Brazilian Journal of Probability and Statistics (2001), 15, pp. 155–168.
©Associação Brasileira de Estatística

USE OF MIXTURE OF DISTRIBUTIONS FOR THE ERROR IN REGRESSION MODELS FOR LIFETIME DATA

Josmar Mazucheli[1], **Jorge A. Achcar**[2], **Robert Kass**[3] and **Roberto Carta**[3]

[1] *Departamento de Estatística, Universidade Estadual de Maringá, Maringá – PR, Brazil. Email: jmazucheli@uem.br*

[2] *ICMC, Universidade de São Paulo, São Carlos – SP, Brazil. Email: jorge@icmc.sc.usp.br*

[3] *Carnegie Mellon University, Dept of Statistics, Pittsburgh, PA, USA. Emails: kass@stat.cmu.edu and roberto@stat.cmu.edu*

Summary
In this paper we present a Bayesian analysis of location-scale regression models assuming standard lifetime distributions and an additional error term with a mixture of normal distributions. Considering a censored lifetime data set, we use Gibbs sampling with Metropolis-Hastings algorithm to get Bayesian quantities of interest. The proposed regression model gives a great flexibility to fit lifetime data as we see in an example.

Key Words: Censored data; lifetime data; location-scale regression models; MCMC algorithms; mixture of normal distributions; semi-parametric models.

1 Introduction

A usual distribution considered for modelling survival time data is the two parameter Weibull distribution with density

$$f(t) = \frac{\beta}{\mu}\left(\frac{t}{\mu}\right)^{\beta-1}\exp\left\{-\left(\frac{t}{\mu}\right)^{\beta}\right\}, \tag{1.1}$$

where $t > 0$ and β, $\mu > 0$ are the parameters of shape and scale respectively. Considering a reparametrization of the form

$$Y = \ln(T) = \alpha + \sigma^* Z, \tag{1.2}$$

where $\alpha = \ln(\mu)$ and $\sigma^* = \beta^{-1}$, we have that Z is a random variable with a standard extreme value density which is given by

$$f(z) = \exp\left(z - e^z\right), -\infty < z < \infty. \tag{1.3}$$

In the presence of a know vector of covariates, $\mathbf{x}$, affecting only the location parameter, α, but not the scale parameter, σ^*, we have a location-scale regression model, so that $\ln(T)$ has constant variance (Lawless, 1982), given by

$$Y = \ln(T) = \alpha(\mathbf{x}) + \sigma^* Z. \tag{1.4}$$

A variety of functional forms for $\alpha(\mathbf{x})$ are often employed, but the most useful form is perhaps the log-linear one, for which

$$\alpha(\mathbf{x}) = \mathbf{x}\beta, \tag{1.5}$$

where $\mathbf{x} = (x_1, \ldots, x_p)$ is the $(1 \times p)$ vector of regressor or concomitant variables that act multiplicatively on the survival time and $\beta = (\beta_1, \ldots, \beta_p)^t$ is a $(p \times 1)$ vector of regression coefficients.

Observe that if $\sigma^* = 1$, in (1.4), we have an exponential distribution for the lifetime T. The error term Z, in (1.4), are usually assumed to be independent and identically distributed and could be considered from other parametric family, for example, the normal, logistic or the log-gamma distribution (Kalbfleisch and Prentice, 1980).

A different form for the location-scale regression model (1.4) is obtained if we introduce an additional error term ϵ (see also Kass, Carta and Achcar, 1999), that is

$$Y = \ln(T) = \alpha(\mathbf{x}) + \sigma^* Z + \epsilon, \tag{1.6}$$

where ϵ has a mixture of normal distributions with mean zero and density

$$\phi\left(\epsilon|\mu, \sigma^2\right) = \sum_{j=1}^{J} p_j \phi_j\left(\epsilon|\mu_j, \sigma_j^2\right). \tag{1.7}$$

Since we have a mixture model, the weights $p_j > 0$ with $\sum_{j=1}^{J} p_j = 1$, $\sum_{j=1}^{J} \mu_j p_j = 0$ and

$$\phi_j\left(\epsilon|\mu_j, \sigma_j^2\right) \propto \sigma_j^{-1} \exp\left\{-\frac{1}{2\sigma_j^2}\left(\epsilon - \mu_j\right)^2\right\}, \tag{1.8}$$

is the normal density with mean μ_j and standard deviation σ_j, $j = 1, \ldots, J$.

To avoid the well recognized identifiability problems we can impose an ordering constraint on the means ($\mu_1 < \cdots < \mu_J$) (Mengersen and Robert, 1996; Carroll, Roeder and Wasserman, 1996; Titterington, Smith and Makov, 1985). In fact, since the primary object of inference is the vector of regression coefficients β, the parameter that explains the relationship between the log survival times and the covariates, the ordering constraint are satisfied by the constraint $\sum_{j=1}^{J} \mu_j p_j = 0$. Alternatives on ordering constraint are presented in Celeux, Hurn and Robert (2000) or Stephens (2000).

Observe that if the number of mixture components in (1.7) is known, we have a parametric mixture model. Stephens (2000), Polymenis and Titterington (1998), Stephens (1997), Richardson and Green (1997) or Dey, Kuo and Sahu (1995) are recent works that address this subject.

The model (1.6) is a random effect regression model where Z could have any standard parametric distribution used for modelling lifetime data and may be used as an alternative to nonparametric regression models or semi-parametric models as the proportional hazards regression model proposed by Cox (1972).

Using censored lifetime data, we consider in this paper a Bayesian analysis of the regression model (1.6) based on Gibbs sampling with Metropolis-Hastings algorithms (Smith and Roberts, 1993) and assuming that the number of mixture components is known. The use of MCMC methods has been recently considered in the literature to analyze data with mixture models (see, for example, Diebolt and Robert, 1994; Robert, 1996 or Roeder and Wasserman, 1997). Section 2 describe a Bayesian formulation for the proposed model considering censored survival times and the introduction of two classes of latent variables used to simplify the conditional posterior needed in the Bayesian estimation. The adopted prior distributions and the conditional posterior considering the mixture of two normal densities are presented in Section 3. A numerical example using a real medical data set is presented in Section 4. Some concluding remarks in Section 5 finalize the paper.

2 Bayesian formulation using MCMC methods

In model (1.6) we can see that $(\mathbf{y}|\beta, \sigma^*, \mathbf{Z})$ has a mixture of normal distributions. Considering a prior distribution $\pi(\theta)$, $\theta = (\beta, \sigma^*, \mu, \sigma, \mathbf{p}, \mathbf{Z})$ where $\beta = (\beta_1, \ldots, \beta_p)$, $\mu = (\mu_1, \ldots, \mu_J)$, $\sigma = (\sigma_1, \ldots, \sigma_J)$, and $\mathbf{p} = (p_1, \ldots, p_J)$, the joint posterior distribution for θ is given by

$$\pi(\theta|\mathbf{y}) \propto \pi(\theta) L(\theta), \tag{2.1}$$

The likelihood function for the survival times exactly observed can be

written as

$$L(\theta) = \prod_{i=1}^{n} \left[\sum_{j=1}^{J} p_j \phi_j \left(\epsilon_i | \mu_j, \sigma_j^2 \right) \right]. \tag{2.2}$$

where $\phi_j \left(\epsilon_i | \mu_j, \sigma_j^2 \right)$ is the normal density (1.8) with $\epsilon_i = y_i - \mathbf{x}_i \beta - \sigma^* Z_i$.

For interval-censored lifetime data (Lindsey and Ryan, 1998; Sun, 1997), considering n individuals in study, let us assume that the first d out of n failure times are interval censored with the ith individual dying between t_{iL} $(y_{iL} = \ln t_{iL})$ and t_{iU} $(y_{iU} = \ln t_{iU})$. The remaining $n-d$ individuals are right censored at y_{iL}. In this case the likelihood function, subject to uninformative censoring, is given by

$$L(\theta) = \prod_{i=1}^{d} \left[F\left(\epsilon_{iU} | \mu, \sigma^2\right) - F\left(\epsilon_{iL} | \mu, \sigma^2\right) \right] \prod_{i=d+1}^{n} \left[1 - F\left(\epsilon_{iL} | \mu, \sigma^2\right) \right], \tag{2.3}$$

where $F\left(\epsilon_i | \mu, \sigma^2\right)$ is the cumulative distribution function for ϵ, derived from (1.6), that is

$$F\left(\epsilon_i | \mu, \sigma^2\right) = \sum_{j=1}^{J} p_j F_j \left(\epsilon_i | \mu, \sigma^2\right),$$

and F_j is the distribution function for the normal density ϕ_j, $j = 1, 2, \ldots, J$.

Note that the likelihood (2.3) includes the grouped data where a number of individuals die within the same interval because we can set y_{iL} to be identical for several i; similarly for y_{iU}.

Following Kuo and Peng 1995, to simplify the conditional distributions needed for the Gibbs sampling algorithm, we introduce latent variables, data augmentation technique (Tanner and Wong, 1987), that allow us to write the likelihood (2.2) and (2.3) as a product of component models for i.i.d. observations.

This is given by augmenting the original data with two classes of latent variables. One is the truncated random variable W that allows us to consider the likelihood for i.i.d. uncensored observations and the other is the index variable, denoted by V, that convert the mixture model to a model of independent components (see, Kuo and Peng 1995).

If the ith individual is censored between t_{iL} and t_{iU}, we can generate a latent variable w_i from the truncated density $\frac{f(w_i)}{F(t_{iU}) - F(t_{iL})} I\left[t_{iL} < w_i < t_{iU}\right]$, where $f(.)$ is the density for the survival time T and $F(.)$ is the cumulative distribution. From the inverse transform method we have that

$$w_i = F^{-1} \left\{ F\left(\epsilon_{iL} | \mu, \sigma^2\right) + U \left[F\left(\epsilon_{iU} | \mu, \sigma^2\right) - F\left(\epsilon_{iL} | \mu, \sigma^2\right) \right] \right\}, \tag{2.4}$$

where $U \sim Uniform\,(0,1)$ and F^{-1} is the inverse function of F.

Similarly, if the ith individual is right-censored at t_{iL}, the latent variable w_i can be generated from the truncated density $\frac{f(w_i)}{1-F(t_{iL})} I\left[t_{iL} < w_i < \infty\right]$ by setting

$$w_i = F^{-1}\left\{F\left(\epsilon_{iL}|\mu,\sigma^2\right) + U\left[1 - F\left(\epsilon_{iL}|\mu,\sigma^2\right)\right]\right\}. \tag{2.5}$$

Then, according to the censoring mechanism, for each i, w_i is generated independently and the likelihood function can be written as

$$L\left(\theta\right) = \prod_{i=1}^{n}\left[\sum_{j=1}^{J} p_j \phi_j\left(\epsilon_i|\mu_j,\sigma_j^2\right)\right], \tag{2.6}$$

where, $\epsilon_i = w_i - \mathbf{x}_i\beta - \sigma^* Z_i$ and $w_i = \ln\left(t_i\right)$ for the uncensored survival times.

We get similar latent variable for any other censoring mechanism (left or right censoring with uncensored data; or a combination of both).

Taking only $J = 2$ normal distributions in the mixture model (1.7), the other class of latent variables is given by $\mathbf{v}_i = \left(v_{i1}, v_{i2}\right)$, $i = 1, 2, \ldots, n$, where $v_{i1}|\theta, w_i, \mathbf{x}_i \sim Be\left(1, h_{i1}\right)$, a Bernoulli distribution with parameter h_{i1}, that is

$$h_{i1} = \frac{p_1\phi_1\left(\epsilon_i|\mu_1,\sigma_1^2\right)}{\sum_{j=1}^{2} p_j\phi_j\left(\epsilon_i|\mu_j,\sigma_j^2\right)}, \tag{2.7}$$

and $\pi\left(\mathbf{v}_i\right) \propto h_{i1}^{v_{i1}}\left(1-h_{i1}\right)^{v_{i2}}$, $v_{i2} = 1$ with probability h_{i1}, $v_{i1} = 0$ with probability $\left(1-h_{i1}\right)$ and $v_{i1} + v_{i2} = 1$. Thus, for each observation we have

$$\pi\left(\mathbf{v}_1,\ldots,\mathbf{v}_n|\theta,\mathbf{w},\mathbf{x}\right) \propto \frac{\prod_{i=1}^{n}\prod_{j=1}^{2}\left[p_j\phi_j\left(\epsilon_i|\mu_j,\sigma_j^2\right)\right]^{v_{ij}}}{\prod_{i=1}^{n}\left[\sum_{j=1}^{2} p_j\phi_j\left(\epsilon_i|\mu_j,\sigma_j^2\right)\right]}. \tag{2.8}$$

When $J > 2$, the vector of latent variables $\mathbf{v}_i$ follows a multinomial distribution.

Combining equation (2.6) with (2.8) the joint posterior distribution for $(\theta, \mathbf{w}, \mathbf{v})$ is given by

$$\pi\left(\theta,\mathbf{w},\mathbf{v}|\mathbf{x}\right) \propto \pi\left(\theta\right)\left\{\prod_{i=1}^{n}\prod_{j=1}^{2}\left[p_j\phi_j\left(\epsilon_i|\mu_j,\sigma_j^2\right)\right]^{v_{ij}}\right\}. \tag{2.9}$$

Similar posterior density is obtained when $J > 2$ and generically, to generate samples of the joint posterior distribution (2.9), use the Gibbs sampling algorithm. That is, starting with the initial values $\theta^{(0)} = \left(\theta_1^{(0)},\ldots,\theta_J^{(0)}\right)$ follow the steps:

i) Generate samples $\mathbf{w}^{(1)} = \left(\mathbf{w}_1^{(1)}, \ldots, \mathbf{w}_n^{(1)}\right)$ from (2.4) or (2.5);

ii) Generate samples $\mathbf{v}^{(1)} = \left(\mathbf{v}_1^{(1)}, \ldots, \mathbf{v}_n^{(1)}\right)$ where $\mathbf{v}_i^{(1)} = \left(v_{i1}^{(1)}, v_{i2}^{(1)}\right)$ from a Bernoulli distribution $Be\,(1, h_{i1})$ where h_{i1} is given in (2.7);

(iii) Generate samples of θ from the conditional distributions

$$\pi\left(\theta_1 | \theta_{-\theta_1}^{(0)}, \mathbf{v}^{(1)}, \mathbf{w}^{(1)}, \mathbf{x}\right)$$
$$\vdots$$
$$\pi\left(\theta_J | \theta_{-\theta_J}^{(1)}, \mathbf{v}^{(1)}, \mathbf{w}^{(1)}, \mathbf{x}\right),$$

where $\theta_{-\theta_i} = (\theta_1, \ldots, \theta_{i-1}, \theta_{i+1}, \ldots, \theta_J)$ and $\theta = (\beta, \sigma^*, \mu, \sigma, \mathbf{p}, \mathbf{Z})$.

3 Conditional posterior distributions

Consider the regression model (1.6) with a mixture of two normal distributions for the error term ϵ and a vector of covariates $\mathbf{x} = (x_1, x_2, \ldots, x_p)$ such that $\mathbf{x}\beta = \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p$.

Assuming that the components of θ are independent, consider the following prior densities

$$\begin{array}{ll} \pi\left(\beta_0, \beta_1, \ldots, \beta_p\right) \propto constant & p_1 \sim Beta\,(a, b) \\ \sigma_j^2 \sim I\Gamma\,(c_j, d_j) & \sigma^* \sim \Gamma\,(e, f) \\ \mu_j \sim N\,(g, h^2) & Z_i \sim Extreme\ Value\,(0, 1) \end{array} \tag{3.1}$$

where a, b, e, f, g, h, c_j, d_j are known hyperparameters and $Beta\,(a, b)$ generically denotes a Beta distribution with mean $a\,(a+b)^{-1}$ and variance $ab\left[(a+b)^2\,(a+b+1)\right]^{-1}$; $I\Gamma\,(a, b)$ denotes an inverse Gamma distribution with mean $b\,(a-1)^{-1}$ and variance $b^2\left[(a-1)^2\,(a-2)\right]^{-1}$; $\Gamma\,(a, b)$ denotes a Gamma distribution with mean ab^{-1} and variance ab^{-2} and $N\,(\mu, \sigma^2)$ generically denotes a normal distribution with mean μ and variance σ^2. In a different way, we also could consider a prior distribution for the hyperparameters.

With the introduction of latent variables $\mathbf{w}$ and $\mathbf{v}$ (see, section 2), where $\mathbf{w} = \mathbf{y}$ for the uncensored survival times, we have from (2.9), with $\pi\,(\theta)$ given in (3.1), the conditional distributions needed for the Gibbs sampling. These conditional distributions considering $J = 2$ are given by

i) $(\beta_0|\theta_{-\beta_0}, \mathbf{v}, \mathbf{w}, \mathbf{Z}, \mathbf{x}) \sim$
$N\left\{\frac{\sigma_1^2\sum_{i=1}^n v_{i2}(r_i-\mu_2)+\sigma_2^2\sum_{i=1}^n v_{i1}(r_i-\mu_1)}{n_2\sigma_1^2+n_1\sigma_2^2}, \frac{\sigma_1^2\sigma_2^2}{n_2\sigma_1^2+n_1\sigma_2^2}\right\}$
where $n_1 = \sum_{i=1}^n v_{i1}$, $n_2 = n - n_1$ and $r_i = w_i - \sum_{s=1}^k \beta_s x_{si} - \sigma^* Z_i$, $i = 1, \ldots, n$;

ii) $\left(\beta_{(s)}|\theta_{-\beta_{(s)}}, \mathbf{v}, \mathbf{w}, \mathbf{Z}, \mathbf{x}\right) \sim$
$N\left\{\frac{\sigma_1^2\sum_{i=1}^n v_{i2}x_{si}(r_{si}-\mu_2)+\sigma_2^2\sum_{i=1}^n v_{i1}x_{si}(r_{si}-\mu_1)}{a_{2s}\sigma_1^2+a_{1s}\sigma_2^2}, \frac{\sigma_1^2\sigma_2^2}{a_{2s}\sigma_1^2+a_{1s}\sigma_2^2}\right\}$
where $s = 1, \ldots, k$, $a_{js} = \sum_{i=1}^n v_{ij}x_{si}^2$, $j = 1, 2$, $r_{si} = w_i - \beta_0 - \sum_{l=1}^k \beta_l x_{li} - \sigma^* Z_i$ and $\beta_{(s)=} (\beta_1, \ldots, \beta_{s-1}, \beta_{s+1}, \ldots, \beta_k)$;

iii) $(p_1|\theta_{-p_1}, \mathbf{v}, \mathbf{w}, \mathbf{Z}, \mathbf{x}) \sim B(a + n_1, b + n_2)$;

iv) $(\mu_1|\theta_{-\mu_1}, \mathbf{v}, \mathbf{w}, \mathbf{Z}, \mathbf{x}) \sim N\left(a_1 - \frac{p_1 b_1 \sum_{j=1}^2 p_j a_j}{\sum_{j=1}^2 p_j^2 b_j}, b_1\left(1 - \frac{p_1^2 b_1}{\sum_{j=1}^2 p_j^2 b_j}\right)\right)$
where
$a_j = \frac{g\sigma_j^2+h^2\sum_{i=1}^n v_{ij}r_i}{\sigma_j^2+n_j h^2}$, $b_j = \frac{h^2\sigma_j^2}{\sigma_j^2+n_j h^2}$ and $r_i = w_i - \beta_0 - \sum_{j=1}^k \beta_{ji} - \sigma^* Z_i$
(observe that we are also conditioning on the constraint $\sum_{j=1}^2 \mu_j p_j = 0$, Roeder and Wasserman (1996));

v) $\left(\sigma_j^2|\theta_{-\sigma_j^2}, \mathbf{v}, \mathbf{w}, \mathbf{Z}, \mathbf{x}\right) \sim$
$I\Gamma\left(c_j + \frac{n_j}{2}; d_j + \frac{1}{2}\sum_{i=1}^n v_{ij}^2 (w_i - \beta_0 - \beta_1 x_i + \sigma^* Z_i - \mu_j)^2\right)$;

vi) $\pi(\sigma^*|\theta_{-\sigma^*}, \mathbf{v}, \mathbf{w}, \mathbf{x}) \propto \sigma^{*e-1} e^{-f\sigma^*} \psi_1(\theta)$ where
$\psi_1(\theta) = \exp\left\{-\frac{1}{2}\sum_{i=1}^n \sum_{j=1}^2 v_{ij}\left(\frac{w_i-\beta_0-\sum_{s=1}^k \beta_s x_{si}+\sigma^* Z_i-\mu_j}{\sigma_j}\right)^2\right\}$;

vii) $\pi(Z_i|\theta_{-z_i}, \mathbf{v}, \mathbf{w}, \mathbf{Z}, \mathbf{x}) \propto \exp\left(Z_i - e^{Z_i}\right)\psi_2(\theta)$ where
$\psi_2(\theta) = \exp\left\{-\frac{1}{2}\sum_{j=1}^2 v_{ij}\left(\frac{w_i-\beta_0-\sum_{s=1}^k \beta_s x_{si}+\sigma^* Z_i-\mu_j}{\sigma_j}\right)^2\right\}$ for $i = 1, \ldots, n$.

From (vi) and (vii) we observe that the parameters $\sigma^*, Z_1, \ldots, Z_n$ should be generated using, for example, the Metropolis-Hastings algorithm since their conditional distributions are not of a known form. This is also true if we consider a normal or a logistic prior distribution from Z_i in (3.1).

4 An example with right censored data

Table 1 gives the remission survival times with additional information about white blood cell count for 21 patients suffering from leukemia on a treatment group (Freireich *et al.*, 1963). The concomitant variable is the *log*(*white blood count*), an important predictor of survival in leukemia patients. In this experiment it was observed 9 (12) uncensored (right censored) observations (observations with + are right censored). The response variable is survival time in weeks.

Table 1
Remission Survival Times and log(*white blood count*).

t_i	x_i	t_i	x_i	t_i	x_i
6	2.31	6	4.06	6	3.28
6^+	3.20	7	4.43	9^+	2.80
10^+	2.96	10^+	2.70	11^+	2.60
13	2.88	16	3.60	17^+	2.16
19^+	2.05	20^+	2.01	22	2.32
23	2.57	25^+	1.78	32^+	2.20
32^+	2.53	34^+	1.47	35^+	1.45

Considering the log-linear model (1.6) with a standard extreme value density (1.3) and a mixture of two normal distributions for the error and the prior (3.1) with hyperparameters $a = 1$, $b = 1$, $c_1 = 0.1$, $c_2 = 0.1$, $d_1 = 0.1$, $d_2 = 0.1$, $e = 0.1$, $f = 0.1$, $g = 0.5$ and $h = 50$, we have in Table 2 the posterior summaries for the model parameters.

Two separate Gibbs chains each of which ran for $S = 51000$ iterations were used. In order to diminish the effect of the starting parameters values we discarded the first 1000 elements of each chain. Convergence of the Gibbs algorithm was observed using diagnostics procedures available in CODA package (Best, Cowles and Vines 1995). For each parameter we considered every 10th draw and so we finally got a sample of size $S = 10000$. The results were generated using Ox matrix language (see, Doornik, 1999).

We also have in Table 2, the posterior summaries for the parameters considering the standard normal density $f(z) \propto e^{\left(-\frac{1}{2}\sigma^2\right)}$ and the standard logistic density $f(z) = e^z (1 + e^z)^{-2}$ for Z_i in the log-linear model (1.6).

Table 3 shows the posterior means for $Z_1, \ldots, Z_{21}$ obtained by the combined chains.

Table 2
Posterior Summaries — Model (1.6).

Density for Z_i	Parameter	Mean	S.D.	95% Credible Interval	
Standard	β_0	3.2401	0.2130	2.8577	3.6981
Extreme Value	β_1	−0.8173	0.2428	−1.3013	−0.3493
	σ^*	0.5159	0.0977	0.3396	0.7120
Standard	β_0	3.0458	0.3771	2.4881	4.0000
Normal	β_1	−0.8188	0.3415	−1.5768	−0.2495
	σ^*	0.4647	0.1013	0.2990	0.6981
Standard	β_0	3.2108	0.3521	2.6646	4.0540
Logistic	β_1	−0.9251	0.3614	−1.7546	−0.3443
	σ^*	0.2494	0.0581	0.1563	0.3860

The sum of squares of the residuals, $\hat{\varepsilon}_i = y_i - \hat{y}_i$, taking the uncensored observations is given by $\sum_{i=1}^{21} \hat{\epsilon}_i^2 = 0.5594$, assuming the standard extreme value density for Z_i. Assuming the normal density for Z_i we have $\sum_{i=1}^{21} \hat{\epsilon}_i^2 = 0.8398$ and $\sum_{i=1}^{21} \hat{\epsilon}_i^2 = 0.4211$ for the logistic density. Based on these quantities, the best model for the remission times of Table 1 is given by the log-linear model (1.6) with the logistic density for Z_i.

In Table 4 we have the posterior means and the standard deviation considering the log-linear model (1.4) with $Z_i \sim$ extreme-value$(0,1)$, $Z_i \sim$ normal$(0,1)$ and $Z_i \sim$ logistic$(0,1)$. In those cases, the sum of squares of the residuals, $\hat{\varepsilon}_i = y_i - \hat{y}_i$, are 5.1204 (standard extreme value), 3.9158 (standard normal) and 3.8450 for Z_i with a standard logistic distribution.

We observe that using the linear model (1.6) with a mixture of two normal distributions for the error, we have better fit for the lifetime data (small values for the sum of squares of the residuals).

Considering some Bayesian model selection techniques such as the marginal likelihood (Raftery, 1996; Chib, 1995) or those presented in Dey, Kuo and Sahu (1995) we have that the best model is the log-linear model (1.6) with a mixture of two logistic densities for the error term.

As a final comment, the hyperparameters in (3.1) and the number of mixture components $J = 2$ were selected analyzing the residuals after the standard models were fitted (extreme value, normal and logistic) that is, models without the additional error term. For some parameters, diffuse proper priors were considered so that the numerical results were focused on the likelihood.

We omitted the estimates for the mixtures components parameters $(\mu_1, \mu_2, \sigma_1, \sigma_2)$ since they were not used to calculate the predicted values or some others quantities of interest.

Table 3
Posterior Means for Z_i.

	Density for Z_i		
$\hat{Z}_i$	Extreme Value	Normal	Logistic
Z_1	−3.0953	−5.3434	−2.0980
Z_2	−0.5020	−0.2716	−0.0808
Z_3	−1.6147	−2.3853	−1.0289
Z_4	0.1573	0.9787	0.5059
Z_5	−1.1695	−1.6915	−0.6625
Z_6	−0.8156	−1.1210	−0.3445
Z_7	0.4357	1.2985	0.8223
Z_8	−0.6861	−1.0185	−0.2128
Z_9	−0.2603	−0.2286	0.1592
Z_{10}	−1.3552	0.0683	0.0175
Z_{11}	−1.5838	0.0619	0.0428
Z_{12}	−1.7373	0.0786	0.0717
Z_{13}	−1.2560	0.0773	0.0834
Z_{14}	−0.9713	0.1232	0.1190
Z_{15}	−0.8542	0.1418	0.1373
Z_{16}	−0.7528	0.1584	0.1583
Z_{17}	−0.6417	0.1719	0.1934
Z_{18}	0.9546	0.8160	0.6540
Z_{19}	1.5087	1.3379	0.9347
Z_{20}	−0.5727	0.2557	0.2370
Z_{21}	−0.4015	0.2492	0.2497

Table 4
Posterior Summaries — Model (1.4).

Density for Z_i	Parameter	Mean	S.D.
Standard	β_0	3.2692	0.1640
Extreme Value	β_1	−0.8419	0.1674
	σ^*	0.3947	0.1011
Standard	β_0	3.0904	3.0904
Normal	β_1	−0.8645	0.2213
	σ^*	0.5923	0.1418
Standard	β_0	3.0863	0.1712
Logistic	β_1	−0.8380	0.1930
	σ^*	0.3124	0.0839

5 Concluding remarks

The use of the proposed log-linear model (1.6) with a mixture of two normal distributions for the error could be a good alternative to analyze uncensored or censored survival data. In the example presented in Section 4, we observe better fit of the log-linear model (1.6) considering a mixture of two normal distributions in comparison with the usual log-linear model (1.4) .

It is important to point out that the use of MCMC methods to get the posterior summaries of interest does not require sophisticated computational expertise and this approach could be extended to survival data with many covariates and also including more than two normal distributions for the mixture distribution of the error ϵ in the log-linear model (1.6) . A more sophisticated treatment of the proposed model could consider the number of the components of the mixture densities unknown and yet consider the incidence probabilities, the p_j parameters, covariate dependent using some link function (see, Kuo and Peng, 1995). This will be addressed in subsequent work.

Acknowledgements

The authors are thankful to the referees for some useful suggestions which improved the paper. The work is partially supported by the Brazilian institutions CAPES and CNPq.

References

Best, N. G., Cowles, M. K. and Vines, S. K. (1995). *CODA: Convergence diagnosis and output analysis software for Gibbs sampling output, Version 0.3*. MRC Biostatistics Unit, Cambridge.

Best, N. G., Spiegelhalter, D. J., Thomas, A. and Brayne, C. E. G. (1996). Bayesian analysis of realistically complex models. *Journal of the Royal Statistical Society*, A, **159**, 323-342.

Carroll, J. R., Roeder K. and Wasserman L. (1996). Flexible parametric measurement error models. Technical Reports #648, Department of Statistics, Carnegie Mellon University.

Celeux, G., Hurn, M., Robert, C. P. (2000). Computational and inferential difficulties with mixture posterior distributions. *Journal of the American Statistical Association*, **95**, 957-970.

Chib, S. (1995). Marginal likelihood from the Gibbs output. *Journal of American Statistical Association*, **90**, 1313-1321.

Cox, D. R. (1972). Regression models and life tables (with discussion). *Journal of the Royal Statistical Society*, B, **24**, 406-423.

Dey, D. K., Kuo, L. and Sahu, S. K. (1995). A Bayesian predictive approach to determining the number of components in a mixture distribution. *Statistics and Computing*, **5**, 297-305.

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

Doornik, J.A. (1999). *Ox: An Object-Oriented Matrix Language (3rd edition*). Timberlake Consultants Press: London.

Freireich, E. J., Gehan, E., Frei, E., Schroeder, L. R. Wolman, I. J., Anbari, R., Burgert, E. O., Mills, S. D., Pinkel, D., Selawry, O. S., Moon, J. H., Gendel, B. R., Spurr, C. L., Storrs, R., Haurani, F., Hoogstraten, B. and Lee, S. (1963). The effect of 6-Mercaptopurine on the duration of steroid-induced remissions in acute leukemia: A model for evaluation of other potentially used therapy. *Blood*, **21**, 699-716.

Gelfand, A. (1996). Model Determination Using Sampling-Based Methods. *In Practical Markov Chain Monte Carlo* (eds. W. R. Gilks, S. Richardson, D. J. Spiegelhalter). Chapman and Hall: London.

Gilks, W. R., Thomas, A. and Spiegelhater, D. J. (1994). A language and program for complex Bayesian modelling. *The Statistician*, **43**, 169-178.

Kass, R. E., Carta, R., Achcar, J. A. (1999). Regression model for lifetime data with extreme value mixture of normal distributions for the error (manuscript). Department of Statistics, Carnegie Mellon University.

Kalbfleisch, J. D., Prentice, R. L. (1980). *The statistical analysis of failure time data*. Wiley: New York.

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

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

Lindsey, J. C. and Ryan, L. M. (1998). Tutorial in biostatistics methods for interval-censored data. *Statistics in Medicine*, **17**, 219-238.

Mengersen, K. and Robert, C. P. (1996). *Testing for mixtures: A Bayesian entropic approach.* In Bayesian Statistics 5 (eds. J. O. Berger, J. M. Bernardo, A. P. Dawid, D. V. Lindley A. F. M. Smith. Oxford University Press: London.

Polymenis, A. and Titterington, D. M. (1998). On the determination of the number of components in a mixture. *Statistics and Probability Letter*, **38**, 295-298.

Raftery, A. E. (1996). Hypothesis testing and model selection. *In Practical Markov Chain Monte Carlo* (eds. W. R. Gilks, S. Richardson, D. J. Spiegelhalter). Chapman and Hall: London.

Richardson, S. and Green, P. J. (1997). On Bayesian analysis of mixtures with an unknown number of components (with discussion). *Journal of the Royal Statistical Society*, B, **59**, 731-792.

Robert, C. P. (1996). Mixtures of distributions: inference and estimation. *In Practical Markov Chain Monte Carlo*, (eds. W. R. Gilks, S. Richardson, D. J. Spiegelhalter). London: Chapman and Hall.

Roeder K. and Wasserman L. (1997). Practical Bayesian density estimation using mixtures of normals. *Journal of the American Statistical Association*, **92**, 894-902.

Smith, A. F. M. and Roberts, G. (1993). Bayesian computation via the Gibbs sampler and related Markov Chain Monte Carlo Methods (with discussion). *Journal of the Royal Statistical Society*, B, **55**, 3-23.

Spiegelhalter, D. J., Thomas, A., Best, N. G. and Gilks W. R. (1995c). BUGS: Bayesian inference using Gibbs sampling, Version 0.50. MRC Biostatistics Unit Cambridge.

Spiegelhalter, D. J., Thomas, A. and Best, N. G. (1995b). Computation on Bayesian graphical models. *In Bayesian Statistics 5.* Clarendon Press, Oxford, UK.

Spiegelhalter, D. J, Best, N. G., Gilks, W. R. and Inskip, H. (1995a). Hepatits: A case study in MCMC Method. *In Practical Markov Chain Monte Carlo*, (eds. W. R. Gilks, S. Richardson, D. J. Spiegelhalter). London: Chapman and Hall.

Stephens, M. (1997). Bayesian methods for mixture of normal distributions. unpublished Ph.D. thesis, University of Oxford.

Stephens, M. (1998). Bayesian analysis of mixtures with an unknown number of components - An alternative to reversible jump methods. *The Annals of Statistics,* **28**,40-74.

Stephens, M. (2000). Dealing with label switching in mixture models. *Journal of the Royal Statistical Society*, B, **62**, 795-809.

Sun, J. (1997). Regression analysis of interval-censored failure time data. *Statistics in Medicine*, **16**, 497-504.

Tanner, M., Wong, W. (1987). The Calculations of posterior Distributions by Data Augmentation. *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*. Wiley: New York.