

Nonparametric Bayesian Inference for Mean Residual Life Functions in Survival Analysis

VALERIE POYNOR¹ and ATHANASIOS KOTTAS²

¹*California State University, Fullerton*

²*University of California, Santa Cruz*

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 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 mean residual life function characterizes the survival distribution. We review key properties of the mean residual life function and investigate its form for some common distributions. We then develop general Bayesian nonparametric inference for mean residual life functions built from a Dirichlet process mixture model for the associated survival distribution. The resulting model for the mean residual life function admits a representation as a mixture of the kernel mean residual life functions with time-dependent mixture weights. This model structure allows for a wide range of shapes for the mean residual life function. Particular emphasis is placed on the selection of the mixture kernel to obtain desirable properties for the mean residual life function arising from the mixture distribution. We advocate for a gamma mixture kernel and dependent baseline distribution for the Dirichlet process prior. The empirical performance of our modeling technique is studied with two simulation examples, a data set of two experimental groups, and a data set involving right censoring.

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 rate function computes the probability of a failure in the next instant given survival up to time t , $h(t) = \lim_{\Delta t \rightarrow 0} \Pr[t < T \leq t + \Delta t \mid T > t]/(\Delta t) = f(t)/S(t)$, with the expression in terms of the density function, $f(t)$, holding for continuous T .

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

$$m(t) = E(T - t \mid T > t) = \frac{\int_t^\infty (u - t)f(u)du}{S(t)} = \frac{\int_t^\infty S(u)du}{S(t)} \quad (1)$$

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 & 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:

$$S(t) = \frac{m(0)}{m(t)} \exp \left[- \int_0^t \frac{1}{m(u)} du \right]. \quad (2)$$

Another important result is the characterization theorem (Hall & 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.

One point of interest in the literature (mainly in reliability analysis) is study of the form for the MRL function of various distributions. 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 (INC) or decreasing (DCR), 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 (UBT) shape. Also well studied is the form of the MRL function in relation to the hazard function (e.g., Gupta & Akman, 1995; Finkelstein, 2002; Xie et al., 2004). In particular, if the hazard function is monotonically increasing (decreasing), then the corresponding MRL function is monotonically decreasing (increasing).

Another point of interest lies in inference for the MRL function. The classical survival analysis literature includes several estimation techniques for MRL functions. The MRL function empirical estimate is defined by $\hat{m}_n(t) = (\int_t^\infty S_n(u)du)/S_n(t)$, for $t \in [0, T_{(n)}]$, where $S_n(t)$ is the empirical survival function and $T_{(n)}$ is the largest observed survival time, and was studied by Yang (1978). Hall & Wellner (1979) extend the empirical estimator by defining it for values on the positive real line. Abdous & Berred (2005) use a local linear fitting technique to find a smooth estimate assuming only that the smoothing kernel is symmetric. A nonparametric hypothesis testing procedure for comparing MRL functions from two groups was introduced by Berger et al. (1988). Classical estimation for the MRL function began to have a semiparametric regression flavor when Oakes & Dasu (1990) extended the class of distributions with linear MRL functions (Hall & Wellner, 1981) to a family having proportional MRL functions, $m_1(t) = \psi m_2(t)$, for $\psi > 0$. Maguluri & Zhang (1994) further extended the proportional MRL model to a regression setting, $m(t; z) = \exp(\psi z)m_0(t)$, where z is the vector of covariates, ψ the vector of regression coefficients, and $m_0(t)$ a baseline MRL function. Chen & Cheng (2005) also extend the proportional MRL model to include inference for the regression parameters with censored data.

In contrast to the classical literature, there has been very little work on modeling and inference for MRL functions under the Bayesian framework. Lahiri & Park (1991) present nonparametric Bayes and empirical Bayes estimators under a Dirichlet process 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 Dirichlet process 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 (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 Dirichlet process prior for the random mixing distribution. 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 Dirichlet process 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 paper is as follows. To set the stage for the nonparametric model, in Section 2, we review properties of MRL functions corresponding to various 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 synthetic and real data sets. Finally, concluding remarks are given in Section 5.

2. REVIEW OF PARAMETRIC MRL 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 & Wellner (1981) studied the class of distributions with linear MRL functions, $m(t) = At + B$, where $A > -1$ and $B > 0$. Using (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 & 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 alternative expression, $m(t) = (\int_t^\infty uf(u)du/S(t)) - t$, which is obtained directly from (1), and/or transformations of T , the MRL function can be expressed in terms of standard integrals; see, for instance, Govilt & Aggarwal (1983) and Gupta et al. (1999). This enables both ready evaluation of the function as well as study of its shape for different combinations of the distribution parameters. Table 1 summarizes results for five 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 in the literature to develop more flexible parametric models with respect to hazard rate and MRL function shapes; see, for instance, Pham & Lai (2007) for generalizations of the Weibull distribution. We focus here on the exponentiated Weibull family of distributions (Mudholkar & Strivasta, 1993) with survival function given by

$$S(t | \alpha, \theta, \sigma) = 1 - [1 - \exp\{-(t/\sigma)^\alpha\}]^\theta, \quad t > 0; \quad \alpha, \theta, \sigma > 0 \quad (3)$$

where α and θ are shape parameters and σ is a scale parameter. As summarized in Table 2, the corresponding MRL function has various shapes, which are controlled by parameters α and θ as

Distribution	Density function	Mean residual life function
Gamma(γ, β) shape $\gamma > 0$, rate $\beta > 0$	$\frac{\beta^\gamma}{\Gamma(\gamma)} t^{\gamma-1} \exp(-\beta t)$	$\gamma < 1$ INC, $\gamma = 1$ constant (β^{-1}), $\gamma > 1$ DCR
Gompertz(γ, λ) shape $\gamma > 0$, scale $\lambda > 0$	$\lambda \gamma \exp\{\gamma + \lambda t - \gamma \exp(\lambda t)\}$	DCR for all γ
Loglogistic(γ, λ) shape $\gamma > 0$, scale $\lambda > 0$	$\frac{(\gamma/\lambda)(t/\lambda)^{\gamma-1}}{[1+(t/\lambda)^\gamma]^2}$	$\gamma \leq 1$ undefined, $\gamma > 1$ BT
Lognormal(μ, σ) location $\mu \in \mathbb{R}$, scale $\sigma > 0$	$\frac{1}{t\sqrt{2\pi\sigma^2}} \exp\{-(\log(t) - \mu)^2/(2\sigma^2)\}$	BT
Weibull(γ, λ) shape $\gamma > 0$, scale $\lambda > 0$	$\frac{\gamma}{\lambda} \left(\frac{t}{\lambda}\right)^{\gamma-1} \exp\{-(t/\lambda)^\gamma\}$	$\gamma < 1$ INC, $\gamma = 1$ constant (λ), $\gamma > 1$ DCR

Table 1: Summary of the shape properties for the MRL function of five common parametric distributions. Shapes are described as increasing (INC), decreasing (DCR), bathtub (BT), constant, or undefined.

α	θ	$\alpha\theta$	form of MRL function
1	1	1	exponential distribution \rightarrow constant
-	1	-	Weibull distribution \rightarrow monotone (INC, DCR or constant)
< 1	$\neq 1$	< 1	INC
> 1	$\neq 1$	> 1	DCR
> 1	< 1	< 1	UBT
< 1	> 1	> 1	BT

Table 2: Shape properties for the MRL function of the exponentiated Weibull distribution. Shapes are described as increasing (INC), decreasing (DCR), upside down bathtub (UBT), bathtub (BT) or constant.

well as by their product. Table 2 is developed based on results from Gupta & Akman (1995) and Xie et al. (2004), and it provides a more general version of a related table given in Mudholkar & Strivasta (1993) for the hazard rate function of the exponentiated Weibull distribution. In Section 4.2, we compare the exponentiated Weibull distribution with the Bayesian nonparametric model proposed in the next section.

3. NONPARAMETRIC MIXTURE MODEL FOR MRL INFERENCE

In this section, we discuss our modeling approach to obtaining inference for the MRL function. Section 3.1 presents the nonparametric Dirichlet process mixture model, including discussion for the choice of the kernel distribution, and study of key model properties. In Section 3.2, we discuss prior specification. Section 3.3 provides the techniques used to obtain posterior inference for the mixture distribution and functionals thereof.

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

$$f(t | G) = \int k(t | \boldsymbol{\theta}) dG(\boldsymbol{\theta}), \quad t \in \mathbb{R}^+; \quad G \sim \text{DP}(\alpha, G_0) \quad (4)$$

where $k(t | \boldsymbol{\theta})$ is a kernel density on \mathbb{R}^+ , with parameter vector $\boldsymbol{\theta}$. Here, $\text{DP}(\alpha, G_0)$ denotes the Dirichlet process (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 $\text{DP}(\alpha, G_0)$ 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. $\text{Beta}(1, \alpha)$.

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

$$m(t | G) = \frac{\int_t^{\infty} S(u | G) du}{S(t | G)} = \frac{\sum_{l=1}^{\infty} w_l \{\int_t^{\infty} S(u | \boldsymbol{\theta}_l) du\}}{\sum_{l=1}^{\infty} w_l S(t | \boldsymbol{\theta}_l)} = \sum_{l=1}^{\infty} q_l(t) m(t | \boldsymbol{\theta}_l) \quad (5)$$

where $m(t | \boldsymbol{\theta})$ is the kernel distribution MRL function, and $q_l(t) = w_l S(t | \boldsymbol{\theta}_l) / \{\sum_{r=1}^{\infty} w_r S(t | \boldsymbol{\theta}_r)\}$ 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 here 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 & 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 $E(T | G) = \int_0^\infty S(t | G) dt$ 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 $k(t | \theta)$ 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.

Starting with finiteness for the expectation of the DP mixture distribution, note that a sufficient condition can be derived extending Theorem 3 of Ferguson (1973) to the DP mixture setting. In particular, if the expectation of the kernel distribution, $E(T | \theta)$, is finite, and if $\int E(T | \theta) dG_0(\theta) < \infty$, then $E(T | G)$ is (almost surely) finite. The condition can be satisfied by a lognormal, Weibull and gamma kernel density $k(t | \theta)$, but in the first two cases it requires restrictions on the parameter space for θ 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 $\theta = (\alpha_0, \beta_0)$ with mean α_0/β_0 , it is easy to ensure finiteness for $\iint E(T | \alpha_0, \beta_0) dG_0(\alpha_0, \beta_0)$, 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 $\theta \equiv (\theta, \phi) = (\log(\alpha_0), \log(\beta_0))$, and use a bivariate normal distribution for G_0 with mean vector μ and covariance matrix Σ . Then, $\iint E(T | \theta, \phi) dG_0(\theta, \phi)$ 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 \mathcal{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 \mathcal{F} , if for any distribution function, $F \in \mathcal{F}$, there exists a sequence of distribution functions, $\{F_n : n = 1, 2, \dots\} \subseteq \mathcal{C}$, 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; see, for instance, Johnson & Taaffe (1988) and Lee & 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. 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, $\{m_n : n = 1, 2, \dots\}$, such that, for any $t_0 \geq 0$, $\lim_{n \rightarrow \infty} m_n(t_0) = m(t_0)$, providing the following denseness result.

Lemma. The set of MRL functions corresponding with the class of 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 paper, we will refer to model (4), with gamma kernel $k(t \mid \theta, \phi) \propto t^{e^\theta - 1} \exp(-e^\phi t)$, $(\theta, \phi) \in \mathbb{R}^2$, and $G_0(\theta, \phi) = N_2(\theta, \phi \mid \boldsymbol{\mu}, \boldsymbol{\Sigma})$, 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, $\boldsymbol{\mu} \sim N_2(a_\mu, B_\mu)$, and an inverse-Wishart prior on the covariance matrix, $\boldsymbol{\Sigma} \sim \text{IWish}(a_\Sigma, B_\Sigma)$, with mean $B_\Sigma/(a_\Sigma - 3)$ provided $a_\Sigma > 3$.

3.2 Prior specification

Here, we discuss specification of the priors for the DP parameters assuming the only available information involves a rough range, R , for the survival distribution. We set the priors for the DP centering distribution parameters, $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$, by estimating the mean, $E(T)$, and variance, $\text{Var}(T)$, of the prior predictive distribution based on R . In particular, we obtain approximations to $E(T)$ and $\text{Var}(T)$, and then set the mean equal to the midrange and estimate the variance by $(R/4)^2$. Note that the prior predictive density conditional on $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ arises by taking the expectation of (4) with respect to the DP prior for G , which yields $E\{f(t \mid G)\} = f(t \mid G_0) = \int k(t \mid \theta, \phi) dN_2(\theta, \phi \mid \boldsymbol{\mu}, \boldsymbol{\Sigma})$.

The (marginal) prior predictive density is defined by the expectation of $f(t | G_0)$ with respect to the prior distribution for $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$.

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 $a_\Sigma = 4$, which is the smallest integer value that ensures finite prior expectation for the inverse-Wishart prior on $\boldsymbol{\Sigma}$. Using the expression for the bivariate normal moment generating function at points $t_1 = (1, -2)'$, $t_2 = (2, -2)'$, $t_3 = (1, -1)'$, and the prior independence for $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$, we approximate the marginal variance as follows:

$$\begin{aligned}
\text{Var}(T) &= \text{E}\{\text{Var}(T | \boldsymbol{\mu}, \boldsymbol{\Sigma})\} + \text{Var}\{\text{E}(T | \boldsymbol{\mu}, \boldsymbol{\Sigma})\} \\
&= \text{E}\{\text{Var}(\text{E}(T | \theta, \phi)) + \text{E}(\text{Var}(T | \theta, \phi))\} + \text{Var}\{\text{E}(\text{E}(T | \theta, \phi))\} \\
&= \text{E}\{\text{Var}(e^{\theta-\phi} | \boldsymbol{\mu}, \boldsymbol{\Sigma})\} + \text{E}\{\text{E}(e^{\theta-2\phi} | \boldsymbol{\mu}, \boldsymbol{\Sigma})\} + \text{Var}\{\text{E}(e^{\theta-\phi} | \boldsymbol{\mu}, \boldsymbol{\Sigma})\} \quad (6) \\
&= \text{E}(e^{t_2' \boldsymbol{\mu}}) \text{E}(e^{0.5 t_2' \boldsymbol{\Sigma} t_2}) + \text{E}(e^{t_1' \boldsymbol{\mu}}) \text{E}(e^{0.5 t_1' \boldsymbol{\Sigma} t_1}) - \{\text{E}(e^{t_3' \boldsymbol{\mu}})\}^2 \{\text{E}(e^{0.5 t_3' \boldsymbol{\Sigma} t_3})\}^2 \\
&\approx \exp(t_2' a_\mu + 0.5 t_2' B_\mu t_2 + 0.5 t_2' B_\Sigma t_2) + \exp(t_1' a_\mu + 0.5 t_1' B_\mu t_1 + 0.5 t_1' B_\Sigma t_1) \\
&\quad - \exp(2 t_3' a_\mu + t_3' B_\mu t_3 + t_3' B_\Sigma t_3)
\end{aligned}$$

where the last expression is obtained by using again the bivariate normal moment generating function, and by approximating $\text{E}\{\exp(0.5 t' \boldsymbol{\Sigma} t)\}$ with $\exp\{0.5 t' \text{E}(\boldsymbol{\Sigma}) t\}$. Analogously, for the marginal mean: $\text{E}(T) \approx \exp(t_3' a_\mu + 0.5 t_3' B_\mu t_3 + 0.5 t_3' B_\Sigma t_3) = \exp(a_{\mu_1} - a_{\mu_2} + b_\mu + b_\Sigma)$, where $a_\mu = (a_{\mu_1}, a_{\mu_2})'$. We now have two equations, $\text{E}(T) = \text{midrange}$ and $\text{Var}(T) = (R/4)^2$, with four unknowns, a_{μ_1} , a_{μ_2} , b_μ and b_Σ . We can finally specify the prior hyperparameters by, for instance, allocating portions of $(R/4)^2$ to the different terms in the variance decomposition in (6), or by setting each term to $(R/4)^2$ to induce a more dispersed prior predictive distribution.

Regarding the prior for the DP precision parameter α , we consider its connection with the number of distinct mixture components, n^* , where n^* increases with α . In particular, for moderately large sample size n , $\text{E}(n^* | \alpha) \approx \alpha \log\{(\alpha + n)/\alpha\}$, 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 & Zarepour, 2000; Ishwaran & James, 2001) for posterior simulation, which is based on a truncation approximation, G_L , to the mixing distribution G . In particular, $G_L = \sum_{l=1}^L p_l \delta_{\theta_l}$, where $\theta_l \stackrel{iid}{\sim} G_0$ for $l = 1, \dots, L$, and $p_1 = v_1$, $p_l = v_l \prod_{r=1}^{l-1} (1 - v_r)$, where $v_r \stackrel{iid}{\sim} \text{Beta}(1, \alpha)$ for $r = 1, \dots, L - 1$, and $p_L = 1 - \sum_{l=1}^{L-1} p_l$. The mixture model for the density function thus becomes $f(t | G_L) = \sum_{l=1}^L p_l k(t | \theta_l)$. The truncation level can be chosen using standard DP properties. For instance, $E(\sum_{l=1}^L w_l | \alpha) = 1 - \{\alpha/(\alpha + 1)\}^L$, which can be averaged over the prior for α to estimate the prior expectation for the partial sum of the original DP weights, $E(\sum_{l=1}^L w_l)$. Then, L can be specified given any desired level of accuracy for the approximation.

The hierarchical model for the data, $\{t_i : i = 1, \dots, n\}$, is augmented with configuration variables $\mathbf{w} = (w_1, \dots, w_n)$ such that $w_i = l$, for $l = 1, \dots, L$, if and only if t_i is assigned to mixture component l . Then, the full model can be expressed as

$$\begin{aligned}
 t_i | \boldsymbol{\theta}, \mathbf{w}_i &\stackrel{iid}{\sim} \Gamma(t_i | e^{\theta_{w_i}}, e^{\phi_{w_i}}), \quad i = 1, \dots, n \\
 \mathbf{w}_i | \mathbf{p} &\stackrel{iid}{\sim} \sum_{l=1}^L p_l \delta_l(w_i), \quad i = 1, \dots, n \\
 \mathbf{p} | \alpha &\sim f(\mathbf{p} | \alpha) \\
 (\theta_l, \phi_l) | \boldsymbol{\mu}, \boldsymbol{\Sigma} &\stackrel{iid}{\sim} N_2((\theta_l, \phi_l) | \boldsymbol{\mu}, \boldsymbol{\Sigma}), \quad l = 1, \dots, L \\
 \alpha, \boldsymbol{\mu}, \boldsymbol{\Sigma} &\stackrel{iid}{\sim} \text{gamma}(\alpha | a_\alpha, b_\alpha) N_2(\boldsymbol{\mu} | a_\mu, B_\mu) \text{IWish}(\boldsymbol{\Sigma} | a_\Sigma, B_\Sigma)
 \end{aligned}$$

where $\mathbf{p} = (p_1, \dots, p_L)$, $\boldsymbol{\theta} = \{(\theta_l, \phi_l) : l = 1, \dots, L\}$, and $f(\mathbf{p} | \alpha) = \alpha^{L-1} p_L^{\alpha-1} (1 - p_1)^{-1} (1 - (p_1 + p_2))^{-1} \times \dots \times (1 - \sum_{l=1}^{L-2} p_l)^{-1}$ is a special case of the generalized Dirichlet distribution (Connor & Mosimann, 1969). 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

$p(\boldsymbol{\theta}, \mathbf{w}, \mathbf{p}, \alpha, \boldsymbol{\mu}, \boldsymbol{\Sigma} \mid \text{data})$. 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 $N_2((\theta_l, \phi_l) \mid \boldsymbol{\mu}, \boldsymbol{\Sigma}) \prod_{\{i: w_i=l\}} \{k(t_i \mid \theta_l, \phi_l)\}^{1-\zeta_i} \left\{ \int_{t_i}^{\infty} k(u \mid \theta_l, \phi_l) du \right\}^{\zeta_i}$. 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 $l = 1, \dots, L$ and associated probabilities proportional to $p_l \Gamma(t_i \mid e^{\theta_l}, e^{\phi_l})$, where again the specific form of $\Gamma(t_i \mid e^{\theta_l}, e^{\phi_l})$ 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, \mathbf{p} and α can be updated as detailed in Ishwaran & Zarepour (2000).

The posterior samples for $G_L \equiv (\mathbf{p}, \boldsymbol{\theta})$ 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. Obtaining the MRL function must be done by numerical integration approximation for the integral over the survival function. Since the MRL function at 0 returns the expected survival time, that is, $m(0) = \mu$, equation (1) can be re-written as

$$m(t) = \frac{\int_t^{\infty} S(u) du}{S(t)} = \frac{\int_0^{\infty} S(u) du - \int_0^t S(u) du}{S(t)} = \frac{\mu - \int_0^t S(u) du}{S(t)}. \quad (7)$$

We can thus avoid having to truncate the upper bound of the integration in the numerator in (1) by using the form above for the MRL function. We obtain posterior point and interval estimates for the MRL function by evaluating this expression at the posterior samples. We do this over a grid of pre-specified time points, z_j for $j = 1, \dots, m$. The survival function is monotone decreasing, so the trapezoid technique is an appropriate method of approximating the integral $\int_0^t S(u) du$. We

evaluate the MRL function at the first grid point by $m(z_1 | G_L) = [E(T | G_L) - 0.5(z_1(1 + S(z_1 | G_L)))]/S(z_1 | G_L)$ and use the following expression for $j = 2, \dots, m$:

$$m(z_j | G_L) = \frac{E(T | G_L) - \frac{1}{2} \left(z_1(1 + S(z_1 | G_L)) + \sum_{i=2}^j (z_j - z_{j-1})(S(z_j | G_L) + S(z_{j-1} | G_L)) \right)}{S(z_j | G_L)}.$$

We save a lower and upper quantile along with the median from the posterior sample at each grid point for each mixture functional to obtain (point-wise) posterior point and interval estimates.

As a final remark, we note that there are alternative posterior simulation methods that avoid truncating the mixing distribution G , including marginal samplers based on the version of the model resulting from integrating G over its DP prior (e.g., Neal, 2000), as well as slice sampling algorithms (e.g., Kalli et al., 2011). These methods are arguably preferable to the blocked Gibbs sampler if the inference objectives are limited to posterior predictive estimation, which does not require posterior samples for G . Evidently, for MRL function inference, posterior simulation must be extended to the mixing distribution, and thus truncation of G is necessary also under the alternative methods. For instance, if the model is fitted with a marginal sampler, such truncation can be developed based on the conditional posterior distribution for G (Gelfand & Kottas, 2002; Ishwaran & Zarepour, 2002).

4. DATA EXAMPLES

In Section 4.1, we use two simulated data examples to illustrate the ability of the gamma DPMM to capture non-standard MRL function shapes as well as the correlation between kernel parameters θ and ϕ . In Section 4.2, we fit a gamma DPMM as well as an 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 4.3, we provide results of fitting the gamma DPMM to a data set of two groups both containing right censored survival times.

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 4.2 we set $L = 50$, which yields $E(\sum_{l=1}^{50} w_l) = 0.99996$ under the gamma(2, 1) prior for α .

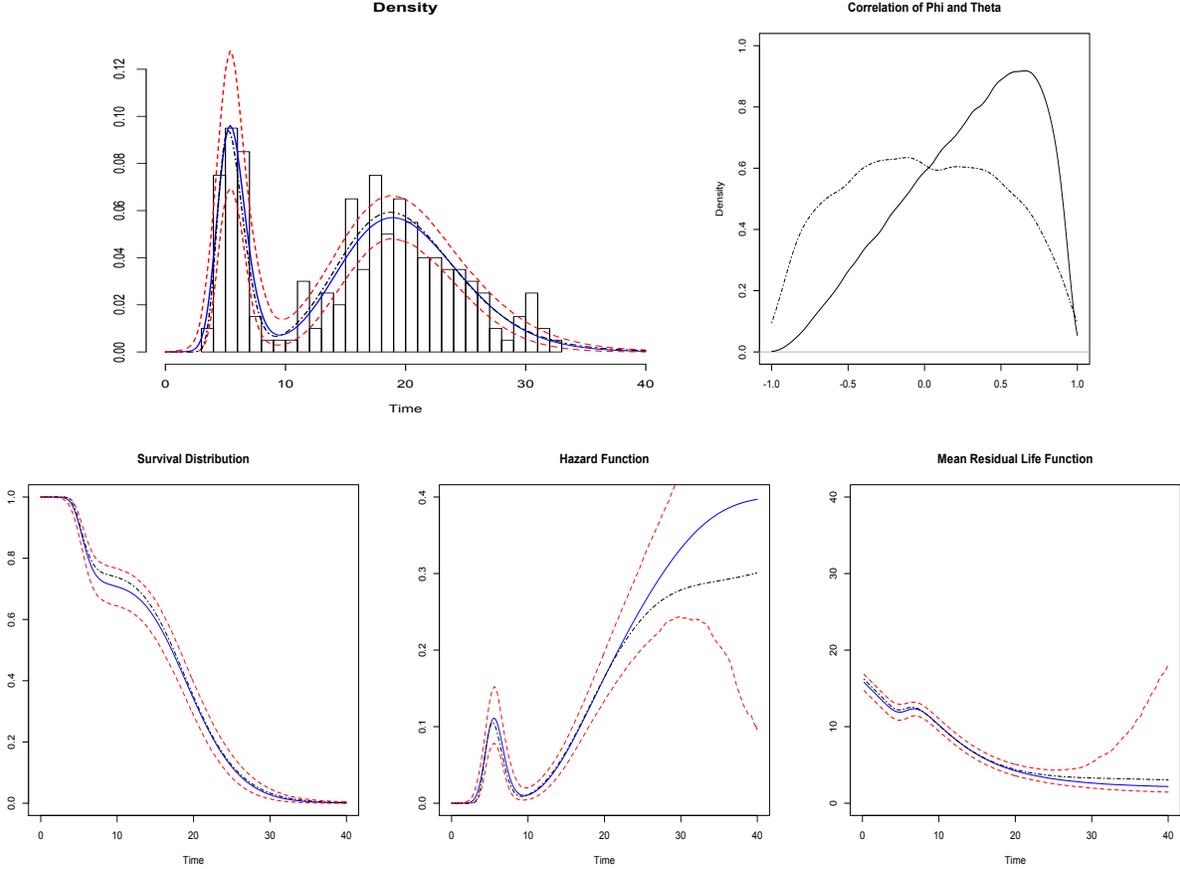


Figure 1: Simulation example 1. Posterior point (solid line) and 95% interval estimates (dashed lines) for the density function (top left panel, overlaying the sample histogram), survival function (bottom left panel), hazard function (bottom middle panel), and MRL function (bottom right panel). In all cases, the black dot-dashed line indicates the corresponding function based on the distribution from which the data were generated. The top right panel plots the posterior density (solid line) and prior density (dot-dashed line) of the correlation between θ and ϕ in the DP centering distribution.

4.1 Simulation examples

The first synthetic data set consists of 200 draws from a mixture distribution of four gamma components in which the shape and scale parameters are positively associated: $0.35\Gamma(10, 0.5) + 0.4\Gamma(20, 1) + 0.15\Gamma(30, 5) + 0.1\Gamma(40, 8)$. We fit the gamma DPMM with priors $a_\mu = (1.6, 0.4)$, $B_\mu = B_\Sigma = \text{diag}(0.39, 0.39)$, and $a_\alpha = 2$, $b_\alpha = 1$. Results are shown in Figure 1. Each of the density, survival, hazard, and MRL function of the true underlying distribution are contained within the gamma DPMM interval estimates, and the point estimate recovers well the shape of the true function within the range of the data. Moreover, the model captures the correlation between the

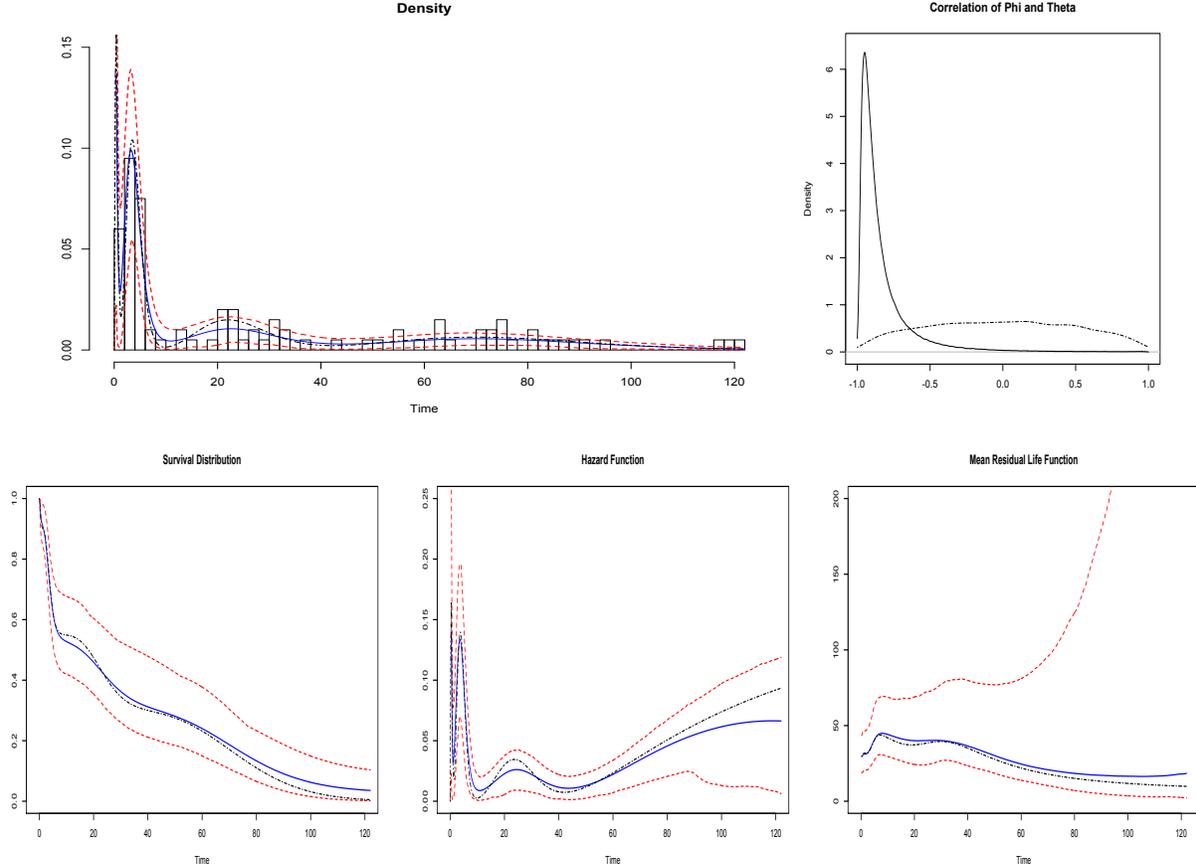


Figure 2: Simulation example 2. Posterior point (solid line) and 95% interval estimates (dashed lines) for the density function (top left panel, overlaying the sample histogram), survival function (bottom left panel), hazard function (bottom middle panel), and MRL function (bottom right panel). In all cases, the black dot-dashed line indicates the corresponding function based on the distribution from which the data were generated. The top right panel plots the posterior density (solid line) and prior density (dot-dashed line) of the correlation between θ and ϕ in the DP centering distribution.

kernel parameters, even though the prior does not favor either negative or positive correlation.

The second data set consists of 100 draws from a gamma mixture distribution with negative correlation between the shape and rate parameters: $0.3\Gamma(15, 0.2) + 0.25\Gamma(12, 0.5) + 0.35\Gamma(8, 2) + 0.1\Gamma(3, 6)$. This population was chosen to test the model's ability to estimate modes that are close together, as well as to capture the tail of a heavy-tailed distribution. A gamma DPMM was fit to the data with priors $a_\mu = (2.4, -1)$, $B_\mu = B_\Sigma = \text{diag}(0.18, 0.18)$, and $a_\alpha = 2$, $b_\alpha = 1$. Results are shown in Figure 2. Again, the model interval estimates contain the true density, survival, hazard, and MRL functional. The point estimates are close to the truth, even in the tail where less data is

available. The uncertainty bands for the MRL estimate show a wider upper bound, which is likely an effect of the smaller sample size and the sparsity of the data from the tail of the underlying distribution. The model estimates a strong negative correlation between the kernel parameters.

4.2 Analysis of survival times from two experimental groups

This data set, considered earlier in Berger et al. (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 first group (ad libitum group) is comprised of 90 rats who were allowed to eat freely, whereas the second group (restricted group) includes 106 rats that were placed on a restricted diet.

We also use this data example to compare results from the gamma DPMM and 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% quantiles of the data 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 3 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 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.

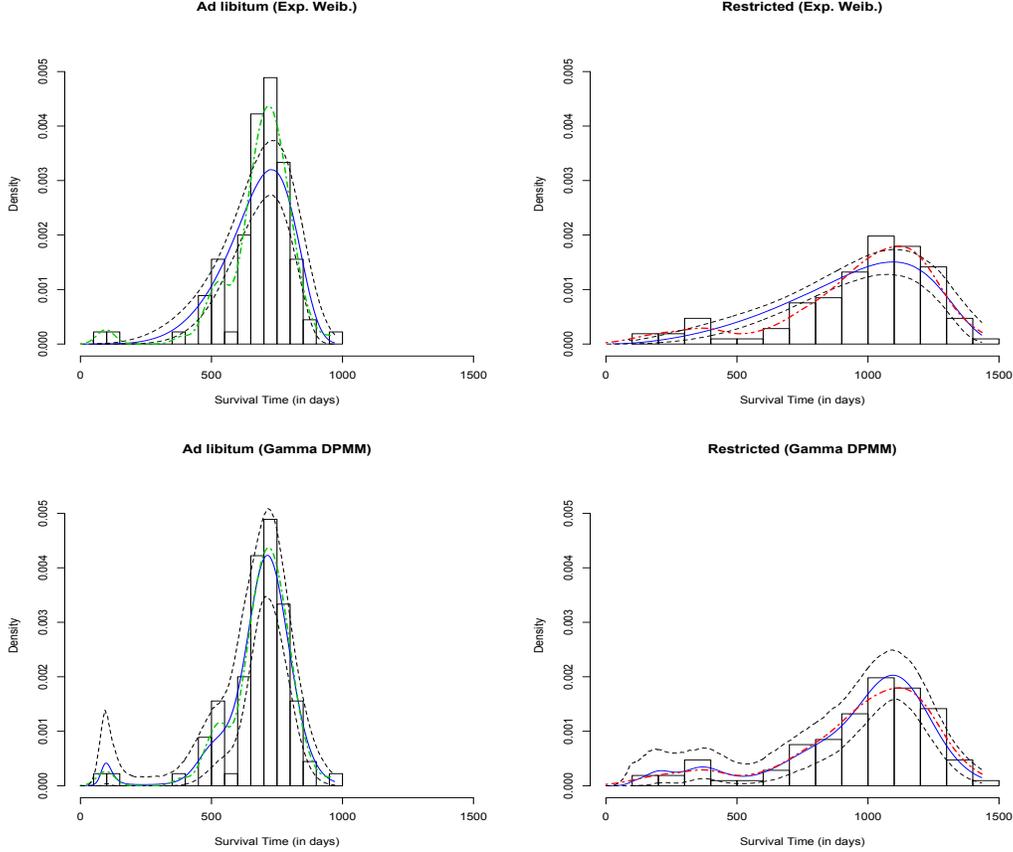


Figure 3: Data from two experimental groups. Posterior point (solid line) and 95% interval estimates (dashed lines) for the density function of the ad libitum group (left panels) and the restricted group (right panels), under the exponentiated Weibull model (top row) and the gamma DPMM (bottom row). Each panel includes the data histogram and the corresponding kernel density estimate (dot-dashed line).

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 & 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 $P(\mathcal{M}) = \sum_{i=1}^n \text{Var}^{\mathcal{M}}(t_{i,rep} | \text{data})$ is a penalty term, and $G(\mathcal{M}) = \sum_{i=1}^n \{t_i - \text{E}^{\mathcal{M}}(t_{i,rep} | \text{data})\}^2$ is a goodness of fit term. Here, $\text{E}^{\mathcal{M}}(t_{i,rep} | \text{data})$ and $\text{Var}^{\mathcal{M}}(t_{i,rep} | \text{data})$ 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

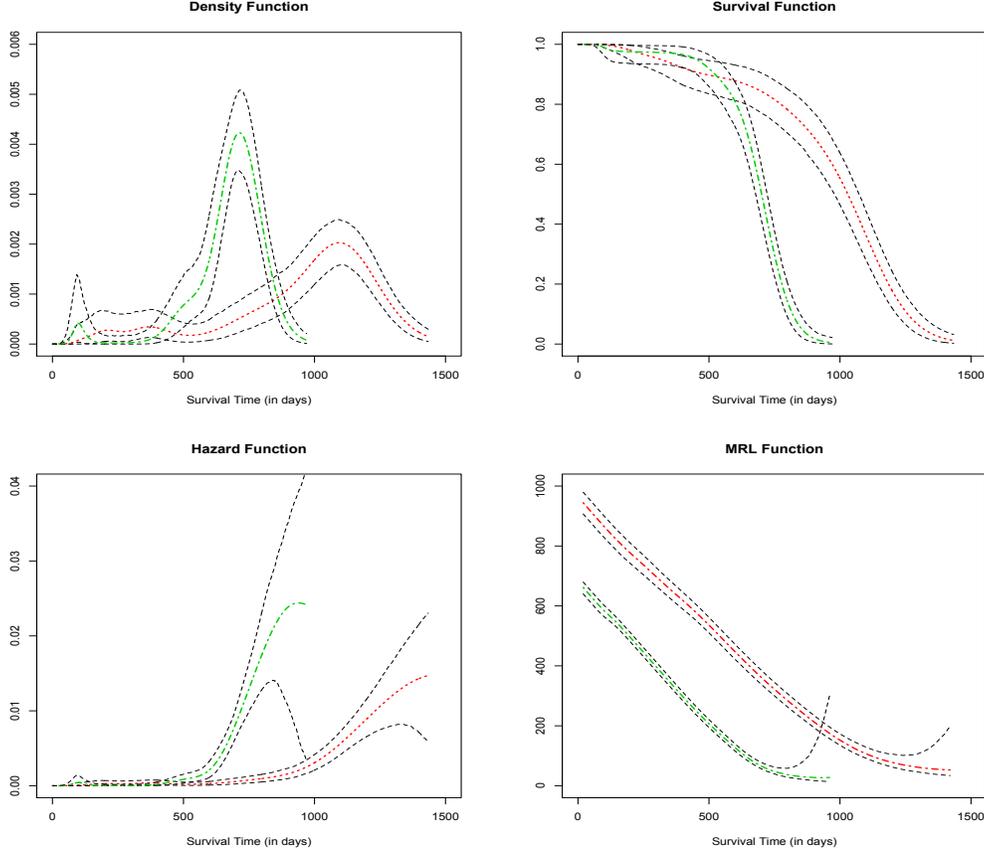


Figure 4: Data from two experimental groups. Under the gamma DPMM, posterior point and interval estimates of the density function (top left panel), survival function (top right panel), hazard function (bottom left panel), and MRL function (bottom right panel) for each group. The point estimates for the ad libitum group and the restricted group are indicated by green and red color, respectively.

\mathcal{M}_1 and \mathcal{M}_2 the exponentiated Weibull model and gamma DPMM, respectively. Then, for the ad libitum group we obtain $G(\mathcal{M}_1) = 1615787$, $P(\mathcal{M}_1) = 1568967$, and $G(\mathcal{M}_2) = 318919$, $P(\mathcal{M}_2) = 684342$. And, for the restricted group, $G(\mathcal{M}_1) = 8542725$, $P(\mathcal{M}_1) = 7917319$, and $G(\mathcal{M}_2) = 739435$, $P(\mathcal{M}_2) = 2247120$. 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 4, 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

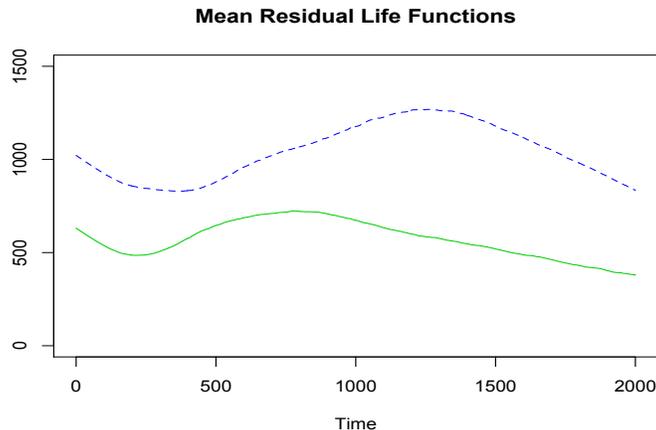


Figure 5: Small cell lung cancer data. Posterior point estimates for the MRL function of Arm A (dashed line) and Arm B (solid line).

to reduce the steepness for 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.

4.3 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 et al., 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 8 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 mean residual life compared to Arm B (see Figure 5), suggesting that the Arm

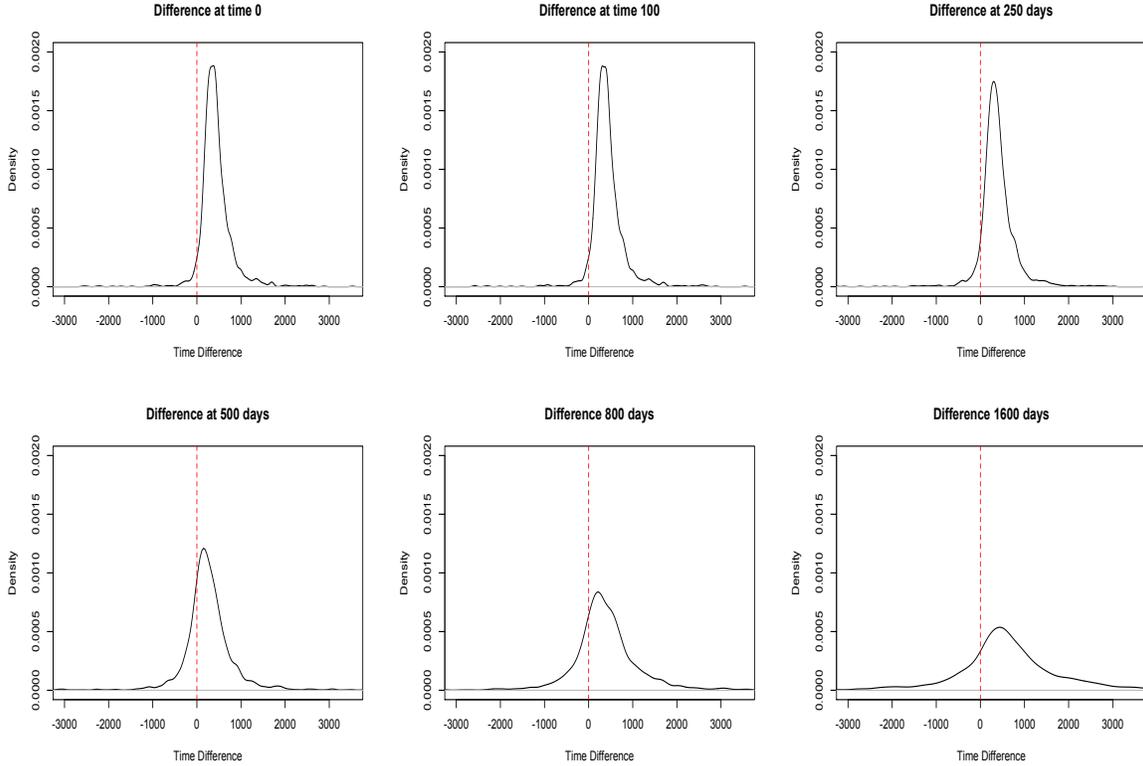


Figure 6: Small cell lung cancer data. Posterior densities for the difference of the MRL functions between Arm A and Arm B, plotted for six time points.

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 6 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. Figure 7 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 poste-

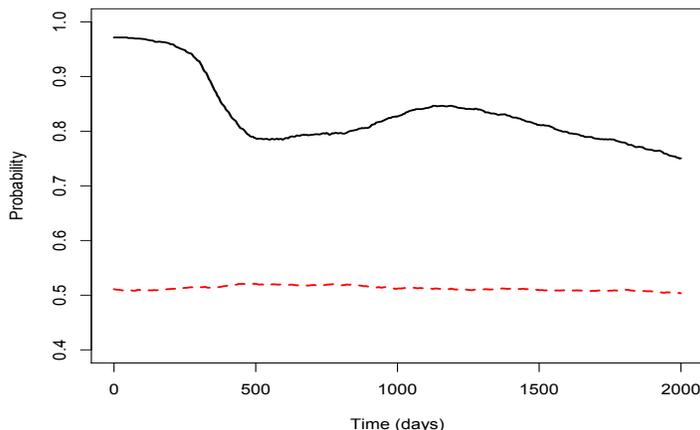


Figure 7: Small cell lung cancer data. The posterior (solid line) and prior (dashed line) probability of the MRL function of Arm A being higher than the MRL function of Arm B, as a function of time.

rior 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 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 flexible modeling approach to obtain inference for mean residual life (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 mixture model for the survival distribution which implies a flexible model structure for the MRL function as a mixture 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 mixture form for the MRL function in (5), which arises from the mixture model in (4) for the corresponding density function, suggests a possible approach to developing nonparametric

priors directly for spaces of MRL functions. In particular, one can start with $\sum q_l(t)m(t | \boldsymbol{\theta}_l)$, as in (5), for a specific choice of $m(t | \boldsymbol{\theta}_l)$ (with corresponding survival function $S(t | \boldsymbol{\theta}_l)$ and density function $k(t | \boldsymbol{\theta}_l)$), and for $\{w_l\}$ that form a (finite or countable) probability vector. Then, if it is shown, using the characterization theorem (Hall & Wellner, 1981), that the particular choice results in a valid MRL function, we have that the density of the corresponding survival distribution is given by $\sum w_l k(t | \boldsymbol{\theta}_l)$, which is practically important for implementation of inference. Such an approach may be fruitful if one seeks more structured modeling, for instance, to incorporate monotonicity restrictions for the shape of the MRL function or to add covariates.

The inclusion of covariates is a particularly important extension, both from a methodological and practical point of view. Mean residual life regression modeling can be explored under either a structured semiparametric setting, using the approach outlined above, or a fully nonparametric framework, based on mixture modeling for the joint response-covariate distribution. We will report on MRL regression methods in a future manuscript.

ACKNOWLEDGMENT

This research is part of the Ph.D. dissertation of the first author, completed at University of California, Santa Cruz. The work of the second author was supported in part by the National Science Foundation under award DMS 1310438.

APPENDIX: Proof of the Lemma

Let \mathcal{F} be the space of absolutely continuous distribution functions on \mathbb{R}^+ with finite mean, $\mu < \infty$, and \mathcal{M} the space of MRL functions for continuous distributions on \mathbb{R}^+ . Using the inversion formula (2), for any MRL function $m \in \mathcal{M}$, we can obtain the corresponding distribution function $F \in \mathcal{F}$.

Consider the class of countable gamma mixture distributions, \mathcal{C} , which is dense in \mathcal{F} . More specifically, for a generic $F \in \mathcal{F}$, consider a sequence of distribution functions, $\{F_n : n = 1, 2, \dots\} \subseteq \mathcal{C}$, defined as follows:

$$F_n(t) = \sum_{j=1}^{\infty} \{F(j/n) - F((j-1)/n)\} F_{\Gamma}(t | j, n)$$

where $F_\Gamma(t | j, n)$ denotes the gamma distribution function with shape parameter j and rate parameter n (thus, its mean is 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 $t_0 > 0$, $\lim_{n \rightarrow \infty} F_n(t_0) = F(t_0)$, that is, the sequence $\{F_n : n = 1, 2, \dots\}$ converges weakly (pointwise) to $F(t)$ (Johnson & Taaffe, 1988; Lee & 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 $\{m_n : n = 1, 2, \dots\}$ 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

$$\begin{aligned}
\mu_n &= \int_0^\infty \{1 - F_n(t)\} dt = \int_0^\infty \sum_{j=1}^\infty \{F(j/n) - F((j-1)/n)\} (1 - F_\Gamma(t | j, n)) dt \\
&= \sum_{j=1}^\infty \{F(j/n) - F((j-1)/n)\} \int_0^\infty (1 - F_\Gamma(t | j, n)) dt \\
&= \sum_{j=1}^\infty \{F(j/n) - F((j-1)/n)\} (j/n) \\
&= \sum_{j=1}^\infty \int_{(j-1)/n}^{j/n} (j/n) dF(t) \\
&\leq \sum_{j=1}^\infty \int_{(j-1)/n}^{j/n} t dF(t) + \sum_{j=1}^\infty \int_{(j-1)/n}^{j/n} n^{-1} dF(t) \\
&= \mu + n^{-1} < \infty
\end{aligned}$$

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 functions $F \in \mathcal{F}$. Analogously to deriving the upper bound for μ_n , we can also obtain a lower bound: $\mu_n = \sum_{j=1}^\infty \int_{(j-1)/n}^{j/n} (j/n) dF(t) \geq \sum_{j=1}^\infty \int_{(j-1)/n}^{j/n} t dF(t) = \mu$. Therefore, $\mu \leq \mu_n \leq \mu + n^{-1}$, which yields $\lim_{n \rightarrow \infty} \mu_n = \mu$, that is, $\lim_{n \rightarrow \infty} m_n(0) = m(0)$.

Next, consider a generic $t_0 > 0$. Using the equivalent MRL function definition in (7), we have

$$\lim_{n \rightarrow \infty} m_n(t_0) = \lim_{n \rightarrow \infty} \frac{\mu_n - \int_0^{t_0} \{1 - F_n(u)\} du}{1 - F_n(t_0)}$$

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, $\lim_{n \rightarrow \infty} \{1 - F_n(t_0)\} = 1 - F(t_0) (> 0)$. Moreover, using the dominated convergence theorem, $\lim_{n \rightarrow \infty} \int_0^{t_0} \{1 - F_n(u)\} du = \int_0^{t_0} \lim_{n \rightarrow \infty} \{1 - F_n(u)\} du = \int_0^{t_0} \{1 - F(u)\} du$. Therefore,

$$\lim_{n \rightarrow \infty} m_n(t_0) = \frac{\lim_{n \rightarrow \infty} \mu_n - \lim_{n \rightarrow \infty} \int_0^{t_0} \{1 - F_n(u)\} du}{\lim_{n \rightarrow \infty} \{1 - F_n(t_0)\}} = \frac{\mu - \int_0^{t_0} \{1 - F(u)\} du}{1 - F(t_0)} = m(t_0)$$

which establishes the result.

References

- ABDOUS, B. & BERRED, A. (2005). Mean residual life estimation. *Journal of Statistical Planning and Inference* **132**, 3–19.
- BERGER, R. L., BOOS, D. D. & GUESS, F. M. (1988). Tests and confidence sets for comparing two mean residual life functions. *Biometrics* **44**, 103–115.
- CHEN, Y. Q. & CHENG, S. (2005). Semiparametric regression analysis of mean residual life with censored data. *Biometrika* **92**, 19–29.
- CONNOR, R. J. & MOSIMANN, J. E. (1969). Concepts of independence for proportions with a generalization of the Dirichlet distribution. *Journal of the American Statistical Association* **64**, 194–206.
- 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. & GHOSH, S. K. (1998). Model choice: A minimum posterior predictive loss approach. *Biometrika* **85**.

- GELFAND, A. E. & 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.
- GOVILT, K. K. & AGGARWAL, K. K. (1983). Mean residual life function of normal, gamma and lognormal densities. *Reliability Engineering* **5**, 47–51.
- GUESS, F. & PROSCHAN, F. (1985). Mean residual life: theory and applications. Technical Report 85-178, North Carolina State University and Florida State University, Tallahassee, Florida.
- GUPTA, R. C. & AKMAN, H. O. (1995). Mean residual life functions for certain types of non-monotonic ageing. *Commun. Statist.-Stochastic Models* **11**, 219–225.
- GUPTA, R. C., AKMAN, O. & LVIN, S. (1999). A study of log-logistic model in survival analysis. *Biometrical Journal* **41**, 431–443.
- HALL, W. J. & WELLNER, J. A. (1979). Estimation of mean residual life. Unpublished manuscript.
- HALL, W. J. & WELLNER, J. A. (1981). Mean residual life. In *Statistics and Related Topics*, Eds. M. Csörgö, D. A. Dawson, J. N. K. Rao & A. K. M. E. Saleh, pp. 169–184. North-Holland Publishing Company.
- HANSON, T. E. (2006). Modeling censored lifetime data using a mixture of gammas baseline. *Bayesian Analysis* **1**, 575–594.
- ISHWARAN, H. & JAMES, L. F. (2001). Gibbs sampling methods for stick-breaking priors. *American Statistical Association* **96**, 161–173.
- ISHWARAN, H. & ZAREPOUR, M. (2000). Markov chain Monte Carlo in approximate Dirichlet and beta two-parameter process hierarchical models. *Biometrika* **87**, 371–390.
- ISHWARAN, H. & ZAREPOUR, M. (2002). Exact and approximate sum representations for the Dirichlet process. *The Canadian Journal of Statistics* **30**, 269–283.
- JOHNSON, M. A. & 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 *Institute for Mathematics and Its Applications: Statistics in the Health Sciences: Diagnosis and Prediction*, Eds. P. Grambsch & S. Geisser, volume 114, pp. 75–90. Springer-Verlag.
- KALLI, M., GRIFFIN, J. & WALKER, S. (2011). Slice sampling mixture models. *Statistics and Computing* **21**, 93–105.
- KOTTAS, A. (2006). Nonparametric Bayesian survival analysis using mixtures of Weibull distributions. *Journal of Statistical Planning and Inference* **136**, 578–596.
- KUO, L. & MALLICK, B. (1997). Bayesian semiparametric inference for the accelerated failure-time model. *The Canadian Journal of Statistics* **25**, 457–472.
- LAHIRI, P. & 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. & LIN, X. S. (2010). Modeling and evaluating insurance losses via mixtures of Erlang distributions. *North American Actuarial Journal* **14**, 107–130.
- MAGULURI, G. & 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. & STRIVASTA, D. K. (1993). Exponentiated Weibull family for analyzing bathtub failure-rate data. *IEEE Transactions of Reliability* **42**, 299–302.
- NEAL, R. (2000). Markov chain sampling methods for Dirichlet process mixture models. *Journal of Computational and Graphical Statistics* **9**, 249–265.
- OAKES, D. & DASU, T. (1990). A note on mean residual life. *Biometrika* **77**, 409–410.
- PHAM, H. & LAI, C.-D. (2007). On recent generalizations of the Weibull distribution. *IEEE Transactions on Reliability* **56**, 454–458.
- 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. Chapman & Hall/CRC.

XIE, M., GOH, T. & 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. & WEI, L. (1995). Survival analysis with median regression models. *Journal of the American Statistical Association* **90**, 178–184.