

NONPARAMETRIC BAYES ESTIMATORS FOR HAZARD FUNCTIONS BASED ON RIGHT CENSORED DATA

IAN W. MCKEAGUE AND MOURAD TIGHIOUART

Abstract. In this article, we analyse right censored survival data by modelling their common hazard function nonparametrically. The hazard rate is assumed to be a stochastic process, with sample paths taking the form of step functions. This process jumps at times that form a time-homogeneous Poisson process, and a class of Markov random fields is used to model the values of these sample paths. Features of the posterior distribution, such as the mean hazard rate and survival probabilities, are evaluated using the Metropolis–Hastings–Green algorithm. We illustrate our methodology by simulation examples.

1. Introduction

Much of nonparametric Bayesian inference restricted the model distributions to conjugate families for their analytical tractability. In some medical studies, such models were found to be too restrictive to describe the prior information in either the cumulative distribution function or the cumulative hazard function. Recent developments in Markov chain Monte Carlo techniques have widened the scopes of nonparametric (as well as parametric) Bayesian analysis for high-dimensional parameter space.

In this article, the hazard function is modeled nonparametrically as a step function that jumps at times that form time-homogeneous Poisson process. The levels of the hazard rate (conditional on these jump times), form a Gaussian Markov random field with a nearest neighbor structure.

The idea of nonparametric Bayesian inference became very popular following the seminal paper of Cox [3]. Ferguson [7] and Susarla and Van Ryzin [19] used a Dirichlet process prior for the cumulative distribution function. This has the attractive feature that the posterior distribution is another Dirichlet process. Mixtures of Dirichlet processes were used by Doss [5], and Polya tree distributions by Muliere [14]. Prior modeling of the cumulative hazard function using beta processes was introduced by Hjort [11]. The latter class of priors was shown to be a conjugate family, and Damien et al. [16] gave an algorithm to generate random variates from the resulting posterior process. Dykstra and

Received July 25, 2001.

2000 *Mathematics Subject Classification.* 62F15.

Key words and phrases. Life history data, Metropolis–Hastings–Green algorithm, right censoring.

Laud [6] used a gamma process to model the hazard function, yielding an IHR property (increasing hazard rate) of the model.

The prior processes used in the above procedures imply an independent increment property for the cumulative hazard function, which is often unrealistic. In most practical situations, a smooth or correlated prior processes for the hazard function is more suitable. In a Cox-type model $h(t|z) = h_0(t) \exp(\beta(t)'z)$, Gamerman [8] introduced classes of auto-correlated processes to describe the prior information. He assumed that the baseline hazard is a step function of the form

$$h_0(t) = \sum_{i=1}^k I(\tau_i < t \leq \tau_{i+1}) h_i, \quad (1.1)$$

where $0 = \tau_1 < \tau_2 < \dots < \tau_{k+1}$ is a fixed grid of jump times. The log-baseline hazard levels $\lambda_i = \log(h_i)$ form a first-order autoregressive process, $\lambda_{i+1}|\lambda_i \sim N(\lambda_i, \sigma^2)$, and a similar structure is used for the covariate effect $\beta(t)$.

Statistical inference based on model (1.1) is dependent upon the choice of the grid of jump times. Putting a prior distribution on these jump times will result in a more complex model but this helps achieve a dense support for the prior. Such an analysis is feasible due to recent developments in Markov chain Monte Carlo (MCMC) techniques, see, e.g., Smith and Roberts [18]. In particular, Arjas and Gasbarra [1] modeled $h(t)$ as a step function of the form (1.1) with $k = \infty$, where the jump times $\tau_2 < \tau_3 < \dots$ form a time homogeneous Poisson process, and the levels of the hazard rate $\{h_i, i \geq 1\}$ form a first-order autoregressive process. Estimation of the predictive hazard and survival functions was carried out using a dynamic version of the Gibbs sampler algorithm. An extensive review of Bayesian methods for survival data can be found in Sinha and Dey [17].

In this article, the hazard function is modeled as a stochastic process, with sample paths taking the form of step functions. We assume that

$$h(t) = \sum_{i=1}^{\infty} I(\tau_i < t \leq \tau_{i+1}) h_i, \quad (1.2)$$

where $\{\tau_i, i \geq 2\}$ is a time homogeneous Poisson process, and the local characteristics of the log-hazard levels $\lambda_i = \log(h_i)$ are normal distributions.

It follows that the conditional mean of λ_k given its two neighbors is

$$\nu_k = \mu_k + s_k(\lambda_{k-1} - \mu_{k-1}) + r_k(\lambda_{k+1} - \mu_{k+1}),$$

where the parameters $\mu_k = E(\lambda_k)$ represent the trend in the levels of the hazard function, and s_k, r_k represent the influences on λ_k of its left and right neighbors, respectively. Arjas and Heikkinen [2] introduced a similar model for the (prior) intensity of a non-homogeneous Poisson process, but they did not study it in the present survival analysis context.

An appropriate choice of the trend parameters μ_k and influence parameters s_k and r_k give rise to the models proposed by Arjas and Gasbara [1] and Gamerman [8]. Features of the posterior distribution of $(h_0(t), \beta(t))$ will be calculated using the Metropolis–Hastings–Green (MHG) algorithm, Metropolis et al. [15], Hastings [10], Green [9].

The article is organized as follows. Section 2 describes the model for simple right censored survival times. The Metropolis–Hastings–Green algorithm used for sampling from the posterior distribution of the parameter of interest is described in the Appendix. In Section 3, we describe the inclusion of covariate effects in the model and present some simulations. Section 4 contains some concluding remarks.

2. Survival Function Estimation

2.1. The model

Let T_1, \dots, T_n be nonnegative independent random variables with a common hazard function h . Assume that the data may be subject to right censoring, i.e., we observe $(X_1, \delta_1), \dots, (X_n, \delta_n)$ where $X_j = \min(T_j, U_j)$, U_j being the censoring time for the j -th individual, and $\delta_j = I\{T_j \leq U_j\}$. Our Bayesian approach consists of putting a prior distribution on the unknown hazard function and then extracting features of the posterior distribution of h given the data using the Metropolis–Hastings–Green algorithm.

The structure of the hazard rate is given by:

$$h(t) = \sum_{i=1}^{\infty} I\{\tau_i < t \leq \tau_{i+1}\} h_i,$$

where $I\{\cdot\}$ is the indicator function, $0 = \tau_1 < \tau_2 < \tau_3 < \dots$ is an increasing sequence of jump times, and the h_i 's represent the levels of the hazard function. Let $\tau_{\max} = \max_{1 \leq j \leq n} X_j$.

We specify the prior distribution on $h(t)$ for $0 < t \leq \tau_{\max}$ as follows:

1. The jump times τ_2, τ_3, \dots form a time-homogeneous Poisson process with rate α .
2. Given that there are $m - 1$ jumps in the interval $[0, \tau_{\max}]$, $(\lambda_1, \lambda_2, \dots, \lambda_m)$ is a Gaussian Markov random field, with a nearest neighbor structure, specified by its local characteristics: $\lambda_k | \{\lambda_i, i \neq k\} \sim N(\nu_k, \sigma_k^2)$, where $\lambda_i = \log(h_i)$.

It can be shown that the conditional mean ν_k is given by

$$\nu_k = \mu_k + s_k(\lambda_{k-1} - \mu_{k-1}) + r_k(\lambda_{k+1} - \mu_{k+1}),$$

where the hyperparameters $\mu_k = E(\lambda_k)$ represent the trend in the levels of the hazard function, and s_k, r_k are the influences of the left and right neighbors of λ_k , respectively. The models proposed by Arjas and Gasbara (1994) and Gamerman (1991) essentially arise as special cases: use a constant trend and let $r_i \rightarrow 0$ and $s_i \rightarrow 1$. The joint distribution of $\lambda_1, \dots, \lambda_m$ (given m fixed) is completely determined by its local characteristics

provided they satisfy the following consistency conditions: s_k, r_k are nonnegative with $s_k + r_k < 1$, and $r_k \sigma_{k+1}^2 = s_{k+1} \sigma_k^2$, see, e.g., Besag and Kooperberg (1995). Now we specify s_k, r_k and σ_k^2 , the aim being to force the corresponding hazard function $h(t)$ to be relatively ‘smooth.’ Of the two neighbors of λ_k , the one corresponding to the longer interval should have the greatest influence on its mean ν_k . We propose to use

$$r_k = \frac{(\Delta_k + \Delta_{k+1})c}{\Delta_{k-1} + 2\Delta_k + \Delta_{k+1}}, \quad s_k = \frac{(\Delta_{k-1} + \Delta_k)c}{\Delta_{k-1} + 2\Delta_k + \Delta_{k+1}},$$

$$\sigma_k^2 = \frac{2\sigma^2}{\Delta_{k-1} + 2\Delta_k + \Delta_{k+1}},$$

where $\Delta_k = \tau_{k+1} - \tau_k$ is the gap between the k -th and $(k+1)$ -st jump times, $2 \leq k \leq m-1$. The influence parameters r_1, s_1, r_m, s_m at the boundaries are defined as above, but identifying the endpoints λ_1 and λ_m as neighbors. It is readily checked that the above consistency conditions are satisfied in this case. Other choices of r_k, s_k, σ_k^2 are also possible.

Arjas and Heikkinen [2] used the Voronoi tessellation of $\{\tau_i\}$ to specify the jump times in their model for the intensity of a non-homogeneous Poisson process. This is technically appealing because of the one-to-one correspondence it induces between τ_1, \dots, τ_m and the log-intensity levels $\lambda_1, \dots, \lambda_m$. However, we prefer using $\{\tau_i\}$ as the the jump times to facilitate comparison with the survival analysis models of Arjas and Gasbara [1] and Gamerman [8].

Given m , the joint distribution of $(\lambda_1, \lambda_2, \dots, \lambda_m)$ is Gaussian with mean vector $\boldsymbol{\mu}_m$ and covariance matrix $(I - C)^{-1}M$, where $\boldsymbol{\mu}_m = (\mu_1, \dots, \mu_m)$, $C = (c_{ij})_{1 \leq i, j \leq m}$, where $c_{ii+1} = r_i, c_{ii-1} = s_i$, $M = \text{diag}(\sigma_1^2, \sigma_2^2, \dots, \sigma_m^2)$, and I is the identity matrix.

Our model has hyperparameters $\alpha, (\mu_k), c$ and σ . The parameter α controls the rate of jump times, (μ_k) is the trend in the log-hazard function, c controls the nearest neighbor interaction ($0 \leq c < 1$), and σ^2 represents the precision of the prior information. Choosing a value of c close to 1 amounts to vague prior knowledge of the trend parameters μ_k . For example, if $\mu_k = \mu$ for all k , then $\nu_k = \mu(1 - c) + s_k \lambda_{k-1} + r_k \lambda_{k+1}$ and the influence of μ on ν_k vanishes as $c \rightarrow 1$. For simplicity of presentation, we restrict attention to the case $\mu_k = \mu$, which indicates a constant prior level in the mean of the log-hazard function.

The likelihood is proportional to

$$\prod_{j=1}^n [h(X_j)^{\delta_j} \exp\{-\int_0^{X_j} h(s) ds\}]$$

$$= \prod_{j=1}^n (h(X_j))^{\delta_j} \exp\{-\int_0^{\tau_{\max}} Y(s)h(s) ds\},$$

where $Y(t) = \sum_{j=1}^n I\{X_j \geq t\}$ is the number of individuals at risk at time t , see, e.g., Cox and Oakes [4, Ch. 3]. Here we have assumed that the censoring mechanism is non-informative so the normalizing constant does not depend on $h(t)$. Let N_i be the number

of observed deaths in the interval $(\tau_i, \tau_{i+1}]$, and denote by $(\boldsymbol{\tau}_m, \boldsymbol{\lambda}_m)$ the $2m$ -dimensional vector $(\tau_1, \dots, \tau_m, \lambda_1, \dots, \lambda_m)$. The posterior density of the parameter $(\boldsymbol{\tau}_m, \boldsymbol{\lambda}_m)$ given the data is proportional to the product of the prior and the likelihood:

$$\alpha^m (2\pi)^{-\frac{m}{2}} |A|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} (\boldsymbol{\lambda}_m - \boldsymbol{\mu}_m)' A (\boldsymbol{\lambda}_m - \boldsymbol{\mu}_m) \right\} \\ \times \exp \left\{ \sum_{i=1}^m N_i \lambda_i - \int_0^{\tau_{\max}} Y(s) h(s) ds \right\}$$

where $A = M^{-1}(I - C)$.

We have devised a Metropolis–Hastings–Green sampler to extract features from this posterior distribution, see the appendix.

2.2. Numerical Examples

The numerical results we present here are based on data generated from two different hazard functions. The censoring times were generated from a uniform distribution on the interval $(0, \tau)$, where τ is chosen to achieve a desired censoring probability. We first simulated 100 failure times from the hazard function shown by the solid line in the second plot of Figure 1. Eight observations were censored, and fifteen observations exceeded 1.0. The hyperparameter values used were $\mu = 0.4$, $c = 0.98$, $\sigma = 0.4$, and $\alpha = 6.0$, which corresponds to a prior mean level for the hazard rate of $\exp(\mu) = 1.49$, and prior mean number of jump times of $\alpha \times 1.0 = 6.0$. In practice, one can choose a value of $\exp(\mu)$ within the range of a smoothed version of the hazard function, using for example the Nelson–Aalen estimator of the cumulative hazard function. When $\tau_{\max} = 1$ and the number of failure times n is between 50 and 300, we recommend values of α in the interval $[5, 10]$. For a more complete Bayesian analysis, second stage priors could be placed on α , μ , c , and σ .

The sampler was run for 1,000,000 updates, after a burn-in period of 100,000 updates. Convergence of the sampler was assessed by observing independent runs of the algorithm, taking different starting sample paths, changing the values of the sampler parameters, and plotting the autocorrelation function at four time points.

The sampler parameters used were $\eta = 0.2$, $\delta = 0.2$, and $\epsilon = 1.2$. The resulting rejection rate of the MHG sampler was about 28%. Based on our experience with the sampler, we suggest using values of η , δ , and ϵ in the intervals $(0.1, 0.3)$, $(\sigma/4, 2\sigma/3)$, and $(2\sigma, 4\sigma)$, respectively. This will result in a rejection rate between 25% and 30%.

The MCMC sample used to approximate the Bayes estimates shown in Figure 1 was obtained by using every 1000-th realization of the Markov chain (to save storage). The proposed estimator of the predictive survival function closely follows the Kaplan–Meier curve, except that it is much smoother. The posterior mean hazard function is a smooth curve due to the averaging of a large number of step functions with random jump locations, and follows the pattern of the true hazard function quite closely. The true hazard curve lies completely within the posterior pointwise quantile curves except in the neighborhood of $t = 0.8$.

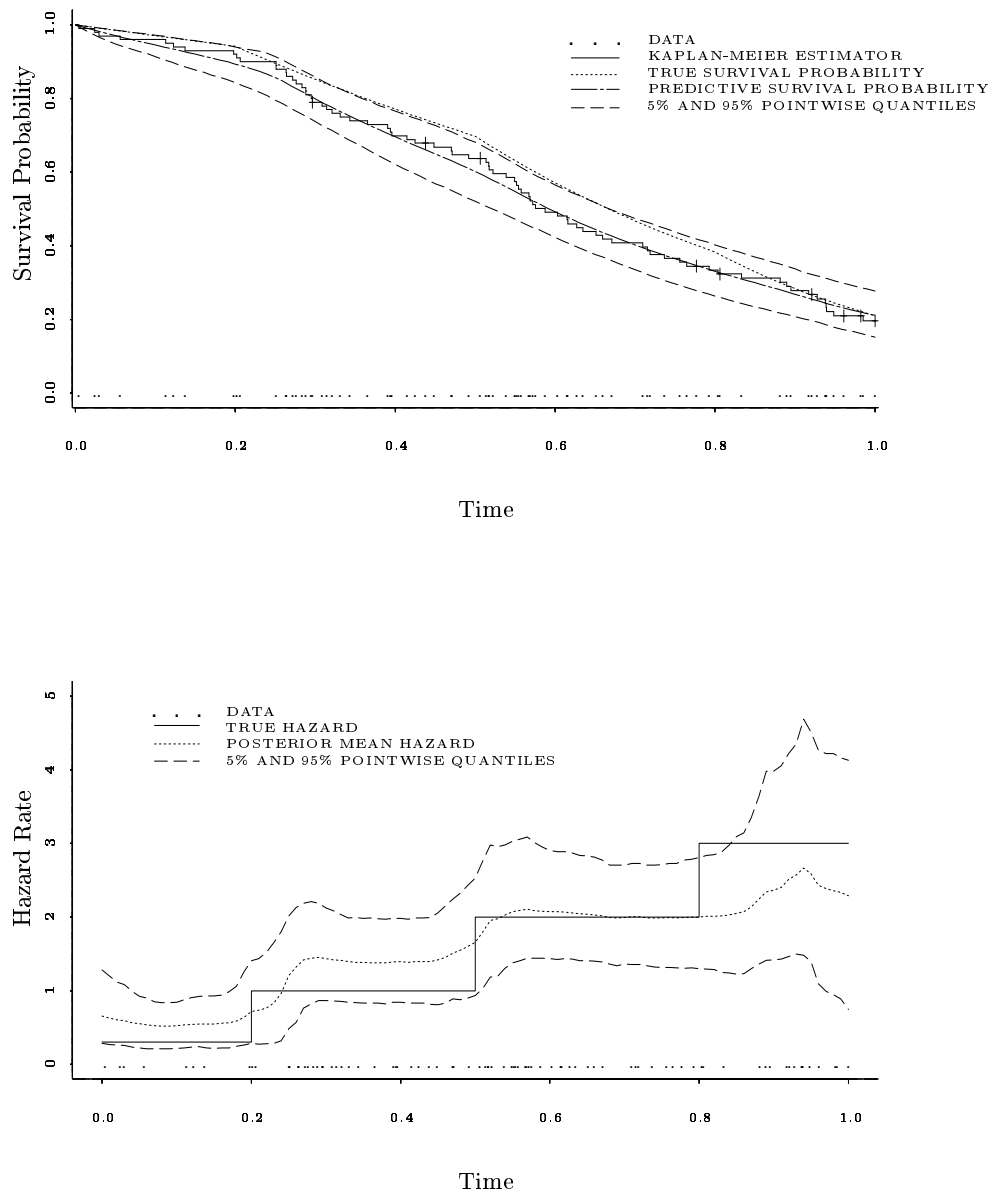


Figure 1. Proposed Bayes estimates based on $n = 100$ failure times.

A sensitivity analysis of the effect of the trend parameter on the Bayes estimates is shown in Figure 2. Three levels of the trend parameter μ were chosen; these correspond to constant log-hazard rates of 0.49 (low), 1.49 (medium), and 2.71 (high). The corresponding estimates are indicated by a dashed line (low trend), dotted line (medium trend), and an irregular dashed line (high trend). The other hyperparameters are as in the first simulation. There is almost no difference in the corresponding posterior mean hazard curves, except in a neighborhood of $t = 0.95$, and the predictive survival curves are indistinguishable.

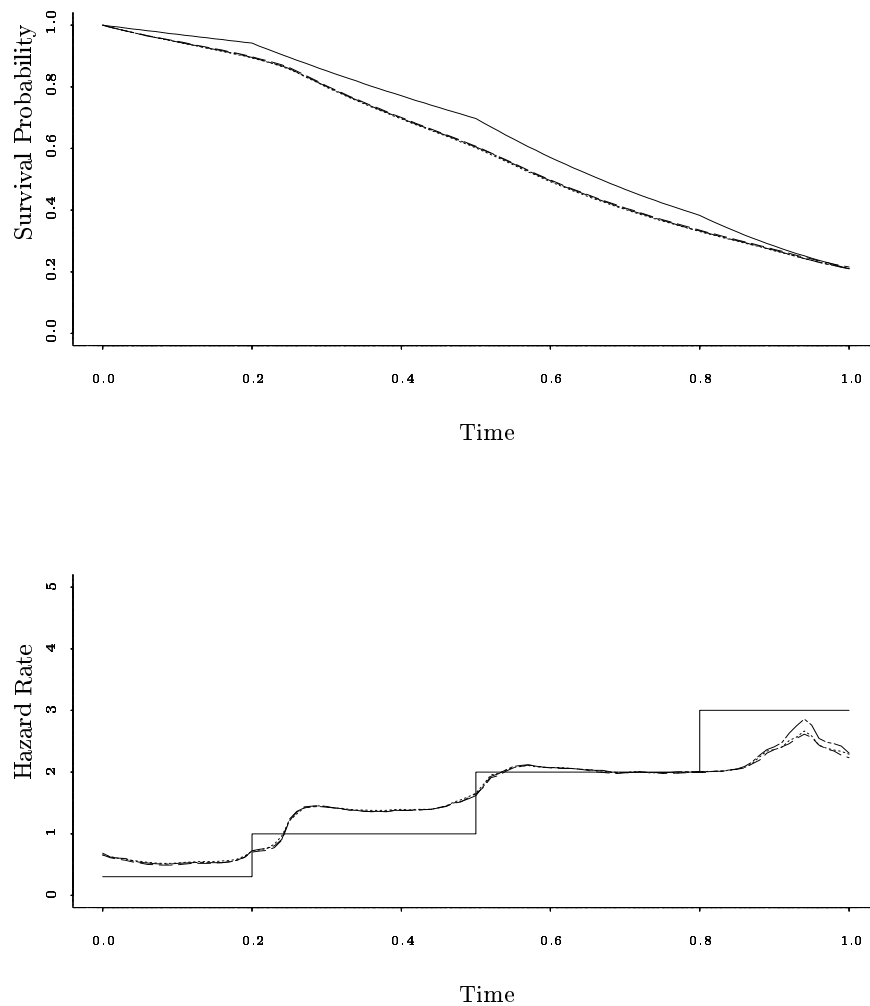


Figure 2. Sensitivity analysis for the trend parameter μ .

As a second illustration, we generated 200 failure times from a rather complex hazard function shown by the solid line at the bottom of Figure 3. From those, 26 observations exceeded $t = 1.0$, and 6 observations were censored in the interval $(0, 1)$.

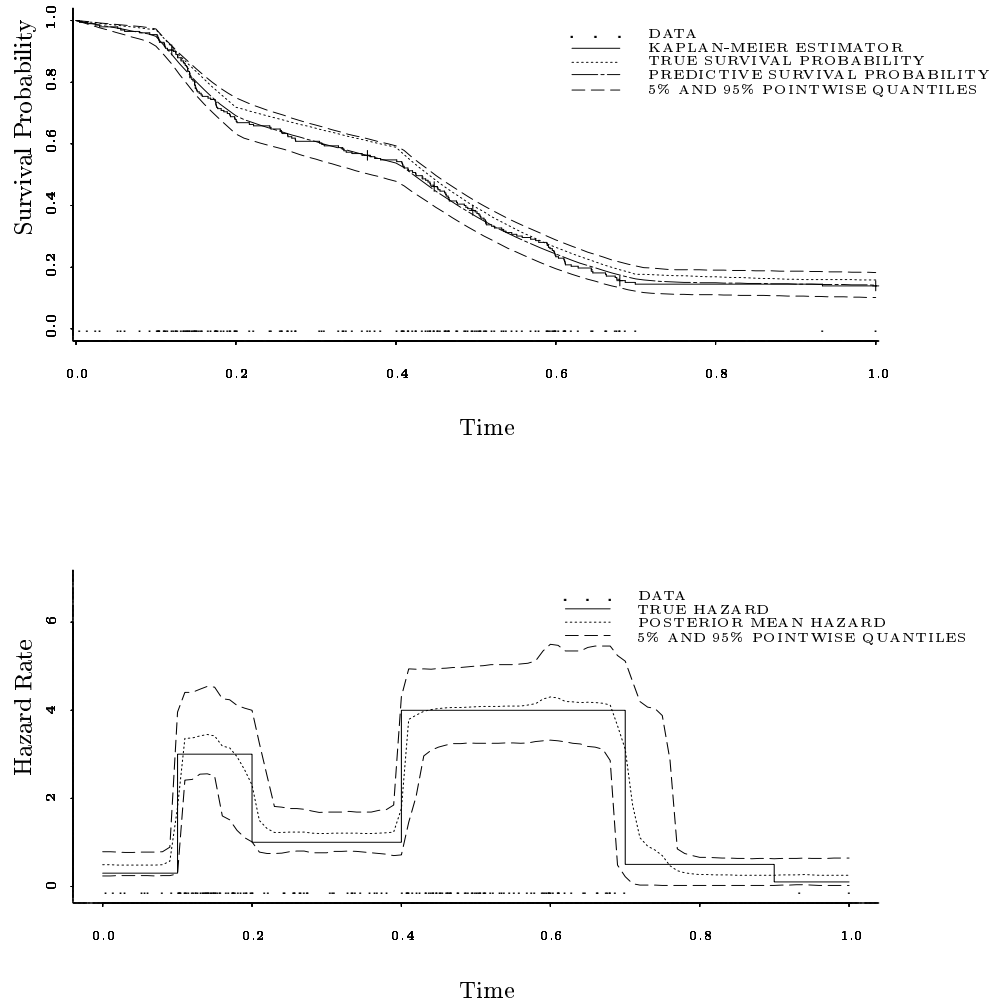


Figure 3. Proposed Bayes estimates based on $n = 200$ failure times.

The hyperparameter values we used are $\mu = 0.7$ which corresponds to a prior mean level of $\exp(\mu) = 2.01$ for the hazard rate, $c = 0.98$, $\sigma = 0.75$, and $\alpha = 6.0$. The predictive survival curve is close to both the Kaplan-Meier curve and the true survival curve. The

posterior mean hazard function is a smooth curve, and captures the key features of the true hazard function remarkably well. The true hazard function lies completely between the pointwise 5% and 95% posterior quantile curves.

3. Adjusting for Covariate Effects

3.1. The model

The data now consist of n independent right-censored survival times and associated p -dimensional covariate vectors z_j , $j = 1, \dots, n$. The prior conditional hazard function now has the structure

$$h(t|z) = \sum_{i=1}^{\infty} I(\tau_i < t \leq \tau_{i+1}) h_i \exp(\beta'_i z),$$

where $\{\beta_i, i \geq 1\} = \{(\beta_{i1}, \dots, \beta_{ip})', i \geq 1\}$ is a p -dimensional process describing the effect of the covariate vector z . The prior distribution of the (τ_i, h_i) is the same as in Section 2.1. A complete prior specification of $h(t|z)$ is then achieved by assuming that, conditional on the first $m - 1$ jump times τ_2, \dots, τ_m in the interval $(0, \tau_{\max})$, and independently of h_1, \dots, h_m , the p covariate effects $(\beta_{1k}, \beta_{2k}, \dots, \beta_{mk}), k = 1, 2, \dots, p$, are independent and for each k , $(\beta_{1k}, \beta_{2k}, \dots, \beta_{mk})$ is a Gaussian Markov random field, specified by its local characteristics. The expressions of the influences r_i, s_i and conditional variances σ_i^2 we used to model the hazard function in Section 2.1 are adopted for each covariate effect. The resulting trend, nearest neighbor interaction, and precision of the prior information hyperparameters are denoted by $(\mu_{ik}), c_k$, and σ_k^2 , respectively.

Simulating sample paths from the resulting posterior distribution is accomplished by adding p transition kernels to the ones defined in the Appendix, corresponding to a change in height of the covariate effects $\beta_1(t), \beta_2(t), \dots, \beta_p(t)$.

3.2. Simulated data example

The numerical results we present here are based on $n = 100$ data generated from model (2) with $p = 2$. First, we generate the covariates z_1, z_2 independently from $U(0, 1)$. Next, we generate a failure time from the conditional hazard function (given z_1, z_2), where the true log-baseline hazard function is shown in the first plot of Figure 4, and the true covariate effects are shown in Figure 5. The censoring time is again generated from $U(0, \tau)$. Seven observations exceeded $t = 1.0$, and 4 observations were censored in the interval $(0, 1)$. The hyperparameters we used were $\alpha = 6$ which corresponds to a prior mean number of jump times of 6, $(\mu, \sigma, c) = (0.15, 0.4, 0.98)$ for the log-baseline hazard function, and $(\mu_1, \sigma_1, c_1) = (\mu_2, \sigma_2, c_2) = (0.8, 0.25, 0.25)$ for the two covariate effects. In practice, one can fit the Cox model, and use the maximum likelihood estimates of the covariate effects for the trend parameters μ_1 and μ_2 . The standard error estimates of the covariate effects can be used for the hyperparameters σ_1, σ_2 .

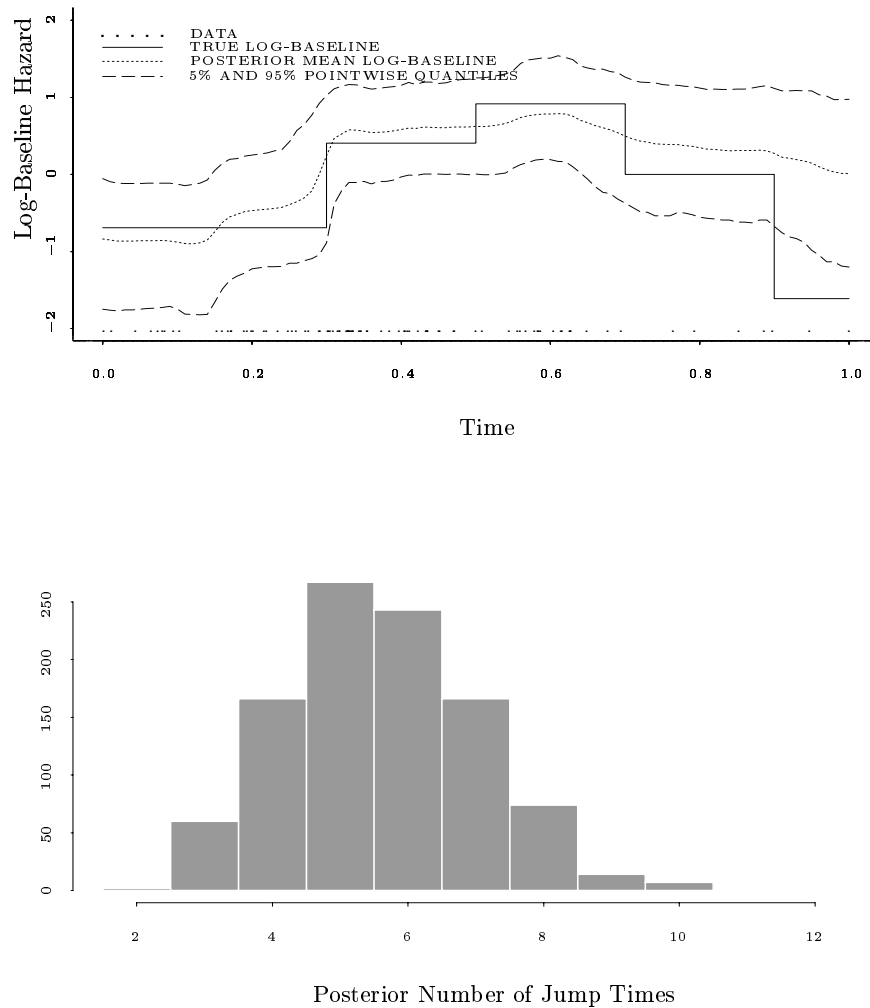


Figure 4. Bayes estimates of the log-baseline hazard function and number of jump times.

The proposed estimate of the posterior mean log-baseline hazard function shown in the first plot of Figure 4 follows the pattern of the true log-baseline hazard function much better for $t \leq 0.6$. The pointwise quantile curves are much wider due the increased number of parameters in the model. The second plot of Figure 4 shows the histogram of the posterior number of jump times, with a posterior mean below the prior mean. The

posterior mean covariate effect shown in the first plot of Figure 5 is centered around the True covariate effect, with a slight tendency to increase for $t \geq 0.6$. The posterior mean covariate effect in the second plot of Figure 5 follows the pattern of the True covariate effect remarkably well for $t \geq 0.3$.

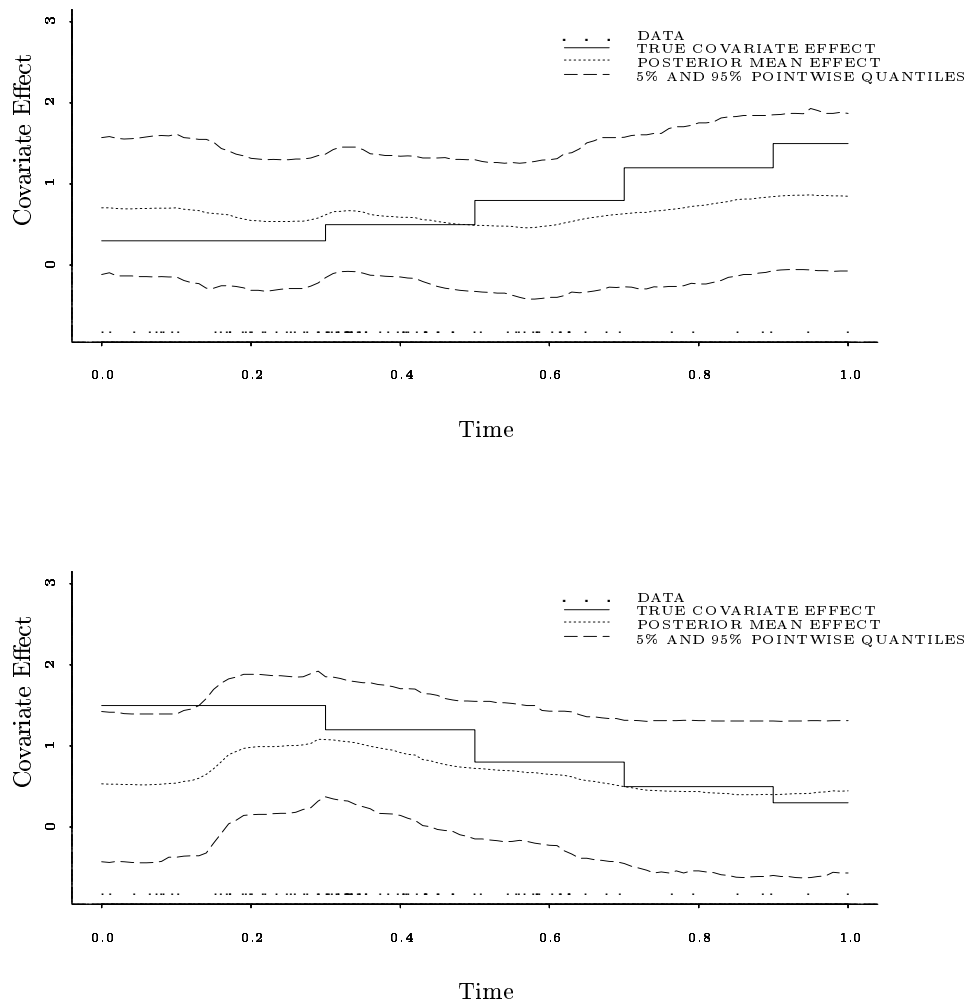


Figure 5. Proposed estimates of the posterior mean covariate effects.

The predictive hazard and survival functions of an individual with covariate vector $z = (z_1, z_2)$ are $h(t|\text{data}, z) = E_\pi\{h(t|z)|\text{data}\}$, and $S(t|\text{data}, z) = E_\pi\{S(t|z)|\text{data}\}$,

respectively. Using the sample paths for the log-baseline hazard function $\lambda_i(t)$, and the two covariate effects $\beta_1^i(t), \beta_2^i(t)$ obtained from our MCMC sample, the predictive hazard function is approximated by

$$\frac{1}{N} \sum_{i=1}^N \exp \{ \lambda_i(t) + \beta_1^i(t) z_1 + \beta_2^i(t) z_2 \},$$

and the predictive survival function by

$$\frac{1}{N} \sum_{i=1}^N \exp \left\{ - \int_0^t \exp(\lambda_i(u) + \beta_1^i(u) z_1 + \beta_2^i(u) z_2) du \right\}.$$

In Figure 6, we compared the predictive hazard and survival curves (dotted lines) for an individual with given covariates $z_1 = 0.5, z_2 = 0.25$ to the true conditional hazard and survival functions (solid lines). The predictive survival function is very close to the true conditional survival function.

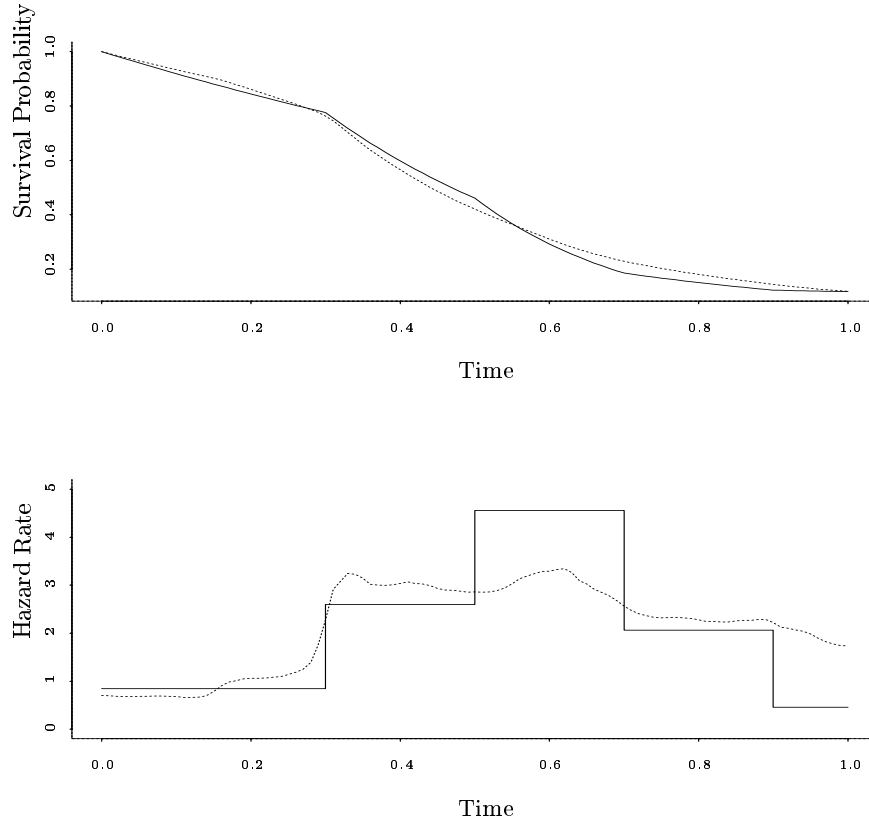


Figure 6. Predictive survival and hazard curves.

McKeague and Tighiouart [12] fitted the model in Section 3 to analyze nasopharynx cancer survival data. They showed the advantage and flexibility of this model over the one used by West [20].

4. Concluding Remarks

We used a simple, yet rich class of prior processes to model the common hazard function for right-censored survival data. This class of prior distributions is more flexible than those of earlier approaches; it extends the model proposed by Arjas and Gasbara [1] by including covariate effects, and is more general than the approach considered by Gamerman [8] in the sense that (1) inference is not conditional on the jump times, and (2) Bayes estimates are not step functions. Moreover, an IHR or DHR (decreasing hazard rate) or a bath-tub shape assumption on the trend of the hazard rate levels can be easily implemented in our model through the trend parameters μ_1, μ_2, \dots . Another advantage of our approach is the computational method we used to extract features of the posterior distribution of the hazard function. It is based on the Metropolis–Hastings–Green algorithm, and is easily extended to models incorporating time-varying covariate effects, whereas it is not clear how the dynamic version of the Gibbs sampler used by Arjas and Gasbara can be used to accommodate for covariate effects.

APPENDIX. Metropolis–Hastings–Green Algorithm

The procedure for calculating features of the posterior distribution of (τ_m, λ_m) (note that here m is random) consists of running a reversible Markov chain on the state space $\mathcal{S} = \cup_{i \geq 1} S_i$, where $S_i = D_i \times \mathbb{R}^i$, and $D_i = \{(x_1, x_2, \dots, x_i) : 0 = x_1 < x_2 < \dots < x_i < \tau_{\max}\}$, using the Metropolis–Hastings–Green algorithm. Suppose that the current state of the chain is $(\tau_m, \lambda_m) \in S_m$, and denote by $(\tau_{m'}, \lambda'_{m'}) \in S_{m'}$ the next state of the chain. When $m' = m$, the update can be done using the classical Metropolis–Hastings algorithm, otherwise some adjustments in the transition kernels are needed when transitions are made between subspaces of different dimension. Green (1995) generalized the classical Metropolis–Hastings algorithm by preserving reversibility of the Markov chain when moving between subspaces of different dimension. To simplify the description of the algorithm, let $x = (\tau_m, \lambda_m)$ and denote by π the posterior distribution of the parameter. We will consider three transition kernels $P_i(x, A)$, $i = 1, 2, 3$, with corresponding state-dependent mixing probabilities $p_i(x)$, satisfying $\sum_{i=1}^3 p_i(x) = 1$. Denote $Q_i(x, A) = p_i(x)P_i(x, A)$. We also need three symmetric measures $\xi_i(dx, dx')$, such that ξ_i dominates $\pi(dx)Q_i(x, dx')$ for each i . Finally, let

$$f_i(x, x') = \frac{\pi(dx)Q_i(x, dx')}{\xi_i(dx, dx')}.$$

The Metropolis–Hastings–Green algorithm updates the current state x of the Markov chain as follows:

1. Select a proposal kernel P_i with probability $p_i(x)$;
2. Generate x' from $P_i(x, \cdot)$;
3. Accept x' with probability $\min\{1, \frac{f_i(x', x)}{f_i(x, x')}\}$, otherwise stay at x .

It can be shown that the resulting MHG transition kernel is reversible with respect to π , see Green (1995). In the context of our problem, transition from (τ_m, λ_m) to a new point (τ'_m, λ'_m) is accomplished by randomly selecting one of three types of moves (H,B,D): a change of height of a randomly selected level of the hazard rate, birth of a new jump time at a randomly selected location in $(0, \tau_{\max})$, and death of a randomly selected jump time, respectively. Denote by p_H^m, p_B^m , and p_D^m the probabilities of selecting the three different types of moves H, B, and D when the current state of the Markov chain is (τ_m, λ_m) . Note that $N(\tau_{\max}) = \#\{i : \tau_i < \tau_{\max}\}$ has a Poisson distribution with parameter $\alpha\tau_{\max}$. The choice of the state-dependent mixing probabilities will be similar to the ones chosen by Green (1995). We take $p_B^m = \gamma \min\{1, \frac{P(N(\tau_{\max})=m+1)}{P(N(\tau_{\max})=m)}\}$, $p_D^m = \gamma \min\{1, \frac{P(N(\tau_{\max})=m-1)}{P(N(\tau_{\max})=m)}\}$, with $p_B^m + p_D^m = \eta$ where η is a sampler parameter, and γ is completely determined by η . Finally, we set $p_D^1 = 0$. When selecting a move of type H, the acceptance probability is the same as in the classical Metropolis–Hastings algorithm:

$$\min\{1, (\text{likelihood ratio}) \times (\text{prior ratio}) \times (\text{proposal ratio})\},$$

whereas if moves of type B or D are selected, the current state (τ_m, λ_m) is mapped onto (τ'_m, λ'_m) by a one-to-one transformation τ . The acceptance probability then takes the form

$$\min\{1, (\text{likelihood ratio}) \times (\text{prior ratio}) \times (\text{proposal ratio}) \times J(\tau)\},$$

where $J(\tau)$ is the Jacobian of the transformation τ . A detailed description of the various transitions and expressions for the acceptance probabilities is given below.

Move of type H. An index k is uniformly selected from $\{1, 2, \dots, m\}$ and V is generated uniformly in the interval $(-\delta, \delta)$, where δ is a sampler parameter. The proposed new level for the log-hazard function is $\lambda'_k = \lambda_k + V$. The proposed new point is (τ'_m, λ'_m) with $\tau'_m = \tau_m$ and $\lambda'_i = \lambda_i$ for $i \neq k$.

The likelihood ratio is

$$\exp \left\{ N_k(\lambda'_k - \lambda_k) + (e^{\lambda_k} - e^{\lambda'_k}) \int_{\tau_k}^{\tau_{k+1}} Y(s) ds \right\}.$$

The prior ratio is

$$\exp\{-\Phi_H(A, \mu, \lambda_m, \lambda'_m)/2\},$$

where

$$\begin{aligned} \Phi_H(A, \mu, \lambda_m, \lambda'_m) &= a_{kk}(\lambda'_k - \lambda_k)(\lambda_k + \lambda'_k - 2\mu) + 2a_{kk-1}(\lambda_{k-1} - \mu)(\lambda'_k - \lambda_k) \\ &\quad + 2a_{kk+1}(\lambda_{k+1} - \mu)(\lambda'_k - \lambda_k). \end{aligned}$$

The proposal ratio is 1 by symmetry of the proposal density.

Move of type B. A new jump time τ^* is drawn uniformly in the interval (τ_1, τ_{\max}) . Suppose that $\tau^* \in (\tau_{k-1}, \tau_k)$. This new jump time induces two new levels λ'_{k-1} and λ'_k for the log-hazard rate constructed using the following transformation $\tau_{\epsilon, V}$: first, draw V uniformly in the interval $(-\epsilon, \epsilon)$, where ϵ is a sampler parameter, then the new log-hazard levels λ'_{k-1} and λ'_k are taken to be convex combinations of the perturbations $\lambda_{k-1} + V$ and $\lambda_{k+1} - V$ with λ_{k-2} and λ_k , respectively. More specifically, we let

$$\begin{aligned}\lambda'_{k-1} &= \frac{\Delta'_{k-1}}{\Delta_{k-1}} \lambda_{k-2} + \frac{\Delta'_k}{\Delta_{k-1}} (\lambda_{k-1} + V), \\ \lambda'_k &= \frac{\Delta'_{k-1}}{\Delta_{k-1}} (\lambda_{k-1} - V) + \frac{\Delta'_k}{\Delta_{k-1}} \lambda_k.\end{aligned}$$

The proposed new point is $(\tau'_{m'}, \lambda'_{m'})$, with $\tau'_i = \tau_i$ for $i \leq k-1$, $\tau'_k = \tau^*$, $\tau'_j = \tau_{j-1}$, $\lambda'_j = \lambda_{j-1}$ for $j \geq k+1$, and $\lambda'_i = \lambda_i$ for $i \leq k-2$.

The likelihood ratio is

$$\exp \left\{ N'_{k-1} \lambda'_{k-1} + N'_k \lambda'_k - N_{k-1} \lambda_{k-1} + e^{\lambda_{k-1}} \int_{\tau_{k-1}}^{\tau_k} Y(s) ds - e^{\lambda'_{k-1}} \int_{\tau_{k-1}}^{\tau'_k} Y(s) ds - e^{\lambda'_k} \int_{\tau'_k}^{\tau_k} Y(s) ds \right\}.$$

The prior ratio is

$$\alpha (2\pi)^{-1/2} \left(\frac{|A'|}{|A|} \right)^{1/2} \exp(\Phi_B(A, A', \mu, \lambda_m, \lambda'_{m'})/2),$$

where $\Phi_B(A, A', \mu, \lambda_m, \lambda'_{m'})$ is given by

$$\begin{aligned}& (\lambda_{k-2} - \mu)^2 (a_{k-2k-2} - a'_{k-2k-2}) + (\lambda_k - \mu)^2 (a_{kk} - a'_{kk}) \\ & + a_{k-1k-1} (\lambda_{k-1} - \mu)^2 - a'_{k-1k-1} (\lambda'_{k-1} - \mu)^2 - a'_{k+1k+1} (\lambda_k - \mu)^2 \\ & - 2(\lambda_{k-2} - \mu) [a'_{k-2k-1} (\lambda'_{k-1} - \mu) - a_{k-2k-1} (\lambda_{k-1} - \mu)] \\ & - 2(\lambda_k - \mu) [a'_{kk+1} (\lambda'_k - \mu) - a_{k-1k} (\lambda_{k-1} - \mu)] \\ & - 2a'_{k-1k} (\lambda'_{k-1} - \mu) (\lambda'_k - \mu).\end{aligned}$$

The proposal ratio for this type of move takes the simple form $2\epsilon/\alpha$.

The Jacobian of the transformation $\tau_{\epsilon, V}$ is $2\epsilon \Delta'_k \Delta'_{k-1} / \Delta_{k-1}^2$.

Move of type D. An index k is uniformly selected from $\{2, 3, \dots, m\}$ corresponding to the removal of the jump time τ_k . The proposed new point is $(\tau'_{m'}, \lambda'_{m'})$, with $\lambda'_i = \lambda_i$ for $i \leq k-2$, $\lambda'_j = \lambda_{j+1}$, $\tau'_j = \tau_{j+1}$ for $j \geq k$, and $\tau'_i = \tau_i$ for $i \leq k-1$. The likelihood, prior, proposal ratios, and the Jacobian for this type of move are the inverse ratios of the ones for the move of type B with the proper labelling of the jump times and the log-hazard levels.

The above type of moves imply that the corresponding Markov chain is π -irreducible. Denote by $\lambda_i(t), i = 1, \dots, N$ the sample paths generated by a run of length N of the MHG sampler. Then for any real valued function $H : \mathcal{S} \rightarrow \mathcal{R}$ integrable with respect to π , we have almost surely

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N H(\lambda_i(t)) = \int_{\mathcal{S}} H((\tau, \lambda)) d\pi((\tau, \lambda) | \text{data}),$$

see Meyn and Tweedie [13, Ch. 17]. In particular, the posterior mean hazard function $E_{\pi}(h(t) | \text{data})$ is approximated by

$$\frac{1}{N} \sum_{i=1}^N \exp(\lambda_i(t)),$$

and the predictive survival function $P_{\pi}(T > t | \text{data})$ by

$$\frac{1}{N} \sum_{i=1}^N \exp\left\{-\int_0^t \exp(\lambda_i(u)) du\right\}.$$

References

- [1] E. Arjas, and D. Gasbarra, *Nonparametric Bayesian inference for right-censored survival data, using the Gibbs sampler*, Statistica Sinica **4**(1994), 505-524.
- [2] E. Arjas and J. Heikkinen, *An algorithm for nonparametric Bayesian estimation of a Poisson intensity*, Journal of Computational Statistics **12**(1997), 385-402.
- [3] D. R. Cox, *Regression models and life-tables (with discussion)*, Journal of the Royal Statistical Society, Series B **34**(1972), 187-220.
- [4] D. R. Cox and D. Oakes, *Analysis of Survival Data*, Chapman and Hall, London, 1984.
- [5] H. Doss, *Bayesian nonparametric estimation for incomplete data via successive substitution sampling*, The Annals of Statistics **22**(1994), 1763-1786.
- [6] R. L. Dykstra and P. Laud, *A Bayesian nonparametric approach to reliability*, The Annals of Statistics **9**(1981), 356-367.
- [7] T. S. Ferguson, *A Bayesian analysis of some nonparametric problems*, The Annals of Statistics **1**(1973), 209-230.
- [8] D. Gamerman, *Dynamic Bayesian models for survival data*, Applied Statistics **40**(1991), 63-79.
- [9] P. J. Green, *Reversible jump Markov chain Monte Carlo computation and Bayesian model determination*, Biometrika **82**(1995), 711-732.
- [10] W. K. Hastings, *Monte Carlo sampling methods using Markov chains and their applications*, Biometrika **57**(1970), 97-109.
- [11] N. L. Hjort, *Nonparametric Bayes estimators based on beta processes in models for life history data*, The Annals of Statistics **18**(1990), 1259-1294.
- [12] I. W. McKeague and M. Tighiouart, *Bayesian estimators for conditional hazard functions*, Biometrics **56**(2000), 213-221.

- [13] S. P. Meyn and R. L. Tweedie, *Markov Chains and Stochastic Stability*, Springer-Verlag, London, 1993.
- [14] P. Muliere, *A Bayesian nonparametric approach to survival analysis using poly trees*, Scandinavian Journal of Statistics **24**(1997), 331-340.
- [15] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller and E. Teller, *Equation of State Calculations by Fast Computing Machines*, Journal of Chemical Physics **21**(1953), 1087-1092.
- [16] P. Damien, P. W. Laud and A. F. Smith, *Implementation of Bayesian nonparametric inference based on beta processes*, Scandinavian Journal of Statistics **23**(1996), 27-36.
- [17] D. Sinha and D. K. Dey, *Semiparametric Bayesian analysis of survival data*, Journal of the American Statistical Association **92**(1997), 1195-1212.
- [18] A. F. M. Smith and G. O. Roberts, *Bayesian computation via the Gibbs sampler and related Markov chain Monte Carlo methods*, Journal of the Royal Statistical Society, Series B **55**(1993), 3-23.
- [19] V. Susarla and J. Van Ryzin, *Nonparametric Bayesian estimation of survival curves from incomplete observations*, Journal of the American Statistical Association **71**(1976), 897-902.
- [20] M. West, *Modelling Time-Varying Hazards and Covariate Effects*. Survival Analysis: State of the Art, J. P. Klein, P. K. Goel, eds. Kluwer, Boston, 1992.

Department of Statistics, Florida State University, Tallahassee, FL 32306.

E-mail: mckeague@stat.fsu.edu

Department of Mathematics and Statistics, Utah State University, Logan, UT 84322.

E-mail: mourad@math.usu.edu