Statistics
in Medicine

Research Article

Received 10 August 2009, Accepted 4 January 2012 Published online 22 March 2012 in Wiley Online Library

(wileyonlinelibrary.com) DOI: 10.1002/sim.5353

Bayesian nonparametric inference on quantile residual life function: Application to breast cancer data

Taeyoung Park,[a*†] Jong-Hyeon Jeong[b] and Jae Won Lee[c]

There is often an interest in estimating a residual life function as a summary measure of survival data. For ease in presentation of the potential therapeutic effect of a new drug, investigators may summarize survival data in terms of the remaining life years of patients. Under heavy right censoring, however, some reasonably high quantiles (e.g., median) of a residual lifetime distribution cannot be always estimated via a popular nonparametric approach on the basis of the Kaplan–Meier estimator. To overcome the difficulties in dealing with heavily censored survival data, this paper develops a Bayesian nonparametric approach that takes advantage of a fully model-based but highly flexible probabilistic framework. We use a Dirichlet process mixture of Weibull distributions to avoid strong parametric assumptions on the unknown failure time distribution, making it possible to estimate any quantile residual life function under heavy censoring. Posterior computation through Markov chain Monte Carlo is straightforward and efficient because of conjugacy properties and partial collapse. We illustrate the proposed methods by using both simulated data and heavily censored survival data from a recent breast cancer clinical trial conducted by the National Surgical Adjuvant Breast and Bowel Project.

Keywords: blocked Gibbs sampler; heavily censored survival data; median residual life function; partially collapsed Gibbs sampler; survival analysis

1. Introduction

The proportional hazards model [1] is the most popular model to analyze censored survival data. Although inference under the model through the partial likelihood method [2] is well understood and is shown to be fully efficient under the null hypothesis of no covariate effect [3], the model is based on the hazard function, that is, the conditional limiting probability of instantaneous failure. Efficacy of a new drug from a clinical trial is often reported as the percentage of reduction in the hazard ratio. The concept of the hazard function is, however, sometimes not well interpretable to biologists or physicians without solid statistical backgrounds. Thus, for more effective communication with investigators and patients when discussing the potential therapeutic effect of a new drug, it may be more intuitively straightforward to use an alternative summary measure of survival data such as the remaining life years of patients. The efficacy can then be evaluated in terms of extending a patient's remaining life years, say, 5 years as a median. In this paper, we thus consider the quantile residual life function with the median residual life function being a special case. The estimate of the $100(q)$th percentile residual life function is interpreted as the value which $100(q)\%$ of subjects have the remaining lifetimes of equal to or less than, at a given time.

As a special case of the quantile (residual) life function, many researchers have studied the median (residual) life function [4–8]. In their classical methods for the median (residual) life function, the estimated Kaplan–Meier curve [9] is inverted to estimate the median (residual) life function or other

[a]*Department of Applied Statistics, Yonsei University, Seoul 120-749, Korea*
[b]*Department of Biostatistics, University of Pittsburgh, Pittsburgh, PA 15260, U.S.A.*
[c]*Department of Statistics, Korea University, Seoul 136-701, Korea*
**Correspondence to: Taeyoung Park, Department of Applied Statistics, Yonsei University, 50 Yonsei-Ro, Seodaemun-Gu, Seoul 120-749, Korea.*
[†]*E-mail: tpark@yonsei.ac.kr*

quantile (residual) life functions nonparametrically. Under heavy right censoring, however, it is not possible to estimate the quantiles of a (residual) lifetime distribution unless the Kaplan–Meier estimate reaches the targeted quantities. In this case, a parametric approach that relies on a parametric assumption of a failure time distribution can be considered. However, the parametric approach may suffer from such strong parametric assumption unless it is correctly specified. These difficulties motivate a Bayesian nonparametric approach that enables estimating any quantile (residual) life functions by flexibly and adaptively modeling the unknown failure time distribution with a Dirichlet process (DP) mixture [10, 11] of a parametric family of probability distributions. Consequently, there is a growing stream of research on Bayesian nonparametric methods developed for modeling survival data. Kuo and Mallick [12] and Kottas and Gelfand [13] developed Bayesian nonparametric methods for the error distribution of the log-transformed survival times. Gelfand and Kottas [14] applied the method of Kottas and Gelfand [13] toward inference on the median residual life function. Instead of dealing with the log-transformed survival times, Merrick *et al.* [15] and Kottas [16] directly modeled the unknown failure time distribution with a DP mixture of Weibull distributions. In the aforementioned Bayesian nonparametric methods, the DP mixture models are fitted by marginalizing over a DP mixture prior, resulting in the Pólya urn Gibbs sampler [17, 18]. As described in Ishwaran and James [19] and Gelfand and Kottas [20], however, the posterior distribution of a mixing probability weight is not a direct output from the Pólya urn approach, and thus inference on the general functionals of the DP mixture model must be made after fitting the model. Kottas [16] extended the work of Gelfand and Kottas [20] and provided a method to approximate the general functions associated with survival data.

When a relationship between a specific quantile of a failure time distribution and covariates is of interest, quantile regression [21] has been used for inference. Researchers [22–25] have popularized Bayesian semiparametric/nonparametric methods to use a DP mixture for the error distribution of a quantile regression model. Bayesian nonparametric inference for a quantile regression model is somewhat related to our inference for a quantile life function in that both deal with the quantiles of a failure time distribution. Although quantile regression focuses on curves for a specific quantile of a lifetime conditional on covariates, however, not much attention has been paid to the regression model on a quantile *residual* life function. Our method focuses on estimating various quantiles of a remaining lifetime at a given survival time in a Bayesian nonparametric framework, which can be extended to the regression model on a quantile residual life function.

The purpose of this paper is threefold. First, we develop a method for efficiently estimating any quantiles of a residual lifetime distribution without resorting to restrictive parametric assumptions. To do so, we model the unknown failure time distribution with a DP mixture of Weibull distributions and devise the blocked Gibbs sampler [19] to fit the Weibull DP mixture model. The blocked Gibbs sampler allows *direct* posterior inference on the general functionals of the DP mixture model, for example, the density function, survival function, hazard function, and cumulative hazard function. In particular, when it comes to estimating the quantile residual life function, it is crucial to obtain essentially all features of the survival function. Thus, our direct posterior inference on the survival function will lead to efficient posterior estimation of the quantile residual life function, taking advantage of the Weibull survival function available in a closed form. Second, we intend to extend the previous work by devising a more efficient MCMC sampler that capitalizes on blocking [26] and partial collapse [27]. A Weibull DP mixture model that mixes on both shape and scale parameters results in a series of nonstandard sampling steps for the shape parameter because of nonconjugacy. To improve the computational efficiency without losing the ability to flexibly and adaptively model the unknown failure time distribution, we use the blocked Gibbs sampler that allows updating blocks of parameters, as opposed to the Pólya urn Gibbs sampler. In addition, we exploit partial collapse to further improve the convergence of the blocked Gibbs sampler. The improved computational efficiency results in more reliable inference on the quantile residual life function. Last, we impose a hierarchical structure on a precision parameter and a parameter characterizing a base distribution in the Weibull DP mixture model. We propose conjugate hyperprior distributions that allow data to inform strongly about the precision parameter and base distribution.

We divide the remainder of this paper into five sections. In Section 2, we define the quantile residual life function and introduce the existing and new estimators of the function. In Section 3, we develop the proposed Weibull DP mixture model, specify prior distributions, and describe efficient posterior computation via the blocked Gibbs sampler. Section 4 validates the proposed modeling approach and methods through a simulation study. In Section 5, we apply our proposed DP mixture models and computational methods to heavily censored survival data from a recent breast cancer study conducted by the National Surgical Adjuvant Breast and Bowel Project. Discussion follows in Section 6.

2. Quantile residual life function

For a patient i, let t_i denote a failure time variable with survival function $S(t)$ for $t > 0$. At a given time t_0, a remaining lifetime t^* has the residual survival function $S_{t_0}(t^*) = S(t^* + t_0)/S(t_0)$ for $t^* > 0$ and $t_0 \geq 0$. The quantile residual life function is defined as the inverse of the cumulative distribution function of a remaining lifetime given that no failure occurs up to time t_0. Thus, the $100(q)$th percentile residual life function t_q^* satisfies

$$S_{t_0}(t_q^*) = 1 - q, \tag{1}$$

for $q \in [0, 1]$. The median residual life function $t_{0.5}^*$ is a special case when $q = 0.5$.

Because of early termination of a study, the failure time t_i may not be observed directly because of censoring at c_i, which is assumed to be independent of t_i. Then, we observe $y_i = \min(t_i, c_i)$ with $d_i = I(t_i \leq c_i)$ for patient i, where $I(\cdot)$ is an indicator function. In this paper, we are interested in estimating the quantile residual life function with such right censored or uncensored survival data.

2.1. Existing nonparametric estimator

In the presence of right censoring, one can estimate the quantile residual life function by substituting the survival function, $S(t)$, with the nonparametric Kaplan–Meier estimator, $\widehat{S}(t)$, and inverting the resulting function in (1). Because the Kaplan–Meier estimators are computable only up to certain uncensored values, however, some quantile residual life functions may not be estimable under heavy right censoring. We illustrate the difficulty of estimating quantile residual life functions under heavy right censoring in Figure 1, where we consider two survival data sets each with 10 patients and 60% censoring. In each panel of Figure 1, we list the 10 patients in the increasing order of observed times at which each patient experiences either failure or censoring, the crosses represent failure and the circles represent censoring, and compute the Kaplan–Meier estimate of a survival probability, shown next to each cross. Although both data sets have the same censoring proportion, the median residual life function at a given time $t_0 = 0$ is estimable under study 1, but not under study 2. In study 1, the last uncensored patient with $\widehat{S}(t) = 46\%$ is followed by only two censored patients so that the median residual life function at $t_0 = 0$ is computable. By contrast, five patients have longer observed times than the last uncensored patient with $\widehat{S}(t) = 57\%$ under study 2 and, as a result, the corresponding median residual life function cannot be estimated. Thus, not all quantile residual life functions may be estimable with the nonparametric method in the presence of heavy right censoring. To evaluate uncertainty, we can derive the asymptotic variance of the nonparametric estimator for the quantile residual lifetime using its martingale properties [8].

2.2. Existing Bayesian parametric estimator

The method of data augmentation [28] can simplify the estimation of the quantile residual life function by creating imputations for censored failure times under a parametric assumption on the unknown failure time distribution. Researchers [29, 30] have taken similar approaches to predict event times in the interim analysis of clinical trials at the occurrence of prespecified event counts. Let $\tilde{\mathbf{t}} = \{\tilde{t}_i\}_{i=1}^n$

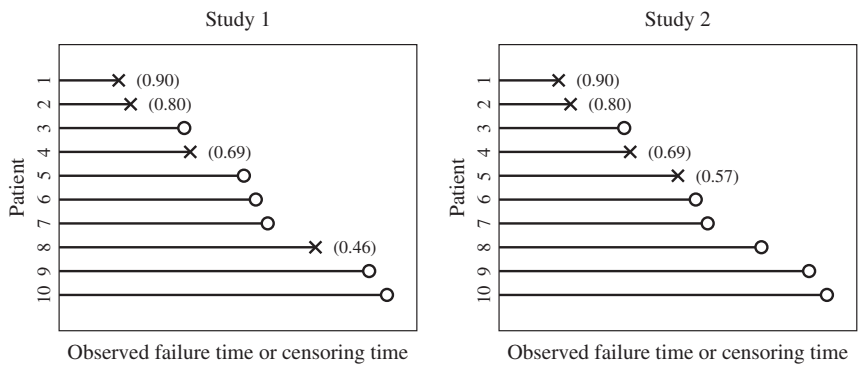

Figure 1. Two different survival data sets with same high-censoring proportion of 60%. In each panel, the crosses represent failure and the circles represent censoring.

denote the augmented failure times with $\tilde{t}_i = y_i$ when $d_i = 1$ for $i = 1, \ldots, n$. That is, observed failure times are augmented by missing failure times for subjects with $d_i = 0$ for $i = 1, \ldots, n$. Then, we simulate the augmented failure times via the posterior predictive distribution,

$$p(\tilde{\mathbf{t}}|\mathbf{Y}) = \int p(\tilde{\mathbf{t}}|\boldsymbol{\theta}, \mathbf{Y}) p(\boldsymbol{\theta}|\mathbf{Y}) \mathrm{d}\boldsymbol{\theta}, \tag{2}$$

for subjects with $d_i = 0$ for $i = 1, \ldots, n$, where $\mathbf{Y} = \{(y_i, d_i)\}_{i=1}^n$ denotes observed data and $\boldsymbol{\theta}$ denotes a collection of model parameters, and then estimate the quantile residual life function t_q^* for each sample of $\tilde{\mathbf{t}}$ by inverting the sample residual survival function

$$\widehat{S}_{t_0}^B(t_q^*) = \frac{\sum_{i=1}^n I(\tilde{t}_i > t_q^* + t_0)}{\sum_{i=1}^n I(\tilde{t}_i > t_0)} = 1 - q. \tag{3}$$

Here, information from censored observations is fully taken advantage of by creating imputations for missing failure times given their observed censored failure times. Note that unlike the nonparametric estimator, the Bayesian estimator is computable under heavy right censoring but may suffer from strong parametric assumptions on the unknown failure time distribution. Here, we assume that the failure times follow the Weibull distribution with survival function $S(t; \rho, \lambda) = \exp(-\lambda t^\rho)$, where ρ denotes a shape parameter and λ denotes a scale parameter. For posterior simulation, we designate independent diffuse gamma prior distributions for the shape and scale parameters.

2.3. *Proposed Bayesian nonparametric estimator*

To circumvent the difficulty of estimating the quantile residual life function under heavy right censoring and the strong parametric assumptions on a failure time distribution, we model the unknown failure time distribution using a DP mixture of Weibull distributions,

$$\begin{array}{lcl} t_i|(\alpha_i, \beta_i) & \overset{\text{ind}}{\sim} & \text{Weibull}(\alpha_i, \beta_i), i = 1, \ldots, n \\ (\alpha_i, \beta_i)|G & \overset{\text{iid}}{\sim} & G, i = 1, \ldots, n \\ G|(\gamma, \eta, \xi) & \sim & \text{DP}(\gamma G_0(\eta, \xi)), \end{array} \tag{4}$$

where Weibull(ρ, λ) denotes the Weibull distribution with survival function $S(t; \rho, \lambda) = \exp(-\lambda t^\rho)$, shape parameter ρ, and scale parameter λ, and the (α_i, β_i)'s are independently and identically distributed according to an unknown prior distribution G, which is drawn from DP with precision parameter γ and base distribution G_0 on $\Re_+$ with parameters η and ξ corresponding to α_i and β_i, respectively. Here, G and G_0 are probability measures, but we refer to them as distributions. The constructive stick-breaking representation [31] implies that G can be represented as

$$G(\cdot) = \sum_{j=1}^{\infty} \pi_j(\mathbf{V}) \delta_{(\theta_j, \vartheta_j)}(\cdot), (\theta_j, \vartheta_j) \overset{\text{iid}}{\sim} G_0, \tag{5}$$

where $\pi_j(\mathbf{V}) = V_j \prod_{l<j}(1 - V_l)$, with $V_j \overset{\text{iid}}{\sim} \text{Beta}(1, \gamma)$, is a probability weight chosen to be independent of (θ_j, ϑ_j) for $j = 1, \ldots, \infty$, $\mathbf{V} = \{V_j\}_{j=1}^\infty$ denotes a set of random variables from a stick-breaking process, and $\delta_{(\theta,\vartheta)}$ is a discrete probability measure concentrated at (θ, ϑ). With the use of the stick-breaking priors, the unknown failure time distribution is essentially modeled by a countably infinite mixture of Weibull distributions and the corresponding survival function can be expressed as

$$\begin{aligned} S^{\text{DP}}(t) &= \int S(t; \alpha, \beta) \mathrm{d}G(\alpha, \beta) \\ &= \sum_{j=1}^{\infty} \pi_j(\mathbf{V}) S(t; \theta_j, \vartheta_j). \end{aligned} \tag{6}$$

At a given time t_0, the proposed Bayesian nonparametric estimator t_q^* for the $100(q)$th percentile residual life function is computed by inverting the residual survival function

$$S_{t_0}^{\text{DP}}(t_q^*) = \frac{S^{\text{DP}}(t_q^* + t_0)}{S^{\text{DP}}(t_0)} = 1 - q, \tag{7}$$

which can be solved by using a root-finding algorithm such as the bisection method. Note that unlike the nonparametric estimator, the Bayesian nonparametric estimator is fully model based and thus allows extrapolation beyond the last observed failure time. Also, unlike the parametric estimator, the Bayesian nonparametric estimator avoids strong parametric assumptions on the unknown failure time distribution. When the Pólya urn Gibbs sampler constructed by integrating out G is used to fit the Weibull DP mixture model in (4), inference for the residual survival function in (7) relies on a Monte Carlo approximation to necessary integrals after fitting the model. Here, we use the blocked Gibbs sampler that enables direct posterior inference on the quantile residual life function as well as the residual survival function, without the need to do such a post process.

3. Dirichlet process mixture of Weibull distributions

3.1. The Weibull Dirichlet process mixture model

Instead of marginalizing G out of the Weibull DP mixture model in (4), we construct a finite dimensional prior by applying an almost sure truncation to the stick-breaking representation of G. As the index j in (5) increases, the probability weight assigned to θ_j decreases rapidly so that most probability weights are essentially assigned to a few dominating θ_j's. Thus, we can achieve an accurate approximation to G by truncating the infinite sum in (5) at the first J terms, that is, by letting $V_J = 1$; see [19] for theoretical arguments.

To facilitate the fitting of the DP mixture model, it is useful to introduce a latent variable Z_i, which indicates cluster membership with $Z_i = j$ for subject i belonging to cluster j. With a finite dimensional prior, the hierarchical DP mixture model in (4) is rewritten as

$$\begin{aligned} t_i|(\boldsymbol{\theta},\boldsymbol{\vartheta},\mathbf{Z}) &\overset{\text{ind}}{\sim} \text{Weibull}(\theta_{Z_i},\vartheta_{Z_i}), i=1,\ldots,n \\ Z_i|\mathbf{V} &\overset{\text{iid}}{\sim} \sum_{j=1}^{J}\pi_j(\mathbf{V})\delta_j(Z_i), i=1,\ldots,n \\ V_j|\gamma &\overset{\text{iid}}{\sim} \text{Beta}(1,\gamma), j=1,\ldots,J-1 \\ (\theta_j,\vartheta_j)|(\eta,\xi) &\overset{\text{iid}}{\sim} G_0(\eta,\xi), j=1,\ldots,J \\ \gamma,\eta,\xi &\sim p(\gamma)p(\eta)p(\xi), \end{aligned} \tag{8}$$

where $\boldsymbol{\theta} = (\theta_1,\ldots,\theta_J)$ and $\boldsymbol{\vartheta} = (\vartheta_1,\ldots,\vartheta_J)$ denote sets of atoms for the Weibull shape and scale parameters, respectively, $\mathbf{Z} = \{Z_i\}_{i=1}^n$ denotes a set of latent variables for allocating each subject to a cluster, $\mathbf{V} = \{V_j\}_{j=1}^J$ denotes a set of random variables from a truncated stick-breaking process, and $\pi_j(\mathbf{V}) = V_j\prod_{l<j}(1-V_l)$ is a probability weight assigned to the jth atom of $(\boldsymbol{\theta},\boldsymbol{\vartheta})$, where $V_J = 1$. The base distribution G_0 is chosen to make posterior computation straightforward through log-concavity and conjugacy properties. In particular, we impose a constraint on the shape parameter such that $\theta_j > 1$, $j = 1,\ldots,J$, to make the posterior simulation of θ_j numerically stable. Thus, we choose $G_0(\eta,\xi) = \text{Trunc-Gamma}(1,\eta,1)\text{Gamma}(1,\xi)$ with rate parameters η and ξ, where Trunc-Gamma(a,b,c) denotes a truncated gamma distribution with shape parameter a, rate parameter b, and left truncation c, and Gamma(a,b) denotes a gamma distribution with shape parameter a, rate parameter b, and mean a/b.

3.2. Prior specification

In a DP mixture model, the precision parameter γ plays a key role in determining the number of clusters. Small values of γ tend to assign most probability weights to a few dominating atoms of $(\boldsymbol{\theta},\boldsymbol{\vartheta})$. On the other hand, for large values of γ, each atom of $(\boldsymbol{\theta},\boldsymbol{\vartheta})$ is assigned a small probability weight, resulting in a bigger number of clusters. Instead of fixing the value of γ, we let the precision parameter be random, allowing data to inform strongly about the level of clustering. Specifically, we use a conjugate gamma prior distribution, $\gamma \sim \text{Gamma}(a_\gamma, b_\gamma)$. Prior knowledge about the level of clustering can be incorporated by changing the values of a_γ and b_γ. Thus, we check how posterior inference is affected according to the choice of a_γ and b_γ; see Section 4.2 for the prior sensitivity issue.

Next, the rate parameters (η,ξ) of the base distribution G_0 are also chosen to be random. A product of independent gamma conjugate prior distributions for (η,ξ) leads to a convenient update in the blocked Gibbs sampler so that we let $(\eta,\xi) \sim \text{Gamma}(a_\eta,b_\eta)\text{Gamma}(a_\xi,b_\xi)$. We choose the values of a_η, b_η, a_ξ, and b_ξ, which result in a diffuse prior distribution for (η,ξ).

3.3. *Efficient posterior computation*

We construct the blocked Gibbs sampler to fit the Weibull DP mixture model in (8). Whereas the Pólya urn Gibbs sampler updates $\boldsymbol{\theta}$ and $\boldsymbol{\vartheta}$ moving one coordinate at a time, the blocked Gibbs sampler allows us to update $\boldsymbol{\theta}$ and $\boldsymbol{\vartheta}$ as a block. By moving all coordinates of $\boldsymbol{\theta}$ and $\boldsymbol{\vartheta}$ at once, the blocked Gibbs sampler is expected to achieve much faster convergence than the Pólya urn Gibbs sampler. In addition, we further improve the convergence of the blocked Gibbs sampler by partially collapsing $\boldsymbol{\theta}$. Reducing the conditioning on $\boldsymbol{\theta}$ only in the sampling step for $\boldsymbol{\vartheta}$ does neither affect the log-concavity of the conditional posterior distribution for $\boldsymbol{\theta}$ nor does it complicate updating of the other model components. Such partial collapse, however, increases the variance of the conditional posterior distribution for $\boldsymbol{\theta}$, thereby mixing a Markov chain faster [27, 32]. Here, the partial collapse of $\boldsymbol{\vartheta}$ in the sampling step for $\boldsymbol{\theta}$ results in combining sampling steps for $\boldsymbol{\theta}$ and $\boldsymbol{\vartheta}$ into a single sampling step, thereby jointly updating $(\boldsymbol{\theta}, \boldsymbol{\vartheta})$ as a block.

To simulate the target posterior distribution $p(\mathbf{Z}, \mathbf{V}, \boldsymbol{\theta}, \boldsymbol{\vartheta}, \gamma, \eta, \xi | \mathbf{Y})$ with $\mathbf{Y} = \{(y_i, d_i)\}_{i=1}^n$, the partially collapsed Gibbs sampler is used to iteratively draw values from the following conditional distributions:

1. Sample Z_i from its complete conditional distribution, that is, a multinomial distribution with probabilities
$$p(Z_i = j | \mathbf{V}, \boldsymbol{\theta}, \boldsymbol{\vartheta}, \gamma, \eta, \xi, \mathbf{Y}) = \frac{\{V_j \prod_{l<j}(1-V_l)\} h(y_i; \theta_j, \vartheta_j)^{d_i} S(y_i; \theta_j, \vartheta_j)}{\sum_{k=1}^{J} \{V_k \prod_{l<k}(1-V_l)\} h(y_i; \theta_k, \vartheta_k)^{d_i} S(y_i; \theta_k, \vartheta_k)},$$
for $i = 1, \ldots, n$ and $j = 1, \ldots, J$, where $h(t; \rho, \lambda)$ is a Weibull hazard function, that is, $h(t; \rho, \lambda) = -d \log S(t; \rho, \lambda)/\mathrm{d}t$.
2. Sample V_j from its complete conditional distribution, which is
$$V_j | (\mathbf{Z}, \boldsymbol{\theta}, \boldsymbol{\vartheta}, \gamma, \eta, \xi, \mathbf{Y}) \overset{\text{ind}}{\sim} \text{Beta}\left(1 + \sum_{i=1}^{n} I(Z_i = j), \gamma + \sum_{i=1}^{n} I(Z_i > j)\right),$$
for $j = 1, \ldots, J-1$.
3. Jointly sample $(\boldsymbol{\theta}, \boldsymbol{\vartheta})$ from their complete conditional distribution, which is
$$p(\boldsymbol{\theta}, \boldsymbol{\vartheta} | \mathbf{Z}, \mathbf{V}, \gamma, \eta, \xi, \mathbf{Y}) = \prod_{j=1}^{J} p(\theta_j | \mathbf{Z}, \mathbf{V}, \gamma, \eta, \xi, \mathbf{Y}) p(\vartheta_j | \mathbf{Z}, \mathbf{V}, \theta_j, \gamma, \eta, \xi, \mathbf{Y}),$$
where the reduced conditional distribution of θ_j after partially collapsing $\boldsymbol{\vartheta}$ out of its complete conditional distribution is proportional to
$$\theta_j^{\sum_{i:Z_i=j} d_i} \exp\left\{-\eta\theta_j + \theta_j \sum_{i:Z_i=j} d_i \log(y_i) - \left(1 + \sum_{i:Z_i=j} d_i\right) \log\left(\xi + \sum_{i:Z_i=j} y_i^{\theta_j}\right)\right\},$$
for $\theta_j > 1$, and the complete conditional distribution of ϑ_j is
$$\vartheta_j | (\mathbf{Z}, \mathbf{V}, \theta_j, \gamma, \eta, \xi, \mathbf{Y}) \overset{\text{ind}}{\sim} \text{Gamma}\left(1 + \sum_{i:Z_i=j} d_i, \xi + \sum_{i:Z_i=j} y_i^{\theta_j}\right),$$
for $j = 1, \ldots, J$.
4. Sample γ from its complete conditional distribution, which is
$$\gamma | (\mathbf{Z}, \mathbf{V}, \boldsymbol{\theta}, \boldsymbol{\vartheta}, \eta, \xi, \mathbf{Y}) \sim \text{Gamma}\left(a_\gamma + J, b_\gamma - \sum_{j=1}^{J-1} \log(1 - V_j)\right).$$
5. Sample η from its complete conditional distribution, which is
$$\eta | (\mathbf{Z}, \mathbf{V}, \boldsymbol{\theta}, \boldsymbol{\vartheta}, \gamma, \xi, \mathbf{Y}) \sim \text{Gamma}\left(a_\eta + J, b_\eta + \sum_{j=1}^{J} \theta_j - J\right).$$

6. Sample ξ from its complete conditional distribution, which is

$$\xi|(\mathbf{Z}, \mathbf{V}, \boldsymbol{\theta}, \boldsymbol{\vartheta}, \gamma, \eta, \mathbf{Y}) \sim \text{Gamma}\left(a_\xi + J, b_\xi + \sum_{j=1}^{J} \vartheta_j\right).$$

Because the reduced conditional distribution of θ_j in step 3 can be shown to be log-concave whenever $\sum_{i=1}^{n} I(Z_i = j)$ and θ_j are positive, we use adaptive rejection sampling [33] with truncation of $\theta_j > 1$ to efficiently implement the step.

4. Simulation study

4.1. Setting

We conduct our simulation study to check the flexibility of the proposed modeling approach and assess the validity of the proposed Bayesian nonparametric estimator for the quantile residual life function. For the first simulation study, we generate failure times of $n = 300$ from a mixture of three lognormal distributions,

$$0.4 \times \text{LN}(0, 0.2) + 0.3 \times \text{LN}(1.2, 0.05) + 0.3 \times \text{LN}(1.4, 0.5), \tag{9}$$

where $\text{LN}(\mu, \sigma^2)$ denotes a lognormal distribution with mean $\exp(\mu + \sigma^2/2)$. The mixture of three lognormal distributions is bimodal and has a much thicker right tail than a single Weibull distribution; see Figure 2. We consider such an irregular failure time distribution to illustrate that our proposed Weibull DP mixture model is rich enough to allow for multimodal shape and general tail behavior and flexible enough to deal with a mixture of densities that are not in the Weibull family. We assume no censoring and consider the quantile residual life function at times $t_0 = 0, 0.2, \ldots, 3$ with $q = 0.25$ and 0.5.

For the second simulation study, we illustrate the robustness of the proposed Bayesian nonparametric estimator for heavily censored data. We generate $n = 1000$ failure times from a single Weibull distribution with $\alpha = 2$ and $\beta = 0.01$. A distribution of censoring time is assumed to follow an independent uniform

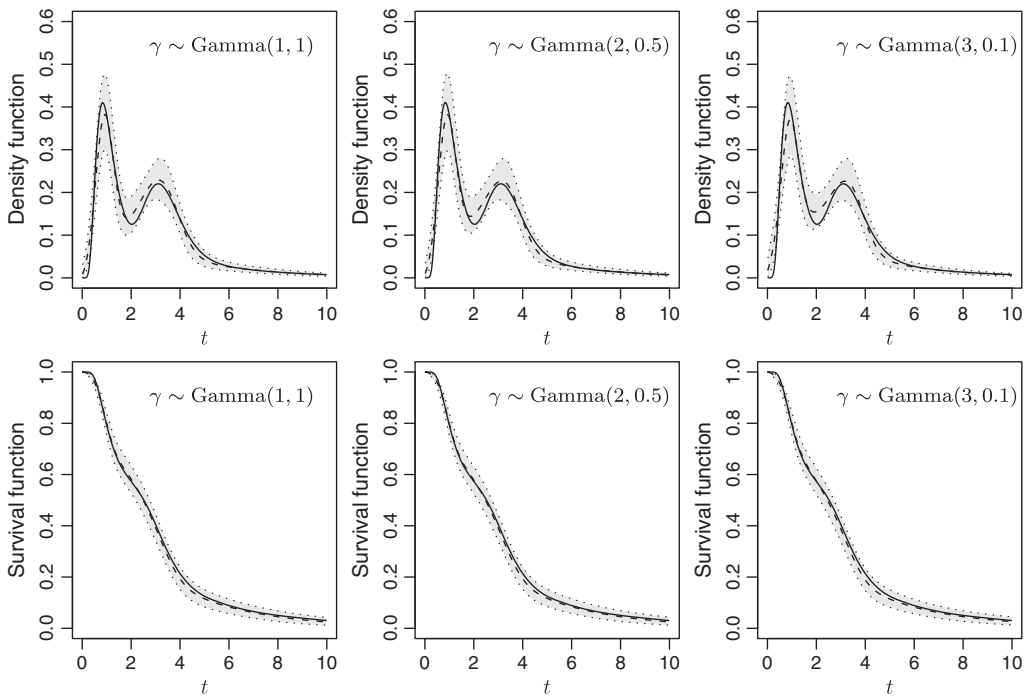

Figure 2. Posterior inference on the density and survival functions under three different prior distributions for γ. The solid lines represent the true curve for either a density or survival function, and the dashed and dotted lines represent pointwise posterior means and 95% credible intervals for the function, respectively.

distribution between 1 and c^* so that censoring does not occur earlier than 1 year. We choose c^* to adjust a censoring proportion to 30% and 50%. The values of c^* corresponding to these censoring proportions are 27.22 and 10.49, respectively. For each c^*, we generate test data and compute Bayesian nonparametric estimates for the quantile residual life function at times $t_0 = 0$, 1, 2, and 3 with $q = 0.25$ and 0.5.

4.2. *Prior sensitivity*

In a DP(γG_0) prior for G, the precision parameter γ controls how subjects will be grouped into clusters so that it affects the number of clusters in a DP mixture. Although it is common to let $\gamma = 1$ in applications, we construct a hierarchical structure by imposing a prior distribution on γ to let the data inform about it. To check the impact of prior knowledge about γ on posterior inference, we consider three prior distributions for γ. Our three choices are Gamma(1, 1), Gamma(2, 0.5), and Gamma(3, 0.1); as the mean of a gamma prior distribution increases, a bigger number of clusters are assumed a priori. The expected number of distinct clusters is approximately $(a_\gamma/b_\gamma)\log(1 + nb_\gamma/a_\gamma)$ a priori [16] so that it is about 72 with $n = 300$, $a_\gamma = 3$, and $b_\gamma = 0.1$ for the first simulation study. Considering Gamma(3, 0.1) is a rather extreme prior distribution for γ, we truncate the stick-breaking representation of G at $J = 100$ terms, which would be sufficient enough to give virtually indistinguishable results from the ones with an infinite dimensional prior when data inform strongly about a precision parameter. As for diffuse prior distributions for η and ξ, we use $\eta \sim$ Gamma(2, 0.1) and $\xi \sim$ Gamma(2, 0.1). As a result of a prior sensitivity analysis, Figure 2 shows that posterior inference is not sensitive to the choice of a prior distribution for γ. In the subsequent analyses, we thus use Gamma(2, 0.5) as a conservative choice for the prior distribution for γ and choose a finite dimensional prior for G that is truncated at $J = 100$ terms. Also, we assume diffuse gamma prior distributions, $\eta \sim$ Gamma(2, 0.1) and $\xi \sim$ Gamma(2, 0.1), for η and ξ.

4.3. *Results*

For each test data set, the blocked Gibbs sampler devised in Section 3 is run with two separate chains of 20,000 iterations with different starting values. After detecting convergence for all fixed-dimensional parameters using the $\hat{R}^{1/2}$ statistic [34], we collect 20,000 posterior draws from the second halves of the two chains and use them for posterior inference.

We aim the first simulation study at checking model flexibility, and the resulting posterior inference is as follows. Figure 2 shows estimated density and survival functions for a failure time variable using the three prior distributions for γ. In Figure 2, the solid lines represent the true curve for a density or survival function of the lognormal mixture in (9), the dashed lines represent pointwise posterior means for the density or survival function, and the dotted lines represent pointwise 95% credible intervals for the density or survival function. The bimodal shape and heavy-tailed behavior of the true density function curve is well specified by posterior estimates. The choice of a prior distribution for γ does not seem to affect posterior inference significantly, producing similar posterior estimates for the density and survival functions. In effect, all probability weights are assigned to a much smaller number of atoms than $J = 100$ under all three prior distributions, showing posterior learning for γ from observed data.

Figure 3 presents Bayesian nonparametric estimates for the 25th percentile and median residual life functions as a function of a given time t_0, resulting from direct posterior simulation of G. The use of the blocked Gibbs sampler devised in Section 3 allows direct inference on the quantile residual life function, as shown in Figure 3. The true curve for the quantile residual life function as a function of given time t_0 is well estimated by its pointwise posterior means and reasonably covered by its pointwise 95% credible intervals. Again, the choice of a prior distribution for γ has a negligible effect on posterior inference about the quantile residual life function.

We conduct the second simulation study to demonstrate the validity of the proposed Bayesian nonparametric estimator for the quantile residual life function under heavy right censoring. After fitting two test data sets with censoring proportion of 30% and 50%, posterior inference is as follows. Table I compares the true values for the 25th percentile and median residual life functions with their posterior estimates. Under all different settings for a censoring proportion, the quantile residual life function seems well estimated by the posterior mean and its true value is contained in the 95% credible interval. As the censoring rate increases, more information is lost in the data. Consequently, with higher censoring, a 95% credible interval is associated with more uncertainty and becomes wider, which is confirmed in Table I.

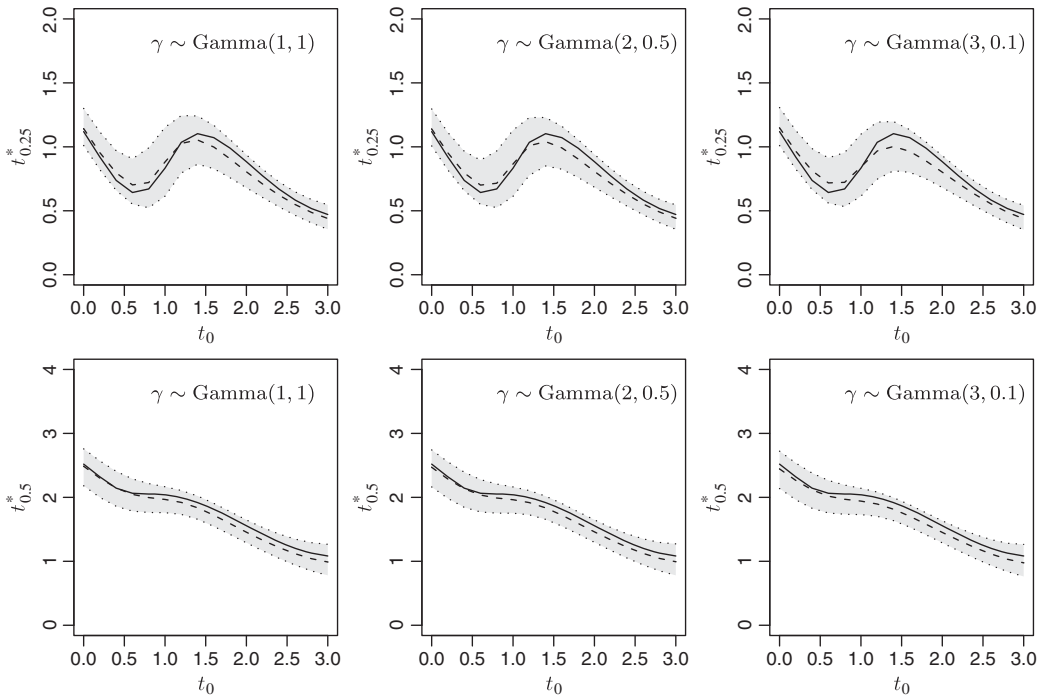

Figure 3. Posterior inference on the quantile residual life function with $q = 0.25$ and 0.5 under three different prior distributions for γ. The solid lines represent the true curve for the quantile residual life function as a function of a given time t_0, and the dashed and dotted lines represent the corresponding nonparametric Bayesian estimates, pointwise posterior means, and 95% credible intervals, respectively.

Table I. Summary statistics of quantiles for residual lifetime under heavy censoring.

			30% censoring		50% censoring	
Given time	q	True value	Estimate	95%CI	Estimate	95%CI
$t_0 = 0$	0.25	5.36	5.35	(5.06, 5.64)	5.42	(5.13, 5.71)
	0.5	8.33	8.31	(7.97, 8.65)	8.21	(7.89, 8.58)
$t_0 = 1$	0.25	4.46	4.44	(4.16, 4.71)	4.49	(4.21, 4.77)
	0.5	7.39	7.37	(7.04, 7.71)	7.26	(6.93, 7.62)
$t_0 = 2$	0.25	3.72	3.71	(3.45, 3.95)	3.72	(3.47, 3.98)
	0.5	6.56	6.55	(6.22, 6.88)	6.40	(6.08, 6.76)
$t_0 = 3$	0.25	3.15	3.13	(2.90, 3.34)	3.11	(2.90, 3.34)
	0.5	5.85	5.84	(5.53, 6.15)	5.66	(5.34, 6.01)

5. Application to breast cancer data

The National Surgical Adjuvant Breast and Bowel Project performed the first study (B-14) to assess the efficacy of a hormonal therapy using tamoxifen in women with lymph node-negative and estrogen receptor-positive breast cancer [35]. In this study, a total of 1450 patients were randomized to the placebo group and 1435 patients to the tamoxifen group over a period of 6 years. The mean time of the study was 13 years. A subset of only 2817 (1413 in placebo; 1404 in tamoxifen) eligible patients with follow-up is analyzed in this paper. The endpoint of overall survival used here is defined as time to any death. Fisher *et al.* [36] reported an update that the tamoxifen-treated group still had a significantly better survival than the placebo group after the long-term follow-up. Because a censoring proportion in this data set is as high as 66%, some reasonably high quantiles, including the median, of a residual lifetime distribution could not be estimated by using the nonparametric Kaplan–Meier estimator. In this breast cancer data set, however, we are mainly interested in the comparison of the 25th percentile and median residual life functions between the tamoxifen-treated and placebo groups.

The blocked Gibbs sampler is run with two separate chains of 20,000 iterations and convergence is assessed by computing the $\hat{R}^{1/2}$ statistic. From the second halves of the two chains, 20,000 posterior draws are retained for inference after discarding the first 10,000 burn-in draws from each chain. The effect of a prior distribution for γ is negligible so that the following results are based on $\gamma \sim \text{Gamma}(2, 0.5)$. We truncate the stick-breaking representation of G at $J = 100$ terms and use $\eta \sim \text{Gamma}(2, 0.1)$ and $\xi \sim \text{Gamma}(2, 0.1)$. Figure 4 shows pointwise posterior means and 95% credible intervals of the density and survival functions for both tamoxifen-treated and placebo groups, along with the Kaplan–Meier curve. Note that the density function for the tamoxifen–treated group is slightly bimodal and heavy tailed, which does not adhere to typical parametric assumptions. The failure time distribution for the tamoxifen-treated group seems to have a local mode around 15 years, which is because quite a few patients are clustered between 10 and 15 years, and censoring rates for these patients are high. Thus, if patients are followed-up beyond the study period, these censored patients will cause the failure time distribution to be even more right skewed. By modeling such an irregular failure time distribution using a DP mixture of Weibull distributions, the pointwise posterior means for the survival function closely follow the Kaplan–Meier curve resulting from a distribution-free nonparametric approach, as shown in the bottom two panels in Figure 4. Unlike the nonparametric method, however, the blocked Gibbs sampler used to fit the Weibull DP mixture model allows directly estimating the entire survival function, extending the estimation range beyond the last observed failure time.

Table II shows summary statistics for the quantile residual life functions with $q = 0.25$ and 0.5 at every other year after surgery, using nonparametric, Bayesian parametric, and Bayesian nonparametric methods. In general, as a given time t_0 increases, more uncertainty is accounted for in estimating the quantile residual life function. As a result, the 95% credible intervals become wider for both tamoxifen-treated and placebo groups as t_0 increases. Because of high censoring, the nonparametric method cannot estimate the median residual life function for both tamoxifen-treated and placebo groups. By contrast, both the Bayesian parametric and Bayesian nonparametric methods are able to estimate the median residual life function because of being model based. The Bayesian parametric method, however, seems

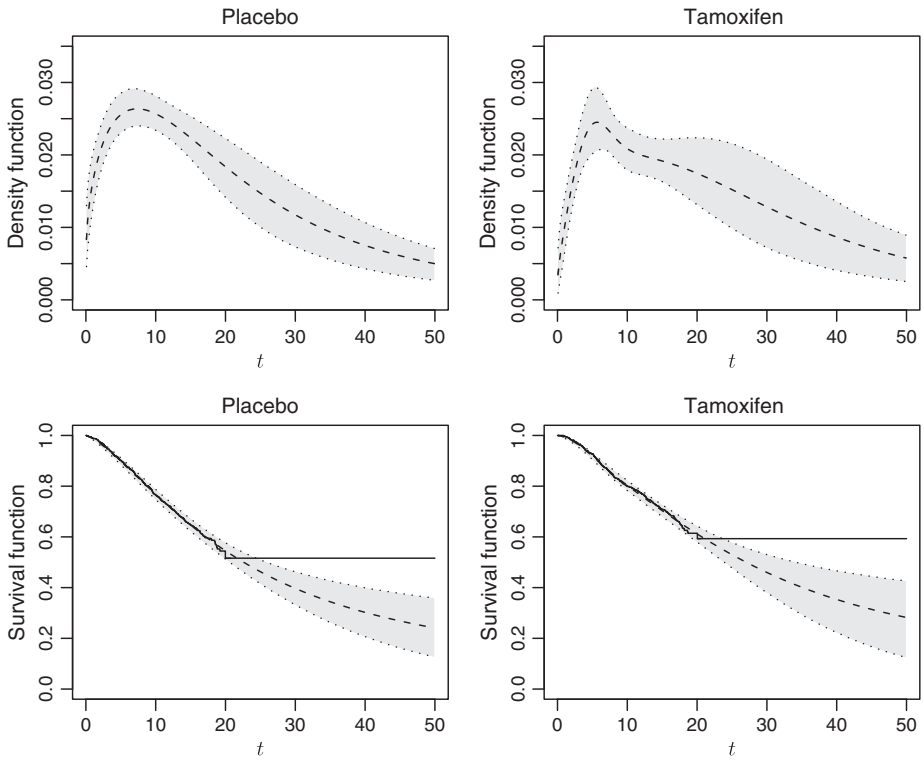

Figure 4. Posterior inference on the density and survival functions for both tamoxifen-treated and placebo groups. The dashed lines represent pointwise posterior means and the dotted lines represent pointwise 95% credible intervals. In the bottom two panels, the Kaplan–Meier curve is shown in solid lines. Note that the maximum follow-up year was about 20 years, but a larger support of the survival function is used to compare the nonparametric Kaplan–Meier curve with the Bayesian nonparametric estimates.

Table II. Summary statistics of quantiles for residual lifetime.

			Placebo		Tamoxifen	
Method	Given time	q	Estimate	95% interval	Estimate	95% interval
NP	$t_0 = 0$	0.25	10.66	(9.65, 11.64)	12.85	(12.02, 14.04)
		0.5	—†	—	—	—
	$t_0 = 2$	0.25	9.50	(8.57, 10.53)	11.63	(10.60, 12.72)
		0.5	—	—	—	—
	$t_0 = 4$	0.25	9.05	(8.24, 10.08)	11.22	(9.94, 12.23)
		0.5	—	—	—	—
	$t_0 = 6$	0.25	8.57	(7.69, 9.79)	10.91	(9.73, 11.63)
		0.5	—	—	—	—
	$t_0 = 8$	0.25	8.38	(7.45, 9.44)	10.13	(9.36, 12.04)
		0.5	—	—	—	—
	$t_0 = 10$	0.25	8.53	(6.82, 9.98)	10.04	(7.64, 10.53)
		0.5	—	—	—	—
BP	$t_0 = 0$	0.25	10.68	(10.57, 10.78)	12.85	(12.69, 13.19)
		0.5	21.57	(20.62, 22.72)	25.02	(23.56, 26.86)
	$t_0 = 2$	0.25	9.51	(9.37, 9.67)	11.62	(11.46, 11.74)
		0.50	20.19	(19.16, 21.44)	23.47	(21.95, 25.39)
	$t_0 = 4$	0.25	9.05	(8.93, 9.26)	11.01	(10.74, 11.47)
		0.5	19.43	(18.25, 20.87)	22.54	(20.87, 24.64)
	$t_0 = 6$	0.25	8.56	(8.31, 8.84)	10.74	(10.30, 11.11)
		0.5	18.66	(17.30, 20.31)	21.82	(19.94, 24.14)
	$t_0 = 8$	0.25	8.35	(8.02, 8.68)	10.51	(9.90, 11.22)
		0.5	17.99	(16.47, 19.86)	21.43	(19.32, 24.09)
	$t_0 = 10$	0.25	8.29	(7.73, 8.80)	10.04	(9.25, 10.99)
		0.5	17.62	(15.87, 19.75)	20.69	(18.42, 23.61)
BNP	$t_0 = 0$	0.25	10.70	(9.89, 11.54)	12.62	(11.57, 13.72)
		0.5	22.77	(20.60, 25.88)	27.60	(23.89, 34.13)
	$t_0 = 2$	0.25	9.64	(8.83, 10.46)	11.37	(10.30, 12.46)
		0.5	21.79	(19.38, 25.40)	26.39	(22.31, 33.71)
	$t_0 = 4$	0.25	9.04	(8.24, 9.88)	10.88	(9.83, 12.02)
		0.5	21.39	(18.52, 25.98)	26.04	(21.28, 35.11)
	$t_0 = 6$	0.25	8.69	(7.88, 9.58)	10.80	(9.68, 12.07)
		0.5	21.36	(17.87, 27.50)	26.27	(20.52, 38.16)
	$t_0 = 8$	0.25	8.51	(7.63, 9.51)	10.76	(9.47, 12.33)
		0.5	21.65	(17.32, 30.04)	26.73	(19.76, 42.69)
	$t_0 = 10$	0.25	8.44	(7.39, 9.75)	10.66	(9.13, 12.70)
		0.5	22.24	(16.83, 33.52)	27.30	(18.91, 48.08)

†Because of heavy right censoring, the median residual lifetime is not estimable with the nonparametric method.
NP, nonparametric; BP, Bayesian parametric; BNP, Bayesian nonparametric.

too restrictive to model an irregular shape of a failure time distribution for the tamoxifen-treated group. As a result, all Bayesian nonparametric estimates of the median residual life function are not within the Bayesian parametric 95% credible intervals. When it comes to the 25th percentile residual life function, the point estimates of three methods agree with one another. Given the bimodal and heavy-tailed failure time distribution for the tamoxifen-treated group, the corresponding residual life distribution becomes more right skewed as a given time t_0 increases. Comparing two flexible modeling approaches, the 95% credible interval constructed by the Bayesian nonparametric method seem to better account for such right skewness than the nonparametric 95% confidence intervals, because the right endpoint of a 95% interval is farther from a point estimate than a left endpoint, especially when t_0 is large.

With the posterior estimates of the survival function, the quantile residual life function can be directly estimated via the relationship in (7). Figure 5 shows the pointwise posterior estimates of the 25th percentile and median residual life functions as a function of a given time t_0. As shown in Figure 5, the tamoxifen-treated group tends to have longer 25th percentile and median residual survivals than the placebo group. The difference between the tamoxifen-treated and placebo groups looks more significant

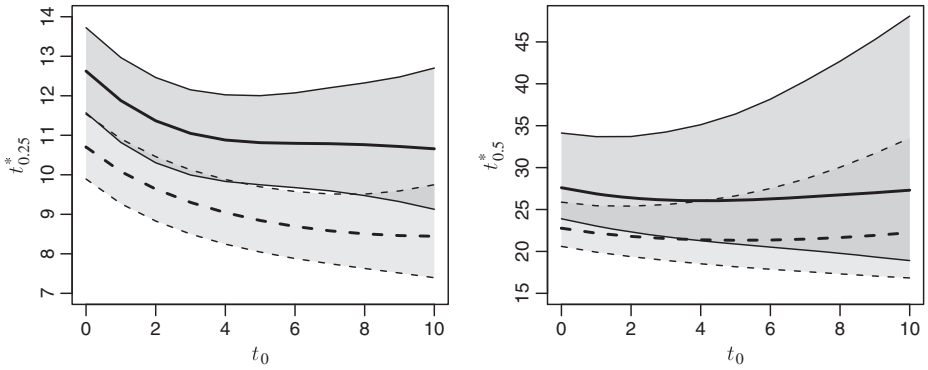

Figure 5. Posterior inference on the quantile residual life function with $q = 0.25$ and 0.5 for both tamoxifen-treated and placebo groups. The solid lines represent the pointwise posterior means (thicker lines) and 95% credible intervals (thinner lines) for the tamoxifen-treated group and the dashed lines for the placebo group.

Table III. Comparison of the tamoxifen-treated group (group 1) and placebo group (group 2).

Given time	q	$\mathrm{E}(t^*_{q1} - t^*_{q2}\|\mathbf{Y})$	$\mathrm{SD}(t^*_{q1} - t^*_{q2}\|\mathbf{Y})$	$p(t^*_{q1} > t^*_{q2}\|\mathbf{Y})$
$t_0 = 0$	0.25	1.92	0.70	0.997
	0.5	4.83	3.29	0.971
$t_0 = 2$	0.25	1.73	0.69	0.994
	0.5	4.60	3.74	0.947
$t_0 = 4$	0.25	1.84	0.69	0.996
	0.5	4.66	4.69	0.911
$t_0 = 6$	0.25	2.10	0.75	0.998
	0.5	4.91	6.18	0.868
$t_0 = 8$	0.25	2.26	0.88	0.996
	0.5	5.08	8.34	0.816
$t_0 = 10$	0.25	2.21	1.12	0.983
	0.5	5.06	11.29	0.754

for the 25th percentile residual survival, because the corresponding 95% credible intervals for both groups are seldom overlapped. As for the median residual life function, the treatment effect of tamoxifen becomes more uncertain as time after surgery progresses.

To closely examine the treatment effect, we provide summary statistics for the difference in quantile residual life functions between the tamoxifen-treated group (group 1) and placebo group (group 2) in Table III. The posterior distribution of the treatment effect is summarized by a posterior mean, $\mathrm{E}(t^*_{q2} - t^*_{q1}|\mathbf{Y})$, a posterior standard deviation, $\mathrm{SD}(t^*_{q2} - t^*_{q1}|\mathbf{Y})$, and the posterior probability of a positive treatment effect, $p(t^*_{q1} > t^*_{q2}|\mathbf{Y})$, where t^*_{qk} denotes the $100(q)$th percentile residual life function of group k. The posterior mean of a difference in the 25th percentile residual life function stays more or less around 2 years with posterior standard deviation of about 0.7 year, as time after surgery progresses. On the other hand, a difference in the median residual life function has a posterior mean of about 5 years, but the corresponding posterior standard deviation is 3.29 years right after surgery and quickly grows to 11.29 years 10 years after surgery, showing more uncertainty as time after surgery progresses. The posterior probability of a positive treatment effect quantifies evidence for a longer residual lifetime of the tamoxifen-treated group than the placebo group. As shown in the last column of Table III, patients in the tamoxifen-treated group have a better 25th percentile residual survival than those in the placebo group with posterior probability greater than 90% at given times $t_0 = 0, \ldots, 10$. The posterior probability that the tamoxifen-treated patients have a better median residual survival than the placebo-controlled patients is also over 90% until 4 years after surgery and then gradually decreases to 75.4% 10 years after surgery.

6. Discussion

In this paper, we propose a Bayesian nonparametric method for efficiently estimating the quantile residual life function in the presence of heavy right censoring. The main idea is to flexibly and adaptively

model the unknown failure time distribution with a DP mixture of Weibull distributions and make direct posterior inference on the quantile residual life function by fitting the Weibull DP mixture model via the blocked Gibbs sampler.

It is well known that the major advantage of the nonparametric method over the parametric method is its flexibility in that it does not need any parametric assumption for the unknown failure time distribution. However, the nonparametric method does not always allow estimating reasonably high quantiles, say the median, of a residual lifetime distribution under heavy right censoring. The proposed Bayesian nonparametric approach is thus developed to estimate any quantile residual life function in a fully model-based framework without losing its ability to flexibly model the unknown failure time distribution. Our simulation study and real data example demonstrate the advantages of the Bayesian nonparametric method over the traditional parametric and nonparametric methods especially when survival data are highly censored.

Acknowledgements

The authors thank the editors and anonymous referees for their valuable comments and suggestions. The Korea Science and Engineering Foundation grant (KOSEF-2011-8-0882) funded by the Korea government, the National Health Institute grants (5-U10-CA69974-09 and 5-U10-CA69651-11), and the Korea Research Foundation grant (KRF-2008-313-C00148) supported this work.

References

1. Cox DR. Regression models and life-tables (with discussion). *Journal of the Royal Statistical Society, Series B* 1972; **34**:187–220.
2. Cox DR. Partial likelihood. *Biometrika* 1975; **62**:269–276.
3. Oakes D. The asymptotic information in censored survival data. *Biometrika* 1977; **64**:441–448.
4. Fligner MA, Rust SW. A modification of Mood's median test for the generalized Behrens-Fisher problem. *Biometrika* 1982; **69**:221–226.
5. Berger RL, Boos DD, Guess FM. Tests and confidence sets for comparing two mean residual life functions. *Biometrics* 1988; **44**:103–115.
6. Wang JL, Herrmansperger TP. Two-sample inference for median survival times based on one-sample procedures for censored survival data. *Journal of the American Statistical Association* 1990; **85**:529–536.
7. Su JQ, Wei LJ. Nonparametric estimation for the difference or ratio of median failure times. *Biometrics* 1993; **49**:603–607.
8. Jeong JH, Jung SH, Costantino JP. Nonparametric inference on median residual life function. *Biometrics* 2008; **64**:157–163.
9. Kaplan EL, Meier P. Nonparametric estimator from incomplete observations. *Journal of the American Statistical Association* 1958; **53**:457–481.
10. Ferguson TS. A Bayesian analysis of some nonparametric problems. *Annals of Statistics* 1973; **1**:209–230.
11. Ferguson TS. Prior distributions on spaces of probability measures. *Annals of Statistics* 1974; **2**:615–629.
12. Kuo L, Mallick B. Bayesian semiparametric inference for the accelerated failure-time model. *The Canadian Journal of Statistics* 1997; **25**:457–472.
13. Kottas A, Gelfand AE. Bayesian semiparametric median regression modeling. *Journal of the American Statistical Association* 2001; **96**:1458–468.
14. Gelfand AE, Kottas A. Bayesian semiparametric regression for median residual life. *Scandinavian Journal of Statistics* 2003; **30**:651–665.
15. Merrick JRW, Soyer R, Mazzuchi TA. A Bayesian semiparametric analysis of the reliability and maintenance of machine tools. *Technometrics* 2003; **45**:58–69.
16. Kottas A. Nonparametric Bayesian survival analysis using mixtures of Weibull distributions. *Journal of Statistical Planning and Inference* 2006; **136**:578–596.
17. Escobar MD. Estimating normal means with a Dirichlet process prior. *Journal of the American Statistical Association* 1994; **89**:268–277.
18. Escobar MD, West M. Bayesian density estimation and inference using mixtures. *Journal of the American Statistical Association* 1995; **90**:577–588.
19. Ishwaran H, James LF. Gibbs sampling methods for stick-breaking priors. *Journal of the American Statistical Association* 2001; **96**:161–173.
20. Gelfand AE, Kottas A. A computational approach for full nonparametric Bayesian inference under Dirichlet process mixture models. *The Journal of Computational and Graphical Statistics* 2002; **11**:289–305.
21. Koenker R, Bassett GJ. Regression quantiles. *Econometrica* 1978; **46**:33–50.
22. Walker SG, Mallick BK. A Bayesian semiparametric accelerated failure time model. *Biometrics* 1999; **55**:477–483.
23. Hanson T, Johnson WO. Modeling regression error with a mixture of poly trees. *Journal of the American Statistical Association* 2002; **97**:1020–1033.
24. Kottas A, Krnjajic M. Bayesian nonparametric modeling in quantile regression. *Scandinavian Journal of Statistics* 2009; **36**:297–319.

25. Taddy MA, Kottas A. A Bayesian nonparametric approach to inference for quantile regression. *Journal of Business and Economic Statistics* 2010; **28**:357–369.
26. Liu JS, Wong WH, Kong A. Covariance structure of the Gibbs sampler with applications to comparisons of estimators and augmentation schemes. *Biometrika* 1994; **81**:27–40.
27. van Dyk D, Park T. Partially collapsed Gibbs samplers: theory and methods. *Journal of the American Statistical Association* 2008; **103**:790–796.
28. van Dyk D, Meng XL. The art of data augmentation (with discussion). *Journal of Computational and Graphical Statistics* 2001; **10**:1–111.
29. Bagiella E, Heitjan DF. Predicting analysis times in randomized clinical trials. *Statistics in Medicine* 2001; **20**:2055–2063.
30. Ying GS, Heitjan DF. Weibull prediction of event times in clinical trials. *Pharmaceutical Statistics* 2008; **7**:107–120.
31. Sethuraman J. A constructive definition of Dirichlet priors. *Statistica Sinica* 1994; **4**:639–650.
32. Park T, van Dyk D. Partially collapsed Gibbs samplers: illustrations and applications. *Journal of Computational and Graphical Statistics* 2009; **18**:283–305.
33. Gilks WR, Wild P. Adaptive rejection sampling for Gibbs sampling. *Applied Statistics* 1992; **41**:337–348.
34. Gelman A, Rubin DB. Inference from iterative simulations using multiple sequences (with discussion). *Statistical Science* 1992; **7**:457–472.
35. Fisher B, Costantino J, Redmond C, Poisson R, Bowman D, Couture J, Dimitrov NV, Wolmark N, Wickerham DL, Fisher ER, *et al.* A randomized clinical trial evaluating tamoxifen in the treatment of patients with node-negative breast cancer who have estrogen-receptor-positive tumors. *New England Journal of Medicine* 1989; **320**:479–484.
36. Fisher B, Jeong JH, Anderson S, Bryant J, Fisher ER, Wolmark N. Twenty-five-year findings from a randomized clinical trial comparing radical mastectomy with total mastectomy and with total mastectomy followed by radiation therapy. *The New England Journal of Medicine* 2002; **347**:567–575.