Journal of Statistical Research *ISSN 0256 - 422 X*
2007, Vol. 41, No. 2, pp. 35-62
Bangladesh

A HIERARCHICAL BAYESIAN APPROACH TO THE ESTIMATION OF MONOTONE HAZARD RATES IN THE RANDOM RIGHT CENSORSHIP MODEL

Hassan Younes

Université du Québec à Montréal

Email: younes.hassan@uqam.ca

Mohan Delampady

Indian Statistical Institute

Email: mohan@isibang.ac.in

Brenda MacGibbon

Université du Québec à Montréal

Email: macgibbon.brenda@uqam.ca

Omar Cherkaoui

Université du Québec à Montréal

Email: cherkaoui.omar@uqam.ca

SUMMARY

Here we study hierarchical Bayesian estimation of a monotone hazard rate for both complete and randomly right censored data. We propose two methods of computation: Monte-Carlo importance sampling and Laplace approximation techniques. These methods are computationally simple and easily implemented on complex hazard functions. They are compared in simulation studies with uncensored and censored data and the methodology is illustrated on two interesting data sets.

Keywords and phrases: Survival analysis, reliability, Laplace approximation, Monte-Carlo importance sampling.

AMS Classification: 62F15

1 Introduction

It can be especially important in the analysis of survival or reliability data to obtain brief summaries of the data which convey the key conclusions. In these situations the hazard function can be extremely useful since it displays risk patterns as a function of time. In the comparison of two sampled populations many papers display the results with plots of the survival function (or Kaplan-Meier curve, *cf.* Kaplan and Meier, 1958). If, as occurs in many cases, the graphs of these functions cross, then the results become difficult to interpret. As suggested by Efron (*cf.* Efron, 1980), many of these papers could benefit from the presentation of the hazard functions, especially if they are simple functions.

The simplest hazard function is the constant function, corresponding to an exponential survival distribution. However, in many real life situations the assumption of monotonicity of the hazard rates is warranted; *e.g.*, in some clinical trials it has been observed that the hazard decreases with time, while in the study of other clinical trials or of the reliability of physical components the hazard is often known to be increasing. Proschan (1963) also provides an explanation of why it is sometimes appropriate to use decreasing hazard rates for physical components. Much statistical research has been directed to the study of inference problems under order or monotonicity constraints. For examples of earlier work done in this area, we cite Ayer *et al.* (1955), Brunk (1955), Barlow *et al.* (1972) and Barlow and Brunk (1972). An excellent account of all these results may be found in the book of Robertson *et al.* (1988). Most of the work in this area has been directed toward the development of maximum likelihood and least squares estimators. Here we prefer to adopt a Bayesian approach and to estimate a monotone hazard function.

A difficulty with the analysis of industrial life testing or medical research data is that the observation of the occurrence of a failure time may be made impossible by the previous occurrence of a censoring event such as the termination of the study or withdrawal from the study; such data is said to be right censored. For frequentist estimation of a monotone hazard rate with randomly right censored data, we cite the original work of Grenander (1956) and that of Prakasa Rao (1970) for uncensored data and for censored data that of Padgett and Wei (1980), Huang and Wellner (1995) and MacGibbon *et al.* (2002) which is based on least concave majorants (greatest convex minorants).

Early Bayesian research in survival analysis mainly concentrated on the estimation of the survival function. Susarla and Van Ryzin (1978) used Dirichlet priors (*cf.* Ferguson, 1973) to estimate the survival function with censored data. Ferguson and Phadia (1979) extended this work to include prior distributions that are neutral to the right, previously studied by Doksum (1974). Kalbfleisch (1978) used a gamma process prior for survival function estimation. Leonard, Hsu and Tsui (1989), following the work of Hasselblad (1969) and Laird (1978) on nonparametric maximum likelihood estimation of finite mixtures, modelled the density function as a finite mixture of exponentials and used an empirical Bayes approach to obtain smooth estimates of the density and survival function. Dey and Kuo (1991) and Kuo and Yiannoustsos (1993) developed empirical Bayes approaches to Type II censored data. Kuo and Smith (1992) found Bayes estimators of the survival function with censored

data using the Gibbs sampler.

Arjas and Liu (1995) used a hierarchical Bayes approach to a nonparametric multiplicative hazard model in order to assess the losses caused by an industrial intervention. Berger and Sun (1996) studied Bayesian inference for a class of poly-Weibull distributions.

Some of the earlier Bayesian research did consider estimating the hazard function. Burridge (1981) presented an empirical Bayes model of the cumulative hazard function as a gamma process. Dykstra and Laud (1981) also modeled the hazard rate. They defined an appropriate prior stochastic process called an extended gamma process whose sample paths are hazard rates, and obtained the posterior distribution of the hazard rates for both exact and censored data.

Broffit (1984) considered the problem of estimating mortality rates over consecutive age intervals. As a model he considered the exponential family of density functions $\{f(x|\theta) : 0 \leq \theta < \infty\}$ where

$$f(x|\theta) = a(x)\theta^{b(x)}e^{-\theta c(x)} \tag{1.1}$$

which, when $a(x), b(x), c(x)$ are appropriately selected, may represent either a continuous or discrete distribution. He assumed random samples from k members of the above family with corresponding parameter values $\theta_1, \ldots, \theta_k$ satisfying

$$\theta_1 < \theta_2 < \ldots < \theta_k \, .$$

Broffit (1984) specified a prior distribution such that the probability of the set $\{(\theta_1, \ldots, \theta_k) : \theta_1 < \theta_2 < \ldots < \theta_k\}$ is 1, thus guaranteeing that the resulting Bayes estimator satisfies the order restriction. He called the technique Bayesian isotonic graduation.

Bayesian nonparametric hazard function estimation methodology in Dykstra and Laud (1981) was generalized in different ways by Ammann (1985), by Thompson and Thavaneswaran (1992) and by Arjas and Gasbarra (1994) who modelled the hazard rate nonparametrically as a jump process having a martingale sturcture with respect to the prior distribution and used Monte-Carlo Markov Chain (MCMC) techniques to implement their method on simulated examples. Hjort (1990) used beta process priors to estimate the cumulative hazard rate process. Further generalizations by Lo and Weng (1989), Ho and Lo (2001) and James (2003, 2005) culminated in the characterization given by Ho (2006) of the posterior distribution of the mixture hazard model of a monotone hazard rate *via* a finite mixture of S-paths. Using a hierarchical model structure, Ho (2006) modelled the hazard rate nonparametrically as a jump process having a martingale sturcture with respect to the prior distribution and described an algorithm that generates sample paths from the posterior by a dynamic Gibbs sampler. He also provided an intricate MCMC path sampler for sampling directly such S-paths. He illustrated the method in a simulation study of a one change-point hazard model with light to moderate censoring (approximately 20%) and a sample size of 100 with 1000 replications.

Our emphasis in this paper is rather different. Here we propose to study hierarchical Bayes estimators of monotone hazard rates and pay particular attention to providing computationally simple numerical algorithms for such estimators which are easily capable of

handling the multi-change-point hazard model. We also provide an exact form for the Bayes estimator which is useful for numerical computations when the sample size is small or moderate. We use Monte-Carlo importance sampling and the Laplace approximation method originally proposed by Tierney and Kadane (1986) and Tierney *et al.* (1989). Simulation studies for the Laplace approximation method and the importance sampling method are also presented. The methods are illustrated on a complete data set from a previously published case study of component reliability edited by Gentleman and Whitmore (1982), previously analyzed by several authors. We also illustrate the method on the Primary Biliary Cirrhosis (PBC) data which is a randomly right censored data set. It has been described in detail in Fleming and Harrington (1991) and analyzed there using frequentist methods.

2 A Hierarchical Bayesian Model for Monotone Hazard Rates

A general form for an increasing hazard rate function $h(t)$ is modelled as follows. We assume that the density function is defined on the interval $[a, b] \subseteq R^+$. For simplicity we assume that $b < \infty$. Let $A_1, \ldots, A_k$ be a strictly decreasing finite sequence of subsets of $[a, b]$ of the form $A_i = [a_i, b]$, with $a_1 = a < a_2 < \ldots < a_k$, and let θ_1, θ_2, ..., θ_k be positive real numbers.

The hazard function $h(x)$ is then modelled as

$$h(x) = \sum_{l=1}^{k} \theta_l I_{A_l}(x), \qquad x \in [a, b], \tag{2.1}$$

where $I_{A_j}(x)$ is the indicator function of the set A_j. The associated density function is clearly,

$$f(x|\theta_1, \ldots, \theta_k) = \left[\sum_{l=1}^{k} \theta_l I_{A_l}(x)\right] \exp\big(-\sum_{l=1}^{k} \theta_l \mu(A_l \cap [0, x])\big),\ x \in [a, b],$$

where $\mu(B)$ = Lebesgue measure of B.

It is clear that the problem of nonparametric estimation of h is thus reduced to the estimation of θ_1, ..., θ_k, a finite dimensional problem. Note that θ_l denotes the increase in the hazard rate in the lth subset $A_l \cap A_{l+1}^C$ where A^C denotes the complement of A.

The prior model for h now proceeds as follows using the hierarchical Bayesian approach. The first stage prior π_1 on θ_1, ..., θ_k is assumed to be a product of independent Gamma (α_i, c_i) distributions. In other words,

$$\begin{aligned} \pi_1(\theta_1, \ldots, \theta_k|\alpha_1, \ldots, \alpha_k, c_1, \ldots, c_k) &= \prod_{i=1}^{k} \pi_1(\theta_i|\alpha_i, c_i) \\ &\propto \prod_{i=1}^{k} \left\{\theta_i^{c_i - 1} \exp(-\alpha_i \theta_i)\right\}. \end{aligned}$$

Note also that $E^{\pi}(\theta_i) = c_i/\alpha_i$ and $Var^{\pi}(\theta_i) = c_i/\alpha_i^2$ represent the prior mean and variance of the increase in the hazard rate in $A_i \cap A_{i+1}^C$. A second stage prior on c_i and α_i completes the prior specifications. In our applications we have chosen $c_i \equiv 1$ and $\alpha_i \equiv \alpha$. Therefore for convenience we use this formulation in our presentation below. Thus

$$\pi_1(\theta_1, \ldots, \theta_k|\alpha) = \alpha^k \exp(-\alpha \sum_{i=1}^{k} \theta_i).$$

At the second stage the hyperparameter α is assigned a prior $\pi_2(\alpha)$. We have used the uniform and exponential distributions in our applications. As is usually the case in hierarchical Bayesian analysis, this second stage prior is not very influential as far as inferences on θ_i are concerned. In fact, we also conduct a sensitivity study on α to measure its influence.

3 Bayesian Estimation of a Monotone Hazard Rate for Complete and Censored Data

Let us suppose that we have a random sample of survival times T_l $(l = 1...n)$ of size n, from a population with density function f. Without censoring, we observe $T_1, ..., T_n$. Under the random right censorship model (*cf.* Klein and Moeschberger (1997)), however, each T_l $(l = 1...n)$ is associated with a random variable, C_l, called a censoring time. The T_i and C_j are assumed independent (i, $j = 1..n$). Let us also assume that the C_l are i.i.d. with distribution function $G(t)$ and density $g(t)$. Under this model we observe $X_l = \min(T_l, C_l)$ and δ_l, an indicator variable, equal to 1 if $T_l \leq C_l$ and 0 if not.

Let n_l denote the number of observations of the X_l $(l = 1...n)$, which fall in the set $A_l \cap A_{l+1}^C$ (i.e. those in A_l but not in A_{l+1}), $l = 1, \ldots k$ respectively. Further, let d_l denote the number of these n_l observations where $\delta_l = 1$, that is, $T_l \leq C_l$. Then note that $n = \sum_{l=1}^{k} n_l$. Let $d = \sum_{l=1}^{k} d_l$ be the number of uncensored observations. It should be noted that $n_l = d_l$, $l = 1, \ldots k$ and that $n = d$ in the uncensored case. (Let $n_0 = d_0 = 0$ for notational convenience.) Also, let $x_{(1)} \leq x_{(2)} \leq \ldots \leq x_{(n)}$ denote the order statistics of the

X_i. The likelihood function can now be expressed in the following form:

$$\begin{aligned} f(\widetilde{x}|\widetilde{\theta}) &= \prod_{l=1}^{n} f(x_l|\widetilde{\theta}) \\ &= \left\{\prod_{\substack{\ell=1\\ \delta_\ell=1}}^{n}\left\{h(x_\ell|\widetilde{\theta})\right\}\right\}\left\{\exp\left(-\sum_{\ell=1}^{n}\int_{-\infty}^{x_\ell} h(u)du\right)\right\}\left\{\prod_{l=1}^{n}\left[g(x_l)^{1-\delta_l}\left(1-G(x_l)\right)^{\delta_l}\right]\right\} \\ &= \left\{\prod_{\substack{\ell=1\\ \delta_\ell=1}}^{n}\left[\sum_{i=1}^{k}\theta_i I_{A_i}(x_l)\right]\right\}\left\{\exp\left(-\sum_{\ell=1}^{n}\sum_{i=1}^{k}\theta_i\mu(A_i\cap[0,x_l])\right)\right\}\times \\ &\qquad \left\{\prod_{l=1}^{n}\left[g(x_l)^{1-\delta_l}\left(1-G(x_l)\right)^{\delta_l}\right]\right\}. \end{aligned}$$

It should be noted that the first product only involves the censored observations while the exponent of the exponential function involves all observations. and the third product does not depend on the parameters to be estimated. Now $f(\widetilde{x}|\widetilde{\theta})$ can be written as:

$$\begin{aligned} f(\widetilde{x}|\widetilde{\theta}) &= \{\theta_1^{d_1}(\theta_1+\theta_2)^{d_2}\ldots(\theta_1+\ldots+\theta_k)^{d_k} \\ &\qquad \times\exp\big(-\sum_{i}^{k}\theta_i\{\sum_{j=n_1+\ldots+n_{i-1}+1}^{n} x_{(j)}-(n-\sum_{l=1}^{i-1}n_l)a_i\}\big)\} \\ &\qquad \times\left\{\prod_{l=1}^{n}\left[g(x_l)^{1-\delta_l}\left(1-G(x_l)\right)^{\delta_l}\right]\right\}. \end{aligned}$$

Since $\pi_1(\widetilde{\theta}|\alpha)=\alpha^k\exp(-\alpha\sum_{i=1}^{k}\theta_i)$, we obtain

$$\begin{aligned} \pi_1(\widetilde{\theta}|\widetilde{x},\alpha) &\propto f(\widetilde{x}|\widetilde{\theta})\pi_1(\widetilde{\theta}|\alpha) \\ &\propto \theta_1^{d_1}(\theta_1+\theta_2)^{d_2}\ldots(\theta_1+\ldots+\theta_k)^{d_k}.\alpha^k \\ &\qquad \times\exp\left(-\sum_{i}^{k}\theta_i\left\{\alpha+\sum_{j=n_1+\ldots+n_{i-1}+1}^{n} x_{(j)}-(n-\sum_{l=1}^{i-1}n_l)a_i\right\}\right) \\ &\qquad \times\left\{\prod_{l=1}^{n}\left[g(x_l)^{1-\delta_l}\left(1-G(x_l)\right)^{\delta_l}\right]\right\}. \end{aligned} \tag{3.1}$$

Computation of $\mathbf{E}(\theta_l|\widetilde{x},\alpha)$ and $\mathbf{Var}(\theta_l|\widetilde{x},\alpha)$ for $1\le l\le k$ are required in order to study the sensitivity of the inferences on $\widetilde{\theta}$ with respect to the prior specifications on α. Clearly, for $1\le l\le k$, and $q>0$,

$$\begin{aligned} \mathbf{E}(\theta_l^q|\widetilde{x},\alpha) &= \int\theta_l^q\pi_1(\widetilde{\theta}|\widetilde{x},\alpha)\,d\widetilde{\theta} \\ &= \frac{\int\theta_l^q f(\widetilde{x}|\widetilde{\theta})\pi_1(\widetilde{\theta}|\alpha)\,d\widetilde{\theta}}{\int f(\widetilde{x}|\widetilde{\theta})\pi_1(\widetilde{\theta}|\alpha)\,d\widetilde{\theta}}. \end{aligned} \tag{3.2}$$

We first assume α known and later introduce a second stage prior on α.

We provide below exact expressions for these posterior expectations with censoring and three numerical approximations which are computationally reliable. Note that the binomial expansion yields,

$$\begin{aligned}\theta_1^{d_1}(\theta_1+\theta_2)^{d_2}\dots(\theta_1+\dots+\theta_k)^{d_k} &= \theta_1^{d_1}\left\{\prod_{h=2}^{k}\left[\sum_{j_1^h+\dots+j_{h-1}^h\le d_h}\binom{d_h}{j_1^h,\dots,j_{h-1}^h}\theta_1^{j_1^h}\dots\theta_{h-1}^{j_{h-1}^h}\theta_h^{d_h-j_1^h-\dots-j_{h-1}^h}\right]\right\}\\ &= \sum_{j_1^2\le d_2}\cdots\sum_{j_1^k+\dots+j_{k-1}^k\le d_k}\binom{d_2}{j_1^2}\cdots\binom{d_k}{j_1^k,\dots,j_{k-1}^k}\theta_1^{m_1}\dots\theta_k^{m_k},\end{aligned}$$

where

$$m_i=\begin{cases} d_1+\sum_{h=2}^{k} j_1^h, & \text{if} \quad i=1\\ d_i-\sum_{h=1}^{i-1} j_h^i+\sum_{h=i}^{k} j_i^h, & \text{if} \quad 2\le i\le k-1\\ d_k-\sum_{h=1}^{k-1} j_h^i, & \text{if} \quad i=k\,.\end{cases}$$

Let z_i be defined by

$$z_i=\alpha+\sum_{j=n_1+\dots+n_{i-1}+1}^{n} x_{(j)}-(n-\sum_{l=1}^{i-1}n_l)a_i. \tag{3.3}$$

Then

$$\begin{aligned}\int\theta_l^q f(\widetilde{x}|\widetilde{\theta})\pi_1(\widetilde{\theta}|\alpha)d\widetilde{\theta} &= \alpha^k\int\sum_{j_1^2\le d_2}\cdots\sum_{j_1^k+\dots+j_{k-1}^k\le d_k}\binom{d_2}{j_1^2}\cdots\binom{d_k}{j_1^k,\dots,j_{k-1}^k}\theta_1^{m_1}\dots\theta_l^{m_l+q}\dots\theta_k^{m_k}e^{-\sum_i^k\theta_i z_i}d\widetilde{\theta}\\ &= \alpha^k\sum_{j_1^2\le d_2}\cdots\sum_{j_1^k+\dots+j_{k-1}^k\le d_k}\binom{d_2}{j_1^2}\cdots\binom{d_k}{j_1^k,\dots,j_{k-1}^k}\int\theta_1^{m_1}\dots\theta_l^{m_l+q}\dots\theta_k^{m_k}e^{-\sum_i^k\theta_i z_i}d\widetilde{\theta}\\ &= \alpha^k\sum_{j_1^2\le d_2}\cdots\sum_{j_1^k+\dots+j_{k-1}^k\le d_k}\binom{d_2}{j_1^2}\cdots\binom{d_k}{j_1^k,\dots,j_{k-1}^k}\\ &\qquad\times\frac{\Gamma(m_1+1)}{z_1^{m_1+1}}\cdots\frac{\Gamma(m_l+q+1)}{z_l^{m_l+q+1}}\cdots\frac{\Gamma(m_k+1)}{z_k^{m_k+1}}\\ &\qquad\times\left\{\prod_{l=1}^{n}\left[g(x_l)^{1-\delta_l}\left(1-G(x_l)\right)^{\delta_l}\right]\right\}.\end{aligned} \tag{3.4}$$

Therefore an exact expression for $\mathbf{E}(\theta_l^q|\widetilde{x},\alpha)$ is

$$\mathbf{E}(\theta_l^q|\widetilde{x},\alpha)=\frac{\sum_{j_1^2\le d_2}\cdots\sum_{j_1^k+\dots+j_{k-1}^k\le d_k}\binom{d_2}{j_1^2}\cdots\binom{d_k}{j_1^k,\dots,j_{k-1}^k}\frac{\Gamma(m_1+1)}{z_1^{m_1+1}}\cdots\frac{\Gamma(m_l+q+1)}{z_l^{m_l+q+1}}\cdots\frac{\Gamma(m_k+1)}{z_k^{m_k+1}}}{\sum_{j_1^2\le d_2}\cdots\sum_{j_1^k+\dots+j_{k-1}^k\le d_k}\binom{d_2}{j_1^2}\cdots\binom{d_k}{j_1^k,\dots,j_{k-1}^k}\frac{\Gamma(m_1+1)}{z_1^{m_1+1}}\cdots\frac{\Gamma(m_l+1)}{z_l^{m_l+1}}\cdots\frac{\Gamma(m_k+1)}{z_k^{m_k+1}}}. \tag{3.5}$$

For small or moderate values of n_i the expression given above for $\mathbf{E}(\theta_l^q|\widetilde{x},\alpha)$ is useful for computations. However, for large n_i, it is obvious that this expression will require too

many computations rendering it quite impractical. We give two alternative methods which can be reliably used for computations.

The first method proposed is the Monte-Carlo importance sampling technique. To use this technique, note first of all that,

$$\begin{aligned}
\pi_1(\widetilde{\theta}|\widetilde{x},\alpha) &\propto \alpha^k \theta_1^{d_1}(\theta_1+\theta_2)^{d_2}\dots(\theta_1+\dots+\theta_k)^{d_k}\exp\left(-\sum_{i=1}^k \theta_i z_i\right)\\
&= \alpha^k \theta_1^{d_1}\theta_2^{d_2}\dots\theta_k^{d_k}\exp\left(-\sum_{i=1}^k z_i\theta_i\right)\prod_{i=1}^k\left(1+\frac{\theta_1}{\theta_i}+\dots\frac{\theta_{i-1}}{\theta_i}\right)^{d_i}\\
&= \alpha^k\left[\prod_{i=1}^k \theta_i^{d_i}\exp(-z_i\theta_i)\right]\prod_{i=1}^k\left(1+\frac{\theta_1}{\theta_i}+\dots\frac{\theta_{i-1}}{\theta_i}\right)^{d_i}\\
&= p(\theta_1,\dots,\theta_k)w(\theta_1,\dots,\theta_k) \qquad (3.6)
\end{aligned}$$

where the function $p(\theta_1,\dots,\theta_k)$ is proportional to the joint density of independent $Gamma(d_i+1,z_i)$ random variables and

$$w(\theta_1,\dots,\theta_k)=\alpha^k\prod_{i=1}^k\left(1+\frac{\theta_1}{\theta_i}+\dots\frac{\theta_{i-1}}{\theta_i}\right)^{d_i}.$$

We have omitted the terms that do not depend on θ or α. Note that although α^k is not necessary here, it will be used later when a second stage prior on α will be employed.

Therefore to compute

$$\begin{aligned}
\mathbf{E}(\theta_l^q|\widetilde{x},\alpha) &= \int \theta_l^q \pi_1(\widetilde{\theta}|\widetilde{x},\alpha)\,d\widetilde{\theta}\\
&= \frac{\int \theta_l^q w(\widetilde{\theta})p(\widetilde{\theta})\,d\widetilde{\theta}}{\int w(\widetilde{\theta})p(\widetilde{\theta})\,d\widetilde{\theta}},
\end{aligned}$$

we can use $p(\widetilde{\theta})$ as the importance function in Monte-Carlo sampling. Then we obtain,

$$\mathbf{E}(\theta_l^q|\widetilde{x},\alpha)=\frac{\frac{1}{M}\sum_{j=1}^M \theta_{l,j}^q w(\widetilde{\theta}_{l,j})}{\frac{1}{M}\sum_{j=1}^M w(\widetilde{\theta}_{l,j})}, \qquad (3.7)$$

where $\widetilde{\theta}_j$, $j=1,\dots M$ are generated from the density $p(\widetilde{\theta})$.

The second technique that can be used is the Laplace approximation method of Tierney and Kadane (1986) and Tierney *et al.* (1989). For large d_i, this technique is especially useful. Rewriting (3.6) in the form

$$\begin{aligned}
\pi_1(\widetilde{\theta}|\widetilde{x},\alpha) &\propto \exp(-\sum_{i=1}^k z_i\theta_i+\sum_{i=1}^k d_i\log(\theta_1+\dots,\theta_i)+k\log(\alpha))\\
&\propto \exp(\mathcal{L}(\widetilde{\theta})),
\end{aligned}$$

where $\mathcal{L}(\widetilde{\theta}) = -\sum_{i=1}^{k} z_i\theta_i + \sum_{i=1}^{k} d_i \log(\theta_1 + \ldots, \theta_i) + k\log(\alpha)$. We also obtain

$$\begin{aligned} \mathbf{E}(\theta_l^q|\widetilde{x},\alpha) &= \frac{\int \theta_l^q \exp(\mathcal{L}(\widetilde{\theta}))\, d\widetilde{\theta}}{\int \exp(\mathcal{L}(\widetilde{\theta}))\, d\widetilde{\theta}} \\ &= \frac{\int \exp(\mathcal{L}^*(\widetilde{\theta}))\, d\widetilde{\theta}}{\int \exp(\mathcal{L}(\widetilde{\theta}))\, d\widetilde{\theta}}, \end{aligned} \tag{3.8}$$

where $\mathcal{L}^*(\widetilde{\theta}) = \mathcal{L}(\widetilde{\theta}) + q\log\theta_l$.

Let $\hat{\theta}$ and $\hat{\theta^*}$, respectively, denote the maxima of $\mathcal{L}(\theta)$ and $\mathcal{L}^*(\theta)$. Then the Laplace approximation yields,

$$\mathbf{E}(\theta_l^q|\widetilde{x},\alpha) \approx \exp(\mathcal{L}^*(\hat{\theta^*}) - \mathcal{L}(\hat{\theta})) \left(\frac{|H(\hat{\theta})|}{|H^*(\hat{\theta^*})|}\right)^{1/2},$$

where $H(\theta)$ and $H^*(\theta)$ are the Hessian matrices of $\mathcal{L}(\theta)$ and $\mathcal{L}^*(\theta)$, respectively. The partial derivatives of $\mathcal{L}$ and $\mathcal{L}^*$ of order 1 and 2 required for this approximation are very easy to obtain. Specifically,

$$\begin{cases} \frac{\partial}{\partial\theta_i}\mathcal{L}(\widetilde{\theta}) = -z_i + \sum_{j=i}^{k} \frac{d_j}{\theta_1+\ldots+\theta_j} \\ \frac{\partial}{\partial\theta_i}\mathcal{L}^*(\widetilde{\theta}) = \begin{cases} \frac{\partial}{\partial\theta_i}\mathcal{L}(\widetilde{\theta}) & \text{if} \quad i \neq l \\ \frac{\partial}{\partial\theta_i}\mathcal{L}(\widetilde{\theta}) + \frac{q}{\theta_l} & \text{if} \quad i = l\,. \end{cases} \end{cases}$$

Setting $\frac{\partial}{\partial\theta}\mathcal{L}(\theta)$ to 0 yields a set of equations for which the solution, $\hat{\theta}$, is easily obtained. To obtain $\hat{\theta^*}$, it is straight forward to solve $\frac{\partial}{\partial\theta}\mathcal{L}^*(\theta) = 0$, iteratively, starting at $\theta = \hat{\theta}$. To compute the Hessian matrices, note that

$$\begin{cases} \frac{\partial^2}{\partial\theta_i\partial\theta_j}\mathcal{L}(\widetilde{\theta}) = & \begin{cases} -\sum_{s=i}^{k} \frac{d_s}{(\theta_1+\ldots+\theta_s)^2}, & \text{if} \quad j \leq i \\ -\sum_{s=j}^{k} \frac{d_s}{(\theta_1+\ldots+\theta_s)^2}, & \text{if} \quad j > i \end{cases} \\ \frac{\partial^2}{\partial\theta_i\partial\theta_j}\mathcal{L}^*(\widetilde{\theta}) = & \begin{cases} \frac{\partial^2}{\partial\theta_i\partial\theta_j}\mathcal{L}(\widetilde{\theta}) & \text{if} \quad i \neq l \text{ or } j \neq l \\ \frac{\partial^2}{\partial\theta_l^2}\mathcal{L}(\widetilde{\theta}) + \frac{q}{\theta_l^2} & \text{if} \quad i = j = l\,. \end{cases} \end{cases}$$

Now let us assume that we have a second stage prior $\pi_2(\alpha)$ on α. The final estimates $\mathbf{E}(\theta_l^q)$ now with a second stage prior on α for each of the methods described can be obtained using analogous methods. The Laplace approximation is very easy to extend to this case since

$$\mathbf{E}(\theta_l^q|\widetilde{x}) = \frac{\int \theta_l^q \pi_1(\widetilde{\theta}|\widetilde{x},\alpha)\pi_2(\alpha)\, d\widetilde{\theta}\, d\alpha}{\int \pi_1(\widetilde{\theta}|\widetilde{x},\alpha)\pi_2(\alpha)\, d\widetilde{\theta}\, d\alpha} \tag{3.9}$$

where $\pi_2(\alpha)$ has an uniform or an exponential density. The expression of $\mathcal{L}(\alpha,\widetilde{\theta})$ and $\mathcal{L}^*(\alpha,\widetilde{\theta})$ with the uniform density for π_2 are the same as those shown above for $\mathcal{L}(\widetilde{\theta})$ and $\mathcal{L}^*(\widetilde{\theta})$.

However for an exponential distribution with parameter a as the prior for π_2, the expressions of $\mathcal{L}(\alpha,\widetilde{\theta})$ and $\mathcal{L}^*(\alpha,\widetilde{\theta})$ $(=\mathcal{L}(\alpha,\widetilde{\theta})+q\log\theta_l)$ become:

$$\mathcal{L}(\alpha,\widetilde{\theta}) = -(a+\sum_{i=1}^{k}\theta_i)\alpha + k\log(\alpha) - \sum_{i=1}^{k} b_i\theta_i + \sum_{i=1}^{k} d_i\log(\theta_1+\ldots+\theta_i)$$

where

$$b_i = \sum_{j=n_1+\ldots+n_{i-1}+1}^{n} x_{(j)} - (n-\sum_{j=1}^{i-1} n_j)a_i\,.$$

The first and second order partial derivatives of $\mathcal{L}$ and $\mathcal{L}^*$ are thus obtained as:

$$\begin{aligned}
\frac{\partial}{\partial\alpha}\mathcal{L}(\alpha,\widetilde{\theta}) &= -(a+\sum_{i=1}^{k}\theta_i)+\frac{k}{\alpha}\\
\frac{\partial}{\partial\theta_i}\mathcal{L}(\alpha,\widetilde{\theta}) &= -\alpha - b_i + \sum_{j=i}^{k}\frac{d_j}{\theta_1+\ldots+\theta_j} \qquad (3.10)\\
\text{and}\qquad \frac{\partial}{\partial\theta_i}\mathcal{L}^*(\alpha,\widetilde{\theta}) &= \begin{cases} \frac{\partial}{\partial\theta_i}\mathcal{L}(\widetilde{\theta}) & \text{for } i\neq l\\ \frac{\partial}{\partial\theta_i}\mathcal{L}(\widetilde{\theta})+\frac{q}{\theta_l} & \text{for } i=l\,. \end{cases}
\end{aligned}$$

Solving $\frac{\partial}{\partial\alpha}\mathcal{L}(\alpha,\widetilde{\theta})=0$ and $\frac{\partial}{\partial\theta_k}\mathcal{L}(\alpha,\widetilde{\theta})=0$ implies that the solution for $\widehat{\alpha}=k/(a+\sum_{i=1}^{k}\theta_i)$ can be obtained by solving the following quadratic equation for $\sum_{i=1}^{k}\theta_i$:

$$-\frac{k}{a+\sum_{s=1}^{k}\theta_s} - b_k + \frac{d_k}{\sum_{s=1}^{k}\theta_s} = 0\,, \qquad (3.11)$$

and let us denote one of the positive solutions of (3.11) by $\widehat{\left(\sum_{s=1}^{k}\theta_s\right)}$.

Now setting $\frac{\partial}{\partial\theta_i}\mathcal{L}(\alpha,\widetilde{\theta})$ to 0, for $i=1,\cdots,k-1$, yields a set of equations for which the solutions $\hat{\theta}_i$ $(i=1,\cdots,k-1)$ is easily obtained. In fact, iterative solutions to (3.10) yields an estimator $\widehat{\left(\sum_{s=1}^{m}\theta_s\right)}$ of $\left(\sum_{s=1}^{m}\theta_s\right)$ for $m=k-1,\cdots,1$. For example, if we let $i=k$ and $i=k-1$ in equation (3.10) and solve simultaneously, we obtain:

$$\widehat{\left(\sum_{s=1}^{k-1}\theta_s\right)} = \frac{d_{k-1}}{b_{k-1}-b_k}\,. \qquad (3.12)$$

In an analogous manner the estimates $\widehat{\left(\sum_{s=1}^{m}\theta_s\right)}$ can be obtained for $m=k-2,\cdots,1$. We then define $\widehat{\theta}_m = \widehat{\left(\sum_{s=1}^{m}\theta_s\right)} - \widehat{\left(\sum_{s=1}^{m-1}\theta_s\right)}$.

We note that the estimator of θ_k can be obtained by the solution in (3.11) and the estimator of $\left(\sum_{s=1}^{k-1}\theta_s\right)$.

To obtain $\hat{\theta^*}$, it is straight forward to solve $\frac{\partial}{\partial\theta}\mathcal{L}^*(\alpha,\widetilde{\theta})=0$, iteratively, starting at $\theta=\hat{\theta}$. To compute the Hessian matrices, note that

$$\begin{cases} \frac{\partial^2}{\partial\theta_i\partial\theta_j}\mathcal{L}(\alpha,\widetilde{\theta}) = \begin{cases} -\sum_{s=i}^{k}\frac{d_s}{(\theta_1+\ldots+\theta_s)^2} \text{ for } & j\le i \\ -\sum_{s=j}^{k}\frac{d_s}{(\theta_1+\ldots+\theta_s)^2} \text{ for } & j>i \end{cases} \\ \frac{\partial^2}{\partial\theta_i\partial\theta_j}\mathcal{L}^*(\alpha,\widetilde{\theta}) = \begin{cases} \frac{\partial^2}{\partial\theta_i\partial\theta_j}\mathcal{L}(\alpha,\widetilde{\theta}) \text{ for } & i\neq l \text{ or } j\neq l \\ \frac{\partial^2}{\partial\theta_l^2}\mathcal{L}(\alpha,\widetilde{\theta})-\frac{q}{\theta_l^2} \text{ for } & i=j=l \end{cases} \\ \frac{\partial^2}{\partial\theta_i\partial\alpha}\mathcal{L}(\alpha,\widetilde{\theta})=-1 \text{ for } i=1\ldots k \\ \frac{\partial^2}{\partial^2\alpha}\mathcal{L}(\alpha,\widetilde{\theta})=-\frac{k}{\alpha^2}. \end{cases}$$

With a second stage prior on α, the Monte-Carlo importance sampling method needs the following modifications.

Note that in (3.3)

$$z_i=\alpha+\sum_{j=n_1+\ldots+n_{i-1}+1}^{n}x_{(j)}-(n-\sum_{j=1}^{i-1}n_j)a_i=\alpha+b_i,$$

so that

$$\pi_1(\widetilde{\theta}|\widetilde{x},\alpha)\propto\theta_1^{d_1}\ldots(\theta_1+\ldots+\theta_k)^{d_k}\exp(-\sum_{i=1}^{k}b_i\theta_i)\alpha^k\exp(-\alpha\sum_{i=1}^{k}\theta_i).$$

Therefore,

$$\begin{aligned} \pi(\widetilde{\theta}|\widetilde{x}) &= \int\pi_1(\widetilde{\theta}|\widetilde{x},\alpha)\pi_2(\alpha)\,d\alpha \\ &\propto \theta_1^{d_1}\ldots(\theta_1+\ldots+\theta_k)^{d_k}\exp(-\sum_{i=1}^{k}b_i\theta_i)\int\alpha^k\exp(-\alpha\sum_{i=1}^{k}\theta_i)\exp(-a\alpha)\,d\alpha \\ &= \theta_1^{d_1}\ldots(\theta_1+\ldots+\theta_k)^{d_k}\exp(-\sum_{i=1}^{k}b_i\theta_i)\left[a+\sum_{i=1}^{k}\theta_i\right]^{-(k+1)} \\ &= \left[\prod_{i=1}^{k}\theta_i^{d_i}\exp(-b_i\theta_i)\right]\prod_{i=2}^{k}\left(1+\frac{\theta_1}{\theta_i}+\ldots+\frac{\theta_{i-1}}{\theta_i}\right)^{d_i}\left[a+\sum_{i=1}^{k}\theta_i\right]^{-(k+1)} \\ &= p(\widetilde{\theta})g(\widetilde{\theta}), \end{aligned}$$

and hence $p(\widetilde{\theta})=\prod_{i=1}^{k}\theta_i^{d_i}\exp(-b_i\theta_i)$ can be used as the new importance function.

Once $\hat{\theta}_l=\mathbf{E}(\theta_l|\widetilde{x})$ are computed, it is easy to obtain

$$\begin{aligned} \mathbf{E}(h(t)|\widetilde{x}) &= \mathbf{E}\left[\sum_{i=1}^{k}\theta_i I_{A_i}(t)|\widetilde{x}\right] \\ &= \sum_{i=1}^{k}\hat{\theta}_i I_{A_i}(t). \end{aligned}$$

To obtain $\mathbf{Var}(h(t)|\widetilde{x})$ we need $\mathbf{E}(\theta_i\theta_j|\widetilde{x})$ also as can be seen below.

$$\begin{aligned}\mathbf{Var}(h(t)|\widetilde{x}) &= \mathbf{Var}\left[\sum_{i=1}^{k}\theta_i I_{A_i}(t)|\widetilde{x}\right]\\ &= \sum_{i=1}^{k}\mathbf{Var}(\theta_1+\ldots+\theta_i|\widetilde{x})I_{A_i\cap A_{i+1}^C}(t)\\ &= \sum_{i=1}^{k}\Big\{\sum_{j=1}^{k}\mathbf{Var}(\theta_j|\widetilde{x})+2\sum_{j=1}^{k}\sum_{l=1}^{j-1}\mathbf{Cov}(\theta_j,\theta_l|\widetilde{x})\Big\}I_{A_i\cap A_{i+1}^C}(t).\end{aligned}$$

However, it is easy to compute $\mathbf{E}(\theta_i\theta_j|\widetilde{x})$, $i\neq j$, following the computational techniques described earlier.

Quite often the estimation of the survival function, $S(y)$, is of interest too. The common estimate is the Kaplan-Meier estimate. We are able to provide a hierarchical Bayes estimate instead. Note that

$$\begin{aligned}S(y) &= \exp(-\int_0^y h(t)dt)\\ &= \exp\left(-\sum_{i=1}^{k}\theta_i\mu(A_i\cap[0,y])\right)\\ &= \exp\left(-\sum_{i=1}^{k}s_i\theta_i\right),\end{aligned}$$

where $s_i=\mu(A_i\cap[0,y])$. Therefore,

$$\mathbf{E}\left[S(y)|\widetilde{x}\right]=\mathbf{E}\left[\exp\left(-\sum_{i=1}^{k}s_i\theta_i\right)|\widetilde{x}\right],$$

which is the moment generating function of $\widetilde{\theta}|\widetilde{x}$ evaluated at $-\widetilde{s}$, where $\widetilde{s}=(s_1,\ldots,s_k)'$. The Monte-Carlo sampling technique and the Laplace approximation technique are especially suited for computing this. To obtain

$$\mathbf{Var}\left[S(y)|\widetilde{x}\right]=\mathbf{E}\left[S^2(y)|\widetilde{x}\right]-\mathbf{E}^2\left[S(y)|\widetilde{x}\right],$$

note that

$$\mathbf{E}\left[S^2(y)|\widetilde{x}\right]=\mathbf{E}\left[\exp\left(-2\sum_{i=1}^{k}s_i\theta_i\right)|\widetilde{x}\right],$$

which is the moment generating function of $\widetilde{\theta}|\widetilde{x}$ evaluated at $-2\widetilde{s}$. It is also possible to to approximate $\mathbf{E}\left[S(y)|\widetilde{x}\right]$ and $\mathbf{Var}\left[S(y)|\widetilde{x}\right]$ by expanding $S(y)=\exp\left(-\sum_{i=1}^{k}s_i\theta_i\right)$ in a Taylor series (with finitely many terms) and replacing the θ_j terms by their corresponding posterior estimates.

4 Data, Examples and Simulations

Here we illustrate the methodology proposed above on a complete and a censored data set as well as by using extensive simulation studies for both cases.

For the Monte-Carlo importance sampling, we have employed a product of independent gamma densities as shown in equation (3.6) as the importance function. Since this involves standard distributions, we found it convenient and easy to handle these computations using standard routines to generate the required gamma variates.

The programming was done on a Pentium 233 MGHZ using Absoft Pro Fortran Version 4.4. The number J of simulations performed in the Monte Carlo was chosen so that the difference between the sums in equation (3.1) for $M = J - 1$ and $M = J$ was less than $\delta = 10^{-7}$, thus guaranteeing convergence.

For the Laplace approximation methods, the following subroutines from the NAG (Numerical Algorithm Group) mark 16 library were used: $G05FFF$ to generate the gamma variates; $D02AHF$ for the integration; $F03ABF$ for calculating the determinants of the Hessian matrices, $C05PBF$ for the solution of the system of nonlinear equations.

4.1 Analysis of the Aluminum Cell Data from Whitmore and Gentleman (1982)

In the uncensored case, we have chosen to illustrate the methods described here on an aluminium cell data set containing both ordinary cells as well as experimental ones. This data set was originally proposed as a case study by Gentleman and Whitmore (1982) and can be briefly described as follows. An ordinary aluminum plant can have several hundreds of cells in operation at the same time. When the strike started in 1967, the electric power was cut. This, according to the company, caused losses in production and associated damages to the cells. Gentleman and Whitmore (1982) explained the data as follows:

> "A 1967 strike at a Quebec aluminium smelter resulted in the uncontrolled shutdown of aluminium-reduction cells in the smelter's potrooms. In a subsequent legal action against the union which was before the courts for more than a decade, the company claimed that the shutdown had reduced the operating lives of the hundreds of cells in service at the time. "

Kalbfleisch and Struthers (1982), Thomas (1982) and later Arjas and Liu (1995) studied the question of whether or not reduction in the length of life of the cells is significantly linked to the strike and consequent monetary loss. Thomas (1982) used a Cox proportional hazard model and Kalbfleisch and Struthers (1982) used nonparametric life table methods and a nonparametric point process approach to estimate the actual operating time lost due to the strike. Arjas and Liu (1995) used a hierarchical Bayes model for the loss suffered by company. Both the first two analyses mentioned above indicated that the data could be appropriately modelled by an increasing hazard rate. Thomas (1982), in addition, found a difference between the cells of ordinary type (labelled type A1-A20 in the original data set)

and those of experimental type (originally labelled B to K). The data consists of 499 cells, of which 395 cells were of the standard design. The other 104 were said to be of experimental design. Here A denotes cells of ordinary type and B of experimental type.

The problem under consideration in the case study was whether the strike had an influence on the duration of the cells. In addition, there are physical reasons for hypothesizing that a cell suffering the strike intervention at an earlier age had a higher risk of failure than those cells that were older at the time of the strike. These ideas are captured and explored here by subdividing groups A and B into I (the control group that did not overlap with the strike) and II (the "intervention" group of cells that lived through the strike). Category II cells were further divided into subgroups of type *(y)* young and *(o)* old, where *(y)* represents those cells where the strike started during thier first year of life and *(o)* those that were at least one year old when the strike occurred.

We decided to use the exponential prior with parameter α. The following five age intervals, $A_1 = [0, b)$, $A_2 = [2, b)$, $A_3 = [3, b)$, $A_4 = [4, b)$ and $A_5 = [5, b)$ were used in equation (2.1). This b was chosen equal to 12, an approximation to the number of years to the time that the legal decision was rendered in this case based on this data.

Our first priority was to do a sensitivity analysis analogous to the one presented in Angers and Delampady (1992) for curve fitting and smoothing. It is very important to establish how sensitive the hazard rate estimates are to changes in the prior.

Several different values of α (both larger and smaller than $\alpha = 1$) in the exponential prior distribution have been tested. Assuming an increasing failure rate for each of these subgroups, as indicated by Kalbfleisch and Struthers (1982) to be appropriate, the hierarchical Bayes estimators of the hazard rates were calculated using the Monte-Carlo importance sampling method.

Figure 1 illustrates some of the results of the sensitivity analysis. It shows graphs of the estimated hazard rate for values of $\alpha = 0.25,\ 0.5$ and 1 in the exponential prior. Similar results were obtained for values of $\alpha > 1$. It should be noted that the estimator does not change except in the last interval $[a_5, b]$ and the change is insignificant. As will be remarked later in the simulation study, the estimator of the hazard for the last interval can be quite variable (even more so than in this sensitivity analysis). Note that the values of the hazard function for the remaining intervals are quite stable as the prior varies, leading us to conclude that the results are insensitive to the choice of exponential prior parameter. We subsequently chose $\alpha = 1$ in the analysis of this data.

We consider each of the hypotheses mentioned above; in Figure 2 the cells of experimental type and ordinary type are contrasted; in Figures 3 and 4 the cells that were functioning during the strike were contrasted with those that were not; and in Figures 5 and 6 the cells that were subject to the intervention in the first year of life were contrasted with those who were subject to a later intervention.

As previously indicated by Efron (1980), the graphs of the hazard functions allow us to easily compare the various subgroups. The results can be described as follows. In Figure 2 the hazard rates of the ordinary and experimental cells are more or less comparable. On the

other hand, in Figure 3, except for the cells of longest duration, the hazard rate is higher for the ordinary cells that were functioning during the strike than those that were not. This same phenomenon is even more pronounced for the cells of experimental type in Figure 3. Figure 4 indicates that the hazard rate is higher if the cell was younger at the time the strike occurred (that is, the cell was less than a year old at the time of the strike). Figure 5 illustrates the same phenomenon for cells of experimental type. The hierarchical Bayesian method of estimation allows us to illustrate fairly clearly that differences in the hazard rate exist between the various subgroups studied in Figures 3 to 6. We were able to conclude that the strike tended to shorten the lifetimes of the cells that were in operation during the intervention. Furthermore this effect is seen to be more detrimental to cells when the strike started during the first year of life.

We also felt it would be important to compare the results obtained by Monte-Carlo sampling and the Laplace approximation methods for moderate sample sizes. We repeated some of the previous analyses to do this comparison and found that the results were rather similar for the two methods; however, there are some slight differences between the results that indicated to us that a detailed simulation study would be useful.

4.2 Simulation Studies

In the simulation studies, to compare the Laplace approximation and the Monte-Carlo with importance sampling methods, the *a priori* measures were taken to be the uniform distribution on$(0, \alpha)$ or the exponential distribution with parameter $\alpha = 4$. First 100 simulations of 100 observations were done and the probability of coverage of the 95% Bayesian credible intervals and the mean square error were used to compare the two methods. It should be noted that with the Laplace approximation, some simulations are rejected because of the fixed level of tolerance in the programme to avoid dividing by zero.

In the first simulation study, the following parameters were used to generate the θ_i: $a_1 = 0.0000, a_2 = 1.5000$, $a_3 = 3.0000$ and $b = 10.0000$. The parameter of $\alpha = 4$ was chosen for the exponential and the uniform distribution. As the results for the exponential and the uniform prior distribution were very similar with a slightly higher proportion of rejections with the uniform one, we present only the results for the exponential in Table 1.

In another simulation study, the following different parameters were used to generate the data: $a_1 = 1.0000, a_2 = 2.0000$, $a_3 = 3.5000$ et $b = 10.0000$. The parameters of the exponential and uniform distributions used to generate the θ_i remained the same. The results of this second study, although not presented here, were similar to the first one with the bias in the estimators of the importance sampling method larger and a slightly higher rejection rate for the Laplace method. We noted that, even if the variances of the estimators by the Monte-Carlo sampling method were smaller, the bias was larger and thus, in the simulations of these parameters, this resulted in the mean square error being smaller for the Laplace method.

The bias of the estimators obtained by the Monte-Carlo sampling method and the number of simulations rejected by the Laplace approximation method were both of some concern.

In order to see if the same phenomena persisted for larger sample sizes, we used 200, 500 and 1000 observations, instead of 100, and repeated the first simulation study. Although the results are not presented here, the bias of the estimators obtained by the Monte-Carlo sampling method remained and the Bayesian credible interval of level 95% for θ_3 did not necessarily contain the true values of θ_3, but the number of simulations rejected by the Laplace approximation method was reduced to 0.2% for n=1000. Even with 200 observations, the Laplace approximation method was superior. Therefore, we conclude that the Laplace approximation method is preferable to the Monte-Carlo sampling method here even for sample sizes as small as 100.

Because of our interest in right censored data, we decided to redo the first simulation study while censoring the data at the rates of 20% and 50%. We kept the same parameters and *a priori* measures used in the first simulation study. We used sample sizes n=200, n=500 and n=1000 and assumed the censoring time was exponentially distributed with parameter μ (where μ was chosen so the censoring would be approximately equal to 20% or 50%).

Note that we haven't included results for n=100 for the censored case because, although we obtained results with the Monte-Carlo importance sampling technique, the Laplace method did not work. For n=200 with censored data, Tables 2 and 3 indicate that the Laplace method gave poor results since it rejected 18.5% and 45% of the simulations for censoring equal to 20% and 50% respectively. Therefore, for n=200 with censored data, we suggest using the Monte-Carlo importance sampling technique.

For n=500 with censored data, Tables 4 and 5 indicate that the Monte-Carlo importance sampling technique gave good results, but the Laplace approximation method was clearly superior with a rejection rate of 5.4% and 5% and the ratios of its MSE to that of the Monte-Carlo sampling technique were .0852 and .4000 when censoring rates were 20% and 50% respectively.

Tables 6 and 7 indicate that the estimator obtained by the Laplace method displays better behavior compared to the one obtained by the Monte-Carlo importance sampling technique for $n = 1000$ and censoring approximatively equal to 20% and 50%. Moreover, rejection rates for the Laplace method are 1.1% and 4% respectively. They also indicate that the Monte-Carlo importance sampling method does not succeed in recovering the true values of θ_2 and θ_3 with 95% Bayesian credible intervals. We therefore recommend the use of the Laplace approximation method over the Monte-Carlo importance sampling one for $n \geq 500$.

4.3 A Censored Data Example: the PBC data, Fleming and Harrington (1991)

Primary Biliary Cirrhosis (PBC) is a very serious liver disease. In the case study presented in Fleming and Harrington (1991), 312 patients were randomized to a clinical trial in primary biliary cirrhosis (PBC) of the liver to compare the drug D-penicillamine with placebo at the Mayo clinic between 1974 and 1984. The observations are incomplete because they are censored if a liver transplantation occurs or if the study ends before the death of the patient.

This right censored data set has been previously analyzed by Fleming and Harrington (1991) and others. The description of the complete set of variables can be found in Fleming and Harrington (1991) and more details of this trial are available in Dickson *et al.* (1989) and Markus *et al.* (1989). It does, however, seem reasonable to model the hazard rate as an increasing function for such a clinical diagnosis. Here we ignore the covariates and use only the survival times in order to illustrate our technique of estimating a monotone hazard rate when the data is randomly right censored.

For the purposes of this study the observations were first divided into 4 groups, each containing approximately the same number of events. Then the fourth group was subdivided into two subgroups, with the subgroup in the tail having slightly more events. The survival time in days was changed to months (by dividing the original data by 30). This resulted in the following intervals $[0\,,b)\,,\ [32\,,b)\,,\ [48\,,b)\,,\ [70\,,b)\,,\ [95\,,b)$, where b was set equal to 200. The results of this analysis using the the Monte-Carlo importance sampling technique are illustrated in Figure 7. Both the estimators of the hazard function in each interval and the 95% Bayes credible intervals are shown in the figure. It should be noted that all the credible intervals overlap at the end points of the intervals with the exception of the intervals $[0\,,32)$ and $[32\,,48)$. This supports the hypothesis that there is a sharp increase in the hazard rate occurring around 32 months.

5 Conclusion

This article is devoted to the study of the hierarchical Bayesian model for the estimation of a monotone hazard rate. Bayesian nonparametric hazard function estimation methodology was first introducted in Dykstra and Laud (1981) and generalized by many authors: Ammann (1985), Thompson and Thavaneswaran (1992), Arjas and Gasbarra (1994), Hjort (1990), Lo and Weng (1989), Ho and Lo (2001), James (2003, 2005) and culminated in the work of Ho (2006), who modelled the hazard rate nonparametrically and described an algorithm that generates sample paths from the posterior by a dynamic Gibbs sampler.

Here, we prefer to take a different approach, extending the method of Broffitt (1984) to include random right censorship. We consider two computationally simple numerical methods to calculate hierarchical Bayesian estimators to estimate a monotone failure rate: Monte-Carlo importance sampling and Laplace approximation. Although our method is more restrictive in some ways since more parameters have to be specified as known, we feel this choice is justified by the ease of the calculations even when estimating hazard functions with many jumps. Most of the previous authors illustrated their techniques on hazard rates with one change point.

The simulation studies showed that for complete data the Laplace method performed well even for sample sizes as small as 100. The Laplace method outperformed the Monte-Carlo importance sampling method since its mean square error was much smaller and the Bayesian credible intervals for the parameters associated with the last interval when using the importance sampling method often failed to cover the true value of the parameter. For

censored data and smaller sample sizes ($n \leq 200$) the Laplace approximation method had difficulties. However, for larger sample sizes and censored data, the Laplace approximation method did very well and outperformed the Monte-Carlo importance sampling one.

The easily computable estimation methods presented here are valuable in elucidating interesting properties of the hazard functions in the two data analyses performed. To conclude, both methods are of interest for the computation of nonparametric hierarchical Bayesian estimators for a monotone hazard rate with complete data and randomly right censored data. For complete data for ($n \geq 100$) and in the censored case for $n \geq 500$ we recommend the Laplace method over the Monte-Carlo importance sampling method.

Acknowledgements

The research of the first and third authors was partially supported by NSERC of Canada and FQRNT of Quebec. We wish to thank Claude Beauchamp for programming assistance. The authors are grateful to the referees for their helpful suggestions which improved this manuscript.

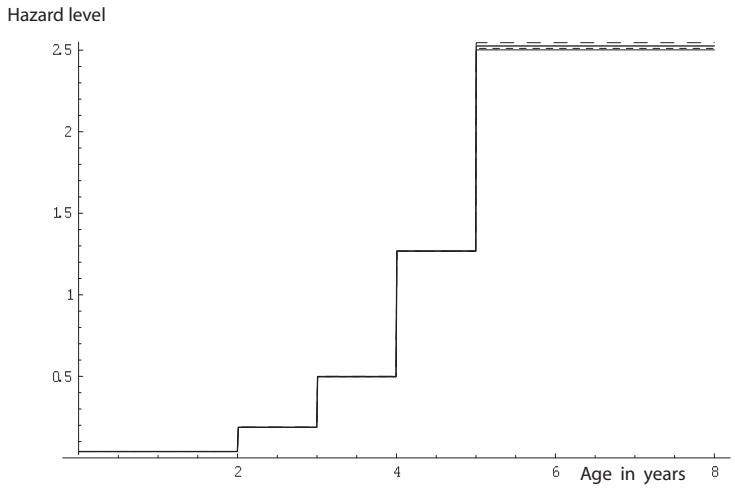

Figure 1: Sensitivity analysis: hazard rates for cells of ordinary type A with 3 different priors.

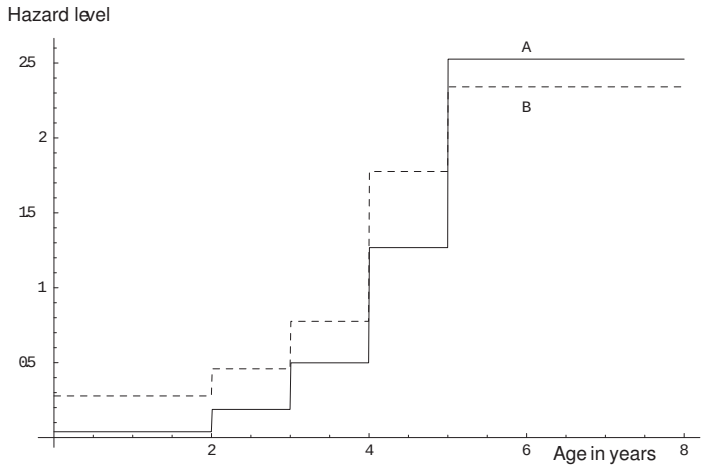

Figure 2: Hazard rates for cells of ordinary type A versus experimental type B.

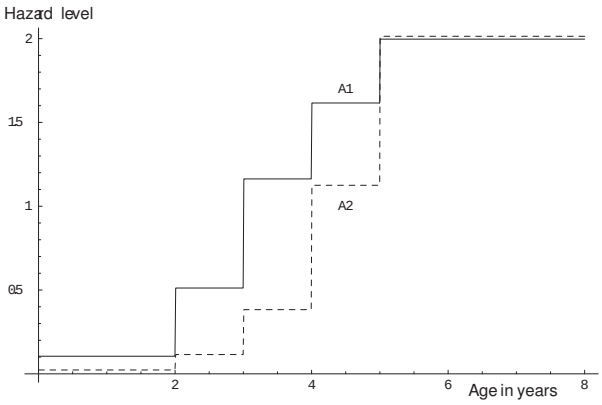

Figure 3: Hazard rates for cells of ordinary type that were (A2) and were not (A1) subject to the intervention of the strike.

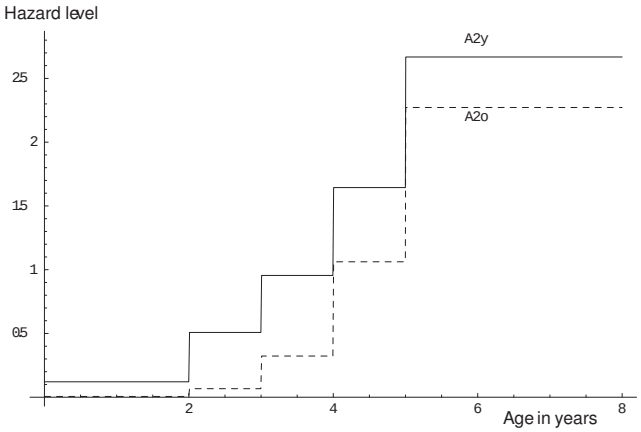

Figure 4: Hazard rates for cells of ordinary type subject to the intervention of the strike at the young (A2y) versus an older age (A2o)

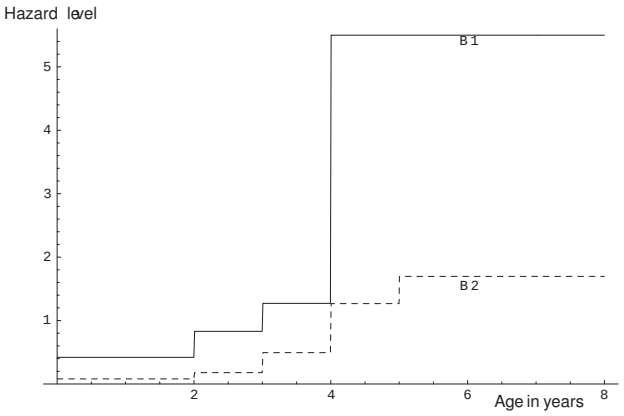

Figure 5: Hazard rates for cells of experimental type that were (B2) and were not (B1) subject to the intervention.

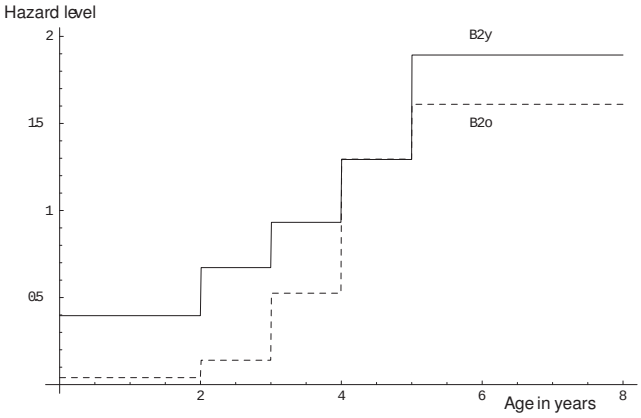

Figure 6: Hazard rates for cells of experimental type subject to the intervention of the strike at a young (B2y) versus an older age (B2o).

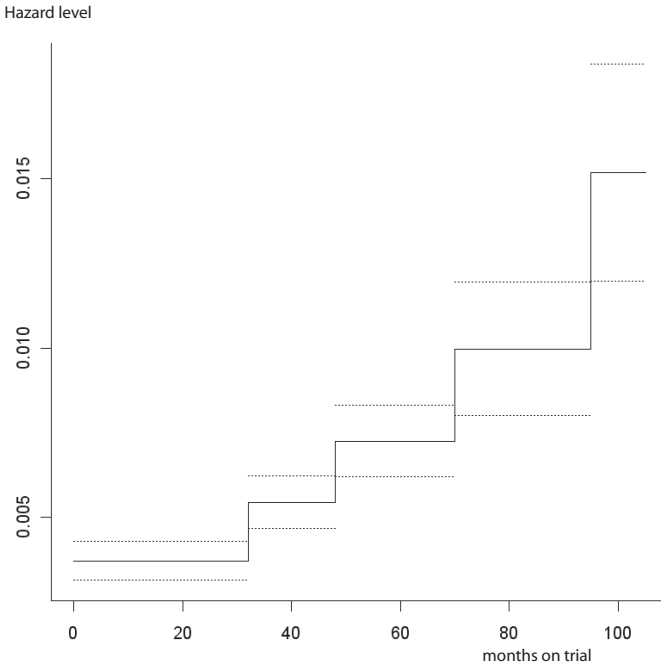

Figure 7: Hazard rates for the PBC data with 95% Bayes credible intervals .

Table 1: Monte-Carlo sampling {Laplace approximation } method with $n = 100$ and a *priori* exponential (α) distribution

θ	$\hat{\theta}$	$Var(\hat{\theta})$	95% confidence interval
0.4201	0.3667 {0.4280}	0.0020{0.0056}	[0.2786 , 0.4549] {[0.2809 , 0.5751]}
0.5123	0.5998 {0.5099}	0.0109{0.0229}	[0.3944 , 0.8051] {[0.2131 , 0.8068]}
0.2323	0.6191 {0.3300}	0.0458{0.0790}	[0.1995 , 1.0386] {[-0.2211 , 0.8811]}
MSE = 0.2138 {0.0553 with rejection rate: 26% }			

Table 2: Monte-Carlo sampling {Laplace approximation } method with $n = 200$ and censoring rate=20%.

θ	$\hat{\theta}$	$Var(\hat{\theta})$	95% confidence interval
0.4201	0.3956 { 0.4402}	0.0012 { 0.0022}	[0.3268 , 0.4644] {[0.3463 , 0.5340]}
0.5123	0.5875 { 0.4380}	0.0087 { 0.0168}	[0.4040 , 0.7709] {[0.1835 , 0.6925]}
0.2323	0.5486 { 0.4127}	0.0561 { 0.1290}	[0.0843 , 1.0129] {[-0.2914 , 1.1169]}
MSE = 0.1915 {0.1352 with rejection rate: 18.5%}			

Table 3: Monte-Carlo sampling {Laplace approximation } method with $n = 200$ and censoring rate=50%.

θ	$\hat{\theta}$	$Var(\hat{\theta})$	95% confidence interval
0.4201	0.4163 { 0.4331}	0.0020 { 0.0026}	[0.3278 , 0.5049]{ [0.3318 , 0.5344]}
0.5123	0.5544 { 0.3380}	0.0197 { 0.0231}	[0.2791 , 0.8297]{ [0.0400 , 0.6361]}
0.2323	0.4918 { 0.4160}	0.1885 { 0.2727}	[-0.3592 , 1.3429]{ [-0.6075 , 1.4396]}
MSE = 0.1818 { 0.0963 with rejection rate: 45%}			

Table 4: Monte-Carlo sampling {Laplace approximation } method with $n = 500$ and censoring rate=20%.

θ	$\hat{\theta}$	$Var(\hat{\theta})$	95% confidence interval
0.4201	0.3697 { 0.4229}	0.0003 { 0.0008}	[0.3358 , 0.4037] { [0.3652 , 0.4807]}
0.5123	0.6297 { 0.4939}	0.0023 { 0.0072}	[0.5339 , 0.7254] { [0.3271 , 0.6607]}
0.2323	0.6896 { 0.2678}	0.0164 { 0.0386}	[0.4382 , 0.9410] { [-0.1173 , 0.6529]}
MSE = 0.2569 { 0.0219 with rejection rate 5.4% }			

Table 5: Monte-Carlo sampling {Laplace approximation } method with $n = 500$ and censoring rate=50%.

θ	$\hat{\theta}$	$Var(\hat{\theta})$	95% confidence interval
0.4201	0.3981 { 0.4236}	0.0006 { 0.0010}	[0.3491 , 0.4471] { [0.3604 , 0.4868]}
0.5123	0.6012 { 0.4709}	0.0056 { 0.0121}	[0.4536 , 0.7487] { [0.2548 , 0.6870]}
0.2323	0.5622 { 0.3570}	0.0613 { 0.1178}	[0.0766 , 1.0477] { [-0.3157 , 1.0299]}
MSE = 0.1830 { 0.0732 with rejection rate: 5%}			

Table 6: Monte-Carlo sampling {Laplace approximation } method with $n = 1000$ and censoring rate=20%.

θ	$\hat{\theta}$	$Var(\hat{\theta})$	95% confidence interval
0.4201	0.3552 { 0.4207}	0.0001 { 0.0004}	[0.3342 , 0.3762]{[0.3800 , 0.4614]}
0.5123	0.6635 { 0.5044}	0.0009 { 0.0036}	[0.6023 , 0.7248]{[0.3859 , 0.6228]}
0.2323	0.7978 { 0.2193}	0.0072 { 0.0183}	[0.6308 , 0.9648]{[-0.0464 , 0.4850]}
MSE = 0.3729{ 0.0195 with rejection rate: 1.1%}			

Table 7: Monte-Carlo sampling {Laplace approximation } method with $n = 1000$ and censoring rate=50%.

θ	$\hat{\theta}$	$Var(\hat{\theta})$	95% confidence interval
0.4201	0.3883 {0.4224 }	0.0002 { 0.0005}	[0.3594 , 0.4173] {[0.3777 , 0.4671]}
0.5123	0.6529 {0.4965 }	0.0019 { 0.0063}	[0.5670 , 0.7389] {[0.3402 , 0.6528]}
0.2323	0.6546 {0.3081 }	0.0262 { 0.0573}	[0.3367 , 0.9724] {[-0.1613 , 0.7776]}
MSE = 0.2524 {0.0536 with rejection rate: 4%}			

References

[1] Ammann, L.P. (1985). Conditional Laplace transforms for Bayesian nonparametric inference in reliability theory. *Stochastic Processes and Applications*, **20**, 197-212.

[2] Angers, J-F. and M. Delampady. (1992). Hierarchical Bayesian curve fitting and smoothing . *Canadian Journal of Statistics,* **20**, 35-49.

[3] Arjas, E. and D. Gasbarra. (1994). Nonparametric Bayesian inference from right censored survival data using the Gibbs sampler. *Statistica Sinica*, **4**, 505-524.

[4] Arjas, E. and L. Liu. (1995). Assessing the losses caused by an industrial intervention: a hierarchical Bayesian approach. *Applied Statistics*, **44**, 357–368.

[5] Ayer, M., H. D. Brunk, G.M. Ewing, W.T. Reid, and E. Silverman. (1955). An empirical distribution function for sampling with incomplete information. *Annals of Mathematical Statistics*, **26**, 641-647.

[6] Barlow, R.E., D.J. Bartholomew, J.M. Bremner, and H.D. Brunk. (1972). *Statistical Inference Under Order Restrictions. The Theory and Application of Isotonic Regression.* John Wiley and Sons, New York.

[7] Barlow, R.E. and H.D. Brunk. (1972). The isotonic regression problem and its dual. *Journal of the American Statistical Association*, **67**, 140-147.

[8] Berger, J.O. and D. Sun. (1996). Bayesian inference for a class of poly-Weibull distributions . In *Bayesian Analysis in Statistics and Econometrics*, (eds: D. A. Berry, K. M. Chaloner, and J. K. Geweke). 101-113. John Wiley and Sons, New York.

[9] Broffitt, J.D. (1984). A Bayes estimator for ordered parameters and isotonic Bayesian graduation. *Scandinavian Actuarial Journal*, **67**, 231-247.

[10] Brunk, H.D. (1955). Maximum likelihood estimates of monotone parameters. *Annals of Mathematical Statistics,* **26**, 607-616.

[11] Burridge, J. (1981). Empirical Bayes Analysis of Survival Time Data. *Journal of the Royal Statistical Society* Series B, **43**, 65-75.

[12] Dey, D.K. and L. Kuo. (1991). A new empirical Bayes estimator with type II censored data. *Computational Statistics and Data Analysis*, **12**, 271-279.

[13] Dickson, E.R., P.M. Grambsch, T.R. Fleming, L.D. Fisher, and A. Langworthy. (1989). Prognosis in primary biliary cirrhosis: model for decision making. *Hepatology*, **10**, 1-7.

[14] Doksum, K. (1974). Tailfree and neutral random probabilities and their posterior distributions. *Annals of Probability*, **2**, 183-201.

[15] Dykstra, R.L. and P.W. Laud. (1981), A Bayesian nonparametric approach to reliability, *Annals of Statistics,* **9**, 356-367.

[16] Efron, B. (1980). Logistic regression, survival analysis and the Kaplan-Meier curve. *Journal of the American Statistical Association,* **83**, 414-420.

[17] Ferguson, T.S. (1973). A Bayesian analysis of some nonparametric problems. *Annals of Statistics*, **1**, 209-230.

[18] Ferguson, T.S. and Phadia, E.G. (1979). Bayesian nonparametric estimation based on censored data. *Annals of Statistics*, **7**, 163-186.

[19] Fleming, T.R. and D.P. Harrington. (1991). *Counting Processes and Survival Analysis*. John Wiley and Sons, New York.

[20] Gentleman, J.F. and G.A. Whitmore. (1982). Case studies in data analysis: measuring the impact of intervention on equipment lives. *Canadian Journal of Statistics,* **10**, 237-259.

[21] Grenander, U. (1956). On the theory of mortality measurement. Part II. *Scandinavisk Aktuarietidskrift*, **39**, 125-153.

[22] Hasselblad, V. (1969). Estimation of finite mixtures of distributions from the exponential family. *Journal of the American Statistical Association,* **64**, 1459-1471.

[23] Hjort, N.L. (1990). Nonparametric Bayes estimators based on beta processes in models for life history data. *Annals of Statistics*, **18**, 1259-1294.

[24] Ho, M.-W. (2006). A Bayes method for a monotone hazard rate via S-Paths. *Annals of Stastistics*, **34**, 820-836.

[25] Ho, M.-W. and A. Y. Lo. (2001). Bayesian nonparametric estimation of a monotone hazard rate. In *System and Bayesian Reliability: Essays in Honor of Professor Richard E. Barlow on His 70th Birthday* (eds: Y. Hayakawa, T. Irony and M. Xie), 301-314, World Scientific.

[26] Huang, J. and J.A. Wellner (1995). Estimation of a monotone density or monotone hazard under random censoring . *Scandinavian Journal of Statistics*, **22**, 3-33.

[27] James, L.F. (2003). Bayesian calculus for gamma processes with applictions to semiparametric models. *Sankhya*, **65**, 159-206.

[28] James, L.F. (2005). Bayesian Poisson process partition calculus with an application to Bayesian Levy moving averages. *Annals of Statistics,* **33**, 1771-1799.

[29] Kalbfleisch, J.D. (1978). Nonparametric Bayesian analysis of survival time data. *Journal of the Royal Statistical Society*, Series B, **40**, 214-221.

[30] Kalbfleisch, J.D. and C.A. Struthers. (1982). An analysis of the Reynolds Metal Company data. *Canadian Journal of Statistics,* **10**, 240-247.

[31] Kaplan, E.L. and P, Meier. (1958). Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association,* **53**, 457-481.

[32] Klein, J.P. and M.L. Moeschberger. (1997). *Survival Analysis: Techiniques for Censored and Truncated Data*. Springer-Verlag, New York.

[33] Kuo, L. and A.F.M. Smith. (1992). Bayesian computations in survival models via the Gibbs sampler . In *Survival Analysis: State of the Art,* (eds: Klein. J.P. and P.Goel). 11-24. Kluwer Academic Publishers, Boston.

[34] Kuo, L. and C. Yiannoustsos. (1993). Empirical Bayes risk evaluation with type II censored data. *Journal of Statistical Computation and Simulation,* **48**, 195-206.

[35] Laird, N.M. (1978). Non-parametric maximum likelihood estimation of a mixing distribution , *Journal of the American Statistical Association,* **73**, 361-379.

[36] Leonard, T., J.S.J. Hsu, and K.W. Tsui. (1989). Bayesian marginal inference. *Journal of the American Statistical Association,* **84**, 1051-1058.

[37] Lo, A.Y. and C.-S. Weng. (1989). On a class of Bayesian nonparametric estimates. II. Hazard rates estimates. *Annals of the Institute of Statistical Mathematics,* **41**, 227-245.

[38] MacGibbon, B., J. Lu and H. Younes. (2002). Limit theorems for asymptotically minimax estimation of a distribution with increasing failure rate under a random mixed censorship/truncation model. *Communications in Statistics: Theory and Methods.* **31**, 1309-1333.

[39] Markus B.H., E.R. Dickson , P.M. Grambsch, T.R. Fleming, V. Mazzaferro, G.B. Klintmalm, R.H. Wiesner, D.H. Van Thiel and T.E. Starzl. (1989). Efficiency of liver transplantation in patients with primary biliary cirrhosis. *New England Journal of Medicine,* **29, 320(26)**, 1709-1713.

[40] Padgett, W.J. and L.J. Wei (1980). Maximum likelihood estimation of a distribution function with increasing failure rate based on censored observations . *Biometrika,* **67**, 470-474.

[41] Prakasa Rao, B.L.S. (1970). Estimation of distributions with monotone failure rate. *Annals of Mathematical Statistics*. **41**, 507-519.

[42] Proschan, F. (1963). Theoretical explanation of observed decreasing failure rate. *Technometrics*, **5**, 375-383.

[43] Robertson, T., F. Wright and R.L. Dykstra. (1988). *Ordered Restricted Statistical Inference*. John Wiley and Sons, New York.

[44] Susarla, V. and J. van Ryzin. (1978). Empirical Bayes estimation of a distribution (survival) function with right censored observations. *Annals of Statistics*, **6**, 740-751.

[45] Thomas, D.C. (1982). Case analysis using Cox's regression model. *Canadian Journal of Statistics,* **10**, 248-253.

[46] Thompson, M.E. and A. Thavaneswaran. (1992). On Bayesian nonparametric estimation for stochastic processes. *Journal of Statistical Planning and Inference*, **33**,131-141.

[47] Tierney, L. and J. Kadane. (1986). Accurate approximations for posterior moments. *Journal of the American Statistical Association*, **81**, 82-86.

[48] Tierney, L., R. Kass, and J. Kadane. (1989). Approximate marginal densities of nonlinear functions. *Biometrika,* **76**, 425-433.