UNIVERSIDADE DE SÃO PAULO

USE OF MIXTURE OF EXPONENTIAL POWER DISTRIBUTIONS FOR INTERVAL-CENSORED SURVIVAL DATA IN PRESENCE OF COVARIATES

JORGE ALBERTO ACHCAR
GILBERTO DE ARAÚJO PEREIRA

Nº 49

NOTAS

Instituto de Ciências Matemáticas de São Carlos

Instituto de Ciências Matemáticas de São Carlos

ISSN - 0103-2577

USE OF MIXTURE OF EXPONENTIAL POWER DISTRIBUTIONS FOR INTERVAL-CENSORED SURVIVAL DATA IN PRESENCE OF COVARIATES

JORGE ALBERTO ACHCAR
GILBERTO DE ARAÚJO PEREIRA

Nº 49

NOTAS DO ICMSC
Série Estatística

São Carlos
Mai./1998

USE OF MIXTURE OF EXPONENTIAL POWER DISTRIBUTIONS FOR INTERVAL-CENSORED SURVIVAL DATA IN 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 mixture of exponential power distributions for interval-censored 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.

Key words: Mixture models, Exponential power distributions, Bayesian analysis, Metropolis-Hastings algorithm.

1 Introduction

The use of mixture models has been considered in the literature as an alternative to nonparametric methods to analyse survival data, since in many applications, the usual parametrical models (see for example, Cox and Oakes, 1984) could not be appropriate for the data set. These data could be observed when a group of subjects may not react to treatments (see for example, Farewell, 1982; or Kuo and Peng, 1995).

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

$$f\left(t|\underline{x},\Theta\right) = \sum_{j=1}^{J} P\left(Y=j|\underline{x},\underline{\gamma}\right)\ f_j\left(t|Y=j,\underline{x},\underline{\beta}_j\right) \tag{1}$$

where T is the lifetime of an individual and $\Theta = \left(\underline{\beta}_1, \underline{\beta}_2, \ldots, \underline{\beta}_J, \underline{\gamma}\right)$ is the vector of all unknown parameters.

The probabilities $P(Y = j|\underline{x}, \underline{\gamma})$, 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,

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

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

$$F(t|\underline{x}, \Theta) = \sum_{j=1}^{J} P\left(j|\underline{x}, \underline{\gamma}\right) F_j\left(t|\underline{x}, \underline{\beta}_j\right) \tag{3}$$

where F_j is the distribution function for f_j.

Classical inference methods for mixture models based on maximum likelihood estimates could be difficult to obtain, even for simple cases considering $J = 2$ mixture distributions (see for example, Titterington et al, 1985).

In this paper, considering interval-censoring survival data, we consider Bayesian methods based on Gibbs sampling with Metropolis-Hastings algorithms (see for example, Smith an Roberts, 1993; or Robert, 1996). These Markov Chain Monte Carlo (MCMC) methods has been explored in the literature considering censored or uncensored observations (see for example, Diebolt and Robert, 1994; or Kuo and Peng, 1995).

2 A Bayesian Analysis Using MCMC Methods

For interval-censored survival data, the time of failure is only known to be between two time points. On developing the likelihood, if the individual fails between the time points $(a_L, a_U]$, then the contribution to the likelihood is, $F(a_U|\underline{x}, \Theta) - F(a_L|\underline{x}, \Theta)$; if it is known that the individual survives beyond the censoring time point c, then the contribution to the likelihood is $1 - F(c|\underline{x}, \Theta)$.

Considering N individuals in study, let us assume that the first n out of N failure times are interval censored with the i^{th} individual dying between t_{iL} and t_{iU}.

The remaining $N-n$ individuals are right censored at t_{iL}.

The likelihood function is given by

$$
\begin{aligned}
L(\Theta|\underline{t},\underline{x}) &= \prod_{i=1}^{n}\{F(t_{iU}|x_i,\Theta)-F(t_{iL}|x_i,\Theta)\} \\
&\quad \prod_{i=n+1}^{N}\{1-F(t_{iL}|x_i,\Theta)\}
\end{aligned} \tag{4}
$$

This likelihood includes the grouped data where a number of individuals die within the same interval, because we can set t_{iL} to be identical for several i; similarly for t_{iU}.

Assuming a prior density $\pi(\Theta)$, the joint posterior density for Θ is given by,

$$
\begin{aligned}
\pi(\Theta|\underline{t},\underline{x}) &\propto \pi(\Theta)\left\{\prod_{i=1}^{n}[F(t_{iU}|x_i,\Theta)-F(t_{iL}|x_i,\Theta)]\right. \\
&\quad \left.\prod_{i=n+1}^{N} S(t_{iL}|x_i,\Theta)\right\}
\end{aligned} \tag{5}
$$

where $F(t_i|\underline{x},\Theta) \;=\; \sum_{j=1}^{J} P\left(Y=j|\underline{x},\underline{\gamma}\right) F_j\left(t_i|Y=j,\underline{x},\underline{\beta}_j\right)$,

and $S(t_i|\underline{x},\Theta) \;=\; 1-F(t_i|\underline{x},\Theta)$

To simplify the conditional distributions needed for the Gibbs sampling algorithm (see for example, Gelfand and Smith, 1990) we introduce latent variables (data augmentation technique; see for example, Tanner and Wong, 1987) and the use of the EM algorithm (see Dempster, Laird and Rubin, 1977) that allow us to consider a likelihood of a product of compnents model for $i.i.d.$ observations as opposed to the mixture likelihood and censored observations. This is given by augmenting the original data with two classes of latent variables: one is the truncated random variable W and other is the index variable denoted by Z that convert the mixture model to a model of independent components (see Kuo and Peng, 1995).

If the i^{th} individual is interval censored between (t_{iL}, t_{iU}), then we can generate a latent variable w_i from the truncated density $f(w_i)\;/\;[F(t_{iU})-F(t_{iL})]$, where f

is given in (1). This can be done by setting

$$w_i = F^{-1}\left\{F\left(t_{iL}|x_i,\Theta\right) + U\left[F\left(t_{iU}|x_i,\Theta\right) - F\left(t_{iL}|x_i,\Theta\right)\right]\right\} \tag{6}$$

where U has an uniform $U(0,1)$ distribution and F^{-1} is the inverse function of F.

Similarly, if the i^{th} individual is right-censored at t_{iL}, we can generate a latent variable w_i from the truncated density $f(w_i)\ /\ [1-F(t_{iL})]$ by setting,

$$w_i = F^{-1}\left\{F\left(t_{iL}|x_i,\Theta\right) + U\left[1 - F\left(t_{iL}|x_i,\Theta\right)\right]\right\} \tag{7}$$

From (6) and (7), we obtain a random sample of generated values $w_1, w_2, \ldots, w_N$. In this case, the likelihood function is given by

$$\begin{aligned} L\left(\Theta|\underline{w},\underline{t},\underline{x}\right) &= \prod_{i=1}^{N} f\left(w_i|x_i,\Theta\right) \\ &= \prod_{i=1}^{N}\left\{\sum_{j=1}^{J} P\left(j|x_i,\underline{\gamma}\right)\ f_j\left(w_i|x_i,\underline{\beta}_j\right)\right\}. \end{aligned} \tag{8}$$

Assuming the special case of $J=2$ distributions in the mixture model (1), the other class of latent variables is given by $\underline{z}_i = (z_{i1}, z_{i2})$, $i=1,2,\ldots,N$ where $z_{i1}|\Theta, w_i, x_i \sim b(1, h_{i1})$ (a Bernoulli distribution) with h_{i1} given by

$$h_{i1} = \frac{P\left(1|x_i,\underline{\gamma}\right)\ f_1\left(w_i|x_i,\underline{\beta}_1\right)}{\sum_{j=1}^{2} P\left(j|x_i,\underline{\gamma}\right)\ f_j\left(w_i|x_i,\underline{\beta}_j\right)} \tag{9}$$

That is,

$$\pi\left(\underline{z}_i\right) \propto h_{i1}^{z_{i1}}\left(1-h_{i1}\right)^{z_{i2}} \tag{10}$$

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

Thus,

$$\pi(z_1,\ldots,z_N) \propto \frac{\prod_{i=1}^{N}\prod_{j=1}^{2}\left\{P\left(j|x_i,\underline{\gamma}\right)\ f_j\left(w_i|x_i,\underline{\beta}_j\right)\right\}^{z_{ij}}}{\prod_{i=1}^{N}\left\{\sum_{j=1}^{2}P\left(j|x_i,\underline{\gamma}\right)\ f_j\left(w_i|x_i,\underline{\beta}_j\right)\right\}} \tag{11}$$

Combining (11) with (8), and considering a prior density $\pi(\Theta)$, the joint posterior density for Θ is given by

$$\pi\left(\Theta|\underline{w},\underline{z},\underline{x}\right) \propto \pi(\Theta)\left\{\prod_{i=1}^{N}\prod_{j=1}^{2}\left[P\left(j|x_i,\underline{\gamma}\right)\ f_j\left(w_i|x_i,\underline{\beta}_j\right)\right]^{z_{ij}}\right\} \tag{12}$$

Similar results are obtained when $J > 2$.

To generate samples of the joint posterior distribution (12), we use the Gibbs sampling algorithm. Starting with initial values $\Theta^{(0)} = (\theta_1^{(0)},\theta_2^{(0)},\ldots,\theta_p^{(0)})$, follow the steps:

(i) Generate samples $\underline{w}^{(1)} = \left(w_1^{(1)},w_2^{(1)},\ldots,w_N^{(1)}\right)$,
from (6) and (7),

(ii) Generate samples $\underline{z}^{(1)} = \left(z_1^{(1)},z_2^{(1)},\ldots,z_N^{(1)}\right)$,
where $\underline{z}_i^{(1)} = (z_{i1}^{(1)},z_{i2}^{(1)})$, from (10).

(iii) Generate a sample of Θ, from the conditional distributions (13)

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

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

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

Using the Gibbs sampling algorithm, we could consider different choices for $f_j(t|j, \underline{x}, \underline{\beta}_j)$ and $P(j|\underline{x}, \underline{\gamma})$ in (1).

3 Mixture of Two Exponential Power Distributions

Assume a mixture of two exponential power distributions (see for example, Box and Tiao, 1973), with density,

$$f_j\left(t|\underline{x}, \underline{\beta}_j\right) = \frac{w(\delta_j)}{\sigma_j} \exp\left\{-c(\delta_j)\left|\frac{t-\theta_j}{\sigma_j}\right|^{2/(1+\delta_j)}\right\} \qquad (14)$$

where $-\infty < t < \infty$; $j = 1, 2$; $\underline{\beta}_j = (\delta_j, \sigma_j, \theta_j)$

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

and,

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

$\sigma_j > 0, -1 < \delta_j \leq 1$ 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 ($\delta_j < 0$). If ($\delta_j = 1$), the distribution (14) is the double exponential. Observe that for lifetime data, the data is concentrated at positive values of t. We also could consider the logarithms of the survival times having the exponential power distribution with density (14).

Let us assume only a covariate x, a mixture of $J = 2$ exponential power distributions (14) with $\theta_j = \alpha_j + \beta_j x$, $j = 1, 2$, and the logistic regression link (2) with,

$$P(1|x,\gamma,\tau) = \frac{e^{\gamma+\tau x}}{1+e^{\gamma+\tau x}}, \text{and}$$
$$P(2|x,\gamma,\tau) = 1 - P(1|\underline{x},\gamma,\tau) = \frac{1}{1+e^{\gamma+\tau x}}. \tag{15}$$

The distribution function of the exponential power distribution with density (14) is given by,

$$F_j(t|\underline{x},\underline{\beta}_j) = \left\{ \frac{(1+\delta_j)w(\delta_j)\Gamma(\frac{\delta_j+1}{2})}{[c(\delta_j)]^{(\delta_j+1)/2}} \right\} I_{\frac{(\delta_j+1)}{2}} \left\{ c(\delta_j) \left| \frac{t-\theta_j}{\sigma_j} \right|^{2/(1+\delta_j)} \right\} \tag{16}$$

where

$$I_k(s) = \frac{1}{\Gamma(k)} \int_0^s x^{k-1} e^{-x} dx,$$

is the incomplete gamma integral, $\theta_j = \alpha_j + \beta_j x,\ j = 1, 2$.

From (16), we generate latent variables w_i in the EM algorithm (see (6) and (7)).

Assuming prior independence among the parameters, consider the following prior densities,

(i) $\alpha_1 \sim N(\alpha_{10}, \sigma_{11}^2); \alpha_{10}, \sigma_{11}^2$ known,

(ii) $\beta_1 \sim N(\beta_{10}, \sigma_{12}^2); \beta_{10}, \sigma_{12}^2$ known,

(iii) $\sigma_1 \sim \Gamma[m_{11}, n_{11}];\ m_{11}, n_{11}$ known,

(iv) $\delta_1 \sim N(\delta_{10}, \sigma_{13}^2); \delta_{10}, \sigma_{13}^2$ known,

(v) $\alpha_2 \sim N(\alpha_{20}, \sigma_{21}^2); \alpha_{20}, \sigma_{21}^2$ known, (17)

(vi) $\beta_2 \sim N(\beta_{20}, \sigma_{22}^2); \beta_{20}, \sigma_{22}^2$ known,

(vii) $\sigma_2 \sim \Gamma[m_{22}, n_{22}];\ m_{22}, n_{22}$ known,

(viii) $\delta_2 \sim N(\delta_{20}, \sigma_{23}^2)$; $\delta_{20}, \sigma_{23}^2$ known,

(ix) $\gamma \sim N(\gamma_0, \sigma_{14}^2)$, γ_0, σ_{14}^2 known,

(x) $\tau \sim N(\tau_0, \sigma_{24}^2)$, τ_0, σ_{24}^2 known,

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 .

Considering the introduction of the latent variables $\underline{w}$ and $\underline{z}$ (see section 2), we have from (12) the joint posterior distribution for $\Theta = (\alpha_1, \beta_1, \sigma_1, \delta_1, \alpha_2, \beta_2, \sigma_2, \delta_2, \gamma, \tau)$:

$$\pi(\Theta \mid \underline{w}, \underline{z}, \underline{x}) \propto \pi(\Theta) \left\{ \frac{exp(r\gamma + a_1\tau)}{\prod_{i=1}^{n} (1 + e^{\gamma + \tau x_i})} \right\} \frac{[w(\delta_1]^r [w(\delta_2]^{n-r}}{\sigma_1^r \sigma_2^{n-r}} \tag{18}$$

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

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

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

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

$$\begin{aligned}
\pi(\underline{\theta}) \propto \; & exp\left\{ -\frac{1}{2\sigma_{11}^2} (\alpha_1 - \alpha_{10})^2 \right\} \; exp\left\{ -\frac{1}{2\sigma_{12}^2} (\beta_1 - \beta_{10})^2 \right\} \\
& exp\left\{ -\frac{1}{2\sigma_{13}^2} (\delta_1 - \delta_{10})^2 \right\} \; exp\left\{ -\frac{1}{2\sigma_{14}^2} (\gamma - \gamma_0)^2 \right\} \\
& exp\left\{ -\frac{1}{2\sigma_{24}^2} (\tau - \tau_0)^2 \right\} \; \sigma_1^{m_{11}-1} \, \sigma_2^{m_{22}-1} \, exp\left\{ -n_{11}\sigma_1 - n_{22}\sigma_2 \right\} \\
& exp\left\{ -\frac{1}{2\sigma_{21}^2} (\alpha_2 - \alpha_{20})^2 \right\} \; exp\left\{ -\frac{1}{2\sigma_{22}^2} (\beta_2 - \beta_{20})^2 \right\} \\
& exp\left\{ -\frac{1}{2\sigma_{23}^2} (\delta_2 - \delta_{20})^2 \right\}
\end{aligned}$$

To generate samples of the joint distribution (18), we use steps (i)-(iii) of the Gibbs algorithm (13), where the conditional distributions for the parameters are given by,

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

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

where,

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

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

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

where,

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

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

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

where,

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

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

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

where,

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

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

$$exp\left\{-\frac{1}{2\sigma_{21}^2}(\alpha_2-\alpha_{20})^2\right\}\Psi_5(\Theta) \tag{19}$$

where,

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

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

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

where,

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

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

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

where,

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

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

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

where,

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

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

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

where,

$$\Psi_9(\Theta) = exp\{r\gamma - \textstyle\sum_{i=1}^n ln\,(1 + e^{\gamma + \tau x_i})\}$$

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

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

where,

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

Observe that we need to use the Metropolis-Hastings algorithm to generate Θ.

3.1 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 (14)), with the same logistic regression links given in (15), 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, (20)

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

We also assume independence among the parameters.

From (12), we get the joint posterior density for $\Theta = (\alpha_1, \beta_1, \sigma_1, \alpha_2, \beta_2, \sigma_2, \gamma, \tau)$,

$$\pi\left(\Theta|\underline{w},\underline{z},\underline{x}\right) \propto \frac{\sigma_1^{-r}\sigma_2^{-(n-r)}}{\left\{\prod_{i=1}^{n}\left(1+e^{\gamma+\tau x_i}\right)\right\}}$$

$$exp\left\{-\frac{1}{2\sigma_{14}^2}\left(\gamma-\gamma_0\right)^2+\gamma r-\frac{1}{2\sigma_{24}^2}\left(\tau-\tau_0\right)^2+\tau a_1\right\} \tag{21}$$

$$\left\{exp\left\{-\frac{1}{2\sigma_1^2}\sum_{i=1}^{n}z_{i1}\left(w_i-\alpha_1-\beta_1x_i\right)^2-\frac{1}{2\sigma_2^2}\sum_{i=1}^{n}z_{i2}\left(w_i-\alpha_2-\beta_2x_i\right)^2\right\}\right\}$$

where r, $(n-r)$ and a_1 are defined in (18).

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

(i) $\pi\left(v|\alpha_1,\beta_1,\alpha_2,\beta_2,\sigma_2,\gamma,\tau,\underline{w},\underline{z},\underline{x}\right) \sim \Gamma\left(\frac{r}{2}+1,\frac{\sum_{i=1}^{n}z_{i1}\left(w_i-\alpha_1-\beta_1x_i\right)^2}{2}\right)$

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

(ii) $\pi\left(\alpha_1,|\beta_1,\sigma_1,\alpha_2,\beta_2,\sigma_2,\gamma,\tau,\underline{w},\underline{z},\underline{x}\right) \sim N\left(\frac{\sum_{i=1}^{n}z_{i1}\left(w_i-\beta_1x_i\right)}{r},\frac{\sigma_1^2}{r}\right)$

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

$$N\left(\frac{\sum_{i=1}^{n}z_{i1}x_i\left(w_i-\alpha_1\right)}{\sum_{i=1}^{n}z_{i1}x_i^2},\frac{\sigma_1^2}{\sum_{i=1}^{n}z_{i1}x_i^2}\right) \tag{22}$$

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

$$\Gamma\left(\frac{(n-r)}{2}+1,\frac{\sum_{i=1}^{n}z_{i2}\left(w_i-\alpha_2-\beta_2x_i\right)^2}{2}\right)$$

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

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

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

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

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

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

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

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

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

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

4 A Normal-Exponential Mixture Model

Assume a mixture of normal-exponential distributions in (1), with,

$$
\begin{aligned}
f_1\left(t_i|x_i,\underline{\beta}_1\right) &= \frac{1}{\sqrt{2\pi}\sigma}\, exp\left\{-\frac{1}{2\sigma^2}\,(t_i-\alpha_1-\beta_1 x_i)^2\right\}, \\
f_2\left(t_i|x_i,\underline{\beta}_2\right) &= (\alpha_2+\beta_2 x_i)^{-1}\; exp\left\{-(\alpha_2+\beta_2 x_i)^{-1}\, t_i\right\},
\end{aligned}
\tag{23}
$$

$\underline{\beta}_1 = (\alpha_1, \beta_1, \sigma)$ and $\underline{\beta}_2 = (\alpha_2, \beta_2)$. Also assume the same logistic regression link (15).

Assuming prior independence among the parameters, consider the following prior densities for α_1, β_1, σ, α_2, β_2, γ, and τ:

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

(ii) $\gamma \sim N(\gamma_0, \sigma_{11}^2), \gamma_0, \sigma_{11}^2$ known, (24)

(iii) $\tau \sim N(\tau_0, \sigma_{12}^2), \tau_0, \sigma_{12}^2$ known

We obtain generated samples of the joint posterior distribution (12) for $\Theta = (\alpha_1, \beta_1, \sigma, \alpha_2, \beta_2, \gamma, \tau)$ using steps (i)-(iii) of the Gibbs sampling algorithm (13), where the conditional distributions for the parameters are given by,

(i) $\pi\left(v \mid \alpha_1, \beta_1, \alpha_2, \beta_2, \gamma, \tau, \underline{z}, \underline{w}, \underline{x}\right) \sim$

$$
\Gamma\left(\frac{r}{2}+1, \frac{\sum_{i=1}^{n} z_{i1}\left(w_i-\alpha_1-\beta_1 x_i\right)^2}{2}\right)
$$

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

(ii) $\pi\left(\alpha_1 \mid \beta_1, \sigma, \alpha_2, \beta_2, \gamma, \tau, \underline{z}, \underline{w}, \underline{x}\right) \sim$

$$N\left(\frac{\sum_{i=1}^{n} z_{i1}\left(w_i - \beta_1 x_i\right)}{r}, \frac{\sigma^2}{r}\right)$$

(iii) $\pi\left(\beta_1 \mid \alpha_1, \sigma, \alpha_2, \beta_2, \gamma, \tau, \underline{z}, \underline{w}, \underline{x}\right) \sim$

$$N\left(\frac{\sum_{i=1}^{n} z_{i1} x_i\left(w_i - \alpha_1\right)}{\sum_{i=1}^{n} z_{i1} x_i^2}, \frac{\sigma^2}{\sum_{i=1}^{n} z_{i1} x_i^2}\right) \quad (25)$$

(iv) $\pi\left(\alpha_2 \mid \alpha_1, \beta_1, \sigma, \beta_2, \gamma, \tau, \underline{z}, \underline{w}, \underline{x}\right) \propto \alpha_2^{-(n-r)}$

$$exp\left\{-\frac{1}{\alpha_2}\sum_{i=1}^{n} z_{i2} w_i\left(1 + \frac{\beta_2}{\alpha_2}x_i\right)^{-1}\right\}\ \Psi_1(\Theta)\ ,$$

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

(v) $\pi\left(\beta_2 \mid \alpha_1, \beta_1, \sigma, \alpha_2, \gamma, \tau, \underline{z}, \underline{w}, \underline{x}\right) \propto \beta_2^{-(n-r)}$

$$exp\left\{-\frac{1}{\beta_2}\sum_{i=1}^{n} z_{i2} w_i\left(\frac{\alpha_2}{\beta_2} + x_i\right)^{-1}\right\}\ \Psi_2(\Theta)\ ,$$

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

(vi) $\pi\left(\gamma \mid \alpha_1, \beta_1, \sigma, \alpha_2, \beta_2, \tau, \underline{z}, \underline{w}, \underline{x}\right) \propto \quad exp\left\{-\frac{1}{2\sigma_{11}^2}\left(\gamma - \gamma_0\right)^2\right\}\Psi_3(\Theta),$

where $\Psi_3(\Theta) = \exp\{\gamma r - \sum_{i=1}^{n} ln(1 + e^{\gamma + \tau x_i})\}$

(vii) $\pi\left(\tau \mid \alpha_1, \beta_1, \sigma, \alpha_2, \beta_2, \gamma, \underline{z}, \underline{w}, \underline{x}\right) \propto \quad exp\left\{-\frac{1}{2\sigma_{12}^2}\left(\tau - \tau_0\right)^2)\right\}\Psi_4(\Theta),$

where $\Psi_4(\Theta) = \exp\{\tau a_1 - \sum_{i=1}^{n} ln(1 + e^{\gamma + \tau x_i})\}$, $r = \sum_{i=1}^{n} z_{i1}$, $(n-r) = \sum_{i=1}^{n} z_{i2}$ and $a_1 = \sum_{i=1}^{n} x_i z_{i1}$.

Observe that, we need to use the Metropolis-Hastings algorithm to generate the variables $\alpha_2, \beta_2, \gamma$ and τ.

5 An Example

Consider the data set of table 1 introduced by Hewlett (1974). The data set consists of interval-censored lifetimes of 317 male adult flour beetles (tribolium castaneum) that were exposed to pyrethrum, a well known plant-based insecticide. Among the 317 insects, 144, 69, 54 and 50 were sprayed of pyrethrum at concentrations of 0.20, 0.32, 0.50 and 0.80 mg/cm^2, respectively. The log-doses (denoted by x_c) are $-1.61, -1.14, -0.69$ and -0.22 respectively.

	(x) Log-dose (mg/cm^2)			
r	-1.61	-1.14	-0.69	-0.22
1	3	7	5	4
2	11	10	8	10
3	10	11	11	8
4	7	16	15	14
5	4	3	4	8
6	3	2	2	2
7	2	1	1	1
8	1	0	1	0
9	0	0	0	0
10	0	0	0	1
11	0	0	0	0
12	1	0	0	0
13	1	0	0	0
14	101	19	7	2

Table 1. Flour beetle data (the numbers indicate the number dead per day, the row 14 gives the number survived after day 13)

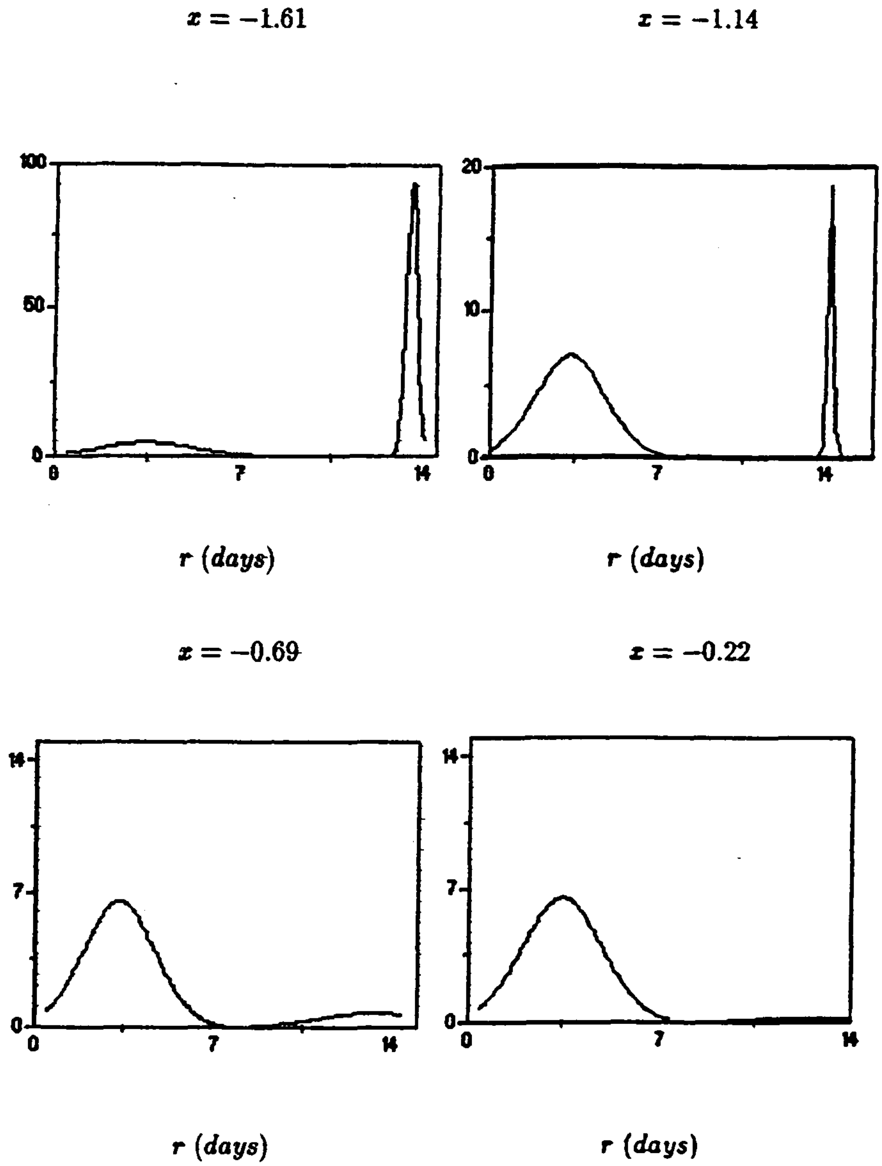

Figure 1. Plots of number dead for different doses x

In figure 1, we have plots for the interval-censored lifetimes of table 1. We observe approximately symmetrical frequency distributions for the different doses of the insecticide, especially for the beetles killed before day 13 that are more susceptible to treatment.

Thus, we assume a mixture of two exponential power distributions (14) with logistic regression links (15) in model (1). For a Bayesian analysis, we considered prior distributions (17) with $\alpha_{10} = 65, \sigma_{11}^2 = 14, \beta_{10} = -3.5, \sigma_{12}^2 = 7.8, \delta_{10} = 0, \sigma_{13}^2 = 0.01, \delta_{20} = 0, \sigma_{23}^2 = 0.01, m_{11} = 180, n_{11} = 5.1, m_{22} = 137, n_{22} = 1.0, \alpha_{20} =$

$351, \sigma_{21}^2 = 3612, \beta_{20} = -127, \sigma_{22}^2 = 1573, \gamma_0 = 4, \sigma_{14}^2 = 2.5, \tau_0 = 3$ 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).

In the generation of latent variables w_i in the EM algorithm in each iteration of the Gibbs sampling algorithm from (6) and (7) with F_j, $j = 1, 2$ given by (16), we considered interval-censored data in hours. Thus, for the interval-censored lifetime of table 1 with $x = -1.61$, we generated 3 observations w_i in the interval $(0, 24]$ from (6), for the first day. For the second day, we generated 11 observations w_i in the interval $(24, 48]$, and so on.

From the Gibbs sampling algorithm (i)-(iii) in (13) with the conditional distributions (19), we generated 3 separate Gibbs chains each of wich ran for 5.000 iterations, and we monitored the convergence of the Gibbs samples using the Gelman and Rubin (1992) method. For each parameter, we considered the $10^{th}, 20^{th}, 30^{th}, \ldots$ iterations, which required a computational time of 85 hours working with the software SAS in a Pentium 166 MHZ. In table 2, we have the obtained posterior summaries for the parameters. We also in table 2, the estimated potential scale reductions $\widehat{R}$ (see Gelman and Rubin, 1992) for all the parameters. In this case, the considered number of iterations were sufficient for approximate convergence ($\sqrt{\widehat{R}} < 1.1$ for all parameters).

Parameter	Mean	95% Credible Interval	$\widehat{R}$
α_1	62.14000	(44.35; 76.14)	1.000241
β_1	-7.55000	(-22.12; -0.25)	1.006190
σ_1	37.53000	(32.26; 43.45)	1.005455
δ_1	1.00500	(0.99; 1.02)	1.000356
α_2	36.76000	(0.72; 152.35)	1.002669
β_2	-368.68000	(-441.76; -310.72)	1.003292
σ_2	150.07000	(126.45; 170.40)	1.003185
δ_2	0.00720	(-0.012; 0.027)	1.003137
γ	3.59000	(2.41; 4.95)	1.003258
τ	3.61000	(2.32; 5.00)	1.003394

Table 2. Posterior summaries (mixture of two exponential power distributions)

Observing the 95 % credible intervals for δ_1 and δ_2 (see table 2), we could consider a reanalysis of the interval-censored data of table 1 assuming a mixture of a double-exponential ($\delta_1 = 1$ in (14)) and a normal distribution ($\delta_2 = 0$ in (14)).

For a reanalysis of the data set of table 1, we assume a mixture of an exponential and a normal distribution with logistic regression links (15) in model (1). Observe that one population consists of beetle susceptible to toxity while the other consists of beetles that are more resistant to the treatment and hence tend to live longer. Therefore, the choice of an exponential distribution for the group more resistant to treatment is a suitable choice. This model also was considered to analyse the data of table 1 by Kuo and Peng (1995).

Considering the exponential-normal mixture model (section 5) and the prior distributions (24) with $\gamma_0 = 4$, $\sigma_{11}^2 = 0.7$, $\tau_0 = 3.6$ and $\sigma_{12}^2 = 1.1$, we generated 3 separate Gibbs chains each of wich ran for 5.000 interations. For each parameter, we considered the $5^{th}, 10^{th}, 15^{th}, \ldots$ iterations, which required a computational time of 53 hours working with the software SAS in a Pentium 166 MHZ. In table 3, we have the obtained posterior summaries for the parameters. We also have in table 3, the estimated potential scale reductions $\hat{R}$ (see Gelman and Rubin, 1992) for all the parameters.

Parameter	Mean	95% Credible Interval	$\hat{R}$
α_1	63.810	(55.94; 71.23)	1.00008700
β_1	-4.340	(-13.24; -0.17)	1.00000000
σ	32.280	(27.53; 37.27)	1.00112400
α_2	0.0037	(0.003; 0.0045)	1.00211000
β_2	0.0003	(0.0003; 0.00043)	1.00000000
γ	4.04	(2.38; 5.59)	1.00015000
τ	3.61	(1.58; 5.65)	1.00629900

Table 3. Posterior summaries (mixture of an exponential and a normal distributions

From the 95% credible interval for β in table 3, we observe that there is a small dose effect on survival times for the susceptible group. This result is in agreement with previous studies by Hewlett (1974) and by Kuo and Peng (1995). This result is also in agreement with the plots of figure 1. We also observe from the 95% credible intervals for τ in table 3, a great dose effect on the mixing probabilities.

6 Concluding Remarks

The use of mixture models is a suitable way to analyse interval-censored lifetime data, since in many applications we can have individuals not susceptible to a given treatment. The mixture of exponential power distributions (14) can give great flexibility to analyse interval-censored data, especially for approximately symmetrical frequency distributions.

A Bayesian analysis for mixture models in the presence of covariates using Markov Chain Monte Carlo methods can give simple and accurate inference results. Classical inference results for these models could be in many cases, very difficult, not accurate or impossible to obtain.

References

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

Cox, D.R.; Oakes, D. (1984). Analysis of survival data. London: Chapman and Hall.

Dempster, A.; Laird, N.; Rubin, D. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society*, Series B, 39, 1-38.

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

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

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.

Hewlett, P.S. (1974). Time from dosage do death in beetles Tribolium castaneum, treated with pyrethrins or DDT, and its bearing on dose-mortality relations. *Journal of Stored Product Research*, 10, 27-41.

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.

Robert, C.P. (1996) Mixture of distributions: inference and estimation, in Markov Chain Monte Carlo in practice (eds Gilks, W.R.et al), Chapman and Hall, 441-464.

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.

Tanner, M.A.; Wong, W.H. (1987). The calculation of posterior distributions via data augmentation (with discussion), *Journal of the American Statistical Association*, 82, 528-550.

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

048/98 CID, J.E.R.; ACHCAR, J.A. - Software reliability considering the super position of non-homogeneous Poisson processes in the presence of a covariate.

047/98 ACHCAR, J.A.; PEREIRA, G.A. - Use of exponential power distributions for mixture models in the presence of covariates.

046/98 ANDRADE, M.G.; HUTTER, C.F.F. - Teste de sazonalidade para função de autocorrelação de processos auto-regressivos periódicos - PAR (pm)

045/98 ACHCAR, J.A.; ANDRADE, M.G.; LOIBEL, S. - Weibull hazard function with a change-point: a bayesian approach using Markov chain Monte carlo methods.

044/97 ACHCAR, J.A.; PEREIRA, G.A. - Bayesian analysis of mixture models for survival data: some computational aspects.

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.