As with any other distribution, many of its interesting characteristics and features can be studied through moments. The EW p.d.f. is given by

$$
f(y;\alpha,\lambda,\beta) = \alpha\lambda\beta[1-\exp(-\lambda y^\beta)]^{\alpha-1}\exp(-\lambda y^\beta)y^{\beta-1}, \qquad y>0, \tag{2.5}
$$

where the parameters α, λ and β are all positive. When $\alpha = 1$, equation (2.5) yields the Weibull distribution. Clearly, the EE distribution is also a sub-model of the EW distribution when $\beta = 1$. The survival and hazard functions are

$$
S(y;\alpha,\lambda,\beta) = 1-[1-\exp(-\lambda y^\beta)]^\alpha
$$

and

$$
h(y;\alpha,\lambda,\beta) = \frac{\alpha\lambda\beta[1-\exp(-\lambda y^\beta)]^{\alpha-1}\exp(-\lambda y^\beta)y^{\beta-1}}{1-[1-\exp(-\lambda y^\beta)]^\alpha},
$$

respectively.

This above hazard function is monotone increasing if $\beta \ge 1$ and $\alpha\beta \ge 1$; monotone decreasing if $\beta \le 1$ and $\alpha\beta \le 1$; unimodal if $\beta < 1$ and $\alpha\beta > 1$ and bathtub shaped if $\beta > 1$ and $\alpha\beta < 1$. Mudholkar and Srivastava (1993) derived some mathematical properties of the EW distribution.

We assume a random sample $\mathbf{y} = (y_1, \ldots, y_n)$ following (2.5). The log-likelihood function $\ell = \ell(\alpha, \lambda, \beta)$ for the model parameters can be written as

$$\begin{aligned}\ell(\alpha, \lambda, \beta) = {} & n \log(\alpha) + n \log(\lambda) + n \log(\beta) \\ & + (\alpha - 1) \sum_{i=1}^{n} \log[1 - \exp(-\lambda y_i^{\beta})] \\ & - \lambda \sum_{i=1}^{n} y_i^{\beta} + (\beta - 1) \sum_{i=1}^{n} \log(y_i).\end{aligned} \tag{2.6}$$

The MLEs $\hat{\alpha}$, $\hat{\lambda}$ and $\hat{\beta}$ are obtained from the nonlinear equations $\partial\ell/\partial\alpha = 0$, $\partial\ell/\partial\lambda = 0$ and $\partial\ell/\partial\beta = 0$ using any iterative algorithm.

2.3 Beta generalized exponential distribution

The four parameter BGE distribution (Barreto-Souza, Santos and Cordeiro (2010)) is defined by taking $G(y)$ in (1.1) to be the c.d.f. $G(y) = (1 - e^{-\lambda y})^{\alpha}$ of the EE distribution. The c.d.f. of the BGE distribution is given by

$$F(y; a, b, \alpha, \lambda) = \frac{1}{B(a, b)} \int_0^{(1-e^{-\lambda y})^{\alpha}} \omega^{a-1} (1 - \omega)^{b-1} \, d\omega, \qquad y > 0$$

for $a > 0$, $b > 0$, $\alpha > 0$ and $\lambda > 0$. The BGE density function does not involve any complicated function and it reduces to

$$\begin{aligned}f(y; a, b, \alpha, \lambda) = {} & \frac{\alpha\lambda}{B(a, b)} \exp(-\lambda y)(1 - e^{-\lambda y})^{\alpha a - 1} \\ & \times [1 - (1 - e^{-\lambda y})^{\alpha}]^{b-1}, \qquad y > 0.\end{aligned} \tag{2.7}$$

The BGE distribution generalizes some well-known distributions in the literature, such as the EE and BE distributions. The EE distribution is a special sub-model for the choice $a = b = 1$. The BE is also a special case for $\alpha = 1$. If, in addition, $a = b = 1$, we obtain the exponential distribution with parameter λ as a special case. Some mathematical properties of the BGE distribution are given by Barreto-Souza, Santos and Cordeiro (2010). It is evident that (2.7) is much more flexible than the EE and BE distributions because of the extra parameters. The corresponding hazard function becomes

$$h(y; a, b, \alpha, \lambda) = \frac{\alpha\lambda e^{-\lambda y}(1 - e^{-\lambda y})^{\alpha a - 1}[1 - (1 - e^{-\lambda y})^{\alpha}]^{b-1}}{B(a, b) I_{1-(1-e^{-\lambda y})^{\alpha}}(a, b)}, \qquad y > 0.$$

The log-likelihood function $\ell = \ell(a, b, \alpha, \lambda)$ for the model parameters of the BGE distribution given a random sample $\mathbf{y} = (y_1, \ldots, y_n)$ can be written from

Table 1 *Summary of the distributions*

	Exponentiated exponential distribution*
Density	$\alpha\lambda[1-\exp(-\lambda y)]^{\alpha-1}\exp(-\lambda y)$
Survival function	$1-[1-\exp(-\lambda y)]^{\alpha}$
Hazard function	$\frac{\alpha\lambda[1-\exp(-\lambda y)]^{\alpha-1}\exp(-\lambda y)}{1-[1-\exp(-\lambda y)]^{\alpha}}$
	Exponentiated Weibull distribution*
Density	$\alpha\lambda\beta[1-\exp(-\lambda y^{\beta})]^{\alpha-1}\exp(-\lambda y^{\beta})y^{\beta-1}$
Survival function	$1-[1-\exp(-\lambda y^{\beta})]^{\alpha}$
Hazard function	$\frac{\alpha\lambda\beta[1-\exp(-\lambda y^{\beta})]^{\alpha-1}\exp(-\lambda y^{\beta})y^{\beta-1}}{1-[1-\exp(-\lambda y^{\beta})]^{\alpha}}$
	Beta generalized exponential distribution*
Density	$\frac{\alpha\lambda}{B(a,b)}\exp(-\lambda y)(1-e^{-\lambda y})^{\alpha a-1}[1-(1-e^{-\lambda y})^{\alpha}]^{b-1}$
Hazard function	$\frac{\alpha\lambda e^{-\lambda y}(1-e^{-\lambda y})^{\alpha a-1}[1-(1-e^{-\lambda y})^{\alpha}]^{b-1}}{B(a,b)I_{1-(1-e^{-\lambda y})^{\alpha}}(a,b)}$

*$y>0$, $a>0$, $b>0$, $\alpha>0$, $\beta>0$, $\lambda>0$.

(2.7) as

$$\ell(a,b,\alpha,\lambda)=n\log\left(\frac{\alpha\lambda}{B(a,b)}\right)-\lambda\sum_{i=1}^{n}y_i+(\alpha a-1)\sum_{i=1}^{n}\log\bigl(1-\exp\{-(\lambda y_i)\}\bigr)$$
$$+(b-1)\sum_{i=1}^{n}\log(1-[1-\exp\{-\lambda y_i\}]^{\alpha}). \tag{2.8}$$

The MLEs $\hat{a}$, $\hat{b}$, $\hat{\alpha}$ and $\hat{\lambda}$ are obtained from the nonlinear equations $\partial\ell/\partial a=0$, $\partial\ell/\partial b=0$, $\partial\ell/\partial\alpha=0$ and $\partial\ell/\partial\lambda=0$ using any iteration procedure such as the Newton–Raphson or Fisher scoring method.

In Table 1, we provide a summary of the three previous distributions. Expressions for the mean, variance, skewness and kurtosis of these distributions are given in the references cited in the article.

3 A Bayesian analysis

3.1 A Bayesian analysis for the EE distribution

For a Bayesian analysis of the EE model, we can use different prior distributions for the model parameters α and λ. The Jeffreys invariant prior (Box and Tiao (1973)) for α and λ is given by

$$\pi_1(\alpha,\lambda)\propto[\det I(\alpha,\lambda)]^{1/2}, \tag{3.1}$$

where $I(\alpha,\lambda)$ is the joint information matrix (2.3). Another prior joint distribution for α and λ can take the form $\pi(\alpha,\lambda)=\pi(\lambda|\alpha)\pi_0(\alpha)$. In this way, using the

Jeffreys rule, we obtain

$$\pi_2(\alpha,\lambda) \propto \sqrt{E\Big(-\frac{\partial^2\ell}{\partial\lambda^2}\Big)}\pi_0(\alpha), \tag{3.2}$$

where $E(-\partial^2\ell/\partial\lambda^2)$ is given by (2.4). Since $\alpha > 0$, a Jeffreys noninformative prior for α becomes $\pi_0(\alpha) \propto 1/\alpha$. Hence,

$$\pi_2(\alpha,\lambda) \propto \frac{1}{\alpha\lambda}\sqrt{A(\alpha)}, \tag{3.3}$$

where

$$\begin{aligned} A(\alpha) = 1 + \frac{\alpha(\alpha-1)}{\alpha-2}[\psi'(1) - \psi'(\alpha-1) + \psi(\alpha-1) - \psi(1)] \\ - \alpha\{\psi'(1) - \psi'(\alpha) + [\psi(\alpha) - \psi(1)]^2\}. \end{aligned} \tag{3.4}$$

Assuming prior independence between the parameters α and λ, we take a noninformative prior distribution expressed as

$$\pi_3(\alpha,\lambda) \propto \frac{1}{\alpha\lambda}, \tag{3.5}$$

where $\alpha > 0$ and $\lambda > 0$.

We consider the reparametrization $\rho_1 = \log(\alpha)$ and $\rho_2 = \log(\lambda)$. We obtain from (3.5) a noninformative prior for ρ_1 and ρ_2, namely $\pi_4(\rho_1,\rho_2) \propto$ *constant*, where $-\infty < \rho_1 < \infty$ and $-\infty < \rho_2 < \infty$. In practical terms, we can consider an uniform prior distribution $U(-a_i, a_i)$ for $i = 1, 2$ with larger values for a_i to produce approximate noninformative priors for ρ_1 and ρ_2 and proper joint posterior distribution.

Further, we assume independence between ρ_1 and ρ_2. Using the reparametrization $\rho_1 = \log(\alpha)$ and $\rho_2 = \log(\lambda)$, the joint posterior distribution for ρ_1 and ρ_2 reduces to

$$\begin{aligned} \pi(\rho_1,\rho_2|y) \propto \pi(\rho_1,\rho_2) \\ \times \exp\Big\{ n\rho_1 + n\rho_2 - n\bar{y}\exp(\rho_2) \\ + [\exp(\rho_1) - 1]\sum_{i=1}^{n}\log\big(1 - \exp[-y_i\exp(\rho_2)]\big)\Big\}. \end{aligned} \tag{3.6}$$

Equation (3.6) is the likelihood function for ρ_1 and ρ_2, that is, $f(\mathbf{y}|\rho_1,\rho_2)$ is the joint distribution for the data in terms of ρ_1 and ρ_2.

Prior summaries of interest could be obtained using MCMC methods such as the Gibbs sampling algorithm (Gelfand and Smith (1990)) or the Metropolis–Hastings algorithm (Smith and Roberts (1993)).

If we assume the prior $\pi_4(\rho_1, \rho_2) \propto$ *constant*, the conditional posterior distributions used in the Gibbs sampling algorithm are given by

$$\pi(\rho_1|\rho_2, \mathbf{y}) \propto \exp\left\{n\rho_1 + [\exp(\rho_1) - 1]\sum_{i=1}^{n}\log\bigl(1 - \exp[-y_i \exp(\rho_2)]\bigr)\right\}. \tag{3.7}$$

and

$$\begin{aligned}\pi(\rho_2|\rho_1, \mathbf{y}) \propto \exp\Biggl\{&n\rho_2 - n\bar{y}\exp(\rho_2)\\ &+ [\exp(\rho_1) - 1]\sum_{i=1}^{n}\log\bigl(1 - \exp[-y_i \exp(\rho_2)]\bigr)\Biggr\}.\end{aligned} \tag{3.8}$$

We can consider the reparametrization $\rho_1 = \log(\alpha)$ and $\rho_2 = \log(\lambda)$ to obtain better performance for the Gibbs sampling algorithm. Equations (3.1)–(3.8) define a Bayesian analysis for the EE distribution.

It is important to point out that we could use alternative algorithms to perform the Bayesian analysis of the proposed models (Rue, Martino and Chopin (2009)) instead of using the standard MCMC methods.

3.2 A Bayesian analysis for the EW distribution

The joint posterior distribution for the parameters α, λ and β of the EW distribution with density (2.5) can be written as

$$\begin{aligned}\pi(\alpha, \lambda, \beta|\mathbf{y}) &\propto \pi_0(\alpha, \lambda, \beta)\\ &\quad\times \exp\Biggl\{n\log\alpha + n\log\lambda + n\log\beta - \lambda\sum_{i=1}^{n} y_i^{\beta}\\ &\qquad + (\alpha - 1)\sum_{i=1}^{n}\log[1 - \exp(-\lambda y_i^{\beta})] + (\beta - 1)\sum_{i=1}^{n}\log y_i\Biggr\},\end{aligned} \tag{3.9}$$

where $\pi_0(\alpha, \lambda, \beta)$ is a joint prior distribution for the model parameters α, λ and β.

If we assume that these parameters are independent with a joint noninformative prior $\pi_0(\alpha, \lambda, \beta) \propto (\alpha\lambda\beta)^{-1}$, $\alpha > 0$, $\lambda > 0$ and $\beta > 0$, and a reparametrization $\rho_1 = \log(\alpha)$, $\rho_2 = \log(\beta)$ and $\rho_3 = \log(\lambda)$, that is, the parameters ρ_1, ρ_2 and ρ_3 have a location uniform prior distribution (Box and Tiao (1973)), the conditional posterior distributions for the Gibbs sampling algorithm are given by

$$\pi(\rho_1|\rho_2, \rho_3, \mathbf{y}) \propto \exp\left\{n\rho_1 + (e^{\rho_1} - 1)\sum_{i=1}^{n}[1 - \exp(-e^{\rho_3} y_i^{e^{\rho_2}})]\right\}, \tag{3.10}$$

$$
\begin{aligned}
\pi(\rho_2|\rho_1, \rho_3, \mathbf{y}) \propto \exp\Bigg\{ & n\rho_2 - e^{\rho_3} \sum_{i=1}^{n} y_i^{e^{\rho_2}} + (e^{\rho_1} - 1) \sum_{i=1}^{n} [1 - \exp(-e^{\rho_3} y_i^{e^{\rho_2}})] \\
& + (e^{\rho_2} - 1) \sum_{i=1}^{n} \log y_i \Bigg\}
\end{aligned}
\tag{3.11}
$$

and

$$
\begin{aligned}
& \pi(\rho_3|\rho_1, \rho_2, \mathbf{y}) \\
& \quad \propto \exp\Bigg\{ n\rho_3 - e^{\rho_3} \sum_{i=1}^{n} y_i^{e^{\rho_2}} + (e^{\rho_1} - 1) \sum_{i=1}^{n} [1 - \exp(-e^{\rho_3} y_i^{e^{\rho_2}})] \Bigg\}.
\end{aligned}
\tag{3.12}
$$

By using uniform prior distributions for ρ_j, over a large interval $(-a_j, a_j)$, $j = 1, 2, 3$, we obtain the proper joint posterior distribution given by

$$
\begin{aligned}
& \pi(\rho_1, \rho_2, \rho_3|\mathbf{y}) \\
& \quad \propto \exp\Bigg\{ n[\rho_1 + \rho_2 + \rho_3] - e^{\rho_3} \sum_{i=1}^{n} y_i^{e^{\rho_2}} \\
& \qquad + (e^{\rho_1} - 1) \sum_{i=1}^{n} \log[1 - \exp(-e^{\rho_3} y_i^{e^{\rho_2}})] + (e^{\rho_2} - 1) \sum_{i=1}^{n} \log y_i \Bigg\},
\end{aligned}
\tag{3.13}
$$

which coincides with the joint density for the data $\mathbf{y}$ reparameterized in terms of ρ_1, ρ_2 and ρ_3.

3.3 A Bayesian analysis for the BGE distribution

We consider the BGE model with density function (2.7) and a noninformative joint prior distribution for a, b, α and λ given by

$$
\pi_0(a, b, \alpha, \lambda) \propto \frac{1}{ab\alpha\lambda},
\tag{3.14}
$$

where $a > 0$, $b > 0$, $\alpha > 0$ and $\lambda > 0$. The joint posterior distribution for these parameters can be written as

$$
\begin{aligned}
\pi(a, b, \alpha, \lambda|\mathbf{y}) \propto \; & \pi_0(a, b, \alpha, \lambda) B(a, b) \\
& \times \exp\Bigg\{ n \log\alpha + n \log\lambda - n \log \\
& \quad - \lambda \sum_{i=1}^{n} y_i + (\alpha a - 1) \sum_{i=1}^{n} \log(1 - e^{-\lambda y_i}) \\
& \quad + (b - 1) \sum_{i=1}^{n} \log[1 - (1 - e^{-\lambda y_i})^{\alpha}] \Bigg\}.
\end{aligned}
\tag{3.15}
$$

We now move to the reparametrization $\rho_1 = \log(a)$, $\rho_2 = \log(b)$, $\rho_3 = \log(\alpha)$ and $\rho_4 = \log(\lambda)$. We can easily obtain a locally uniform prior distribution for the new parameters ρ_1, ρ_2, ρ_3 and ρ_4 from equation (3.14).

Some posterior summaries of interest can be derived from the generated samples for the joint posterior distribution for the new parameters by using the Gibbs sampling algorithm. A considerable simplification can be achieved using the Winbugs software (Spiegelhalter et al. (1995)) which requires only the specification of the joint distribution for the data and the prior distributions for the model parameters. We observe that using uniform prior distributions over large intervals for the parameter ρ_1, ρ_2, ρ_3 and ρ_4, we have proper joint posterior distribution.

4 Model selection

Different model selection methods to choose the most adequate model could be adopted under the Bayesian paradigm (Berg, Meyer and Yu (2004)). We consider the Deviance Information Criterion (DIC) which is a specifically useful for selecting models under the Bayesian approach, where samples of the posterior distribution for the model parameters are obtained by using MCMC methods.

The deviance can be expressed as

$$D(\boldsymbol{\theta}) = -2 \log L(\boldsymbol{\theta}|\mathbf{y}) + c, \tag{4.1}$$

where $L(\boldsymbol{\theta}|\mathbf{y})$ is the likelihood function for the unknown parameters in $\boldsymbol{\theta}$ given the observed data $\mathbf{y}$ and c is a constant not required for comparing models.

Spiegelhalter, Best and Vander Linde (2000) defined the DIC criterion by

$$\mathrm{DIC} = D(\hat{\boldsymbol{\theta}}) + 2n_D, \tag{4.2}$$

where $D(\hat{\boldsymbol{\theta}})$ is the deviance evaluated at the posterior mean $\hat{\boldsymbol{\theta}}$ and n_D is the effective number of parameters in the model, namely $n_D = \bar{D} - D(\hat{\boldsymbol{\theta}})$, where $\bar{D} = E[D(\boldsymbol{\theta})]$ is the posterior deviance measuring the quality of the goodness-of-fit of the current model to the data. Smaller values of DIC indicate better models. Note that these values could be negative.

For model selection, we could also consider the conditional predictive ordinate (CPO) for each observation (Gelfand, Dey and Chang (1992)). CPO is a cross-validated predictive approach, the predictive density for y_i given $\mathbf{y}_{(i)} = (y_1, \dots, y_{i-1}, y_{i+1}, \dots, y_n)$, can be expressed as

$$c_i = p(y_i|\mathbf{y}_{(i)}) = \int p(y_i|\boldsymbol{\theta}) p(\boldsymbol{\theta}|\mathbf{y}_{(i)})\, d\boldsymbol{\theta}, \tag{4.3}$$

where $p(y_i|\boldsymbol{\theta})$ is the density proposed for the data and $p(\boldsymbol{\theta}|\mathbf{y}_{(i)})$ is the posterior density for the vector $\boldsymbol{\theta}$ of parameters given the data $\mathbf{y}_{(i)}$. We could obtain Monte Carlo estimates for c_i based on the generated MCMC sample.

Larger values of c_i (in average) indicate a better model. Alternatively, we could also select the model with larger value for the sum (or product) of the c_i's, $i = 1, \ldots, n$.

Another commonly used measure of goodness-of-fit is the Akaike information criterion (AIC) Akaike (1973, 1974) given by

$$\text{AIC} = -2 \log L(\hat{\boldsymbol{\theta}}|\mathbf{y}) + 2p, \tag{4.4}$$

where $L(\hat{\boldsymbol{\theta}}|\mathbf{y})$ is the maximized likelihood value and p is the number of parameters in the model. Smaller values of AIC indicate better models.

Although DIC values are given automatically by Winbugs software, which leads to a great simplification in practical work, there is some controversy about the use of DIC for model comparison in Bayesian context, as pointed out in the literature. In this way, it is recommended using some criteria such as CPO or AIC. In terms of MCMC methods, we adopt the expected AIC rather than the AIC criteria used in the classical approach. Other proposals have been suggested in the literature (Brooks (2002), Celeux et al. (2005)).

5 An example

We now consider a real data set introduced by Lawless (1982, p. 228) related to tests on the endurance of deep groove ball bearings Gupta and Kundu (2001), Lieblein and Zelen (1956). The data represent the number of million revolutions before failure of each of the 23 ball bearings in the life test. The data are: 17.88, 28.92, 33.00, 41.52, 42.12, 45.60, 48.80, 51.84, 51.96, 54.12, 55.56, 67.80, 68.64, 68.64, 68.88, 84.12, 93.12, 98.64, 105.12, 105.84, 127.92, 128.04 and 173.40.

First, we consider the EE distribution with density (2.1) under the reparametrization $\rho_1 = \log(\alpha)$ and $\rho_2 = \log(\lambda)$. Further, we assume approximate noninformative prior uniform $U(-10, 10)$ and $U(-10, 10)$ distributions for ρ_1 and ρ_2, respectively.

The use of uniform $U(-10, 10)$ prior distributions for ρ_1 and ρ_2 was considered to have approximate noninformative priors and convergence of the MCMC algorithm using the Winbugs software.

We generate 3,000 Gibbs samples taking every 10th sample after a "burn-in-sample" of size 5,000 to eliminate the initial values considered for the Gibbs sampling algorithm. All the calculations were performed using the Winbugs software. Convergence of the Gibbs sampling algorithm was verified from time series plots for the simulated samples. Table 2 lists the posterior summaries of interest for the EE model, the MLEs of α and λ and their corresponding standard errors and 95% confidence intervals (the Winbugs code is listed in Appendix).

Table 2 *Posterior summaries of interest and MLEs for some fitted models*

Model	Parameter	Posterior mean (SD)	95% credible interval	MLE (SD)	95% confidence interval
GE model	ρ_1	1.604 (0.387)	(0.843, 2.352)	1.664 (0.388)	(0.862, 2.467)
DIC = 229.901	ρ_2	−3.468 (0.206)	(−3.902, −3.099)	−3.433 (0.200)	(−3.844, −3.021)
AIC = 230.0	α	5.359 (2.133)	(2.324, 10.510)	5.283(2.050)	(1.044, 9.522)
	λ	0.032 (0.006)	(0.020, 0.045)	0.032 (0.006)	(0.019, 0.046)
EW model	ρ_1	1.543 (0.248)	(1.072, 1.965)	1.556 (1.627)	(−1.810, 4.922)
DIC = 228.974	ρ_2	0.057 (0.036)	(0.003, 0.129)	0.044 (0.647)	(−1.295, 1.382)
AIC = 231.9	ρ_3	−3.719 (0.180)	(−3.986, −3.339)	−3.673 (3.619)	(−11.159, 3.813)
	α	4.823 (1.177)	(2.922, 7.132)	4.740 (7.714)	(−11.217, 20.698)
	β	1.059 (0.038)	(1.003, 1.138)	1.045 (0.676)	(−0.354, 2.443)
	λ	0.025 (0.005)	(0.019, 0.035)	0.025 (0.092)	(−0.165, 0.216)
BGE model	ρ_1	1.385 (0.259)	(1.016, 1.923)	1.437 (75.083)	(−153.880, 156.760)
DIC = 229.053	ρ_2	0.034 (0.020)	(0.002, 0.068)	0.053 (3.054)	(−6.265, 6.370)
AIC = 234.0	ρ_3	0.390 (0.252)	(0.018, 0.926)	0.209 (75.296)	(−155.550, 155.970)
	ρ_4	−3.412 (0.158)	(−3.740, −3.115)	−3.473 (2.342)	(−8.318, 1.373)
	a	4.134 (1.135)	(2.763, 6.839)	4.207 (315.900)	(−649.290, 657.710)
	b	1.035 (0.021)	(1.002, 1.070)	1.054 (3.219)	(−5.604, 7.712)
	α	1.526 (0.409)	(1.018, 2.525)	1.233 (92.841)	(−190.820, 193.290)
	λ	0.033 (0.005)	(0.024, 0.044)	0.031 (0.073)	(−0.119, 0.181)

Secondly, we move to the EW distribution with density (2.5) and consider the reparametrization $\rho_1 = \log(\alpha)$, $\rho_2 = \log(\beta)$ and $\rho_3 = \log(\lambda)$. We adopt the same generating procedure to simulate Gibbs samples used for the GE distribution. We consider approximate noninformative prior uniform $U(1, 2)$, $U(0, 1)$ and $U(-4, -3)$ distributions for ρ_1, ρ_2 and ρ_3, respectively. Table 2 lists the posterior summaries of interest for the EW model based on 3,000 simulated Gibbs samples using the Winbugs software, the MLEs of the parameters α, β and λ and the corresponding standard errors and confidence intervals. The choices of the values of hyper-parameters of the uniform priors were required to obtain convergence of the Gibbs sampling algorithm.

Thirdly, we assume the BGE distribution with density (2.7) under the reparametrization $\rho_1 = \log(a)$, $\rho_2 = \log(b)$, $\rho_3 = \log(\alpha)$ and $\rho_4 = \log(\lambda)$. We consider approximate noninformative prior uniform $U(1, 2)$, $U(0, 0.1)$, $U(0, 1)$ and $U(-4, -3)$ distributions for ρ_1, ρ_2, ρ_3 and ρ_4, respectively. We adopt the same Gibbs sampling approach used for the EE and EW models. The calculations were done with the Winbugs software. Table 2 gives the posterior summaries of interest for the BGE model, the MLEs and the corresponding standard errors and confidence intervals.

For discrimination of the proposed models, we first use the DIC criterion. From the results of Table 2, we conclude that the EW model yields better fit to the data set, since its DIC value is smaller. Using the AIC criterion, we conclude (see Table 2) that the GE model is the best model fitted to the data. Note that the three models (GE, EW and BGE) yield similar fits, since their DIC or AIC values are quite close. Figure 1 plots the CPO's for the three fitted models, thus indicating similar fits with a small improvement for the BGE model, since their CPO's have, in average, bigger values.

6 Concluding remarks

Exponentiated and generalized beta distributions have been proved to be very versatile and a variety of uncertainties can be usefully modeled by them. Many of the classical distributions encountered in practice can be easily extended into the exponentiated and generalized beta forms. We show that the use of Bayesian methods to analyze these types of extended distributions is a suitable alternative to produce accurate inference for the parameters of interest. In the example considered in the article, we conclude that the usual maximum likelihood inference using classical asymptotic results could lead to larger confidence intervals when compared to the posterior summaries. The use of a software like Winbugs provides great facility to generate Gibbs samples for the joint posterior distribution of interest.

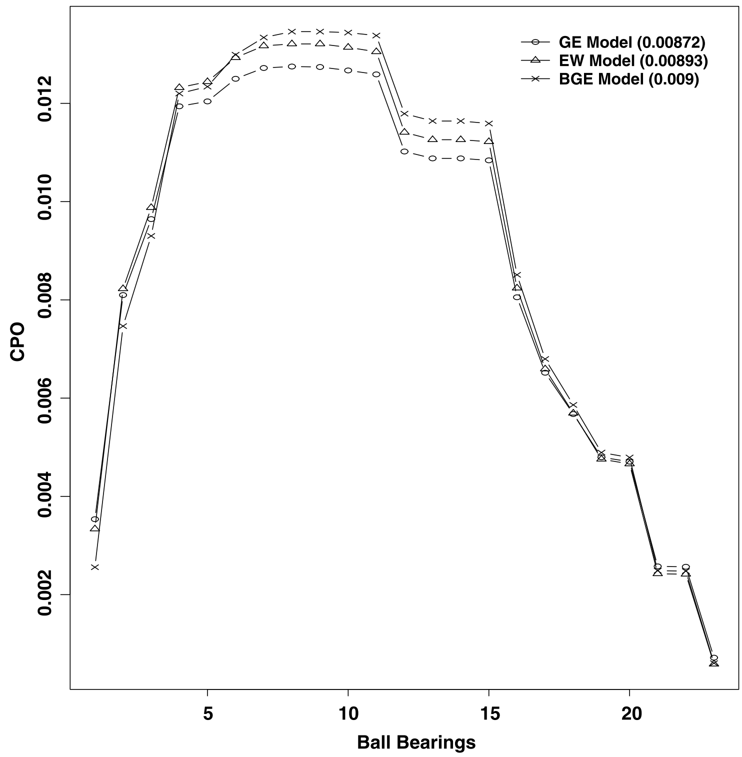

Figure 1 *CPO plots for the three fitted models.*

Appendix: Winbugs code for the three models

Listing 1 *Exponentiated exponential distribution codes*

```
model
{
  for (i in 1:N)
  {
    zeros[i] <- 0
    phi[i]      <- -log(L[i])
    zeros[i] ~ dpois(phi[i])

    L[i]<- exp(rho1+rho2-exp(rho2)*y[i]+
           (exp(rho1)-1)*log(1-exp(-exp(rho2)*y[i])))
  }

rho1 ~ dunif(-10,10)
rho2 ~ dunif(-10,10)
```

```
alpha <- exp(rho1)
lambda <- exp(rho2)
}
```

Listing 2 *Exponentiated weibull distribution codes*

```
model
{
  for (i in 1:N)
  {
   zeros[i] <- 0
   phi[i] <- -log(L[i])
   zeros[i] ~ dpois(phi[i])

   L[i]<- exp(rho1+rho2+rho3+(exp(rho1)-1)*
          log(1-exp(-exp(rho3)*pow(y[i],exp(rho2))))
          -exp(rho3)*pow(y[i],exp(rho2))
          +(exp(rho2)-1)*log(y[i]))
  }
rho1 ~ dunif(1,2)
rho2 ~ dunif(0,1)
rho3 ~ dunif(-4,-3)
alpha<-exp(rho1)
beta<-exp(rho2)
lambda<-exp(rho3)
}
```

Listing 3 *Beta generalized exponential distribution codes*

```
model
{
  for (i in 1:N)
  {
    zeros[i] <- 0
    phi[i] <- -log(L[i])
    zeros[i] ~ dpois(phi[i])

    L[i]<- exp(loggam(exp(rho1)+exp(rho2))
           -loggam(exp(rho1))-loggam(exp(rho2))+rho3+
           rho4-exp(rho4)*y[i]+(exp(rho3)*
           exp(rho1)-1)*log(1-exp(-exp(rho4)*y[i]))
           +(exp(rho2)-1)*log(1-pow(1-exp
           (-exp(rho4)*y[i]),exp(rho3))))
```

```
  }
rho1 ~ dunif(1,2)
rho2 ~ dunif(0,0.07)
rho3 ~ dunif(0,1)
rho4 ~ dunif(-4,-3)
a <- exp(rho1)
b <- exp(rho2)
alpha <- exp(rho3)
lambda <- exp(rho4)
}
```

Acknowledgments

The authors would like to thank two referees for their valuable suggestions.

References

Akaike, H. (1973). *Information Theory and an Extension of the Maximum Likelihood Principle*: *Proceedings of the 2nd International Symposium on Information Theory* (N. Petrov and F. Caski, eds.). Budapest: Adadémiai Kiadó. MR0483125

Akaike, H. (1974). A new look at statistical model identification. *IEEE Transactions Automatic Control* **19**, 716–722. MR0423716

Barreto-Souza, W., Santos, A. and Cordeiro, G. M. (2010). The beta generalized exponential distribution. *Journal of Statistical Computation and Simulation* **80**, 159–172. MR2603623

Berg, A., Meyer, R. and Yu, J. (2004). Deviance information criterion for comparing stochastic volatility models. *Journal of Business & Economic Statistics* **22**, 107–120. MR2019793

Box, G. E. and Tiao, G. C. (1973). *Bayesian Inference in Statistical Analysis*. New York: Addison-Wesley. MR0418321

Brooks, S. P. (2002). Discussion on the paper by Spiegelhalter, Best, Carlin and Van de Linde (2002). *Journal of the Royal Statistical Society, Ser. B* **64**, 616–618.

Celeux, G., Forbes, F., Robert, C. P. and Titterington, D. M. (2005). Deviance information criteria for missing data models. *Bayesian Analysis* **1**, 651–674.

Cordeiro, G. M., Simas, A. B. and Stosic, B. (2011). Closed form expressions for moments of the beta Weibull distribution. *Anais da Academia Brasileira de Ciências* **83**, 357–373.

Eugene, N., Lee, C. and Famoye, F. (2002). Beta-normal distribution and its applications. *Communication in Statistics—Theory and Methods* **31**, 497–512. MR1902307

Gelfand, A. E., Dey, D. K. and Chang, H. (1992). Model determination using predictive distributions with implementation via sampling-based methods. In *Bayesian Statistics* 147–167. New York: Oxford Univ. Press. MR1380275

Gelfand, A. E. and Smith, A. F. M. (1990). Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association* **85**, 398–409. MR1141740

Gupta, R. C., Gupta, P. L. and Gupta, R. D. (1998). Modeling failure time data by Lehman alternatives. *Communication in Statistics—Theory and Methods* **27**, 887–904. MR1613497

Gupta, R. D. and Kundu, D. (2001). Exponentiated exponential family: An alternative to gamma and Weibull distributions. *Biometrical Journal* **43**, 117–130.

Gupta, R. D. and Kundu, D. (1999). Generalized exponential distributions. *Australian & New Zealand Journal of Statistics* **41**, 173–188. MR1705342

Gupta, R. D. and Kundu, D. (2001). Exponentiated exponential family: An alternative to gamma and Weibull distributions. *Biometrical Journal* **43**, 117–130. MR1847768

Gupta, A. K. and Nadarajah, S. (2004). On the moments of the beta normal distribution. *Communication in Statistics—Theory and Methods* **33**, 1–13. MR2040549

Jones, M. C. (2004). Families of distributions arising from distributions of order statistics. *Test* **13**, 1–43. MR2065642

Kundu, D. and Gupta, R. D. (2007). A convenient way of generating gamma random variables using generalized exponential distribution. *Computational Statistics & Data Analysis* **51**, 2796–2802. MR2345608

Kundu, D. and Gupta, R. D. (2008). Generalized exponential distribution: Bayesian estimations. *Computational Statistics & Data Analysis* **52**, 1873–1883. MR2418477

Lawless, J. F. (1982). *Statistical Models and Methods for Lifetime Data*. New York: Wiley. MR0640866

Lieblein, J. and Zelen, M. (1956). Statistical investigation of the fatigue life of deep groove ball bearings. *Journal of Research of the National Bureau Standards* **57**, 273–316.

Mudholkar, G. S. and Hutson, A. D. (1996). The exponentiated Weibull family: Some properties and a flood data application. *Communication in Statistics—Theory and Methods* **25**, 3059–3083. MR1422323

Mudholkar, G. S. and Srivastava, D. K. (1993). Exponentiated Weibull family for analysing bathtub failure data. *IEEE Transactions on Reliability* **42**, 299–302.

Mudholkar, G. S., Srivastava, D. K. and Freimer, M. (1995). The exponentiated Weibull family. *Technometrics* **37**, 436–445.

Nadarajah, S. and Kotz, S. (2004). The beta Gumbel distribution. *Mathematical Problems in Engineering* **10**, 323–332. MR2109721

Nadarajah, S. and Kotz, S. (2005). The beta exponential distribution. *Reliability Engineering and System Safety* **91**, 689–697.

Nadarajah, S. and Kotz, S. (2006). The exponentiated type distributions. *Acta Applicandae Mathematicae* **92**, 97–111. MR2265333

Nassar, M. M. and Eissa, F. H. (2003). On the exponentiated Weibull distribution. *Communication in Statistics—Theory and Methods* **32**, 1317–1336. MR1985853

Raqab, M. Z. (2002). Inferences for generalized exponential distribution based on record statistics. *Journal of Statistical Planning and Inference* **104**, 339–350. MR1906016

Raqab, M. Z. and Ahsanullah, M. (2001). Estimation of the location and scale parameters of generalized exponential distribution based on order statistics. *Journal of Statistical Computation and Simulation* **69**, 109–124. MR1897171

Rue, H., Martino, S. and Chopin, N. (2009). Approximete Bayesian inference for latent Gaussian models using integrated nested Laplace approximations (with discussion). *Journal of the Royal Statistical Society, Ser. B* **71**, 319–392.

Smith, A. F. M. and Roberts, G. O. (1993). Bayesian computation via the Gibbs sampler and related Markov chain Monte Carlo methods. *Journal of the Royal Statistical Society, Ser. B* **55**, 3–23. MR1210421

Spiegelhalter, D. J., Best, N. G. and Vander Linde, A. (2000). A Bayesian measure of model complexity and fit (with discussion). *Journal of the Royal Statistical Society, Ser. B* **64**, 583–639. MR1979380

Spiegelhalter, D. J., Thomas, A., Best, N. G. and Gilks, W. R. (1995). *BUGS: Bayesian Inference Using Gibbs Sampling, Version 0.50*. Cambridge: MRC Biostatistics Unit.

Surles, J. G. and Padgett, W. J. (2001). Inference for reliability and stress-strength for a scaled Burr type X distribution. *Lifetime Data Analysis* **7**, 187–200. MR1842327

Zheng, G. (2002). Fisher information matrix in type-II censored data from exponentiated exponential family. *Biometrical Journal* **44**, 353–357. MR1898411

J. A. Achcar
Departamento de Medicina Social, FMRP/USP
Universidade de São Paulo
Ribeirão Preto, SP
Brazil
E-mail: achcar@fmrp.usp.br

E. A. Coelho-Barros
Departamento de Estatística, DEs/UEM
Universidade Estadual de Maringá
Maringá, PR
Brazil
E-mail: eacbarros2@uem.br

G. M. Cordeiro
Departamento de Informática, DEINFO/UFRPE
Universidade Federal Rural de Pernambuco
Recife, PE
Brazil
E-mail: gausscordeiro@uol.com.br