RT-MAE-8810

SOME APPLICATIONS OF BAYESIAN METHODS IN ANALYSIS OF LIFE DATA

by

J.A. Achcar, H. Bolfarine

and

L.R. Pericchi

Palavras Chaves: Censored data, generalized Gamma
(Key words) regression model, power transformation, predictive density, design of clinical trials.

Classificação AMS: 62F15
(AMS Classification) 65J99

- Abril de 1988 -

SOME APPLICATIONS OF BAYESIAN METHODS IN ANALYSIS OF LIFE DATA

Jorge Alberto Achcar
Universidade Federal de São Carlos, BRAZIL

Heleno Bolfarine
Universidade de São Paulo, B

Luiz Raul Pericchi
University Simon Bolivar, VEN

1. INTRODUCTION

Usually, researches in the medical area should analyse data given by the survival times T of patients submitted to a new treatment. These life data can be censored.

In this situation, researchers usually use nonparametrical techniques. One of the most popular of these nonparametrical methods is the product-limit estimator of Kaplan and Meier (1958) of the survival function $S(t) = P(T > t)$. Also, it is common the use of nonparametrical tests in the comparison of two or more treatments.

With covariates, the proportional hazards model of Cox (1972) is very popular in the medical area. We can also mention, some nonparametrical linear models as proposed by Buckley and James (1979), Miller (1976) and Koul, Susarla and Van Ryzin (1981).

In the parametrical approach, we usually use graphical methods in the identification and checking of a proposed model and asymptotical results for the inferences of interest.

Bayesian methods are not very common in the analysis of medical data, but they are used in engineering applications (see for example, Martz and Waller, 1982).

As an alternative to these classical existing methods, we present Bayesian methods in the identification, inferences and checking of a proposed model in survival analysis. We also present a Bayesian application in the design of clinical trials.

2. LOG-LINEAR MODEL WITH A GENERALIZED GAMMA DISTRIBUTION FOR THE ERROR

Let T be a nonnegative random variable denoting the survival time of a patient.

The generalized gamma distribution has the following density:

$$g(t) = \frac{\theta t^{\theta k-1}}{\Gamma(K)\alpha^{\theta k}} \exp\{-(\frac{t}{\alpha})^{\theta}\} \qquad (1)$$

where $\alpha, k, \theta > 0$.

We observe that:

(i) If $\theta = k = 1$, we have an exponential distribution for T.

(ii) If $k = 1$, we have a Weibull distribution for T.

(iii) If $\theta = 1$, we have a Gamma distribution for T.

(iv) If $k \to \infty$, we have a log-normal distribution for T.

Some authors (Prentice (1974), Farewell and Prentice (1977), Lawless (1980)) consider the following reparametrization of

(1):

$$Y = \ln(T) = \mu + \sigma W \qquad (2)$$

where $\mu = \ln(\alpha)$, $\sigma = 1/\theta$ and the error variable W has the density:

$$g(w) = \frac{1}{\Gamma(K)} \exp\{kw - e^{w}\}. \qquad (3)$$

A generalization of the log-lienar model (2) is given by:

$$Y = \ln(T) = \mu + \beta X + \sigma W \qquad (4)$$

where μ is a location parameter, σ is a scale parameter and β is a regression parameter associated with the covariate X.

Considering $T_1, T_2, \ldots, T_n$, a random sample of size n, the likelihood function for μ, σ, β and k is given by:

$$\ell(\mu,\sigma,\beta,k) = \prod_{i=1}^{n} \{\sigma^{-1} g(w_i)\} \qquad (5)$$

$$= \{\frac{1}{\sigma\Gamma(k)}\}^{n} \exp\{k \sum_{i=1}^{n} w_i - \sum_{i=1}^{n} e^{w_i}\}$$

where $w_i = (y_i - \mu - \beta x_i)/\sigma$, $i = 1,2,\ldots,n$.

That is,

$$\ell(\mu,\sigma,\beta,k) = \{\frac{1}{\sigma\Gamma(k)}\}^{n} . \exp\{\frac{k}{\sigma} \sum_{i=1}^{n} y_i - \frac{kn\mu}{\sigma} - \frac{\beta k \sum_{i=1}^{n} x_i}{n}\}. \qquad (6)$$

$$. \exp\{-e^{-\mu/\sigma} \sum_{i=1}^{n} e^{(y_i - \beta x_i)/\sigma}\}.$$

The Jeffreys prior for μ, σ, β and k is given by:

$$\pi(\mu,\sigma,\beta,k) = \pi(\mu,\sigma,\beta|k)\pi_o(k) \quad (7)$$

$$\propto \{\det I_k(\mu,\sigma,\beta)^{1/2}\}\pi_o(k)$$

where $I_k(\mu,\sigma,\beta)$ is the Fisher information matrix given k.

Since $E(W) = \psi(k) = \frac{d\ell n\Gamma(k)}{dk}$ is the digamma function and $\text{var}(W) = \psi^{(1)}(k) = \frac{d^2\ell n\Gamma(k)}{dk^2}$ is the trigamma function, the Fisher information matrix given k is given by

$$I_k(\mu,\sigma,\beta) = \{a_{ij}\},\ i,j = 1,2,3,$$

where:

$$a_{11} = \frac{nk}{\sigma^2}$$

$$a_{12} = a_{21} = \frac{k \sum_{i=1}^{n} x_i}{\sigma^2}$$

$$a_{13} = a_{31} = \frac{n}{\sigma^2}(1 + k\psi(k))$$

$$a_{22} = \frac{k \sum_{i=1}^{n} x_i^2}{\sigma^2}$$

$$a_{23} = a_{32} = \frac{\sum_{i=1}^{n} x_i}{\sigma^2}(1 + k\psi(k))$$

and $$a_{33} = \frac{n}{\sigma^2}[1 + k\,\psi^{(1)}(k+1) + k(\psi(k+1)^2].$$

Assuming $\psi^{(1)}(k+1) \simeq 1/(k+1)$, and a locally uniform prior for $\ln(k)$, we find:

$$\pi(\mu,\sigma,\beta,k) \propto \sigma^{-3}\,\xi(k) \tag{8}$$

where $\sigma,k > 0$; $-\infty < \mu,\ \beta < \infty$ and $\xi(k) = (1 + \frac{k}{k+1})^{1/2}$.

The marginal posterior density for σ and k using Laplace's methods (see Tierney and Kadane, 1986) is given by:

$$\pi(\sigma,k|\text{data}) \propto$$

$$\propto \frac{\chi^{*1/2}\,\xi(k)k^{(n-1)/2}\,n^{nk-1/2}\,e^{ka_1/\sigma}\,\hat{v}^{*\frac{ka_2}{\sigma}-1}}{\sigma^{n+2}\{\sum_{i=1}^{n} e^{y_i/\sigma}\,\hat{v}^{*x_i/\sigma}\}^{kn}}$$

where $a_1 = \sum_{i=1}^{n} y_i$; $a_2 = \sum_{i=1}^{n} x_i$, $\hat{v}^*$ satisfies the relationship,

$$ka_2 - \sigma = \frac{kn \sum_{i=1}^{n} x_i\, e^{y_i/\sigma}\,\hat{v}^{*x_i/\sigma}}{\sum_{i=1}^{n} e^{y_i/\sigma}\,\hat{v}^{*x_i/\sigma}}$$

and $\chi^* = -1 \Big/ \left.\frac{d^2 L_{k,\sigma}(v)}{d v^2}\right|_{v=\hat{v}^*}$

where $\frac{d^2 L_{k,\sigma}(v)}{d v^2} = \frac{1}{n}\{-(\frac{ka_2}{\sigma} - 1)\frac{1}{v^2} -$

$$-kn\left[\frac{\left(\sum_{i=1}^{n} e^{y_i/\sigma}\frac{x_i}{\sigma}\left(\frac{x_i}{\sigma}-1\right)v^{\frac{x_i}{\sigma}-2}\right)\left(\sum_{i=1}^{n} e^{y_i/\sigma}v^{x_i/\sigma}\right)-\left(\sum_{i=1}^{n} e^{y_i/\sigma}\frac{x_i}{\sigma}v^{\frac{x_i}{\sigma}-1}\right)^2}{\left(\sum_{i=1}^{n} e^{y_i/\sigma}v^{x_i/\sigma}\right)^2}\right]>$$

(see Achcar and Bolfarine, 1986).

In Table 1, we have the survival times of patients in two treatment groups (data in Lee, 1980).

TABLE 1 - SURVIVAL DATA IN WEEKS

TREATMENT 1 (n_1=40)	5, 10, 17, 32, 32, 33, 34, 36, 43, 44, 44, 48, 48, 61, 64, 65, 65, 66, 67, 68, 82, 85, 90, 92, 92, 102, 103, 106, 107, 114, 114, 116, 117, 124, 139, 142, 132, 151, 148, 195.
TREATMENT 2 (n_2=40)	20.9, 32.2, 33.2, 39.4, 40.0, 46.8, 54.3, 57.3, 58.0, 59.7, 61.1, 61.4, 66.0, 66.3, 67.4, 68.5, 69.9, 72.4, 73.0, 73.2, 88.7, 89.3, 91.6, 93.1, 94.2, 97.7, 101.6, 101.9, 107.6, 108.0, 109.7, 110.8, 114.1, 117.5, 119.2, 120.3, 133.0, 133.8, 163.3, 165.1

Considering x=0 for treatment 1 and x=1 for treatment 2, we have $a_1 = \sum_{i=1}^{80} y_i = 341.4$, $a_2 = \sum_{i=1}^{80} x_i = 40$, $n_1 = n_2 = 40$ and n=80.

From (9), the marginal posterior density for k and σ is given by:

$\pi(k,\sigma|\text{data}) \propto$

$$\propto \frac{\sqrt{2k+1}\ (40k+\sigma)^{40k-1/2}\ (40k-\sigma)^{40k-1/2}\ e^{341.4\ k/\sigma}}{\sqrt{k+1}\ k^{80(k-1/2)}\ \sigma^{81}\left(\sum_{i=1}^{40} T_{1i}^{1/\sigma}\right)^{40k}\left(\sum_{i=1}^{40} T_{2i}^{1/\sigma}\right)^{40k}}.$$

In figure 1, we have contours of posterior density for k and σ. From figure 1, we observe that the mode of $\pi(k,\sigma|\text{data})$ is approximately $\hat{k}=1.0$ and $\hat{\sigma}=0.4$. Assuming k=1.0 and $\sigma=0.4$ known, and a locally prior density for μ and β, the marginal posterior density for β is given by:

$$\pi(\beta|\text{data}) \propto \frac{\exp\{-100\beta\}}{\{3736745 + 3435657\ e^{-2.5\beta}\}^{80}} \qquad (11)$$

where $-\infty < \beta < \infty$.

In figure 2, we observe that the marginal posterior density for β has a mode close to zero, that is, we conclude that both treatments have same effect on the survival times of the patients.

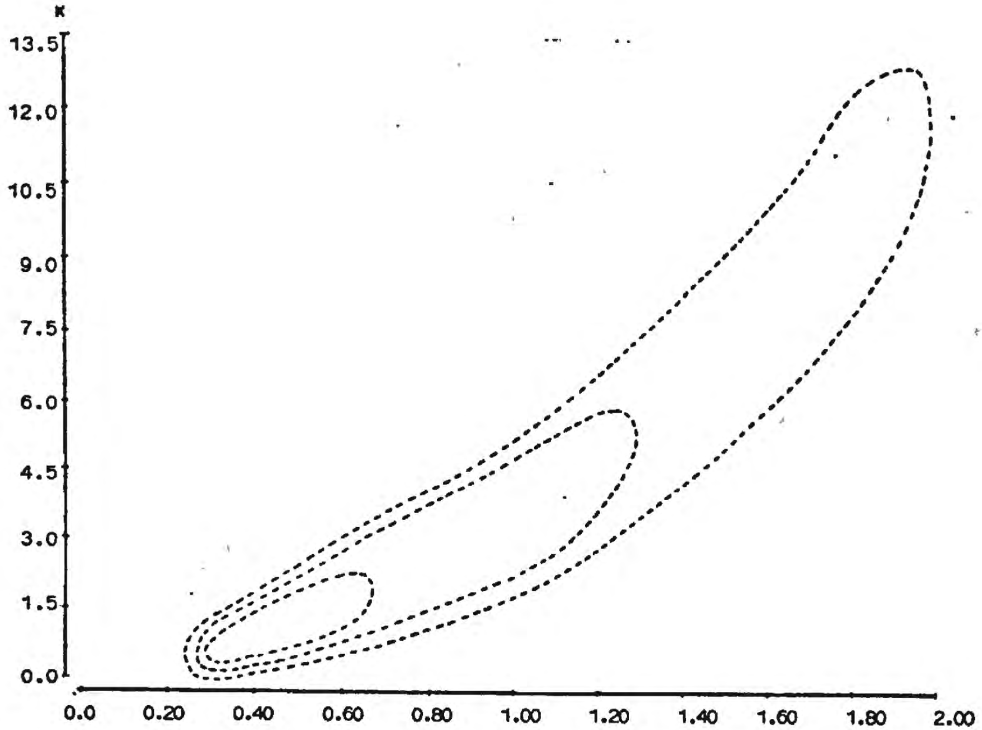

FIGURE 1 - CONTOURS OF POSTERIOR DENSITY FOR k AND σ

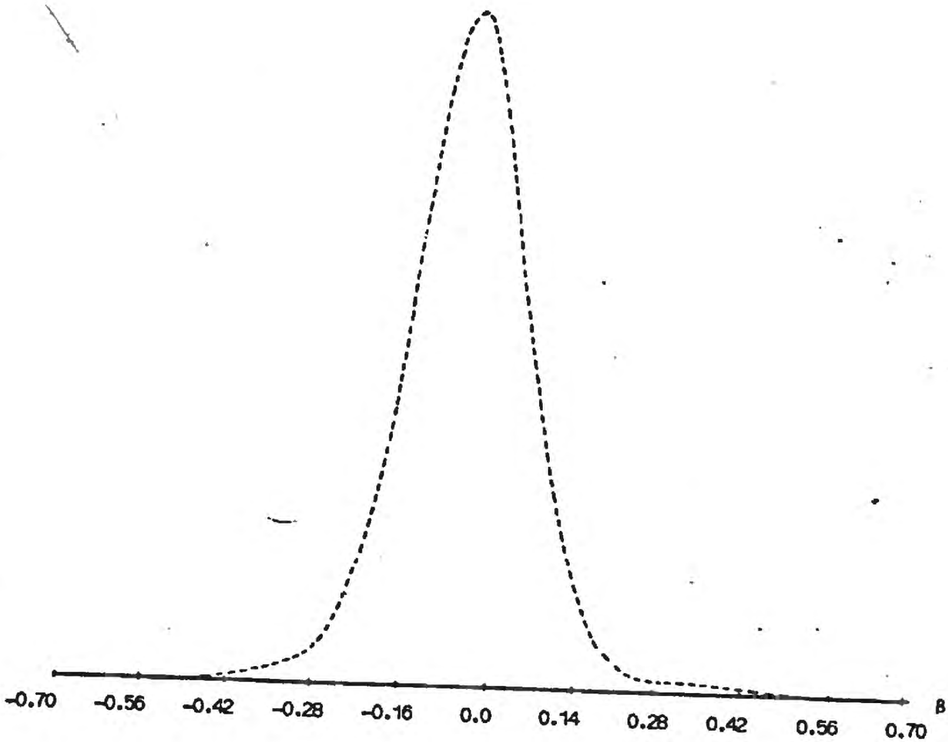

FIGURE 2 - POSTERIOR DENSITY FOR β

(k=1.0; σ=0.4)

3. TRANSFORMATION OF SURVIVAL DATA TO AN EXTREME VALUE DISTRIBUTION

Let us consider the power transformation of Box and Cox (1964) given by:

$$y^{(\lambda)} = \begin{cases} \dfrac{(T^{\lambda} - 1)}{\lambda} & \text{if} \quad \lambda \neq 0 \\ \ell\,(T) & \text{if} \quad \lambda = 0 \end{cases} \tag{12}$$

where $y^{(\lambda)} = \mu + \sigma W$ and W has an extreme value density $\exp\{w-e^{w}\}$, $-\infty < w < \infty$.

We observe that if $\lambda = 0$, the survival time T has an Weibull distribution with parameters $\alpha = e^{\mu}$ and $\theta = 1/\sigma$.

The log-likelihood function for μ, σ and λ considering $W = (y^{(\lambda)} - \mu)/\sigma$, with an extreme value density, is given by:

$$L(\mu,\sigma,\lambda) = (\lambda-1) \sum_{i=1}^{n} \ell n(t_i) - n\ell n(\sigma) + \frac{1}{\sigma} \sum_{i=1}^{n} y_i^{(\lambda)} - \frac{n\mu}{\sigma} - e^{-\mu/\sigma} \sum_{i=1}^{n} e^{y_i^{(\lambda)}/\sigma} . \tag{13}$$

The Jeffreys prior density for μ, σ and k is given by:

$$\pi(\mu,\sigma,k) = \pi(\mu,\sigma|k)\pi_0(k) \tag{14}$$

$$\alpha\{\det I_{\lambda}(\mu,\sigma)\}^{1/2}\, \pi_0(\lambda)$$

where $I_\lambda(\mu,\sigma)$ is the Fisher information matrix given λ.

Since $E(W) = \psi(1)$ (digamma function and $var(W) = \psi^{(1)}(1)$ (trigamma function, we have $I_\lambda(\mu,\sigma) = (b_{ij})$, $i,j=1,2$, where

$$b_{11} = \frac{n}{\sigma^2}$$

$$b_{12} = b_{21} = \frac{n}{\sigma^2}(1 + \psi(1))$$

$$b_{22} = \frac{n}{\sigma^2}(1 + \psi^{(1)}(2) + (\psi(2))^2).$$

Thus, since $\psi(2) = 1 + \psi(1)$, $\det I_\lambda(\mu,\sigma) = n^2\sigma^{-4}(1+\psi^{(1)}(2))$. Considering an uniform prior for λ, we have (from (14)):

$$\pi(\mu,\sigma,\lambda) \propto 1/\sigma^2 \tag{15}$$

where $-\infty < \mu, \lambda < \infty$ and $\sigma > 0$.

The joint marginal posterior for σ and λ is given by:

$\pi(\sigma,\lambda|\text{data}) \propto$

$$\propto \frac{\{\prod_{i=1}^{n} t_i^{\lambda-1}\}\exp\{\frac{1}{\sigma}\sum_{i=1}^{n} y_i^{(\lambda)}\}}{\sigma^{n+1}\{\sum_{i=1}^{n} e^{y_i^{(\lambda)}/\sigma}\}^n} \tag{16}$$

where $\sigma > 0$ and $-\infty < \lambda < \infty$.

The marginal posterior density for λ (using Laplace's method) is given by:

$\pi(\lambda|\text{data}) \propto$

$$\frac{\mathcal{I}^{*1/2} \left(\prod_{i=1}^{n} t_i^{\lambda-1} \right) \exp\left(\frac{1}{\hat{\sigma}^*} \sum_{i=1}^{n} y_i^{(\lambda)}\right)}{\hat{\sigma}^{*n+1} \left\{ \sum_{i=1}^{n} e^{y_i^{(\lambda)}/\hat{\sigma}^*} \right\}^n} \qquad (17)$$

where $\hat{\sigma}^* = \dfrac{n \sum_{i=1}^{n} y_i^{(\lambda)} e^{y_i^{(\lambda)}/\hat{\sigma}^*}}{(n+1) \sum_{i=1}^{n} e^{y_i^{(\lambda)}/\hat{\sigma}^*}} - \dfrac{\sum_{i=1}^{n} y_i^{(\lambda)}}{(n+1)}$,

$$\mathcal{I}^* = -1/ \left. \frac{d^2 L_\lambda(\sigma)}{d\sigma^2} \right|_{\sigma = \hat{\sigma}^*}$$

and $L_\lambda(\sigma) = \frac{1}{n} \{(\lambda-1) \sum_{i=1}^{n} \ln(t_i) - (n+1)\ln(\sigma) -$

$- n\ln\left(\sum_{i=1}^{n} e^{y_i^{(\lambda)}/\sigma} \right) + \frac{1}{\sigma} \sum_{i=1}^{n} y_i^{(\lambda)} \}$ (see Achcar, Bolfarine and Pericchi, 1987).

As a numerical example, consider the survival times of 21 patients submitted to a new treatment (times in weeks): 1, 1, 4, 4, 9, 16, 25, 25, 64, 64, 64, 64, 121, 121, 144, 144, 225, 289, 484, 529.

In figure 3, we have the graph of $\ln\{-\ln \hat{G}(w_i)\}$ versus t_i, where $\hat{G}(w)$ is the empirical distribution function. We conclude that the extreme value density is not appropriate for the original data set.

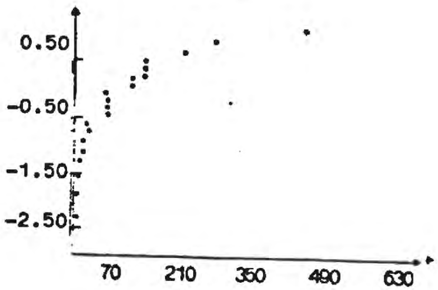

FIGURE 3

In figure 4, we have the graph of the marginal posterior density for λ. We observe that the mode of the marginal posterior density for λ is $\hat{\lambda} \approx 0.0$. In figure 5, we see that the transformed data $\ln(t_i)$ is well fitted to an extreme value distribution.

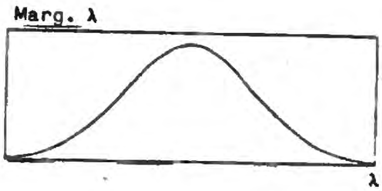

FIGURE 4

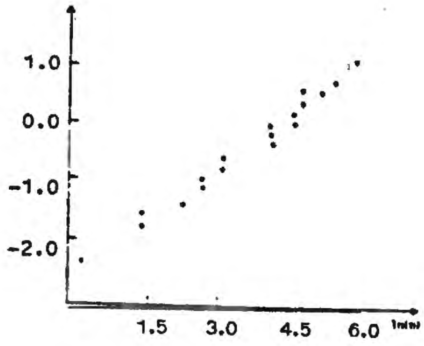

FIGURE 5

4. A BAYESIAN ANALYSIS CONSIDERING AN EXTREME VALUE DENSITY FOR W

Assuming $k=1$, $\beta=0$ and σ known in (4), and a locally unfiorm prior for μ, the posterior density for μ is given by:

$$\pi(\mu|\text{data}) = \frac{V^n}{\sigma\Gamma(n)} \exp -\{\frac{\mu n}{\sigma} - V e^{-\mu/\sigma}\} \tag{18}$$

where $-\infty < \mu < \infty$; $V = \sum_{i=1}^{n} \exp\{y_i/\sigma\}$.

The posterior density for the median survival time $m = (\ell n\, 2)^\sigma e^\mu$ is given by:

$\pi(m|\text{data}) =$

$$= \frac{(V \ell n\, 2)^n}{\sigma\,\Gamma(n)}\, m^{-(\frac{n}{\sigma}+1)} \exp\{-\frac{V \ell n\, 2}{m^{1/\sigma}}\} \tag{19}$$

where $m > 0$.

Assuming σ unknown and a joint prior for μ and σ proportional to σ^{-1}, the joint posterior density for μ and σ is given by:

$\pi(\mu, \mid \text{data}) \propto$

$$\propto \sigma^{-(n+1)} \exp\{\sigma^{-1} \sum_{i=1}^{n} y_i - \mu n \sigma^{-1} - V e^{-\mu/\sigma}\}$$

where $-\infty < \mu < \infty$ and $\sigma > 0$.

The marginal posterior density for the shape parameter $\theta = \sigma^{-1}$ of the Weibull distribution is given by:

$$\pi(\theta \mid \text{data}) \propto \frac{\theta^{n-2}}{V^n} \exp\{\theta \sum_{i=1}^{n} y_i\} \tag{21}$$

As an example, consider the Gehan's (1963) data, where we have the survival times (in weeks) of 42 patients in two treatment groups: 21 patients in group 1 (treatment GMP) with 12 censored observations and 21 patients in group 2 (treatment placebo) with no censorings.

Considering an exponential model, we have in figure 6 the graphs of posterior densities for the median m for both treatments. We conclude that there is a great treatment effect of the drug GMP in the survival times of the patients.

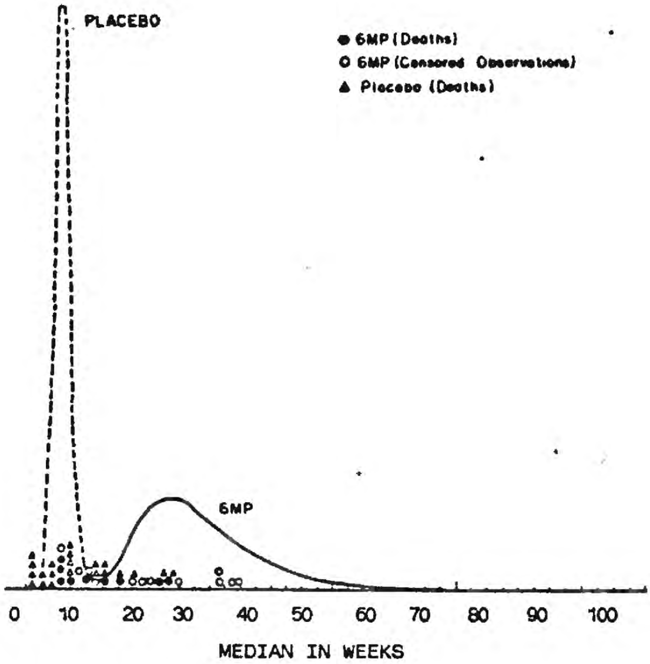

FIGURE 6

We can also consider the posterior density for the ratio of medians $R = m_1/m_2$ given by:

$$\pi(R|\text{data}) = \frac{\Gamma(d_1+d_2)V_1^{d_1} V_2^{d_2} R^{\frac{d_2}{\sigma}-1}}{\sigma\Gamma(d_1)\Gamma(d_2)\ (V_1+R^{1/\sigma} V_2)^{d_1+d_2}} \quad (22)$$

where d_1 is the number of deaths in group 1, d_2 is the number of deaths in group 2, $V_i = \sum_{j=1}^{n_i} \exp\{y_{ij}/\sigma\}$, $i = 1,2$, n_1 is the sample size in group 1 and n_2 is the sample size in group 2.

In figure 7, we have the graph of the posterior density

π(R|data) with $\sigma = 1$. We also conclude that there is strong treatment effect, since $P(R \leq 1) \approx 0$.

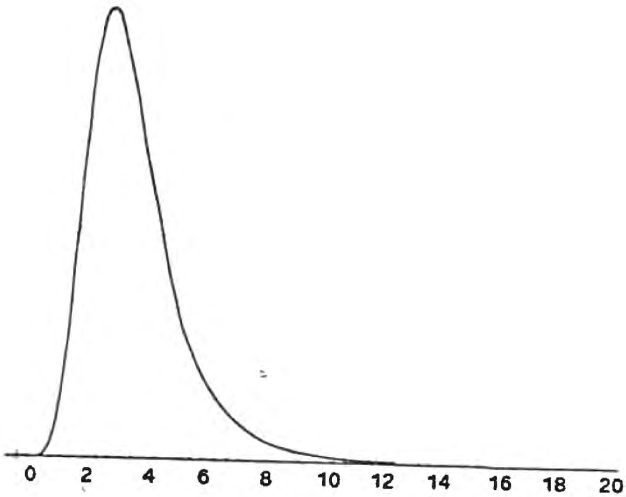

FIGURE 7

Considering σ unknown, the contours of posterior density for $\theta = \sigma^{-1}$ and the median m are given in figure 8. From figure 8, we observe that the exponential model ($\theta = \sigma = 1$) is adequate to analyse the data for both treatment groups (see Achcar, Brookmeyer and Hunter, 1985).

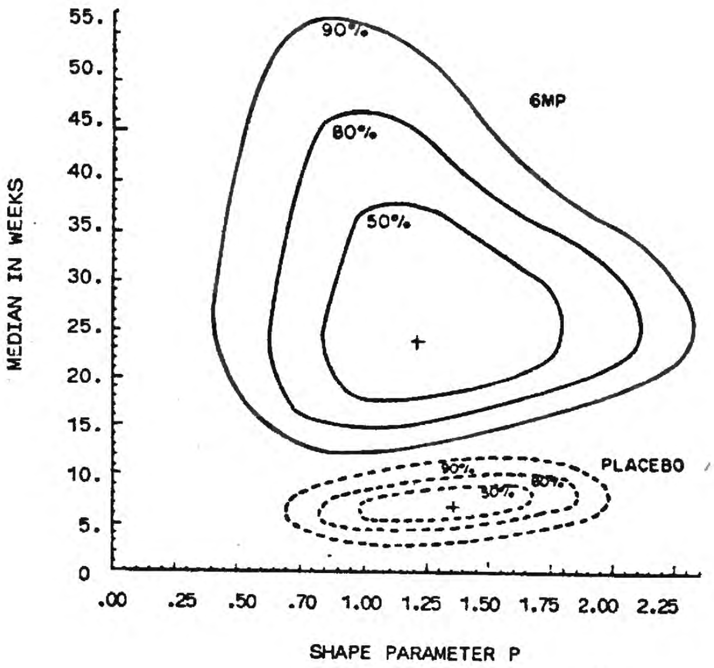

FIGURE 8

5. USE OF PREDICTIVE DENSITIES IN SURVIVAL ANALYSIS

The predictive density for $\underset{\sim}{y}' = (y_1, y_2, \ldots, y_n)$ given a proposed model A is given by:

$$\pi(\underset{\sim}{y}|A) = \int_{\theta} \pi(\underset{\sim}{y}|\theta, A)\pi(\theta|A)d\theta \, . \tag{23}$$

Considering the log-linear model (4) with density (3) for the error variable W, and σ and k known, the likelihood function for μ and β is given by:

$$\ell(\mu,\beta) = \xi_1(k,\sigma)\exp\{-\frac{kn\mu}{\sigma} - \frac{\beta k\sum_{i=1}^{n} x_i}{\sigma}\}.$$

$$\cdot \exp\{-e^{-\mu/\sigma}\sum_{i=1}^{n} e^{(y_i-\beta x_i)/\sigma}\} \tag{24}$$

where $\xi_1(k,\sigma) = \{\frac{1}{\sigma\Gamma(k)}\}^n \exp\{\frac{k}{\sigma}\sum_{i=1}^{n} y_i\}$.

Considering a locally uniform prior for μ and β, we have (from (23)):

$$\pi(\underset{\sim}{y}|A) = \int_{-\infty}^{\infty} \frac{\xi_1(k,\sigma)\sigma\ \Gamma(kn)\ e^{-\beta k a_2/\sigma}\ d\beta}{\{\sum_{i=1}^{n} e^{(y_i-\beta x_i)/\sigma}\}^{kn}} \tag{25}$$

where $a_2 = \sum_{i=1}^{n} x_i$.

Using Laplace's method, we find:

$$\pi(\underset{\sim}{y}|A) = \frac{\sqrt{2\pi}\ \sigma^2\Gamma(kn)\exp\{-\frac{\hat{\beta} k a_2}{\sigma}\}\ \xi_1(k,\sigma)}{\sqrt{nk}\ \{\sum_{i=1}^{n} e^{(y_i-\hat{\beta}x_i)/\sigma}\}^{kn-1}}. \tag{26}$$

$$\cdot \left[(\sum_{i=1}^{n} x_i^2 e^{(y_i-\hat{\beta}x_i)/\sigma})(\sum_{i=1}^{n} e^{(y_i-\hat{\beta}x_i)/\sigma}) - (\sum_{i=1}^{n} x_i e^{(y_i-\hat{\beta}x_i)/\sigma})^2 \right]^{-1/2}$$

where $\hat{\beta}$ satisfies the relation,

$$\frac{\sum_{i=1}^{n} x_i \, e^{(y_i-\hat{\beta}x_i)/\sigma}}{\sum_{i=1}^{n} e^{(y_i-\hat{\beta}x_i)/\sigma}} = \frac{a_2}{n}$$

In the selection of a model between two possible ones A_1 or A_2, we can find the ratio of predictives $\pi(\underset{\sim}{y}|A_1)/\pi(\underset{\sim}{y}|A_2)$ to decide by the best one. In the diagnostic checking of a proposed model A, we can calculate $\alpha = P\{\pi(\underset{\sim}{y}|A) < \pi(\underset{\sim}{y}_d|A)\}$ (Box, 1980), where $\underset{\sim}{y}_d$ is the vector of observed data. If α is too small (e.g., $\alpha < 0.001$), we have strong evidence against the use of model A.

In the two-sample problem, consider $x = 0$ (treatment 1) and $x = 1$ (treatment 2). Also consider $T_{11}, T_{12}, \ldots, T_{1n_1}$ a random sample of size n_1 representing the survival times for treatment 1 and $T_{21}, T_{22}, \ldots, T_{2n_2}$ a random sample of size n_2 representing the survival times for treatment 2. From (26), we have $\hat{\beta} = \sigma\{\ell n(n_1 \sum_{i=1}^{n_2} T_{2i}^{1/\sigma}) - \ell_n(n_2 \sum_{i=1}^{n_1} T_{1i}^{1/\sigma})\}$ and the predictive density for $\underset{\sim}{t}' = (t_1, t_2, \ldots, t_n)$, $n = n_1 + n_2$ is given by:

$$\pi(\underset{\sim}{t}|A) \approx$$

$$\approx \frac{\sqrt{2\pi}\ \sigma^2 \Gamma(kn) n_1^{kn_1-1/2} \ n_2^{kn_2-1/2} \ \xi_1(k,\sigma)}{\sqrt{kn}\ n^{kn-1} \left(\sum_{i=1}^{n_1} T_{1_i}^{1/\sigma}\right)^{kn_1} \left(\sum_{i=1}^{n_2} T_{2_i}^{1/\sigma}\right)^{kn_2}}$$

Considering the data of table 1, where $n_1 = 40$, $x = 0$ (group 1), $n_2 = 40$ and $x = 1$ (group 2), we have from figure 1 two possible models:

A_1 with $k^{(1)} = 1$ and $\sigma^{(1)} = 0.4$ (Weibull model) for A_2 with $k^{(2)} = 3$ and $\sigma^{(2)} = 1$ (Gamma model). The ratio of predictives (from (27)) is given by:

$$\frac{\pi(\underset{\sim}{t}|A_1)}{\pi(\underset{\sim}{t}|A_2)} =$$

$$\simeq \frac{\sqrt{3}\ \Gamma(80)\ 80^{160}\ (0.4)^2\ e^{-170.7}\ 2^{80}\ (3381.50)^{120}}{(0.4)^{80}\ \Gamma(240)\ 40^{160}\ (3736745)^{40}\ (3435657)^{40}\ (3254)^{-120}}$$

That is, $\pi(\underset{\sim}{t}|A_1)/\pi(\underset{\sim}{t}|A_2) \simeq 57.103$ and we conclude that model A_1 (Weibull) is preferable to model A_2 (Gamma).

We also can calculate $\alpha = P\{X^2_{2n-1} > c\}$ where

$$c = 4 \exp\{-\frac{1}{\sigma}(\mu + \frac{\beta}{2})\}\ (\sum_{i=1}^{n} t_{d1i}^{1/\sigma})^{1/2}\ (\sum_{i=1}^{40} t_{d2i}^{1/\sigma})^{1/2}.$$

With $k = 1$, $\sigma = 0.4$, $\beta = -0.045$ and $\mu = 4.578$ (see Achcar and Bolfarine, 1986), we find $\alpha \simeq P\{X^2_{159} > 161.24\} \simeq 0.56$, that is, we conclude that the Weibull model is appropriate.

6. BAYESIAN METHODS IN THE DESIGN OF CLINICAL TRIALS

Usually, researchers in the medical area should determine the required number of patients in one clinical trial before starting the experiment.

Let us consider two periods of study. In the first period, we consider n_1 patients and a fixed period of time t_1. Let us call this data as "data 1". From the first period of time, we have n_{1s} survivors and $d_1 = n_1 - n_{1s}$ deaths. In the second period of time, we consider n_{2n} new patients and the n_{1s} survivors of the first period. Thus, we have $n_2 = n_{2n} + n_{1s}$ (sample size) and t_2 the fixed follow-up period of time for the second period ("data 2").

We assume the log-linear model (2) given by $y_i = \ln(T_i^o) = \mu + \sigma W_i$ where T_i^o is the time survival time and W_i is the error variable with an extreme value density. The observed survival time is given by $T_i = \min(T_i^o, C_i)$ where C_i is the fixed censoring time associated to T_i^o.

With σ known and a locally uniform pior for μ, the posterior density for the logarithm of the median $m = (\ln 2)^\sigma e^\mu$ considering the "data 1" is given by (from (19)):

$$\pi(\ell \mid \text{data } 1, \sigma) \propto \exp\{-\frac{d_1 \phi}{\sigma} - \frac{V_1 \ln 2}{e^{\ell/\sigma}}\} \tag{28}$$

where $-\infty < \phi = \ln(m) < \infty$, $V_1 = \sum_{i=1}^{n_1} T_{1i}^{1/\sigma}$ and d_1 is the number of deaths in period 1.

Approximately, we have:

$$\phi \mid \text{data } 1, \sigma \overset{a}{\sim} N\{\sigma \ln(\frac{V_1 \ln^2}{d_1}); \frac{\sigma^2}{d_1}\} \tag{29}$$

(see Achcar, 1984).

We consider the posterior density (29) in period 1 as the prior density for period 2.

In the second period of study, the likelihood function for ϕ is approximated by:

$$\ell(\phi \mid \text{data } 2,\sigma) \sim N(\sigma \ln(\frac{V_2 \ln 2}{d_2}); \frac{\sigma^2}{d_2}) \qquad (30)$$

where d_2 is the number of deaths in period 2 and $V_2 = \sum_{i=1}^{n_2} T_{2i}^{1/\sigma}$.

Combining (29) and (30), we find:

$$\phi \mid \text{data } 1, \text{ data } 2,\sigma \sim N(\bar{\phi}; \bar{\Sigma}) \qquad (31)$$

where $\bar{\phi} = \frac{\sigma}{d_1+d_2} \{d_1 \ln(\frac{V_1 \ln 2}{d_1}) + d_2 \ln(\frac{V_2 \ln 2}{d_2})\}$

and $\bar{\Sigma} = \sigma^2/(d_1+d_2)$.

The length of a 100(1-α)% HPD Interval for $\phi = \ln(m)$ is given by $\ell = 2Z_{1-\alpha/2}\, \sigma/\sqrt{d_1+d_2}$ where $Z_{1-\alpha/2}$ is a percentile of the standard normal distribution.

If we specify the length ℓ of the HPD interval for $\phi =$ $= \ln(m)$ and σ is known, we can find an estimate of the number n_{2n} of new patients required in the second part of the study by solving the following equation for n_{2n}:

$$P\left\{\left[\frac{2Z_{1-\alpha/2}\,\sigma}{\sqrt{d_1 + d_2}} \le \ell \mid \text{data } 1,\sigma\right]\right\} = 1 - \beta \qquad (32)$$

;

where $(1-\beta)$ is a given confidence coefficient.

That is,

$$P\{ d_2 \geq \frac{4Z^2_{1-\alpha/2}\ \sigma^2}{\ell^2} - d_1 \mid \text{data } 1,\sigma\} = 1 - \beta \tag{33}$$

where d_1 is fixed (number of deaths in the first period) and $d_2 =$ $= d_{2s} + d_{2n}$ where d_{2s} is the number of deaths among the n_{1s} survivors of the first period of study and d_{2n} is the number of deaths among the n_{2n} new patients.

Using "delta method" and assuming that the entry times e_i of the patients are uniform in the second period of time, we find (see Achcar, 1984):

$$d_2 \mid \text{data } 1,\sigma \overset{a}{=} N\{ \sum_{i=1}^{n_{1s}} \hat{p}_s(e_i) + n_{2n}\hat{p}_{2n};$$

$$\sum_{i=1}^{n_s} \hat{p}_s(e_i)(1-\hat{p}_s(e_i)) + n_{2n}\hat{p}_{2n}(1-\hat{p}_{2n}) + C\} \tag{34}$$

where

$$\hat{p}_s(e_i) = 1 - \exp\{-\mathfrak{Z}^{1/\sigma}[(t_1+t_2-e_i)^{1/\sigma} - (t_1-e_i)^{1/\sigma}]\}$$

is the estimated probability of death for survivors of the first period and conditional on the entry times e_i, $i = 1,2,\ldots,n_{1s}$, $\mathfrak{Z} =$ $= 1/\hat{\alpha} = e^{-\hat{\mu}}$ (maximum likelihood estimator), and

$$\hat{p}_{2n} = 1 - \frac{\sigma\,\Gamma(\sigma)}{\mathfrak{Z}t_2}\, I_\sigma\{(\mathfrak{Z}t_2)^{1/\sigma}\}$$

is the estimated probability of death in the second follow-up period t_2 assuming that they enter uniformly on study, $I_k(s)$ is the incomplete gamma integral, and

$$C = \frac{\sigma \mathfrak{J}^2 [\sum_{i=1}^{n_{1s}} \hat{A}_i + n_{2n}^2 \hat{B}^2 + \sum_{i \neq j}^{n_{1s}} \sum^{n_{1s}} \sqrt{\hat{A}_i \hat{A}_j} + 2n_{2n}\hat{B} \sum_{i=1}^{n_{1s}} \sqrt{\hat{A}_i}\ ]}{[d_1 + (1/\sigma - 1) V_1 \mathfrak{J}^{1/\sigma}]}$$

and

$$\hat{A}_i = \frac{1}{\sigma^2} \mathfrak{J}^{2(\frac{1}{\sigma}-1)} [(t_1+t_2-e_i)^{1/\sigma} - (t_1-e_i)^{1/\sigma}]^2 .$$

$$\cdot \exp\{-2\mathfrak{J}^{1/\sigma}[(t_1+t_2-e_i)^{1/\sigma} - (t_1-e_i)^{1/\sigma}]\},$$

and

$$\hat{B} = \frac{\sigma \Gamma(\sigma)}{\mathfrak{J}^2 t_2} I_\sigma\{(\mathfrak{j}t_2)^{1/\sigma}\} \Big|_{\mathfrak{j}=\mathfrak{J}} - \frac{\sigma \Gamma(\sigma)}{\mathfrak{J}\, t_2} \frac{dI_\sigma\{(\mathfrak{j}t_2)^{1/\sigma}\}}{d\mathfrak{j}} \Big|_{\mathfrak{j}=\mathfrak{J}} .$$

Thus, we find n_{2n} such that:

$$\Phi < \left[\frac{4Z_{1-\alpha/2}^2\ \sigma^2/\ell^2 - d_2 - g_1(n_{2n})}{\sqrt{g_2(n_{2n})}} \right] >= \beta \qquad (35)$$

where Φ is the distribution function of a standard normal distribution,

$$g_1(n_{2n}) = \sum_{i=1}^{n_{1s}} \hat{p}_s(e_i) + n_{2n}\hat{p}_{2n}$$

and $g_2(n_{2n}) = \sum_{i=1}^{n_{1s}} \hat{p}_s(e_i)(1-\hat{p}_s(e_i)) +$

$+ n_{2n} \hat{p}_{2n} (1-\hat{p}_{2n}) + C.$ (see Achcar, 1984).

In table II, we have the survival data of thirty-eight patients with non-small cell carninoma of the lung (data obtained from the Wisconsin Clinical Cancer Center). Using the information of the thirty-eight patients in table II, we want to find an estimate of the number of new patients n_{2n} needed in a second part of the study.

TABLE II - SURVIVAL TIMES IN DAYS

i	t_i	DATE OF ENTRY	e_i	i	t_i	DATE OF ENTRY	e_i
1	181	05/17/79	0	20	62	05/14/80	362
2	11	06/01/79	15	21	147	05/20/80	368
3	64	06/06/79	20	22	723+	05/29/80	377
4	216	08/08/79	83	23	146	06/06/80	385
5	374	08/22/79	97	24	130	06/24/80	403
6	216	08/31/79	106	25	67	07/08/80	417
7	227	09/04/79	110	26	87	07/12/80	421
8	237	12/06/79	203	27	169	07/21/80	430
9	229	12/12/79	209	28	201	07/28/80	437
10	264	12/27/79	224	29	510	07/29/80	438
11	97	01/22/80	250	30	661+	07/30/80	439
12	53	02/15/80	274	31	543	08/10/80	441
13	361	02/20/80	279	32	600+	09/29/80	500
14	214	02/21/80	280	33	38	10/16/80	517
15	799+	03/14/80	301	34	18	11/04/80	536
16	786+	03/27/80	314	35	15	11/07/80	539
17	158	04/09/80	327	36	561+	11/07/80	539
18	75	04/11/80	329	37	527+	12/11/80	573
19	754+	04/28/80	346	38	193	01/28/81	622

The survival times t_i are given in days and the entry times e_i, in terms of calendar time from start of study, also are given in days. We observe that the patient 22 has a survival time $t = 723+$ and entry time $e_{22} = 377$. Thus, we assume a follow-up period $t_1 = 1100$ (that is, the end of study is 05/22/82). We also assume that all censored patients are alive by the end of the sutdy. From table II, we have $n_1 = 38$, $n_{1s} = 8$ and $d_1 = 30$.

Assuming a fixed period of time $t_2 = 500$ days for the second period of study, and considering $\alpha = \beta = 0.05$, we have in table III some estimated values of n_{2n} required for each given length ℓ.

TABLE III

NUMBER OF PATIENTS IN SECOND PERIOD

LENGTH ℓ	n_{2n}
0.53	10
0.44	20
0.34	38
0.26	64

We observe that we need to add 38 new patients in the second period of study to have a length $\ell = 0.34$ for the HPD interval for $\phi = \ell n(m)$. That is we need a total of 76 patients in both periods.

REFERENCES

ACHCAR, J.A. (1984). "Use of Bayesian Analysis to design of clinical trials with one treatment", COMMUN. STATIST. -- THEOR.METH., 13(14), 1693-1707.

ACHCAR, J.A.; BROOKMEYER, R.; HUNTER, W.G. (1985)."An application of Bayesian analysis to medical follow-up data", STATISTICS IN MEDICINE, 4, 509-520.

ACHCAR, J.A.; BOLFARINE, H. (1986). "The log-linear model with a generalized Gamma distribution for the error: a Bayesian approach", STATISTICS & PROBABILITY LETTERS, 4, 325-332.

ACHCAR, J.A.; BOLFARINE, H. (1986). "Predictive densities in survival analysis with a generalized Gamma regression model", REBRAPE (Brazilian Journal of Probability and Statistics) (to appear).

ACHCAR, J.A.; BOLFARINE, H.; PERICCHI, L.R. (1987). "Transformation of survival data to an extreme value distribution", THE STATISTICIAN, 36, 229-234.

AITCHISON, J.; DUNSMORE, I.R. (1975). STATISTICAL PREDICTION ANALYSIS, Cambridge University Press.

BOX, G.E. (1980). "Sampling and Bayes' Inference in scientific modelling and robustness", J.R.Stat.Soc., A, 143, part A, 383-430.

BOX, G.E.; COX, D.R. (1964). "An analysis of transformations", J.R.Stat.Soc., Series B, 26, 211-250.

BOX, G.E.; TIAO, G.C. (1973). BAYESIAN INFERENCE IN STATISTICAL ANALYSIS, New York: Addison-Wesley.

BUCKLEY, J.; JAMES, I. (1979). "Linear Regression with Censored Data", BIOMETRIKA, 66, 429-436.

COX, D.R. (1972). "Regression Models and Life-Tables", J.R.Stat. Soc., Series B, 34, 187-202.

FAREWELL, V.T.; PRENTICE, R.L. (1977). "A study of distributional shape in life testing", TECHNOMETRICS, 1979, 69-75.

GEHAN, E. et all (1963). "The effect of 6-mercaptopurine on the duration of steroid-induced remission in acute leukemia: a model for evaluation of other potentially useful therapy", BLOOD, 21(6), 699-716.

KAPLAN, E.L.; MEIER, P. (1958). "Nonparametric estimation from incomplete observations", JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION, 53, 457-481.

KOUL, H.; SUSARLA, V.; VAN RYZIN, J. (1981). "Regression Analysis with Randomly Right-Censored Data". ANNALS OF STATISTICS, 9, 1276-1288.

LAWLESS, J.F. (1980). "Inference in the Generalized gamma and log-Gamma distributions", TECHNOMETRICS, 22, 409-419.

LEE, L.T. (1980). STATISTICAL METHODS FOR SURVIVAL DATA ANALYSIS, LIFETIME LEARNING PUBLICATIONS, Belmont, California, 1980.

MARTZ, H.F.; WALLER, R.A. (1982). BAYESIAN RELIABILITY ANALYSIS, Wiley, New York.

MILLER, R. (1976). "Least-squares Regression with Censored Data", BIOMETRIKA, 63, 449-464.

PRENTICE, R.L. (1974). "A log-gamma model and its maximum likelihood estimation", BIOMETRIKA, 61, 539-544.

TIERNEY, L.; KADANE, J.B. (1986). "Accurate Apporximations for Posterior Moments and Marginal Densities", Journal of the American Statistical Association, 81, 393, 82-86.

RELATÓRIO TÉCNICO
DO
DEPARTAMENTO DE ESTATÍSTICA

TÍTULOS PUBLICADOS EM 1987

8701 - ACHCAR, J.A. & BOLFARINE, H.; Constant Hazard Against a Change-Point Alternative: A Bayesian Approach with Censored Data, São Paulo, IME-USP, 1987, 20p.

8702 - RODRIGUES, J.; Some Results on Restricted Bayes Least Squares Predictors for Finite Populations, São Paulo, IME-USP, 1987, 16p.

8703 - LEITE, J.G., BOLFARINE, H. & RODRIGUES, J.; Exact Expression for the Posterior Mode of a Finite Population Size: Capture-Recapture Sequential Sampling, São Paulo, IME-USP, 1987, 14p.

8704 - RODRIGUES, J., BOLFARINE, H. & LEITE, J.G.; A Bayesian Analysis in Closed Animal Populations from Capture Recapture Experiments with Trap Response, São Paulo, IME-USP, 1987, 21p.

8705 - PAULINO, C.D.M.; Analysis of Categorical Data with Full and Partial Classification: A Survey of the Conditional Maximum Likelihood and Weighted Least Squares Approaches, São Paulo, IME-USP, 1987, 52p.

8706 - CORDEIRO, G.M. & BOLFARINE, H.; Prediction in a Finite Population under a Generalized Linear Model, São Paulo, IME-USP, 1987, 21p.

8707 - RODRIGUES, J. & BOLFARINE, H.; Nonlinear Bayesian Least-Squares Theory and the Inverse Linear Regression, São Paulo, IME-USP, 1987, 15p.

8708 - RODRIGUES, J. & BOLFARINE, H.; A Note on Bayesian Least-Squares Estimators of Time-Varying Regression Coefficients, São Paulo, IME-USP, 1987, 11p.

8709 - ACHCAR, J.A., BOLFARINE, H. & RODRIGUES, J.; Inverse Gaussian Distribution: A Bayesian Approach, São Paulo, IME-USP, 1987, 20p.

8710 - CORDEIRO, G.M. & PAULA, G.A.; Improved Likelihood Ratio Statistics for Exponential Family Nonlinear Models, São Paulo, IME-USP, 1987, 26p.

8711 - SINGER, J.M.; PERES, C.A.& HARLE, C.E.; On the Hardy -Weinberg Equilibrium in Generalized ABO Systems, São Paulo, IME-USP, 1987, 16p.

8712 - BOLFARINE, H. & RODRIGUES, J.; A Review and Some Extensions on Distribution Free Bayesian Approaches for Estimation and Prediction, São Paulo, IME-USP, 1987, 19p.

8713 - RODRIGUES, J.; BOLFARINE, H. & LEITE, J.G.; A Simple Nonparametric Bayes Solution to the Estimation of the Size of a Closed Animal Population, São Paulo, IME-USP, 1987, 11p.

8714 - BUENO, V.C.; Generalizing Importance of Components for Multistate Monotone Systems, São Paulo, IME-USP, 1987, 12p.

8801 - PEREIRA, C.A.B. & WECHSLER, S.; On the Concept of P-value, São Paulo, IME-USP, 1988, 22p.

8802 - ZACKS, S., PEREIRA, C.A.B. & LEITE, J.G.; Bayes Sequential Estimation of the Size of a Finite Population, São Paulo,IME-USP, 1988, 23p.

8803 - BOLFARINE, H.; Finite Population Prediction Under Dynamic Generalized Linear Models, São Paulo, IME-USP, 1988, 21p.

8804 - BOLFARINE, H.; Minimax Prediction in Finite Populations, São Paulo, IME-USP, 1988, 18p.

8805 - SINGER, J.M. & ANDRADE, D.F.; On the Choice of Appropriate Error Terms for Testing the General Linear Hypothesis in Profile Analysis, São Paulo, IME-USP, 1988, 23p.

8806 - DACHS, J.N.W. & PAULA, G.A.; Testing for Ordered Rate Rations in Follow-up Studies with Incidence Density Data, São Paulo, IME-USP, 1988, 18p.

8807 - CORDEIRO, G.M. & PAULA, G.A.; Estimation, Significance Tests and Diagnostic Methods for the Non-Exponential Family Nonlinear Models, São Paulo, IME-USP, 1988, 29p.

8808 - RODRIGUES, J. & ELIAN, S.N.; The Coordinate - Free Estimation in Finite Population Sampling, São Paulo, IME-USP, 1988, 5p.

8809 - BUENO, V.C. & CUADRADO, R.Z.B.; On the Importance of Components for Continuous Structures, São Paulo, IME-USP, 1988, 14p.