UNIVERSIDADE DE SÃO PAULO

BAYESIAN ANALYSIS OF MIXTURE FOR SURVIVAL DATA: SOME COMPUTATIONAL ASPECTS

JORGE ALBERTO ACHCAR
GILBERTO DE ARAÚJO PEREIRA

Nº 44

NOTAS

Instituto de Ciências Matemáticas de São Carlos

Instituto de Ciências Matemáticas de São Carlos

ISSN - 0103-2577

BAYESIAN ANALYSIS OF MIXTURE FOR SURVIVAL DATA: SOME COMPUTATIONAL ASPECTS

JORGE ALBERTO ACHCAR
GILBERTO DE ARAÚJO PEREIRA

Nº 44

NOTAS DO ICMSC
Série Estatística

São Carlos
Out./1997

BAYESIAN ANALYSIS OF MIXTURE MODELS FOR SURVIVAL DATA:SOME COMPUTATIONAL ASPECTS

Jorge Alberto Achcar*
Gilberto de Araújo Pereira*

■ ABSTRACT: In this paper, we present a Bayesian analysis of mixture models for survival data in the presence of one covariate. Considering Gibbs with Metropolis-Hastings algorithms, we get Monte Carlo estimates for the posterior quantities of interest, assuming different choices for the densities in the mixture model. We also present some considerations on model selection and we introduce a numerical example, to illustrate the proposed methodology.

■ KEYWORDS: Mixture models, regression, 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 lifetime data, since in many applications the usual parametrical survival models could not be appropriate for some data sets (see for example, Farewell, 1982; or Kuo and Peng, 1995). These data could be observed when a group of subjects may not react to treatments. A generalization of these mixture models for survival data is given when there is one or more covariates which may influence both the incidence probabilities and the conditional latency distributions (see for example, Kuo and Peng, 1995).

Let $\underset{\sim}{x}$ denote the covariate vector associated with a subject of lifetime T. Let Y be an index variable for the subpopulations. The mixture model assumes the density

$$f(t \mid \underset{\sim}{x}, \underset{\sim}{\theta}) = \sum_{j=1}^{J} P(Y = j \mid \underset{\sim}{x}, \underset{\sim}{\gamma}) \; f_j(t \mid Y = j, \underset{\sim}{x}, \underset{\sim}{\beta_j}) \tag{1}$$

where $\underset{\sim}{\theta} = (\underset{\sim}{\beta_1}, \underset{\sim}{\beta_2}, \ldots, \underset{\sim}{\beta_J}, \ldots, \underset{\sim}{\gamma})$ is the collection of all unknown parameters.

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

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

* ICMSC, USP, C. Postal 668, 13560-970 - São Carlos - SP.

$$h_{i1} = \frac{P(1 \mid x_i, \underset{\sim}{\gamma})\ f_1(t_i \mid x_i, \underset{\sim}{\beta_1})}{\sum_{j=1}^{2} P(j \mid x_i, \underset{\sim}{\gamma})\ f_j(t_i \mid x_i, \underset{\sim}{\beta_j})} \tag{6}$$

where $i = 1, 2, ..., n$.

That is,

$$\pi(\underset{\sim}{z_i}) \propto h_{i1}^{z_{i1}} (1 - h_{i1})^{z_{i2}} \tag{7}$$

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

Thus,

$$\pi(\underset{\sim}{z_1}, ..., \underset{\sim}{z_n}) \propto \frac{\prod_{i=1}^{n} \prod_{j=1}^{2} \left\{ P(j \mid x_i, \underset{\sim}{\gamma})\ f_j(t_i \mid x_i, \underset{\sim}{\beta_j}) \right\}^{z_{ij}}}{\prod_{i=1}^{n} \left\{ \sum_{j=1}^{2} P(j \mid x_i, \underset{\sim}{\gamma})\ f_j(t_i \mid x_i, \underset{\sim}{\beta_j}) \right\}} \tag{8}$$

Combining (8) with (5), we get the joint posterior distribution for $\underset{\sim}{z_1}, ..., z_n, \underset{\sim}{\theta}$,

$$\pi(\underset{\sim}{z_1}, ..., \underset{\sim}{z_n}, \underset{\sim}{\theta} \setminus \underset{\sim}{t}, \underset{\sim}{x}) \propto \pi(\underset{\sim}{\theta}) \left\{ \prod_{i=1}^{n} \prod_{j=1}^{2} \left\{ P(j \mid x_i, \underset{\sim}{\gamma})\ f_j(t_i \mid x_i, \underset{\sim}{\beta_j}) \right\}^{z_{ij}} \right\} \tag{9}$$

To generate samples of the joint posterior distribution (9), we use the Gibbs sampling algorithm. Starting with initial values $\underset{\sim}{\theta}^{(0)} = (\theta_1^{(0)}, \ldots, \theta_p^{(0)})$, follow the steps:

(i) Generate a sample $\underset{\sim}{z}^{(1)} = (\underset{\sim}{z}_1^{(1)}, \ldots, \underset{\sim}{z}_n^{(1)})$ from (7).
(ii) Generate a sample of $\underset{\sim}{\theta}$, from the conditional distribuitions
$\pi(\theta_1 \mid \theta_2^{(0)}, ..., \theta_p^{(0)}, \underset{\sim}{z}^{(1)}, \underset{\sim}{t}, \underset{\sim}{x})$, $\pi(\theta_2 \mid \theta_1^{(0)}, \theta_3^{(0)} \ldots, \theta_p^{(0)}, \underset{\sim}{z}^{(1)}, \underset{\sim}{t}, \underset{\sim}{x}), \ldots,$
$\pi(\theta_p \mid \theta_1^{(1)}, \ldots, \theta_{p-1}^{(1)}, \underset{\sim}{z}^{(1)}, \underset{\sim}{t}, \underset{\sim}{x})$.

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

Now, we introduce some special cases for the mixture model (1) assuming one covariate variable x and uncensored observations.

2.1. A normal - exponential mixture model

Let us assume a mixture of normal- exponential distributions in (1) , one covariate x and the logistic regression link (2) ; that is,

$$f_1(t_i \mid x_i, \underset{\sim}{\beta_1}) = \frac{1}{\sqrt{2\pi}\ \sigma} exp\{ -\frac{1}{2\sigma^2} (t_i - \alpha_1 - \beta_1 x_i)^2 \},$$

$$\underset{\sim}{\beta_1} = (\alpha_1, \beta_1, \sigma), \quad f_2\left(t_i \mid x_i, \underset{\sim}{\beta_2}\right) = \lambda_i \, exp\{-\lambda_i t_i\},$$

where $\lambda_i = (\alpha_2 + \beta_2 x_i)^{-1}$, $\underset{\sim}{\beta_2} = (\alpha_2, \beta_2)$,

$$P(1 \mid x_i, \underset{\sim}{\gamma}) = \frac{e^{\gamma + \tau x_i}}{1 + e^{\gamma + \tau x_i}}, \quad \underset{\sim}{\gamma} = (\gamma, \tau) \text{ and}$$
$$P(2 \mid x_i, \underset{\sim}{\gamma}) = 1 - P(1 \mid x_i, \underset{\sim}{\gamma}) = \frac{1}{1 + e^{\gamma + \tau x_i}}, \quad i = 1, 2, \ldots, n.$$

Assuming prior independence among the parameters, consider the following prior densities for $\alpha_1, \beta_1, \sigma, \alpha_2, \beta_2, \gamma, \tau$:

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

(ii) $\gamma \sim N(\gamma_0, s_1^2)$, γ_0, s_1^2 known, (12)

(iii) $\tau \sim N(\tau_0, s_2^2)$, τ_0, s_2^2 known

where $N(a,b)$ denotes a normal distribution with mean a and variance b.

The joint posterior density (9) for $\underset{\sim}{z}$ and $\underset{\sim}{\theta} = (\alpha_1, \beta_1, \sigma, \alpha_2, \beta_2, \gamma, \tau)$ is given by,

$$\pi(\underset{\sim}{z}, \underset{\sim}{\theta} \mid \underset{\sim}{t}, \underset{\sim}{x}) \propto \frac{\left\{\prod_{i=1}^{n} (\alpha_2 + \beta_2 x_i)^{-z_{i2}}\right\}}{\left\{\prod_{i=1}^{n} (1 + e^{\gamma + \tau x_i})\right\}} \frac{e^{\gamma r + \tau a_1}}{\sigma^r}$$
$$exp\left\{-\frac{1}{2 s_1^2}(\gamma - \gamma_0)^2 - \frac{1}{2 s_2^2}(\tau - \tau_0)^2\right\} \tag{13}$$
$$\left\{exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n} z_{i1} (t_i - \alpha_1 - \beta_1 x_i)^2 - \sum_{i=1}^{n} z_{i2} t_i (\alpha_2 + \beta_2 x_i)^{-1}\right\}\right\}$$

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

To generate samples of the joint posterior distribution (13), we use steps (i) and (ii) of the Gibbs algorithm (10), where the conditional distributions for the parameters are given by,

(i) $\pi(v \mid \alpha_1, \beta_1, \alpha_2, \beta_2, \gamma, \tau, \underset{\sim}{z}, \underset{\sim}{t}, x) \sim \Gamma\left(\frac{r}{2} + 1, \frac{\sum_{i=1}^{n} z_{i1} (t_i - \alpha_1 - \beta_1 x_i)^2}{2}\right)$

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

(ii) $\pi(\alpha_1 \mid \beta_1, \sigma, \alpha_2, \beta_2, \gamma, \tau, \underset{\sim}{z}, \underset{\sim}{t}, x) \sim N\left(\frac{\sum_{i=1}^{n} z_{i1} (t_i - \beta_1 x_i)}{r}, \frac{\sigma^2}{r}\right)$

$$\text{(iii)}\pi(\beta_1 \mid \alpha_1, \sigma, \alpha_2, \beta_2, \gamma, \tau, \underset{\sim}{z}, \underset{\sim}{t}, x) \sim N\left(\frac{\sum_{i=1}^{n} z_{i1}x_i(t_i-\alpha_1)}{\sum_{i=1}^{n} z_{i1}x_i^2}, \frac{\sigma^2}{\sum_{i=1}^{n} z_{i1}x_i^2}\right) \qquad (14)$$

$$\text{(iv)}\pi(\alpha_2 \mid \alpha_1,\beta_1,\sigma,\beta_2,\gamma,\tau,\underset{\sim}{z}, \underset{\sim}{t}, x) \propto \alpha_2^{-(n-r)} exp\left\{-\frac{1}{\alpha_2}\sum_{i=1}^{n} z_{i2}t_i(1+\frac{\beta_2}{\alpha_2}x_i)^{-1}\right\}\Psi_1(\underset{\sim}{\theta}),$$

where $\Psi_1(\underset{\sim}{\theta}) = \prod_{i=1}^{n}(1+\frac{\beta_2}{\alpha_2}x_i)^{-z_{i2}}$,

$$\text{(v)}\ \pi(\beta_2 \mid \alpha_1,\beta_1,\sigma,\alpha_2,\gamma,\tau,\underset{\sim}{z}, \underset{\sim}{t}, x) \propto \beta_2^{-(n-r)} exp\left\{-\frac{1}{\beta_2}\sum_{i=1}^{n} z_{i2}t_i(\frac{\alpha_2}{\beta_2}+x_i)^{-1}\right\}\Psi_2(\underset{\sim}{\theta}),$$

where $\Psi_2(\underset{\sim}{\theta}) = \prod_{i=1}^{n}(\frac{\alpha_2}{\beta_2}+x_i)^{-z_{i2}}$

$$\text{(vi)}\ \pi(\gamma \mid \alpha_1,\beta_1,\sigma,\alpha_2,\beta_2,\tau,\underset{\sim}{z}, \underset{\sim}{t}, x) \propto exp\left\{-\frac{1}{2\,s_1^2}(\gamma-\gamma_0)^2\right\}\Psi_3(\underset{\sim}{\theta}),$$

where $\Psi_3(\underset{\sim}{\theta}) = exp\left\{\gamma r - \sum_{i=1}^{n} ln(1+e^{\gamma+\tau x_i})\right\}$

$$\text{(vii)}\ \pi(\tau \mid \alpha_1,\beta_1,\sigma,\alpha_2,\beta_2,\gamma,\underset{\sim}{z}, \underset{\sim}{t}, x) \propto exp\left\{-\frac{1}{2\,s_2^2}(\tau-\tau_0)^2\right\}\Psi_4(\underset{\sim}{\theta}),$$

where $\Psi_4(\underset{\sim}{\theta}) = exp\left\{\tau a_1 - \sum_{i=1}^{n} ln(1+e^{\gamma+\tau x_i})\right\}$.

Here, $\Gamma(a,b)$ denotes a gamma distribution with mean $\frac{a}{b}$ and variance $\frac{a}{b^2}$.

Observe that, we need to use the Metropolis-Hastings algorithm to generate the variables α_2, β_2, γ and τ. In this way, we could generate candidates of inverse gamma distributions for the variables α_2 and β_2, and generate candidates for the variables γ and τ from the normal distributions $N(\gamma_0,\ s_1^2)$ and $N(\tau_0,\ s_2^2)$, respectively.

2.2. A normal-normal mixture model

Considering now the normal-normal mixture in (1), with densities

$f_1(t_i \mid x_i, \underset{\sim}{\beta_1}) = \frac{1}{\sqrt{2\pi}\,\sigma_1} exp\left\{-\frac{1}{2\sigma_1^2}(t_i-\alpha_1-\beta_1 x_i)^2\right\}$,
$\underset{\sim}{\beta_1} = (\alpha_1, \beta_1, \sigma_1)$, and
$f_2(t_i \mid x_i, \underset{\sim}{\beta_2}) = \frac{1}{\sqrt{2\pi}\,\sigma_2} exp\left\{-\frac{1}{2\sigma_2^2}(t_i-\alpha_2-\beta_2 x_i)^2\right\}$,
$\underset{\sim}{\beta_2} = (\alpha_2, \beta_2, \sigma_2)$ and the same logistic regression links given in (11), assume the following prior densities for α_1, β_1, σ_1, α_2, β_2, σ_2, γ and τ:

(i) α_1, β_1, σ_1, α_2, β_2, σ_2 locally uniform,

(ii) $\gamma \sim N(\gamma_0,\ s_1^2)$, γ_0, s_1^2 known, (15)

(iii) $\tau \sim N(\tau_0,\ s_2^2)$, τ_0, s_2^2 known

We also assume independence among the parameters.

From (9), we obtain the joint posterior density for $\underset{\sim}{z}$ and $\underset{\sim}{\theta} = (\alpha_1, \beta_1, \sigma_1, \alpha_2, \beta_2, \sigma_2, \gamma, \tau)$:

$$\pi(\underset{\sim}{z}, \underset{\sim}{\theta} \mid \underset{\sim}{t}, \underset{\sim}{x}) \propto \frac{\sigma_1^{-r}\,\sigma_2^{-(n-r)}}{\{\prod_{i=1}^{n}(1+e^{\gamma+\tau x_i})\}}$$
$$exp\left\{-\frac{1}{2\,s_1^2}(\gamma-\gamma_0)^2 + \gamma r - \frac{1}{2\,s_2^2}(\tau-\tau_0)^2 + \tau a_1\right\}$$
$$\left\{exp\left\{-\frac{1}{2\sigma_1^2}\sum_{i=1}^{n} z_{i1}\,(t_i-\alpha_1-\beta_1 x_i)^2 - \frac{1}{2\sigma_2^2}\sum_{i=1}^{n} z_{i2}\,(t_i-\alpha_1-\beta_1 x_i)^2\right\}\right\} \tag{16}$$

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

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

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

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

(ii) $\pi(\alpha_1 \mid \beta_1, \sigma_1, \alpha_2, \beta_2, \sigma_2, \gamma, \tau, \underset{\sim}{z}, \underset{\sim}{t}, x) \sim N\left(\frac{\sum_{i=1}^{n} z_{i1}(t_i-\beta_1 x_i)}{r},\ \frac{\sigma_1^2}{r}\right)$

(iii) $\pi(\beta_1 \mid \alpha_1, \sigma_1, \alpha_2, \beta_2, \sigma_2, \gamma, \tau, \underset{\sim}{z}, \underset{\sim}{t}, x) \sim N\left(\frac{\sum_{i=1}^{n} z_{i1} x_i(t_i-\alpha_1)}{\sum_{i=1}^{n} z_{i1} x_i^2},\ \frac{\sigma^2}{\sum_{i=1}^{n} z_{i1} x_i^2}\right)$ (17)

(iv) $\pi(u \mid \alpha_1, \beta_1, \sigma_1, \alpha_2, \beta_2, \gamma, \tau, \underset{\sim}{z}, \underset{\sim}{t}, x) \sim \Gamma\left(\frac{(n-r)}{2}+1,\ \frac{\sum_{i=1}^{n} z_{i2}\,(t_i-\alpha_2-\beta_2 x_i)^2}{2}\right)$

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

(v) $\pi(\alpha_2 \mid \alpha_1, \beta_1, \sigma_1, \beta_2, \sigma_2, \gamma, \tau, \underset{\sim}{z}, \underset{\sim}{t}, x) \sim N\left(\frac{\sum_{i=1}^{n} z_{i2}(t_i-\beta_2 x_i)}{(n-r)},\ \frac{\sigma_2^2}{(n-r)}\right)$

(v) $\pi(\beta_2 \mid \alpha_1, \beta_1, \sigma_1, \alpha_2, \sigma_2, \gamma, \tau, \underset{\sim}{z}, \underset{\sim}{t}, x) \sim N\left(\frac{\sum_{i=1}^{n} z_{i2} x_i (t_i - \alpha_2)}{\sum_{i=1}^{n} z_{i2} x_i^2}, \frac{\sigma^2}{\sum_{i=1}^{n} z_{i2} x_i^2} \right)$

(vi) $\pi(\gamma \mid \alpha_1, \beta_1, \sigma_1, \alpha_2, \beta_2, \sigma_2, \tau, \underset{\sim}{z}, \underset{\sim}{t}, x) \propto exp\left\{ -\frac{1}{2\, s_1^2}(\gamma - \gamma_0)^2 \right\} \Psi_1(\underset{\sim}{\theta}),$

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

(vii) $\pi(\tau \mid \alpha_1, \beta_1, \sigma_1, \alpha_2, \beta_2, \sigma_2, \gamma, \underset{\sim}{z}, \underset{\sim}{t}, x) \propto exp\left\{ -\frac{1}{2\, s_2^2}(\tau - \tau_0)^2 \right\} \Psi_2(\underset{\sim}{\theta}),$

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

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

2.3. A gamma-normal mixture model

Now, consider the gamma-normal mixture in (1), with densities,

$f_1(t_i \mid x_i, \underset{\sim}{\beta_1}) = \frac{1}{\Gamma(\alpha_0)} (\alpha_1 + \beta_1 x_i)^{\alpha_0} t_i^{\alpha_0 - 1} exp\{ -(\alpha_1 + \beta_1 x_i) t_i \},$
$\underset{\sim}{\beta_1} = (\alpha_0, \alpha_1, \beta_1)$, and
$f_2(t_i \mid x_i, \underset{\sim}{\beta_2}) = \frac{1}{\sqrt{2\pi}\, \sigma} exp\{ -\frac{1}{2\sigma^2} (t_i - \alpha_2 - \beta_2 x_i)^2 \},$

$\underset{\sim}{\beta_2} = (\alpha_2, \beta_2, \sigma)$ and the same logistic regression links given in (11). Also, assume the following prior densities for $\alpha_0, \alpha_1, \beta_1, \alpha_2, \beta_2, \sigma, \gamma$ and τ,

(i) $\alpha_0, \alpha_1, \beta_1, \alpha_2, \beta_2, \sigma$ locally uniform,

(ii) $\gamma \sim N(\gamma_0, s_1^2), \gamma_0, s_1^2$ known, (18)

(iii) $\tau \sim N(\tau_0, s_2^2), \tau_0, s_2^2$ known

We further assume independence among the parameters.

From (9), the joint posterior distribution for $\underset{\sim}{z}$ and $\underset{\sim}{\theta} = (\alpha_0, \alpha_1, \beta_1, \alpha_2, \beta_2, \sigma, \gamma, \tau)$ is given by,

$$\pi(\underset{\sim}{z}, \underset{\sim}{\theta} \mid \underset{\sim}{t}, \underset{\sim}{x}) \propto \frac{e^{\gamma r + \tau a_1}}{\sigma^{(n-r)} \{\Gamma(\alpha_0)^r\}} \left\{ \frac{1}{\{ \prod_{i=1}^{n} (1 + e^{\gamma + \tau x_i}) \}} \right\}$$
$$exp\left\{ -\frac{1}{2\, s_1^2}(\gamma - \gamma_0)^2 - \frac{1}{2\, s_2^2}(\tau - \tau_0)^2 \right\}$$

$$\left\{\prod_{i=1}^{n}(\alpha_1+\beta_1 x_i)^{\alpha_0 z_{i1}}\right\}\left\{\prod_{i=1}^{n} t_i^{z_{i1}(\alpha_0-1)}\right\}$$

$$exp\left\{-\sum_{i=1}^{n} z_{i1} t_i(\alpha_1+\beta_1 x_i)^2\right\} exp\left\{-\sum_{i=1}^{n} z_{i2}(t_i-\alpha_2-\beta_2 x_i)^2\right\} \tag{19}$$

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

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

(i) $\pi(\alpha_0 \mid \alpha_1, \beta_1, \alpha_2, \beta_2, \sigma, \gamma, \tau, \underset{\sim}{z}, \underset{\sim}{t}, x) \propto \alpha_0^{\sum_{i=1}^{n} z_{i1}} exp\left\{-\alpha_0 \sum_{i=1}^{n} z_{i1} x_i\right\} \Psi_1(\underset{\sim}{\theta})$,

where $\Psi_1(\underset{\sim}{\theta}) = \dfrac{\left\{\prod_{i=1}^{n}(\alpha_1+\beta_1 x_i)^{\alpha_0 z_{i1}}\right\}\left\{\prod_{i=1}^{n} t_i^{z_{i1}(\alpha_0-1)}\right\}}{\alpha_0^{\sum_{i=1}^{n} z_{i1}} exp\left\{-\alpha_0 \sum_{i=1}^{n} z_{i1} x_i\right\}\{\Gamma(\alpha_0)\}^r}$

(ii) $\pi(\alpha_1 \mid \alpha_0, \beta_1, \alpha_2, \beta_2, \sigma, \gamma, \tau, \underset{\sim}{z}, \underset{\sim}{t}, x) \propto \alpha_1^{\alpha_0 \sum_{i=1}^{n} z_{i1}} exp\left\{-\alpha_1 \sum_{i=1}^{n} z_{i1} t_i (1+\frac{\beta_1}{\alpha_1} x_i)\right\} \Psi_2(\underset{\sim}{\theta})$,

where $\Psi_2(\underset{\sim}{\theta}) = \prod_{i=1}^{n}\left(1+\frac{\beta_1}{\alpha_1} x_i\right)^{\alpha_0 z_{i1}}$

(iii) $\pi(\beta_1 \mid \alpha_0, \alpha_1, \alpha_2, \beta_2, \sigma, \gamma, \tau, \underset{\sim}{z}, \underset{\sim}{t}, x) \propto \beta_1^{\alpha_0 \sum_{i=1}^{n} z_{i1}} exp\left\{-\beta_1 \sum_{i=1}^{n} z_{i1} t_i (\frac{\alpha_1}{\beta_1} + x_i)\right\} \Psi_3(\underset{\sim}{\theta})$,

where $\Psi_3(\underset{\sim}{\theta}) = \prod_{i=1}^{n}\left(\frac{\alpha_1}{\beta_1}+x_i\right)^{\alpha_0 z_{i1}}$ (20)

(iv) $\pi(v \mid \alpha_0, \alpha_1, \beta_1, \alpha_2, \beta_2, \gamma, \tau, \underset{\sim}{z}, \underset{\sim}{t}, x) \sim \Gamma\left(\frac{(n-r)}{2}+1, \frac{\sum_{i=1}^{n} z_{i2}(t_i-\alpha_2-\beta_2 x_i)^2}{2}\right)$

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

(v) $\pi(\alpha_2 \mid \alpha_0, \alpha_1, \beta_1, \beta_2, \sigma, \gamma, \tau, \underset{\sim}{z}, \underset{\sim}{t}, x) \sim N\left(\frac{\sum_{i=1}^{n} z_{i2}(t_i-\beta_2 x_i)}{(n-r)}, \frac{\sigma_2^2}{(n-r)}\right)$

(vi) $\pi(\beta_2 \mid \alpha_0, \alpha_1, \beta_1, \alpha_2, \sigma, \gamma, \tau, \underset{\sim}{z}, \underset{\sim}{t}, x) \sim N\left(\frac{\sum_{i=1}^{n} z_{i2} x_i (t_i-\alpha_2)}{\sum_{i=1}^{n} z_{i2} x_i^2}, \frac{\sigma^2}{\sum_{i=1}^{n} z_{i2} x_i^2}\right)$

(vii) $\pi(\gamma \mid \alpha_0, \alpha_1, \beta_1, \sigma_1, \alpha_2, \beta_2, \sigma, \tau, \underset{\sim}{z}, \underset{\sim}{t}, x) \propto exp\left\{-\frac{1}{2 s_1^2}(\gamma - \gamma_0)^2\right\}\Psi_4(\underset{\sim}{\theta})$,

where $\Psi_4(\underset{\sim}{\theta}) = exp\left\{\gamma r - \sum_{i=1}^{n} ln(1 + e^{\gamma + \tau x_i})\right\}$

(vii) $\pi(\tau \mid \alpha_0, \alpha_1, \beta_1, \sigma_1, \alpha_2, \beta_2, \sigma_2, \gamma, \underset{\sim}{z}, \underset{\sim}{t}, x) \propto exp\left\{-\frac{1}{2 s_2^2}(\tau - \tau_0)^2\right\}\Psi_5(\underset{\sim}{\theta})$,

where $\Psi_5(\underset{\sim}{\theta}) = exp\left\{\tau a_1 - \sum_{i=1}^{n} ln(1 + e^{\gamma + \tau x_i})\right\}$.

Observe that the variables $\alpha_0, \alpha_1, \beta_1, \gamma$ and τ should be generated using the Metropolis-Hastings algorithm.

3. Some considerations on model selection

For model selection, we could use the predictive density for t_i given $\underset{\sim}{t}_{(i)} = \{t_1, ..., t_{i-1}, t_{i+1}, ..., t_n\}$.

The predictive density for t_i given $\underset{\sim}{t}_{(i)}$ is given by

$$c_i = f\left(t_i \mid t_{(i)}, x_i\right) = \int f(t_i \mid \underset{\sim}{\theta}, x_i)\, \pi\left(\underset{\sim}{\theta} \mid t_{(i)}, x_{(i)}\right) d\underset{\sim}{\theta}, \tag{21}$$

where $\pi\left(\underset{\sim}{\theta} \mid t_{(i)}; x_{(i)}\right)$ is the posterior density for a vector of parameters $\underset{\sim}{\theta}$, given the data $\underset{\sim}{t}_{(i)}$.

Using the Gibbs samples, (21) can be approximated by its Monte Carlo estimate,

$$\hat{f}\left(t_i \backslash t_{(i)}, x_i\right) = \frac{2}{RS} \sum_{r=1}^{R} \sum_{s=\frac{S}{2}+1}^{S} f\left(t_i \mid x_i, \underset{\sim}{\theta}^{(r,s)}\right)$$

where $\underset{\sim}{\theta}^{(r,s)}$ are generated for S iterations in each of R chains considering different initial values for $\underset{\sim}{\theta}$.

We can use $c_i = f\left(t_i \mid t_{(i)}, x_i\right)$ in model selection. In this way, we consider plots of c_i versus $i\ (i = 1, 2, ..., n)$ for different models; large values of c_i (in average) indicates the better model. We also could choose the model such that $c(l) = \prod_{i=1}^{n} c_i(l)$ is maximum (l indexes models).

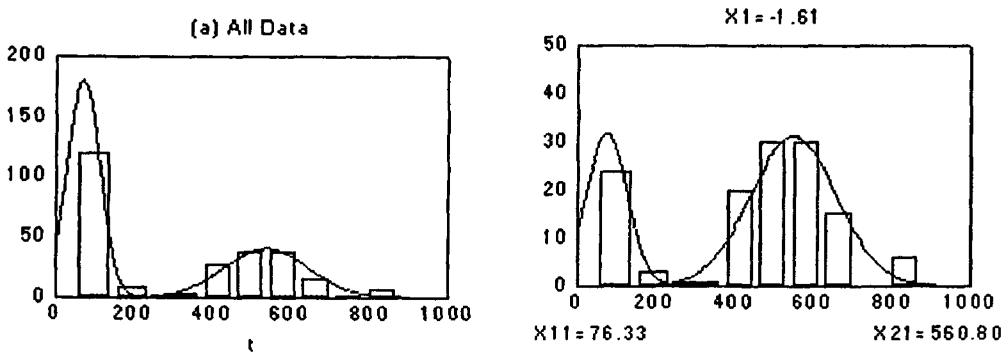
4. An example

Consider the survival data given in table 1, where we have the lifetimes of $n = 317$ beetles receiving four dosages of a toxity (generated data). Among the 317 beetles, 144, 69, 54 and 50 were 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 $n = 317$ beetle 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),30(600),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),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 histograms of all survival data (a), and for each individual dosage $x_i, i = 1, 2, 3, 4$. In these graphs, we clearly observe bimodal frequency distributions, indicating the need of fit of mixture distributions of the form (1).



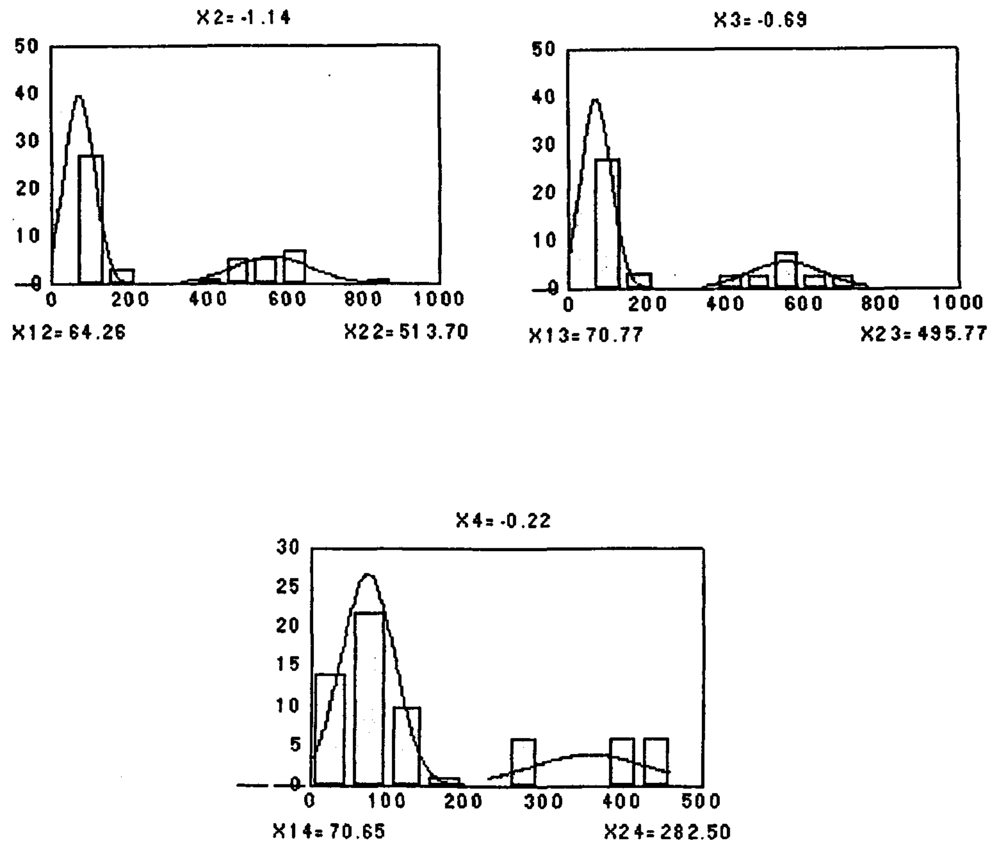


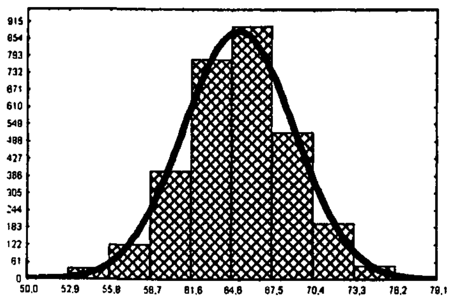
Figure 1. Histograms of Survival Data of Table 1.

To analyse the survival data of table 1, we assume some mixture distributions (1), with ligistic regression link: a exponential-normal model, a normal-normal distribution and a normal-gamma distribution.

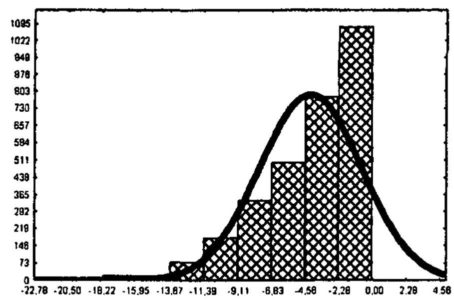
Considering the exponential-normal mixture model (11), and the prior distributions (12) with $\gamma_0 = 4.0$, $s_1^2 = 1.1$, $\tau_0 = 3.5$, $s_2^2 = 0.9$, we generated 3 separate Gibbs chains each of wich ran for 26000 iterations, and we monitored the convergence of the Gibbs samples using the Gelman and Rubin (1992) method that uses the analysis of variance technique to determine if further iterations are needed. For each parameter we considered the 26^{th}, 52^{th}, 78^{th},...iterations, which required a computational time of 6 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 $S = 3000$ Gibbs samples. We also have in table 2, the estimated potential scale reductions $\widehat{\mathbf{R}}$ (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 (Exponential-Normal Distribution)

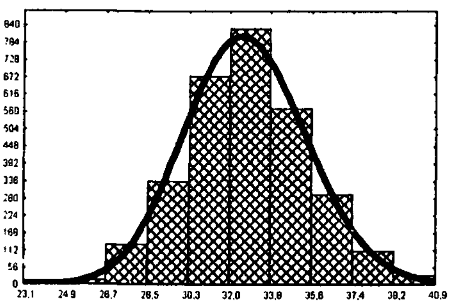
Parameter	Mean	95% Credible Interval	$\hat{R}$
α_1	65.03000	(57.030; 72.431)	1.00002500
β_1	- 4.21600	(-12.447; -0.102)	1.00146300
σ	32.76500	(27.822; 38.161)	1.00008087
α_2	0.00400	(0.0031; 0.0045)	1.00199600
β_2	0.00034	(0.00028; 0.00421)	1.00000000
γ	3.98400	(2.022; 5.921)	1.00342300
τ	3.54600	(1.865; 5.371)	1.00061300



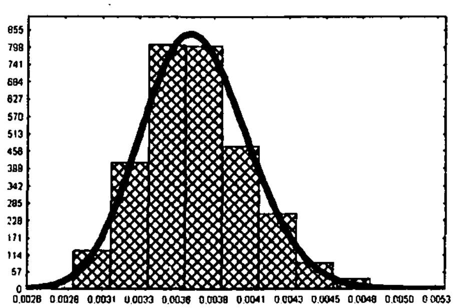
α_1



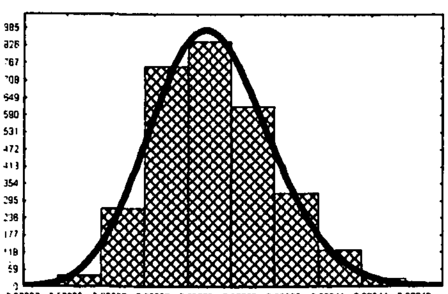
β_1



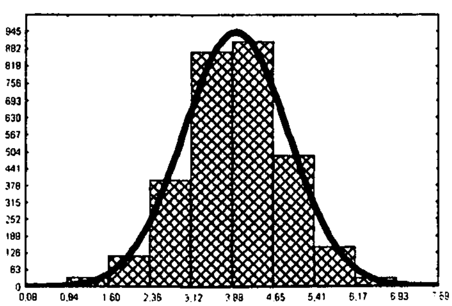
σ



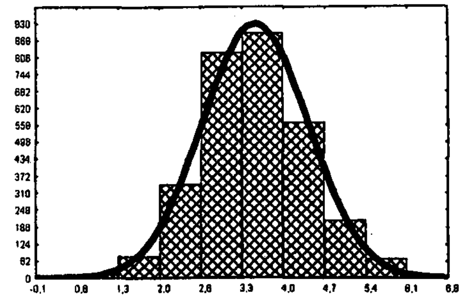
α_2



β_2



γ



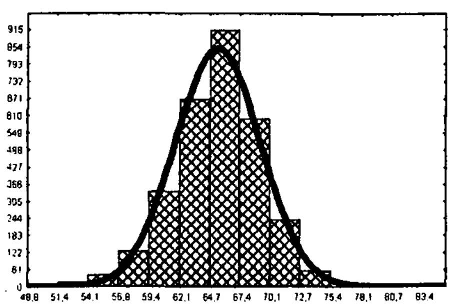
τ

Figure 2. Approximate marginal posterior densities (exponential-normal distribution).

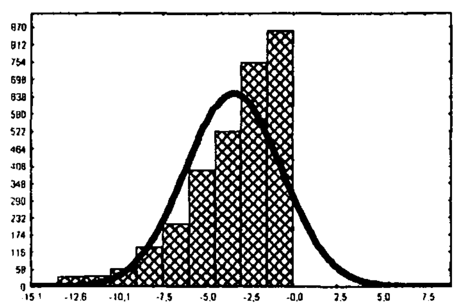
Considering the normal-normal mixture model (see section 2.2), and the prior densities (15) with $\gamma_0 = 4.0$, $s_1^2 = 0.57$, $\tau_0 = 3.0$, $s_2^2 = 0.37$, we also generated 3 separate Gibbs chains each of which ran for 26000 iterations, considering the 26^{th}, 52^{th}, 78^{th},...iterations, 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	$\widehat{R}$
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



β_1

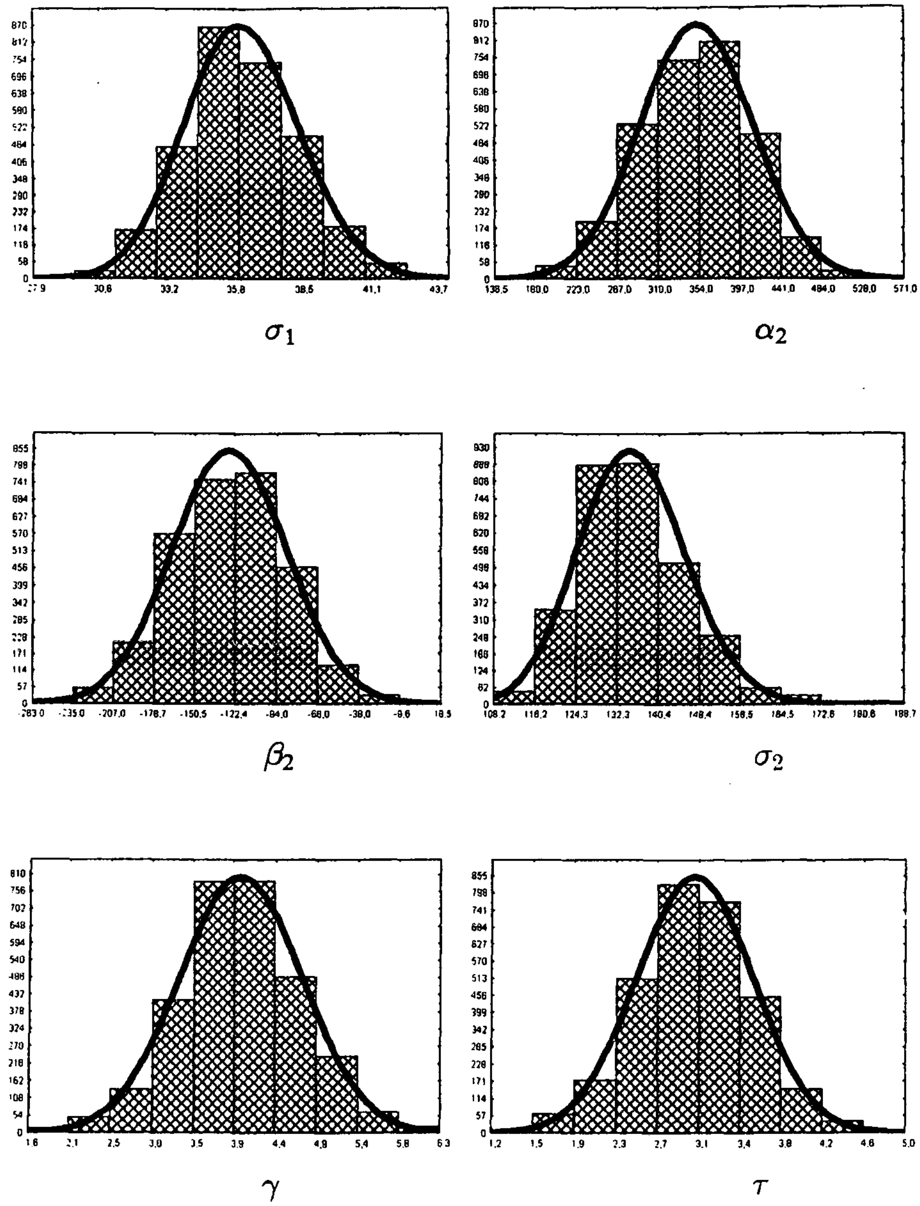


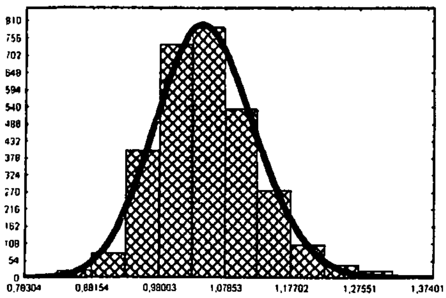
Figure 3. Approximate marginal posterior densities (normal-normal distribution).

Considering now, the gamma-normal mixture model (see section 2.3), and the prior densities (18) with $\gamma_0 = 4.0$, $s_1^2 = 0.1$, $\tau_0 = 3.0$, $s_2^2 = 0.09$, we generated 3 separate Gibbs chains each of wich ran for 26000 iterations, considering the 26^{th}, 52^{th}, 78^{th}, ...iterations, which required a computational time of 6 hours working with the software SAS in a Pentium 166 MHZ. In table 4, we have the obtained posterior summaries, and in figure 4, we have the approximate marginal posterior densities considering the $S = 3000$ Gibbs samples. For all parameters, we observe (see table 4), $\sqrt{\widehat{R}} < 1.1$, indicating approximate convergence.

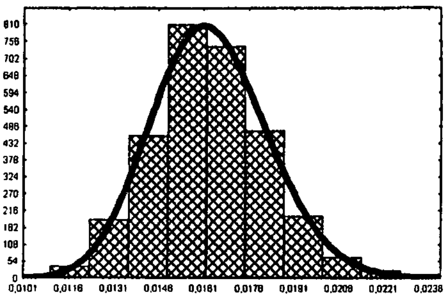
From the generated Gibbs samples, we also could get Monte Carlo estimates for the predictive densities $c_i = f\left(t_i \mid t_{(i)}, x_i\right)$ (see section 3) to be considered in the selection of the best model for the survival data of table 1.

Table 4. Posterior Summaries (Gamma-Normal Distribution).

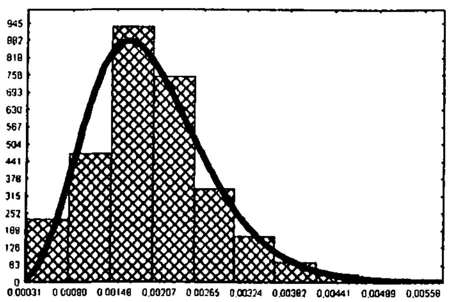
Parameter	Mean	95% Credible Interval	$\widehat{R}$
α_0	1.049	(0.921; 1.216)	1.00174800
α_1	0.016	(0.0128; 0.0204)	1.00632700
β_1	0.002	(0.00064; 0.0039)	1.00305900
α_2	342.525	(144.002; 508.232)	1.00416400
β_2	- 130.596	(- 248.680; -22.776)	1.00466600
σ	140.621	(116.796; 180.982)	1.00111200
γ	3.991	(3.459; 4.580)	1.00022700
τ	3.000	(2.441; 3.606)	1.00119100



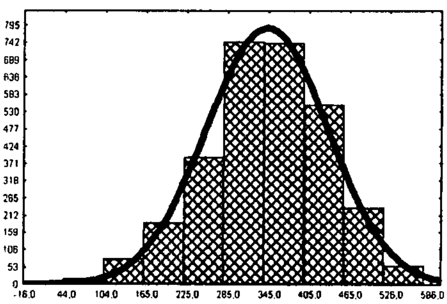
α_0



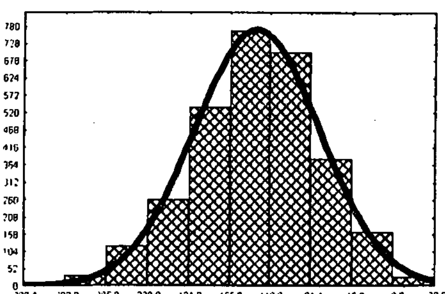
α_1



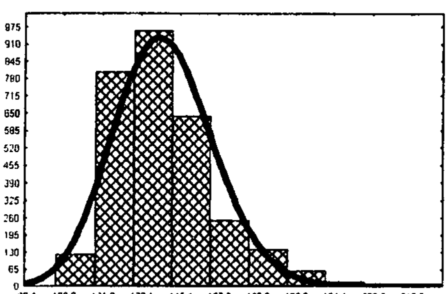
β_1



α_2



β_2



σ

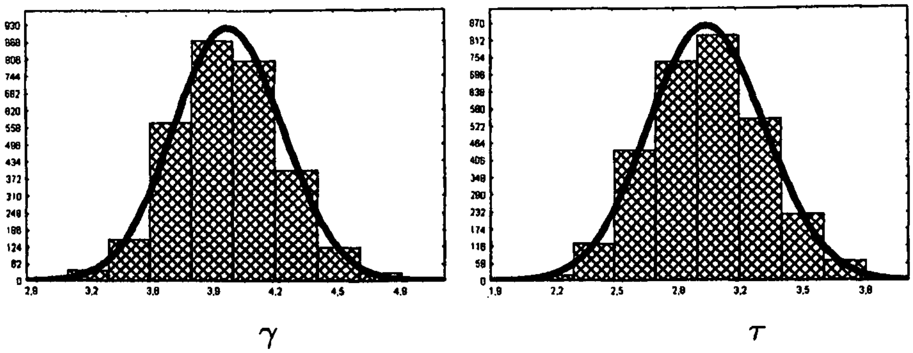


Figure 4. Approximate marginal posterior densities (gamma-normal distribution).

The values of $c(l) = \prod_{i=1}^{n} c_i(l)$, where l indexes models, considering the values of the predictive densities evaluated at the observed values t_i and x_i, are given by $c(1) = 2.56\text{x}100^{-501}$, (exponential-normal distribution), $c(2) = 3.14\text{x}100^{-390}$, (normal-normal distribution) and $c(3) = 2.11\text{x}100^{-478}$, (gamma-normal distribution), which indicates that the normal-normal mixture model has better fit for the survival data of table 1.

5. Some conclusions

The use of mixture models in the presence of one or more covariates is a suitable way to analyse survival data in many applications. Usually, a preliminary data analysis indicates that the standard parametrical models commonly used to analyse survival data could be not appropriate, as it was seen for the survival data of table 1.

Classical inference approaches for these models usually are difficult and the obtained asymptotic 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. The implementation of Gibbs and Metropolis-Hastings algorithms do not require very sophisticated computational expertise and the computational time required for simulation of the Gibbs samples is not high.

■ RESUMO: Neste artigo, apresentamos uma análise Bayesiana de modelos de misturas para dados de sobrevivência na presença de uma covariável. Considerando algoritmos de Gibbs com Metropolis-Hastings, obtemos estimadores de Monte Carlo para quantidades à posteriori de interesse, assumindo diferentes escolhas para as densidades no modelo de misturas. Também apresentamos algumas considerações na seleção de modelos e introduzimos um exemplo numérico, para ilustrar a metodologia proposta.

■ PALAVRA-CHAVE: Modelos de misturas, regressão, análise Bayesiana, algoritmo amostrador de Gibbs.

References

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

2. Farewell, V.T.,. The use of mixture models for the analysis of survival data with long-term survivors. *Biometrics* (Raleigh), 38, 1041-1046, 1982.

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

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

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

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

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

NOTAS DO ICMSC

SÉRIE ESTATÍSTICA

043/97 ACHCAR, J.A.; PEREIRA, G.A. - Mixture models for type II censored survival data in the presence of covariates.

042/97 MOALA, F.A.; RODRIGUES, J. - A note on the prior distributions for the Weibull reliability function.

041/97 RODRIGUES, J. - Diagnostic of convergence of a Rao-Black Wellised estimate of the marginal density via calibrated divergence measures.

040/97 BARATELA, D.S.; RODRIGUES, J. - Uma caracterização da existência da posteriori marginal do parâmetro N do modelo de Jelinski-Moranda.

039/97 ACHCAR, J.A.; BRASSOLATI, D. - Use of markov chain Monte Carlo methods for a bayesian analysis of software reliability models.

038/97 FRANCELIN, R.A.; BALLINI, R.; ANDARDE, M.G. - Back-propagation vs. Box and Jenkins model to streamflow forescasting.

037/97 ACHCAR, J.A.; LOIBEL, S. - Constant hazard function models with a change-point: a bayesian analysis using markov chain Monte Carlo methods

036 /97 MOALA, F..A.; RODRIGUES, J. - Bayesian inference of the Weibull reliability function via Laplace approximation

035/96 ACHCAR, J.A.; STORANI, K. - Nonhomogeneous poisson processes assuming a inverse Gaussian order statistics model for software reliability data: a bayesian approach.

034/96 ACHCAR, J.A.; LEANDRO, R.A. - Regression models for bivariate survival data: a bayesian approach.