Biostatistics (2018) **00**, 00, *pp.* 1–16
doi:10.1093/biostatistics/kxx075

Nonparametric Bayesian inference for mean residual life functions in survival analysis

VALERIE POYNOR

Department of Mathematics, California State University at Fullerton, 800 N State College Blvd, Fullerton, CA 92831, USA

ATHANASIOS KOTTAS*

Department of Applied Mathematics and Statistics, University of California at Santa Cruz, 1156 High Street, Santa Cruz, CA 95064, USA

thanos@soe.ucsc.edu

Summary

Modeling and inference for survival analysis problems typically revolves around different functions related to the survival distribution. Here, we focus on the mean residual life (MRL) function, which provides the expected remaining lifetime given that a subject has survived (i.e. is event-free) up to a particular time. This function is of direct interest in reliability, medical, and actuarial fields. In addition to its practical interpretation, the MRL function characterizes the survival distribution. We develop general Bayesian nonparametric inference for MRL functions built from a Dirichlet process mixture model for the associated survival distribution. The resulting model for the MRL function admits a representation as a mixture of the kernel MRL functions with time-dependent mixture weights. This model structure allows for a wide range of shapes for the MRL function. Particular emphasis is placed on the selection of the mixture kernel, taken to be a gamma distribution, to obtain desirable properties for the MRL function arising from the mixture model. The inference method is illustrated with a data set of two experimental groups and a data set involving right censoring. The supplementary material available at *Biostatistics* online provides further results on empirical performance of the model, using simulated data examples.

Keywords: Bayesian nonparametrics; Dirichlet process mixture models; Mean residual life; Right censoring; Survival function.

1. Introduction

Survival data describe the time to a particular event. The event may represent the failure of some machine, death of a person, relapse of a patient, duration of unemployment, or life expectancy of a product. The survival function of a random variable T with support on $\mathbb{R}^+$ defines the probability of survival beyond time t, $S(t) = \Pr(T > t) = 1 - F(t)$, where $F(t)$ is the distribution function. The hazard function computes the probability of a failure in the next instant given survival up to time t, $h(t) = \lim_{\Delta t \to 0} \Pr[t < T \leq t + \Delta t \mid T > t]/(\Delta t) = f(t)/S(t)$, with the expression in terms of the density function, $f(t)$, valid for continuous T.

*To whom correspondence should be addressed.

Our focus is on the mean residual life (MRL) function, which at any time point t defines the expected remaining survival time given survival up to time t. Provided $F(0) = 0$ and $\mu \equiv \mathrm{E}(T) = \int_0^\infty S(t)\mathrm{d}t < \infty$, the MRL function for continuous T is defined as:

$$m(t) = \mathrm{E}(T - t \mid T > t) = \frac{\int_t^\infty (u-t)f(u)\mathrm{d}u}{S(t)} = \frac{\int_t^\infty S(u)\mathrm{d}u}{S(t)} \tag{1.1}$$

with $m(t) \equiv 0$ whenever $S(t) = 0$. Note that $m(0) = \mu$. The MRL function is of particular interest in various application areas because of its easy interpretability (Guess and Proschan, 1985). Moreover, it characterizes the survival distribution via the "Inversion Formula" (Smith, 2002). More specifically, for continuous T with finite mean, the survival function is defined through the MRL function:

$$S(t) = \frac{m(0)}{m(t)} \exp\left[-\int_0^t \frac{1}{m(u)}\mathrm{d}u\right]. \tag{1.2}$$

Another important result is the characterization theorem (Hall and Wellner, 1981), which provides necessary and sufficient conditions such that a positive-valued function $m(t)$ defined on $\mathbb{R}^+$ is the MRL function for a survival distribution; the key conditions are that $m(t)$ is right-continuous, and that $m(t) + t$ is a non-decreasing function.

The form of the MRL function for various distributions has been studied in the reliability analysis literature. In Section 2, we review some results for the MRL function of standard parametric distributions. The shape of parametric MRL functions is often limited to be monotonically increasing or decreasing, which may not be suitable for certain applications. For instance, biological lifetime data tend to support lower MRL during infancy and elderly age while there is a higher MRL during the middle ages. The shape of such a MRL function is unimodal and commonly referred to as upside-down bathtub shape. Also well studied is the form of the MRL function in relation to the hazard function (e.g. Gupta and Akman, 1995; Finkelstein, 2002; Xie *and others*, 2004). In particular, if the hazard function is monotonically increasing (decreasing), then the corresponding MRL function is monotonically decreasing (increasing).

Regarding inference for MRL functions, the classical survival analysis literature includes several estimation techniques. The MRL function empirical estimate is defined by $\hat{m}_n(t) = (\int_t^\infty S_n(u)\mathrm{d}u)/S_n(t)$, for $t \in [0, T_{(n)}]$, where $S_n(t)$ is the empirical survival function and $T_{(n)}$ is the largest observed survival time (Yang, 1978). Abdous and Berred (2005) use a local linear fitting technique to find a smooth estimate, assuming a symmetric smoothing kernel. Berger *and others* (1988) develop a nonparametric hypothesis test for comparing two MRL functions. Classical estimation for the MRL function began to have a semiparametric regression flavor when Oakes and Dasu (1990) extended the class of distributions with linear MRL functions (Hall and Wellner, 1981) to a family having proportional MRL functions, $m_1(t) = \psi m_2(t)$, for $\psi > 0$. Maguluri and Zhang (1994) and Chen and Cheng (2005) further extended the proportional MRL model to a regression setting, $m(t; \boldsymbol{x}) = \exp(\boldsymbol{\psi x}) m_0(t)$, where $\boldsymbol{x}$ is the vector of covariates, $\boldsymbol{\psi}$ the vector of regression coefficients, and $m_0(t)$ a baseline MRL function.

There is by now a rich literature on Bayesian nonparametric methods for various survival analysis problems; see, for instance, Ibrahim *and others* (2001) and Müller *and others* (2015) for related references. However, in contrast to the classical literature, there has been very little work on Bayesian modeling and inference for MRL functions. Lahiri and Park (1991) present nonparametric Bayes and empirical Bayes estimators under a Dirichlet process (DP) prior (Ferguson, 1973) for the distribution function. They show that the Bayes estimator becomes a weighted average of the prior guess for the MRL function and the empirical MRL function of the data. Johnson (1999) discusses a Bayesian method for estimation of the MRL function under interval and right censored data, also using a DP prior for the corresponding survival function.

Our objective is to develop inference tools for MRL functions arising from a flexible probabilistic modeling framework. Under the Bayesian nonparametric approach to modeling, it is natural to consider defining nonparametric priors directly for the space of MRL functions. This is possible utilizing the characterization theorem, but there is a challenge in updating the prior to the posterior distribution given the data. The issue arises from the complicated fashion in which the MRL function enters the likelihood, as can be seen from equation (1.2). In order to achieve computational feasibility, the flexibility of possible MRL function shapes under the prior has to be substantially limited. We therefore instead build the inference approach from a mixture model for the density function of the survival distribution, with a parametric kernel density and a DP prior for the random mixing distribution. (Hence, the approach is similar to Gelfand and Kottas (2002) and Kottas (2006), where inference for survival and hazard functions was obtained through DP mixture priors for the density function.) Interestingly, this approach results in a prior model for the corresponding MRL function that retains interpretability as a mixture of the kernel MRL functions with time-dependent mixture weights. We place particular emphasis on the choice of the kernel for the DP mixture model to ensure a well-defined MRL function (thus, focusing on finiteness for the mean of the mixture distribution) and to achieve denseness of the mixture model in the space of MRL functions for continuous distributions. We develop approaches to prior specification and posterior simulation, and investigate empirically the inference method for MRL functions with both simulated and real data examples.

The outline of the article is as follows. To set the stage for the nonparametric model, in Section 2, we review properties of MRL functions for parametric distributions from the survival/reliability analysis literature. Section 3 develops the methodology, including discussion of model properties, prior specification, and posterior inference. In Section 4, we illustrate the modeling approach using two data examples that have been previously considered in the literature. Concluding remarks are given in Section 5. The supplementary material available at *Biostatistics* online includes additional details on prior specification and computation, as well as results from simulated data examples.

2. Review of parametric mean residual life functions

In this section, we collect the key results on the shape of MRL functions corresponding to common distributions and review some of the work on parametric distributions that have been defined to provide more flexible hazard and MRL functions.

The most basic shape for the MRL function is linear. Although this is evidently a restrictive assumption from a modeling perspective, it is of theoretical interest to identify distributions with linear MRL functions. Hall and Wellner (1981) studied the class of distributions with linear MRL functions, $m(t) = At + B$, where $A > -1$ and $B > 0$. Using (1.2), the corresponding survival function admits the form $S(t) = [B/(At + B)]_+^{1/A+1}$. When $A = 0$, the survival distribution is exponential with mean B. For $A > 0$, the survival function corresponds to a Pareto distribution under the linear transformation, $Z = AT + B$, with shape parameter $(1/A) + 1$ and scale parameter B. For $-1 < A < 0$, the survival function corresponds to a rescaled beta distribution under the transformation $Z = -AT$. Oakes and Dasu (1990) provided further characterizations for this family of distributions, including the result that if two survival functions have both proportional MRL functions and proportional hazard rate functions, then they have linear MRL functions.

The parametric distributions commonly used in survival and reliability analysis do not admit a closed form for their MRL function. However, using the expression, $m(t) = (\int_t^\infty uf(u)\mathrm{d}u/S(t)) - t$, obtained directly from (1.1), and/or transformations of T, the MRL function can be expressed in terms of standard integrals (e.g. Govil and Aggarwal, 1983; Gupta *and others*, 1999). This enables both ready evaluation of the function as well as study of its shape for different parameter combinations. Note that, even for standard distributions, the MRL function is not defined for all parameter combinations, an example being

Table 1. *Summary of the shape properties for the MRL function of four common parametric distributions (first four rows), and of the three-parameter exponentiated Weibull model (bottom row). Shapes are described as constant, increasing (INC), decreasing (DCR), bathtub (BT), or upside-down bathtub (UBT)*

Distribution	Constant	INC	DCR	BT	UBT
Gamma (γ, β) shape$\gamma > 0$, rate $\beta > 0$	$\gamma = 1$	$\gamma < 1$	$\gamma > 1$	—	—
Gompertz (γ, λ) shape $\gamma > 0$, scale $\lambda > 0$	—	—	$\forall\,(\gamma, \lambda)$	—	—
Lognormal (μ, σ) mean $\exp(\mu + 0.5\sigma^2)$, $\mu \in \mathbb{R}$, $\sigma > 0$	—	—	—	$\forall\,(\mu, \sigma)$	—
Weibull(γ, λ) shape $\gamma > 0$, scale $\lambda > 0$	$\gamma = 1$	$\gamma < 1$	$\gamma > 1$	—	—
ExpWeib(α, θ, σ) shape $\alpha > 0$, $\theta > 0$, scale $\sigma > 0$	$\alpha = 1$ $\theta = 1$	$\alpha < 1$ $\forall\,\theta$ $\alpha\theta < 1$	$\alpha > 1$ $\forall\,\theta$ $\alpha\theta > 1$	$\alpha < 1$ $\theta > 1$ $\alpha\theta > 1$	$\alpha > 1$ $\theta < 1$ $\alpha\theta < 1$

the log-logistic distribution with shape parameter less than or equal to 1. Table 1 summarizes results for four common survival distributions, indicating the restrictions imposed on the shape of the MRL function by selecting a particular parametric model. Among these models, the gamma and Weibull distributions are more versatile in terms of allowing both increasing and decreasing MRL functions, although the rate of increase/decrease is controlled by a single parameter and neither of the distributions allows for change points.

Several extensions of standard survival distributions have been considered to develop more flexible parametric models with respect to hazard rate and MRL function shapes; see, for instance, Pham and Lai (2007) for generalizations of the Weibull distribution. We focus here on the exponentiated Weibull distribution (Mudholkar and Strivasta, 1993) with survival function

$$S(t \mid \alpha, \theta, \sigma) = 1 - [1 - \exp\{-(t/\sigma)^{\alpha}\}]^{\theta}, \quad t > 0; \;\; \alpha > 0, \theta > 0, \sigma > 0 \tag{2.1}$$

where α and θ are shape parameters and σ is a scale parameter. As shown in Gupta and Akman (1995) and Xie *and others* (2004), the corresponding MRL function has various shapes, which are controlled by parameters α and θ and by their product; see Table 1. In Section 4.1, we compare the exponentiated Weibull distribution with the model proposed in the next section.

3. NONPARAMETRIC MIXTURE MODEL FOR MRL INFERENCE

Section 3.1 presents the nonparametric DP mixture model, including discussion for the choice of the kernel distribution, and study of key model properties. In Section 3.2, we briefly discuss prior specification, with more details included in the supplementary material available at *Biostatistics* online. Section 3.3 provides the techniques used to obtain posterior inference for the mixture distribution and the MRL function.

3.1. *Model formulation*

As discussed in the Introduction, it appears particularly difficult to develop a nonparametric prior model directly for the space of MRL functions that supports general MRL functional shapes and is feasible to

implement under a probabilistic inferential framework. We thus propose a model for the density of the survival distribution, building on the flexibility of nonparametric mixture models while focusing attention on properties of the resulting MRL function. More specifically, we model the density function of the survival distribution through

$$f(t \mid G) = \int k(t \mid \boldsymbol{\theta})\, \mathrm{d}G(\boldsymbol{\theta}), \quad t \in \mathbb{R}^+; \qquad G \sim \mathrm{DP}(\alpha, G_0) \tag{3.1}$$

where $k(t \mid \boldsymbol{\theta})$ is a kernel density on $\mathbb{R}^+$, with parameter vector $\boldsymbol{\theta}$. Here, $\mathrm{DP}(\alpha, G_0)$ denotes the DP prior for the mixing distribution G, defined in terms of the baseline (centering) distribution G_0 and total mass (precision) parameter $\alpha > 0$. Recall the DP constructive definition (Sethuraman, 1994), according to which a distribution G generated from $\mathrm{DP}(\alpha, G_0)$ is almost surely of the form $\sum_{l=1}^{\infty} w_l \delta_{\boldsymbol{\theta}_l}$, where the atoms $\boldsymbol{\theta}_l$ are independent and identically distributed (i.i.d.) from G_0, and the weights w_l are constructed through stick-breaking. In particular, $w_1 = v_1$ and $w_l = v_l \prod_{r=1}^{l-1}(1 - v_r)$, for $l \geq 2$, where the v_l are i.i.d. $\mathrm{Beta}(1, \alpha)$.

Based on the DP constructive definition, the density function can be expressed as $f(t \mid G) = \sum_{l=1}^{\infty} w_l k(t \mid \boldsymbol{\theta}_l)$, and the survival function $S(t \mid G) = \sum_{l=1}^{\infty} w_l S(t \mid \boldsymbol{\theta}_l)$, where $S(t \mid \boldsymbol{\theta})$ is the parametric survival function of the mixture kernel distribution. Then, using equation (1.1), we can obtain the implied model structure for the MRL function of the DP mixture:

$$m(t \mid G) = \frac{\int_t^{\infty} S(u \mid G)\, \mathrm{d}u}{S(t \mid G)} = \frac{\sum_{l=1}^{\infty} w_l \{\int_t^{\infty} S(u \mid \boldsymbol{\theta}_l)\, \mathrm{d}u\}}{\sum_{l=1}^{\infty} w_l S(t \mid \boldsymbol{\theta}_l)} = \sum_{l=1}^{\infty} q_l(t)\, m(t \mid \boldsymbol{\theta}_l) \tag{3.2}$$

where $m(t \mid \boldsymbol{\theta})$ is the kernel distribution MRL function, and $q_l(t) = w_l S(t \mid \boldsymbol{\theta}_l) / \{\sum_{r=1}^{\infty} w_r S(t \mid \boldsymbol{\theta}_r)\}$ are normalized weights defined through the DP weights adjusted by the kernel survival function at time t with parameter vector given by the corresponding atom $\boldsymbol{\theta}_l$. Hence, even though the prior model is not placed directly on the MRL function, we obtain an interpretable model structure for the mixture MRL function as a mixture of the kernel MRL functions with time-dependent weights. The latter enables local structure (in time) to be captured, and thus potentially a wide range of MRL functional shapes to be achieved by the model.

The remaining effort for the model formulation focuses on the choice of the DP mixture kernel. We note that various parametric families have been used to build DP mixture models for survival data analysis, including lognormal, Weibull, and gamma distributions (e.g. Kuo and Mallick, 1997; Kottas, 2006; Hanson, 2006), though none of this earlier work studied inference for MRL functions. Under our setting, the minimal requirements are a well-defined kernel MRL function (thus, we need kernel distributions with finite expectation) and a well-defined MRL function for the mixture distribution (thus, we need $\mathrm{E}(T \mid G) = \int_0^{\infty} S(t \mid G)\, \mathrm{d}t$ to be finite almost surely). We also study the concept of denseness of the mixture model in the space of MRL functions. In all cases, we considered standard parametric lifetime densities as possible choices for $k(t \mid \boldsymbol{\theta})$ with particular attention to the gamma and Weibull densities based on the review of Section 2. As detailed next, the gamma distribution emerges as our preferred choice for the DP mixture kernel.

A sufficient condition for finiteness of the expectation of the DP mixture distribution can be derived extending Theorem 3 of Ferguson (1973) to the DP mixture setting: if the expectation of the kernel distribution, $\mathrm{E}(T \mid \boldsymbol{\theta})$, is finite, and if $\int \mathrm{E}(T \mid \boldsymbol{\theta})\, \mathrm{d}G_0(\boldsymbol{\theta}) < \infty$, then $\mathrm{E}(T \mid G)$ is (almost surely) finite. The condition can be satisfied by a lognormal, Weibull, and gamma kernel density $k(t \mid \boldsymbol{\theta})$, but in the first two cases it requires restrictions on the parameter space for $\boldsymbol{\theta}$ which would potentially limit model flexibility and/or complicate posterior simulation.

Verifying the sufficient condition is much simpler for a gamma kernel density. Under the gamma distribution parameterization $\boldsymbol{\theta} = (\alpha_0, \beta_0)$ with mean α_0/β_0, it is easy to ensure finiteness for $\int\int \mathrm{E}(T \mid \alpha_0, \beta_0)\, \mathrm{d}G_0(\alpha_0, \beta_0)$, especially for a choice of G_0 that comprises independent components for α_0 and β_0. To encourage more efficient estimation of mixture components, we favor a dependent G_0. To facilitate such a choice, we consider the re-parameterization $\boldsymbol{\theta} \equiv (\theta, \phi) = (\log(\alpha_0), \log(\beta_0))$, and use a bivariate normal distribution for G_0 with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. Then, $\int\int \mathrm{E}(T \mid \theta, \phi)\, \mathrm{d}G_0(\theta, \phi)$ is recognized as the bivariate normal moment generating function at point $(1, -1)$, readily establishing the condition.

The choice of the gamma distribution for the kernel facilitates also the study of the support of the proposed model, which we explore through the concept of denseness. Let $\mathscr{F}$ represent the space of absolutely continuous distribution functions on $\mathbb{R}^+$ with finite mean. Formally, a class of distributions, $\mathcal{C}$, is said to be dense in $\mathscr{F}$, if for any distribution function, $F \in \mathscr{F}$, there exists a sequence of distribution functions, $\{F_n : n = 1, 2, ...\} \subseteq \mathcal{C}$, that converges to F. The type of convergence implies a measure of distance between the limiting sequence and F. In our context, the key result is the denseness of countable mixtures of Erlang distributions under weak convergence (e.g. Johnson and Taaffe, 1988; Lee and Lin, 2010). The specific result is included in the Appendix, but note that the Erlang distribution is a special case of the gamma distribution with the shape parameter constrained to take positive integer values. The literature includes also results on Kullback-Leibler support and posterior concentration rates specifically for gamma DP mixtures in density estimation (e.g. Wu and Ghosal, 2008; Bochkina and Rousseau, 2017).

More interesting from our prospective, however, is the denseness of the class of MRL functions arising from a gamma mixture for the corresponding density functions. In the Appendix, we show that for any MRL function of a continuous distribution, m, there exists a corresponding sequence of MRL functions for a mixture of gamma distributions, $\{m_n : n = 1, 2, ...\}$, such that, for any $t_0 \geq 0$, $\lim_{n\to\infty} m_n(t_0) = m(t_0)$, providing the following denseness result.

LEMMA. The set of MRL functions corresponding to gamma mixture distributions is dense, in the pointwise sense, in the space of MRL functions for continuous distributions on $\mathbb{R}^+$.

In the remainder of the article, we will refer to model (3.1), with gamma kernel $k(t \mid \theta, \phi) \propto t^{e^{\theta}-1} \exp(-e^{\phi} t)$, $(\theta, \phi) \in \mathbb{R}^2$, and $G_0(\theta, \phi) = \mathrm{N}_2(\theta, \phi \mid \boldsymbol{\mu}, \boldsymbol{\Sigma})$, as the gamma DPMM. The full Bayesian model is completed with a gamma(a_α, b_α) prior (with mean a_α/b_α) for the DP precision parameter α, and independent priors for the parameters of the DP centering distribution. In particular, we place a normal prior on the mean vector, $\boldsymbol{\mu} \sim \mathrm{N}_2(a_\mu, B_\mu)$, and an inverse-Wishart prior on the covariance matrix, $\boldsymbol{\Sigma} \sim \mathrm{IWish}(a_\Sigma, B_\Sigma)$, with mean $B_\Sigma/(a_\Sigma - 3)$ provided $a_\Sigma > 3$.

3.2. *Prior specification*

To specify the priors for the DP parameters, we assume the only available information involves a range, R, for the survival distribution. To reduce the number of hyperparameters, we take B_μ and B_Σ to be diagonal with the same diagonal element, b_μ and b_Σ, respectively. We also set $a_\Sigma = 4$, the smallest integer value that ensures finite prior expectation for $\boldsymbol{\Sigma}$, although we recommend larger values of a_Σ (thus, less dispersed priors) for particularly heavy tailed survival distributions; see the supplementary material available at *Biostatistics* online for more details on prior sensitivity analysis. We set the priors for the DP centering distribution parameters, $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$, by estimating the mean, $\mathrm{E}(T)$, and variance, $\mathrm{Var}(T)$, of the prior predictive distribution based on R. In particular, as detailed in the supplementary material available at *Biostatistics* online, we obtain approximations to $\mathrm{E}(T)$ and $\mathrm{Var}(T)$, and then set the mean equal to the midrange and estimate the variance by $(R/4)^2$. Note that the prior predictive density arises by taking the expectation of (3.1) with respect to the DP prior for G, which yields $\mathrm{E}\{f(t \mid G)\} = f(t \mid G_0) =$

$\int k(t \mid \theta, \phi)\, \mathrm{dN}_2(\theta, \phi \mid \boldsymbol{\mu}, \boldsymbol{\Sigma})$. The (marginal) prior predictive density is given by the expectation of $f(t \mid G_0)$ with respect to the prior distribution for $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$.

Regarding DP precision parameter α, we consider its connection with the number of distinct mixture components, n^*, which increases with α. In particular, for moderately large sample size n, $\mathrm{E}(n^* \mid \alpha) \approx \alpha \log\{(\alpha + n)/\alpha\}$, which can be used to suggest an appropriate range of α values.

This relatively automatic approach to prior specification is based on a small amount of prior information regarding the survival distribution. In general, we recommend studying the implied prior distribution for important survival functionals, in particular, obtaining prior point estimates and prior uncertainty bands for the MRL function.

3.3. *Posterior inference*

We use blocked Gibbs sampling (Ishwaran and Zarepour, 2000; Ishwaran and James, 2001) for posterior simulation, which is based on a truncation approximation, G_L, to G. In particular, $G_L = \sum_{l=1}^{L} p_l \delta_{\boldsymbol{\theta}_l}$, where $\boldsymbol{\theta}_l = (\theta_l, \phi_l) \overset{iid}{\sim} G_0$ for $l = 1, ..., L$, and $p_1 = v_1$, $p_l = v_l \prod_{r=1}^{l-1}(1 - v_r)$, where $v_r \overset{iid}{\sim} \mathrm{Beta}(1, \alpha)$ for $r = 1, ..., L-1$, and $p_L = 1 - \sum_{l=1}^{L-1} p_l$. The mixture model for the density function thus becomes $f(t \mid G_L) = \sum_{l=1}^{L} p_l k(t \mid \boldsymbol{\theta}_l)$. The truncation level can be chosen using standard DP properties. For instance, $\mathrm{E}(\sum_{l=1}^{L} w_l \mid \alpha) = 1 - \{\alpha/(\alpha+1)\}^L$, which can be averaged over the prior for α to estimate the prior expectation for the partial sum of the DP weights, $\mathrm{E}(\sum_{l=1}^{L} w_l)$. Then, L can be specified given any desired level of accuracy for the approximation.

The hierarchical model for the data, $\{t_i : i = 1, ..., n\}$, is augmented with configuration variables $\mathbf{w} = (\mathrm{w}_1, ..., \mathrm{w}_n)$ such that $\mathrm{w}_i = l$, for $l = 1, ..., L$, if and only if t_i is assigned to mixture component l. Let $\boldsymbol{p} = (p_1, ..., p_L)$ and $\boldsymbol{\theta} = \{(\theta_l, \phi_l) : l = 1, ..., L\}$. Then, the model is given by:

$$
\begin{aligned}
t_i \mid \boldsymbol{\theta}, \mathrm{w}_i &\overset{ind}{\sim} \mathrm{gamma}(t_i \mid e^{\theta_{\mathrm{w}_i}}, e^{\phi_{\mathrm{w}_i}}),\ i = 1, ..., n \\
\mathrm{w}_i \mid \boldsymbol{p} &\overset{iid}{\sim} \sum\nolimits_{l=1}^{L} p_l\, \delta_l(\mathrm{w}_i),\ i = 1, ..., n \\
\boldsymbol{p} \mid \alpha &\sim f(\boldsymbol{p} \mid \alpha) = \alpha^{L-1} p_L^{\alpha-1} (1 - p_1)^{-1} (1 - (p_1 + p_2))^{-1} \times ... \times (1 - \sum\nolimits_{l=1}^{L-2} p_l)^{-1} \\
(\theta_l, \phi_l) \mid \boldsymbol{\mu}, \boldsymbol{\Sigma} &\overset{iid}{\sim} \mathrm{N}_2((\theta_l, \phi_l) \mid \boldsymbol{\mu}, \boldsymbol{\Sigma}),\ l = 1, ..., L
\end{aligned}
$$

with the priors for $\boldsymbol{\mu}$, $\boldsymbol{\Sigma}$ and α given in Section 3.1. We use generic notation for the first stage of the model to account for censored observations. The contribution to the likelihood from each t_i is either a gamma kernel density ordinate (for observed survival times) or an appropriate integral of the density function (for censored observations).

Now, we can utilize the blocked Gibbs sampler to obtain samples from the posterior distribution $p(\boldsymbol{\theta}, \mathbf{w}, \boldsymbol{p}, \alpha, \boldsymbol{\mu}, \boldsymbol{\Sigma} \mid \text{data})$. At each Gibbs sampler iteration, we have an active number of mixture components, $n^* \leq L$, with the allocation of the data points to those components recorded in vector $\mathbf{w}$. The update for each (θ_l, ϕ_l) depends on whether l corresponds to an active component or not. In the latter case, (θ_l, ϕ_l) is drawn from the normal DP centering distribution given the currently imputed values for $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$. If l is an active component, the posterior full conditional for (θ_l, ϕ_l) involves a contribution from the first stage of the hierarchical model from all data points allocated to component l. In particular, for a data set that comprises observed and right censored survival times (indicated by $\zeta_i = 0$ if t_i is observed, and $\zeta_i = 1$ if t_i is right censored), the full conditional for (θ_l, ϕ_l) is proportional to $\mathrm{N}_2((\theta_l, \phi_l) \mid \boldsymbol{\mu}, \boldsymbol{\Sigma}) \prod_{\{i:\mathrm{w}_i = l\}} \{k(t_i \mid \theta_l, \phi_l)\}^{1-\zeta_i} \left\{ \int_{t_i}^{\infty} k(u \mid \theta_l, \phi_l) \mathrm{d}u \right\}^{\zeta_i}$. We use Metropolis-Hastings steps for these updates, with a bivariate normal proposal distribution centered on the current state for (θ_l, ϕ_l),

and with covariance matrix estimated from initial runs based on a product of two normals for the proposal distribution. The posterior full conditional for each w_i is a discrete distribution with values $l = 1, ..., L$ and associated probabilities proportional to $p_l \, \text{gamma}(t_i \mid e^{\theta_l}, e^{\phi_l})$, where again the specific form of $\text{gamma}(t_i \mid e^{\theta_l}, e^{\phi_l})$ depends on whether t_i is an observed or censored survival time. Based on their conditionally conjugate priors, the posterior full conditional for $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ is given by a normal and an inverse-Wishart distribution, respectively. Finally, $\boldsymbol{p}$ and α can be updated as detailed in Ishwaran and Zarepour (2000).

The posterior samples for $G_L \equiv (\boldsymbol{p}, \boldsymbol{\theta})$ can be used to obtain inference for the density, survival, and hazard functions at any time point t, by directly evaluating the expressions for these functions under the gamma DPMM. We next discuss two distinct methods for estimating the MRL function.

Following directly the MRL definition, the first approach is based on numerical integration of the survival function. Recalling that the MRL function at 0 returns the expected survival time, that is, $m(0) = \mu$, equation (1.1) can be re-written as

$$m(t) = \frac{\int_t^\infty S(u)\,\mathrm{d}u}{S(t)} = \frac{\int_0^\infty S(u)\,\mathrm{d}u - \int_0^t S(u)\,\mathrm{d}u}{S(t)} = \frac{\mu - \int_0^t S(u)\,\mathrm{d}u}{S(t)}. \tag{3.3}$$

We can thus avoid truncating the upper bound of the integration in the numerator of (1.1). We obtain posterior realizations for the MRL function by evaluating this expression at the posterior samples for the survival function. Since the survival function is monotone decreasing, the trapezoid technique is a natural method of approximating the integral $\int_0^t S(u)\mathrm{d}u$.

The approach above is generic as it is applicable to any prior probability model for the survival distribution. An alternative, and more numerically efficient, method utilizes the representation in (3.2) for the MRL function of the gamma DPMM. Under the DP truncation approximation, $m(t \mid G_L) = \sum_{l=1}^{L} q_l^*(t)\, m_\Gamma(t \mid \boldsymbol{\theta}_l)$, where $q_l^*(t) = p_l S_\Gamma(t \mid \boldsymbol{\theta}_l) / \{\sum_{r=1}^{L} p_r S_\Gamma(t \mid \boldsymbol{\theta}_r)\}$, with $S_\Gamma(t \mid \boldsymbol{\theta})$ and $m_\Gamma(t \mid \boldsymbol{\theta})$ denoting the survival and MRL function for the gamma kernel distribution. The gamma distribution MRL function can be written in terms of the Gamma function, $\Gamma(z) = \int_0^\infty u^{z-1} e^{-u}\mathrm{d}u$, and $S_\Gamma(t)$ (Govil and Aggarwal, 1983). Under our parameterization,

$$m_\Gamma(t \mid \theta, \phi) = \frac{t^{e^\theta} \exp(-e^\phi t) \exp\{\phi(e^\theta - 1)\}}{\Gamma(e^\theta) S_\Gamma(t \mid \theta, \phi)} + \exp(\theta - \phi) - t.$$

Hence, this approach provides MRL posterior realizations by evaluating $m(t \mid G_L)$ over the posterior samples for model parameters, and it thus overcomes the need for numerical integration.

4. DATA EXAMPLES

In Section 4.1, we fit the gamma DPMM as well as the exponentiated Weibull model to a data set involving survival times for subjects from two groups, including formal model comparison between the two models. In Section 4.2, we provide results of fitting the gamma DPMM to a data set of two groups both containing right censored survival times. In addition, the supplementary material available at *Biostatistics* online includes simulated data examples designed to illustrate the ability of the gamma DPMM to capture non-standard MRL function shapes, as well as to study the effect of the extent of censoring to inference results.

For all data examples, we followed the prior specification approach of Section 3.2, using a value for the range R which was taken to be about 2 times the data range. Moreover, we used the prior expectation for the partial sum of DP weights to set the DP truncation level (see Section 3.3). For instance, for the analysis of Section 4.1 we set $L = 50$, which yields $\text{E}(\sum_{l=1}^{50} w_l) = 0.99996$ under the gamma(2, 1) prior for α.

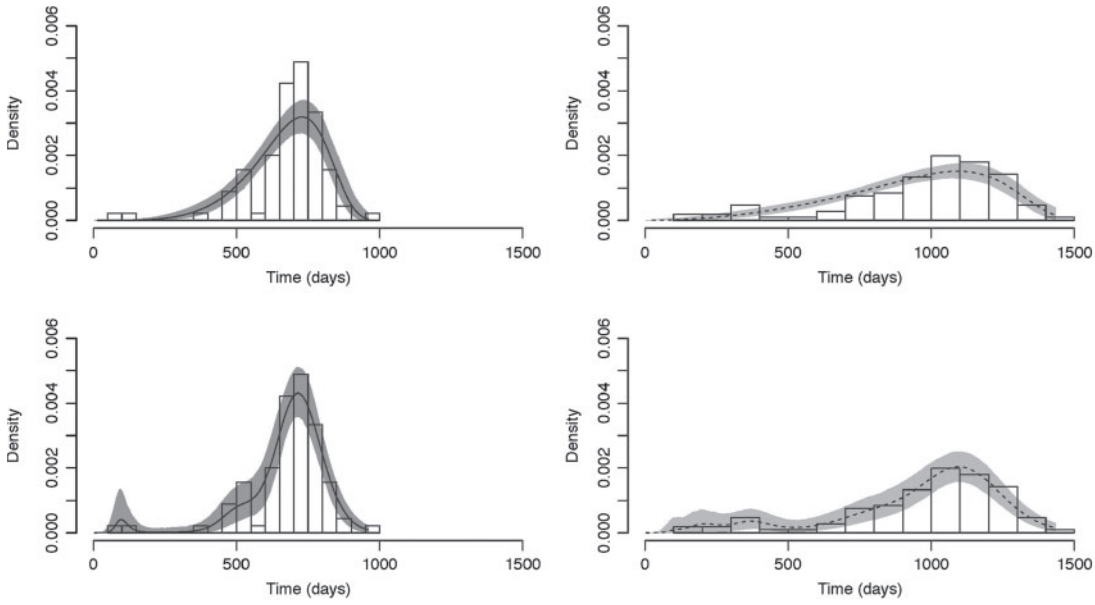

Fig. 1. Data from two experimental groups. Posterior point estimates and 95% uncertainty bands for the density function of the *ad libitum* group (left panels) and the restricted group (right panels), under the exponentiated Weibull model (top row) and the gamma DPMM (bottom row). The point estimates for the *ad libitum* group are plotted by solid lines and the interval estimates with dark gray bands. The corresponding estimates for the restricted group are given by dashed lines and light gray bands. Each panel includes the data histogram.

4.1. *Analysis of survival times from two experimental groups*

This data set, considered earlier in Berger *and others* (1988), is used to illustrate comparative inferences for two MRL functions. The data consists of survival times (in days) of rats from two experimental groups: the "*ad libitum* group" is comprised of 90 rats who were allowed to eat freely, whereas the "restricted group" includes 106 rats that were placed on a restricted diet.

We also use this data example to compare the gamma DPMM with the exponentiated Weibull model discussed in Section 2. Seeking to favor the parametric model in terms of the amount of prior information, we used a data-based approach to specify the priors for the three parameters of the exponentiated Weibull model, which are taken to be exponential distributions. In particular, we used the 10%, 50%, and 90% data quantiles to obtain a system of three equations from the distribution function, $p = [1 - \exp\{-(q/\sigma)^{\alpha}\}]^{\theta}$, where $p = 0.1, 0.5, 0.9$ and q is the corresponding data quantile. The (approximate) solution for this system of equations is used to specify the prior means for α, σ and θ. Posterior samples for the exponentiated Weibull model were obtained using a Metropolis-Hastings algorithm with a trivariate normal proposal distribution on the log-scale.

The gamma DPMM model was applied using $a_{\mu} = (4.1, 3.6)$, $B_{\mu} = B_{\Sigma} = \text{diag}(0.1, 0.1)$, and $a_{\alpha} = 2$, $b_{\alpha} = 1$ for the restricted group, and $a_{\mu} = (4.16, 3.8)$, $B_{\mu} = B_{\Sigma} = \text{diag}(0.095, 0.095)$, and $a_{\alpha} = 2$, $b_{\alpha} = 1$ for the *ad libitum* group.

Figure 1 plots point and interval estimates for the density of each of the two groups, under the gamma DPMM and the exponentiated Weibull model. The data from both groups suggest a heavy left tail for the underlying densities, and possibly a second, less pronounced mode. Restricted by its unimodal density shape, the parametric model is challenged in attempting to capture such local features. For both groups, its density estimates reach their left tail to the smaller observed survival times at the cost of underestimating

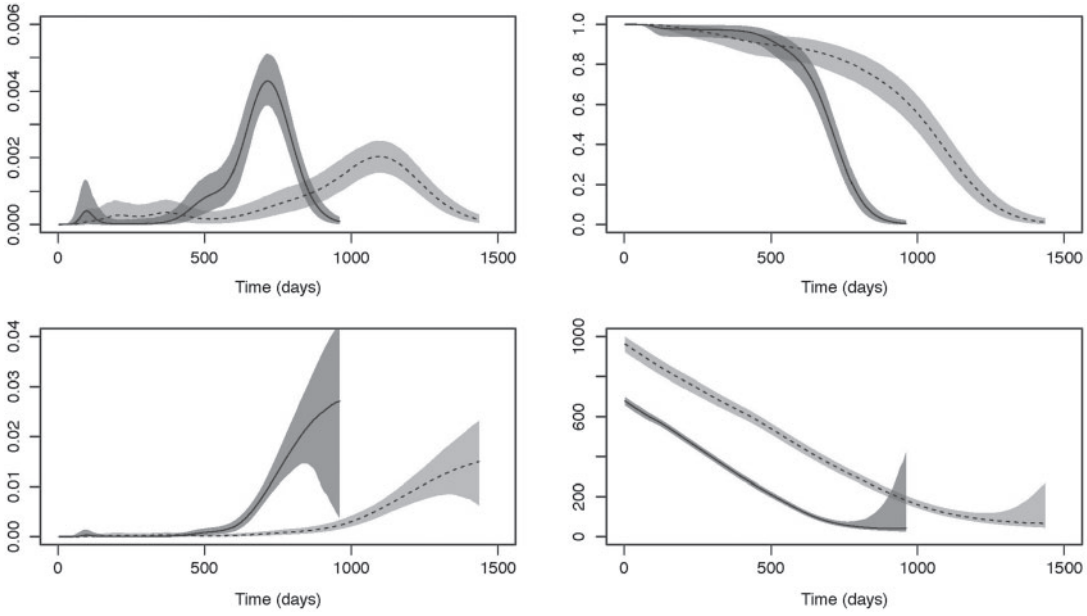

Fig. 2. Data from two experimental groups. Under the gamma DPMM, posterior point estimates and uncertainty bands for the density function (top left panel), survival function (top right panel), hazard function (bottom left panel), and MRL function (bottom right panel) for each group. The point estimates for the *ad libitum* group are plotted by solid lines and the interval estimates with dark gray bands. The corresponding estimates for the restricted group are given by dashed lines and light gray bands. The uncertainty bands for the MRL function are reported at the 80% posterior probability level, whereas for the other functions at the 95% level.

the density where most of the data lie, and overestimating the density where there is little or no data. The gamma DPMM density estimates are more successful in capturing the local features suggested by the data without compromising the quality of estimation in the time interval where the majority of the data fall.

To supplement the graphical evidence for the superiority of the nonparametric mixture model with more formal model comparison, we use the posterior predictive loss criterion from Gelfand and Ghosh (1998). The criterion favors the model $\mathcal{M}$ that minimizes the predictive loss measure $D_k(\mathcal{M}) = P(\mathcal{M}) + \{k/(k+1)\}G(\mathcal{M})$, where $P(\mathcal{M}) = \sum_{i=1}^{n} \mathrm{Var}^{\mathcal{M}}(t_{i,rep} \mid \text{data})$ is a penalty term, and $G(\mathcal{M}) = \sum_{i=1}^{n}\{t_i - \mathrm{E}^{\mathcal{M}}(t_{i,rep} \mid \text{data})\}^2$ is a goodness of fit term. Here, $\mathrm{E}^{\mathcal{M}}(t_{i,rep} \mid \text{data})$ and $\mathrm{Var}^{\mathcal{M}}(t_{i,rep} \mid \text{data})$ is the mean and variance of the posterior predictive distribution under model $\mathcal{M}$ for replicated response $t_{i,rep}$. The value of k provides the desired weight for the goodness of fit term relative to the penalty term. Note that for applications that do not involve covariates, such as the one here, the posterior predictive mean and variance is the same for all data points. Denote by $\mathcal{M}_1$ and $\mathcal{M}_2$ the exponentiated Weibull model and gamma DPMM, respectively. Then, for the *ad libitum* group we obtain $G(\mathcal{M}_1) = 1\,615\,787$, $P(\mathcal{M}_1) = 1\,568\,967$, and $G(\mathcal{M}_2) = 318\,919$, $P(\mathcal{M}_2) = 684\,342$. And, for the restricted group, $G(\mathcal{M}_1) = 8\,542\,725$, $P(\mathcal{M}_1) = 7\,917\,319$, and $G(\mathcal{M}_2) = 739\,435$, $P(\mathcal{M}_2) = 2\,247\,120$. Hence, regardless of the value of k, the gamma DPMM outperforms the exponentiated Weibull model for both groups.

To compare inference results for the *ad libitum* and restricted diet groups, we focus on the gamma DPMM. In Figure 2, we plot estimates for the density, survival, hazard, and MRL functions. Note that the interval estimates for the MRL function correspond to 80% probability bands as opposed to the other interval estimates which are 95% probability bands. The reason for this is to reduce the steepness for

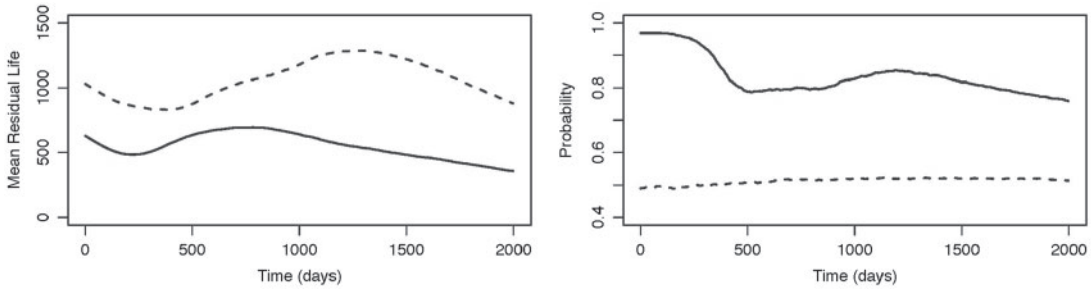

Fig. 3. Small cell lung cancer data. The left panel shows posterior point estimates for the MRL function of Arm A (dashed line) and Arm B (solid line). The right panel plots the posterior (solid line) and prior (dashed line) probability of the MRL function of Arm A being higher than the MRL function of Arm B, as a function of time.

which the upper bound increases toward the end of the range of the data. All the inference results support an improvement in survival time under the restricted diet. In particular, the MRL function point estimate for the restricted group is larger than the one for the *ad libitum* group throughout the effective range of survival times. The interval estimates also do not cross until about 800 days. The results strongly suggest that the remaining life expectancy for the restricted diet group is higher than the remaining life expectancy for the *ad libitum* group.

4.2. *Analysis of survival times of patients with small cell lung cancer*

For a data illustration involving right censoring, we fit the gamma DPMM to the survival times (in days) of patients with small cell lung cancer (Ying *and others*, 1995). The patients were randomly assigned to one of two treatments referred to as Arm A and Arm B. Arm A patients received cisplatin (P) followed by etoposide (E), while Arm B patients received (E) followed by (P). There were a total of 62 patients in Arm A with 15 right censored survival times, while Arm B consisted of 59 patients with eight right censored survival times. We fit the model independently to the two groups. For the Arm A data, we used $a_\mu = (2.5, -3)$, $B_\mu = B_\Sigma = \text{diag}(0.21, 0.21)$, and for the Arm B data, $a_\mu = (2.6, -2.9)$, $B_\mu = B_\Sigma = \text{diag}(0.21, 0.21)$. Moreover, we set $a_\alpha = 3$ and $b_\alpha = 1$.

The MRL function point estimates for the two treatment groups show Arm A to have a consistently higher MRL compared with Arm B (Figure 3, left panel), suggesting that the Arm A treatment is the more effective. For a more comprehensive inspection of the difference of MRL functions, we explore the posterior density of $m_A(t) - m_B(t)$, where $m_A(t)$ and $m_B(t)$ are the MRL functions corresponding to Arm A and Arm B, for a number of fixed time points. Figure 4 includes results for six time points, $t = 0, 100, 250, 500, 800$, and 1600 days. Time zero, which corresponds to the difference of the means, depicts a clear difference between the two treatments in favor of Arm A. The same can be observed at 100 days, and to a somewhat smaller extent, at 250 days. At the larger time points, although the indication of a difference is still present, it becomes less emphatic.

A further means of comparing the MRL functions of Arm A and Arm B is through the probability of the MRL function of one group being higher than that of the other, computed over a fine grid of survival times. The right panel of Figure 3 plots as a function of time the prior probability, $\Pr(m_A(t) > m_B(t))$, and the posterior probability, $\Pr(m_A(t) > m_B(t) \mid \text{data})$. The prior probability is relatively flat around 0.5, indicating a prior specification that does not favor either of the two groups. The posterior probabilities strongly suggest Arm A has the higher MRL function. The posterior probability is particularly high during the early time period, decreases slightly around 500 days, followed by another

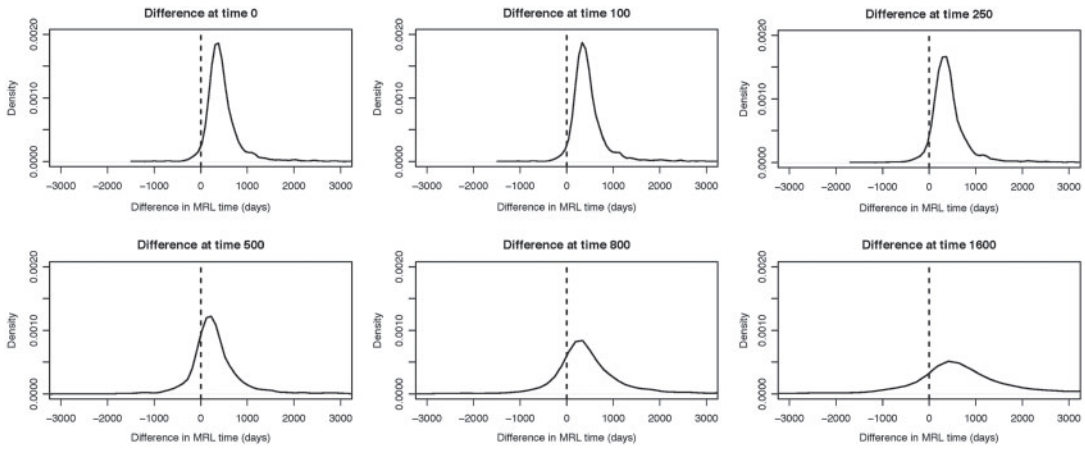

Fig. 4. Small cell lung cancer data. Posterior densities for the difference of the MRL functions between Arm A and Arm B, $m_A(t) - m_B(t)$, plotted for six time points, $t = 0, 100, 250, 500, 800$, and 1600 days.

peak around 1200 days. The posterior probability remains above 0.7 across the range of survival times in the data.

5. DISCUSSION

We have proposed a modeling approach to obtain inference for MRL functions. Although this functional is of key importance in survival analysis, it has received little attention in the Bayesian literature. The approach builds from a Dirichlet process (DP) mixture model for the survival distribution which yields flexible MRL functions as mixtures of the kernel MRL functions with time-dependent mixture weights. With the focus on inference for this particular functional, the choice of the mixture kernel plays an important role, and we have found the gamma kernel to possess the most desirable properties among the distributions we investigated. The practical utility of the proposed nonparametric mixture model was demonstrated through analysis of simulated data examples and real data sets from the literature.

The inclusion of covariates to the model is an important extension, both from a methodological and practical point of view. MRL regression modeling can be explored under the density regression framework (e.g. Müller *and others*, 1996; Taddy and Kottas, 2010; DeYoreo and Kottas, 2015) which builds from DP mixture modeling for the joint response-covariate distribution. This approach is natural for covariates, $\boldsymbol{x}$, that can be meaningfully modeled as random, and it provides an interesting extension of the model structure for the MRL function. Now, the DP mixture model for the response-covariate density is given by $f(t, \boldsymbol{x} \mid G) = \int k(t, \boldsymbol{x} \mid \boldsymbol{\theta})\, \mathrm{d}G(\boldsymbol{\theta})$, where the kernel density can be chosen following the considerations of Section 3.1; the simplest form involves a product kernel with independent components for the response and the covariates. Under the truncation approximation to the mixing distribution G, the covariate-dependent MRL function becomes $m(t \mid \boldsymbol{x}, G_L) = \sum_{l=1}^{L} q_l(t, \boldsymbol{x})\, m(t \mid \boldsymbol{x}, \boldsymbol{\theta}_l)$, where $m(t \mid \boldsymbol{x}, \boldsymbol{\theta})$ is the MRL function for the conditional response distribution of the mixture kernel (given by $m(t \mid \boldsymbol{\theta})$ under a product kernel), and $q_l(t, \boldsymbol{x}) = p_l k(\boldsymbol{x} \mid \boldsymbol{\theta}_l) S(t \mid \boldsymbol{x}, \boldsymbol{\theta}_l) / \{\sum_{r=1}^{L} p_r k(\boldsymbol{x} \mid \boldsymbol{\theta}_r) S(t \mid \boldsymbol{x}, \boldsymbol{\theta}_r)\}$ are covariate-dependent and time-dependent weights. Hence, the MRL form in (3.2) is extended to allow for local structure in time as well as across the covariate space. The DP mixture prior model can be further elaborated to incorporate dependence across experimental groups, such as treatment and control groups, using a dependent DP

prior for the group-specific mixing distributions. Poynor and Kottas (2017) reports results under this MRL regression modeling approach.

6. Software

Software, in the form of R code, to implement the gamma DPMM for the censored survival data example of Section 4.2 is available on GitHub (https://github.com/vpoynor/BNPInferenceForMRL).

Supplementary material

Supplementary material is available at http://biostatistics.oxfordjournals.org.

Acknowledgments

This research is part of the Ph.D. dissertation of the first author, completed at University of California, Santa Cruz. The authors wish to thank an Associate Editor and two reviewers for their positive feedback and for useful comments.

Conflict of Interest: None declared.

Funding

The work of the second author was supported in part by the National Science Foundation under award DMS 1310438.

Appendix

Here, we provide the proof of the Lemma of Section 3.1. The proof builds from the denseness of mixtures of Erlang distributions in the space of continuous distributions on $\mathbb{R}^+$. The structure of Erlang mixtures and the definition of the MRL function through the distribution function allow us to work with the limit of the approximating sequence of MRL functions to establish the pointwise convergence result of the Lemma.

Let $\mathscr{F}$ be the space of absolutely continuous distribution functions on $\mathbb{R}^+$ with finite mean, $\mu < \infty$, and $\mathscr{M}$ the space of MRL functions for continuous distributions on $\mathbb{R}^+$. Using (1.2), for any MRL function $m \in \mathscr{M}$, we can obtain the corresponding distribution function $F \in \mathscr{F}$.

Consider the class of countable gamma mixture distributions, $\mathcal{C}$, which is dense in $\mathscr{F}$. More specifically, for a generic $F \in \mathscr{F}$, consider a sequence of distribution functions, $\{F_n : n = 1, 2, ...\} \subseteq \mathcal{C}$, with $F_n(t) = \sum_{j=1}^{\infty} \{F(j/n) - F((j-1)/n)\} F_\Gamma(t \mid j, n)$, where $F_\Gamma(t \mid j, n)$ denotes the gamma distribution function with shape parameter j and mean j/n. Hence, each F_n is a countable mixture of Erlang distributions with the same rate parameter, mixing on the (integer) shape parameters, and with mixture weights defined through increments of the target distribution function F. Then, for any $t_0 > 0$, $\lim_{n\to\infty} F_n(t_0) = F(t_0)$, that is, the sequence $\{F_n : n = 1, 2, ...\}$ converges weakly (pointwise) to $F(t)$ (Johnson and Taaffe, 1988; Lee and Lin, 2010).

Denote by m_n the MRL function associated with F_n, and by m the MRL function of the target distribution function F. We seek to show that the sequence of MRL functions $\{m_n : n = 1, 2, ...\}$ converges pointwise to m. We first verify that m_n is well defined, that is, the expectation μ_n of F_n is finite. To this end, we have

$$\mu_n = \int_0^\infty \{1 - F_n(t)\}\, \mathrm{d}t = \int_0^\infty \sum_{j=1}^{\infty} \{F(j/n) - F((j-1)/n)\}\, (1 - F_\Gamma(t \mid j, n))\, \mathrm{d}t$$

$$= \sum_{j=1}^{\infty} \{F(j/n) - F((j-1)/n)\} \int_0^\infty (1 - F_\Gamma(t \mid j, n))\, \mathrm{d}t$$

$$= \sum_{j=1}^{\infty} \{F(j/n) - F((j-1)/n)\}(j/n) = \sum_{j=1}^{\infty} \int_{(j-1)/n}^{j/n} (j/n)\, \mathrm{d}F(t)$$

$$\leq \sum_{j=1}^{\infty} \int_{(j-1)/n}^{j/n} t\, \mathrm{d}F(t) + \sum_{j=1}^{\infty} \int_{(j-1)/n}^{j/n} n^{-1}\, \mathrm{d}F(t) = \mu + n^{-1} < \infty$$

where in the second line, we can exchange the order of integration and summation because the function involved takes positive values, and in the last line, we use the finite mean restriction for distribution function $F \in \mathscr{F}$. Analogously to deriving the upper bound for μ_n, we can also obtain a lower bound: $\mu_n = \sum_{j=1}^{\infty} \int_{(j-1)/n}^{j/n} (j/n)\, \mathrm{d}F(t) \geq \sum_{j=1}^{\infty} \int_{(j-1)/n}^{j/n} t\, \mathrm{d}F(t) = \mu$. Therefore, $\mu \leq \mu_n \leq \mu + n^{-1}$, which yields $\lim_{n\to\infty} \mu_n = \mu$, that is, $\lim_{n\to\infty} m_n(0) = m(0)$.

Next, consider a generic $t_0 > 0$. Using the MRL function definition in (3.3), we have

$$\lim_{n\to\infty} m_n(t_0) = \lim_{n\to\infty} \frac{\mu_n - \int_0^{t_0} \{1 - F_n(u)\}\, \mathrm{d}u}{1 - F_n(t_0)}$$

The limit can be distributed to the numerator and denominator, provided the corresponding limits exist (and the denominator limit is not zero). Based on the denseness result for the distribution functions, $\lim_{n\to\infty}\{1 - F_n(t_0)\} = 1 - F(t_0)$ (> 0). Moreover, using the dominated convergence theorem, $\lim_{n\to\infty} \int_0^{t_0} \{1 - F_n(u)\}\, \mathrm{d}u = \int_0^{t_0} \lim_{n\to\infty} \{1 - F_n(u)\}\, \mathrm{d}u = \int_0^{t_0} \{1 - F(u)\}\, \mathrm{d}u$. Therefore,

$$\lim_{n\to\infty} m_n(t_0) = \frac{\lim_{n\to\infty} \mu_n - \lim_{n\to\infty} \int_0^{t_0} \{1 - F_n(u)\}\, \mathrm{d}u}{\lim_{n\to\infty}\{1 - F_n(t_0)\}} = \frac{\mu - \int_0^{t_0} \{1 - F(u)\}\, \mathrm{d}u}{1 - F(t_0)} = m(t_0).$$

REFERENCES

ABDOUS, B. AND BERRED, A. (2005). Mean residual life estimation. *Journal of Statistical Planning and Inference* **132**, 3–19.

BERGER, R. L., BOOS, D. D. AND GUESS, F. M. (1988). Tests and confidence sets for comparing two mean residual life functions. *Biometrics* **44**, 103–115.

BOCHKINA, N. AND ROUSSEAU, J. (2017). Adaptive density estimation based on a mixture of Gammas. *Electronic Journal of Statistics* **11**, 916–962.

CHEN, Y. Q. AND CHENG, S. (2005). Semiparametric regression analysis of mean residual life with censored data. *Biometrika* **92**, 19–29.

DEYOREO, M. AND KOTTAS, A. (2015). A fully nonparametric modeling approach to binary regression. *Bayesian Analysis* **10**, 821–847.

FERGUSON, T. S. (1973). A Bayesian analysis of some nonparametric problems. *The Annals of Statistics* **1**, 209–230.

FINKELSTEIN, M. S. (2002). On the shape of the mean residual lifetime function. *Applied Stochastic Models in Business and Industry* **18**, 135–146.

GELFAND, A. E. AND GHOSH, S. K. (1998). Model choice: a minimum posterior predictive loss approach. *Biometrika* **85**, 1–11.

GELFAND, A. E. AND KOTTAS, A. (2002). A computational approach for full nonparametric Bayesian inference under Dirichlet process mixture models. *Journal of Computational and Graphical Statistics* **11**, 289–305.

GOVIL, K. K. AND AGGARWAL, K. K. (1983). Mean residual life function for normal, gamma and lognormal densities. *Reliability Engineering* **5**, 47–51.

GUESS, F. AND PROSCHAN, F. (1985). Mean residual life: theory and applications. *Technical Report* 85–178. Tallahassee, FL: North Carolina State University and Florida State University.

GUPTA, R. C. AND AKMAN, H. O. (1995). Mean residual life functions for certain types of non-monotonic ageing. *Communications in Statistics. Stochastic Models* **11**, 219–225.

GUPTA, R. C., AKMAN, O. AND LVIN, S. (1999). A study of log-logistic model in survival analysis. *Biometrical Journal* **41**, 431–443.

HALL, W. J. AND WELLNER, JON A. (1981). Mean residual life. In: Csörgö, M., Dawson, D. A., Rao, J. N. K. and Saleh, A. K. Md. E. (editors), *Statistics and Related Topics*. Amsterdam: North-Holland Publishing Company, pp. 169–184.

HANSON, T. E. (2006). Modeling censored lifetime data using a mixture of gammas baseline. *Bayesian Analysis* **1**, 575–594.

IBRAHIM, J. G., CHEN, M.-H. AND SINHA, D. (2001). *Bayesian Survival Analysis*. New York: Springer.

ISHWARAN, H. AND JAMES, L. F. (2001). Gibbs sampling methods for stick-breaking priors. *American Statistical Association* **96**, 161–173.

ISHWARAN, H. AND ZAREPOUR, M. (2000). Markov chain Monte Carlo in approximate Dirichlet and beta two-parameter process hierarchical models. *Biometrika* **87**, 371–390.

JOHNSON, M. A. AND TAAFFE, M. R. (1988). The denseness of Phase distributions. *Technical Report*. School of Industrial Engineering, Purdue University. Research Memorandum No 88-20.

JOHNSON, W. O. (1999). Survival analysis for interval data. In: Grambsch, P. and Geisser, S. (editors), *Institute for Mathematics and Its Applications: Statistics in the Health Sciences: Diagnosis and Prediction*, Volume 114. New York: Springer-Verlag, pp. 75–90.

KOTTAS, A. (2006). Nonparametric Bayesian survival analysis using mixtures of Weibull distributions. *Journal of Statistical Planning and Inference* **136**, 578–596.

KUO, L. AND MALLICK, B. (1997). Bayesian semiparametric inference for the accelerated failure-time model. *The Canadian Journal of Statistics* **25**, 457–472.

LAHIRI, P. AND PARK, D. H. (1991). Nonparametric Bayes and empirical Bayes estimators of mean residual life. *Journal of Statistical Planning and Inference* **29**, 125–136.

LEE, S. C. K. AND LIN, X. S. (2010). Modeling and evaluating insurance losses via mixtures of Erlang distributions. *North American Actuarial Journal* **14**, 107–130.

MAGULURI, G. AND ZHANG, C.-H. (1994). Estimation in the mean residual life regression model. *Journal of the Royal Statistical Society Series B* **56**, 477–489.

MUDHOLKAR, G. S. AND STRIVASTA, D. K. (1993). Exponentiated Weibull family for analyzing bathtub failure-rate data. *IEEE Transactions of Reliability* **42**, 299–302.

MÜLLER, P., ERKANLI, A. AND WEST, M. (1996). Bayesian curve fitting using multivariate normal mixtures. *Biometrika* **83**, 67–79.

MÜLLER, P., QUINTANA, F. A., JARA, A. AND HANSON, T. (2015). *Bayesian Nonparametric Data Analysis*. Cham, Switzerland: Springer.

OAKES, D. AND DASU, T. (1990). A note on mean residual life. *Biometrika* **77**, 409–410.

PHAM, H. AND LAI, C.-D. (2007). On recent generalizations of the Weibull distribution. *IEEE Transactions on Reliability* **56**, 454–458.

POYNOR, V. AND KOTTAS, A. (2017). Bayesian nonparametric modeling for mean residual life regression. *Technical Report* UCSC-SOE–17–08. Santa Cruz: Jack Baskin School of Engineering, University of California.

SETHURAMAN, J. (1994). A constructive definition of Dirichlet priors. *Statistica Sinica* **4**, 639–650.

SMITH, P. J. (2002). *Analysis of Failure and Survival Data*, Texts in Statistical Science Series. Boca Raton: Chapman & Hall/CRC.

TADDY, M. AND KOTTAS, A. (2010). A Bayesian nonparametric approach to inference for quantile regression. *Journal of Business and Economic Statistics* **28**, 357–369.

WU, Y. AND GHOSAL, S. (2008). Kullback Leibler property of kernel mixture priors in Bayesian density estimation. *Electronic Journal of Statistics* **2**, 298–331.

XIE, M., GOH, T. N. AND TANG, Y. (2004). On changing points of mean residual life and failure rate function for some generalized Weibull distributions. *Reliability Engineering and System Safety* **84**, 293–299.

YANG, G. L. (1978). Estimation of a biometric function. *The Annals of Statistics* **6**, 112–116.

YING, Z., JUNG, S.H. AND WEI, L.J. (1995). Survival analysis with median regression models. *Journal of the American Statistical Association* **90**, 178–184.

[*Received September 4, 2016; revised December 20, 2017; accepted for publication December 29, 2017*]