Les Cahiers du GERAD ISSN: 0711–2440

Bayesian Estimation of the Hazard Function with Randomly Right Censored Data

J.-F. Angers
B. MacGibbon

G–2006–24

April 2006
Revised: February 2007

Bayesian Estimation of the Hazard Function with Randomly Right Censored Data

Jean-François Angers
Département de mathématiques et de statistique
Université de Montréal
C.P. 6128, Succ. Centre-ville
Montréal (Québec) Canada H3C 3J7
angers@dms.umontreal.ca

Brenda MacGibbon
GERAD and Département de mathématiques
Université du Québec à Montréal
C.P. 8888, Succ. Centre-ville
Montréal (Québec) Canada H3C 3P8
brenda@math.uqam.ca

April 2006
Revised: February 2007

Les Cahiers du GERAD

G–2006–24

Abstract

We develop a nonparametric Bayesian functional estimation method, using monotone wavelet approximation, for hazard estimation from randomly right censored data. The proposed methodology is compared with that of Arjas and Gasbarra (1994) and Antoniadis et al. (1999) in a simulation study and with real data.

Key Words: Monotone wavelet approximation, Bayesian hierarchical model, generalized linear model.

Résumé

Nous proposons une méthode d'estimation fonctionelle bayésienne non paramétrique pour la fonction de risque en utilisant une approximation par ondelettes monotones pour à partir de données censurées à droite de façon aléatoire en utilisant une approche non paramétrique. Une étude de simulation est faite pour comparer la méthodologie présentée dans cet article à celles introduites dans Arjas et Gasbarra (1994) et Antoniadis et al. (1999). Enfin, un exemple pratique est présenté.

Acknowledgments: The authors wish to acknowledge the support of the Natural Sciences and Engineering Research Council of Canada and to thank the referees for their helpful comments which improved this manuscript. They are also grateful to Dr. R.J. O'Reilly of the Sloan Kettering Memorial Cancer Center and to Dr. Susan Groshen of the University of Southern California, Keck School of Medicine for making available to us the bone marrow transplant data (see Brochstein et al., 1987) that we used to illustrate our method here.

1 Introduction

For survival analysis in medical research, it is useful to have clear summaries of the data for clinicians and, as advocated by Efron (1988), this can often be achieved by a graphical presentation of the hazard function. The data may consist solely of observed survival times or with each time there may be associated a vector of covariates. In the latter case the hazard is often modeled as a product of a hazard function that depends only on time and a function of the covariates which is presumed independent of the time. In analysing such data the two components are often analaysed separately and such an analysis is called semiparametric. For an excellent review of Bayesian semiparametric analysis for even more complex models, we cite Sinha and Dey (1997). Here we prefer to concentrate on the problem of estimating the hazard function without covariates, although indications will be given on how the methodology can be extended to the semiparametric model.

Our purpose here is, threefold: (1) to give an overview of the relevant literature concerning both Bayesian and frequentist nonparametric estimation of the hazard rate, mainly in the case without covariates; (2) to introduce a new nonparametric Bayesian method using monotone wavelets; (3) to compare our estimators with both a frequentist and a Bayesian nonparametric estimator.

We first give a brief overview of frequentist approaches to this problem. Various authors have used splines for estimating the survival function and the hazard rate in the random right censoship model. We cite, in particular, Bloxom (1985), Klotz and Yu (1986), Whittemore and Keller (1986), Efron (1988), O'Sullivan (1988), Jarjoura (1988), Kooperberg and Stone (1992), Kooperberg et al. (1995). Senthilselvan (1987) proposed penalized likelihood methods and Loader (1999) used local likelihood methods for hazard rate estimation with censored data. Kernel estimation of the hazard rate has also proved to be a useful method (see Ramlau-Hansen, 1983; Roussas, 1989, 1990; Izenman and Tran, 1990; Hall et al., 2001). 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 (19), Huang and Wellner (1992) and MacGibbon et al. (2002) which is based on least concave majorants (greatest convex minorants).

Some researchers have previously used orthogonal series methods; in particular Patil (1997) used orthogonal wavelet methods for hazard rate estimation in the uncensored case and Antoniadis et al. (1999) in the random right censorship model. We also cite the theoretical research on wavelet density and hazard estimation by Li (2002, 2006) and Liang et al. (2005).

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. Kuo and Smith (1992) found Bayes estimators of the survival function with censored data using the Gibbs sampler. We also cite other interesting Bayesian research related to hazard rate estimation such as Arjas and Liu (1995) and Berger and Sun (1996).

Among the first to estimate the hazard directly were Dykstra and Laud (1981) and Broffitt (1984). Dykstra and Laud (1981) 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. Bayesian nonparametric hazard function estimation methodology in Dykstra and Laud (1981) was generalized in different ways by Ammann (1985), who used conditional Laplace transforms and Thompson and Thavaneswaran (1992). Hjort (1990) used beta process priors to estimate the cumulative hazard rate process. Further generalizaitons 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.

One of the most interesting methods perhaps, for estimating the hazsard rate, influenced by Dykstra and Laud (1981) is that proposed by Arjas and Gasbarra (1994). Using a hierarchical model structure, they modelled the hazard rate nonparametrically as a jump process having a martingale sturcture with respect to the prior distribution. They describe an algorithm that generates sample paths from the posterior by a dynamic Gibbs sampler and illustrate the method on simulated examples. We have chosen here to compare our proposed method with theirs.

Angers and MacGibbon (2004) developed a Bayesian adaptation of the Antoniadis et al. (1999) method by employing Bayesian nonparametric estimation techniques with Fourier series methods in order to obtain a procedure that is easier to implement. Their method did not perform as well as the method of Antoniadis et al. (1999) for the estimation of the sub-density but in simulations, it was shown to be as good or superior to the method of Antoniadis et al. (1999) for the estimation of the hazard rate. Here we propose the use of monotone wavelet approximation introduced by Anastassiou and Yu (1992) to estimate the sub-density and hazard function.

We proceed as in Antoniadis et al. (1999) to estimate the number of events and the survival functions separately. In order to describe our method here, we follow the description as given by Antoniadis et al. (1999) in Section 1.1. For ease of presentation, Section 1.2 is devoted to recalling the Bayesian approach to linear models. In Section 2, the Bayesian model using monotone wavelet approximation is introduced. In Section 3 we develop our method of Bayesian functional estimation for the hazard rate problem with right censored data. Section 4 contains a simulation study and the comparison of our results with those of Antoniadis et al. (1999). Section 5 presents an application of our method as well as those of Antoniadis et al. (1999) and Arjas and Gasbarra (1994) to a bone marrow transplantation data set and to the Standford heart data. Section 6 consists of some concluding remarks.

1.1 The random right censorship model

Survival analysis is usually based on the study of a group of individuals of size n for which we assume their failure times, the non-negative random variables $T_1, \cdots, T_n$, are independent and identically distributed with distribution function $F(t)$, survival function $S(t) = 1 - F(t)$ and density $f(t)$. However, one of the features that distinguishes the analysis of survival data from classical statistical analysis is the possibility that the data may be incomplete; that is, some individuals may not be observed until failure. For example, some patients will survive to the end of a clinical trial and thus their failure times cannot be observed. If this happens in a random fashion then this type of incompleteness is modeled by assuming that there exist $C_1, \cdots, C_n$ independent and identically distributed random variables with distribution function G and density g representing the censoring mechanism. Instead of observing the complete data $T_1, \cdots, T_n$, we observe $X_i = \min(T_i, C_i)$, $i = 1, \cdots, n$ and an indicator function $\delta_i = 1$ if $T_i \leq C_i$ and $= 0$ if not.

Since the density function of T exists, the hazard rate function can also be defined as

$$\lambda(t) = \frac{f(t)}{1 - F(t)} \qquad F(t) < 1.$$

With T_j, C_j, δ_j defined as above, the observed random variables are then X_j and δ_j. Henceforth we assume that

(a) $T_1, T_2, \cdots, T_n$ are non-negative, independent and identically distributed with distribution function F and density f,

(b) $C_1, C_2, \cdots, C_n$ are non-negative, independent and identically distributed with distribution function G and density g, and

(c) the T's and C's are independent.

In the censored case, if $G(t) < 1$, we have

$$\lambda(t) = \frac{f(t)\{1 - G(t)\}}{\{1 - F(t)\}\{1 - G(t)\}}, \qquad F(t) < 1.$$

If we let $L(t) = P(X_i \leq t)$, then

$$1 - L(t) = \{1 - F(t)\}\{1 - G(t)\}.$$

Letting

$$f^*(t) = f(t)\{1 - G(t)\},$$

be the sub-density of those observations that are still to fail, then clearly

$$\lambda(t) = \frac{f^*(t)}{1 - L(t)}, \qquad L(t) < 1.$$

1.2 The Bayesian model

The estimator for the hazard rate proposed in the next section is obtained by writing the estimation problem using a Bayesian linear model. Hence, for ease of presentation, we first recall the Bayesian linear model as found in Lindley and Smith (1972) and Robert (2001). Let

$$Y = X\theta + \varepsilon,$$

where

$$\begin{aligned}
&Y \text{ is a } n \times 1 \text{ vector of observations,}\\
&X \text{ is a } n \times p \text{ known matrix,}\\
&\theta \text{ is a } p \times 1 \text{ vector of regression coefficients,}\\
&\varepsilon \sim N_n(0, \sigma^2 \mathbf{I}_n).
\end{aligned}$$

Note that X is assumed to be of full rank but even if X is singular the theory holds. Furthermore, σ^2 might be known or unknown. If σ^2 is unknown, it will also be considered as a random variable.

Given this model, the likelihood function is given by

$$\ell(\theta, \sigma^2) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left\{ -\frac{1}{2\sigma^2}(Y - X\theta)'(Y - X\theta) \right\}.$$

The loss function typically used is:

$$L(\theta, \widehat{\theta}) = (\theta - \widehat{\theta})'\mathbf{Q}(\theta - \widehat{\theta}), \tag{1}$$

where $\mathbf{Q}$ is a positive definite matrix.

Let

$$\begin{aligned}
\theta_{LS} &= (X'X)^{-1}X'Y,\\
S &= (Y - X\theta_{LS})'(Y - X\theta_{LS}),
\end{aligned}$$

then the likelihood function can be rewritten as:

$$\begin{aligned}
\ell(\theta, \sigma^2) &\propto \frac{1}{(\sigma^2)^{n/2}} \exp\left\{ -\frac{1}{2\sigma^2} \left[(\theta - \theta_{LS})'X'X(\theta - \theta_{LS}) + S \right] \right\}\\
&= \left(\frac{1}{(\sigma^2)^{p/2}} \exp\left\{ -\frac{1}{2\sigma^2}(\theta - \theta_{LS})'X'X(\theta - \theta_{LS}) \right\} \right)\\
&\qquad \times \left(\frac{1}{(\sigma^2)^{(n-p)/2}} \exp\left\{ -\frac{S}{2\sigma^2} \right\} \right)\\
&= \left[\theta \mid \sigma^2, Y \sim N_p(\theta_{LS}, \sigma^2(X'X)^{-1}) \right]\\
&\qquad \times \left[\sigma^2 \mid Y \sim I\Gamma((n-p-2)/2, S/2) \right],
\end{aligned} \tag{2}$$

where $N_p(\mu, \Sigma)$ denotes the multivariate normal density with mean μ and covariance matrix Σ while $I\Gamma(a, b)$ represents the inverse gamma density with shape parameter a and scale parameter b. (Note that if $\lambda \sim I\Gamma(a, b)$

$$\pi(\lambda) = \left(\begin{cases} \frac{b^a}{\Gamma(a)\lambda^{a+1}}\, \mathrm{e}^{-b/\lambda} & \text{if } \lambda > 0, \\ 0 & \text{otherwise.} \end{cases}\right)$$

Now a conjugate prior for (θ, σ^2) is given by:

$$\theta \mid \sigma^2 \sim N_p(\eta, \sigma^2 C), \tag{3}$$

$$\sigma^2 \sim I\Gamma(\alpha/2, \gamma/2), \tag{4}$$

where η, C, α and γ are assumed to be known.

With the prior model given in equations (3) and (4) and the likelihood function given in equation (2), the posterior density on (θ, σ^2) and the marginal of the least squares estimator θ_{LS} are:

$$\begin{aligned} \theta \mid \sigma^2, Y &\sim N_p(\theta_*, \sigma^2 C_*), \\ \sigma^2 \mid Y &\sim I\Gamma\left(\frac{n+\alpha}{2}, \frac{\gamma_*}{2}\right), \\ \theta_{LS} &\sim T_p\left(n+\alpha-p, \theta_0, \frac{S+\gamma}{n+\alpha-p} A_*\right), \end{aligned} \tag{5}$$

where

$$\begin{aligned} \theta_* &= \theta_{LS} - C_* C^{-1}(\theta_{LS} - \eta), \\ C_* &= (X'X + C^{-1})^{-1} = C - C(C + (X'X)^{-1})^{-1} C, \\ \gamma_* &= S + \gamma + (\theta_{LS} - \eta)' A_*^{-1} (\theta_{LS} - \eta), \\ A_* &= (X'X)^{-1} + C. \end{aligned}$$

Then, under the loss function given in equation (1), the Bayes estimator of θ is given by:

$$\widehat{\theta} = \mathbb{E}[\theta \mid Y] = \theta_{LS} - C_* C^{-1}(\theta_{LS} - \eta).$$

Another intersting loss function is given by

$$L((\theta, \sigma^2), (\widehat{\theta}, \widehat{\sigma}^2)) = (\theta - \widehat{\theta})' \mathbf{Q} (\theta - \widehat{\theta}) + (\sigma^2 - \widehat{\sigma}^2)^2,$$

The associated Bayes estimators of β and σ^2 are given by:

$$\begin{aligned} \widehat{\theta} &= \theta_* = \theta_{LS} - (X^t X + C^{-1})^{-1} C^{-1} (\theta_{LS} - \theta_0), \\ \widehat{\sigma}^2 &= \frac{\gamma_*}{n+\alpha-2} = \frac{S + \gamma + (\theta_{LS} - \theta_0)' A_*^{-1} (\theta_{LS} - \theta_0)}{n+\alpha-2}. \end{aligned}$$

Remark 1. *If we use the reference prior $\pi(\theta,\sigma^2) \propto 1/\sigma^2$, the posterior densities become*

$$\theta \mid \sigma^2, Y \sim N_p(\theta_{LS}, \sigma^2(X'X)^{-1}),$$
$$\sigma^2 \mid Y \sim I\Gamma\left(\frac{n}{2}, \frac{S}{2}\right).$$

(The reference prior is the limit case of equations (3) *and* (4) *with $C^{-1} \to 0$, $\alpha \to 0$ and $\gamma \to 0$.)*

2 Bayesian functional model using monotone wavelet approximation

We first consider the estimation of the cumulative distribution function $L(t)$ and $F^*(t)$ which are monotone functions. Let us consider the general case and assume that $H(t)$ is a monotone nondecreasing function. ($H(t)$ stands for $L(t)$ or $F^*(t)$ depending on the observations considered.) Now, we can develop this Bayesian functional estimation model using the Bayesian linear model described in the previous section as a basis. As many authors including Antoniadis et al. (1999) have indicated, wavelet estimators are ideal for estimating functions with inhomogeneous spatial smoothness. This is often the case with hazard functions. Here we introduce the terminology from Anastassiou and Yu (1992).
Let $\varphi(x)$ denote a bounded right-continuous function on $\mathbb{R}$ with compact support, that is supp $\varphi(x) \subseteq [-a, a]$, $0 < a < +\infty$ and define:

$$\varphi_{kj}(x) := 2^{\frac{k}{2}}\varphi(2^k x - j) \quad \text{for} \quad k, j \in Z.$$

If H is continuous, then define

$$B_k(H)(x) := \sum_j H(2^{-k}j)\varphi_{kj}(x) \text{ for } k \in Z. \tag{6}$$

Since $\varphi(x)$ is compatly supported, for any fixed $x \in \mathbb{R}$ the summation in (6) only involves a finite number of terms (see the appendix), so $B_k(H)(x)$ is well-defined on $\mathbb{R}$, that is

$$B_k(H)(x) = \sum_{j=j_0}^{j_1} H(2^{-k}j)\varphi_{kj}(x).$$

Theorem 6 of Anastasiou and Yu (1992) states that if $\varphi(x)$ satisfies the conditions C.1 to C.4 given below and if $H(x) \in C(\mathbb{R})$ is a non-decreasing function, then the linear wavelet operator $B_k(H)(x)$ given by equation (6), is also non-decreasing on $\mathbb{R}$ and satisfy

$$|B_k(H)(x) - H(x)| \leq C\omega_2(H, 2^{-k+1}a), \quad \text{for } x \in \mathbb{R}, k \in \mathbb{Z},$$

where C is an absolute constant and

$$\omega_2(H,\delta) = \sup_{h<\delta}\sup_x |H(x+2h) - 2H(x+h) + H(x)|.$$

The conditions on $\varphi(x)$ are:

C.1: $\sum_{j\in\mathbb{Z}} \varphi(x-j) = 1 \ \forall x \in \mathbb{R}$,
C.2: there exist a number b such that $\varphi(x)$ is non-decreasing if $x \leq b$ and is non-increasing if $x \geq b$,
C.3: $\int_{-\infty}^{\infty} \varphi(x)dx = 1$,
C.4: $\sum_{j\in\mathbb{Z}} j\varphi(x-j) = x \ \forall x \in \mathbb{R}$.

Note that if $a = 1$, the conditions C.1 and C.4 can be written as

C.1:

$$\begin{cases} \varphi(x) + \varphi(x+1) = 1 \text{ if } -1 \leq x \leq 0, \\ \varphi(x) + \varphi(x-1) = 1 \text{ if } 0 < x \leq 1, \end{cases}$$

C.4:

$$\begin{cases} x + \varphi(x+1) = 0 \text{ if } -1 \leq x \leq 0, \\ x - \varphi(x-1) = 0 \text{ if } 0 < x \leq 1. \end{cases}$$

It can be easily shown that the only function satisfying these two conditions is

$$\varphi(x) = \begin{cases} 1+x \text{ if } -1 \leq x \leq 0, \\ 1-x \text{ if } 0 < x \leq 1. \end{cases} \tag{7}$$

Since $H(\cdot)$ is unknown, we cannot compute $H(2^{-k}j)$ directly. Consequently, let $\{\theta_j\}_{j=j_0}^{j_1}$ be a sequence of real numbers. Hence, renumbering the θ_j, $B_k(H)$ can be written as

$$\begin{aligned} B_k(H)(x) &= \sum_{j=j_0}^{j_1} \theta_j \varphi_{kj}(x) \\ &= \sum_{j=j_0}^{j_1} \theta_j 2^{k/2} \varphi(2^k x - j) \\ &= 2^{k/2} \sum_{l=0}^{j_1-j_0} \theta_l \varphi(2^k x + j_0 - l). \end{aligned} \tag{8}$$

However, since $H(x)$ is a non-decreasing function, the θ_j's should also be non-decreasing. Hence, given θ_0, θ_l $(l = 1, 2 \ldots, j_1 - j_0)$ can be written as

$$
\begin{aligned}
\theta_1 &= \theta_0 + \zeta_1, \\
&\vdots \\
\theta_l &= \theta_0 + \zeta_1 + \ldots, \zeta_l \\
&= \sum_{q=0}^{l} \zeta_q,
\end{aligned}
$$

where $\zeta_0 = \theta_0$ and $\zeta_l \geq 0$ for $l = 1, 2, \ldots, j_1 - j_0$. Consequently, equation (8) can be written as

$$
\begin{aligned}
B_k(H)(x) &= 2^{k/2} \sum_{l=0}^{j_1-j_0} \left[\sum_{q=0}^{l} \zeta_q \right] \varphi(2^k x + j_0 - l) \\
&= 2^{k/2} \sum_{q=0}^{j_1-j_0} \zeta_q \left[\sum_{l=q}^{j_1-j_0} \varphi(2^k x + j_0 - l) \right] \\
&= 2^{k/2} \sum_{q=0}^{j_1-j_0} \zeta_q \Phi_q(2^k x + j_0),
\end{aligned}
\tag{9}
$$

where

$$
\Phi_q(2^k x + j_0) = \sum_{l=q}^{j_1-j_0} \varphi(2^k x + j_0 - l).
$$

Using standard techniques, we can obtain a linear model as in equation (2) with

$$
\begin{aligned}
Y &= (H(x_1), H(x_2), \ldots, H(x_n))', \\
(X)_{i,j} &= 2^{k/2} \Phi_{j-1}(2^k x_i) \text{ for } i = 1, 2, \ldots, n \text{ and } j = 1, 2, \ldots, j_1 - j_0 + 1, \\
\zeta &= (\zeta_0, \zeta_1, \ldots, \zeta_{j_1-j_0})'.
\end{aligned}
$$

However, the prior on ζ is different from equation (3). To account for the non-negativity of ζ_q for $j = 1, 2, \ldots, j_1 - j_0$, the prior is then

$$
\begin{aligned}
\zeta_0 &\sim N(\eta_0, \sigma^2/n_0), \\
\zeta_q &\sim N(\eta_q, \sigma^2/n_0) \mathbb{I}_{[0,\infty)}(\zeta_q) \text{ for } q = 1, 2, \ldots, j_1 - j_0, \\
\sigma^2 &\sim I\Gamma(\alpha/2, \gamma/2),
\end{aligned}
$$

where $\mathbb{I}_{[0,\infty)}(\zeta_q)$ represents the indicator function of the set $[0, \infty)$. The posterior density of ζ is similar to equation (5) but we have to account for the non-negativity of $\zeta_1, \zeta_2, \ldots, \zeta_{j_1-j_0}$. The Bayes estimator of ζ is then given by

$$\widehat{\zeta} = \int_0^\infty \int_0^\infty \dots \int_0^\infty \int_{-\infty}^\infty \zeta \times [\zeta \mid \sigma^2, Y \sim N_{j_1-j_0+1}(\zeta_*, \sigma^2 C_*) \prod_{q=1}^{j_1-j_0} \mathbb{I}_{[0,\infty)}(\zeta_q)] \\ \times \left[\sigma^2 \mid Y \sim I\Gamma\left(\frac{n+\alpha}{2}, \frac{\gamma_*}{2}\right)\right] d\zeta_0 d\zeta_1 \dots d\zeta_{j_1-j_0} d\sigma^2. \quad (10)$$

3 Estimation of the sub-density f^*

To obtain the estimator of $f^*(t)$, we start by estimating the cumulative distribution function (cdf) F^* using equation (10) based only on the uncensored observations (values of i such that $\delta_i = 1$). The vector Y, described at the beginning of Section 1.2, will then be based on the empirical cdf of the uncensored observations; that is,

$$Y = \frac{1}{n_o+1}(1, 2, \dots, n_o - 1, n_o)^t,$$

where n_o represent the number of uncensored observations. Referring to equation (9), it is clear that the estimator of Y is also an estimator of the sub-distribution which we denote as $\widehat{F}^*(x)$ given by

$$\widehat{F}^*(x) = 2^{k/2} \sum_{q=0}^{j_1-j_0} \widehat{\zeta}_q \Phi_q(2^k x + j_0),$$

where $\widehat{\zeta}$ is defined by equation (10). To obtain the estimator of $f^*(t)$, we proceed as follows :

$$\begin{aligned} \widehat{f^*}(x) &= \frac{\partial}{\partial x}\widehat{F}^*(x) \\ &= 2^{3k/2} \sum_{q=0}^{j_1-j_0} \widehat{\zeta}_q \Phi'_q(2^k x + j_0). \end{aligned}$$

To estimate $L(x) = P(X_i \leq x)$, we proceed as for $F^*(x)$ in order to obtain $\widehat{L}(x)$, but this time all the observations (censored and uncensored) are used. The estimator of the hazard function is then given by

$$\widehat{\lambda}(x) = \frac{\widehat{f^*}(x)}{1 - \widehat{L}(x)}.$$

Remark 2. *This method is easy to adapt to hazard estimation in more general models such as the Cox proportional risk model:*

$$\lambda(t) = \lambda_0(t) \exp\{-\underset{\sim}{\beta} \cdot \underset{\sim}{x}\},$$

where $\underset{\sim}{x}$ is a vector of covariates, $\underset{\sim}{\beta}$ a vector of parameters and $\lambda_0(t)$ is the baseline hazard function and the usual assumption of non-informative censoring (cf. *Fleming and Harrington, 1991) is made. We proceed with the usual semiparametric Bayesian approach to find an estimator $\widehat{\underset{\sim}{\beta}}$ of $\underset{\sim}{\beta}$. The vector Y, described in Section 2, will now be based on the empirical cumulative hazard function given by*

$$\widehat{\Lambda}(t) = \sum_{t_i \leq t} d_i \Big/ \sum_{k \in R(t_i)} \exp\{-\sum_{m=1}^{p} \widehat{\beta}_m x_{km}\},$$

where t_i represent the observed "event" times and d_i, the observed number of events occurring at time t_i and $R(t_i)$, the risk set associated with t_i (cf. *Klein and Moeschberger, 1997).*

Again, referring to equation (9)*, the estimator of Y is also an estimator $\widehat{\Lambda}$ of the cumulative hazard function Λ and the survival function is estimated by* $\exp\{-\widehat{\Lambda}(t)\}$ *and an estimator of the density function is found by differentiation. Proceeding in a manner analogous to the above, we obtain an estimator of the baseline hazard function $\widehat{\lambda}_0(t)$ and consequently of $\lambda(t)$.*

Remark 3. *We conjecture that in an analogous way, the methodology presented here can be extended to include more general situations such as some of those mentioned in Sinha and Dey (1997). In particular, we feel that our methodology can be extended to such models as a Cox model with informataive censoring or one with a cure fraction as in Ibrahim et al. (2001).*

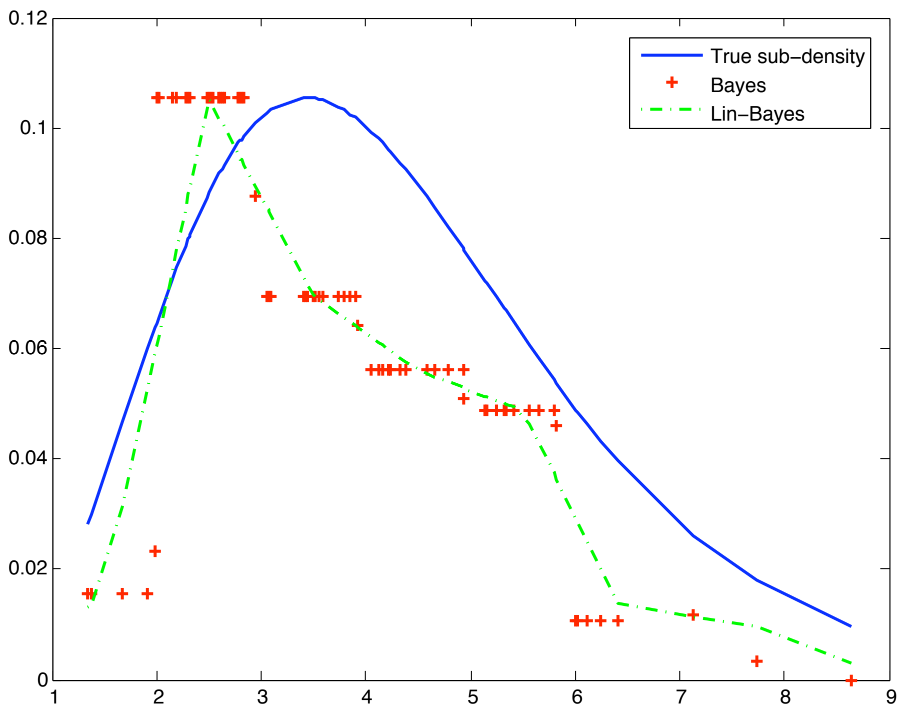
4 Simulations

We consider here the first simulation study proposed by Antoniadis et al. (1999). Samples of size n, $T_i, 1 \leq i \leq n$, from the gamma distribution with shape parameter 5 and scale parameter 1, denoted by f_1 and an independent sample $C_i, 1 \leq i \leq n$, from the exponential distribution with mean 6 (the mean was chosen to yield about 50% censoring) were generated. The performance measure used to compare the different estimators is the average mean-squared error obtained by averaging the mean-squared errors given by :

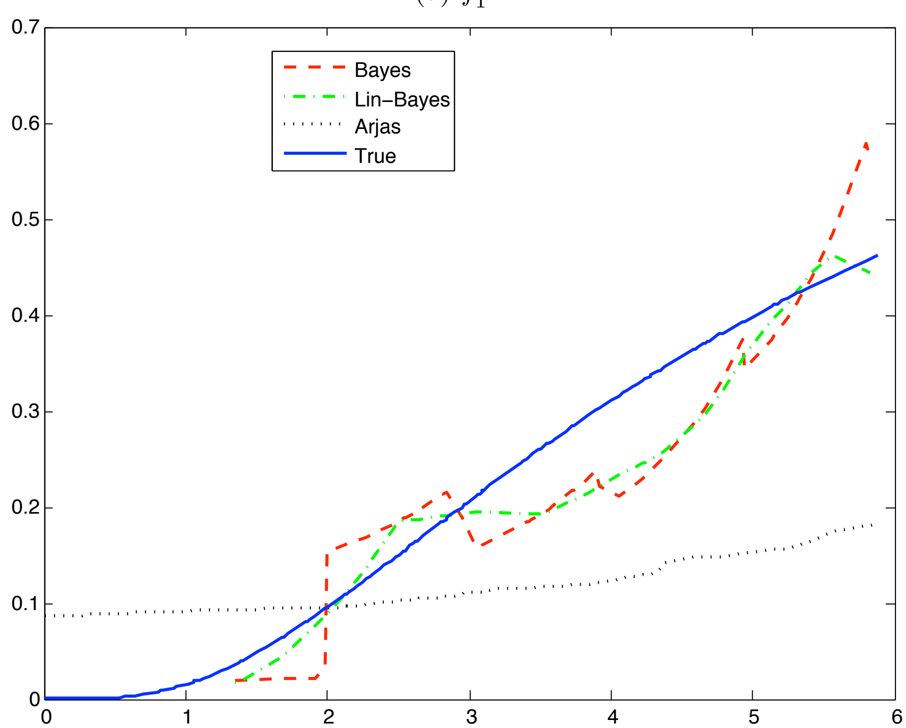
$$\text{ASE}(f^*) = n_*^{-1} \sum_{i=1}^{n_*} [\widehat{f}^*(x_i) - f^*(x_i)]^2,$$
$$\text{ASE}(\lambda) = n_*^{-1} \sum_{i=1}^{n_*} [\widehat{\lambda}(x_i) - \lambda(x_i)]^2,$$

where n_* represents the number of observations with $x_i \leq 6$ Two values of n are considered, that is, $n = 200$ and $n = 500$.

The simulation results are given in Table 1. The true function along with the proposed estimator and the one given in Antoniadis et al. (1999) are given in Figure 1. Because



(a) f_1^*



(b) λ_1

Figure 1: Estimate of the sub-density $\widehat{f}_1^*$ and the hazard λ_1

Table 1: Average mean-squared errors ($\times 10^{-5}$ for the sub-density and $\times 10^{-3}$ for the hazard function) in the first simulation set-up of Antoniadis et al. (1999) based on 200 repetitions.

	f_1^*		λ_1	
	$n = 200$	$n = 500$	$n = 200$	$n = 500$
Antoniadis	[14.6; 20.5]	[5.2; 13.6]	[2.5; 5.8]	[1.6; 5.9]
Bayes	50.1	38.6	11.6	7.0
Linear Bayes	33.4	21.9	3.9	2.0

$\varphi'(x)$ is a step function (see equation (7)), the estimator of the sub-density is also a step function. Hence, we did a linear interpolation based on the center of each interval to smooth $\widehat{f}^*$. In Table 1, the results for this interpolation is denoted by linear Bayes, that is, "Lin-Bayes" in the different figures.

From the table, it can be seen that our proposed estimators are not as efficient as the one proposed in Antoniadis et al. (1999) for the sub-density. However, the linear interpolation performs as well as the Antoniadis estimator for the hazard function.

5 Examples

In this section, two real data sets are considered and the sub-densities along with the corresponding hazard functions are estimated using the monotone wavelet estimator described here as well as the estimators proposed by Antoniadis et al. (1999) and Arjas and Gasberra (1994).

We have chosen to illustrate our method on a data set consisting of a follow-up study of acute leukaemia patients after allogenic bone marrow transplantion. The survival times are given in months. The data set consist of 162 patients (including 63 deaths) For a more complete description of this data, see Brochstein et al. (1987).

Various subsets of this data have been previously used by Mueller and Wang (1990) and Antoniadis et al. (2000) to illustrate different change points methods. Here as a preliminary step in a more complete data analysis to be pursued elsewhere we have chosen to use the complete data set with death due to any cause as the end point and leukaemic relapse or end of study as the censoring mechanism.

Figure 2 illustrates our estimate of the sub-density along with the estimator proposed in Antionadis et al. (1999). (This estimator has been computed using Rice Wavelet Toolbox, version 2.4. with 64 bins and hard thresholding. The resolution was set at $k = 6$. The threshold level was chosen to yield the smoother graph of f^* (threshold level at 0.0032) and λ (0.023).) Figure 3 gives our estimate of the actual hazard rate for this example, the one using wavelets proposed by Antioniadis et al. (1999) and the Bayesian one of Arjas and Gasberra (1994).

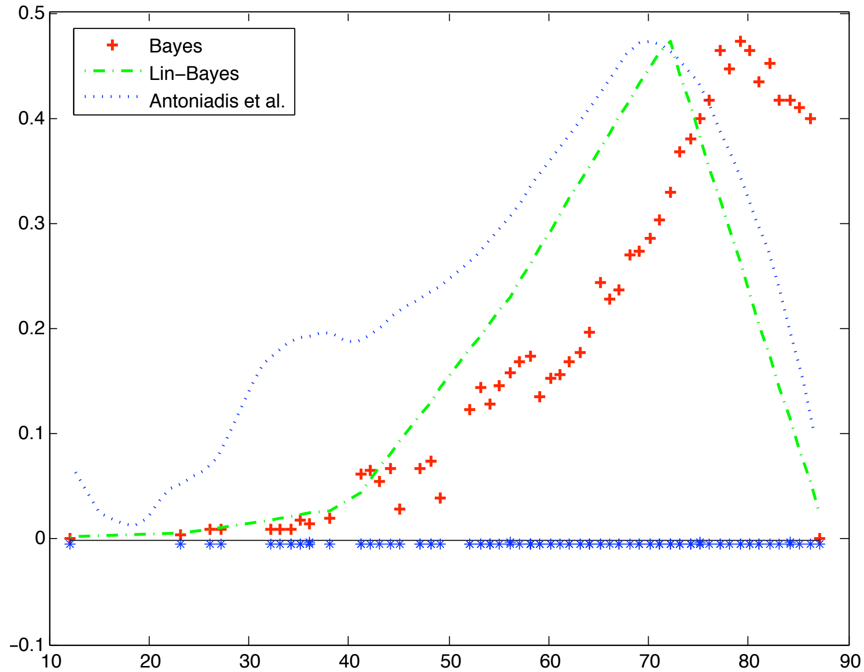


Figure 2: Estimate of the sub-density $\widehat{f}^*$ for the bone marrow transplantation example

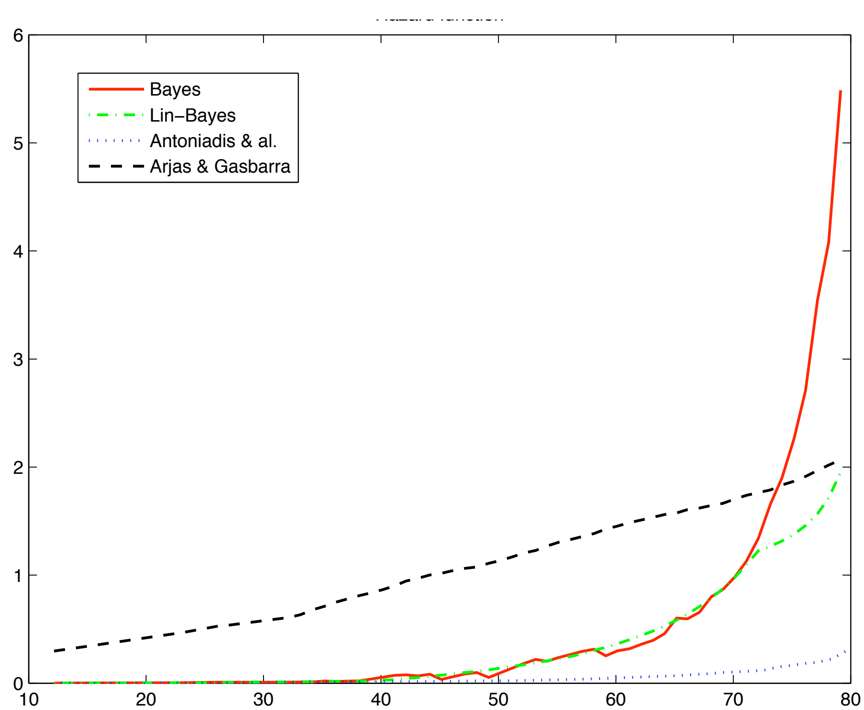


Figure 3: Estimate of the hazard rate function $\widehat{\lambda}$ for the bone marrow transplantation example

From Figure 2 it can be seen that the linearized Bayesian estimator is similar to the one obtained using the Antoniadis et al. (1999) approach, althougth theirs gives more weight to smaller survival times. However, there is much less data manipulation required in order to obtain our estimator. From Figure 3, all estimators of the hazard function except the one from Arjas and Gasberra (1994), are similar for $x \leq 50$. Starting from this point, the direct Bayes estimator is similar to that of Antoniadis et al. (1999) while the linearized version yields a smaller hazard.

The second data set that we consider here is the February 1980 version of the Stanford heart transplant data, published in Cox and Oakes (1984). This data set has previously been analysed by many authors including Loader (1991) and Antoniadis et al. (2000) who considered a change point model for it. Here Figure 4 shows the graphs of the sub-density functions given by our methods (Bayes and linear Bayes) and that of Antoniadis et al. (1999). It should be noted that the sub-density function of Antoniadis et al. (1999) is negative over a small interval and then increases dramatically after that. The linear Bayes graph seems a more reasonable representation of a sub-density. Figure 5 compares the four different hazard function estimators: those of Arjas and Gasberra (1994), Antoniadis et al. (1999) and our Bayes and linear Bayes estimators. Up until $t = 2400$ the four estimators are similar with the Bayes and the linear Bayes representing a compromise between Antoniadis et al. (1999) and Arjas and Gasberra (1994). The estimator of Antoniadis et al. (1999) does have the disadvantage of being negative for large survival times while ours have a sharp peak between 2500 and 3000. The estimator of Arjas and Gasberra (1994) is the most stable.

6 Concluding remarks

Our objective here, as well as to review some of the relevant literature on frequentist and Bayesian nonparametric hazard estimation, was to use Bayesian functional estimation techniques combined with the monotone wavelet approximation methods of Anastassiou and Yu (1992) to estimate the hazard rate with randomly right censored data by a relatively easy method to implement. This has been accomplished.

Although our model is not as effective in the simulation study as the frequentist method of Antoniadis et al. (1999) for the sub-density estimation, the performance of the linear Bayes estimator for the hazard function is comparable and ours is much easier to implement and extremely flexible. Because of the monotonicity of the wavelet approximation, our estimators of the sub-density are theoretically always positive. This is not the case for the estimator of Antoniadis et al. (1999). In fact, in one of the real data examples (see Figure 4) it is negative.

We have also chosen the Bayesian nonparametric method of Arjas and Gasbarra (1994) for purposes of comparison. Arjas and Gasbarra (1994) have an excellent Bayesian nonparametric method which performs very well on the examples here. It is more stable than ours or that of Antoniadis et al. (1999). However, it does not provide an estimate of

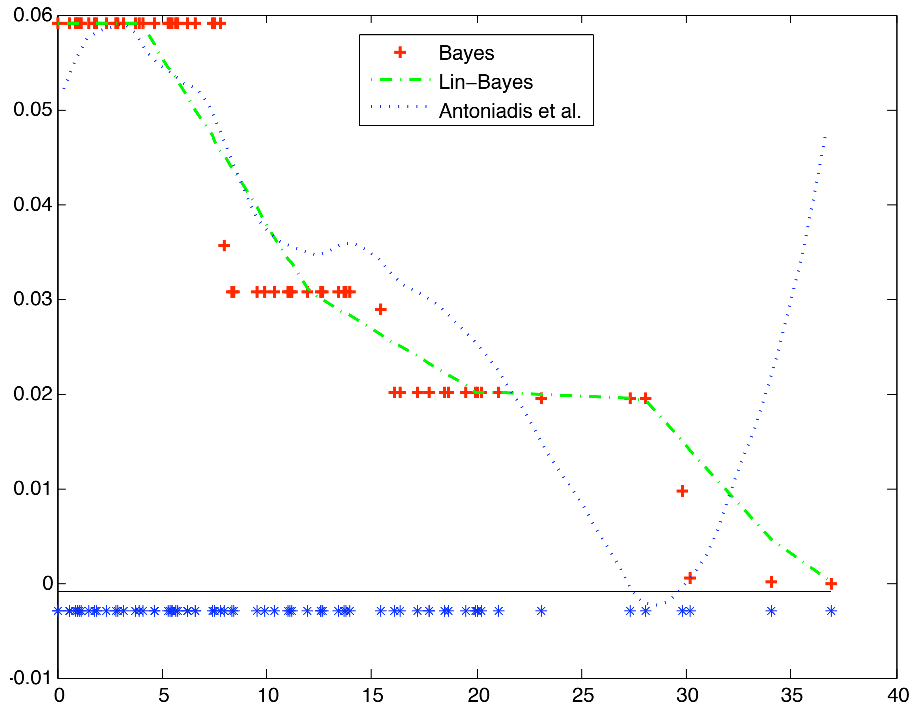


Figure 4: Estimate of the sub-density $\widehat{f}^*$ for the Standford heart transplant example

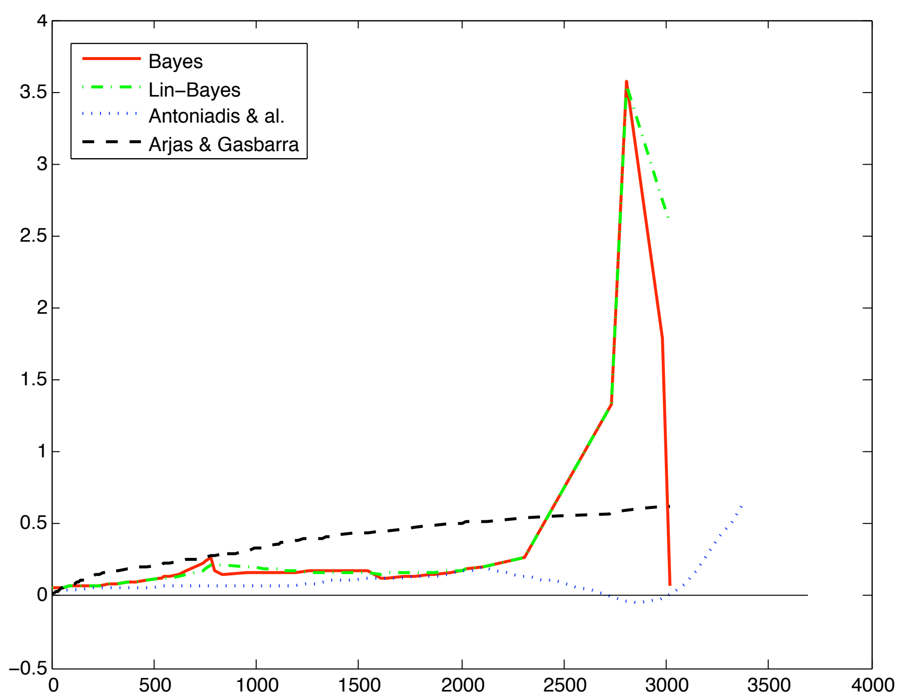


Figure 5: Estimate of the hazard rate function $\widehat{\lambda}$ for the Standford heart transplant example

the sub-density, which we consider an interesting function in its own right. We, therefore conclude that ours, that is, the linear Bayes one, is a good compromise method between the fully frequentist wavelet one and that of Arjas and Gasberra (1994). It is our hope, however, to somehow combine ours with the latter and have a fully Bayesian nonparametric method which combines the good qualities of both.

Appendix : Choice of the resolution level

In this appendix, we will discuss the choice of k and the bounds j_0 and j_1. Suppose that the observed times are $0 < x_1 \leq x_2 \leq ... \leq x_n < T$. Since the support of φ is $[-a; a]$, then, for a fixed k,

$$\begin{aligned}\varphi_{k,j}(t) = 0 &\Leftrightarrow 2^{k/2}\varphi(2^k x - j) = 0\\ &\Leftrightarrow 2^k x - j < -a \text{ or } 2^k x - j > a\\ &\Leftrightarrow 2^k x + a < j \text{ or } j < 2^k x - a.\end{aligned}$$

Since $x \in [0, T]$, then

$$\begin{aligned}\varphi_{k,j}(t) = 0 &\Rightarrow j \notin [2^k x - a, 2^k x + a]\\ &\Rightarrow j \notin [-a, 2^k T + a].\end{aligned}$$

Hence $j_0 = -a$ and $j_1 = 2^k T + a$.

To choose k, we proceed as follows. For each value of j, we want a series of m observations such that $\varphi_{k,j}(x_i) > 0$, $\varphi_{k,j}(x_{i+1}) > 0$, $\ldots$, $\varphi_{k,j}(x_{i+m-1}) > 0$. Since the support of φ is $[-a; a]$, k should be such that

$$\frac{j-a}{2^k} \leq x_i \leq x_{i+1} \leq x_{i+m-1} \leq \frac{j+a}{2^k}.$$

Let $\Delta_m = max_i(x_{i+m-1} - x_i)$. This condition is satisfied if

$$\begin{aligned}\Delta_m &\leq \frac{j+a}{2^k} - \frac{j-a}{2^k}\\ &\Rightarrow \Delta_m \leq \frac{a}{2^{k-1}}\\ &\Rightarrow 2^{k-1} \leq \frac{a}{\Delta_m}\\ &\Rightarrow k \leq 1 + \frac{\log{(a/\Delta_m)}}{\log(2)}\\ &\Rightarrow k \leq 1 + \log_2\left(\frac{a}{\Delta_m}\right).\end{aligned}$$

References

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

[2] Anastassiou, G.A. and Yu, X.M. (1992), Monotone and probabilistic wavelet approximation. *Stochastic Analysis and Applications*, **10**, pp. 251–264.

[3] Angers, J.-F. and MacGibbon, B. (2004), Bayesian fuctional estimation of the hazard rate for randomly right censored data using Fourier series methods. In *Statistical Modeling and Analysis for Complex Data Problems*, (Eds P. Duchesne et B. Rémillard) Klewer, pp. 53–69.

[4] Antoniadis, A., Gijbels, I. and MacGibbon, B. (2000). Nonparametric estimation for the location of a change-point in an otherwise smooth hazard function under random censoring. *Scandinavian Journal of Statistics*, **27**, pp. 501–519.

[5] Antoniadis, A., Grégoire, G. and Nason, G. (1999), Density and hazard rate estimation for right-censored data by using wavelet methods. *Journal of the Royal Statistical Society* Series B, **61**, pp. 63–84.

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

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

[8] Berger, J.O. and Sun, D. (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), Wiley, New York, pp. 101–113.

[9] Bloxom, B. (1985), A constrained spline estimator of a hazard function, *Psychometrika*, **50**, pp. 301–321.

[10] Brochstein, J.A., Kernan, N.A., Groshen, S., Cirrincione, C., Shank, B., Emanual, D., Laver, J. and O'Reilly, R.J. (1987), Allogenic bone marrow transplantation after hyperfractionated total body irradiation and cyclophosphamide in children with acute leukaemia. *New England Journal of Medicine*, **317**, pp. 1618–1624.

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

[12] Cox, D.R. and Oakes, D. (1984), *Analysis of Survival Data.* Chapman and Hall, London.

[13] Digital Signal Processing Group, *Rice Wavelet Toolbox*, version 2.4, Rice University. URL http://www.dsp.rice.edu/software/RWTl.

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

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

[16] Efron, B. (1988), Logistic regression, survival analysis, and the Kaplan-Meier curve. *Journal of the American Stastistical Association*, **83**, pp. 414–425.

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

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

[19] Fleming, T.R. and Harrington, D.P. (1991), *Counting processes and survival analysis.* Wiley, New York.

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

[21] Hall, P. Huang, L.-S., Gifford, J.A. and Gijbels, I. (2001), Nonparametric estimation of hazard rate under the constraint of monotonicity. *Journal of Computational and Graphical Statistics*, **10**, pp. 592–614.

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

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

[24] Ho, M.-W. and Lo, A.Y. (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), pp. 301–314, World Scientific.

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

[26] Ibrahim, J.G., Chen, M-H, and Sinha, D. (2001), Bayesian semiparametric models for survival data with a cure fraction. *Biometrics*, **57**, pp. 383–388.

[27] Izenman, A.J. and Tran, L.T. (1990), Kernel estimation of the survival function and hazard rate under weak dependence. *Journal of Statistical Planning and Inference*, **24**, pp. 233–247.

[28] James, L.F. (2003), Bayesian calculus for Gamma processes with applictions to semi-parametric models. *Sankhya*, **65**, pp. 159–206.

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

[30] Jarjoura, D. (1988), Smoothing hazard rates with cubic splines. *Communications in Statistics – Simulation and Computation*, **17**, pp. 377–392.

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

[32] Klein, J.P. and Moeschberger, M.L. (1997), *Survival analysis: techiniques for censored and truncated data*, Springer-Verlag, New York.

[33] Klotz, J. and Yu, R.-Y. (1986), Small sample relative performance of the spline smooth survival estimator, *Communications in Statistics, Part B – Simulation and Computation*, **15**, pp. 815–818.

[34] Kooperberg, C. and Stone, C.J. (1992), Logspline density estimation for censored data. *Journal of Computational and Graphical Statistics*, **1**, pp. 301–328.

[35] Kooperberg, C., Stone, C.J. and Truong, Y.K. (1995), The L_2 rate of convergence for hazard regression. *Scandinavian Journal of Statistics*, **22**, pp. 143–157.

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

[37] Li, L. (2002), Hazard rate estimation for censored data by wavelet methods. *Communications in Statistics: Theory and Methods*, **31**, pp. 943–960.

[38] Li, L.Y. (2006), On the minimax optimality of wavelet estimators with censored data. *Journal of Statistical Planning and Inference* (in press).

[39] Liang, H.-Y., Mammitzsch, V. and Steinebach, J. (2005), Nonlinear wavelet density and hazard rate estimation for censored data under dependent observations. *Statitics and Decisions*, **23**, pp. 161–180.

[40] Lindley, D.V. and Smith, A.F.M. (1972), Bayes estimates for the linear model. *Journal of the Royal Statistical Society* Series B, **34**, pp. 1–41.

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

[42] Loader, C. (1991), Inference for a hazard rate change point. *Biometrika*, **74**, pp. 301–209.

[43] Loader, C. (1999), *Local regression and likelihood*, Statistics and Computing, Springer-Verlag, New York.

[44] MacGibbon, B., Lu, J. and Younes, H. (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**, pp. 1309–1333.

[45] Mueller, H.G. and Wang, J.L. (1990), Nonparametric analysis of changes in hazard rates for censored survival data: An alternative to change-point models. *Biometrika*, **77**, pp. 305–314.

[46] O'Sullivan, F. (1988), Fast computation of fully automated log-density and log-hazard estimators. *SIAM Journal on Scientific and Statistical Computing*, **9**, pp. 363–379.

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

[48] Patil, P. (1997), Nonparametric hazard rate estimation by orthogonal wavelet methods. *Journal of Statistical Planning and Inference*, **60**, pp. 53–168.

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

[50] Ramlau-Hansen, H. (1983), Smoothing counting process intensities by means of kernel functions. *Annals of Statistics*, **11**, pp. 453–466.

[51] Robert, C.P. (2001), *The Bayesian Choice*, 2nd Edition, Springer, New York.

[52] Roussas, G.G. (1989), Hazard rate estimation under dependence conditions. *Journal of Statistical Planning and Inference*, **22**, pp. 81–93.

[53] Roussas, G.G. (1990), Asymptotic normality of the kernel estimate under dependence conditions: Application to hazard rate. *Journal of Statistical Planning and Inference*, **25**, pp. 81–104.

[54] Senthilselvan, A. (1987), Penalized likelihood estimation of hazard and intensity functions. *Journal of the Royal Statistical Society* Series B, **49**, pp. 170–174.

[55] Sinha, D. and Dey, D.K. (1997), Semiparametric Bayesian analysis of survival data, *Journal of the American Statistical Association*, **92**, pp. 1195–1212.

[56] Susarla, V., and Van Ryzin, J. (1976), Empirical Bayes estimation of a distribution (survival) function with right censored observations. *Annals of Statistics*, **6**, pp. 740–751.

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

[58] Whittemore, A.S. and Keller, J.B. (1986), Survival estimation using splines. *Biometrics*,**16**, pp. 1–11.