UNIVERSIDADE DE SÃO PAULO

USE OF EXPONENTIAL POWER DISTRIBUTIONS FOR MIXTURE MODELS IN THE PRESENCE OF COVARIATES

JORGE ALBERTO ACHCAR
GILBERTO DE ARAÚJO PEREIRA

Nº 47

NOTAS

Instituto de Ciências Matemáticas de São Carlos

Instituto de Ciências Matemáticas de São Carlos

ISSN - 0103-2577

USE OF EXPONENTIAL POWER DISTRIBUTIONS FOR MIXTURE MODELS IN THE PRESENCE OF COVARIATES

JORGE ALBERTO ACHCAR
GILBERTO DE ARAÚJO PEREIRA

Nº 47

NOTAS DO ICMSC
Série Estatística

São Carlos
Abr./1998

USE OF EXPONENTIAL POWER DISTRIBUTIONS FOR MIXTURE MODELS IN THE PRESENCE OF COVARIATES

Jorge Alberto Achcar
Gilberto de Araújo Pereira
University of SãoPaulo
ICMSC, USP, C.P. 668
13560-970, SãoCarlos, S.P., Brazil

Abstract

In this paper, we present a Bayesian analysis of exponential power mixture models in the presence of a covariate. Considering Gibbs with Metropolis-Hastings algorithms, we get Monte Carlo estimates for the posterior quantities of interest.
Key words: exponential power mixture model, Bayesian analysis, Gibbs sampling algorithm.

1 Introduction

The use of mixture models has been considered in the literature as an alternative to nonparametric methods to analyse data, since in many applications, the usual parametrical models could not be appropriate for the particular data set. These data could be observed when a group of subjects may not react to a given treatment.

Considering the introduction of a covariate vector $\underline{x}$ which may influence both the incidence probabilities and the conditional latency distribution, the mixture model (see for example, Kuo and Peng, 1995) assumes the density.

$$f\left(y|\underline{x},\underline{\theta}\right) = \sum_{j=1}^{J} P\left(j|\underline{x},\underline{\gamma}\right) \; f_j\left(y|j,\underline{x},\underline{\beta_j}\right) \tag{1}$$

where y is the value of a randon variable Y and $\underline{\theta} = \left(\underline{\beta_1},\underline{\beta_2},\ldots,\underline{\beta_J},\underline{\gamma}\right)$ is the vector of all unknown parameters.

The probabilities $P(j|\underline{x},\underline{\gamma})$, assumes that $\sum_{j=1}^{J} P(j|\underline{x},\underline{\gamma}) = 1$, where $\underline{\gamma}$ is the vector of parameters in the incidence probabilities.

Logistic regression links could be considered for the incidence probabilities, that is,

$$P(j|\underline{x},\underline{\gamma}) = \frac{e^{\underline{x}'\underline{\gamma_j}}}{\sum_{j=1}^{J} e^{\underline{x}'\underline{\gamma_j}}} \qquad (2)$$

The cumulative distribution function for Y, derived from (1), is given by

$$F(y|\underline{x},\underline{\theta}) = \sum_{j=1}^{J} P\left(j|\underline{x},\underline{\gamma}\right) F_j\left(y|\underline{x},\underline{\beta_j}\right) \qquad (3)$$

where F_j is the distribution function for f_j.

In this paper, we assume mixture of exponential power distributions (see for example, Box and Tiao, 1973) with density,

$$f_j\left(y|\underline{x},\underline{\beta_j}\right) = \frac{w(\delta_j)}{\sigma_j} \exp\left\{-c(\delta_j)\left|\frac{y-\theta_j}{\sigma_j}\right|^{2/(1+\delta_j)}\right\} \qquad (4)$$

where $-\infty < y < \infty$; j = 1,2;

$$c(\delta_j) = \left\{\frac{\Gamma\left[\frac{3}{2}(1+\delta_j)\right]}{\Gamma\left[\frac{1}{2}(1+\delta_j)\right]}\right\}^{1/(1+\delta_j)}$$

and,

$$w(\delta_j) = \frac{\left\{\Gamma\left[\frac{3}{2}(1+\delta_j)\right]\right\}^{1/2}}{(1+\delta_j)\left\{\Gamma\left[\frac{1}{2}(1+\delta_j)\right]\right\}^{3/2}},$$

$\sigma_j > 0, -1 < \delta_j < 1$ and $-\infty < \theta_j < \infty$.

This distribution includes a wider class of symmetric distributions which includes the normal distribution ($\delta_j = 0$), together with other distributions more leptokurtic ($\delta_j > 0$) or more platykurtic ($\delta_j < 0$).

Classical inference for mixture models based on the maximum likelihood estimators could be difficult even for simple cases considering $J = 2$ (see for example, Titterington et al, 1985). Thus, we consider a Bayesian approach based on Gibbs sampling with Metropolis-Hastings algorithms (see for example, Gelfand and Smith, 1990; or Smith and Roberts, 1993).

These Markov Chain Monte Carlo (MCMC) methods has been explored in the literature by many authors for finite mixture of distributions (see for example, Diebolt and Robert, 1994; or Kuo and Peng, 1995).

2 A Bayesian Analysis for the Model

Assuming $\underline{y} = (y_1, y_2, \ldots, y_n)$ a random sample of size n, the likelihood function for $\underline{\theta}$ is given by,

$$L\left(\underline{\theta}|\underline{y}, \underline{x}\right) = \prod_{i=1}^{n} \sum_{j=1}^{J} P\left(j|\underline{x}, \underline{\gamma}\right) \; f_j\left(y_i|\underline{x}, \underline{\beta_j}\right) \tag{5}$$

Considering the special case $J = 2$ and assuming a prior density $\pi(\underline{\theta})$, the joint posterior density for $\underline{\theta}$ is given by,

$$\pi\left(\underline{\theta}|\underline{y}, \underline{x}\right) \propto \pi\left(\underline{\theta}\right) \left\{ \prod_{i=1}^{n} \sum_{j=1}^{2} P\left(j|\underline{x}, \underline{\gamma}\right) \; f_j\left(y_i|\underline{x}, \underline{\beta_j}\right) \right\} \tag{6}$$

To get better performance for the Gibbs sampling algorithm, we consider the introduction of latent variables (see for example, Kuo and Peng, 1995) given by $(z_{i1}, z_{i2}), i = 1, 2, \ldots, n$ where $z_{i1}|\underline{\theta}, y_i, x_i \sim b(1, h_{i1})$ (a Bernoulli distribution) with h_{i1} given by

$$h_{li1} = \frac{P\left(1|x_i, \underline{\gamma}\right) \; f_1\left(y_i|x_i, \underline{\beta_1}\right)}{\sum_{j=1}^{2} P\left(j|x_i, \underline{\gamma}\right) \; f_j\left(y_i|x_i, \underline{\beta_j}\right)} \tag{7}$$

That is,

$$\pi\left(\underline{z_i}\right) \;\propto\; h_{i1}^{z_{i1}}\left(1-h_{i1}\right)^{z_{i2}} \tag{8}$$

where $z_{i1} = 1$ with probability $h_{i1}(z_{i1} = 0$ with probability $1 - h_{i1})$. Observe that $z_{i1} + z_{i2} = 1$.

Thus,

$$\pi\left(\underline{z_1}, \ldots, \underline{z_n}\right) \;\propto\; \frac{\prod_{i=1}^{n}\prod_{j=1}^{2}\left\{P\left(j|x_i,\underline{\gamma}\right)\; f_j\left(y_i|x_i,\underline{\beta_j}\right)\right\}^{z_{ij}}}{\prod_{i=1}^{n}\left\{\sum_{j=1}^{2} P\left(j|x_i,\underline{\gamma}\right)\; f_j\left(y_i|x_i,\underline{\beta_j}\right)\right\}} \tag{9}$$

Combining (9) with (6), we get,

$$\pi\left(\underline{z_1}, \underline{z_2}, \ldots, \underline{z_n}, \underline{\theta}|\underline{y}, \underline{x}\right) \;\propto\; \pi\left(\underline{\theta}\right)\left\{\prod_{i=1}^{n}\prod_{j=1}^{2}\;\left\{P\left(j|x_i,\underline{\gamma}\right)\; f_j\left(y_i|x_i,\underline{\beta_j}\right)\right)\right\}^{z_{ij}}\right\} \tag{10}$$

To generate samples of the joint posterior distribution (10), we use the Gibbs sampling algorithm. Starting with initial values $\underline{\theta}^{(0)} = (\theta_1^{(0)}, \ldots, \theta_p^{(0)})$, follow the steps:

(i) Generate a sample $\underline{z}^{(1)} = \left(\underline{z}_1^{(1)}, \underline{z}_2^{(1)}, \ldots, \underline{z}_n^{(1)}\right)$,

from (8). (11)

(ii) Generate a sample of θ, from the conditional distributions

$\pi\left(\theta_1|\theta_2^{(0)}, \ldots, \theta_p^{(0)}, \underline{z}^{(1)}, \underline{x}\right), \pi\left(\theta_2|\theta_1^{(1)}, \theta_3^{(0)} \ldots, \theta_p^{(0)}, \underline{z}^{(1)}, \underline{x}\right), \ldots,$

$\pi\left(\theta_p|\theta_1^{(1)}, \ldots, \theta_{p-1}^{(1)}, \underline{z}^{(1)}, \underline{x}\right).$

Then, continue iteration by repeating steps (i) and (ii).

2.1 Mixture of Exponential Power Distributions

Let us assume only a covariate x, a mixture of $J = 2$ exponential power distributions (1), with $\theta_j = \alpha_j + \beta_j x, j = 1, 2$; and the logistic regression link (2) with,

$$P(1|x,\gamma,\tau) = \frac{e^{\gamma+\tau x}}{1+e^{\gamma+\tau x}}, \text{and}$$

$$P(2|x,\gamma,\tau) = 1 - P(1|\underline{x},\gamma,\tau) = \frac{1}{1+e^{\gamma+\tau x}}.$$

Assuming prior independence among the parameters, consider the following prior densities for $\alpha_1, \beta_1, \sigma_1, \delta_1, \alpha_2, \beta_2, \sigma_2, \delta_2, \gamma$, and τ:

$$
\begin{aligned}
&\text{(i) } \alpha_1 \sim N(\alpha_{10}, \sigma_{11}^2); \alpha_{10}, \sigma_{11}^2 \text{ known},\\
&\text{(ii) } \beta_1 \sim N(\beta_{10}, \sigma_{12}^2); \beta_{10}, \sigma_{12}^2 \text{ known},\\
&\text{(iii) } \sigma_1 \sim \Gamma[m_{11}, n_{11}];\ m_{11}, n_{11} \text{ known},\\
&\text{(iv) } \delta_1 \sim N(\delta_{10}, \sigma_{13}^2); \delta_{10}, \sigma_{13}^2 \text{ known},\\
&\text{(v) } \alpha_2 \sim N(\alpha_{20}, \sigma_{21}^2); \alpha_{20}, \sigma_{21}^2 \text{ known},\\
&\text{(vi) } \beta_2 \sim N(\beta_{20}, \sigma_{22}^2); \beta_{20}, \sigma_{22}^2 \text{ known},\\
&\text{(vii) } \sigma_2 \sim \Gamma[m_{22}, n_{22}];\ m_{22}, n_{22} \text{ known},\\
&\text{(viii) } \delta_2 \sim N(\delta_{20}, \sigma_{23}^2); \delta_{20}, \sigma_{23}^2 \text{ known},\\
&\text{(ix) } \gamma \sim N(\gamma_0, \sigma_{14}^2), \gamma_0, \sigma_{14}^2 \text{ known},\\
&\text{(x) } \tau \sim N(\tau_0, \sigma_{24}^2), \tau_0, \sigma_{24}^2 \text{ known},
\end{aligned}
\tag{12}
$$

where $N(\mu, \sigma^2)$ denotes a normal distribution with mean μ and variance σ^2; $\Gamma[a, b]$ denotes a gamma distribution with mean a/b and variance a/b^2 .

The joint posterior distribution for $\underline{\theta} = (\alpha_1, \beta_1, \sigma_1, \delta_1, \alpha_2, \beta_2, \sigma_2, \delta_2, \gamma, \tau)$ is (from (6)) given by,

$$\pi\left(\underline{\theta} \mid \underline{y}, \underline{x}\right) \propto \pi(\underline{\theta})\, L\left(\underline{\theta} \mid \underline{y}, \underline{x}\right) \tag{13}$$

where

$$\pi(\underline{\theta}) \propto exp\left\{-\frac{1}{2\sigma_{11}^2}(\alpha_1 - \alpha_{10})^2\right\} \; exp\left\{-\frac{1}{2\sigma_{12}^2}(\beta_1 - \beta_{10})^2\right\}$$

$$exp\left\{-\frac{1}{2\sigma_{13}^2}(\delta_1 - \delta_{10})^2\right\} \; exp\left\{-\frac{1}{2\sigma_{14}^2}(\gamma - \gamma_0)^2\right\}$$

$$exp\left\{-\frac{1}{2\sigma_{24}^2}(\tau - \tau_0)^2\right\} \; \sigma_1^{m_{11}-1}\, \sigma_2^{m_{22}-1}\, exp\left\{-n_{11}\sigma_1 - n_{22}\sigma_2\right\}$$

$$exp\left\{-\frac{1}{2\sigma_{21}^2}(\alpha_2 - \alpha_{20})^2\right\} \; exp\left\{-\frac{1}{2\sigma_{22}^2}(\beta_2 - \beta_{20})^2\right\}$$

$$exp\left\{-\frac{1}{2\sigma_{23}^2}(\delta_2 - \delta_{20})^2\right\}$$

and

$$L\left(\underline{\theta} \mid \underline{y}, \underline{x}\right) = \prod_{i=1}^{n}\left\{\frac{e^{\gamma+\tau x_i}}{1+e^{\gamma+\tau x_i}}\left[w(\delta_1)\sigma_1^{-1} exp\left(-c(\delta_1)\left|\frac{y_i - \alpha_1 - \beta_1 x_i}{\sigma_1}\right|^{2/(1+\delta_1)}\right)\right] + \right.$$

$$\left.\frac{1}{(1+e^{\gamma+\tau x_i})}\left[w(\delta_2)\sigma_2^{-1} exp\left(-c(\delta_2)\left|\frac{y_i - \alpha_2 - \beta_2 x_i}{\sigma_2}\right|^{2/(1+\delta_2)}\right)\right]\right\}$$

With the introduction of latent variables $(z_{i1}, z_{i2}), i = 1, 2, \ldots, n,$ we get (from(10)),

$$\pi\left(\underline{z_1}, \ldots, \underline{z_n}, \underline{\theta} \mid \underline{y}, \underline{x}\right) \propto \pi(\underline{\theta})\left\{\frac{exp(r\gamma + a_1\tau)}{\prod_{i=1}^{n}(1+e^{\gamma+\tau x_i})}\right\}\frac{[w(\delta_1]^r[w(\delta_2]^{n-r}}{\sigma_1^r\sigma_2^{n-r}}$$

$$exp\left\{-\frac{c(\delta_1)}{\sigma_1^{2/(1+\delta_1)}}B_1(\alpha_1, \beta_1, \delta_1) - \frac{c(\delta_2)}{\sigma_2^{2/(1+\delta_2)}}B_2(\alpha_2, \beta_2, \delta_2)\right\} \tag{14}$$

where $\pi(\underline{\theta})$ is given in (13),

$r = \sum_{i=1}^{n} z_{i1}, n - r = \sum_{i=1}^{n} z_{i2}, a_1 = \sum_{i=1}^{n} x_i z_{i1},$

$B_1(\alpha_1, \beta_1, \delta_1) = \sum_{i=1}^{n} z_{i1} \left| y_i - \alpha_1 - \beta_1 x_i \right|^{2/(1+\delta_1)},$

$B_2(\alpha_2, \beta_2, \delta_2) = \sum_{i=1}^{n} z_{i2} \left| y_i - \alpha_2 - \beta_2 x_i \right|^{2/(1+\delta_2)}.$

To generate samples of the joint distribution (14), we use steps (i) and (ii) of the Gibbs algorithm (11), where the conditional distributions for the parameters are given by,

(i) $\quad \pi\left(\alpha_1 \mid \beta_1, \sigma_1, \delta_1, \alpha_2, \beta_2, \sigma_2, \delta_2, \gamma, \tau, \underline{z}, \underline{y}, \underline{x}\right) \propto$

$$exp\left\{-\frac{1}{2\sigma_{11}^2}(\alpha_1 - \alpha_{10})^2\right\} \Psi_1(\underline{\theta})$$

where,

$$\Psi_1(\underline{\theta}) = exp\left\{-\frac{c(\delta_1)}{\sigma_1^{2/(1+\delta_1)}} B_1(\alpha_1, \beta_1, \delta_1)\right\}$$

(ii) $\quad \pi\left(\beta_1 \mid \alpha_1, \sigma_1, \delta_1, \alpha_2, \beta_2, \sigma_2, \delta_2, \gamma, \tau, \underline{z}, \underline{y}, \underline{x}\right) \propto$

$$exp\left\{-\frac{1}{2\sigma_{12}^2}(\beta_1 - \beta_{10})^2\right\} \Psi_2(\underline{\theta})$$

where,

$$\Psi_2(\underline{\theta}) = exp\left\{-\frac{c(\delta_1)}{\sigma_1^{2/(1+\delta_1)}} B_1(\alpha_1, \beta_1, \delta_1)\right\}$$

(iii) $\quad \pi\left(\sigma_1 \mid \alpha_1, \beta_1, \delta_1, \alpha_2, \beta_2, \sigma_2, \delta_2, \gamma, \tau, \underline{z}, \underline{y}, \underline{x}\right) \propto$

$$\sigma_1^{m_{11}-1} exp\left\{-n_{11}\sigma_1\right\} \Psi_3(\underline{\theta})$$

where,

$$\Psi_3(\underline{\theta}) = exp\left\{-rln(\sigma_1) - \sigma_1^{-2/(1+\delta_1)}c(\delta_1)B_1(\alpha_1,\beta_1,\delta_1)\right\}$$

(iv) $$\pi\left(\delta_1 \mid \alpha_1,\beta_1,\sigma_1,\alpha_2,\beta_2,\sigma_2,\delta_2,\gamma,\tau,\underline{z},\underline{y},\underline{x}\right) \propto$$

$$exp\left\{-\frac{1}{2\sigma_{13}^2}(\delta_1-\delta_{10})^2\right\}\Psi_4(\underline{\theta})$$

where,

$$\Psi_4(\underline{\theta}) = exp\left\{rln(w(\delta_1)) - \frac{c(\delta_1)}{\sigma_1^{2/(1+\delta_1)}}B_1(\alpha_1,\beta_1,\delta_1)\right\}$$

(v) $$\pi\left(\alpha_2 \mid \alpha_1,\beta_1,\sigma_1,\delta_1,\beta_2,\sigma_2,\delta_2,\gamma,\tau,\underline{z},\underline{y},\underline{x}\right) \propto$$

$$exp\left\{-\frac{1}{2\sigma_{21}^2}(\alpha_2-\alpha_{20})^2\right\}\Psi_5(\underline{\theta}) \tag{15}$$

where,

$$\Psi_5(\underline{\theta}) = exp\left\{-\frac{c(\delta_2)}{\sigma_2^{2/(1+\delta_2)}}B_2(\alpha_2,\beta_2,\delta_2)\right\}$$

(vi) $$\pi\left(\beta_2 \mid \alpha_1,\beta_1,\sigma_1,\delta_1,\alpha_2,\sigma_2,\delta_2,\gamma,\tau,\underline{z},\underline{y},\underline{x}\right) \propto$$

$$exp\left\{-\frac{1}{2\sigma_{22}^2}(\beta_2-\beta_{20})^2\right\}\Psi_6(\underline{\theta})$$

where,

$$\Psi_6(\underline{\theta}) = exp\left\{-\frac{c(\delta_2)}{\sigma_2^{2/(1+\delta_2)}}B_2(\alpha_2,\beta_2,\delta_2)\right\}$$

(vii) $$\pi\left(\sigma_2 \mid \alpha_1,\beta_1,\delta_1,\sigma_1,\alpha_2,\beta_2,\delta_2,\gamma,\tau,\underline{z},\underline{y},\underline{x}\right) \propto$$

$$\sigma_2^{m_{22}-1}exp\left\{-n_{22}\sigma_2\right\}\Psi_7(\underline{\theta})$$

where,

$$\Psi_7(\underline{\theta}) = exp\left\{-(n-r)ln(\sigma_2) - \sigma_2^{-2/(1+\delta_2)}c(\delta_2)B_2(\alpha_2,\beta_2,\delta_2)\right\}$$

(viii) $\pi\left(\delta_2 \mid \alpha_1,\beta_1,\sigma_1,\delta_1,\alpha_2,\beta_2,\sigma_2,\gamma,\tau,\underline{z},\underline{y},\underline{x}\right) \propto$

$$exp\left\{-\frac{1}{2\sigma_{23}^2}(\delta_2-\delta_{20})^2\right\}\Psi_8(\underline{\theta})$$

where,

$$\Psi_8(\underline{\theta}) = exp\left\{(n-r)ln(w(\delta_2)) - \frac{c(\delta_2)}{\sigma_2^{2/(1+\delta_2)}}B_2(\alpha_2,\beta_2,\delta_2)\right\}$$

(ix) $\pi\left(\gamma \mid \alpha_1,\beta_1,\sigma_1,\delta_1,\alpha_2,\beta_2,\sigma_2,\delta_2,\tau,\underline{z},\underline{y},\underline{x}\right) \propto$

$$exp\left\{-\frac{1}{2\sigma_{14}^2}(\gamma-\gamma_0)^2\right\}\Psi_9(\underline{\theta})$$

where,

$$\Psi_9(\underline{\theta}) = exp\left\{r\gamma - \textstyle\sum_{i=1}^n ln\left(1+e^{\gamma+\tau x_i}\right)\right\}$$

(x) $\pi\left(\tau \mid \alpha_1,\beta_1,\sigma_1,\delta_1,\alpha_2,\beta_2,\sigma_2,\delta_2,\gamma,\underline{z},\underline{y},\underline{x}\right) \propto$

$$exp\left\{-\frac{1}{2\sigma_{24}^2}(\tau-\tau_0)^2\right\}\Psi_{10}(\underline{\theta})$$

where,

$$\Psi_{10}(\underline{\theta}) = exp\left\{a_1\tau - \textstyle\sum_{i=1}^n ln\left(1+e^{\gamma+\tau x_i}\right)\right\}$$

Observe that we need to use the Metropolis-Hastings algorithm to generate $\underline{\theta}$.

2.2 Mixture of Normal Distributions

In the special case of a mixture of two normal distributions ($\delta_1=\delta_2=0$ in the exponential power distributions given in (4)), with the same logistic regression links given in (2), consider the following prior densities for $\alpha_1,\beta_1,\sigma_1,\alpha_2,\beta_2,\sigma_2,\gamma$ and τ:

(i) $\alpha_1,\beta_1,\sigma_1,\alpha_2,\beta_2,\sigma_2$ locally uniform,

(ii) $\gamma \sim N(\gamma_0, \sigma_{14}^2), \gamma_0, \sigma_{14}^2$ known, (16)

(iii) $\tau \sim N(\gamma_0, \sigma_{24}^2), \tau_0, \sigma_{24}^2$ known,

We also assume independence among the parameters.

From (10), we get the joint posterior density for $\underline{z}$ and $\underline{\theta} = (\alpha_1, \beta_1, \sigma_1, \alpha_2, \beta_2, \sigma_2, \gamma, \tau)$,

$$\pi\left(\underline{\theta}|\underline{y}, \underline{z}, \underline{x}\right) \propto \frac{\sigma_1^{-r}\sigma_2^{-(n-r)}}{\{\prod_{i=1}^n (1+e^{\gamma+\tau x_i})\}}$$
$$exp\left\{-\frac{1}{2\sigma_{14}^2}(\gamma-\gamma_0)^2 + \gamma r - \frac{1}{2\sigma_{24}^2}(\tau-\tau_0)^2 + \tau a_1\right\} \quad (17)$$
$$\left\{exp\left\{-\frac{1}{2\sigma_1^2}\sum_{i=1}^n z_{i1}(y_i-\alpha_1-\beta_1 x_i)^2 - \frac{1}{2\sigma_2^2}\sum_{i=1}^n z_{i2}(y_i-\alpha_2-\beta_2 x_i)^2\right\}\right\}$$

where $r, n-r$ and a_1 are defined in (14).

The conditional distributions for the Gibbs sampling algorithm are given by,

(i) $\pi\left(v|\alpha_1, \beta_1, \alpha_2, \beta_2, \sigma_2, \gamma, \tau, \underline{z}, \underline{y}, \underline{x}\right) \sim \Gamma\left(\frac{r}{2}+1, \frac{\sum_{i=1}^n z_{i1}(y_i-\alpha_1-\beta_1 x_i)^2}{2}\right)$

where $v = \sigma_1^{-2}$

(ii) $\pi\left(\alpha_1, |\beta_1, \sigma_1, \alpha_2, \beta_2, \sigma_2, \gamma, \tau, \underline{z}, \underline{y}, \underline{x}\right) \sim \Gamma\left(\frac{\sum_{i=1}^n z_{i1}(y_i-\beta_1 x_i)}{r}, \frac{\sigma_1^2}{r}\right)$

(iii) $\pi\left(\beta_1|\alpha_1, \sigma_1, \alpha_2, \beta_2, \sigma_2, \gamma, \tau, \underline{z}, \underline{y}, \underline{x}\right) \sim$

$$N\left(\frac{\sum_{i=1}^n z_{i1}x_i(y_i-\alpha_1)}{\sum_{i=1}^n z_{i1}x_i^2}, \frac{\sigma_1^2}{\sum_{i=1}^n z_{i1}x_i^2}\right) \quad (18)$$

(iv) $\pi\left(u|\alpha_1, \beta_1, \sigma_1, \alpha_2, \beta_2, \gamma, \tau, \underline{z}, \underline{y}, \underline{x}\right) \sim$

$$N\left(\frac{(n-r)}{2}+1, \frac{\sum_{i=1}^n z_{i2}(y_i-\alpha_2-\beta_2 x_i)^2}{2}\right)$$

where $u = \sigma_2^{-2}$

(v) $\pi\left(\alpha_2|\alpha_1,\beta_1,\sigma_1,\beta_2,\sigma_2,\gamma,\tau,\underline{z},\underline{y},\underline{x}\right) \sim$

$$N\left(\frac{\sum_{i=1}^{n} z_{i2}\left(y_i-\beta_2 x_i\right)}{(n-r)}, \frac{\sigma_2^2}{(n-r)}\right)$$

(vi) $\pi\left(\beta_2|\alpha_1,\beta_1,\sigma_1,\alpha_2,\sigma_2,\gamma,\tau,\underline{z},\underline{y},\underline{x}\right) \sim$

$$N\left(\frac{\sum_{i=1}^{n} z_{i2}x_i\left(y_i-\alpha_2\right)}{\sum_{i=1}^{n} z_{i2}x_i^2}\ ,\ \frac{\sigma_2^2}{\sum_{i=1}^{n} z_{i2}x_i^2}\right)$$

(vii) $\pi\left(\gamma|\alpha_1,\beta_1,\sigma_1,\alpha_2,\beta_2,\sigma_2,\tau,\underline{z},\underline{y},\underline{x}\right) \propto\ exp\left\{-\frac{1}{2\sigma_{14}^2}\left(\gamma-\gamma_0\right)^2\right\}\Psi_1\left(\underline{\theta}\right),$

where $\Psi_1\left(\underline{\theta}\right) = exp\left\{\gamma r - \sum_{i=1}^{n} ln\left(1+e^{\gamma+\tau x_i}\right)\right\}$

(viii) $\pi\left(\tau|\alpha_1,\beta_1,\sigma_1,\alpha_2,\beta_2,\sigma_2,\gamma,\underline{z},\underline{y},\underline{x}\right) \propto\ exp\left\{-\frac{1}{2\sigma_{24}^2}\left(\tau-\tau_0\right)^2\right\}\Psi_2\left(\underline{\theta}\right),$

where $\Psi_2\left(\underline{\theta}\right) = exp\left\{\tau a_1 - \sum_{i=1}^{n} ln\left(1+e^{\gamma+\tau x_i}\right)\right\}.$

Observe that the variables γ and τ should be generated using the Metropolis-Hastings algorithm.

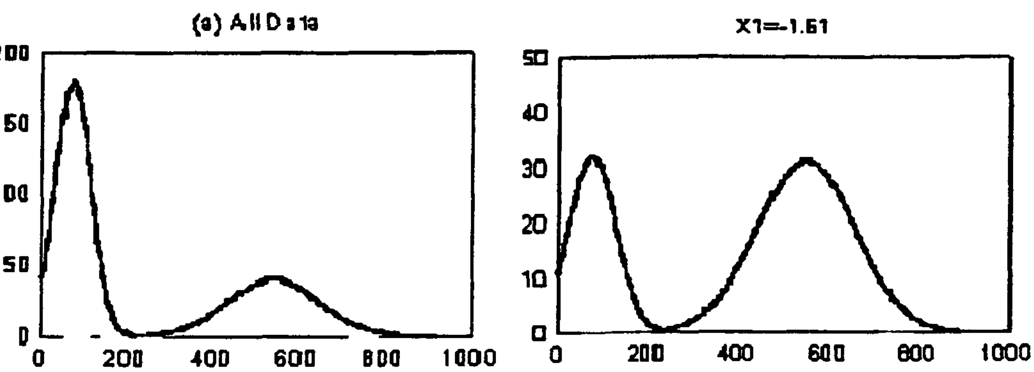
3 A Numerical Illustration

In table 1, we have the lifetimes of $n = 317$ insects receiving four dosages of a toxity. Among the 317 insects, 144, 69, 54 and 50 where sprayed with a insecticide at concentrations of 0.20, 0.32, 0.50 and 0.80 mg/cm^2, respectively. The log-doses (denoted by x) are -1.61, -1.14, -0.69 and -0.22, respectively.

Table 1. Survival times (in hours) of $n = 317$ insects exposed to 4 dosages of a insecticide.

Log-Dosage(x)	Survival Times (t)
$x_1 = -1.61$	12,2(16),5(30),4(36),2(40),3(52),2(60),4(65),70,2(76), 2(80),3(90),2(100),2(110),130,2(140),150,160,180,280 300,20(400),30(500),15(700),6(900)
$x_2 = -1.14$	3(10),2(16),2(20),3(30),3(35),2(40),2(45),4(50),3(56), 2(60),2(65),5(80),3(85),4(90),4(92),2(100),115,130, 160,340,5(400),5(500),4(580),3(600),2(800)
$x_3 = -0.69$	2(10),2(18),20,3(30),2(32),2(40),45,4(50),3(60),2(65),2(68) 5(80),5(85),5(85),3(90),2(92),2(100),2(118),130,140,160,180 340,400,3(500),580,650
$x_4 = -0.22$	2(10),2(18),3(30),3(38),2(40),2(45),2(50),2(60),3(68), 70,4(80),5(86),4(88),90,3(100),3(110),2(118),130,138 160,220,350,400

In figure 1, we have the histograms of all survival data (a), and for each individual dosage $x_{ik}, i = 1, 2, \ldots, n, k = 1, 2, 3, 4$.In these graphs, we clearly observe bimodal frequency distributions, indicating the need of fit of mixture distributions of the form (1).



t $\qquad$ $\bar{t}_{11} = 76.33$ $\qquad$ $\bar{t}_{21} = 560.80$

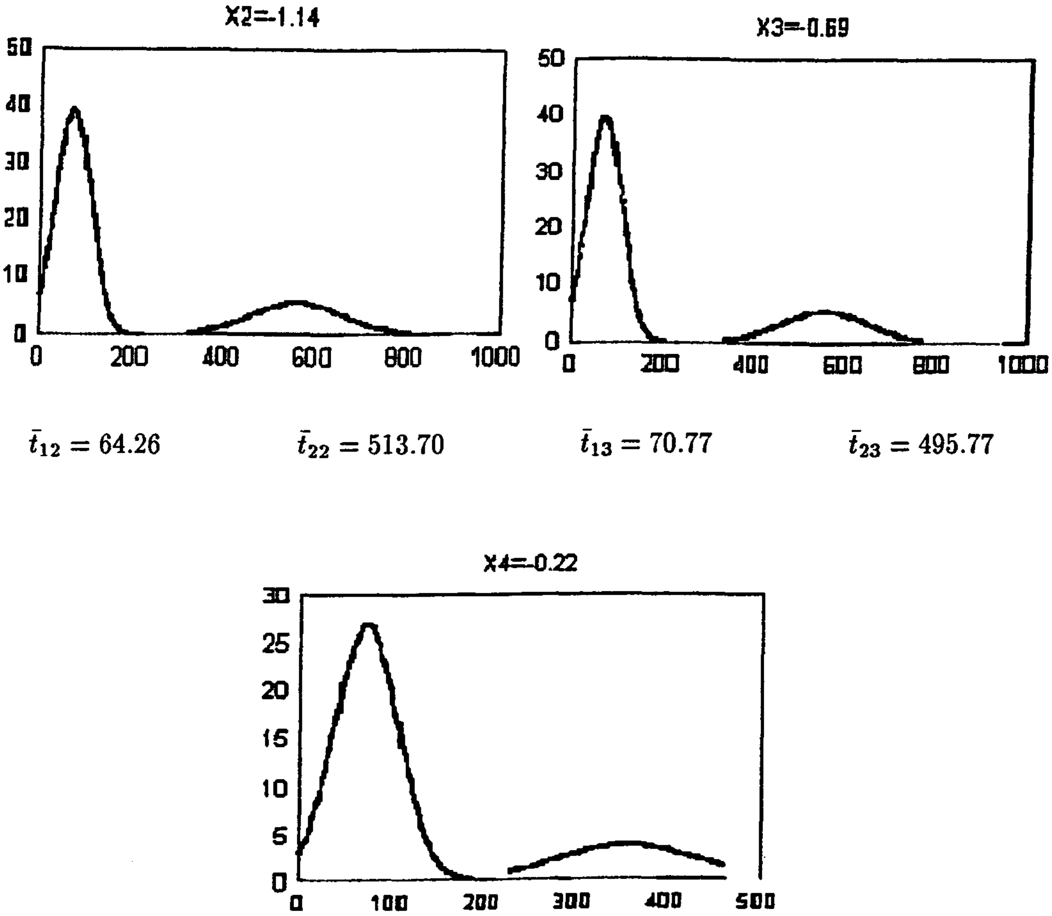


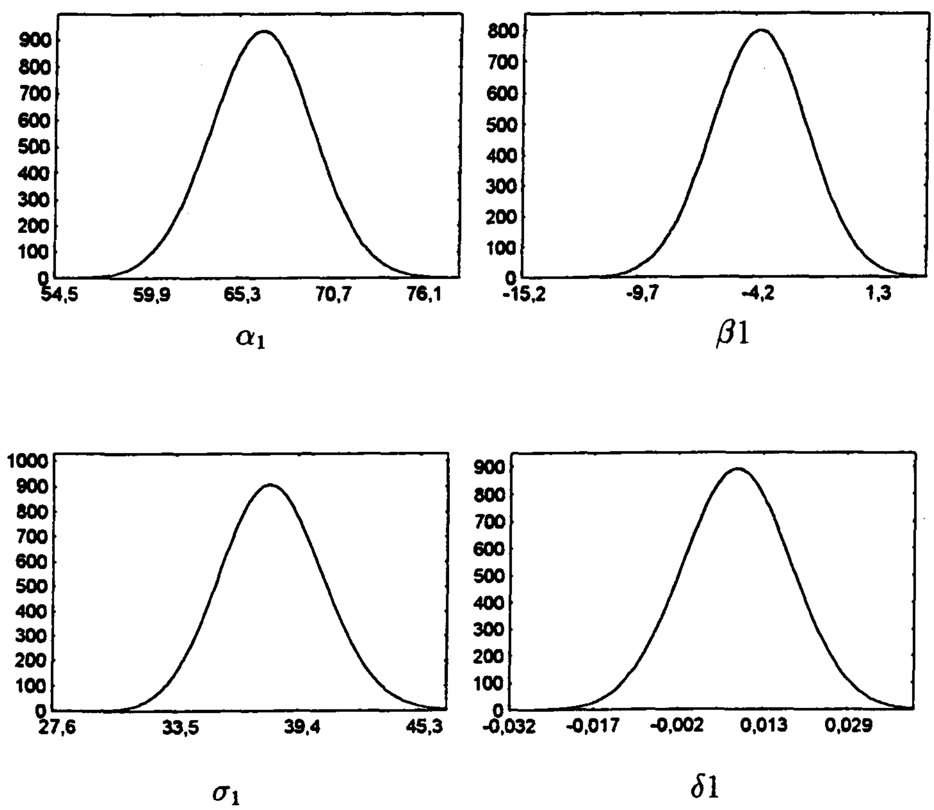
Figure 1. Histograms of Survival Data of Table 1.

To analyse the survival data of table 1, we first assume a mixture of exponential power distributions (4), with a ligistic regression link. Considering the prior densities for $\alpha_1, \beta_1, \sigma_1, \delta_1, \alpha_2, \beta_2, \sigma_2, \delta_2, \gamma$ and τ given in (12) with $\alpha_{10} = 65.4, \sigma_{11}^2 = 14.1, \beta_{10} = -3.5, \sigma_{12}^2 = 7.7, \delta_{10} = 0, \sigma_{13}^2 = 0.01, \delta_{20} = 0, \sigma_{23}^2 = 0.01, m_{11} = 180, n_{11} = 5.1, m_{22} = 137, n_{22} = 1.0, \alpha_{20} = 351.3, \sigma_{21}^2 = 3611.9, \beta_{20} = -126.8, \sigma_{22}^2 = 1573.4, \gamma_0 = 4.0, \sigma_{14}^2 = 2.5, \tau_0 = 3.0$ and $\sigma_{24}^2 = 2.3$ (the choice of these values for the parameters of the prior densities was based on expert opinion combined with a preliminary analysis of the data), we generated 3 separate Gibbs chains each of wich ran for 15.000 iteration, and we monitored the convergence of the Gibbs samples using the Gelman and Rubin (1992) method that uses the analysis of variance technique to determine if further iterations are needed. For each parameter, we considered the $15^{th}, 30^{th}, 45^{th}, \ldots$ iterations, which required a computational time of 7 hours working with the software

SAS in a Pentium 166 MHZ. In table 2, we have the obtained posterior summaries for the parameters, and in figure 2 we have the approximate marginal posterior densities considering the $S = 1000$ Gibbs samples. We also have in table 2, the estimated potential scale reductions $\hat{R}$ (see Gelman and Rubin, 1992) for all the parameters. In this case, the considered number of iterations were sufficient for approximate convergence ($\sqrt{\hat{R}} < 1.1$ for all parameters).

Table 2. Posterior Summaries (mixture of two exponential power distribution)

Parameter	Mean	95% Credible Interval	$\hat{R}$
α_1	66.65000	(60.42 ; 72.61)	1.000711
β_1	-3.99000	(-8.87 ; -0.23)	1.000117
σ_1	38.21000	(32.74 ; 42.83)	1.003883
δ_1	0.00870	(-0.00106 ; 0.028)	1.000158
α_2	337.73000	(249.16 ; 429.11)	1.001783
β_2	-124.12000	(-182.2 ; -64.82)	1.001530
σ_2	148.84000	(125.66 ; 169.58)	1.000429
δ_2	0.00710	(-0.0123 ; 0.027)	1.000107
γ	4.02000	(1.53 ; 6.41)	1.001148
τ	3.11000	(1.04 ; 5.20)	1.001449



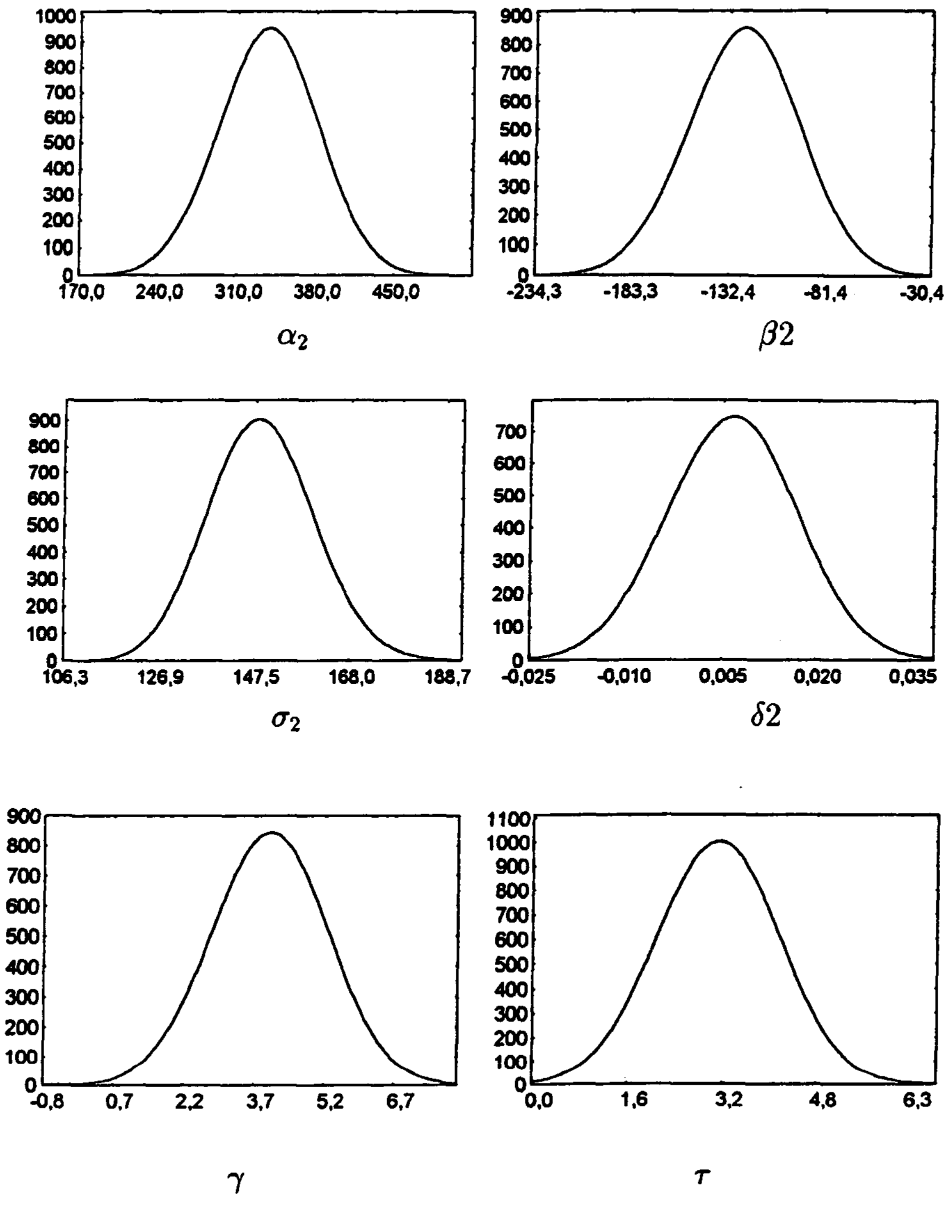


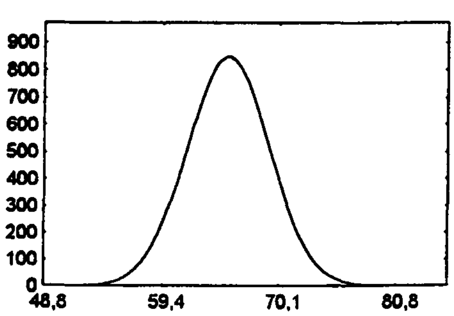
Figure 2. Approximate marginal posterior densities (mixture of two exponential power distributions).

From the observed results of table 2 (see also figure 2), we could consider a mixture of two normal distributions to analyse the data of table 1, since the Bayes estimators for δ_1 and δ_2 based on Monte Carlo approximations of the posterior means are close to zero (also observe that both 95 % credible intervals for δ_1 and

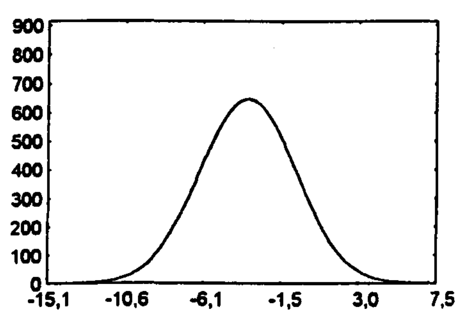
δ_2 includes zero). Thus, considering a normal-normal mixture model, and the prior densities (16) with $\gamma_0 = 4.0, \sigma_{14}^2 = 0.60, \tau_0 = 3.0$ and $\sigma_{24}^2 = 0.4$, we also generated 3 separate Gibbs chains each of which ran for 26000 iterations, considering the $26^{th}, 52^{th}, 78^{th}, \ldots$ iterations, which required a computational time of 6 hours working with the software SAS in a Pentium 166 MHZ. In table 3, we have the obtained posterior summaries and in figure 3, we have the approximate marginal posterior densities considering the $S = 3000$ Gibbs samples. We also observe approximate convergence, since the estimated potential scale reductions introduced by Gelman and Rubin (1992) are close to one for all parameters.

Table 3. Posterior Summaries (Normal-Normal Distribution).

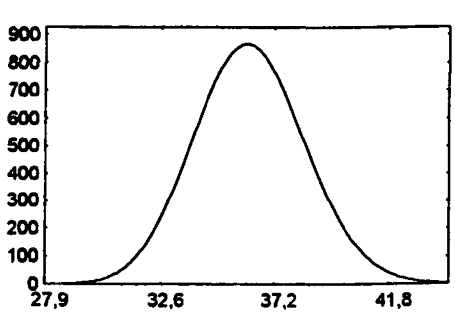
Parameter	Mean	95% Credible Interval	$\hat{R}$
α_1	65.447	(57.248 ; 72.217)	1.00006700
β_1	-3.465	(-10.590 ; -0.142)	1.00002900
σ_1	36.181	(31.993 ; 40.891)	1.00246600
α_2	351.260	(235.177 ; 468.857)	1.00353600
β_2	-126.806	(-203.062 ; -50.101)	1.00398100
σ_2	135.419	(117.943 ; 158.697)	1.00196700
γ	4.020	(2.620 ; 5.422)	1.00001800
τ	3.021	(1.923 ; 4.085)	1.00342600



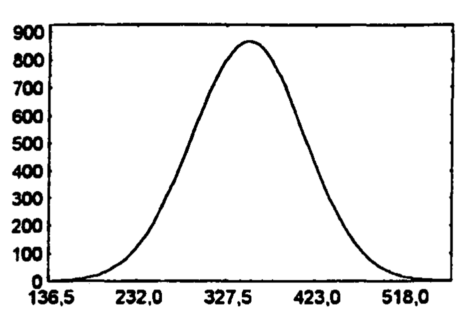
α_1



$\beta 1$



σ_1



$\alpha 2$

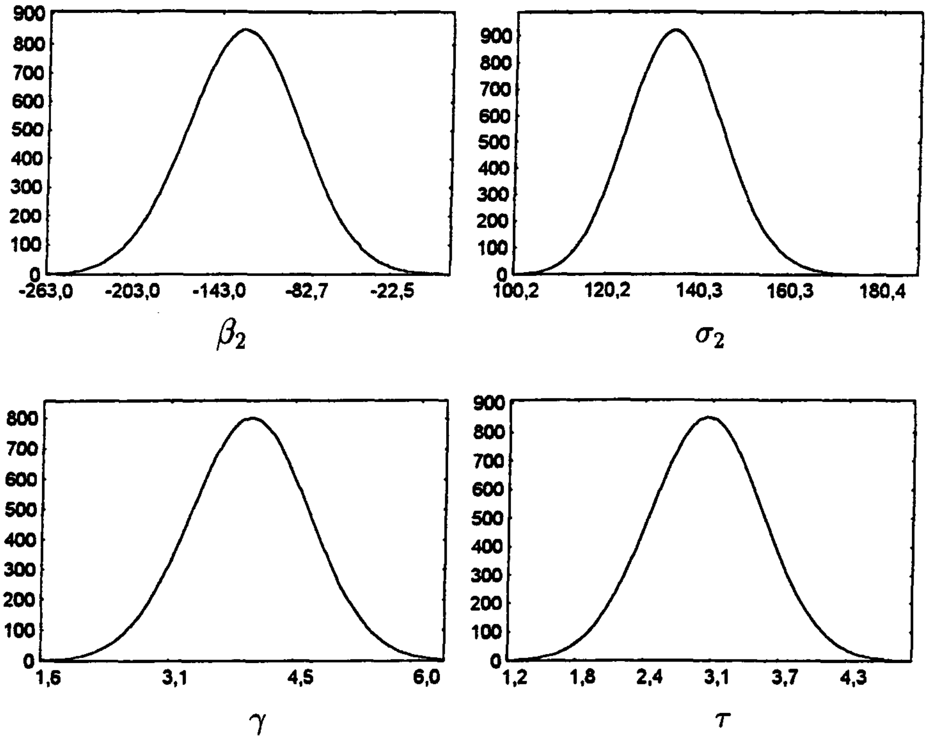


Figure 3. Approximate marginal posterior densities (normal-normal distribution).

We observe that the inference results for the parameters $\alpha_1, \beta_1, \sigma_1, \alpha_2, \beta 2, \sigma_2, \gamma$ and τ are very similar, considering both mixture models. For the lifetime data of table 1, it is reasonable to consider a mixture of two normal distributions.

4 Concluding Remarks

The use of mixture of exponential power distributions in the presence of one or more covariates is a suitable way to analyse data in many applications. Usually, a preliminary data analysis indicates that the standard parametrical models commonly used to analyse data could not be appropriate, as it was seen for the lifetime data of table 1. Classical inference approach for this model usually is difficult and the obtained results could be not accurate, especially for small or moderate sample sizes. The use of Bayesian methods considering MCMC methods is a suitable way to analyse this family of models.

References

Box, G.E.P.; Tiao, G.C.(1973). Bayesian inference in statistical analysis, Addison-Wesley Publishing Company.

Diebolt, J.; Robert, C. (1994). Estimation of finite mixture distributions through Bayesian sampling. *Journal of the Royal Statistical Society*, Series B, 56, 363-375.

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

Gelman, A.; Rubin, D.B. (1992) . Inference from iterative simulation using multiple sequences (with discussion). *Statistical Science*, 7,457-511.

Kuo, L.; Peng, F. (1995) . A mixture-model approach to the analysis of survival data. Technical Report Number 95-31, dep. of Statistics, University of Connecticut, Storrs.

Smith, A.F.M.; Roberts, G.O. (1993) . Bayesian computation via the Gibbs sampler and related Markov Chain Monte Carlo methods, *Journal of the Royal Statistical Society*, B, 55, 3-24.

Titterington, D.M.; Smith, A.F.M.; and Makov, U.E. (1985) . Statistical Analysis of Finite Mixture Distributions. New York: John Wiley.

NOTAS DO ICMSC

SÉRIE ESTATÍSTICA