Ann Inst Stat Math (2011) 63:827–850
DOI 10.1007/s10463-009-0253-1

On Bayes inference for a bathtub failure rate via S-paths

Man-Wai Ho

Received: 24 October 2008 / Revised: 1 May 2009 / Published online: 31 July 2009

Abstract A class of semi-parametric hazard/failure rates with a bathtub shape is of interest. It does not only provide a great deal of flexibility over existing parametric methods in the modeling aspect but also results in a closed-form and tractable Bayes estimator for the bathtub-shaped failure rate. Such an estimator is derived to be a finite sum over two **S**-paths due to an explicit posterior analysis in terms of two (conditionally independent) **S**-paths. These, newly discovered, explicit results can be proved to be Rao–Blackwell improvements of counterpart results in terms of partitions that are readily available by a specialization of James' work (Ann Stat 33:1771–1799, 2005). Both iterative and non-iterative computational procedures are introduced for evaluating the hazard estimates. Two applications of the proposed methodology are discussed, of which one is about a Bayesian test for bathtub-shaped failure rates and the other is related to modeling with covariates.

Keywords Random partition · Completely random measure · Rao–Blackwellization · Sequential importance sampling · Gibbs sampler · Accelerated path sampler

1 Introduction

In reliability theory and survival analysis it is important to understand the propensity of failure of an item or death of a human being in the instant future given its survival until time t, which is called a hazard rate (or failure rate). Different shapes for the

M.-W. Ho (✉)
Department of Statistics and Applied Probability, National University of Singapore, 6 Science Drive 2, Singapore 117546, Singapore
e-mail: stahmw@nus.edu.sg

hazard as a function of t correspond to different life distributions for the subjects. In particular, a class of life distributions which corresponds to a bathtub-shaped failure rate (BFR) has received considerable attention as most of the electronic, eletromechanical, and mechanical products and human beings are subject to a high risk for failures/deaths initially in an "infant mortality" phase, then to a lower and constant risk in the so-called "useful life" period and finally to an increasing risk with time during the so-called "wearout" phase. Many parametric families of distributions for BFRs have been proposed over the last few decades. Most of which typically involving three or more parameters are based on mixtures or generalizations of some common probability distributions, such as exponential, gamma, Weibull and Pareto distributions; see Rajarshi and Rajarshi (1988) and Lai and Murthy (2001, Section 4) for an extensive and collective review. For discussion of parametric models for other typical hazard functions, see Kalbfleisch and Prentice (1980) and Lawless (1982). Also see Singpurwalla (2006) for a comprehensive discussion on reliability and risk from a Bayesian perspective.

The literature on nonparametric estimation of BFRs is rather limited. Early works include, for example, Bray et al. (1967a,b). Amman (1984) and Laud et al. (2006) studied a U-shaped process by combining two random processes, of which one is the increasing random hazard rates based on extended gamma processes first considered by Dykstra and Laud (1981) and the other one is the decreasing counterpart defined analogously. However, the combined process does not necessarily generate BFRs. Reboul (2005) introduced a data-driven nonparametric estimator of BFRs which, though is not in a closed form, can be computed by applying the "Pool Adjacent Violators Algorithm" discussed in Barlow et al. (1972). Some other recent works can be found in Jankowski and Wellner (2008a,b).

The present paper considers a semi-parametric family of hazard rates on $\mathcal{H} = (0, \infty)$ which has a mixture form, for $t, \theta \in \mathcal{H}$,

$$\lambda(t|\mu, \theta) = \int_{\mathcal{R}} K_\theta(t, u)\mu(\mathrm{d}u), \tag{1}$$

where the kernel $K_\theta(t, u) = \mathbb{I}(t - \theta \leq u < 0) + \mathbb{I}(0 < u \leq t - \theta)$ with $\mathbb{I}(\cdot)$ being an indicator function and μ is a completely random measure (CRM) without drift (Kingman 1967, 1993) on $\mathcal{R} = (-\infty, \infty)$. This class of hazard rates generalize and include the class of monotone failure rates (MFRs) discussed in Ho (2006a) as subclasses when $\theta = 0$ or $\theta = \infty$. Based on the mixture representation of MFRs considered there, analogous argument of constructing unimodal densities on the real line with mode θ in Brunner (1992) applies and verifies that (1) gives an BFR on $\mathcal{H}$ with a minimum point, or a change point called by Mitra and Basu (1995), at $\theta \in \mathcal{H}$.

James (2005) studied nonparametric hazard rates with general shapes expressible as in (1), wherein the kernel is any known positive measurable kernel on a Polish space $\mathcal{X} \times \mathcal{U}$ and μ is a CRM on $\mathcal{U}$. An explicit posterior analysis in terms of random partitions was derived. Partition-based posterior analyses for the subclasses of MFRs considered in Ho (2006a), as well as BFRs in (1), is readily available as special cases. Ho (2006a), exploiting special structure of an indicator kernel defined by $K_\theta(t, u)$ with

$\theta = 0$ in (1), improves James' general result in such a special case by giving a tractable and less complex characterization in terms of one random **S**-path for all MFRs, and shows via simulation studies that an efficiently designed algorithm for sampling an **S**-path, called accelerated path (AP) sampler, results in less variable Bayes estimates of MFRs, compared to a partition-based algorithm introduced by James (2005).

In this work, we show that all BFRs defined in (1) possess special structures that naturally arise in relation to two conditionally independent **S**-paths given θ in Sect. 2, rather than one in the case of MFRs. A novel and explicit characterization of the posterior distribution in terms of two random **S**-paths for all such BFRs, which generalizes the corresponding characterization of MFRs discussed in Ho (2006a) that depends on only one path, results in a closed and tractable posterior mean of BFRs as a finite sum over two **S**-paths that serve as a viable estimator of any BFR and, hence, an alternative to most existing parametric inferences which suffer from intractability problems (Lawless 1982, pp. 255) and often resort to extensive iterative procedure (Haupt and Schabe 1997). Understanding these characterization and estimator for BFRs is of statistical importance as they are shown to be Rao–Blackwell improvements of the partition-based counterparts, which in turn suggests that more parsimonious methods for inference, compared with partition-based methods introduced in James (2005), would be available if one could efficiently sample the two paths in this context, as shown in Ho (2006a) for the special case of MFRs. In this regard, Sect. 3 proposes an iterative Monte Carlo procedure extending the AP sampler. Furthermore, a sequential importance sampling (SIS) (Kong et al. 1994; Liu and Chen 1998) scheme for sampling one path at a time is introduced and its extension applicable in our model is discussed. Simulation studies are reported in Sect. 4 to demonstrate practicality and effectiveness of our methodology. Two applications of the methodology are presented in the last two sections in which the proposed algorithms can be applied to approximate the posterior quantities of interest. A test of an MFR versus an BFR based on models in (1) is illustrated in Sect. 5. Section 6 shows that a posterior distribution explicitly characterized by two random **S**-paths also exists in modeling with covariates by a proportional hazards model.

2 Posterior analysis via two S-paths

Suppose that we collect independent failure times $\mathbf{T} = (T_1, \ldots, T_N)$ from N items with a common continuous life distribution which corresponds to an BFR with change point at θ, specified by (1), until time τ, so that $0 < T_1 < \cdots < T_c < \tau$ denote c completely observed failure times, and $T_{c+1} = \cdots = T_N \equiv \tau$ are $N - c$ number of right-censored times. Assuming a multiplicative intensity model discussed in (Aalen 1975, 1978), the likelihood of the data $\mathbf{T}$ is proportional to

$$\exp[-\mu(g_{N,\theta})] \prod_{k=1}^{c} \int K_\theta(T_k, u_k)\mu(\mathrm{d}u_k), \tag{2}$$

where $g_{N,\theta}(u) = \int_0^\tau [\sum_{i=1}^N \mathbb{I}(T_i \geq t)] K_\theta(t,u)\mathrm{d}t$ is a piecewise linear function of u (see 27) and $\mu(g_{N,\theta}) = \int_{\mathcal{R}} g_{N,\theta}(u)\mu(\mathrm{d}u) = \int_0^\tau [\sum_{i=1}^N \mathbb{I}(T_i \geq t)]\lambda(t|\mu,\theta)\mathrm{d}t$ with $\sum_{i=1}^N \mathbb{I}(T_i \geq t)$ called the total time on test (TTT) transform (Barlow et al. 1972).

Suppose that μ, which is uniquely characterized by the Laplace functional

$$\mathcal{L}_\mu(g|\rho,\eta) = \exp\left[-\int_{\mathcal{R}}\int_{\mathcal{H}}\left(1-\mathrm{e}^{-g(u)x}\right)\rho(\mathrm{d}x|u)\eta(\mathrm{d}u)\right], \tag{3}$$

for any non-negative function g on $\mathcal{R}$, is a CRM with a Lévy measure $\rho(\mathrm{d}x|u)\eta(\mathrm{d}u)$ such that $\int_B \int_{\mathcal{H}} \min(x,1)\rho(\mathrm{d}x|u)\eta(\mathrm{d}u) < \infty$ for any bounded set $B \subset \mathcal{R}$; see more discussion on CRM in Kingman (1967, 1993). The posterior law of the pair (μ,θ) given $\mathbf{T}$ according to the likelihood (2) with respect to (w.r.t.) any prior $\pi(\mathrm{d}\theta)$ for $\theta \in \mathcal{H}$ can always be determined by the double expectation formula,

$$E[h(\mu,\theta)|\mathbf{T}] = E\{E[h(\mu,\theta)|\theta,\mathbf{T}]|\mathbf{T}\} = \int_{\mathcal{H}}\int_{\mathcal{M}} h(\mu,\theta)\mathcal{P}(\mathrm{d}\mu|\theta,\mathbf{T})\mathcal{P}(\mathrm{d}\theta|\mathbf{T}), \tag{4}$$

where h is any nonnegative or integrable function, $\mathcal{M}$ is the space of measures over $\mathcal{R}$, and, $\mathcal{P}(\mathrm{d}\mu|\theta,\mathbf{T})$ and $\mathcal{P}(\mathrm{d}\theta|\mathbf{T})$ denote the law of μ given $(\theta,\mathbf{T})$ and the law of θ given $\mathbf{T}$, respectively.

One can first look at the posterior law of μ for a fixed value of θ. Suppose that $0 < \theta < \tau$ is given. We can always assume that

$$(T_1-\theta,\ldots,T_c-\theta) = \hat{\mathbf{Z}} \cup \bar{\mathbf{Z}} = (\hat{Z}_1,\hat{Z}_2,\ldots,\hat{Z}_{\hat{n}}) \cup (\bar{Z}_1,\bar{Z}_2,\ldots,\bar{Z}_{\bar{n}}), \tag{5}$$

where $\hat{n}$ and $\bar{n}$ are non-negative integers such that $\hat{n}+\bar{n} = c$, $-\theta \equiv \hat{Z}_0 < \hat{Z}_1 < \cdots < \hat{Z}_{\hat{n}} < \hat{Z}_{\hat{n}+1} \equiv 0$ and $0 \equiv \bar{Z}_{\bar{n}+1} < \bar{Z}_{\bar{n}} < \cdots < \bar{Z}_1 < \bar{Z}_0 \equiv \tau - \theta$ are referred to as "negative" and "positive" observations. In the sequel, more pairs of symbols and expressions with "hat" and "overline" are defined to represent objects to the left and to the right of a given value of θ, respectively.

Lemma 1 *Once an observation T_k is completely observed from (1), for $k = 1,\ldots,c$, the kernel $K_\theta(T_k,u_k)$ in (2) simplifies substantially to two different expressions according to either one of two mutually exclusive situations governed by $T_k - \theta$, or equivalently, the BFR in (1) reduces to*

$$\lambda(T_k|\mu,\theta) = \begin{cases} \int \mathbb{I}(\hat{Z}_i \leq u_k < 0)\mu(\mathrm{d}u_k), & \hat{Z}_i = T_k - \theta < 0, \\ \int \mathbb{I}(0 < u_k \leq \bar{Z}_j)\mu(\mathrm{d}u_k), & \bar{Z}_j = T_k - \theta > 0, \end{cases} \tag{6}$$

for some $i = 1,\ldots,\hat{n}$ and $j = 1,\ldots,\bar{n}$.

The above lemma reveals that when θ is known such that the difference between an observation T_k and θ is greater than zero (resp. less than zero), the latent variable u_k associated with T_k in (1) must be greater than zero (resp. less than zero). This plays a crucial role in characterizing the posterior distribution of models in (1) in terms of a pair of **S**-paths (see the proof of Theorem 1), where an **S**-path (Ho 2006a,b) is defined as follows.

Definition 1 For any integer $c > 0$, an **S**-path (of $c+1$ coordinates) is an integer-valued vector $\mathbf{S} = (S_0, S_1, \ldots, S_{c-1}, S_c)$, satisfying

(i) $S_0 \equiv 0$ and $S_c \equiv c$;
(ii) $S_j \leq j$, $j = 1, \ldots, c-1$; and
(iii) $S_j \leq S_{j+1}$, $j = 0, 1, \ldots, c-1$.

Dykstra and Laud (1981) and Brunner and Lo (1989) respectively made use of **e**-vectors and **m**-vectors, identical to each other, in Bayes estimation of MFRs and monotone or symmetric unimodal densities, of which there exists a one-to-one relationship between each vector and an **S**-path through accumulation of (or incremental difference between) the entries. In general, an **S**-path can be viewed as or created via a combinatorial reduction from partitions; in particular, a path **S** of $c+1$ coordinates is said to *correspond to* one or many partitions $\mathbf{p} = \{C_1, \ldots, C_{n(\mathbf{p})}\}$ of the integers $\{1, \ldots, c\}$, provided that (i) indices of the maximal elements of the $n(\mathbf{p})$ clusters C_i's in **p** coincide with locations j at which $S_j > S_{j-1}$, and (ii) number of indices e_i in cluster C_i for all $i = 1, \ldots, n(\mathbf{p})$ with a maximal index j, $j = 1, \ldots, c$, is identical to $S_j - S_{j-1}$. Brunner and Lo (1989) showed that there are

$$|\mathbb{C}_{\mathbf{S}}| \equiv \sum_{\mathbf{p}\in\mathbb{C}_{\mathbf{S}}} 1 = \prod_{\{j\,|\mathbf{S}\}} \binom{j-1-S_{j-1}}{j-S_j} \tag{7}$$

total number of partitions in the collection of all partitions of c integers that correspond to a path **S** of $c+1$ coordinates, denoted by $\mathbb{C}_{\mathbf{S}}$, where $\prod_{\{j\,|\mathbf{S}\}}$ stands for $\prod_{j=1:S_j>S_{j-1}}^{c}$ if there are $c+1$ coordinates in **S**. Similarly, $\sum_{\{j\,|\mathbf{S}\}}$ will stand for $\sum_{j=1:S_j>S_{j-1}}^{c}$.

Define $f_{N,\theta}(x,u) = g_{N,\theta}(u)x$, for any $(x,u) \in (\mathcal{H}, \mathcal{R})$, and assume that

$$\kappa_\ell(\mathrm{e}^{-f_{N,\theta}}\rho|u) = \int_{\mathcal{H}} x^\ell \mathrm{e}^{-g_{N,\theta}(u)x}\rho(\mathrm{d}x|u) < \infty, \tag{8}$$

for any positive integer $\ell \leq c$ and a fixed $u \in \mathcal{R}$. Now, as in James (2005), re-writing

$$\mu(\mathrm{d}u_k) = \int_{\mathcal{H}} x_k \mathcal{N}(\mathrm{d}x_k, \mathrm{d}u_k), \quad k = 1, \ldots, c, \tag{9}$$

in (2), where $\mathcal{N}(\mathrm{d}x, \mathrm{d}u)$ is a Poisson random measure taking on points (x,u) in $\mathcal{H} \times \mathcal{R}$ with mean intensity measure $E[\mathcal{N}(\mathrm{d}x, \mathrm{d}u)] = \rho(\mathrm{d}x|u)\eta(\mathrm{d}u)$, and augmenting the latent variables $\mathbf{u} = (u_1, \ldots, u_c)$ and $\mathbf{x} = (x_1, \ldots, x_c)$ yield that given the value of θ, the law of μ given the data can be described by the law of μ given $(\mathbf{u}, \mathbf{x})$ and **T** mixed over by that of $(\mathbf{u}, \mathbf{x})$ given **T**.

Theorem 1 *Suppose that the value of θ is given. The likelihood of the data* **T** *is given by* (2) *with μ represented as in* (9)*. Then, the distribution of* $(\mathbf{u}, \mathbf{x})$ *given* **T** *can be described by a joint distribution of* $(\hat{\mathbf{y}}, \hat{\mathbf{Q}}, \hat{\mathbf{S}})$ *and* $(\bar{\mathbf{y}}, \bar{\mathbf{Q}}, \bar{\mathbf{S}})$ *given* **T** *defined as follows:*

(i) *Given* **T***, there are two paths* $\hat{\mathbf{S}} = (0, \hat{S}_1, \ldots, \hat{S}_{\hat{n}-1}, \hat{n})$ *and* $\bar{\mathbf{S}} = (0, \bar{S}_1, \ldots, \bar{S}_{\bar{n}-1}, \bar{n})$ *of $\hat{n}+1$ coordinates and of $\bar{n}+1$ coordinates to the left and to the right of θ, respectively. They are independently distributed as*

$$\hat{W}(\hat{\mathbf{S}}|\theta,\mathbf{T}) \propto \hat{\phi}_\theta(\hat{\mathbf{S}},\mathbf{T}) = |\mathbb{C}_{\hat{\mathbf{S}}}| \prod_{\{j|\hat{\mathbf{S}}\}} \int_{\hat{Z}_j}^{0} \kappa_{\hat{m}_j}(\mathrm{e}^{-f_{N,\theta}}\rho|y)\eta(\mathrm{d}y) \tag{10}$$

and

$$\bar{W}(\bar{\mathbf{S}}|\theta,\mathbf{T}) \propto \bar{\phi}_\theta(\bar{\mathbf{S}},\mathbf{T}) = |\mathbb{C}_{\bar{\mathbf{S}}}| \prod_{\{j|\bar{\mathbf{S}}\}} \int_{0}^{\bar{Z}_j} \kappa_{\bar{m}_j}(\mathrm{e}^{-f_{N,\theta}}\rho|y)\eta(\mathrm{d}y), \tag{11}$$

where $|\mathbb{C}_{\hat{\mathbf{S}}}|$ *and* $|\mathbb{C}_{\bar{\mathbf{S}}}|$ *are defined in (7),* $\hat{m}_j \equiv \hat{S}_j - \hat{S}_{j-1}$, $j = 1,\ldots,\hat{n}$ *and* $\bar{m}_j \equiv \bar{S}_j - \bar{S}_{j-1}$, $j = 1,\ldots,\bar{n}$.

(ii) *Given* $(\hat{\mathbf{S}},\bar{\mathbf{S}})$ *and* $\mathbf{T}$*, there exist* $\sum_{\{j|\hat{\mathbf{S}}\}} 1$ *and* $\sum_{\{j|\bar{\mathbf{S}}\}} 1$ *independent pairs of* $(\hat{y}_j,\hat{Q}_j)$ *and* $(\bar{y}_j,\bar{Q}_j)$*, respectively denoted by* $(\hat{\mathbf{y}},\hat{\mathbf{Q}}) = \{(\hat{y}_j,\hat{Q}_j) : \hat{m}_j > 0, j = 1,\ldots,\hat{n}\}$ *and* $(\bar{\mathbf{y}},\bar{\mathbf{Q}}) = \{(\bar{y}_j,\bar{Q}_j) : \bar{m}_j > 0, j = 1,\ldots,\bar{n}\}$*. They are distributed as*

$$\hat{\eta}_j(\mathrm{d}\hat{y}_j|\hat{\mathbf{S}},\theta,\mathbf{T}) \propto \mathbb{I}(\hat{Z}_j \le \hat{y}_j < 0)\kappa_{\hat{m}_j}(\mathrm{e}^{-f_{N,\theta}}\rho|\hat{y}_j)\eta(\mathrm{d}\hat{y}_j), \tag{12}$$

$$\Pr\{\hat{Q}_j \in \mathrm{d}q|\hat{y}_j,\hat{\mathbf{S}},\theta,\mathbf{T}\} \propto q^{\hat{m}_j}\mathrm{e}^{-g_{N,\theta}(\hat{y}_j)q}\rho(\mathrm{d}q|\hat{y}_j), \tag{13}$$

and

$$\bar{\eta}_j(\mathrm{d}\bar{y}_j|\bar{\mathbf{S}},\theta,\mathbf{T}) \propto \mathbb{I}(0 < \bar{y}_j \le \bar{Z}_j)\kappa_{\bar{m}_j}(\mathrm{e}^{-f_{N,\theta}}\rho|\bar{y}_j)\eta(\mathrm{d}\bar{y}_j), \tag{14}$$

$$\Pr\{\bar{Q}_j \in \mathrm{d}q|\bar{y}_j,\bar{\mathbf{S}},\theta,\mathbf{T}\} \propto q^{\bar{m}_j}\mathrm{e}^{-g_{N,\theta}(\bar{y}_j)q}\rho(\mathrm{d}q|\bar{y}_j), \tag{15}$$

respectively, with existences guaranteed by (8).

The above characterization of the posterior distribution of $(\mathbf{u},\mathbf{x})$ for models in (1) in terms of two $\mathbf{S}$-paths, though look more involved symbolically, are less complex than (or as complex as only when $\hat{n},\bar{n} < 4$) the characterization in terms of partitions specialized from James (2005), which is proportional to (34) or (35). See Remark 1 for discussion in detail. Applications of the same arguments as in James (2005) result in the following two corollaries.

Corollary 1 *Suppose that the value of* θ *is given. The likelihood of the data* $\mathbf{T}$ *is given by (2) and* μ *is a CRM characterized by (3). Then, the law of* μ *given* $\mathbf{T}$ *can be described as the law of the random measure,*

$$\mu^* = \mu_{g_{N,\theta}} + \sum_{\{j|\hat{\mathbf{S}}\}} \hat{Q}_j\delta_{\hat{y}_j} + \sum_{\{j|\bar{\mathbf{S}}\}} \bar{Q}_j\delta_{\bar{y}_j},$$

mixed over by the joint distribution of $(\hat{\mathbf{y}},\hat{\mathbf{Q}},\hat{\mathbf{S}})$ *and* $(\bar{\mathbf{y}},\bar{\mathbf{Q}},\bar{\mathbf{S}})$ *given* $\mathbf{T}$ *defined in Theorem 1, where* $\mu_{g_{N,\theta}}$ *is a CRM with Lévy measure* $\mathrm{e}^{-g_{N,\theta}(u)x}\rho(\mathrm{d}x|u)\eta(\mathrm{d}u)$*, conditionally independent of* $(\hat{\mathbf{y}},\hat{\mathbf{Q}},\hat{\mathbf{S}})$ *and* $(\bar{\mathbf{y}},\bar{\mathbf{Q}},\bar{\mathbf{S}})$ *given* $\mathbf{T}$.

Corollary 2 *A Bayes estimate of the BFRs in* (1) *with a known value of* θ *is given by the posterior expectation, for* $t \in [0, \tau]$,

$$E[\lambda(t|\mu,\theta)|\theta,\mathbf{T}] = \sum_{\hat{\mathbf{S}}}\sum_{\bar{\mathbf{S}}} a_\lambda(t|\hat{\mathbf{S}},\bar{\mathbf{S}},\theta,\mathbf{T})\, W(\hat{\mathbf{S}},\bar{\mathbf{S}}|\theta,\mathbf{T}), \tag{16}$$

where $\sum_{\mathbf{S}}$ *represents summing over all paths* $\mathbf{S}$ *of the same number of coordinates,* $W(\hat{\mathbf{S}},\bar{\mathbf{S}}|\theta,\mathbf{T}) = \hat{W}(\hat{\mathbf{S}}|\theta,\mathbf{T})\bar{W}(\bar{\mathbf{S}}|\theta,\mathbf{T})$, *and*

$$\begin{aligned} a_\lambda(t|\hat{\mathbf{S}},\bar{\mathbf{S}},\theta,\mathbf{T}) \;=\; & \left[\int_{t-\theta}^{0} \kappa_1(\mathrm{e}^{-f_{N,\theta}}\rho|y)\eta(\mathrm{d}y) + \sum_{\{j|\hat{\mathbf{S}}\}} \hat{\lambda}_{\theta,j}(t|\hat{\mathbf{S}})\right]\mathbb{I}(t<\theta) \\ & + \left[\int_{0}^{t-\theta} \kappa_1(\mathrm{e}^{-f_{N,\theta}}\rho|y)\eta(\mathrm{d}y) + \sum_{\{j|\bar{\mathbf{S}}\}} \bar{\lambda}_{\theta,j}(t|\bar{\mathbf{S}})\right]\mathbb{I}(t>\theta), \end{aligned}$$

wherein $\hat{\lambda}_{\theta,j}(t|\hat{\mathbf{S}}) = \int_{\max(t-\theta,\hat{Z}_j)}^{0} \kappa_{\hat{m}_j+1}(\mathrm{e}^{-f_{N,\theta}}\rho|y)\eta(\mathrm{d}y) / \int_{\hat{Z}_j}^{0} \kappa_{\hat{m}_j}(\mathrm{e}^{-f_{N,\theta}}\rho|y)\eta(\mathrm{d}y)$, *for* $j = 1,\ldots,\hat{n}$, *and* $\bar{\lambda}_{\theta,j}(t|\bar{\mathbf{S}}) = \int_{0}^{\min(t-\theta,\bar{Z}_j)} \kappa_{\bar{m}_j+1}(\mathrm{e}^{-f_{N,\theta}}\rho|y)\eta(\mathrm{d}y) \,/\, \int_{0}^{\bar{Z}_j} \kappa_{\bar{m}_j}(\mathrm{e}^{-f_{N,\theta}}\rho|y)\eta(\mathrm{d}y)$, *for* $j = 1,\ldots,\bar{n}$.

The above results generalize Ho (2006a) explicit posterior analysis via one single **S**-path for the class of MFRs defined in (1) when $\theta = 0$ or $\theta = \infty$. From a statistical point of view, the above Bayes estimate of the unknown hazard in terms of two **S**-paths should always be preferred to the counterpart result in terms of partitions specialized from Corollary 4.1 in James (2005) if one can directly sample $(\hat{\mathbf{S}},\bar{\mathbf{S}})$ from $W(\hat{\mathbf{S}},\bar{\mathbf{S}}|\theta,\mathbf{T})$ as a consequence of the Rao–Blackwell theorem (see Appendix for discussion in detail).

When θ is not known, posterior analysis of models in (1) follows from (4) with $\mathcal{P}(\mathrm{d}\theta|\mathbf{T})$ defined below.

Theorem 2 *Suppose that the likelihood of the data* $\mathbf{T}$ *given* (μ,θ) *is proportional to* (2). *Assume that* μ *is a CRM characterized by* (3) *and the prior of* θ *is* $\pi(\mathrm{d}\theta)$, *independent of* μ. *The posterior distribution of* θ, $\mathcal{P}(\mathrm{d}\theta|\mathbf{T})$, *is characterized by, for any Borel set* $B \in \mathcal{H}$,

$$P(\theta \in B|\mathbf{T}) = \int_B \sum_{\hat{\mathbf{S}}}\sum_{\bar{\mathbf{S}}} \pi(\hat{\mathbf{S}},\bar{\mathbf{S}},\mathrm{d}\theta|\mathbf{T}), \tag{17}$$

where

$$\pi(\hat{\mathbf{S}},\bar{\mathbf{S}},\mathrm{d}\theta|\mathbf{T}) \propto \mathcal{L}_\mu(g_{N,\theta}|\rho,\eta)\hat{\phi}_\theta(\hat{\mathbf{S}},\mathbf{T})\,\bar{\phi}_\theta(\bar{\mathbf{S}},\mathbf{T})\pi(\mathrm{d}\theta) \tag{18}$$

defines the posterior distribution of $(\hat{\mathbf{S}},\bar{\mathbf{S}},\theta)$ *given* $\mathbf{T}$, *with a normalization constant* $\int_{\mathcal{H}} \mathcal{L}_\mu(g_{N,\vartheta}|\rho,\eta)[\sum_{\hat{\mathbf{S}}}\hat{\phi}_\vartheta(\hat{\mathbf{S}},\mathbf{T})]\,[\sum_{\bar{\mathbf{S}}}\bar{\phi}_\vartheta(\bar{\mathbf{S}},\mathbf{T})]\pi(d\vartheta)$, *and* $\mathcal{L}_\mu(\cdot|\rho,\eta)$, $\hat{\phi}_\theta(\hat{\mathbf{S}},\mathbf{T})$ *and* $\bar{\phi}_\theta(\bar{\mathbf{S}},\mathbf{T})$ *are defined in* (3), (10) *and* (11), *respectively.*

Table 1 Complexities between characterizations of BFRs in (1) in terms of two **S**-paths (in Theorem 1) and in terms of partitions (specialized from James 2005) versus sample sizes $\bar{n}$ and $\hat{n} = 20 - \bar{n}$

$\bar{n}$	$20 - \bar{n}$	$\Lambda_{\bar{n}} \times \Lambda_{20-\bar{n}}$	$B_{\bar{n}} \times B_{20-\bar{n}}$	Ratio (%)
10	10	282,105,616	13,450,200,625	2.097
8	12	297,457,160	17,444,291,580	1.705
6	14	353,026,080	38,752,562,366	0.911
4	16	495,007,380	157,202,132,205	0.315
2	18	1,432,916,100	2,046,230,418,477	0.070
0	20	6,564,120,420	51,724,158,235,372	0.013

For instance, the posterior mean of hazard rates in (1) is given by

$$E[\lambda(t|\mu,\theta)|\mathbf{T}] = \int_{\mathcal{H}} \sum_{\hat{\mathbf{S}}} \sum_{\bar{\mathbf{S}}} a_\lambda(t|\hat{\mathbf{S}},\bar{\mathbf{S}},\theta,\mathbf{T})\pi(\hat{\mathbf{S}},\bar{\mathbf{S}},\mathrm{d}\theta|\mathbf{T}), \tag{19}$$

where $a_\lambda(t|\hat{\mathbf{S}},\bar{\mathbf{S}},\theta,\mathbf{T})$ is defined in Corollary 2.

Remark 1 As the total number of partitions of c integers given by the Bell's exponential number B_c is roughly equal to $c!$, the complexity of the partition-based characterization summarized by (35) relies on the total numbers of pairs of partitions $\hat{\mathbf{p}}$ and $\bar{\mathbf{p}}$ (defined in the proof of Theorem 1 in Appendix) and is roughly equal to $\hat{n}! \times \bar{n}!$ (that is, the number of summands in $\sum_{\mathbf{p}}$). Meanwhile, the complexity of the new characterization described in Theorem 1 is based on the total numbers of **S**-paths of $\hat{n}+1$ and $\bar{n}+1$ coordinates and is given by $\Lambda_{\hat{n}} \times \Lambda_{\bar{n}}$ (that is, the number of summands in the double sum $\sum_{\hat{\mathbf{S}}} \sum_{\bar{\mathbf{S}}}$) where Λ_c, for any integer $c > 0$, denotes the total number of **S**-paths of $c+1$ coordinates. Hence, its complexity is less than that in (35) except when both $\hat{n}$ and $\bar{n}$ are less than 4 because $\Lambda_c \leq B_c$ for all integers $c > 0$, with equality only when $c < 4$ (Brunner and Lo 1989; Ho 2006a). Table 1 shows a ratio between the complexities of (35) and the new characterization in terms of two **S**-paths to be at most 0.02097 when $c = 20$. This upper bound on the ratio drops quickly as c increases; for example, the bound is given by $0.00013^2 = 1.69 \times 10^{-8}$ when $c = 40$.

3 Monte Carlo procedures

This section introduces Monte Carlo procedures for evaluating/approximating posterior quantities of models in (1), like (16), (17) and (19), which are expressible as finite sums over two **S**-paths, based on sampling the triplets $(\hat{\mathbf{S}}, \bar{\mathbf{S}}, \theta)$ in light of the data. For brevity, conditioning statements on the data **T** will be suppressed throughout this section as all sampling procedures are designed w.r.t. distributions conditioning on **T**. First, when θ is given, both iterative and non-iterative procedures for sampling the paths $(\hat{\mathbf{S}}, \bar{\mathbf{S}})$ will be discussed. Then, a sequential importance sampling (SIS) scheme for drawing the triplets from the posterior distribution $\pi(\hat{\mathbf{S}}, \bar{\mathbf{S}}, \mathrm{d}\theta|\mathbf{T})$ in (18) is

proposed. Derivations of these algorithms naturally follow from one of the nice probabilistic features of the posterior distribution for models in (1), namely, conditional independence between $\hat{\mathbf{S}}$ and $\bar{\mathbf{S}}$ conditional on θ and $\mathbf{T}$ in statement (i) of Theorem 1.

3.1 When θ is known

3.1.1 A Gibbs sampler

As a consequence of conditional independence between $\hat{\mathbf{S}}$ and $\bar{\mathbf{S}}$ given θ and $\mathbf{T}$, an iterative scheme, dubbed as *accelerated paths* (APs) *sampler*, for sampling a pair of $(\hat{\mathbf{S}}, \bar{\mathbf{S}})$ from the posterior distribution $W(\hat{\mathbf{S}}, \bar{\mathbf{S}}|\theta, \mathbf{T})$ defined in Corollary 2, can be naturally designed as two independent implementations of the AP sampler (Algorithm 3.1 in Ho 2006a), or equivalently, by cycling through the following two interchangeable steps in a cycle:

(G1) Determine $\hat{\mathbf{S}}$ by applying the AP sampler with n, $\phi(\mathbf{S})$, and $T_1, \ldots, T_n$ there replaced by $\hat{n}$, $\hat{\phi}_\theta(\hat{\mathbf{S}}, \mathbf{T})$ in (10), $\hat{Z}_1, \ldots, \hat{Z}_{\hat{n}}$, respectively.

(G2) Determine $\bar{\mathbf{S}}$ by applying the AP sampler with n, $\phi(\mathbf{S})$ and $T_1, \ldots, T_n$ there replaced by $\bar{n}$, $\bar{\phi}_\theta(\bar{\mathbf{S}}, \mathbf{T})$ in (11) and $\bar{Z}_1, \ldots, \bar{Z}_{\bar{n}}$, respectively.

A Markov chain $(\hat{\mathbf{S}}_{(0)}, \bar{\mathbf{S}}_{(0)}), \ldots, (\hat{\mathbf{S}}_{(M)}, \bar{\mathbf{S}}_{(M)})$ with a unique stationary distribution $W(\hat{\mathbf{S}}, \bar{\mathbf{S}}|\theta, \mathbf{T})$ can be obtained by starting with an arbitrary pair of paths $\hat{\mathbf{S}}_{(0)}$ and $\bar{\mathbf{S}}_{(0)}$, and repeating M cycles of steps (G1) and (G2). Then, expectations of any functional $f(\hat{\mathbf{S}}, \bar{\mathbf{S}})$ w.r.t. the probability distribution $W(\hat{\mathbf{S}}, \bar{\mathbf{S}}|\theta, \mathbf{T})$ can be approximated by the ergodic average (Meyn and Tweedie 1993)

$$\nu^M_{f,\theta} = \frac{1}{M} \sum_{i=1}^{M} f(\hat{\mathbf{S}}_{(i)}, \bar{\mathbf{S}}_{(i)}).$$

For instance, the posterior expectation $\mathbb{E}[\lambda(t|\mu, \theta)|\theta, \mathbf{T}]$ in (16) can be approximated by

$$\frac{1}{M} \sum_{i=1}^{M} a_\lambda(t|\hat{\mathbf{S}}_{(i)}, \bar{\mathbf{S}}_{(i)}, \theta, \mathbf{T}). \tag{20}$$

Due to the independence between any pair of $\hat{\mathbf{S}}_{(i)}$ and $\bar{\mathbf{S}}_{(j)}$, $i, j = 1, \ldots, M$, one may achieve better efficiency by defining

$$\zeta^M_{f,\theta} = \frac{1}{M^2} \sum_{i=1}^{M} \sum_{j=1}^{M} f(\hat{\mathbf{S}}_{(i)}, \bar{\mathbf{S}}_{(j)})$$

as an alternative to $\nu^M_{f,\theta}$ with the same number of iterations.

3.1.2 A sequential importance sampling scheme

In analogy to construction of the APs sampler, a sequential importance sampling (SIS) (Kong et al. 1994; Liu and Chen 1998) scheme for sampling the two conditional

independent paths from the target distribution $W(\hat{\mathbf{S}}, \bar{\mathbf{S}}|\theta, \mathbf{T})$ is designed as two independent implementations of another SIS scheme for sampling one path at a time, called the *sequential importance path* (SIP) *sampler*. Specifically, the SIP sampler is an SIS scheme that allows us to draw an $\mathbf{S}$-path of $c+1$ coordinates according to any probability distribution $\pi(\mathbf{S}) \propto \phi(\mathbf{S})$ through $c-1$ determinations of the random entries one at a time. The order in which the entries are determined follows $\mathcal{I}^{c-1} = \{I_1, \ldots, I_{c-1}\}$, a random permutation of $\{1, 2, \ldots, c-1\}$, such that all possible sequences are equally likely with probability $1/(c-1)!$. Let $\Xi_0^c = \{I_0, I_c\}$ with $I_0 \equiv 0$ and $I_c \equiv c$.

Algorithm 1 (The SIP sampler) Given Ξ_0^c and $\mathcal{I}^{c-1}$, an SIS scheme for sampling $\mathbf{S} = (0, S_1, \ldots, S_{c-1}, c)$ from a target distribution $\pi(\mathbf{S}) \propto \phi(\mathbf{S})$ consists of recursive applications of the following SIS steps for $r = 1, \ldots, c-1$:

A. From $\Xi_r^c = \Xi_{r-1}^c \cup \{I_r\}$, which is the index set of all determined coordinates of a randomly selected $\mathbf{S}$-path after step r, obtain $p = \max\{j \in \Xi_{r-1}^c : j < I_r\}$ and $q = \min\{j \in \Xi_{r-1}^c : j > I_r\}$. Set $S_{I_r} = k$, for $k = S_p, S_p + 1, \ldots, \min(I_r, S_q)$, according to a probability distribution

$$\sigma_r(k|\{S_\ell : \ell \in \Xi_{r-1}^c\}) \propto \phi(\widetilde{\mathbf{S}}_{I_r,k}),$$

where $\widetilde{\mathbf{S}}_{I_r,k} = (0, \widetilde{S}_1, \ldots, \widetilde{S}_{c-1}, c)$ is a path of $c+1$ coordinates such that $\widetilde{S}_{I_r} = k$, and, for $i = 1, \ldots, I_r - 1, I_r + 1, \ldots, c-1$, $\widetilde{S}_i = S_\ell$ if $i = \ell \in \Xi_{r-1}^c$; otherwise, $\widetilde{S}_i = \widetilde{S}_{i-1}$.

B. Compute $\sigma_r(k|\{S_\ell : \ell \in \Xi_{r-1}^c\})$, which equals $\phi(\widetilde{\mathbf{S}}_{I_r,k})$ multiplied by the appropriate constant of proportionality, for the chosen value k of S_{I_r}.

After step $c-1$, a random path $\mathbf{S} = (0, S_1, \ldots, S_{c-1}, c)$ distributed as

$$\boldsymbol{\sigma}_{c-1}(\mathbf{S}) = \prod_{r=1}^{c-1} \sigma_r(S_{I_r}|\{S_\ell : \ell \in \Xi_{r-1}^c\}) \tag{21}$$

is obtained. The importance weight of this realized path $\mathbf{S}$ is given by $\boldsymbol{\omega}_c(\mathbf{S}) = \phi(\mathbf{S})/\boldsymbol{\sigma}_{c-1}(\mathbf{S})$. Or, $\mathbf{S}$ is said to be *properly weighted* by a weighting function $\boldsymbol{\omega}_c(\mathbf{S})$ w.r.t. the distribution $\pi(\mathbf{S}) \propto \phi(\mathbf{S})$ (Liu and Chen 1998).

Algorithm 2 (SIPs sampler) For a given value of θ, obtain $\hat{\mathbf{Z}}$ and $\bar{\mathbf{Z}}$ according to (5). Choose random permutations $\mathcal{I}^{\hat{n}-1}$ and $\mathcal{I}^{\bar{n}-1}$ of the integers $\{1, \ldots, \hat{n}-1\}$ and $\{1, \ldots, \bar{n}-1\}$, respectively, and define accordingly $\Xi_r^{\hat{n}}$, $r = 0, 1, \ldots, \hat{n}$, and $\Xi_r^{\bar{n}}$, $r = 0, 1, \ldots, \bar{n}$. An SIS scheme for sampling a random pair of $(\hat{\mathbf{S}}, \bar{\mathbf{S}})$ from $W(\hat{\mathbf{S}}, \bar{\mathbf{S}}|\theta, \mathbf{T}) = \hat{W}(\hat{\mathbf{S}}|\theta, \mathbf{T})\bar{W}(\bar{\mathbf{S}}|\theta, \mathbf{T})$ consists of the following two interchangeable steps:

(S1) Determine $\hat{\mathbf{S}}$ of $\hat{n}+1$ coordinates by applying Algorithm 1 with c and $\phi(\mathbf{S})$ replaced by $\hat{n}$ and $\hat{\phi}_\theta(\hat{\mathbf{S}}, \mathbf{T})$, respectively. Obtain $\hat{\boldsymbol{\sigma}}_{\hat{n}-1}(\hat{\mathbf{S}}|\theta)$ according to (21).

(S2) Determine $\bar{\mathbf{S}}$ of $\bar{n}+1$ coordinates by applying Algorithm 1 with c and $\phi(\mathbf{S})$ replaced by $\bar{n}$ and $\bar{\phi}_\theta(\bar{\mathbf{S}}, \mathbf{T})$, respectively. Obtain $\bar{\boldsymbol{\sigma}}_{\bar{n}-1}(\bar{\mathbf{S}}|\theta)$ according to (21).

Then, the selected $\hat{\mathbf{S}}$ and $\bar{\mathbf{S}}$ are properly weighted by importance weights

$$\hat{\boldsymbol{\omega}}_{\hat{n}}(\hat{\mathbf{S}}|\theta) = \frac{\hat{\phi}_\theta(\hat{\mathbf{S}}, \mathbf{T})}{\hat{\boldsymbol{\sigma}}_{\hat{n}-1}(\hat{\mathbf{S}}|\theta)} \quad \text{and} \quad \bar{\boldsymbol{\omega}}_{\bar{n}}(\bar{\mathbf{S}}|\theta) = \frac{\bar{\phi}_\theta(\bar{\mathbf{S}}, \mathbf{T})}{\bar{\boldsymbol{\sigma}}_{\bar{n}-1}(\bar{\mathbf{S}}|\theta)} \tag{22}$$

w.r.t. $\hat{W}(\hat{\mathbf{S}}|\theta, \mathbf{T})$ and $\bar{W}(\bar{\mathbf{S}}|\theta, \mathbf{T})$, respectively. Replicating the above algorithm M times (preferably based on different random permutations $\mathcal{I}^{\hat{n}-1}$ and $\mathcal{I}^{\bar{n}-1}$) results in two independent sequences of M realizations, $\hat{\mathbf{S}}_{(1)}, \ldots, \hat{\mathbf{S}}_{(M)}$ and $\bar{\mathbf{S}}_{(1)}, \ldots, \bar{\mathbf{S}}_{(M)}$, from $\hat{W}(\hat{\mathbf{S}}|\theta, \mathbf{T})$ and $\bar{W}(\bar{\mathbf{S}}|\theta, \mathbf{T})$, respectively. Expectations of any functional $f(\hat{\mathbf{S}}, \bar{\mathbf{S}})$ w.r.t. the probability distribution $W(\hat{\mathbf{S}}, \bar{\mathbf{S}}|\theta, \mathbf{T})$ can be approximated by a weighted average defined by the M^2 possible pairs of paths from the two sequences. For example, the posterior mean $E[\lambda(t|\mu, \theta)|\theta, \mathbf{T}]$ in (16) can be approximated by

$$\frac{\sum_{i=1}^{M}\sum_{j=1}^{M} a_\lambda(t|\hat{\mathbf{S}}_{(i)}, \bar{\mathbf{S}}_{(j)}, \theta, \mathbf{T})\, \hat{\boldsymbol{\omega}}_{\hat{n}}(\hat{\mathbf{S}}_{(i)}|\theta)\bar{\boldsymbol{\omega}}_{\bar{n}}(\bar{\mathbf{S}}_{(j)}|\theta)}{\sum_{i=1}^{M}\sum_{j=1}^{M} \hat{\boldsymbol{\omega}}_{\hat{n}}(\hat{\mathbf{S}}_{(i)}|\theta)\bar{\boldsymbol{\omega}}_{\bar{n}}(\bar{\mathbf{S}}_{(j)}|\theta)}. \tag{23}$$

3.2 When θ is unknown: the SIPs(θ) sampler

When $\theta \in \mathcal{H}$ is unknown, an SIS scheme generalizing the SIPs sampler (Algorithm 2) for sampling the triplets from $\pi(\hat{\mathbf{S}}, \bar{\mathbf{S}}, d\theta|\mathbf{T})$ in (18), dubbed as SIPs(θ) sampler, is defined by inserting the following step,

(S0) Sample θ according to a density $\rho(\theta) > 0, \theta \in \mathcal{H}$,

before implementing Algorithm 2. Specifically, select $\theta_{(1)}, \ldots, \theta_{(\Theta)}$ by step (S0) for a large number of times $\Theta > 0$, which results in $\hat{n}_{(k)}$ and $\bar{n}_{(k)}, k = 1, \ldots, \Theta$, numbers of positive and negative observations. For each $\theta_{(k)}$, implement Algorithm 2 M number of times to obtain two independent sequences of paths, denoted by $\hat{\mathbf{S}}_{(k,1)}, \ldots, \hat{\mathbf{S}}_{(k,M)}$ and $\bar{\mathbf{S}}_{(k,1)}, \ldots, \bar{\mathbf{S}}_{(k,M)}$, such that they are properly weighted by importance weights, for $i = 1, \ldots, M$,

$$\hat{\boldsymbol{\omega}}_{\hat{n}_{(k)}}(\hat{\mathbf{S}}_{(k,i)}|\theta_{(k)}) = \frac{\hat{\phi}_{\theta_{(k)}}(\hat{\mathbf{S}}_{(k,i)}, \mathbf{T})}{\boldsymbol{\sigma}_{\hat{n}_{(k)}-1}(\hat{\mathbf{S}}_{(k,i)}|\theta_{(k)})}$$

and

$$\bar{\boldsymbol{\omega}}_{\bar{n}_{(k)}}(\bar{\mathbf{S}}_{(k,i)}|\theta_{(k)}) = \frac{\bar{\phi}_{\theta_{(k)}}(\bar{\mathbf{S}}_{(k,i)}, \mathbf{T})}{\boldsymbol{\sigma}_{\bar{n}_{(k)}-1}(\bar{\mathbf{S}}_{(k,i)}|\theta_{(k)})},$$

respectively. As a consequence, a random triplet $(\hat{\mathbf{S}}_{(k,i)}, \bar{\mathbf{S}}_{(k,j)}, \theta_{(k)})$, for $k = 1, \ldots, \Theta$ and $i, j = 1, \ldots, M$, is properly weighted by a weighting function

$$\boldsymbol{\omega}(\hat{\mathbf{S}}_{(k,i)}, \bar{\mathbf{S}}_{(k,j)}, \theta_{(k)}) = \mathcal{L}_\mu(g_{N,\theta_{(k)}}|\rho, \eta)\hat{\boldsymbol{\omega}}_{\hat{n}_{(k)}}(\hat{\mathbf{S}}_{(k,i)}|\theta_{(k)})\bar{\boldsymbol{\omega}}_{\bar{n}_{(k)}}(\bar{\mathbf{S}}_{(k,i)}|\theta_{(k)})\frac{\pi(\theta_{(k)})}{\rho(\theta_{(k)})}$$

w.r.t. $\pi(\hat{\mathbf{S}}, \bar{\mathbf{S}}, d\theta|\mathbf{T})$ if $\pi(\mathrm{d}\theta) = \pi(\theta)\mathrm{d}\theta$. For any function $f(\hat{\mathbf{S}}, \bar{\mathbf{S}}, \theta)$, the posterior expectation

$$E[f(\hat{\mathbf{S}}, \bar{\mathbf{S}}, \theta)|\mathbf{T}] \equiv \int_{\mathcal{H}} \sum_{\hat{\mathbf{S}}} \sum_{\bar{\mathbf{S}}} f(\hat{\mathbf{S}}, \bar{\mathbf{S}}, \theta)\pi(\hat{\mathbf{S}}, \bar{\mathbf{S}}, \mathrm{d}\theta|\mathbf{T})$$

can be approximated by $[\sum_{i,j,k}^{\Theta,M} f(\hat{\mathbf{S}}_{(k,i)}, \bar{\mathbf{S}}_{(k,j)}, \theta_{(k)})\, \boldsymbol{\omega}(\hat{\mathbf{S}}_{(k,i)}, \bar{\mathbf{S}}_{(k,j)}, \theta_{(k)})]/\sum_{i,j,k}^{\Theta,M} \boldsymbol{\omega}(\hat{\mathbf{S}}_{(k,i)}, \bar{\mathbf{S}}_{(k,j)}, \theta_{(k)})$, where $\sum_{i,j,k}^{\Theta,M}$ denotes the triple sum $\sum_{k=1}^{\Theta}\sum_{i=1}^{M}\sum_{j=1}^{M}$. Hence, the posterior probability (17) can be approximated by setting $f(\hat{\mathbf{S}}, \bar{\mathbf{S}}, \theta) = \mathbb{I}(\theta \in B)$, that is,

$$P(\theta \in B|\mathbf{T}) = E[\mathbb{I}(\theta \in B)|\mathbf{T}] \approx \frac{\sum_{i,j,k}^{\Theta,M} \mathbb{I}(\theta_{(i)} \in B)\, \boldsymbol{\omega}(\hat{\mathbf{S}}_{(k,i)}, \bar{\mathbf{S}}_{(k,j)}, \theta_{(k)})}{\sum_{i,j,k}^{\Theta,M} \boldsymbol{\omega}(\hat{\mathbf{S}}_{(k,i)}, \bar{\mathbf{S}}_{(k,j)}, \theta_{(k)})}. \quad (24)$$

Similarly, regarding the Bayes estimate of the BFRs in (1) given by (19), we have

$$E[\lambda(t|\mu, \theta)|\mathbf{T}] \approx \frac{\sum_{i,j,k}^{\Theta,M} a_{\lambda}(t|\hat{\mathbf{S}}_{(k,i)}, \bar{\mathbf{S}}_{(k,j)}, \theta_{(k)}, \mathbf{T})\, \boldsymbol{\omega}(\hat{\mathbf{S}}_{(k,i)}, \bar{\mathbf{S}}_{(k,j)}, \theta_{(k)})}{\sum_{i,j,k}^{\Theta,M} \boldsymbol{\omega}(\hat{\mathbf{S}}_{(k,i)}, \bar{\mathbf{S}}_{(k,j)}, \theta_{(k)})}. \quad (25)$$

4 Simulation studies

This section illustrates the methodology with some numerical examples. For purpose of illustration, μ is selected to be a gamma process with shape measure as a uniform density on $[-2\tau, 2\tau]$, that is, a CRM with Lévy measure

$$\rho(\mathrm{d}x|u)\eta(\mathrm{d}u) = x^{-1}\mathrm{e}^{-x}\mathbb{I}(x > 0)\mathrm{d}x \times \frac{1}{4\tau}\mathbb{I}(-2\tau < u < 2\tau)\mathrm{d}u,$$

as it results in closed and easily manageable expressions for most of the quantities that appear so far. For instance,

$$\kappa_{\ell}(\mathrm{e}^{-f_{N,\theta}}\rho|u) = (\ell - 1)! \times [1 + g_{N,\theta}(u)]^{-\ell}, \quad \ell = 1, 2, \ldots, c, \quad (26)$$

where

$$g_{N,\theta}(u) = \begin{cases} \sum_{k=1}^{i} T_k + (N - i)(u + \theta), \\ \qquad u \in (T_i - \theta, \min(0, T_{i+1} - \theta)], i = 0, 1, \ldots, \hat{n}, \\ (N - c)\tau + \sum_{k=c-(i-1)}^{c} T_k - (N - c + i)(u + \theta), \\ \qquad u \in (\max(0, T_{c-i} - \theta), T_{c-(i-1)} - \theta], i = 0, 1, \ldots, \bar{n}. \end{cases} \quad (27)$$

In turn, a closed-form solution for the integral, for $a > b > 0$ or $b < a < 0$,

$$\int_b^a \kappa_{\ell}(\mathrm{e}^{-f_{N,\theta}}\rho|u)\, \eta(\mathrm{d}u)$$

is available. However, this choice of gamma process here does not mean that our methodology together with the proposed algorithms can only be applied in this special case of μ. It is important to note that $\kappa_\ell(\mathrm{e}^{-f_{N,\theta}\rho}|u)$, defined in (8), in the last integral is the conditional cumulant of an (exponentially tilted) infinitely divisible random variable with density of an available form for any given ρ. Then, it follows that the integral may be calculated using a result of T. N. Thiele, which gives a recursive relation between cumulants and moments of a random variable (see, for instance, Hald 1981 and McCullagh 1987, Section 2.3).

The prior $\pi(\mathrm{d}\theta)$ is chosen to be uniformly distributed on a reasonably large interval on $\mathcal{H}$ to "deflate" the prior belief. Simulated data are generated from two bathtub-shaped life distributions to test the methodology. The life distributions correspond to BFRs given by

$$\lambda_1(t) = \begin{cases} 1, & 0 < t \le 0.5, \\ \mathrm{e}^{-1}, & 0.5 < t \le 3, \\ \mathrm{e}^{-2/3}, & t > 3. \end{cases} \quad \text{and} \quad \lambda_2(t) = \begin{cases} \mathrm{e}^{-2.5t}, & 0 < t \le 1, \\ \mathrm{e}^{-2.5}, & 1 < t \le 5, \\ \mathrm{e}^{-6+0.7t}, & t > 5, \end{cases} \tag{28}$$

respectively. The censoring rates in the data sets governed by the above two hazard rates are about 15 and 20% by setting termination times $\tau = 4$ and $\tau = 8$, respectively. Monte Carlo size $M = 10{,}000$ is chosen for implementations of the proposed SIS methods in all results that follow.

Our attention is to first investigate whether the iterative method and the SIS scheme work well when θ is fixed. The APs sampler discussed in Sect. 3.1.1 and the SIPs sampler (Algorithm 2) are implemented to approximate (16) based on a fixed value of θ, wherein the APs sampler is initialized by paths $\hat{\mathbf{S}}_{(0)}$ and $\bar{\mathbf{S}}_{(0)}$ with coordinates $\hat{S}_i = \bar{S}_i = i$, for all i, to produce totally $M = 1{,}000$ pairs of paths in the sense that samples are taken once in every 5 cycles after a "burn-in" period of 5,000 cycles. As there is a long interval in which either test BFR attains its minimum value, both the algorithms are implemented with three different values of θ to study the effect of different choices of θ on the performance. For fitting $\lambda_1(t)$, θ is fixed at 0.5, 1.75 and 3, whereas for fitting $\lambda_2(t)$, 1, 3 and 5 are selected. In particular, the convergence property of the approximated hazard rate estimates as the total number of observations N increases is studied. Figures 1 and 2 depict ergodic averages (20) produced by the APs sampler with the aforementioned different values of θ based on nested samples of sizes $N = 500$, 1,000 and 3,000 from the life distributions governed by the two BFRs, respectively. Corresponding weighted average estimates (23) produced by the SIPs sampler are graphed in the first three rows of Figs. 3 and 4.

To investigate the performance of the SIPs(θ) sampler when θ is not known, we set $\rho(\theta)$ to be uniform on an interval which includes all the complete observations and draw $\Theta = 1{,}000$ realizations of θ. For each value of θ, independent random samples of $(\hat{\mathbf{S}}, \bar{\mathbf{S}})$ of size $M = 10{,}000$ are generated by implementing the SIPs sampler as discussed before. For the sake of a better comparison between results by the SIPs(θ) sampler based on an unknown θ and those by the SIPs sampler with a fixed θ, the resulting Bayes estimates of the two test BFRs, given by the weighted average (25), are presented in the last rows of Figs. 3 and 4, respectively.

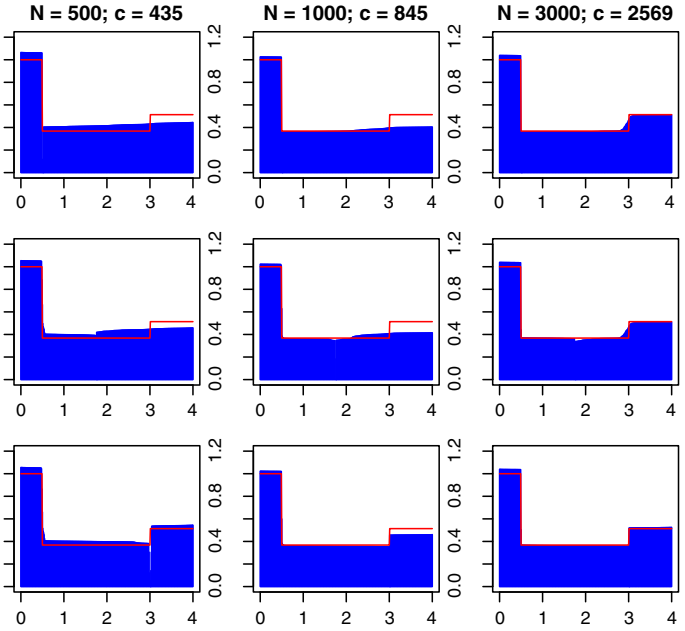

Fig. 1 The true bathtub-shaped hazard rate $\lambda_1(t)$ (*solid line*) given in (28) and the Bayes estimates produced by the APs sampler based on total number of observations, $N = 500$ (*left column*), 1,000 (*middle column*) and 3,000 (*right column*), with $\theta = 0.5$, 1.75 and 3 (*from top row to bottom row*)

In summary, the graphs echo the fact that approximations for Bayes estimates of the BFRs in (1) by all the proposed algorithms tend to the "true" hazard rates, $\lambda_1(t)$ and $\lambda_2(t)$, as sample size increases. We remark that some other simulations we have carried out applying the APs and the SIPs samplers based on fixed values of θ other than those stated above reveal that there is not much difference between simulation results based on different values of θ. It is important to point out that the large-sample simulation study carried in Ho (2006a) which compares monotone hazard estimates produced by the AP sampler based on sampling **S**-paths and a Gibbs sampler for sampling partitions defined in Lo et al. (1996) applies to this situation and suggests that the BFR estimates produced by the proposed APs sampler are always less variable than those produced by any other numerical methods based on sampling partitions, for instance, by Lo et al. (1996), Ishwaran and James (2003) and James (2005).

5 A test of an MFR versus an BFR

Early references devoted to testing for a constant hazard rate versus an MFR include Proschan and Pyke (1967), Bickel and Doksum (1969) and Gail and Gastwirth (1978a,b), among others. Without relying on exponentiality assumption, Gijbels and Heckman (2004) develop a testing procedure via normalized spacings for testing an MFR against alternatives of some local departures. For testing an MFR versus other general alternatives, Hall and Van Keilegom (2005) propose a calibration method

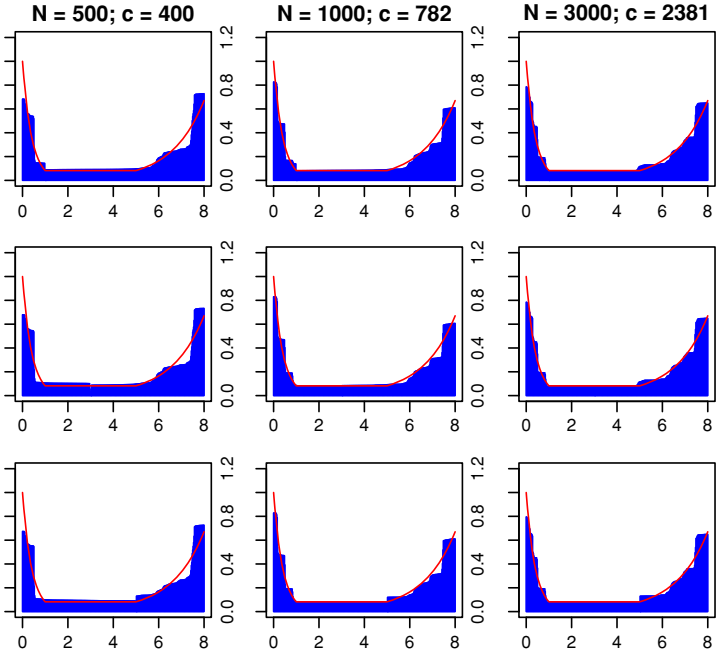

Fig. 2 The true bathtub-shaped hazard rate $\lambda_2(t)$ (*solid line*) given in (28) and the Bayes estimates produced by the APs sampler based on total number of observations, $N = 500$ (*left column*), 1,000 (*middle column*) and 3,000 (*right column*), with $\theta = 1, 3$ and 5 (*from top row to bottom row*)

related to the "increasing bandwidth" approach suggested by Silverman (1981) in the case of density estimation. Testing procedures involving BFRs can be found in, for example, Aarset (1985), who discussed the test statistic proposed by Bergman (1979) for testing a constant hazard rate against an BFR, and Vaurio (1999), who proposed a few test statistics for testing between an MFR and other non-monotone alternatives including BFRs.

A Bayesian test of monotone versus bathtub-shaped hazard rates can be readily defined in terms of a test between different values of θ based on the models in (1) with μ being a nuisance parameter as follows: Suppose that we are interested in testing whether a set of observations $\mathbf{T}$, defined as in Sect. 2, is generated according to a non-decreasing hazard rate or an BFR. Based on (1), it is equivalent to choose between two hypotheses $H_0 : \theta = 0$ and $H_1 : \theta \in (0, \infty)$ as when $\theta = 0$, models in (1) correspond to a class of non-decreasing hazard rates; otherwise, they give a class of BFRs with a change point $\theta > 0$. In particular, the likelihood of the data given (μ, θ) under H_1 is given by (2) when $\theta \neq 0$ or ∞, while the likelihood of the data given μ_0, which is another CRM distributed same as μ and independent of μ, under H_0 follows from (2) with $\theta = 0$ as

$$\mathrm{e}^{-\mu(g_{N,0})} \prod_{k=1}^{c} \int \mathbb{I}(0 < u_k \leq T_k) \mu(\mathrm{d}u_k). \tag{29}$$

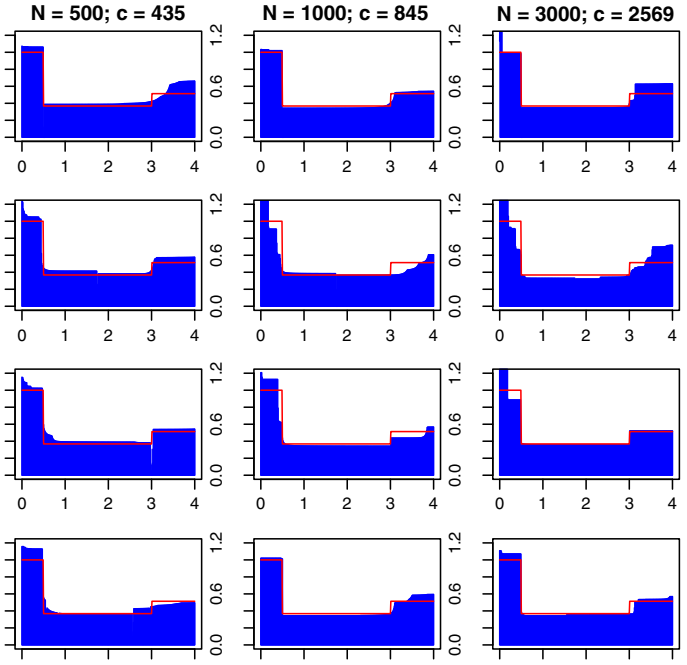

Fig. 3 The true bathtub-shaped hazard rate $\lambda_1(t)$ (*solid line*) given in (28) and the Bayes estimates produced by the SIS methods based on total number of observations, $N = 500$ (*left column*), 1,000 (*middle column*) and 3,000 (*right column*), wherein estimates in the *first three rows from top to bottom* are obtained by the SIPs sampler (Algorithm 2) with $\theta = 0.5, 1.75$ and 3, respectively, and those in the *last row* are obtained by the SIPs(θ) sampler with an unknown θ

Corollary 3 *Suppose μ is a CRM characterized by (3). It follows from Theorem 2 that the likelihood of the data* $\mathbf{T}$ *given θ is proportional to*

$$m_\theta(\mathbf{T}) = \mathcal{L}_\mu(g_{N,\theta}|\rho, \eta) \times \sum_{\hat{\mathbf{S}}} \hat{\phi}_\theta(\hat{\mathbf{S}}, \mathbf{T}) \times \sum_{\bar{\mathbf{S}}} \bar{\phi}_\theta(\bar{\mathbf{S}}, \mathbf{T}). \tag{30}$$

Let p_0 denote the prior probability of H_0, and then $1 - p_0$ denotes the prior probability of H_1; furthermore, suppose that θ has a prior $\pi(\mathrm{d}\theta)$ restricted on H_1. Then, the marginal density of $\mathbf{T}$ is given by

$$m(\mathbf{T}) = p_0 \times \mathcal{L}_\mu(g_{N,0}|\rho, \eta) \sum_{\mathbf{S}} \bar{\phi}_0(\mathbf{S}, \mathbf{T}) + (1 - p_0) \times \int_{\mathcal{H}} m_\theta(\mathbf{T}) \pi(\mathrm{d}\theta), \tag{31}$$

where $\mathbf{S}$ is a path of $c + 1$ coordinates defined as in Definition 1. It implies that the posterior probability of H_0 is given by

$$P(H_0|\mathbf{T}) = \frac{p_0 \times \mathcal{L}_\mu(g_{N,0}|\rho, \eta) \sum_{\mathbf{S}} \bar{\phi}_0(\mathbf{S}, \mathbf{T})}{m(\mathbf{T})},$$

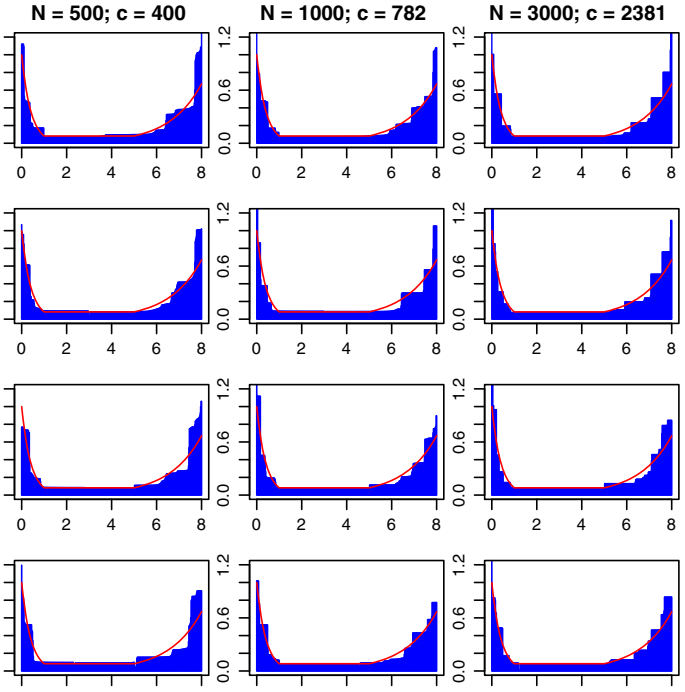

Fig. 4 The true bathtub-shaped hazard rate $\lambda_2(t)$ (*solid line*) given in (28) and the Bayes estimates produced by the SIS methods based on total number of observations, $N = 500$ (*left column*), 1,000 (*middle column*) and 3,000 (*right column*), wherein estimates in the *first three rows from top to bottom* are obtained by the SIPs sampler (Algorithm 2) with $\theta = 1$, 3 and 5, respectively, and those in the *last row* are obtained by the SIPs(θ) sampler with an unknown θ

and that of H_1 is equal to $1 - P(H_0|\mathbf{T})$. Also of interest is the posterior odds of H_0 to H_1, which is given by

$$\frac{p_0}{1-p_0} \times \frac{\mathcal{L}_\mu(g_{N,0}|\rho,\eta)\sum_{\mathbf{S}}\bar{\phi}_0(\mathbf{S},\mathbf{T})}{\int_{\mathcal{H}} m_\theta(\mathbf{T})\pi(\mathrm{d}\theta)},$$

wherein $p_0/(1-p_0)$ is the prior odds and the latter ratio is the Bayes factor for H_0 versus H_1 (see Kass and Raftery 1995 for a review of Bayes factors).

Implementation of the above Bayesian test can be carried out by applying Algorithm 1 and the SIPs(θ) sampler. On one hand, the sum $\sum_{\mathbf{S}}\bar{\phi}_0(\mathbf{S},\mathbf{T})$ can be approximated by

$$\frac{1}{M}\sum_{i=1}^{M}\frac{\bar{\phi}_0(\mathbf{S}_{(i)},\mathbf{T})}{\boldsymbol{\sigma}_{c-1}(\mathbf{S}_{(i)})},$$

if $\mathbf{S}_{(0)}, \mathbf{S}_{(1)}, \ldots, \mathbf{S}_{(M)}$ are independent samples obtained via implementing Algorithm 1 with $\phi(\mathbf{S}) = \bar{\phi}_0(\mathbf{S},\mathbf{T})$ in (11). On the other hand, the integral $\int_{\mathcal{H}} m_\theta(\mathbf{T})\pi(\mathrm{d}\theta)$ is approximated by

$$\frac{1}{\Theta M^2}\sum_{i,j,k}^{\Theta,M}\boldsymbol{\omega}(\hat{\mathbf{S}}_{(k,i)},\bar{\mathbf{S}}_{(k,j)},\theta_{(k)}),$$

given random samples $(\hat{\mathbf{S}}_{(k,i)},\bar{\mathbf{S}}_{(k,j)},\theta_{(k)})$, for $k=1,\ldots,\Theta$ and $i,j=1,\ldots,M$, obtained via implementation of the SIPs(θ) sampler as described in Sect. 3.2.

6 Proportional hazards

The Cox regression model (Cox 1972) is an important example of the multiplicative intensity model that can allow incorporation of covariates, together with right independent censoring, in survival analysis. For Bayes inference of general hazard rates with presence of covariates, see Kalbfleisch (1978), Ibrahim et al. (1999), James (2003) and Ishwaran and James (2004), among others. Suppose we collect failure data until time τ, which are governed by an underlying hazard rate on $\mathcal{H}$ associated with a p-dimensional covariate vector $\mathbf{X}\in\mathcal{R}^p$,

$$\lambda(t|\mathbf{X},\boldsymbol{\beta},\mu,\theta)=\lambda(t|\mu,\theta)\exp(\boldsymbol{\beta}^\top\mathbf{X}),$$

where $\lambda(t|\mu,\theta)$ defined in (1) is an unknown baseline hazard rate of a bathtub shape and $\boldsymbol{\beta}\in\mathcal{R}^p$ is an unknown parameter vector. The data $\mathbf{D}=((T_1,\mathbf{X}_1),\ldots,(T_N,\mathbf{X}_N))$ summarize completely observed failure times $T_1<\cdots<T_c$ and right-censored times $T_i=\tau$, $i=c+1,\ldots,N$, associated with covariate vector $\mathbf{X}_i$, $i=1,\ldots,N$, respectively. Define $f_{N,\boldsymbol{\beta},\theta}(x,u)=g_{N,\boldsymbol{\beta},\theta}(u)x$, for any $(x,u)\in(\mathcal{H},\mathcal{R})$, where

$$g_{N,\boldsymbol{\beta},\theta}(u)=\int_0^\tau\left[\sum_{i=1}^N\mathbb{I}(T_i\geq t)\exp\left(\boldsymbol{\beta}^\top\mathbf{X}_i\right)\right]K_\theta(t,u)\mathrm{d}t. \tag{32}$$

Then, the Cox proportional hazards likelihood may be written as

$$\left[\prod_{i=1}^c\exp\left(\boldsymbol{\beta}^\top\mathbf{X}_i\right)\lambda(T_i|\mu,\theta)\right]\exp\left[-\mu(g_{N,\boldsymbol{\beta},\theta})\right], \tag{33}$$

where $\mu(g_{N,\boldsymbol{\beta},\theta})=\int_{\mathcal{R}}g_{N,\boldsymbol{\beta},\theta}(u)\mu(\mathrm{d}u)=\int_0^\tau[\sum_{i=1}^N\mathbb{I}(T_i\geq t)\exp(\boldsymbol{\beta}^\top\mathbf{X}_i)]\lambda(t|\mu,\theta)\mathrm{d}t$. Assume $\int_{\mathcal{R}}x^\ell\mathrm{e}^{-g_{N,\boldsymbol{\beta},\theta}(u)x}\rho(\mathrm{d}x|u)<\infty$, for $\ell=1,\ldots,c$ and a fixed $u>0$. Suppose that $\boldsymbol{\beta}$ are fixed. Applying the same arguments in showing Theorem 1 and Corollaries 1 and 2 for model (1) leads to the following analogous results under this Cox model (33) by noticing that the resulting likelihood given by (33) without the leading term $\prod_{i=1}^c\exp(\boldsymbol{\beta}^\top\mathbf{X}_i)$ takes the same form as (2).

Proposition 1 *Suppose that the likelihood of the data is given by* (33) *with* μ *being a CRM characterized by* (3)*. When* $\boldsymbol{\beta}$ *is given,*

(i) *for a fixed value of* θ*, the law of* μ *given* **D** *is equivalent to that of a random measure*

$$\mu_{g_{N,\boldsymbol{\beta},\theta}} + \sum_{\{j|\hat{\mathbf{S}}\}} \hat{Q}_j \delta_{\hat{y}_j} + \sum_{\{j|\bar{\mathbf{S}}\}} \bar{Q}_j \delta_{\bar{y}_j},$$

mixed over by the joint law of $(\hat{\mathbf{y}}, \hat{\mathbf{Q}}, \hat{\mathbf{S}})$ *and* $(\bar{\mathbf{y}}, \bar{\mathbf{Q}}, \bar{\mathbf{S}})$ *given* **D**, *wherein* $\mu_{g_{N,\boldsymbol{\beta},\theta}}$, *conditionally independent of* $(\hat{\mathbf{y}}, \hat{\mathbf{Q}}, \hat{\mathbf{S}})$ *and* $(\bar{\mathbf{y}}, \bar{\mathbf{Q}}, \bar{\mathbf{S}})$, *is a completely random measure with Lévy measure* $\mathrm{e}^{-g_{N,\boldsymbol{\beta},\theta}(u)x}\rho(\mathrm{d}x|u)\eta(\mathrm{d}u)$ *and the mixing distribution of* $(\hat{\mathbf{y}}, \hat{\mathbf{Q}}, \hat{\mathbf{S}})$ *and* $(\bar{\mathbf{y}}, \bar{\mathbf{Q}}, \bar{\mathbf{S}})$ *given* **D** *is described as in Theorem 1 with* $f_{N,\theta}(\cdot,\cdot)$ *and* $g_{N,\theta}(\cdot)$ *replaced by* $f_{N,\boldsymbol{\beta},\theta}(\cdot,\cdot)$ *and* $g_{N,\boldsymbol{\beta},\theta}(\cdot)$, *respectively; and*

(ii) *if* θ *has a prior* $\pi(\mathrm{d}\theta)$, *independent of* μ, *the law of* θ *given* **D** *is characterized by*

$$P(\theta \in B|\boldsymbol{\beta}, \mathbf{D}) = \int_B \sum_{\hat{\mathbf{S}}} \sum_{\bar{\mathbf{S}}} \pi(\hat{\mathbf{S}}, \bar{\mathbf{S}}, \mathrm{d}\theta|\boldsymbol{\beta}, \mathbf{D})$$

for any Borel set $B \in \mathcal{H}$, *where* $\pi(\hat{\mathbf{S}}, \bar{\mathbf{S}}, \mathrm{d}\theta|\boldsymbol{\beta}, \mathbf{D}) \propto \mathcal{L}_\mu(g_{N,\boldsymbol{\beta},\theta}|\rho, \eta) \times \pi(\mathrm{d}\theta) \times |\mathbb{C}_{\hat{\mathbf{S}}}| \big[\prod_{\{j|\hat{\mathbf{S}}\}} \int_{\hat{Z}_j}^0 \kappa_{\hat{m}_j}(\mathrm{e}^{-f_{N,\boldsymbol{\beta},\theta}\rho|y})\eta(\mathrm{d}y) \big] \times |\mathbb{C}_{\bar{\mathbf{S}}}| \big[\prod_{\{j|\bar{\mathbf{S}}\}} \int_0^{\bar{Z}_j} \kappa_{\bar{m}_j}(\mathrm{e}^{-f_{N,\boldsymbol{\beta},\theta}\rho|y}) \eta(\mathrm{d}y) \big]$.

To evaluate any posterior quantities of model (33), such as the posterior mean of the underlying bathtub-shaped baseline hazard rate and the posterior mean of the covariate parameters $\boldsymbol{\beta}$, run the following Gibbs sampler to obtain random samples from the posterior distribution of $((\hat{\mathbf{y}}, \hat{\mathbf{Q}}, \hat{\mathbf{S}}), (\bar{\mathbf{y}}, \bar{\mathbf{Q}}, \bar{\mathbf{S}}), \theta, \boldsymbol{\beta})$ given the data **D**:

1. Draw $(\hat{\mathbf{S}}, \bar{\mathbf{S}})$ conditional on $(\theta, \boldsymbol{\beta}, \mathbf{D})$ by implementing the AP sampler independently as in steps (G1) and (G2) in Sect. 3.1.1.
2. Draw $(\hat{\mathbf{y}}, \hat{\mathbf{Q}})$ and $(\bar{\mathbf{y}}, \bar{\mathbf{Q}})$ independently conditional on $(\hat{\mathbf{S}}, \bar{\mathbf{S}}, \theta, \boldsymbol{\beta}, \mathbf{D})$ according to the analogues of the conditional distributions (12–14) in Theorem 1 with $f_{N,\theta}(\cdot,\cdot)$ and $g_{N,\theta}(\cdot)$ replaced by $f_{N,\boldsymbol{\beta},\theta}(\cdot,\cdot)$ and $g_{N,\boldsymbol{\beta},\theta}(\cdot)$, respectively.
3. Draw θ conditional on $((\hat{\mathbf{y}}, \hat{\mathbf{Q}}, \hat{\mathbf{S}}), (\bar{\mathbf{y}}, \bar{\mathbf{Q}}, \bar{\mathbf{S}}), \boldsymbol{\beta}, \mathbf{D})$ from the density proportional to

$$\pi(\mathrm{d}\theta)\mathcal{L}_\mu(g_{N,\boldsymbol{\beta},\theta}|\rho, \eta) \left[\prod_{\{j|\hat{\mathbf{S}}\}} \mathrm{e}^{-g_{N,\boldsymbol{\beta},\theta}(\hat{y}_j)\hat{Q}_j} \mathbb{I}(\hat{Z}_j - \theta \le \hat{y}_j) \right] \times \left[\prod_{\{j|\bar{\mathbf{S}}\}} \mathrm{e}^{-g_{N,\boldsymbol{\beta},\theta}(\bar{y}_j)\bar{Q}_j} \mathbb{I}(\bar{y}_j \le \bar{Z}_j - \theta) \right].$$

4. Draw $\boldsymbol{\beta}$ conditional on $((\hat{\mathbf{y}}, \hat{\mathbf{Q}}, \hat{\mathbf{S}}), (\bar{\mathbf{y}}, \bar{\mathbf{Q}}, \bar{\mathbf{S}}), \theta, \mathbf{D})$ from the density proportional to

$$\pi(\mathrm{d}\boldsymbol{\beta})\mathcal{L}_{\mu}(g_{N,\boldsymbol{\beta},\theta}|\rho,\eta)\prod_{i=1}^{c}\exp(\boldsymbol{\beta}^{\top}\mathbf{X}_i)\prod_{\{j|\hat{\mathbf{S}}\}}\mathrm{e}^{-g_{N,\boldsymbol{\beta},\theta}(\hat{y}_j)\hat{Q}_j}\prod_{\{j|\bar{\mathbf{S}}\}}\mathrm{e}^{-g_{N,\boldsymbol{\beta},\theta}(\bar{y}_j)\bar{Q}_j},$$

when $\boldsymbol{\beta}$ has a prior $\pi(\mathrm{d}\boldsymbol{\beta})$.

Note that $g_{N,\boldsymbol{\beta},\theta}(u)$ is again a piecewise linear function of u as $g_{N,\theta}(u)$ given by (27) in the case without covariates. This does not create any complexities in evaluating integrals at steps 1 and 2 of the above Gibbs sampler; see discussion of Remark 5.1 in Ho (2006a). Step 4 above, which is of the same form as the step 4 (for conditional draws of regression parameters $\boldsymbol{\beta}$) of the Blocked Gibbs algorithm suggested by Ishwaran and James (2004, pp. 184), can be dealt with via a Metropolis step, while step 3 can also be done similarly as the density looks like the one in step 4.

7 Appendix

7.1 Proof of Theorem 1

When θ is given, Theorem 4.1 in James (2005) specializes and summarizes the law of $(\mathbf{u}, \mathbf{x})|\theta, \mathbf{T}$ as a joint distribution of a partition $\mathbf{p}$ and the distinct pairs among $(\mathbf{u}, \mathbf{x})$, which is proportional to

$$\prod_{i=1}^{n(\mathbf{p})}\left[J_i^{e_i}\mathrm{e}^{-g_{N,\theta}(v_i)J_i}\rho(\mathrm{d}J_i|v_i)\prod_{k\in C_i}K_{\theta}(T_k,v_i)\eta(\mathrm{d}v_i)\right], \tag{34}$$

where $\mathbf{v} = (v_1, \ldots, v_{n(\mathbf{p})})$ and $\mathbf{J} = (J_1, \ldots, J_{n(\mathbf{p})})$ denote the distinct pairs among $(\mathbf{u}, \mathbf{x})$, respectively. Re-write $\mathbf{T}$ as $\hat{\mathbf{Z}}$ and $\bar{\mathbf{Z}}$ as defined in (5) and observe that Lemma 1 implies that the $\hat{n}$-negative observations $\hat{\mathbf{Z}}$ can cluster only with one another but not with any of the positive observations $\bar{\mathbf{Z}}$, or vice versa. It is eligible to "split" $\mathbf{p}$ into two non-overlapping partitions $\hat{\mathbf{p}}$ and $\bar{\mathbf{p}}$. Write $\mathbf{p} = \hat{\mathbf{p}} \cup \bar{\mathbf{p}}$. Without loss of generality, let $\hat{\mathbf{p}} = \{C_1, \ldots, C_{n(\hat{\mathbf{p}})}\}$ and $\bar{\mathbf{p}} = \{C_{n(\hat{\mathbf{p}})+1}, \ldots, C_{n(\mathbf{p})}\}$ denote the partition of the $\hat{n}$-negative observations $\hat{\mathbf{Z}}$ and that of the remaining $\bar{n}$-positive observations $\bar{\mathbf{Z}}$ in relation to negative and positive distinct values in $\mathbf{v}$, respectively. Hence, (34) is alternatively expressible as

$$\begin{aligned}&\prod_{i=1}^{n(\hat{\mathbf{p}})}\left[J_i^{e_i}\mathrm{e}^{-g_{N,\theta}(v_i)J_i}\rho(\mathrm{d}J_i|v_i)\mathbb{I}(\hat{Z}_{\max_{k\in C_i}k}\le v_i<0)\eta(\mathrm{d}v_i)\right]\\ &\quad\times\prod_{i=n(\hat{\mathbf{p}})+1}^{n(\mathbf{p})}\left[J_i^{e_i}\mathrm{e}^{-g_{N,\theta}(v_i)J_i}\rho(\mathrm{d}J_i|v_i)\mathbb{I}(0<v_i\le\bar{Z}_{\max_{k\in C_i}k})\eta(\mathrm{d}v_i)\right],\end{aligned} \tag{35}$$

depending on only (i) the number of elements and (ii) the maximal index but not the remaining indices of each cluster in both $\hat{\mathbf{p}}$ and $\bar{\mathbf{p}}$. Note that each of the two product terms above takes the same form as expression (26) without the leading term $\mathbb{P}(\mathrm{d}\mu_N^*)$ in Ho (2006a), implying that the sufficient information in defining the law of $(\mathbf{v}, \mathbf{J}), \mathbf{p}|\theta, \mathbf{T}$ can be summarized by the intrinsic characteristics of two paths $\hat{\mathbf{S}}$ and $\bar{\mathbf{S}}$ of respectively $\hat{n}+1$ and $\bar{n}+1$ coordinates. By virtue of the proof of Theorem 2.1 in Ho (2006a), for any $\mathbf{p} = \hat{\mathbf{p}} \cup \bar{\mathbf{p}}$ such that $\hat{\mathbf{p}} \in \mathbb{C}_{\hat{\mathbf{S}}}$ and $\bar{\mathbf{p}} \in \mathbb{C}_{\bar{\mathbf{S}}}$, (34) or (35) can be written as

$$\prod_{\{j|\hat{\mathbf{S}}\}} \left[(\hat{Q}_j)^{\hat{m}_j} \mathrm{e}^{-g_{N,\theta}(\hat{y}_j)\hat{Q}_j} \rho(\mathrm{d}\hat{Q}_j|\hat{y}_j) \mathbb{I}(\hat{Z}_j \le \hat{y}_j < 0) \eta(\mathrm{d}\hat{y}_j) \right] \times \prod_{\{j|\bar{\mathbf{S}}\}} \left[(\bar{Q}_j)^{\bar{m}_j} \mathrm{e}^{-g_{N,\theta}(\bar{y}_j)\bar{Q}_j} \rho(\mathrm{d}\bar{Q}_j|\bar{y}_j) \mathbb{I}(0 < \bar{y}_j \le \bar{Z}_j) \eta(\mathrm{d}\bar{y}_j) \right] \tag{36}$$

by observing two analogues of property ($\star$) in Ho (2006a) that

$$\prod_{i=1}^{n(\hat{\mathbf{p}})} \mathbb{I}(\hat{Z}_{\max_{k\in C_i} k} \le v_i < 0) = \prod_{\{j|\hat{\mathbf{S}}\}} \mathbb{I}(\hat{Z}_j \le \hat{y}_j < 0) \tag{37}$$

and

$$\prod_{i=n(\hat{\mathbf{p}})+1}^{n(\mathbf{p})} \mathbb{I}(0 < v_i \le \bar{Z}_{\max_{k\in C_i} k}) = \prod_{\{j|\bar{\mathbf{S}}\}} \mathbb{I}(0 < \bar{y}_j \le \bar{Z}_j), \tag{38}$$

where accordingly $(\hat{\mathbf{y}}, \hat{\mathbf{Q}})$ and $(\bar{\mathbf{y}}, \bar{\mathbf{Q}})$ are permutations of $\{(v_1, J_1), \ldots, (v_{n(\hat{\mathbf{p}})}, J_{n(\hat{\mathbf{p}})})\}$ and $\{(v_{n(\hat{\mathbf{p}})+1}, J_{n(\hat{\mathbf{p}})+1}), \ldots, (v_{n(\mathbf{p})}, J_{n(\mathbf{p})})\}$, respectively. Then, summing over all $\hat{\mathbf{p}} \in \mathbb{C}_{\hat{\mathbf{S}}}$ and $\bar{\mathbf{p}} \in \mathbb{C}_{\bar{\mathbf{S}}}$ in (36) yields that the distribution of $(\hat{\mathbf{y}}, \hat{\mathbf{Q}}, \hat{\mathbf{S}}), (\bar{\mathbf{y}}, \bar{\mathbf{Q}}, \bar{\mathbf{S}})|\theta, \mathbf{T}$ is proportional to $|\mathbb{C}_{\hat{\mathbf{S}}}| \times |\mathbb{C}_{\bar{\mathbf{S}}}|$ multiplies the expression in (36) as a result of constancy of all the summands for $\hat{\mathbf{p}} \in \mathbb{C}_{\hat{\mathbf{S}}}$ and $\bar{\mathbf{p}} \in \mathbb{C}_{\bar{\mathbf{S}}}$, where $|\mathbb{C}_{\hat{\mathbf{S}}}|$ and $|\mathbb{C}_{\bar{\mathbf{S}}}|$ are the total numbers of partitions $\hat{\mathbf{p}}$ and $\bar{\mathbf{p}}$ corresponding to $\hat{\mathbf{S}}$ and $\bar{\mathbf{S}}$, respectively. Now, the laws given by (10–14), together with conditional independence between $(\hat{\mathbf{y}}, \hat{\mathbf{Q}}, \hat{\mathbf{S}})$ and $(\bar{\mathbf{y}}, \bar{\mathbf{Q}}, \bar{\mathbf{S}})$, follow from Bayes' theorem and multiplication rule. □

7.2 Rao–Blackwellization

The following analogue of Lemma 2.1 in Ho (2006a) for MFRs defined as in (1) with $\theta = 0$ or ∞ can be established by dividing (34) by the distribution of $(\hat{\mathbf{y}}, \hat{\mathbf{Q}}, \hat{\mathbf{S}}), (\bar{\mathbf{y}}, \bar{\mathbf{Q}}, \bar{\mathbf{S}})|\theta, \mathbf{T}$.

Corollary 4 *Suppose* $\hat{\mathbf{S}}, \bar{\mathbf{S}}|\theta, \mathbf{T} \sim W(\hat{\mathbf{S}}, \bar{\mathbf{S}}|\theta, \mathbf{T})$. *Then, there exists a conditional distribution* $\pi(\mathbf{p}|\hat{\mathbf{S}}, \bar{\mathbf{S}}, \theta, \mathbf{T}) = 1/(|\mathbb{C}_{\hat{\mathbf{S}}}||\mathbb{C}_{\bar{\mathbf{S}}}|)$ *for* $\mathbf{p} = \hat{\mathbf{p}} \cup \bar{\mathbf{p}}$, *defined in the proof of Theorem 1, such that* $\hat{\mathbf{p}} \in \mathbb{C}_{\hat{\mathbf{S}}}$ *and* $\bar{\mathbf{p}} \in \mathbb{C}_{\bar{\mathbf{S}}}$, *where* $|\mathbb{C}_{\mathbf{S}}|$ *is given by* (7).

Note that the posterior mean of the BFR (1) with a given value of θ in terms of partitions, followed from James (2005), is given by $\sum_{\mathbf{p}} b_\lambda(t|\mathbf{p}, \theta, \mathbf{T})\pi(\mathbf{p}|\theta, \mathbf{T})$, where, with $\mathbf{p} = \hat{\mathbf{p}} \cup \bar{\mathbf{p}}$, $b_\lambda(t|\mathbf{p}, \theta, \mathbf{T}) \equiv b_\lambda(t|\hat{\mathbf{p}}, \hat{\mathbf{p}}, \theta, \mathbf{T})$ equals

$$\left[\int_{t-\theta}^{0}\kappa_1(\mathrm{e}^{-f_N}\rho|y)\eta(\mathrm{d}y)+\sum_{i=1}^{n(\hat{\mathbf{p}})}\frac{\int_{\max(t-\theta,\hat{Z}_{\max_{k\in C_i}k})}^{0}\kappa_{e_j+1}(\mathrm{e}^{-f_{N,\theta}}\rho|y)\eta(\mathrm{d}y)}{\int_{\hat{Z}_{\max_{k\in C_i}k}}^{0}\kappa_{e_j}(\mathrm{e}^{-f_{N,\theta}}\rho|y)\eta(\mathrm{d}y)}\right]\mathbb{I}(t<\theta)$$
$$+\left[\int_{0}^{t-\theta}\kappa_1(\mathrm{e}^{-f_N}\rho|y)\eta(dy)+\sum_{i=n(\hat{\mathbf{p}})+1}^{n(\hat{\mathbf{p}})+n(\bar{\mathbf{p}})}\frac{\int_{0}^{\min(t-\theta,\bar{Z}_{\max_{k\in C_i}k})}\kappa_{e_j+1}(\mathrm{e}^{-f_{N,\theta}}\rho|y)\eta(\mathrm{d}y)}{\int_{0}^{\bar{Z}_{\max_{k\in C_i}k}}\kappa_{e_j}(\mathrm{e}^{-f_{N,\theta}}\rho|y)\eta(\mathrm{d}y)}\right]\mathbb{I}(t>\theta),$$

and $\pi(\mathbf{p}|\theta,\mathbf{T})$ is given by marginalizing $(\mathbf{v},\mathbf{J})$ in (34). By virtue of (37) and (38), one can show that the path-sum estimator (16) is identical to this partition counterpart in the sense that

$$\begin{aligned}a_\lambda(t|\hat{\mathbf{S}},\bar{\mathbf{S}},\theta,\mathbf{T})&=E[b_\lambda(t|\mathbf{p},\theta,\mathbf{T})|\hat{\mathbf{S}},\bar{\mathbf{S}},\theta,\mathbf{T}]\\&=\sum_{\hat{\mathbf{p}}\in\mathbb{C}_{\hat{\mathbf{S}}}}\sum_{\bar{\mathbf{p}}\in\mathbb{C}_{\bar{\mathbf{S}}}}b_\lambda(t|\hat{\mathbf{p}},\bar{\mathbf{p}},\theta,\mathbf{T})\pi(\mathbf{p}|\hat{\mathbf{S}},\bar{\mathbf{S}},\theta,\mathbf{T})\end{aligned}$$

defined in Corollary 2 is a Rao–Blackwell improvement of $b_\lambda(t|\mathbf{p},\theta,\mathbf{T})$ and, hence, it has a smaller variability than the partition counterpart by the Rao–Blackwell theorem; see similar results in density estimation problems in Lo et al. (1996) and Ishwaran and James (2003) for discussion of Rao–Blackwell improvements by using partitions instead of latent variables in Dirichlet process and species sampling mixture models.

7.3 Proof of Theorem 2

Applying Proposition 2.1 in James (2005) and following the same argument as in proving Theorem 1 yield a joint distribution of $(\mathbf{v},\mathbf{J},\mathbf{p},\theta)$ given $\mathbf{T}$, which is proportional to a product of $\mathcal{L}_\mu(g_{N,\theta}|\rho,\eta)\pi(\mathrm{d}\theta)$ and the expression in (34) or (35). Integrating out $(\mathbf{v},\mathbf{J})$, which is equivalent to integrating out $(\hat{\mathbf{y}},\bar{\mathbf{y}},\hat{\mathbf{Q}},\bar{\mathbf{Q}})$ in (36), gives a joint distribution of $(\hat{\mathbf{S}},\bar{\mathbf{S}},\theta)$ given $\mathbf{T}$ as in (18). Result follows from further marginalizing $\hat{\mathbf{S}}$ and $\bar{\mathbf{S}}$. □

Acknowledgments This work was supported in part by National University of Singapore research grant R-155-050-067-101 and R-155-050-067-133. Study of this problem was suggested by Albert Lo. The author has greatly benefited from discussions with him in preparing the manuscript. Thanks are also due to the referees for their helpful suggestions.

References

Aalen, O. (1975). Statistical inference for a family of counting processes. Ph.D. dissertation. University of California, Berkeley.

Aalen, O. (1978). Nonparametric inference for a family of counting processes. *The Annals of Statistics, 6*, 701–726.

Aarset, M. V. (1985). The null distribution of a test of constant versus bathtub-failure rate. *Scandinavian Journal of Statistics, 12*, 55–62.

Amman, L. (1984). Bayesian nonparametric inference for quantal response data. *The Annals of Statistics, 12*, 636–645.

Barlow, R. E., Bartholomew, D. J., Bremner, J. M., Brunk, H. D. (1972). *Statistical inference under order restrictions*. New York: Wiley.

Barron, A., Schervish, M. J., Wasserman, L. (1999). The consistency of posterior distributions in nonparametric problems. *The Annals of Statistics, 27*, 536–561.

Bergman, B. (1979). On age replacement and the total time on test concept. *Scandinavian Journal of Statistics, 6*, 161–168.

Bickel, P. J., Doksum, K. A. (1969). Tests for monotone failure rate based on normalized spacings. *The Annals of Mathematical Statistics, 40*, 1216–1235.

Bray, T., Crawford, G., Proschan, F. (1967a). Maximum likelihood estimation of a failure rate function initially increasing and then decreasing, In *Proceedings 1967 Spring Seminar on Reliability, Maintainability, IEEE*.

Bray, T., Crawford, G., Proschan, F. (1967b). Maximum likelihood estimation of a U-shaped failure rate function. *Mathematical Note* 534. Mathematics Research Laboratory, Boeing Scientific Research Laboratories, Seattle, WA. http://www.stat.washington.edu/jaw/RESEARCH/OLD-PAPERS-OTHERS/UMLE.pdf

Brunner, L. J. (1992). Bayesian nonparametric methods for data from a unimodal density. *Statistics and Probability Letters, 14*, 195–199.

Brunner, L. J., Lo, A. Y. (1989). Bayes methods for a symmetric unimodal density and its mode. *The Annals of Statistics, 17*, 1550–1566.

Cox, D. R. (1972). Regression models and life-tables (with discussion). *Journal of the Royal Statistical Society. Series B. Statistical Methodology, 34*, 187–220.

Drăgichi, L., Ramamoorthi, R. V. (2003). Consistency of Dykstra–Laud priors. *Sankhyā, 65*, 464–481.

Dykstra, R. L., Laud, P. (1981). A Bayesian nonparametric approach to reliability. *The Annals of Statistics, 9*, 356–367.

Gail, M. H., Gastwirth, J. L. (1978a). A scale-free goodness-of-fit test for the exponential distribution based on the Gini statistic. *Journal of the Royal Statistical Society Series B, Statistical Methodology, 40*, 350–357.

Gail, M. H., Gastwirth, J. L. (1978b). A scale-free goodness-of-fit test for the exponential distribution based on the Lorenz curve. *Journal of the American Statistical Association, 73*, 787–793.

Ghosh, J. K., Ramamoorthi, R. V. (2003). *Bayesian nonparametrics*. New York: Springer.

Gijbels, I., Heckman, N. (2004). Nonparametric testing for a monotone hazard function via normalized spacings. *Journal of Nonparametric Statistics, 16*, 463–478.

Hald, A. (1981). T. N. Thieles contributions to statistics. *International Statistical Review, 49*, 1–20.

Hall, P., Van Keilegom, I. (2005). Testing for monotone increasing hazard rate. *The Annals of Statistics, 33*, 1109–1137.

Haupt, E., Schabe, H. (1997). The TTT transformation and a new bathtub distribution model. *Journal of Statistical Planning and Inference, 60*, 229–240.

Ho, M.-W. (2006a). A Bayes method for a monotone hazard rate via **S**-paths. *The Annals of Statistics, 34*, 820–836.

Ho, M.-W. (2006b). Bayes estimation of a symmetric unimodal density via **S**-paths. *Journal of Computational and Graphical Statistics, 15*, 848–860.

Ibrahim, J. G., Chen, M.-H., MacEachern, S. N. (1999). Bayesian variable selection for proportional hazards models. *The Canadian Journal of Statistics, 37*, 701–717.

Ishwaran, H., James, L. F. (2003). Generalized weighted Chinese restaurant processes for species sampling mixture models. *Statistica Sinica, 13*, 1211–1235.

Ishwaran, H., James, L. F. (2004). Computational methods for multiplicative intensity models using weighted gamma processes: proportional Hazards, marked point processes, and panel count data. *Journal of the American Statistical Association, 99*, 175–190.

James, L. F. (2003). Bayesian calculus for gamma processes with applications to semiparametric intensity models. *Sankhyā, 65*, 196–223.

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

Jankowski, H., Wellner, J. A. (2008a) Nonparametric estimation of a convex bathtub-shaped hazard function. Technical Report 521, Department of Statistics, University of Washington. http://arXiv:0801.0712v1[math.ST].

Jankowski, H., Wellner, J. A. (2008b) *Computation of Nonparametric Convex Hazard Estimators via Profile Methods*. Technical Report 542, Department of Statistics, University of Washington.

Kalbfleisch, J. D. (1978). Non-parametric Bayesian analysis of survival time data. *Journal of the Royal Statistical Society Series B, Statistical Methodology, 40*, 214–221.

Kalbfleisch, J. D., Prentice, R. L. (1980). *The statistical analysis of failure time data*. New York: Wiley.
Kass, R. E., Raftery, A. E. (1995). Bayes factors and model uncertainty. *Journal of the American Statistical Associaton, 90*, 773–795.
Kingman, J. F. C. (1967). Completely random measures. *Pacific Journal of Mathematics, 21*, 59–78.
Kingman, J. F. C. (1993). *Poisson processes*. Oxford: Oxford University Press.
Kong, A., Liu, J. S., Wong, W. H. (1994). Sequential imputations and Bayesian missing data problems. *Journal of the American Statistical Associaton, 89*, 278–288.
Lai, C. D., Xie, M., Murthy, D. N. P. (2001). Bathtub-shaped failure rate distributions. In N. Balakrishnan, C. R. Rao (Eds), *Handbook of statistics 20* (pp. 69–104). London: Elsevier.
Laud, P., Damien, P., Walker, S. G. (2006). Computations via auxiliary random functions for survival models. *Scandinavian Journal of Statistics, 33*, 219–226.
Lawless, J. F. (1982). *Statistical methods and methods for life time data*. New York: Wiley.
Liu, J., Chen, R. (1998). Sequential Monte Carlo methods for dynamic systems. *Journal of the American Statistical Associaton, 93*, 1032–1044.
Lo, A. Y., Brunner, L. J., Chan, A. T. (1996). *Weighted Chinese restaurant processes and Bayesian mixture models,* Research report 1, The Hong Kong University of Science and Technology, Department of Information and Systems Management. http://www.erin.utoronto.ca/~jbrunner/papers/wcr96.pdf
McCullagh, P. (1987). *Tensor methods in statistics*. London: Chapman & Hall.
Meyn, S. P., Tweedie, R. L. (1993). *Markov chains and stochastic stability*. Berlin: Springer.
Mitra, M., Basu, S. K. (1995). Change point estimation in non-monotonic aging models. *Annals of the Institute of Statistical Mathematics, 47*, 483–491.
Proschan, F., Pyke, R. (1967). Tests for monotone failure rate. In *Fifth Berkeley Symposium, 3*, 293–313.
Rajarshi, S., Rajarshi, M. B. (1988). Bathtub distributions: a review. *Communications in Statistics. Theory and Methods, 17*, 2597–2621.
Reboul, L. (2005). Estimation of a function under shape restrictions. Applications to reliability. *The Annals of Statistics, 33*, 1330–1356.
Silverman, B. W. (1981). Using kernel density estimates to investigate multimodality. *Journal of the Royal Statistical Society. Series B. Statistical Methodology, 43*, 97–99.
Singpurwalla, N. D. (2006). *Reliability and risk: a Bayesian perspective*. Chichester: Wiley.
Vaurio, J. K. (1999). Indentification of process and distribution characteristics by testing monotonic and non-monotonic trends in failure intensities and hazard rates. *Reliability Engineering and System Safety, 64*, 345–357.