UNIVERSITY OF CALIFORNIA
SANTA CRUZ

FLEXIBLE BAYESIAN NONPARAMETRIC MODELING FOR TIME-TO-EVENT DATA

A dissertation submitted in partial satisfaction of the
requirements for the degree of

DOCTOR OF PHILOSOPHY

in

STATISTICAL SCIENCE

by

Yunzhe Li

September 2024

The Dissertation of Yunzhe Li
is approved:

Professor Athanasios Kottas, Co-chair

Professor Juhee Lee, Co-chair

Professor Bruno Sansó

Professor Zehang Richard Li

Dean Peter Biehl
Vice Provost and Dean of Graduate Studies

Table of Contents

List of Figures

List of Tables

Abstract

Flexible Bayesian Nonparametric Modeling for Time-to-event Data

by

Yunzhe Li

This dissertation presents the development of innovative Bayesian nonparametric models tailored for the complex demands of analyzing time-to-event data, specifically for problems in survival analysis. These models offer flexibility and computational efficiency in estimating various functionals of the survival distribution. The first thesis component introduces a flexible Erlang mixture model for survival analysis, structured on a weighted combination of Erlang densities with integer shape parameters, and a common scale parameter. The mixture weights are constructed through increments of a distribution function on the positive real line, which is assigned a Dirichlet process prior. The model balances general inference for survival functionals with efficient posterior simulation. The modeling approach is extended to accommodate multiple experimental groups through a dependent Dirichlet process prior. Moving to the second part of the dissertation, a Dirichlet process mixture model with a log-logistic kernel is proposed. The model incorporates covariates through a density regression framework, allowing variations in mixture weights and mixing parameters as functions of covariates. The model yields flexible inference for density, survival, and hazard functions across the covariate space. The final dissertation component explores a joint modeling approach for recurrent events and survival time, relevant for medical studies where the

recurrent events process and the risk of death are related. Here, the density functions for the survival times and the gap times of recurrent events are modeled by dependent Dirichlet process mixtures with a log-logistic kernel. This modeling approach builds dependence between survival times and recurrent events through bivariate random effects. The joint modeling framework aims to provide flexibility in inferring marginal and conditional functionals of survival and gap times. For all proposed models, we discuss model properties, prior specification, and posterior simulation techniques, illustrating their effectiveness through synthetic and real data examples.

Acknowledgments

The research presented in this thesis, portions of which have been previously published in *Computational Statistics & Data Analysis* (2024), would not have been possible without the support and guidance of many individuals. I would like to extend my sincere thanks to the National Science Foundation (NSF) for their financial support under award DMS 2015428, which played a crucial role in enabling this research. The published work is as follows: Li, Y., Lee, J., and Kottas, A. (2024), "Bayesian Nonparametric Erlang Mixture Modeling for Survival Analysis," *Computational Statistics & Data Analysis*, 191, 107874.

First and foremost, I would like to express my deepest gratitude to my advisors, Athanasios Kottas and Juhee Lee, for their unwavering support, guidance, and encouragement throughout my research and the writing of this thesis. Their insightful feedback and dedication have been invaluable to my academic growth.

I am also incredibly grateful to the members of my thesis committee, Bruno Sansó and Richard Zehang Li, for their time, constructive criticism, and valuable suggestions that significantly improved the quality of my work.

Special thanks go to my colleagues and friends in the Statistics department of University of California, Santa Cruz, whose collaboration, discussions, and moral support have made this journey enjoyable and intellectually stimulating.

I would also like to extend my appreciation to Herbie Lee, who kindly provided me with GSR opportunities for over three years.

I am deeply indebted to my family for their unconditional love, patience, and encouragement throughout my academic journey. Their support has been a constant source of motivation for me.

Chapter 1

Introduction

Time-to-event data, also known as survival data or failure time data, arise in many fields of study, including medicine, biology, engineering, public health, epidemiology, and economics. The focus is on the time duration until the occurrence of an event of interest, which could range from death and disease recurrence to equipment failure or any event signifying a change in status or outcome. From a frequentist standpoint, Klein & Moeschberger (1997) provide a comprehensive review of survival analysis methods. Ibrahim et al. (2001) offer a detailed review of Bayesian modeling and inference approaches to survival data.

A more complex scenario in survival analysis involves recurrent events, where subjects may experience multiple non-fatal events within a certain period before a terminal event occurs. Cook & Lawless (2007) discuss such scenarios, emphasizing the importance of jointly modeling terminal and recurrent events to accurately capture the dependencies and provide meaningful insights.

This dissertation is focused on developing flexible Bayesian methods for estimation of functionals in survival analysis, including density, survival, and hazard functions. Section 1.1 provides an introduction to survival analysis and joint analysis of survival times and recurrent events. Subsequently, Section 1.2 outlines our research objectives and provides an overview of this dissertation.

1.1 Background

1.1.1 Survival analysis

Let T be a continuous random variable representing survival time, defined on $\mathbb{R}^+$. The probability density function is denoted by $f(t)$. The survival function, denoted as $S(t)$, represents the probability that the time to an event exceeds t and is defined as

$$S(t) = Pr(T > t) = \int_t^\infty f(u)du.$$

Additionally, the hazard function, denoted by $h(t)$, describes the instantaneous failure rate and is defined as

$$h(t) = \lim_{\Delta t \to 0} P(t < T \leq t + \Delta t \mid T > t),$$

and it can be calculated by the ratio of density to survival function, $h(t) = f(t)/S(t)$. Conversely, the survival function can also be derived from a hazard function by $S(t) = \exp\left(-\int_0^t h(u)du\right)$.

A common challenge in survival analysis is right censoring, where survival times are observed for only a subset of the subjects under study, while for others, survival

times are only known to exceed certain values. Specifically, an observation is considered right-censored at c if it is only known that the survival time is greater than c.

Let n denote the total number of subjects under study. For the i^{th} subject, let t_i be its survival time, and c_i be the right censored time. Thus, the final follow-up time is defined as $\tau_i = \min(t_i, c_i)$. Let ν_i be a binary indicator, where $\nu_i = 1$ indicates an observed survival time.

Assume t_i's are independent and identically distributed (i.i.d.) with density $f(t)$ and survival function $S(t)$, then the likelihood function for all subjects is given by

$$L = \prod_{i=1}^{n} \{f(t_i)\}^{\nu_i} \{S(t_i)\}^{1-\nu_i}.$$

The observed survival times and censored times contribute to the likelihood through the density and survival function, respectively.

Survival data are often collected along with covariates that provide additional information about the subjects. These covariates might include demographic details, treatment types, genetic markers, or other relevant factors that can influence the time-to-event outcome. Incorporating covariates into survival analysis allows for more nuanced and precise modeling, leading to better inference and prediction of survival times. Two popular regression model formulations for survival data are the proportional hazards (PH) model and the accelerated failure time model (AFT).

The PH model assumes that the hazard function $h(t \mid \boldsymbol{x})$ for a subject with covariate vector $\boldsymbol{x}$ is given by:

$$h(t \mid \boldsymbol{x}) = h_0(t) \exp(\boldsymbol{\beta}' \boldsymbol{x}),$$

where $h_0(t)$ is a baseline hazard function, which can be modeled through parametric or nonparametric approaches. And $\exp(\boldsymbol{\beta}'\boldsymbol{x})$ is a multiplicative effect on baseline hazard function. As the name indicates, for two different covariate vectors $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$, the hazards ratio is expressed as

$$\frac{h(t \mid \boldsymbol{x}_1)}{h(t \mid \boldsymbol{x}_2)} = \exp(\boldsymbol{\beta}'(\boldsymbol{x}_1 - \boldsymbol{x}_2)),$$

which is constant, in time, meaning crossing hazards are not allowed in the PH model.

As an useful alternation to the PH model, the AFT model assume that the log-transformed survival time is linearly related to the covariates. For survival times t, $i = 1, \ldots, N$, the basic form of the AFT model might be written as:

$$\log(t) = \boldsymbol{\beta}'\boldsymbol{x}_i + \epsilon_i,$$

where ϵ_i are i.i.d. random variables with a unspecified common distribution function F, which can be modeled through parametric or nonparametric approaches. The AFT model indicates that the survival function of a subject with covariate $\boldsymbol{x}$ at time t is

$$S(t \mid \boldsymbol{x}) = S_0(t \exp(\boldsymbol{\beta}'\boldsymbol{x})),$$

where S_0 is the baseline survival function. However, the crossing survival functions over the covariate space is unavailable.

Both the PH and AFT models impose a parametric linear component in the model formulations, which can be too restrictive under certain circumstances. For instance, the hazard and survival functions for two different covariates might not maintain a consistent stochastic ordering over time. Covariates effects might be negative

at the beginning of the study and positive by the end. Specifically, for two different covariates, x_1 and x_2 with $x_1 < x_2$, it is possible to have $h(t \mid x_1) < h(t \mid x_2)$ and $S(t \mid x_1) < S(t \mid x_2)$, for small t, but $h(t \mid x_1) > h(t \mid x_2)$ and $S(t \mid x_1) > S(t \mid x_2)$ for large t. Thus, a more sophisticated model is needed.

The first two projects in this dissertation aim to provide new statistical tools for analyzing survival times, both with and without covariates, under the Bayesian nonparametric modeling framework, which allows for flexibility regarding the survival response distribution and the regression relationship.

1.1.2 Joint analysis for survival times and recurrent events

In more complex scenarios, survival times are often observed alongside a series of recurrent events. Subjects may experience multiple non-fatal events repeatedly within a certain time frame until a terminal event occurs. In this context, recurrent events and terminal events are considered two different types of events. One example is in medical studies where patients might have multiple hospital readmission before an eventual death. For more more detailed reviews, we refer to Cook & Lawless (2007) for a discussion of general framework and Sinha et al. (2008) for a more recent review of Bayesian methodologies.

When modeling recurrent events, two primary types of point processes are often considered: the Poisson process and the renewal process. The Poisson process describes situations where events occur randomly, with the number of events in nonoverlapping time intervals being statistically independent. It is defined through its intensity function,

$\lambda(t)$, for $t > 0$, a nonnegative integrable function. In this case, the Poisson process, where intensity function is constant over time, is referred as the homogeneous Poisson process. In this case, the gap times between successive events follow an exponential distribution. When the intensity function varies over time, the process is known as a nonhomogeneous Poisson process, which is suitable when the event rate changes over time. Based on the intensity function, the likelihood for the Poisson process model observed in time window $(0, t)$ with point pattern $\{0 < y_1 < y_2 < \cdots < y_N < t\}$ is proportional to $\prod_{i=1}^{N} \lambda(y_i) \exp(-\int_0^t \lambda(u)du)$.

The renewal process is another extension to the homogeneous Poisson process, but instead of focusing on the intensity function, it models the distribution of gap times between events. In a renewal process, these gap times are assumed to be i.i.d. Let $w_i = y_i - y_{i-1}$ $(i = 1, 2, \dots)$ with y_0 defined as 0, and let $\{w_1, w_2, \dots, w_N\}$ be a collection of gap times within a time window $(0, t)$. The likelihood function can be written as $\prod_{i=1}^{N} f_R(w_i) S_R(t - \sum_{j=1}^{N} w_j)$, where f_R and S_R are density and survival functions of gap times. This formulation accounts for the observed gap times and the gap between the last observed event and the upper bound of the time window.

The joint modeling of survival times and recurrent events typically employs a random effects framework to establish the dependence between these two types of events. For subject i, let ϵ_i and ξ_i be positive-valued random effects associated with survival times and recurrent events, respectively. Suppose that, conditional on (ϵ_i, ξ_i) and covariate vector $\boldsymbol{x}_i$, survival times are independent of recurrent events and have a density function $f_S(\cdot \mid \epsilon_i, \boldsymbol{x}_i)$, which is typically modeled by either PH or AFT models.

For recurrent events under the Poisson-process assumption, the intensity function is modified to $\lambda_i(t \mid \xi_i, \boldsymbol{x}_i)$, where random effects and covariates term $\exp(\boldsymbol{\beta}'\boldsymbol{x}_i)$ are usually modeled as multiplicative effects on a baseline intensity function. Alternatively, under renewal-process assumption, recurrent events are modeled through the density function of the gap times. Similar to the survival analysis, the gap times conditional on random effects and covariates are modeled through PH or AFT models.

1.2 Research Objectives

Bayesian nonparametric (BNP) models are known for their remarkable flexibility and have been studied in many fields. Unlike traditional parametric models, which rely on a fixed number of parameters, BNP models employ priors that are probability models for infinite-dimensional parameters. This approach allows BNP models to adapt to the complexity of the data without being constrained by predetermined functional forms.

For a comprehensive overview of BNP inference and computational strategies, Müller et al. (2015) provide details discussions on practical implementation. On the theoretical perspective, Ghosal & Vaart (2016) cover the foundational aspects of BNP models.

This dissertation introduces new Bayesian nonparametric models designed to meet the complex demands of analyzing time-to-event data. These models offer flexibility and computational efficiency in estimating various statistical functionals.

In Chapter 2, we construct a flexible Erlang mixture model for survival analysis. This model is composed of Erlang densities with common scale parameter and integer-valued shape parameters. The mixture weights are defined by increments of a distribution function on the positive real line $\mathbb{R}^+$. By assigning it a Dirichlet process (DP) prior, the model allows for flexible estimation for density, survival, and hazard functions. The model can be viewed as a basis representation through Erlang densities, with the number of basis densities being random. A key feature of this model is its ability to enable efficient posterior computation via Markov chain Monte Carlo (MCMC) methods. The model is then extended to accommodate multiple experimental groups through a dependent Dirichlet process (DDP), with a focus on control-treatment groups. This model extension retains the flexibility for the time-to-event distributions, avoiding specific parametric assumptions, such as those in proportional hazards (PH) and accelerated failure times models.

Moving to the second part of the dissertation in Chapter 3, we propose a Dirichlet process mixture (DPM) model with a log-logistic density kernel, referred by DPM-LL. Similar to Erlang mixture models, this model is able to flexibly capture a wide range of functional shapes for density, survival, and hazard functions. A major advantage of the DPM-LL model is that it allows a relatively straightforward extension to accommodate multiple covariates through a density regression framework. Unlike the model in Chapter 3 with the basis representation, this approach is constructed through traditional mixture model setting, allowing both mixture weights and locations to vary as functions of covariates.

In Chapter 4, we develop a joint model for survival times and gap times of recurrent events. This model integrates Bayesian nonparametric mixture models for both survival and gap times. By employing the DDP mixture model with a log-logistic kernel for the mixing distributions across the covariate space, we incorporate covariates into both the mixture atoms and weights. This model connects nonparametric mixture models for survival and gap times through bivariate random effects, allowing it to capture non-standard density and survival functions across various covariates and to estimate conditional survival probabilities given no occurrence of recurrent events. We explore the model properties and posterior simulation, illustrating the methodology with real data. The model's efficacy is assessed through out-of-sample predictions, validating its practical applicability and flexibility.

Chapter 2

Bayesian Nonparametric Erlang Mixture Modeling for Survival Analysis

2.1 Introduction

The Erlang mixture model is defined as a weighted combination of M Erlang densities $f(t \mid M, \theta, \boldsymbol{\omega}) = \sum_{m=1}^{M} \omega_m \operatorname{Er}(t \mid m, \theta)$, for $t \in \mathbb{R}^+$, where $\operatorname{Er}(t \mid m, \theta)$ represents the Erlang densities with integer shape parameters, m, and scale parameter, θ, shared by all densities. In the formulation, M and $\boldsymbol{\omega} = (\omega_1, \ldots, \omega_M)$ are the number of mixture components and the mixing weights, respectively. The model is completely specified by M, $\boldsymbol{\omega}$ and θ. Hence, in contrast to traditional mixture models, Erlang mixtures comprise identifiable mixture components and a parsimonious model formulation built from kernels that involve a single parameter that needs to be estimated. Indeed, it is more natural to view the model as a basis representation for densities on $\mathbb{R}^+$, where

the $\text{Er}(t \mid m, \theta)$ densities play the role of the basis densities and the ω_m provide the corresponding weights. Erlang mixtures are flexible, making them particularly useful in providing approximations for general densities on $\mathbb{R}^+$. For example, Venturini et al. (2008) used the Erlang mixture model for density estimation, with fixed M and with a Dirichlet prior distribution for $\boldsymbol{\omega}$. Another example is Ayala et al. (2022) that developed a Dirichlet process mixture model (Antoniak 1974) with Erlang kernels for phase-type distributions of a Markov process. The resulting model can be expressed as a mixture of Erlang distributions in a form similar to $f(t \mid M, \theta, \boldsymbol{\omega})$.

Different from the aforementioned approaches, the weights $\boldsymbol{\omega}$ in Erlang mixtures can be constructed as increments of a distribution function G on $\mathbb{R}^+$. In particular, we let $\omega_m = G(m\theta) - G((m-1)\theta)$, for $m = 1, \ldots, M-1$, and $\omega_M = 1 - G((M-1)\theta)$. This formulation yields an important theoretical result for Erlang mixtures: as $M \to \infty$ and $\theta \to 0$, the Erlang mixture density converges pointwise to the density of G (e.g., Butzer 1954, Lee & Lin 2010).

The Erlang mixture structure, in conjunction with the theoretical support from the convergence result, provides an appealing setting for nonparametric Bayesian modeling and inference. The key ingredient for such modeling is a nonparametric prior for distribution G, which, along with priors for θ and M, yields the full Bayesian model. Regarding relevant existing approaches, we are only aware of Xiao et al. (2021) where the Erlang mixture is used as a prior model for inter-arrival densities of homogeneous renewal processes. Also related is the prior model for Poisson process intensities in Kim & Kottas (2022), although for that model the weights are defined as increments of a

cumulative intensity function.

To our knowledge, Erlang mixtures have not been explored as a general methodological tool for nonparametric Bayesian survival analysis, and this is our motivation for the work in this article. The nonparametric Bayesian model is built from a Dirichlet process (DP) prior (Ferguson 1973) for distribution G, which defines the mixture weights, and from parametric priors for θ and M, which control the effective support and smoothness in the shape of the Erlang mixture density. The modeling approach is sufficiently flexible to handle non-standard shapes for important functionals of the time to event distribution, including the survival function and the hazard function. We discuss prior specification for the model hyperparameters, and design an efficient posterior simulation method that draws from well-established techniques for DP mixture models. The model is extended to incorporate survival responses from multiple experimental groups, using a dependent Dirichlet process prior (MacEachern 2000, Quintana, Müller, Jara & MacEachern 2022) for the group-specific distributions that define the mixture weights. The model extension retains the flexibility in the group-specific survival densities, and it also allows for general relationships between groups that bypass restrictive assumptions, such as proportional hazards.

Survival analysis is among the earliest application areas of Bayesian nonparametrics. The literature includes modeling and inference methods based on priors on the space of survival functions, survival densities, cumulative hazard functions, or hazard functions. Reviews can be found, for instance, in Ibrahim et al. (2001), Phadia (2013), Müller et al. (2015), and Mitra & Müller (2015). The part of this literature that is more

closely related to our proposed methodology involves DP mixture models for the survival density. Such mixture models have been developed using kernels that include the Weibull distribution (e.g., Kottas 2006), log-normal distribution (e.g., De Iorio, Johnson, Müller & Rosner 2009), and gamma distribution (e.g., Hanson 2006, Poynor & Kottas 2019). The convergence property for Erlang mixtures is the only mathematical result we are aware of that supports the choice of a particular parametric kernel in mixture modeling for densities on $\mathbb{R}^+$.

Our main objective is to add a new practical tool to the collection of nonparametric Bayesian survival analysis methods. The DP-based Erlang mixture model may be attractive for its modeling perspective that involves a representation of basis densities, its parsimonious mixture structure, and efficient posterior simulation algorithms (comparable to the ones for standard DP mixtures).

The rest of the article is organized as follows. Section 2 introduces the methodology, including approaches to prior specification and posterior simulation (with details for the latter given in the Appendixes). Sections 3 and Section 4 present results from synthetic and real data examples, respectively. Finally, Section 5 concludes with a summary.

2.2 Methodology

2.2.1 The modeling approach

Erlang Mixture Model We propose a structured mixture model of Erlang densities for the density, $f(t)$, of the time to event distribution, aiming at more general inference for survival functionals than what specific parametric distributions can provide. Specifically, let

$$f(t) \equiv f(t \mid M, \theta, \boldsymbol{\omega}) = \sum_{m=1}^{M} \omega_m \, \mathrm{Er}(t \mid m, \theta), \quad t \in \mathbb{R}^+, \tag{2.1}$$

where $\boldsymbol{\omega} = \{\omega_m : m = 1, \ldots, M\}$ denotes the vector of mixture weights, and $\mathrm{Er}(\cdot \mid m, \theta)$ the density of the Erlang distribution, that is, the gamma distribution with integer shape parameter m and scale parameter θ, such that the mean is $m\theta$ and the variance $m\theta^2$. Given the number of the Erlang mixture components, M, the kernel densities in (2.1) are fully specified up to the common scale parameter θ. Hence, compared with standard mixture models, for which the number of unknown parameters increases with M, the model in (2.1) offers a parsimonious mixture representation.

A key component of the model specification revolves around the mixture weights. These are defined through increments of a distribution function G with support on $\mathbb{R}^+$, such that $\omega_m = G(m\theta) - G((m-1)\theta)$, for $m = 1, \ldots, M-1$, and $\omega_M = 1 - G((M-1)\theta)$. This formulation for the mixture weights provides appealing theoretical results for the Erlang mixture model in (2.1). In particular, as $M \to \infty$ and $\theta \to 0$, $f(t \mid M, \theta, \boldsymbol{\omega})$ converges pointwise to the density function of G. The convergence property for the density can be derived from more general probabilistic results (e.g.,

Butzer 1954); a proof of the convergence of the distribution function of $f(t \mid M, \theta, \boldsymbol{\omega})$ to G can be found in Lee & Lin (2010). This result highlights that using a prior with wide support for G is crucial to achieve the generality of the model in (2.1) required to capture non-standard shapes for the time to event distribution. We provide details below on the nonparametric prior for G, as well as on the priors for parameters θ and M.

The model in (2.1) also offers a flexible, albeit parsimonious mixture representation for the survival function, $S(t \mid M, \theta, G)$, and the hazard function, $h(t \mid M, \theta, G)$. Note that, having defined the mixture weights $\boldsymbol{\omega}$ through distribution G, we use the latter in the notation for model parameters. Denote by $S_{\mathrm{Er}}(\cdot \mid m, \theta)$ and $h_{\mathrm{Er}}(\cdot \mid m, \theta)$ the survival and hazard function, respectively, of the Erlang distribution with parameters m and θ. Then, the survival function associated with the model in (2.1) is given by

$$S(t \mid M, \theta, G) \;=\; \sum_{m=1}^{M} \omega_m \, S_{\mathrm{Er}}(t \mid m, \theta), \tag{2.2}$$

that is, it has the same weighted combination representation as the density, replacing the Erlang basis densities by the corresponding survival functions. Moreover, the hazard function under the Erlang mixture model can be expressed as

$$h(t \mid M, \theta, G) \;=\; \sum_{m=1}^{M} \omega^{\star}_m(t) \, h_{\mathrm{Er}}(t \mid m, \theta), \tag{2.3}$$

where $\omega^{\star}_m(t) = \omega_m S_{\mathrm{Er}}(t \mid m, \theta) / \{\sum_{m'=1}^{M} \omega_{m'} \, S_{\mathrm{Er}}(t \mid m', \theta)\}$. The hazard function is a weighted combination of the hazard functions associated with the Erlang basis densities, and, importantly, the mixture weights in (2.3) vary with t. Such time-dependent weights allow for local adjustment, and thus $h(t \mid M, \theta, G)$ can achieve general shapes, despite

the fact that the basis hazard functions, $h_{\text{Er}}(t \mid m, \theta)$, are non-decreasing in t (constant for $m = 1$, and increasing for $m \geq 2$).

We note that Erlang mixtures can be viewed as extensions of Bernstein polynomial densities, defined on the unit interval to $\mathbb{R}^+$. The Bernstein polynomial density is given by $f^*(y \mid K, \boldsymbol{q}) = \sum_{k=1}^{K} q_k \, \text{Be}(y \mid k, K - k + 1)$, where $\text{Be}(\cdot \mid a, b)$ is the beta density with mean $a/(a + b)$. The mixture weights are defined through increments of a distribution function G^* on the unit interval, such that $q_k = G^*(k/K) - G^*((k-1)/K)$, for $k = 1, \ldots, K$, where G^* is a probability distribution over the unit interval. Here, as $K \to \infty$, $f^*(y \mid K, \boldsymbol{q})$ converges uniformly to the density of G^*. Bernstein polynomials have been explored for density estimation on compact sets following the work of Petrone (1999*a*,*b*). Applications and extensions of the basic model include density estimation on higher dimensional spaces (Zheng et al. 2010), density estimation with multiscale mixtures of Bernstein polynomials (Canale & Dunson 2016), density regression (Barrientos et al. 2017), and modeling for bivariate stable distributions (Richardson et al. 2020).

Dirichlet Process Prior for G As previously discussed, a key model component is distribution G as it defines the mixture weights ω_m through discretization of its distribution function on intervals $B_m = ((m-1)\theta, m\theta]$, for $m = 1, \ldots, M-1$, and $B_M = ((M-1)\theta, \infty)$. We place a DP prior on G, i.e., $G \mid \alpha, G_0 \sim \text{DP}(\alpha, G_0)$, where $\alpha > 0$ is the total mass parameter and G_0 the centering distribution (Ferguson 1973). We work with an exponential distribution, $\text{Exp}(\zeta)$, for G_0, with random mean ζ assigned an inverse-gamma hyperprior, $\zeta \sim \text{inv-Ga}(a_\zeta, b_\zeta)$. We further assume a gamma hy-

perprior for the total mass parameter, $\alpha \sim \text{Ga}(a_\alpha, b_\alpha)$. Given M, the DP prior for G implies a Dirichlet prior distribution for the vector of mixture weights, $\boldsymbol{\omega} \mid M, \alpha, \zeta \sim \text{Dir}(\alpha G_0(B_1), \ldots, \alpha G_0(B_M))$.

The nonparametric prior for G is of primary importance. The DP prior allows the corresponding distribution function realizations to admit general shapes that can concentrate probability mass on different time intervals B_m, thus favoring different Erlang basis densities through the associated ω_m. The key parameter in this respect is α, as it controls the extent of discreteness for realizations of G and the variability of such realizations around G_0. As an illustration, Figure 2.1 plots prior realizations for the mixture weights and the corresponding Erlang mixture density under three values of α ($\alpha = 1, 10$ or 100), using in all cases $M = 50$, $\theta = 0.5$, and $G_0 = \text{Exp}(5)$. The smaller α gets, the smaller the number of effective mixture weights becomes. Also, for larger α the Erlang mixture density becomes similar to the density of G_0, which is to be expected from the pointwise convergence result and the fact that larger α values imply smaller variability of G around G_0.

Priors for θ and M Under the model construction for the mixture weights, θ controls the step size of the increments and thus how fine the discretization of G is. Moreover, θ controls the location and dispersion of the Erlang basis densities in (2.1). With smaller θ, the Erlang densities are more concentrated around their mean $m\theta$, and the discretization of G becomes finer. Hence, as the pointwise convergence result suggests, smaller θ values may be needed to accommodate non-standard density shapes. Also,

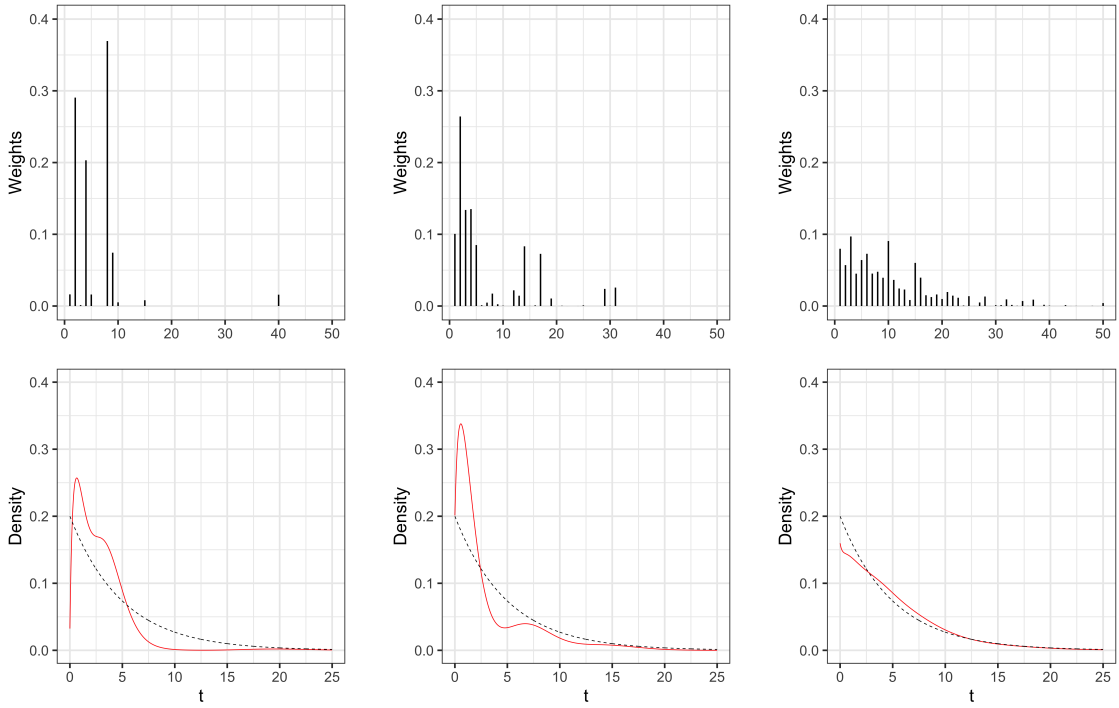


Figure 2.1: Prior realizations of the mixture weights $\boldsymbol{\omega}$ (top row) and the corresponding densities $f(t \mid M, \theta, G)$ given by the red solid lines (bottom row), under $\alpha = 1, 10, 100$ (left, middle, right columns). In all cases, $M = 50$, $\theta = 0.5$, and $G_0 = \text{Exp}(5)$. The black dotted line in the bottom row panels is the density of G_0.

the last component in (2.1) has mean $M\theta$ (with variance $M\theta^2$), and thus the effective support of $f(t \mid M, \theta, G)$ is jointly determined by M and θ; with smaller θ, a greater value of M is needed to achieve the same effective support. Figure 1 in the Supplementary Material plots prior realizations of the Erlang mixture density for different combinations of (M, θ) to illustrate how M and θ jointly affect $f(t \mid M, \theta, G)$.

We work with a joint prior for θ and M, $p(\theta, M) = p(\theta)p(M \mid \theta)$. We assume $\theta \sim \text{Ga}(a_\theta, b_\theta)$, and conditional on θ, assign to M a discrete uniform distribution,

$M \mid \theta \sim \text{Unif}(\lceil M_1/\theta \rceil, \ldots, \lceil M_2/\theta \rceil)$, where $\lceil a \rceil$ is the smallest integer that is larger or equal to a. To specify the hyperparameters a_θ, b_θ, M_1 and M_2, we use a relatively conservative approach, based on the range of the data. For M_1, we choose a value greater than the largest value in the data, and set $M_2 = cM_1$ for a relatively small integer c. The motivation for this choice is to ensure that the effective support of the Erlang mixture model is sufficiently large for the particular data application. To specify the prior hyperparameters for θ, we notice that $M_1/\theta \sim \text{inv-Ga}(a_\theta, M_1/b_\theta)$, based on which we recommend selecting values for a_θ and b_θ such that $\mathbb{E}(M_1/\theta)$ is between 10 and 50. Also, we use $c = 3$ or 4 that imply $\mathbb{E}(M_2/\theta)$ is between 20 and 150 or between 40 and 200, for the simulation studies in Section 3 and the real data analyses in Section 4. This specification provides an adequate number of basis densities a priori for sufficient flexibility for the examples.

Posterior Simulation The data point for the i^{th} subject is recorded as $y_i = \min(t_i, c_i)$, where t_i is the survival time and c_i the (independent) administrative censoring time, for $i = 1, \ldots, n$. The data set can be represented through $\mathcal{D} = \{(y_i, \nu_i) : i = 1, \ldots, n\}$, where the ν_i are binary censoring indicators such that $\nu_i = 1$ if t_i is observed, and $\nu_i = 0$ otherwise. Then, the likelihood function can be written as

$$L(M, \theta, G; \mathcal{D}) = \prod_{i=1}^{n} \{f(y_i \mid M, \theta, G)\}^{\nu_i} \{S(y_i \mid M, \theta, G)\}^{1-\nu_i}, \tag{2.4}$$

where $f(\cdot \mid M, \theta, G) \equiv f(\cdot \mid M, \theta, \boldsymbol{\omega})$ and $S(\cdot \mid M, \theta, G)$ are given in (2.1) and (2.2), respectively.

We implement posterior inference via Markov chain Monte Carlo (MCMC)

simulation, using standard posterior simulation methods for DP mixture models (e.g., Escobar & West 1995, Neal 2000). The Erlang mixture density in (2.1) can be expressed as a DP mixture by exploiting the definition of the weights ω_m through distribution G, resulting in the following alternative mixture representation:

$$f(t \mid M, \theta, G) \;=\; \sum_{m=1}^{M} \omega_m \,\mathrm{Er}(t \mid m, \theta) \;=\; \int_0^\infty \left\{ \sum_{m=1}^{M} \mathbb{1}_{B_m}(\phi)\, \mathrm{Er}(t \mid m, \theta) \right\} \mathrm{d}G(\phi).$$

Here, $\mathbb{1}_B(\cdot)$ is the indicator function for set B, and, as before, $B_m = ((m-1)\theta, m\theta]$, for $m = 1, \ldots, M-1$, and $B_M = ((M-1)\theta, \infty)$.

For posterior simulation, we augment the likelihood in (2.4) with subject-specific latent variables, $\phi_i \mid G \overset{i.i.d.}{\sim} G$, which indicate the mixture component for the associated observations. In particular, if ϕ_i falls into interval B_m, the i^{th} observation corresponds to the m^{th} Erlang basis density. The posterior distribution involves G, M, θ, the set of latent variables $\boldsymbol{\phi} = \{\phi_i : i = 1, ..., n\}$, and the DP hyperparameters (α, ζ). We marginalize G over its DP prior and work with the prior full conditionals for the ϕ_i, implied by the DP Pólya urn representation (Blackwell & MacQueen 1973*a*), to sample from the marginal posterior distribution for all model parameters except G. To this end, we employ the MCMC method in Escobar & West (1995); the details are given in Appendix A.

Although we do not sample the mixture weights $\boldsymbol{\omega}$ during the MCMC simulation, it is straightforward to obtain posterior samples for $\boldsymbol{\omega}$, using their definition in terms of distribution G.

The conditional posterior distribution for G, given (α, ζ) and $\boldsymbol{\phi}$, is character-

ized by a DP with updated total mass parameter $\alpha^\star = \alpha + n$, and centering distribution $G_0^\star = \alpha(\alpha+n)^{-1}\text{Exp}(\zeta) + (\alpha+n)^{-1}\sum_{i=1}^n \delta_{\phi_i}$. Hence, using the DP definition, the conditional posterior distribution for $\boldsymbol{\omega}$, given M, (α, ζ), and $\boldsymbol{\phi}$, is a Dirichlet distribution with parameter vector $(\alpha^\star G_0^\star(B_1), \ldots, \alpha^\star G_0^\star(B_M))$.

Two points about the posterior simulation method are worth making. First, note that the model parameters do not explicitly contain the vector of mixture weights. The mixture weights are estimated through the posterior distribution of G, which plays the role of the relevant parameter. This is practically important in that the dimension of the parameter space does not change with M, and we thus do not need to resort to more complex trans-dimensional MCMC algorithms. Second, the DP-based Erlang mixture model offers an interesting example where full posterior inference can be obtained from a DP mixture model without the need to truncate or approximate the DP prior. This is a result of the use of a marginal MCMC method, as well as of the fact that distribution G enters the model only through increments of its distribution function, which define the mixture weights.

2.2.2 Model extension for control-treatment studies

A practically relevant scenario in studies where survival responses are collected involves data from multiple experimental groups, typically associated with different treatments. Evidently, it is of interest in these settings to compare time to event distributions across different groups. We develop an extension of the Erlang mixture model in this direction, focusing on the case of two groups for, say, a generic control-treatment

study. Our objective is to retain the flexible modeling approach for the time to event distributions, avoiding restrictions to specific parametric shapes or rigid relationships, such as proportional hazards. We also seek a prior probability model that allows for dependence, and thus borrowing of information, between the two distributions.

We use the dependent DP (DDP) prior structure (MacEachern 2000) that extends the DP prior for distribution G to a prior model for a collection of covariate-dependent distributions, G_x, where x indexes the distributions in terms of values in the covariate space. Our context involves a binary covariate $x \in \mathcal{X} = \{ctr, trt\}$, where *ctr* and *trt* represent control and treatment groups, respectively. The DDP prior builds from the DP stick-breaking representation (Sethuraman 1994) by utilizing covariate-dependent weights and/or atoms. We work with a common-weights DDP prior model:

$$G_x = \sum_{\ell=1}^{\infty} p_\ell \, \delta_{\varphi^*_{x\ell}}, \quad \text{for } x \in \mathcal{X}, \tag{2.5}$$

with $p_1 = v_1$, $p_\ell = v_\ell \prod_{r=1}^{\ell-1}(1 - v_r)$, for $\ell \geq 2$, where the v_ℓ are i.i.d. from a Beta$(1, \alpha)$ distribution, and the atoms $\boldsymbol{\varphi}^\star_\ell = (\varphi^\star_{ctr\ell}, \varphi^\star_{trt\ell})$ arise i.i.d. from a bivariate distribution G_0; moreover, $\{v_\ell\}$ and $\{\boldsymbol{\varphi}^\star_\ell\}$ are independent sequences of random variables. Note that, under this construction, G_x follows marginally a DP(α, G_{0x}) prior, where G_{0x}, for $x \in \mathcal{X}$, are the marginals of G_0 associated with the control and treatment groups. For G_0, we consider a bivariate log-normal distribution, such that $\boldsymbol{\varphi}^\star_\ell \mid \boldsymbol{\mu}, \Sigma \overset{i.i.d.}{\sim} \text{LN}_2(\boldsymbol{\mu}, \Sigma)$. We place a bivariate normal, $\text{N}_2(\bar{\boldsymbol{\mu}}, \Sigma_0)$, hyperprior on $\boldsymbol{\mu}$, with $\bar{\boldsymbol{\mu}}$ and Σ_0 fixed, an inverse-Wishart, inv-Wishart(r, R), hyperprior on Σ, with r and R fixed, and a gamma hyperprior on the total mass parameter α.

Allowing also for group-specific number of Erlang basis densities, M_x, as well as group-specific Erlang scale parameter, θ_x, the extension of the Erlang mixture model in (2.1) can be expressed as

$$f_x(t) \equiv f(t \mid M_x, \theta_x, G_x) = \sum_{m=1}^{M_x} \omega_{xm} \, \mathrm{Er}(t \mid m, \theta_x), \quad t \in \mathbb{R}^+, \tag{2.6}$$

where $\omega_{xm} = G_x(m\theta_x) - G_x((m-1)\theta_x)$, $m = 1, \ldots, M_x - 1$, and $\omega_{xM_x} = 1 - G_x((M_x - 1)\theta_x)$. Similar to the model in (2.1), the group-specific Erlang basis densities are fully specified given M_x and θ_x. The prior probability model in (2.6) induces dependence between the control and treatment group densities through the dependent distributions G_{ctr} and G_{trt} in (2.5). These random discrete distributions have common weights and dependent group-specific atoms. The effect of the common weights (total mass parameter α) and of the dependent atoms (parameters of distribution G_0) can be studied, for instance, through the covariance between random probabilities under the time to event distributions associated with (2.6). The mathematical derivation of this covariance is given in the Supplementary Material.

The survival functions, $S_x(t)$, and hazard functions, $h_x(t)$, under the extended model have a mixture representation similar to (2.2) and (2.3),

$$S_x(t) = \sum_{m=1}^{M_x} \omega_{xm} \, S_{\mathrm{Er}}(t \mid m, \theta_x) \quad \text{and} \quad h_x(t) = \sum_{m=1}^{M_x} \omega^\star_{xm}(t) \, h_{\mathrm{Er}}(t \mid m, \theta_x), \tag{2.7}$$

where $\omega^\star_{xm}(t) = \omega_{xm} \, S_{\mathrm{Er}}(t \mid m, \theta_x) / \{\sum_{m'=1}^{M_x} \omega_{xm'} \, S_{\mathrm{Er}}(t \mid m', \theta_x)\}$. Note that both the mixture components and weights are indexed by x. Again, the time-dependent weights in the hazard mixture form allow for local adjustment, and thus for flexible group-specific hazard rate shapes. Importantly, the prior model allows for general relationships

between the control and treatment group hazard functions. In particular, inference is not restricted by the proportional hazards assumption, implied by several commonly used parametric or semiparametric survival regression models.

To complete the full Bayesian model, we place priors on θ_x and M_x, using again the role of these parameters (discussed in Section 2.2.1). More specifically, for each x, the joint prior, $p(\theta_x, M_x) = p(\theta_x)p(M_x \mid \theta_x)$. We further assume $\theta_x \overset{ind}{\sim} \text{Ga}(a_{x\theta}, b_{x\theta})$, and $M_x \mid \theta_x \overset{ind.}{\sim} \text{Unif}\,(\lceil M_{x1}/\theta_x \rceil, \ldots, \lceil M_{x2}/\theta_x \rceil)$. We use an approach similar to the one described in Section 2.2.1 to specify M_{x1} and M_{x2}, and the hyperparameters for θ_x.

Posterior simulation for the DDP-based Erlang mixture model proceeds with a relatively straightforward extension of the MCMC simulation method in Section 2.2.1. The details are provided in Appendix B.

The primary focus of this paper is on the DP-based Erlang mixture model for survival analysis and its extension for the control-treatment setting. We note however that the DDP-based Erlang mixture model can be further extended to accommodate a general p-variate covariate vector $\boldsymbol{x}$. For example, we may consider a linear-DDP structure (De Iorio, Johnson, Müller & Rosner 2009) to extend G_x in (2.5) to $G_{\boldsymbol{x}} = \sum_{\ell=1}^{\infty} p_\ell \, \delta_{\psi^\star_\ell(\boldsymbol{x})}$, where $\psi^\star_\ell(\boldsymbol{x}) = \exp((1, \boldsymbol{x}')\boldsymbol{\beta}_\ell)$ with the $\boldsymbol{\beta}_\ell$ i.i.d. from a baseline distribution. The structured DDP prior for $G_{\boldsymbol{x}}$ yields covariate-dependent mixture weights, and thus a nonparametric prior model for covariate-dependent survival densities and hazard functions. A regression model may also be used for M and/or θ. Different from the linear-DDP mixture of log-normal distributions in De Iorio, Johnson, Müller & Rosner (2009), the extended model retains the parsimonious Erlang mixture structure.

2.3 Simulation Study

We use three simulation scenarios to illustrate the models developed in Section 2.2. For the Erlang mixture model for a single distribution, we consider simulated data from: a two-component log-normal mixture to demonstrate the model's capacity to estimate non-standard density and hazard function shapes (Section 2.3.1); and a log-normal distribution sampled with different levels of censoring (Section 2.3.2). The DDP-based extension of the model is illustrated in Section 2.3.3 with a synthetic data example based on a log-normal control distribution and a two-component log-normal mixture treatment distribution, specified such that the corresponding hazard functions cross each other.

For all data examples considered here and in Section 2.4, we used the approach discussed in Section 2.2 to specify the prior hyperparameters. Consistent with inference results obtained from DP mixture models, we have observed some sensitivity to the prior choice for α. The effect on the posterior distribution for α is more noticeable for the small cell lung cancer data of Section 2.4.2 (involving the smallest sample size among our data examples). However, posterior inference results for survival functionals are largely unaffected even under fairly different priors for α. When the sample size is relatively small for each group, we recommend applying the DDP-based Erlang mixture model with a prior for α that supports small to moderate values, such as the $\text{Ga}(5, 1)$ prior used in Sections 2.3.3 and 2.4.2.

We examined convergence and mixing of the MCMC algorithms using standard

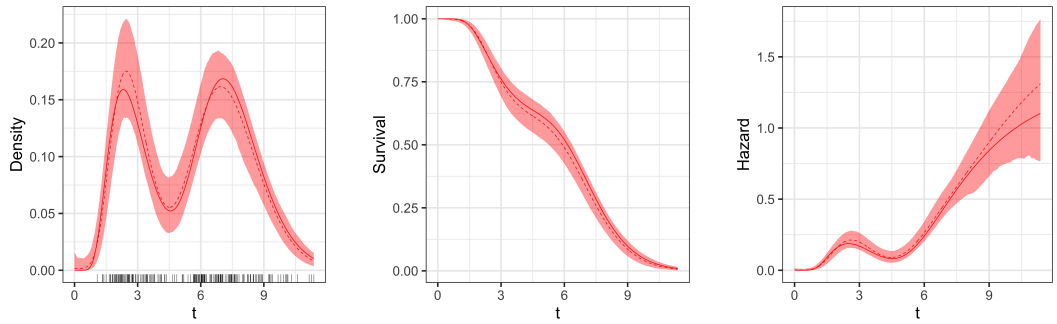
diagnostic techniques. In our experiments, we observed that parameters θ and M are highly correlated, and moderate thinning was used to improve efficiency. A general approach we take is to run the MCMC chain for 100,000 iterations, then discard the first 25% posterior samples and keep every 38th iteration for posterior inference.

2.3.1 Example 1: Bimodal density

We simulate $n = 200$ survival times from a mixture of two log-normal distributions, $0.4\,\mathrm{LN}(1, 0.4) + 0.6\,\mathrm{LN}(2, 0.2)$, which yields a bimodal density and a non-monotonic hazard function. The true underlying functions $f(t)$, $S(t)$ and $h(t)$ are plotted with solid lines in Figure 2.2. Regarding prior specification, we used: $\alpha \sim \mathrm{Ga}(2, 1)$; $\zeta \sim \text{inv-Ga}(3, 4)$; $\theta \sim \mathrm{Ga}(1, 1)$; and, $M \mid \theta \sim \mathrm{Unif}\left(\lceil M_1/\theta \rceil, \ldots, \lceil M_2/\theta \rceil\right)$, with $M_1 = 13$ and $M_2 = 3 \times M_1$.

Posterior inference is summarized in Figure 2.2. The complex features of the underlying survival functionals are captured well by the model. In particular, the inference results for the hazard function demonstrate the effectiveness of the model structure in (2.3) with the time-dependent weights allowing for local adjustment and estimation of a non-standard hazard function shape.

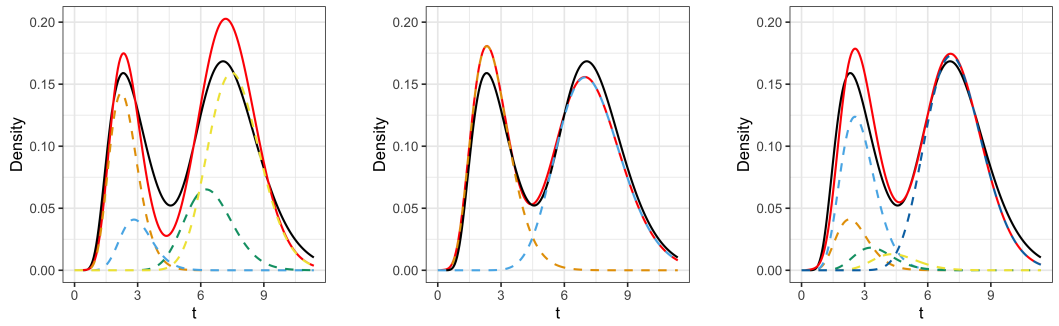
The posterior distribution for the common scale parameter θ is substantially concentrated on smaller values relative to its prior, in particular, the posterior mean and 95% credible interval estimates for θ are 0.28 and $(0.13, 0.39)$. Recalling the definition of the mixture weights, this indicates the level of partitioning needed to accommodate the non-standard, bimodal shape of the underlying density. The posterior mean and



(a) Density function (b) Survival function (c) Hazard function

Figure 2.2: Simulation Example 1. Posterior mean (dashed lines) and 95% interval estimates (shaded regions) for the density function (left panel), survival function (middle panel) and hazard function (right panel). The red solid line in each panel corresponds to the true underlying function. The black marks on the x-axis in the left panel show the observed survival times.

95% credible interval estimates of the number M of mixture components are 101 and $(44, 223)$. However, the number of effective mixture components (i.e., effective basis densities) is considerably smaller than M. As an informal rule, we identify an effective Erlang basis density through its corresponding mixture weight taking value greater than a threshold of 0.01. Then, the number of effective mixture components is about 4 (on average across posterior samples). For a graphical illustration, Figure 2.3 plots three randomly selected posterior realizations of $f(t \mid M, \theta, G)$. The associated posterior draws for (θ, M) are $(0.2, 153)$, $(0.33, 78)$, and $(0.25, 54)$, whereas the number of effective Erlang basis densities is only 4, 2, and 5, respectively. The weighted effective basis densities (i.e., $\omega_m \times \text{Er}(t \mid m, \theta)$ for m such that $\omega_m > 0.01$) are also plotted in



(a) $\theta = 0.20$, $M = 153$ (b) $\theta = 0.33$, $M = 78$ (c) $\theta = 0.25$, $M = 54$

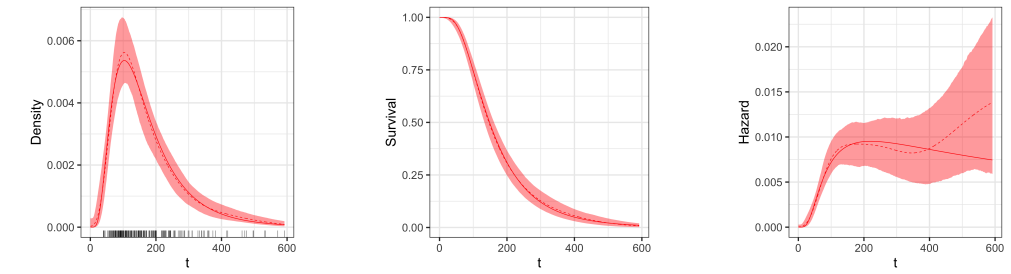
Figure 2.3: Simulation Example 1. Plots (a)-(c) show the posterior realization of $f(t \mid M, \theta, G)$ (red solid line), based on three randomly chosen posterior samples. Each dashed line represents the Erlang basis density $\text{Er}(t \mid m, \theta)$ for components with $\omega_m > 0.01$, multiplied by its corresponding weight. The black solid line is the true underlying density.

Figure 2.3. This example highlights the critical importance of the nonparametric prior for distribution G that defines the weights for the Erlang mixture model.

2.3.2 Example 2: Unimodal density with censoring

For the second synthetic data example, we generate survival times from a log-normal distribution, $t_i \overset{i.i.d.}{\sim} \text{LN}(5, 0.6)$, $i = 1, \ldots, n$ with $n = 200$. The priors for the model parameters are: $\alpha \sim \text{Ga}(2, 1)$; $\zeta \sim \text{inv-Ga}(3, 1000)$; $\theta \sim \text{Ga}(2, 25)$; and $M \mid \theta \sim \text{Unif}\left(\lceil 1000/\theta \rceil, \ldots, \lceil 3000/\theta \rceil\right)$.

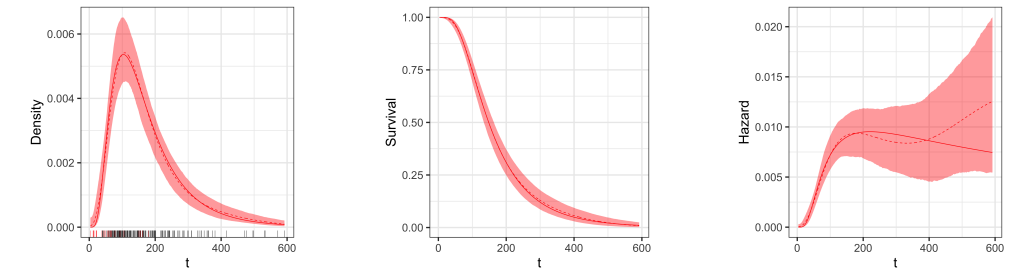
As shown in Figure 2.4(a)-(c), the model estimates well the density, survival and hazard function. The point estimate for the hazard function is less accurate beyond $t = 400$, which is to be expected given the very few observations that are greater than



(a) Density, no censoring

(b) Survival, no censoring

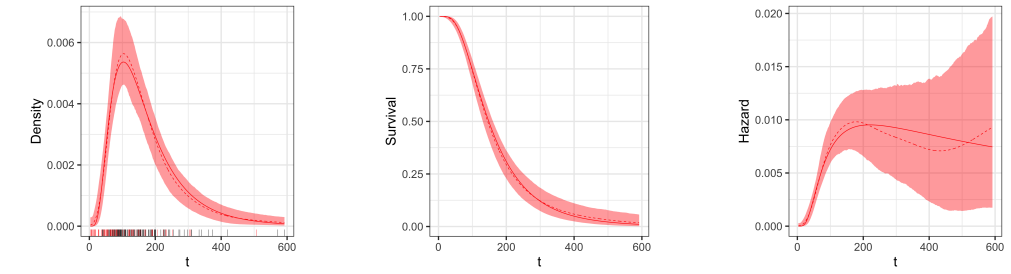
(c) Hazard, no censoring



(d) Density, $g = 12\%$

(e) Survival, $g = 12\%$

(f) Hazard, $g = 12\%$



(g) Density, $g = 33.5\%$

(h) Survival, $g = 33.5\%$

(i) Hazard, $g = 33.5\%$

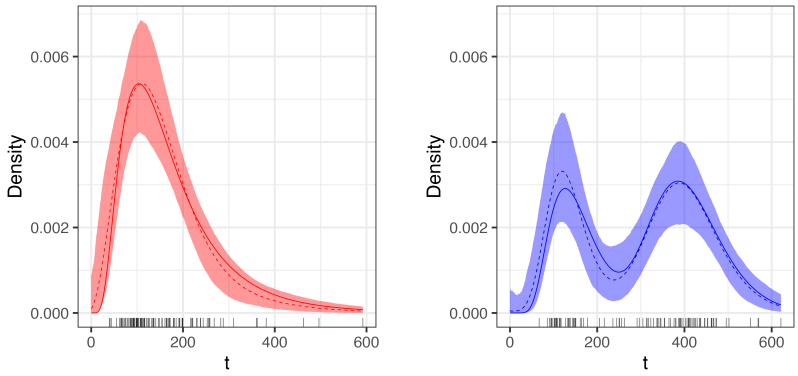
Figure 2.4: Simulation Example 2. Posterior mean (dashed lines) and 95% interval estimates (shaded regions) for the density function (left panel), survival function (middle panel) and hazard function (right panel). The red solid line in each panel corresponds to the true underlying function, and the black and red rugs in the left panel show the survival and censoring times, respectively. For the results in the first row, the survival times are fully observed, whereas 12% and 33.5% of the observations are censored for those in the second and third rows, respectively.

that time point, although the interval estimate contains the true function throughout the observation time window.

In addition, we examine the model's performance for data with censored observations. We simulate censoring times c_i from an exponential distribution with mean parameter κ, and define the observed times as $y_i = \min(t_i, c_i)$, with binary censoring indicators $\nu_i = 1(y_i \leq c_i)$. We generate the c_i under two different values of κ, resulting in two datasets with different proportions of censored observations, $g = 12\%$ and 33.5%. Figure 2.4(d)-(i) plot posterior mean and 95% interval estimates for the density, survival and hazard functionals. We note that censoring does not substantially affect the quality of the inference results, with the true function contained in all cases within the posterior interval estimates. The width of the posterior uncertainty bands increases with the larger censoring proportion. The increase is more noticeable for the hazard function estimates.

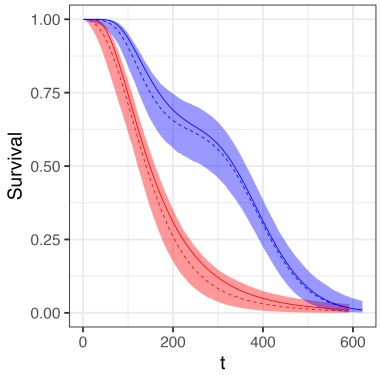
2.3.3 Example 3: A control-treatment synthetic data set

Here, we examine the performance of the DDP-based Erlang mixture model of Section 2.2.2. We consider a binary covariate, $x_i = ctr$ or trt, with 100 responses in each group, such that $n = 200$. We generate $t_i \overset{i.i.d.}{\sim} \text{LN}(5, 0.6)$ for subjects with $x_i = C$, and $t_i \overset{i.i.d.}{\sim} 0.4\,\text{LN}(5, 0.4) + 0.6\,\text{LN}(6, 0.2)$ for subjects with $x_i = T$. The true density, survival and hazard functions are shown in Figure 2.5. The control group density is unimodal, whereas the treatment group has a bimodal density and a non-standard, non-monotonic hazard function. The truth is specified such that we have

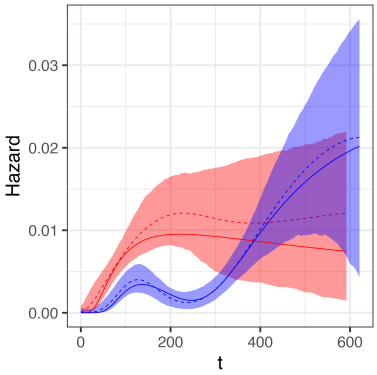


(a) Density (control group)

(b) Density (treatment group)



(c) Survival function



(d) Hazard function

Figure 2.5: Simulation Example 3. Panels (a) and (b) plot the estimates for the control and treatment group density, respectively (the rug plots show the corresponding survival times). Panels (c) and (d) compare the estimates for the survival and hazard function, respectively. In each panel, the dashed lines denote the posterior mean estimates, the solid line the true underlying function, and the shaded regions indicate the 95% credible intervals. Red and blue color is used for the control and treatment group, respectively.

crossing hazard functions for the two groups, a scenario that traditional proportional hazards models cannot accommodate.

Regarding the prior hyperparameters, we set: $\alpha \sim \text{Ga}(2, 1)$; $\boldsymbol{\mu} \sim \text{N}_2((5, 5.5)', 10\,\text{I}_2)$;

$\Sigma \sim$ inv-Wishart$(4, 3\,\mathrm{I}_2)$; $\theta_x \overset{ind.}{\sim} \mathrm{Ga}(2, 20)$; and $M_x \mid \theta_x \overset{ind.}{\sim} \mathrm{Unif}\,(\lceil 1000/\theta_x \rceil, \ldots, \lceil 3000/\theta_x \rceil)$. As shown in Figure 2.5, the model captures effectively the shape of the survival functionals, despite the fact that the functions vary greatly across the two groups, and it successfully recovers the non-proportional hazards relationship between the groups. Again, with respect to hazard estimation, the point estimates are generally less accurate and the interval bands are wider for larger time points where data is scarce.

For comparison, we apply the linear-DDP (LDDP) model with log-normal kernels in De Iorio, Johnson, Müller & Rosner (2009). R package *DPpackage* (Jara et al. 2011) is used to fit the LDDP model to the dataset. While the LDDP model estimates of $f_x(t)$ are generally reasonable, they show relatively poor performance, especially for $t < 300$, where both conditions have a reasonable number of observations. This misfit may be attributed to the linear-DDP model structure, which assumes shared weights for the conditions and linearity for the locations. Note that the Erlang-DDP model in (2.6) has condition-specific weights and does not assume any particular structure on the locations. More details are provided in Section 3 of the Supplementary Material.

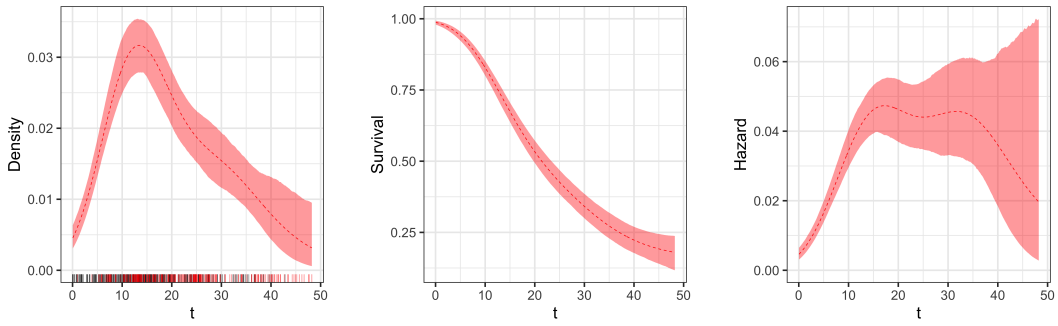
For additional sensitivity analysis, we explore different specifications of the fixed hyperparameters for the priors of $\boldsymbol{\mu}$ and Σ and refit the data. In particular, we change the specification of Σ_0, r and R. We observe that these changes had a minimal impact on the posterior inference. Details are given in Section 4 of the Supplementary Material.

2.4 Data Examples

2.4.1 Liver metastases data

We consider data on survival times (in months) from 622 patients with liver metastases from a colorectal primary tumor without other distant metastases, available from the R package "locfit". The censoring proportion is high, with 259 censored responses. The data set has been used in earlier work to illustrate classical and Bayesian nonparametric methods for density and hazard estimation; see, e.g., Antoniadis et al. (1999) and Kottas (2006).

To apply the DP-based Erlang mixture model, we set the priors as follows: $\alpha \sim \text{Ga}(5, 1)$; $\zeta \sim \text{inv-Ga}(3, 80)$; $\theta \sim \text{Ga}(2, 2)$; and, $M \mid \theta \sim \text{Unif}(\lceil 100/\theta \rceil, \ldots, \lceil 300/\theta \rceil)$. Inference results for the density, survival, and hazard function are reported in Figure 2.6. The model estimates a unimodal survival density (with mode at about 13 months), with a non-standard, skewed right tail. The hazard rate estimate increases up to about 17 months, stays roughly constant between 17 to 35 months, and then decreases. The width of the posterior uncertainty bands for the hazard function increases considerably beyond 40 months, which is consistent with the fact that there are very few responses beyond that time point, and almost all of them are censored. Density and hazard rate estimates with similar shapes were obtained from the previous analyses in Antoniadis et al. (1999) and Kottas (2006). Overall, this example supports the findings from the simulation study regarding the Erlang mixture model's capacity to effectively estimate non-standard density and hazard function shapes.



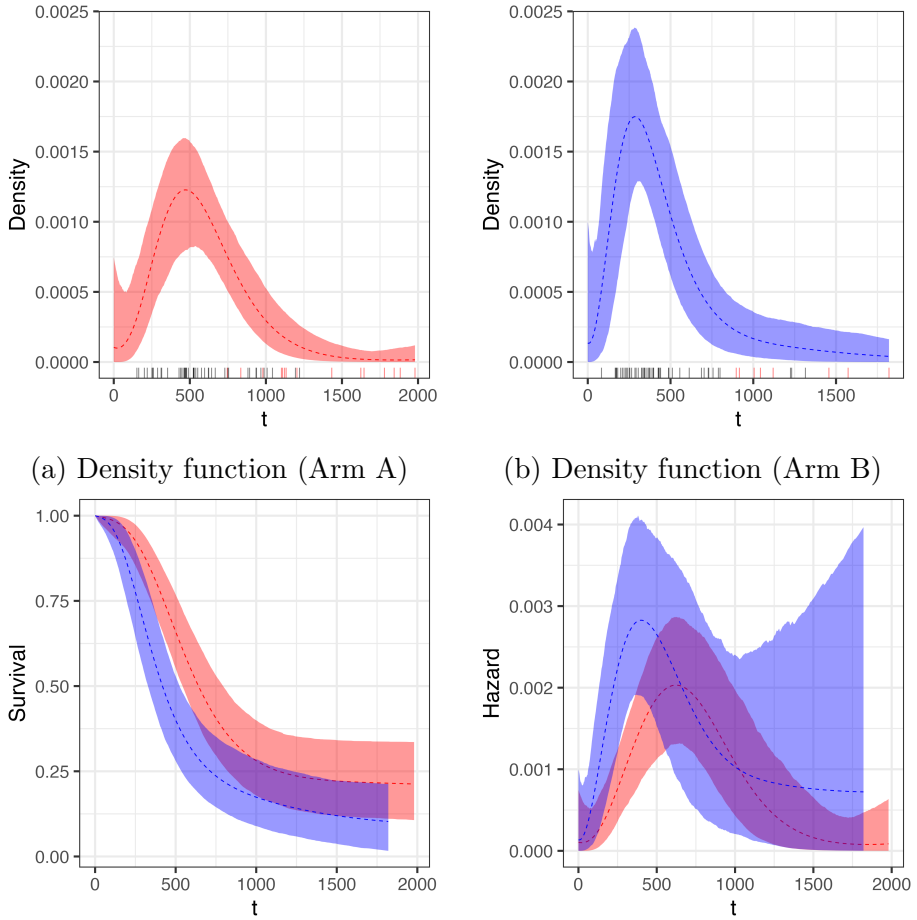
(a) Density function (b) Survival function (c) Hazard function

Figure 2.6: Liver metastases data. Panels (a), (b) and (c) plot posterior mean (dashed lines) and 95% interval estimates (shaded regions) for the density, survival and hazard function, respectively. The rug plot in panel (a) shows observed (black) and censored (red) survival times.

2.4.2 Small cell lung cancer data

To illustrate the DDP-based Erlang mixture model with real data, we consider the data set from Ying et al. (1995) on survival times (in days) of patients with small cell lung cancer. The data correspond to a study designed to evaluate two treatment regimens of drugs, etoposide (E) and cisplatin (P), given with a different sequence, with Arm A denoting the regimen where P is followed by E, and Arm B the regimen where E is followed by P. A total of 121 patients were randomly assigned to one of the treatment arms, resulting in 62 patients in Arm A, and 59 in Arm B. The survival times of 23 patients (15 in Arm A and 8 in Arm B) are administratively right censored.

The DDP-based Erlang mixture model is applied with $x \in \mathcal{X} = \{A, B\}$. The priors are set as follows: $\theta_x \overset{ind.}{\sim} \text{Ga}(50, 2)$; $M_x \mid \theta_x \overset{ind.}{\sim} \text{Unif}\left(\lceil 2500/\theta_x \rceil, \ldots, \lceil 10000/\theta_x \rceil\right)$;



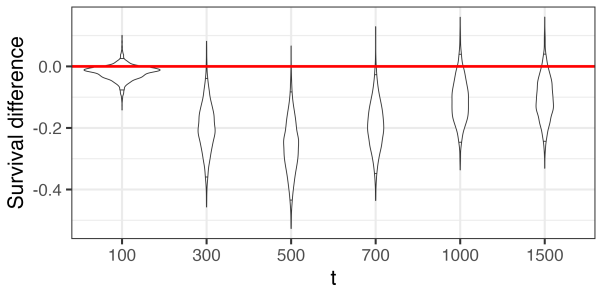
(a) Density function (Arm A)

(b) Density function (Arm B)

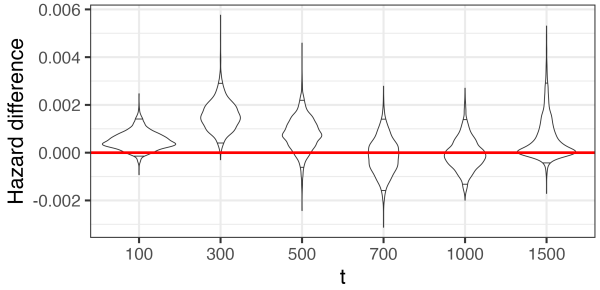
(c) Survival functions

(d) Hazard functions

Figure 2.7: Small cell lung cancer data. Panels (a) and (b) plot estimates for the Arm A and Arm B density; the rug plots show observed (black) and censored (red) survival times. Panels (c) and (d) compare the estimates for the survival and hazard function. In each panel, the dashed lines denote the posterior mean estimates, and the shaded regions indicate the 95% credible intervals. Red and blue color is used for the Arm A and Arm B group, respectively.



(a) $S_B(t) - S_A(t)$



(b) $h_B(t) - h_A(t)$

Figure 2.8: Small cell lung cancer data. Panels (a) and (b) show, through violin plots, the posterior distributions of the difference between the two treatment survival and hazard functions at six specific time points, $t = 100$, 300, 500, 700, 1000, and 1500 days. The short black solid lines within each violin plot indicate the 95% posterior credible interval.

$\alpha \sim \text{Ga}(2, 1)$; $\boldsymbol{\mu} \sim \text{N}_2((6.7, 6.3)', 10\,\text{I}_2)$; and $\Sigma \sim \text{inv-Wishart}(13, 30\,\text{I}_2)$.

Posterior mean and interval estimates for the density, survival, and hazard function are compared across the two treatments in Figure 2.7. The Arm B density estimate is more peaked, and the mode under Arm B is estimated to be smaller than that under Arm A. The posterior mean estimates for the survival functions indicate that

survival time under Arm B is stochastically smaller than that under Arm A. However, we note the overlap in the interval estimates for the two treatment survival functions for smaller time points and, more emphatically, for time points beyond about $t = 700$ days. Based on the hazard function posterior mean estimates, the hazard rate under arm B is larger than that under arm A, with the exception of the time interval from about 700 to 1100 days that corresponds to a crossing of the estimated hazard functions. In this case, there is even more substantial overlap of the interval estimates, driven by the large posterior uncertainty for the arm B hazard rate estimate. Nonetheless, the estimates strongly suggest that the proportional hazards assumption is not suitable for this study.

For a more focused comparison of the two treatments, Figure 2.8 plots the entire posterior distribution for the difference between the survival and hazard functions at six specific time points, $t = 100$, 300, 500, 700, 1000, and 1500 days. The lines within each violin plot indicate the 95% posterior credible interval for $S_B(t) - S_A(t)$ and $h_B(t) - h_A(t)$, and can thus be contrasted with the horizontal reference line at 0. Based on the 95% interval estimate, treatment A outperforms treatment B at $t = 300$, 500 and 700 days with respect to survival probability, and at $t = 300$ days according to hazard rate.

2.5 Summary

We have developed a parsimonious Erlang mixture model as a general methodological tool for nonparametric Bayesian survival analysis. The model is built from a

basis representation for the survival density, using Erlang basis densities with a common scale parameter. The weights are defined through increments of a random distribution function, which is flexibly modeled with a Dirichlet process prior. Utilizing a common-weights dependent Dirichlet process prior, the model has been extended to accommodate a categorical covariate associated with a generic control-treatment setting. The proposed methodology provides a useful balance between model flexibility and computational efficiency. The models were illustrated with synthetic and real data examples.

Chapter 3

Bayesian Nonparametric Survival Regression Model using Log-logistic kernel

3.1 Introduction

The log-logistic model, a distribution that belongs to the family of generalized linear models, has become increasingly relevant for analyzing time-to-event data. Originating from life data analysis, its application has spanned various disciplines, including medical studies, engineering, and social sciences. Historically, the log-logistic distribution was regarded as an alternative to the more frequently used Weibull and log-normal distributions, particularly in the realm of survival analysis. The log-logistic model offers certain benefits, such as its flexible shape for the hazard function, a closed-form expression for both survival and hazard functions, and heavier tails (Cox & Oakes

1984; Bennett 1983). Over time, numerous studies have focused on the applications and extensions of the log-logistic model. For example, Collett (2003) highlighted its implementation in medical research, while Klein & Moeschberger (1997) emphasized the model's suitability for censored and truncated data.

In this study, we present an innovative approach that combines the log-logistic model with the BNP methods. The aim is to augment existing survival analysis tools by enhancing both their flexibility and computational efficiency. While parametric and semi-parametric methodologies have predominantly filled the literature, the BNP techniques introduced here stand out for their minimal reliance on distributional assumptions. Historically linked with survival analysis, extensive literature exists on BNP, delving into topics like survival densities, functions, and hazard mechanisms. For an in-depth exploration of these core methodologies, readers can consult seminal works such as those by Ibrahim et al. (2001), Phadia (2013), Müller et al. (2015), and Mitra & Müller (2015).

DPM models occupy a significant position in the BNP literature. These models serve as bridges between traditional parametric distributions and the more expansive world of BNP. In essence, they form a countable mixtures of parametric distributions, enabling greater flexibility from a modeling perspective and efficient computation via Markov chain Monte Carlo (MCMC) approaches (Escobar & West 1995; Neal, 2000). In particular, if G is a $DP(\alpha, G_0)$ with total mass parameter α and baseline distribution

G_0, the DPM models can be represented as (Sethuraman 1994),

$$f(t \mid G) = \sum_{l=1}^{\infty} p_l k(t \mid \theta_l),$$

where $k(\cdot)$ and θ_l are generic notations for kernel function and mixing parameters. p_l are mixture weights follow a stick-breaking process, $p_l = v_h \prod_{r<l}(1 - v_r)$, with $v_r \overset{i.i.d.}{\sim}$ Be$(1, \alpha)$, and atoms θ_l are i.i.d. from the baseline distribution. Our investigation is specifically on combining a log-logistic kernel with DPM.

The DPM models allow a straightforward extension to incorporate covariates through either atoms or mixture weights, or both, in the DP prior, referred as dependent DP (DDP) (MacEachern 2000). Specifically, the density function is written as:

$$f(t \mid G, x) = \sum_{l=1}^{\infty} p_l(x) k(t \mid \theta_l(x)),$$

where x is a generic notation for covariates. Various methods have been explored in the literature to introduce covariates through atoms or weights. Pioneering work of covariates-dependent mixture weights has been introduced by Müller et al. (1996) where a joint framework modeling both response and covariates via a mixture of multivariate normals under DPM. This methodology has since been expanded upon by researchers as Taddy & Kottas (2010), Wade et al. (2014), DeYoreo & Kottas (2018), and has found applications across various disciplines. Product partition models (PPM) share much in common with the joint modeling methodologies, which were first studied by Hartigan (1990) and Barry & Hartigan, and later extended by Müller et al. (2011) and Park & Dunson (2010) to account for covariates. This congruence is particularly evident when the PPM-based response partition aligns with a DP-induced model and

the covariate model is probabilistic in nature. Another branch of extensions in atoms has been explored by De Iorio, Müller, Rosner & MacEachern (2004) and De Iorio, Johnson, Müller & Rosner (2009), where a linear combination of covariates replaces the mean parameter in the normal kernel, thus referred as the linear DDP (LDDP) model. These methodologies have also been studied in the work of Shi et al. (2021), which built upon the DPM model with a Weibull kernel (Kottas 2006), for competing risks regression. In this study, we extend the DPM-based log-logistic model by combing the response-covariates joint modeling approach with LDDP.

The remainder of this article is structured as follows: Section 2 details the methodology, encompassing approaches to prior specification and posterior computation (with the appendix offering in-depth simulation details). Sections 3 and 4 present results from synthetic and real data examples, respectively. We conclude with a summary in Section 5.

3.2 Methodology

This section begins by presenting the intrinsic characteristics of the log-logistic distribution in Section 3.2.1 followed by a discussion of its central role as a kernel function under the DPM framework for survival analysis in Section 3.2.2. Then we describe the prior specification mechanism and posterior computation in Section 3.2.3.

3.2.1 Log-logistic kernel

The log-logistic distribution has demonstrated its value in the field of survival analysis (Bennett 1983). This distributions is similar with the log-normal density in shape but has a heavier tail. Under specific parameter conditions, its form can decrease monotonically. Moreover, there are closed-form expression for its density, survival, and hazard functions.

The functions for the log-logistic distribution are as follows:

$$f_{\mathrm{LL}}(t \mid \theta, \phi) = \frac{\frac{\phi}{\theta}\left(\frac{t}{\theta}\right)^{\phi-1}}{\left\{1+\left(\frac{t}{\theta}\right)^{\phi}\right\}^2} \text{ and } S_{\mathrm{LL}}(t \mid \theta, \phi) = \frac{1}{1+\left(\frac{t}{\theta}\right)^{\phi}} \text{ and } h_{\mathrm{LL}}(t \mid \theta, \phi) = \frac{\frac{\phi}{\theta}\left(\frac{t}{\theta}\right)^{\phi-1}}{1+\left(\frac{t}{\theta}\right)^{\phi}},$$

where $f_{\mathrm{LL}}(\cdot \mid \theta, \phi)$, $S_{\mathrm{LL}}(\cdot \mid \theta, \phi)$ and $h_{\mathrm{LL}}(\cdot \mid \theta, \phi)$ respectively represent density, survival, and hazard functions, parameterized by a scale parameter θ and a shape parameter ϕ. The median of the log-logistic model is equivalent to the model's scale parameter.

The density and hazard function exhibit a unimodal shape when the shape parameter ϕ is greater than 1; otherwise, the density decreases monotonically. The shape parameter is also associated with the model's dispersion in that a larger value of ϕ indicates a more tightly concentrated model.

Besides the similarities in shape between the log-logistic and log-normal distributions, there's a deeper relationship between them. Our research reveals a log-normal scale mixture representation for the log-logistic model.

Remark 1 *Let $f_{LL}(t \mid \theta, \phi)$ denote the density of a log-logistic distribution with scale parameter $\theta > 0$ and shape parameter $\phi > 0$, and let $f_{LN}(t \mid \mu, \sigma^2)$ denote the density*

of a log-normal distribution with parameters $\mu \in \mathbb{R}$ *and* $\sigma^2 > 0$*. Consequently, the log-logistic distribution can be represented as a scale mixture of log-normal distributions:*

$$f_{LL}(t \mid \theta, \phi) = \int_0^\infty f_{LN}\left(t \mid \log(\theta), (u\phi^2)^{-1}\right) p(u)\, du,$$

where $p(u)$ *is an alternating-sign sum of inverse-Gamma densities given by*

$$p(u) = \sum_{n=1}^{\infty} 2(-1)^n f_{\text{inv-}Ga}\left(u \mid 1, \frac{(n+1)^2}{2}\right),$$

and $f_{\text{inv-}Ga}(u \mid \alpha, \beta)$ *represents the density of an Inverse-Gamma distribution with shape parameter* $\alpha > 0$ *and scale parameter* $\beta > 0$*.*

This finding mirrors the work of Stafanski (1991), who convincingly showed that a standard logistic distribution can be represented as a scale mixture of the standard normal distribution, intricately linked with a Kolmogorov-Smirnov-related distribution.

3.2.2 Model for Survival Analysis

We now advance our discussion to the integration of the log-logistic kernel within the DPM framework for survival analysis, referred as the DPM-LL model. The proposed model has the following formulation,

$$f(t \mid G) = \int f_{\text{LL}}(t \mid \theta, \phi) dG(\theta, \phi) \quad \text{and} \quad G \sim DP(\alpha, G_0),$$

where α is the DP total mass parameter and G_0 represents the baseline distribution. The versatility of this model arises from the mixing on both scale and shape parameters of the log-logistic distribution, thereby enabling the modeling of a wide spectrum of

distributional shapes on $\mathbb{R}^+$. The density function and survival function have an almost sure representation as discussed in Sethuraman (1994), using the stick-breaking process definition. They are expressed as follows:

$$f(t \mid G) = \sum_{l=1}^{\infty} p_l f_{\mathrm{LL}}(t \mid \theta_l^*, \phi_l^*) \text{ and } S(t \mid G) = \sum_{l=1}^{\infty} p_l S_{\mathrm{LL}}(t \mid \theta_l^*, \phi_l^*), \tag{3.1}$$

where stick-breaking weights $(p_1, p_2, \dots)$ are generated through latent i.i.d. random variables $v_r \overset{i.i.d.}{\sim} \mathrm{Be}(1, \alpha)$, with $p_1 = v_1$, $p_l = v_l \prod_{r=1}^{l-1}(1 - v_r)$ for $l = 1, 2, \dots$. The distinct values (θ_l^*, ϕ_l^*) are independently generated from the baseline distribution.

In survival analysis, the hazard function is a fundamental quantity. It is defined as the instantaneous failure rate for an event (such as death) to occur, given that the event has not yet happened up to that time. Under the stick-breaking process construction, we can the hazard function of the DPM-LL model as a time-dependent mixture of log-logistic hazard functions:

$$h(t \mid G) \quad = \quad \sum_{l=1}^{\infty} \tilde{p}_l(t) h_{\mathrm{LL}}(t \mid \theta_l^*, \phi_l^*), \tag{3.2}$$

where $\tilde{p}_l(t) = p_l S_{\mathrm{LL}}(t \mid \theta_l^*, \phi_l^*) / \sum_{r=1}^{\infty} p_r S_{\mathrm{LL}}(t \mid \theta_r^*, \phi_r^*)$. This mixture representation with time-varying weights facilitates local adjustment for capturing various shapes of the hazard function, thus overcoming the shape restriction of individual log-logistic hazard functions.

A vital component of this DPM-LL model is the baseline distribution G_0. It determines the expected distribution of the proposed model as $E(f(\cdot \mid G)) = \int f_{\mathrm{LL}}(\cdot \mid \theta, \phi) dG_0(\theta, \phi)$. The choice of the baseline distribution aims to enable efficient computation, allow model's flexibility, and incorporate prior information through its hyper-

parameters. To achieve conditional conjugacy with the hyperparameters, we place the prior base measure on $\log(\theta)$ and ϕ^2 as:

$$G_0(\theta, \phi \mid \mu_\theta, \sigma_\theta^2, a_\phi, b_\phi) = \mathrm{N}(\log(\theta) \mid \mu_\theta, \sigma_\theta^2)\mathrm{Ga}(\phi^2 \mid a_\phi, b_\phi),$$

where $\mathrm{Ga}(\cdot \mid a, b)$ denotes the Gamma distribution with a mean of ab. In the following section, the prior specification and posterior computation method will be discussed.

3.2.3 Prior specification and posterior computation

The DP prior for random mixing distribution G, given by $DP(\alpha, G_0)$, is crucial in the context of our proposed model construction. The total mass parameter α of the DP determines the variability of the generated realizations from a prior distribution. As recommended by Escobar & West (1995), we assign a Gamma prior distribution to α, denoted as $\alpha \sim \mathrm{Ga}(a_\alpha, b_\alpha)$. In our experience in simulation studies, letting $a_\alpha = 2$ and $b_\alpha = 1$ is sufficient for functional estimations tasks. The prior density function given the choice of the baseline distribution is affected by four hyperparameters, including μ_θ, σ_θ^2, a_ϕ and b_ϕ. Each of the hyperparameter could be assigned a hyper prior to enhance the model's effectiveness.

For simplicity, σ_θ^2 and a_ϕ can be pre-fixed. Let μ_θ have a $\mathrm{N}(s_\theta, S_\theta)$ prior with mean s_θ and variance S_θ, and b_ϕ have a inv-$\mathrm{Ga}(r_\phi, R_\phi)$ prior.

This results in a marginal prior base measure of $\log(\theta)$ following a normal distribution, $\mathrm{N}(s_\theta, S_\theta + \sigma_\theta^2)$, and the marginal prior base measure of ϕ^2 is proportional to $(\phi^2)^{a_\phi - 1}(\phi^2 + R_\phi)^{-(a_\phi + r_\phi)}$ with mean value equal to $a_\phi R_\phi/(r_\phi - 1)$. When $R_\phi = 1$,

the distribution is a Beta prime. In practice, the hyperparameter r_ϕ can be set as the smallest integer value (3) that produces and an inverse-gamma distribution with finite mean and finite but largest variance. The selection of values for a_ϕ, r_ϕ and R_ϕ allows incorporation of prior information about the density function's shape and dispersion. Specifically, smaller values of ϕ^2 correspond to more decreasing mixture components with large dispersion.

To validate the hyperparameter setting, it is helpful to visualize prior estimates. Under a given set of hyperparameters, the prior estimates for the density, survival, and hazard functions can be approximated using the truncated version of Equations (3.1) and (3.2) at level of N where N is chosen large enough to ensure that $\sum_{l=1}^{N} p_l \approx 1$.

Our method to posterior computation is founded on an augmentation model that enables a Gibbs sampler. Each subject is assigned a membership indicator, L_i, which takes values in $1, \ldots, k$. Here, k represents the total number of distinct atoms arising from the DP prior, and t_i denotes the survival times of the n subjects. As such, the following hierarchical model can be structured:

$$\begin{aligned}
\mathrm{P}(L_1, \ldots, L_n \mid \alpha) &= \alpha^k \frac{\prod_{l=1}^{k} \Gamma(n_l)}{\prod_{i=1}^{n} (\alpha + i - 1)}, \\
\log(\theta_l^*) \mid \mu_\theta, \sigma_\theta^2 &\overset{i.i.d.}{\sim} \mathrm{N}(\mu_\theta, \sigma_\theta^2) \text{ for } l = 1, \ldots, k, \\
\phi_l^2 \mid a_\phi, b_\phi &\overset{i.i.d.}{\sim} \mathrm{Ga}(a_\phi, b_\phi) \text{ for } l = 1, \ldots, k, \\
t_i \mid L_i, \boldsymbol{\theta^*}, \boldsymbol{\phi^*} &\overset{ind.}{\sim} \mathrm{LL}(\theta_{L_i}^*, \phi_{L_i}^*) \text{ for } i = 1, \ldots, n,
\end{aligned}$$

where $\boldsymbol{\theta^*} = (\theta_l^* : l = 1, \ldots, k)$, and $\boldsymbol{\phi^*} = (\phi_l^* : l = 1, \ldots, k)$.

Survival analysis often contends with right censoring, a common issue where

observations of study participants are curtailed due to non-informative reasons such as administrative constraints. Generally, data are recorded as $y_i = \min(t_i, c_i)$ for $i = 1, \ldots, n$, where t_i indicates survival time and c_i indicates the right censoring time. Given this, the entire dataset can be expressed as $\mathcal{D} = (y_i, \nu_i) : i = 1, \ldots, n$, where ν_i is a binary censoring indicator, with $\nu_i = 1$ if t_i is observed and $\nu_i = 0$ if t_i is right-censored.

Then the likelihood function of the augmented model can be written as

$$L(\mathbf{\Phi}; \mathcal{D}) = \sum_{i=1}^{n} \{f_{\text{LL}}(y_i \mid \theta^*_{L_i}, \phi^*_{L_i})\}^{\nu_i} \{S_{\text{LL}}(y_i \mid \theta^*_{L_i}, \phi^*_{L_i})\}^{1-\nu_i},$$

where $\mathbf{\Phi}$ indicates all parameters.

We implement posterior inference via Markov chain Monte Carlo (MCMC) simulation, using standard posterior simulation methods for DPM models (e.g., Escobar & West (1995), Neal (2000)). The computational details can be found in Appendix B.

3.3 Model extension for regression

In survival analysis, survival responses are often collected with covariates. There might be a complex relationship between responses and covariates. To address this challenge, we extend the DPM-LL model to incorporate with covariates and add it to the regression toolkit. In practical applications, there are cases where certain covariates should not be treated as random variables. An example would be in a controlled experiment where group assignments are predetermined, not random. Suppose we have a p-dimensional covariate vector, $\boldsymbol{x}$, which is random, and a q-dimensional covariate vector, $\boldsymbol{z}$, which we consider as fixed. As discussed, we proposed a regression model

combing the response-covariates joint model and LDDP in the following way:

$$f(t, \boldsymbol{x} \mid G, \boldsymbol{z}) = \sum_{l=1}^{\infty} p_l f_{\mathrm{LL}}(t \mid \theta_l^* \exp\{(\boldsymbol{x}', \boldsymbol{z}')\boldsymbol{\beta}_l^*\}, \phi_l^*) f(\boldsymbol{x} \mid \boldsymbol{\Omega}_l^*),$$

where $f(\boldsymbol{x} \mid \boldsymbol{\Omega})$ is a general representation for the joint probability density/mass function of covariates. $\{p_l, l = 1, 2, \dots\}$ are stick-breaking weights and $\{(\theta_l^*, \boldsymbol{\beta}_l^*, \phi_l^*, \boldsymbol{\Omega}_l^*) : l = 1, \dots\}$ are i.i.d. from the baseline distribution G_0, where $\boldsymbol{\beta}_l^*$ is $p + q$ dimensional. The resulting conditional density function of t is then

$$f(t \mid \boldsymbol{x}, G, \boldsymbol{z}) = \sum_{l=1}^{\infty} \omega_l(\boldsymbol{x}) f_{\mathrm{LL}}(t \mid \theta_l^* \exp((\boldsymbol{x}', \boldsymbol{z}')\boldsymbol{\beta}_l^*), \phi_l^*), \tag{3.3}$$

where $\omega_l(\boldsymbol{x}) = p_l f(\boldsymbol{x} \mid \boldsymbol{\Omega}_l^*) / \sum_{r=1}^{\infty} p_r f(\boldsymbol{x} \mid \boldsymbol{\Omega}_r^*)$. The conditional survival and hazard functions can also be obtained in the same manner, expressed as,

$$S(t \mid \boldsymbol{x}, G, \boldsymbol{z}) = \sum_{l=1}^{\infty} \omega_l(\boldsymbol{x}) S_{\mathrm{LL}}(t \mid \theta_l^* \exp((\boldsymbol{x}', \boldsymbol{z}')\boldsymbol{\beta}_l^*), \phi_l^*), \tag{3.4}$$

and

$$h(t \mid \boldsymbol{x}, G, \boldsymbol{z}) = \sum_{l=1}^{\infty} \tilde{\omega}_l(t, \boldsymbol{x}) h_{\mathrm{LL}}(t \mid \theta_l^* \exp((\boldsymbol{x}', \boldsymbol{z}')\boldsymbol{\beta}_l^*), \phi_l^*), \tag{3.5}$$

where

$$\frac{\tilde{\omega}_l(t, \boldsymbol{x}) = \omega_l(\boldsymbol{x}) S_{\mathrm{LL}}(t \mid \theta_l^* \exp((\boldsymbol{x}', \boldsymbol{z}')\boldsymbol{\beta}_l^*), \phi_l^*)}{\sum_{r=1}^{\infty} \omega_r(\boldsymbol{x}) S_{\mathrm{LL}}(t \mid \theta_r^* \exp((\boldsymbol{x}', \boldsymbol{z}')\boldsymbol{\beta}_r^*), \phi_r^*)}.$$

The choice of covariates kernel functions, $f(\boldsymbol{x} \mid \boldsymbol{\Omega})$, and their corresponding baseline distribution, $G_0(\boldsymbol{\Omega})$, depends on the nature of the covariates, which may be continuous, count, or categorical. Considering the computational ease, we assume conditional independence between covariates, as, $f(\boldsymbol{x} \mid \boldsymbol{\Omega}) = \prod_{j=1}^{p} f(x_j \mid \Omega_j)$ and $G_0(\boldsymbol{x}) = \prod_{j=1}^{p} G_0^{(j)}(\Omega_j)$. Suitable conjugate models are selected for each type: a

normal-normal model for continuous covariates, a Poisson-Gamma for a count-type, and Beta-Binomial for categorical ones,

Recall from Section 3.2.2 that prior base measures for $\log(\theta)$ and ϕ^2 are defined as $\mathrm{N}(\mu_\theta, \sigma_\theta^2)$ and $\mathrm{Ga}(\phi^2 \mid a_\phi, b_\phi)$, respectively. We extend these to include independent baseline distributions for the mixing parameter $\boldsymbol{\beta}$ as a product of normal distributions, $\prod_{j=1}^{p+q} \mathrm{N}(\beta_j \mid \mu_{\beta j}, \sigma_{\beta j}^2)$. This configuration ensures the capability of dealing with various types of covariates while retaining the desired computational efficiency.

To complete the Bayesian framework, we begin by fixing $\sigma_{\beta_j}^2$, for $j = 1, \ldots, p+q$, and assign independent normal hyper prior distributions to each $\mu_{\beta_j} \overset{ind.}{\sim} \mathrm{N}(s_{\beta_j}, S_{\beta_j})$, which yields a marginal multivariate normal baseline distribution of $\boldsymbol{\beta}$, expressed as $\mathrm{N}_{p+q}(\boldsymbol{s}_\beta, \boldsymbol{S}_\beta + \Sigma_\beta)$, where $\boldsymbol{s}_\beta = (s_{\beta_1}, \ldots, s_{\beta_{p+q}})'$, $\boldsymbol{S}_\beta = \mathrm{Diag}(S_{\beta_1}, \ldots, S_{\beta_{p+q}})$, $\Sigma_\beta = \mathrm{Diag}(\sigma_{\beta_1}^2, \ldots, \sigma_{\beta_{p+q}}^2)$. As for parameter settings, we suggest a $\boldsymbol{s}_\beta = (0, \ldots, 0)'$, based on our experiences with synthetic data. Also $\sigma_{\beta_j}^2$ and S_{β_j} are recommended to be set to a small number between 0.5 and 1.5, although the specific selection may vary depending on the context.

Similar to DPM-LL model, prior estimates can be used to validate hyperparameter specification. For a specified set of hyperparameters and covariates $\boldsymbol{x}$, the prior estimates of the conditional density (3.3), survival (3.4), hazard (3.5) functions can be approximated through truncation.

The implementation of posterior simulation for the BNP log-logistic regression model is based on its DPM representation of the model for both survival responses and covariates. Thus, the computation extends naturally from the previous algorithm.

Detailed procedures are provided in Appendix C.

3.4 Simulation study

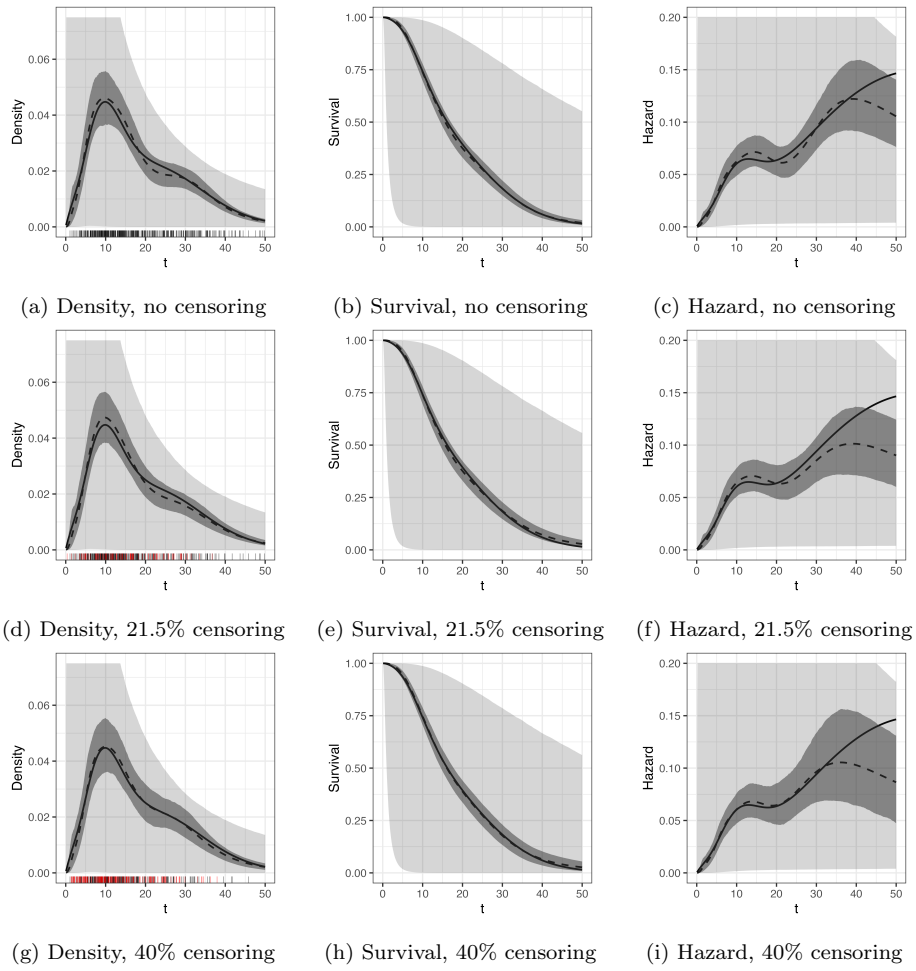
In this section, we demonstrate the application of the proposed model, detailed in Section 3.2.2 through synthetic data examples. This first dataset is constructed from a mixture of three Gamma distributions, manifesting in a non-standard unimodal shape for its density function. We further examine the model's performance in the presence of right-censored observations due to independent censoring.

In addition, we will illustrate the extension of the model in Section 3.4.2, using synthetic data with one binary and one continuous covariate. The goal here is to exhibit how our model can be adeptly adapted as a tool for quantile regression, thereby demonstrating its multifaceted utility.

For all data examples considered in Sections 3.4 and 3.5, we specify the prior hyperparameters using the approach described in Sections 3.2 and 3.3. Standard diagnostic techniques are employed to examine convergence and mixing of the MCMC algorithms. We run the MCMC chain for 20,000 iterations, then discard the first 5,000 samples, and use every seventh sample from the remaining iterations for the posterior inference.

3.4.1 Simulation example 1: survival times without covariate

In this example, 400 survival times are simulated from a mixture of three Gamma distributions, given by $0.3\text{Ga}(6,2) + 0.4\text{Ga}(2,8) + 0.3\text{Ga}(10,3)$, which yields



(a) Density, no censoring (b) Survival, no censoring (c) Hazard, no censoring

(d) Density, 21.5% censoring (e) Survival, 21.5% censoring (f) Hazard, 21.5% censoring

(g) Density, 40% censoring (h) Survival, 40% censoring (i) Hazard, 40% censoring

Figure 3.1: Synthetic data example 1. Each row presents functional estimates on datasets with varying degrees of censoring (0%, 21.5%, and 40%). Each column corresponds to different functional estimates: density, survival, and hazard function, respectively. Each subfigure shows the posterior mean estimates (dashed lines), the 95% pointwise posterior credible intervals (dark shaded regions), the 95% prior pointwise credible intervals (light shaded regions), and simulation truth (solid lines). Rug plots in the first column display observed (black) and censored (red) survival times.

a complex unimodal density. With respect to the hazard function, a subtle decline is observable from $t = 10$ to $t = 20$. The true underlying functions for density, survival, and hazard are displayed as solid lines in Figure 3.1(a)-(c). The prior distributions are designed as follows: $\alpha \sim \text{Ga}(2, 1)$, $\mu_\theta \sim \text{N}(2, 0.5)$, $b_\phi \sim \text{inv-Ga}(3, 20)$, while a_ϕ and σ_θ^2 are fixed at 4 and 1 respectively. The corresponding 95% pointwise prior credible interval estimates are also illustrated with light shaded regions. With these hyperparameter values, the prior interval estimates exhibit wide uncertainty bands over the range of observed survival times, as illustrated by the rug plot in black in each subfigure.

Figures 3.1(a)-(c) present the posterior point and interval estimates as dashed lines and dark shaded regions, respectively. The non-standard unimodal shape of the true underlying functions is effectively captured by the model. The model's substantial learning from the data is visibly represented through the transition from light to dark regions in each panel. In particular, the inference results for the hazard function demonstrate the local adjustment, where the decreasing hazards from $t = 10$ to 20 is captured.

To examine the model's performance with censored data, we employed a random censoring mechanism and generated 400 censoring times from a Gamma distribution, denoted by c_i for $i = 1, \ldots, n$. As discussed, we defined y_i as $\min(t_i, c_i)$, and ν_i as 1 if $c_i > t_i$, and 0 otherwise. By changing the parameter in the Gamma distribution for censoring times, we obtained two separate sets of data that contain 21.5% and 40% censoring, respectively. The same set of priors and the same MCMC implementation were employed for each of these censored datasets.

As visualized in panels (d)-(i) of Figure 3.1, censoring does not substantially impact the inference. This is evidenced by the consistent containment of the true function within the posterior interval estimates across in all instances, particularly for density and survival functions. Notably, the hazard function is well captured given a sufficient data, in these particular data examples, $t < 40$. As the proportion of censoring expands, the posterior uncertainty bands concurrently widen, a trend especially evident in the estimates for the hazard function.

3.4.2 Simulation example 2: survival times with two covariates

Here, we evaluate the performance of the extended model for handling non-linear regression problems. Consider a case with one binary (z) covariate indicating experimental group assignments and one continuous (x) covariate. Let $z_i = 0$, for $i = 1, \ldots, 300$, and $z_i = 1$ for $i = 301, \ldots, 600$. And let $x_i \sim \mathrm{N}(0, 1)$ for $i = 1, \ldots, 600$. Conditional on z and x, we simulate survival times from a log-normal distribution, $\mathrm{LN}(\mu_z(x), \sigma_z^2)$, where its median, $\exp(\mu_z(x))$, is a nonlinear function of x for each value of z, in detail, $4\exp(-0.3(x + 0.5)^2) + 0.5$ and $4\exp(-0.2x^2) + 2.5$, and σ_z^2 is set as $\sigma_0^2 = 0.6^2$ and $\sigma_1^2 = 0.8^2$.

In line with our previous discussion, we used a normal kernel for the continuous covariate x. A hierarchical representation can be constructed with the use of a latent variable, such as $x \mid \mu \sim N(\mu, \tau^2)$, $\mu \sim \mathrm{N}(\mu \mid s_\mu, S_\mu)$. τ^2, s_μ, and S_μ are fixed at prespecified values. The marginal normal model is obtained for μ as $\mathrm{N}(\mu \mid s_\mu, S_\mu + \tau^2)$. If no further prior knowledge is provided for value assignments for τ^2, s_μ, and S_μ, we

recommend the standard approach that sets $s_\mu = 0$, and τ^2 to be comparably smaller than the sample variance of x, and let S_μ be large to increase prior variance. In this example, we let $s_\mu = 0$, $S_\mu = 10$, and $\tau^2 = 0.5$.

The hyper priors that are related to responses are set in the following way, $a_\phi = 10$, $r_\phi = 3$, $R_\phi = 20$, $s_\theta = 0.5$, $S_\theta = 0.5$, $\sigma_\theta^2 = 1$, $s_\beta = 0$, $S_\beta = 0.5$, and $\sigma_\beta^2 = 1$. In addition, we let $a_\alpha = 2$ and $b_\alpha = 1$. The pointwise prior 95% credible interval estimates are illustrated in Figure 3.2 in light gray shades providing large support for all the density, survival, and hazard functions.

Posterior inferences are summarized in Figure 3.2, with dark gray shades indicating 95% pointwise posterior estimates and dashed lines denoting posterior mean estimates, illustrating the model's success in capturing the true underlying function (solid lines) across the range of covariates. The transition from the prior to posterior estimates clearly demonstrates the model's substantial learning from the data. Our extended model can be readily adopted to a quantile regression model for any chosen percentile of survival times. Given covariates x_0, z_0, and one draw of posterior samples $\Theta^{(b)}$, the posterior Pth percentile survival time Q can be obtained through equation $0.01P = 1 - S(Q \mid x_0, \Theta^{(b)}, z_0)$ by searching through the discretization of $S(Q \mid x_0, \Theta^{(b)}, z_0)$, the details for functional evaluation are described at the end of Appendix C. Figure 3.2 displays estimated median and 75th percentile of survival times over grids of the continuous covariate x for $z = 0$ (in red) and $z = 1$ (in blue). The extended model is capable of capturing the nonlinear shape of two different percentiles of survival times over the grid of x, alongside the correct stochastic order of for $z = 0$

and 1, with its point estimates enclosing the true values within the 95% uncertainty bands.

3.5 Real data example

In this section, we examine the model on a real dateset comprising survival times (measured in days) of 121 patients diagnosed with small cell lung cancer, as detailed in Ying et al. (1995). These patients were randomly allocated to one of two treatments regimes, denoted as Arm A and Arm B. Arm A's protocol involved administering cisplatin (P), followed by etoposide (E), to the patients, while Arm B's treatment plan involved reversing this order, with etoposide (E) given first, followed by cisplatin (P). Of the 62 patients under Arm A's treatment, 15 experienced right censored survival times, and similarly, of the 59 patients in Arm B, 8 experienced right censored survival times. Additionally, patient age at the time of study entry is provided, which, for the purpose of this analysis, we have standardized to have a mean of zero and a variance of one.

For the purpose of demonstrating our model, we represent treatment assignments as z, and the standardized age as x. Similar to the simulation example 2 (in Section 3.4.2), we assume a normal kernel for x, thus $x \mid \mu, \tau^2 \sim \mathrm{N}(\mu, \tau^2)$, and $\mu \sim \mathrm{N}(s_\mu, S_\mu)$. The hyperparameters are specified as follows: we assign b_ϕ an inverse Gamma distribution, inv-Ga$(3, 10)$, and μ_θ a normal distribution, N$(6, 0.5)$. Similarly, μ_β is assigned a normal distribution, N$((0, 0)', 0.5\mathrm{I}_2)$, and α is assigned a Gamma dis-

tribution, Ga(2, 1). In addition, a_ϕ, σ^2_θ, σ^2_β, s_μ, S_μ, and τ^2 are fixed at 10, 1, 1, 0, 10, and 0.5, respectively. The 95% pointwise prior interval estimates for density given different covariates are displayed in Figure 3.4 with light gray shades with reasonable wide coverage.

Figure 3.4 also showcases the model's adaptability in capturing the conditional density estimates, where pointwise posterior point and interval estimates are shown in dark gray shades and dashed lines. The model consistently implies a unimodal shape for the conditional density across different values of the covariate. Notably, the conditional density for Arm A exhibits more discrepancy than Arm B at the same age. In Figure 3.5(a) and 3.5(b), posterior point estimates for the conditional survival and hazard functions are presented. For a given age, Arm A exhibits a higher survival curve and lower hazard curve than the Arm B. Within the same treatment group, there is a stochastic order of hazards and survival probabilities for ages 54, 62, and 70 groups, however, this order doesn't hold for age 45. In Arm A group, age 45 has crossing hazards with age 62 at $t = 1800$ and crossing survival probabilities with age 62 at $t = 1000$. The crossing hazards and survival probabilities also appear in Arm B group. This finding indicates a nonlinear relation between survival times and age given a treatment group.

To reveal the nonlinearity between age and survival times, we represent Figure 3.6 where the median and 75th percentile regression functions over the grids of age, separately for Arm A and Arm B are displayed. Both pointwise posterior median and 50% credible interval estimates are shown. The model infers a nonlinear relationship between survival times and age for both treatment arms, pointing the peak of the median

and 75th percentile survival times when age falls between $(55, 60)$. For both median and 75th percentile survival times, Arm A consistently exhibits higher survival times than Arm B across all ages.

3.6 Summary

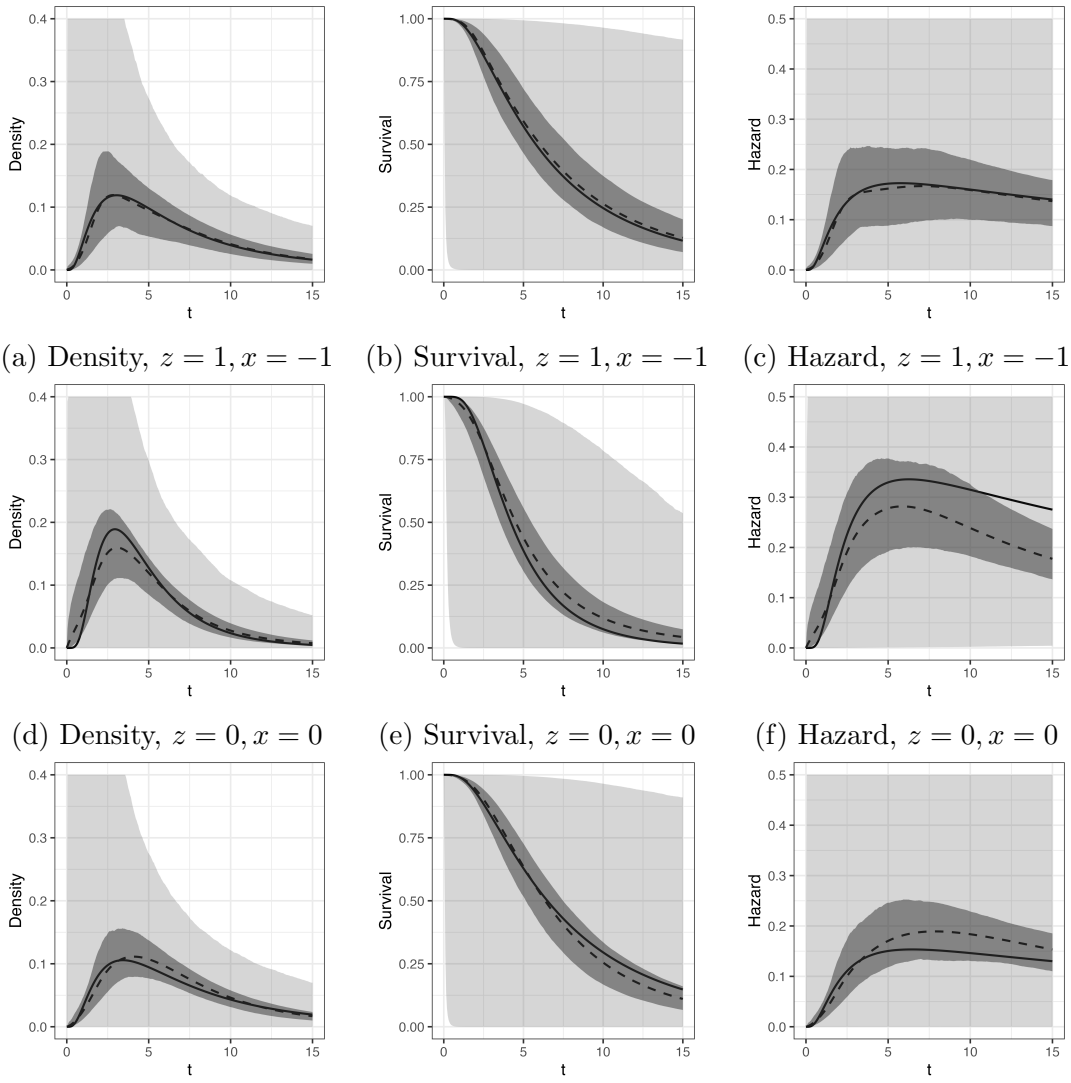
In summary, this research presents an advanced Bayesian nonparametric method tailored for comprehensive inference in survival data analysis. By integrating a Dirichlet process mixture model with a log-logistic kernel, which is a well-established tool in survival analysis but underrepresented in Bayesian nonparametric studies, we offer a nuanced understanding of the complexities inherent in survival data. Our approach is computationally efficient, facilitated by the use of Markov chain Monte Carlo techniques and Pólya-gamma data augmentation method for posterior computations. We also revealed a log-normal scale mixture representation of a log-logistic model.

A key innovation in our study is the extension of the core model to incorporate covariates via a density regression framework. This is achieved through core components.

This novel approach allows local adjustment in mixture weights of covariates which can be considered random variables. Moreover, in randomized experimental studies, experimental group assignments may not be considered random. The extended model consider those covariates as a part of the log-linear model. First, we introduce a log-linear model for the scale parameter of the log-logistic kernel, following the linear

dependent Dirichlet process model as outlined by De Iorio, Johnson, Müller & Rosner (2009). Second, we extend the model to include a joint kernel function that simultaneously models covariates and responses, building on the foundational work by Müller et al. (1996).

In conclusion, this study enriches toolkit available for advanced and flexible survival analysis approaches.

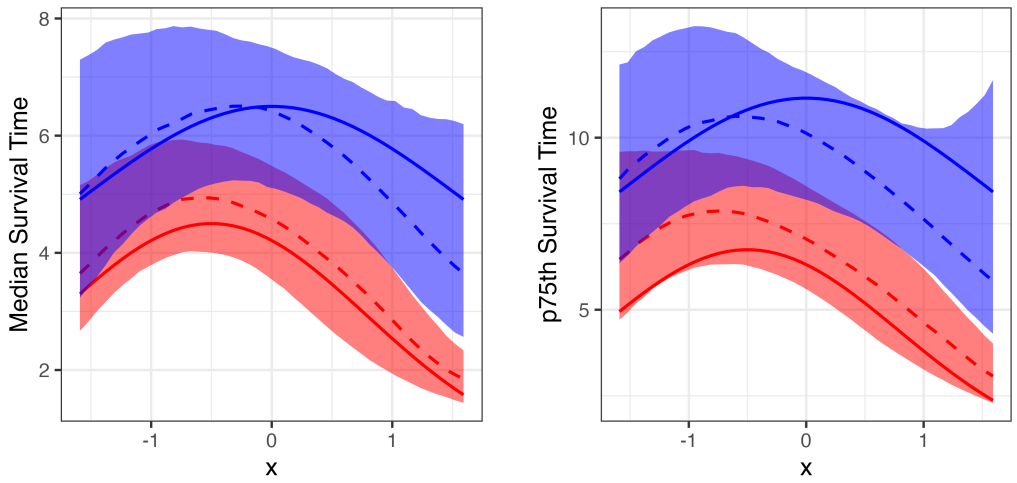


(a) Density, $z = 1, x = -1$ (b) Survival, $z = 1, x = -1$ (c) Hazard, $z = 1, x = -1$

(d) Density, $z = 0, x = 0$ (e) Survival, $z = 0, x = 0$ (f) Hazard, $z = 0, x = 0$

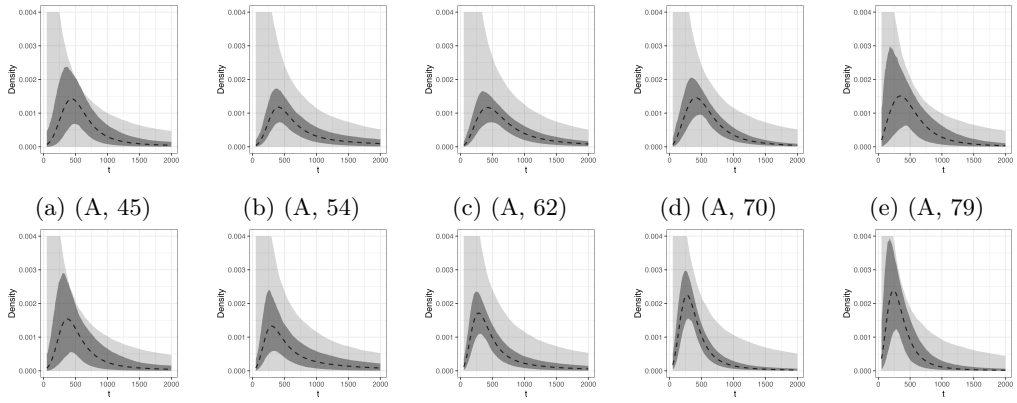
(g) Density, $z = 1, x = 0$ (h) Survival, $z = 1, x = 0$ (i) Hazard, $z = 1, x = 0$

Figure 3.2: Simulation 2. Each row presents functional estimates with varying values of covariate (z, x) ((1,-1), (0,0), (1,0)). Each column corresponds to different functional estimates: density, survival, and hazard function, respectively. Each subfigure shows the posterior mean estimates (dashed lines), the 95% pointwise posterior credible intervals (dark shaded regions), the 95% pointwise prior credible intervals (light shaded regions), and the simulation truth (solid lines).



(a) Median regression (b) 75th percentile regression

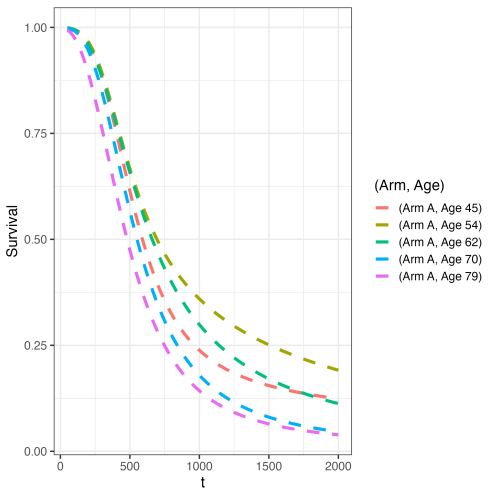
Figure 3.3: Simulation 2. The Figure displays pointwise posterior estimates for the median (panel a), and 75th percentile (panel b) of survival times across the continuous covariate x range (-1.64, 1.64). Point estimates are represented by dashed lines and 95% credible intervals are highlighted with shaded regions. The red color denotes $z = 0$ while the blue color indicates $z = 1$. The underlying truth is represented by solid lines.



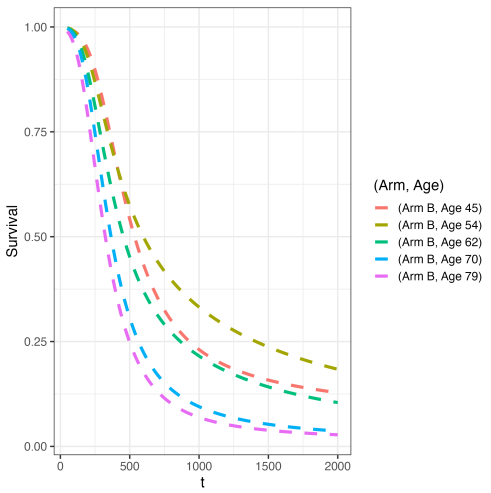
(a) (A, 45) (b) (A, 54) (c) (A, 62) (d) (A, 70) (e) (A, 79)

(f) (B, 45) (g) (B, 54) (h) (B, 62) (i) (B, 70) (j) (B, 79)

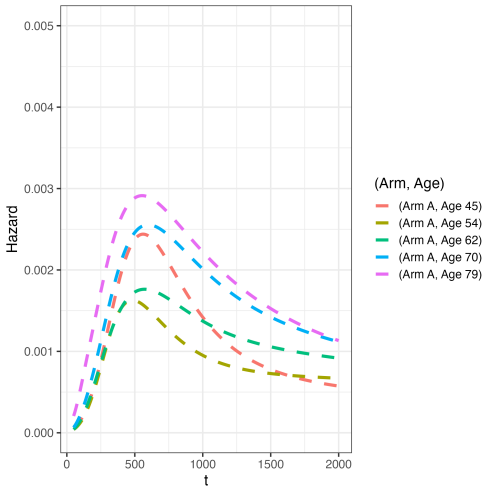
Figure 3.4: Small cells lung data. Each row presents survival estimates with different treatment Arm (A and B). Each column corresponds to different values of age (45, 54, 62, 70, 79). Each subfigure shows the posterior mean estimates (dashed lines), 95% pointwise posterior credible intervals (dark shaded regions), and 95% prior interval estimates (light shaded regions).



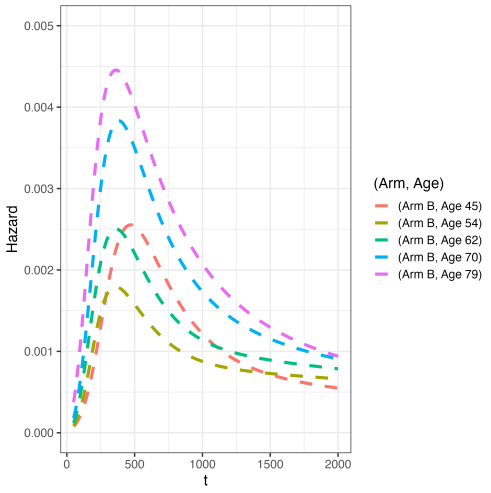
(a) Survival, Arm A



(b) Survival, Arm B

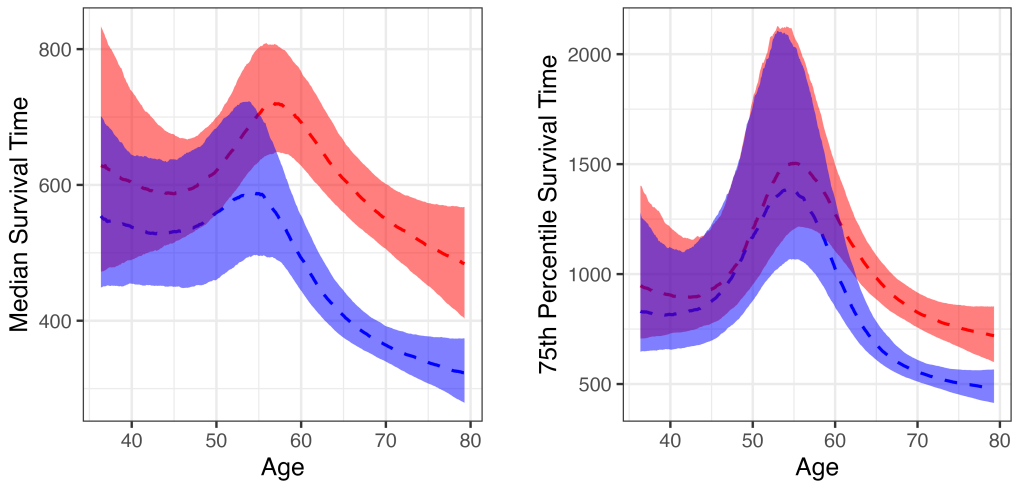


(a) Hazard, Arm A



(b) Hazard, Arm B

Figure 3.5: Small cells lung data. The first row shows the posterior mean estimates of the conditional survival function, while the second panel displays posterior estimates of the conditional hazard function. The left and right columns are corresponding to Arm A and Arm B, respectively. Each color corresponds to a distinct age of patients.



(a) Median regression (b) 75th percentile regression

Figure 3.6: Small cells lung data. The left panel shows the pointwise posterior mean estimates (dashed lines) and the 50% pointwise posterior credible intervals for the median survival times over the range of the standardized age. The right panel shows the posterior point and interval estimates for the 75th percentile of survival times. The treatment groups are color-coded, with Arm A represented in red and Arm B in blue.

Chapter 4

Bayesian Nonparametric Regression Model for Joint Modeling of Recurrent Events and Survival Times

4.1 Introduction

Recurrent events commonly arise in clinical trials, reliability studies, and other fields, for instances, recurrent infections, rehospitalizations, and machine repairs. These sequences of events can often be terminated by some fatal events, such as death or a critical system failure. It is possible that the frequency of recurrent events and timing of termination events (survival times) are related. Thus, the development of a nuance understanding of the dependence between recurrent events and survival times is crucial. Addressing this need effectively requires a sophisticated and flexible joint modeling framework that can intricately model the dependencies between recurrent and terminal

events.

Historically, the statistical modeling of recurrent events has been approached by two primary frameworks, as reviewed by Cook & Lawless (2007). The first framework assumes a continuous intensity function, typically employing various Poisson-type processes to model for recurrent events. Alternatively, the second framework makes a renewal-process assumption. This modeling approach concentrates on the density function of the gap (interval-arrival) times between events, offering a different perspective on the temporal structure of the data.

Survival analysis is a well-studied topic in statistics, with proportional hazards (PH) models being the most well-known for survival regression, where covariates contribute multiplicatively to the hazard rate. Accelerated failure time (AFT) models offer an alternative approach, assuming a probabilistic model for log-transformed survival times. Comprehensive reviews of both frequentist and Bayesian methods can be found in Klein & Moeschberger (1997) and Ibrahim et al. (2001).

Under the context of the joint analysis of recurrent events and survival times, a commonly used modeling approach for establishing dependency is through the use of random effects. Conditional on the random effects, recurrent events and survival times are assumed to be independent. Typically, recurrent events are modeled using a nonhomogeneous Poisson process (NHPP) regression model, while survival times are often approached with PH or AFT models. This method has been widely adopted across various studies, including those by Liu et al. (2004), Rondeau et al. (2007), and Ye et al. (2007), with a notable focus on Bayesian implementations by Sinha et al.

(2008), Ouyang et al. (2013), and Lee et al. (2019). Specifically, in the studies by Sinha et al. (2008) and Ouyang et al. (2013), a PH approach is used. Both studies assume a shared Gamma random effects model, where the NHPP model and survival model share the same effects. On the other hand, Lee et al. (2019) employs an AFT approach and enhances model flexibility by placing a Dirichlet process (DP) prior on random effects model.

An alternative strategy is to utilize the second approach for recurrent events under the joint modeling framework. For instance, Huang & Liu (2007) applied PH models for both gap and survival times within a frequentist framework. This model was further enhanced by Yu & Liu (2011) who incorporated a nonlinear covariate function via penalized splines within the recurrent events component. Tawiah et al. (2019) incorporated a cure fraction into their model development, expanding its applicability.

From a Bayesian perspective, Paulon et al. (2020) adopted AFT models for conditional survival and gap times. Instead of modeling the density function of gap times directly, Boom et al. (2022) chose to model the number of recurrent events, then implemented an autoregressive model to relax the independence assumption among gap times. Both studies employed a DP prior on random effects. In a more specialized context, Wen et al. (2016) addressed the complexity of having multiple types of recurrent events (local and distant) by using three PH models to simultaneously model these recurrent events along with survival times, incorporating a bivariate random effects model for each type of recurrent event and a combination of two random effects for survival times. Similar to these approaches, we assume subject-specific random effects

to introduce dependencies between survival times and recurrent events. Additionally, we relax the PH and AFT assumptions and model the density functions of survival and gap times within a Bayesian nonparametric framework.

The main novel methodological contribution of this work is to model conditional densities of gap and survival times using Bayesian nonparametric regression models. We place dependent Dirichlet process (DDP) mixture priors on the density functions of survival and gap times. The DDP, as initially conceptualized by MacEachern (2000), provides flexibility in modeling a collection of distributions varying across covariate space. In the DDP, each set of random measures is configured to be marginally a DP-distributed random measure for any given covariate value $\boldsymbol{x} \in \chi$. This is achieved by adapting the stick-breaking process construction (Sethuraman 1994) of a DP, which can be expressed as:

$$G_{\boldsymbol{x}}(\cdot) \quad = \quad \sum_{l=1}^{\infty} \left\{ V_l(\boldsymbol{x}) \prod_{r<l} [1 - V_r(\boldsymbol{x})] \right\} \delta_{\theta_l(\boldsymbol{x})}(\cdot),$$

where $V_l(\boldsymbol{x})$ are independent stochastic processes across the covariate space χ with Beta marginal distributions. The $\theta_l(\boldsymbol{x})$ are also independent stochastic processes with index set χ and $G_{\boldsymbol{x}0}$ marginal distributions. This Bayesian nonparametric regression modeling approach extends traditional joint models to accommodate more complex data structures across the covariates space.

The rest of the article is structured to in the following way. In Section 4.2, we briefly reviews the joint modeling framework with survival and gap times. The detailed description of methodological development is introduced in Section 4.3. The

model is then illustrated with an Esophageal cancer patients dataset in Section 4.4. We summarize the paper in Section 4.5

4.2 Joint Model for Gap and Survival Times

Consider a dataset comprising n individuals, where $Y_{i0} := 0$ marks the initiation of recurrent events. The observation endpoint τ_i represents either the survival time T_i or an independent right-censored time C_i, hence $\tau_i = \min(C_i, T_i)$. The censoring status is indicated by a binary variable ν_i, with 0 for censored observations and 1 for recorded deaths.

Denote $N_i \geq 0$ as the count of recurrent events observed within $(0, \tau_i]$ at times $Y_{i0} = 0 < Y_{i1} < \cdots < Y_{iN_i} < \tau_i$. The intervals between recurrent events, termed gap times, are defined as $W_{ij} = Y_{ij} - Y_{i,j-1}$ for $j = 1, \ldots, N_i$. It is noteworthy that the final gap time for each subject is censored by their survival or censored time, $\tau_i - Y_{i,N_i}$.

As mentioned, our approach for modeling the dependency between survival and gap times involves subject-specific random effects. This framework generally posits conditional independence between survival and gap times given random effects. Unlike the popular shared random effects, we assume correlated random effects, (ϵ_i, ξ_i), for i^{th} subject. This model enhances flexibility by enabling simultaneous control over the magnitude of individual random effects and the degree of dependency between them. A similar random effects model has been used in Tawiah et al. (2019). Suppose $\boldsymbol{x}$ is a covariates vector. The models for survival times and gap times are formulated as follows

for $i = 1, \ldots, n$:

$$T_i \mid \epsilon_i, \boldsymbol{x}_i \overset{ind.}{\sim} f_S(t_i \mid \epsilon_i, \boldsymbol{x}_i),$$

$$W_{ij} \mid \xi_i, \boldsymbol{x}_i \overset{i.i.d.}{\sim} f_R(w_{ij} \mid \xi_i, \boldsymbol{x}_i), j = 1, \ldots, N_i,$$

where f_S and f_R denote regression models for survival and gap times, respectively. The likelihood function for all observations is written as,

$$\mathrm{L} = \prod_{i=1}^{n} \{f_S(\tau_i \mid \epsilon_i, \boldsymbol{x}_i)\}^{\nu_i} \{S_S(\tau_i \mid \epsilon_i, \boldsymbol{x}_i)\}^{1-\nu_i} \prod_{j=1}^{N_i} f_R(w_{ij} \mid \xi_i, \boldsymbol{x}_i) S_R(\tau_i - w_{i,N_i} \mid \xi_i, \boldsymbol{x}_i),$$

where S_S and S_R denote the survival functions corresponding to survival and gap times, respectively. This formulation accounts for censored observations through their survival function, ensuring that the likelihood function incorporates both the observed events and the censored information.

In the existing literature, the functions f_S and f_R are commonly modeled using the PH models (Huang & Liu 2007, Yu & Liu 2011) or AFT models (Paulon et al. 2020). These models impose constraints that prevent the hazard or survival functions from crossing. To address the need for greater modeling flexibility across the covariate space, we introduce DDP mixture models in the subsequent section.

4.3 Joint Model with Bayesian Nonparametric Mixtures

4.3.1 Model formulation

In this work, we introduce a novel framework for the joint analysis of survival and gap times, expanding upon existing methodologies by employing a Bayesian

nonparametric approach. This approach allows for the modeling of general shapes of density and survival functions within and across the covariate space.

We employ DDP mixture models to articulate the conditional density of survival and gap times as follows:

$$\begin{aligned} f_S(t \mid G_{\boldsymbol{x}}, \epsilon) &= \int k(t \mid \theta(\boldsymbol{x}), \epsilon) dG_{\boldsymbol{x}},\ G_{\boldsymbol{x}} \sim DDP(\Phi, G_{0\boldsymbol{x}}), \\ f_R(w \mid H_{\boldsymbol{x}}, \xi) &= \int k(w \mid \lambda(\boldsymbol{x}), \xi) dH_{\boldsymbol{x}},\ H_{\boldsymbol{x}} \sim DDP(\Lambda, H_{0\boldsymbol{x}}), \end{aligned}$$

where $k(\cdot)$ represents a generic mixture kernel density function, $G_{\boldsymbol{x}}$ and $H_{\boldsymbol{x}}$ are mixing distributions, Φ and Λ are parameters associated with mixture weights of $G_{\boldsymbol{x}}$ and $H_{\boldsymbol{x}}$. And $G_{0\boldsymbol{x}}$ and $H_{0\boldsymbol{x}}$ are dependent baseline distributions.

As mentioned, the DDP defines the set of random distributions that are marginally DP-distributed measures. $G_{\boldsymbol{x}}$ has a stick-breaking representation (Sethuraman 1994) as: $G_{\boldsymbol{x}}(\cdot) = \sum_{l=1}^{\infty} p_l(\boldsymbol{x}) \delta_{\theta_l(\boldsymbol{x})}(\cdot)$. We start to build the model from extending atoms by imposing a linear model with covariates on point masses, $\theta_l(\boldsymbol{x}) = \beta_{l0} + \boldsymbol{\beta}_l' \boldsymbol{x}$, as in De Iorio, Johnson, Müller & Rosner (2009). Developing covariate-dependent mixture weights that maintain marginal Beta distributions poses a significant challenge. Our modeling approach simplifies this by focusing on a binary component in $\boldsymbol{x}$, typically representing group assignments in control-treatment studies. The groups are denoted by the variable z, with *ctr* indicating control group and *trt* for the treatment group. This leads to dependent mixture weights for each group, formulated as $p_l(\boldsymbol{x}) \equiv p_{zl} = V_{zl} \prod_{r=1}^{l-1} (1 - V_{zr})$. Here, we incorporate a latent bivariate Beta distribution for the generation of paired mixture weights, a methodological choice that preserves

the marginal distribution as Beta.

We construct the dependent weights of mixing distribution through a bivariate Beta distribution introduced by Olkin & Trikalinos (2015), which is built upon a Dirichlet distribution and can provide correlation in $(-1, 1)$. Specifically, $(\Delta_{11}, \Delta_{10}, \Delta_{01}, \Delta_{00}) \sim \text{Dir}(\alpha_{11}, \alpha_{10}, \alpha_{01}, \alpha_{00})$, then define $V_{trt} = \Delta_{11} + \Delta_{10}$ and $V_{ctr} = \Delta_{11} + \Delta_{01}$. Thus, the bivariate distribution, $(V_{ctr}, V_{trt}) \sim \text{Bi-Be}(\Delta_{00}, \Delta_{01}, \Delta_{10}, \Delta_{11})$, results two marginal Beta distributions, $V_{trt} \sim \text{Be}(\alpha_{11} + \alpha_{10}, \alpha_{01} + \alpha_{00})$ and $V_{ctr} \sim \text{Be}(\alpha_{11} + \alpha_{01}, \alpha_{10} + \alpha_{00})$. The desired marginals are obtained by setting $\alpha_{11} = 1 - \alpha_0$, $\alpha_{10} = \alpha_{01} = \alpha_0$, and $\alpha_{00} = \alpha - \alpha_0$, such that the random mixing distributions have the same marginal DP prior. As outlined by Olkin & Trikalinos (2015), this construction for the bivariate beta distribution can be extended to $q > 2$ dimensions. However, such extensions necessitate a complex setup involving $2^q - 1$ latent variables.

The level of dependence between survival times and recurrent events may vary across treatment groups. To address this, we assume group-dependent random effects and choose to use a bivariate lognormal distribution, formulated as $(\epsilon_{iz}, \xi_{iz})' \mid \Sigma_{ez} \overset{i.i.d.}{\sim} \text{LN}_2((0,0)', \Sigma_{ez})$, where the median of the distribution is fixed at $(0,0)'$ for avoiding identifiability issue. The structure of Σ_{ez}, with its diagonal elements quantifying the random effects' magnitude and the off-diagonal elements introducing dependency between survival and gap times, is pivotal in capturing the complexity interrelations inherent in the data.

For the kernel of our mixture models, we select the log-logistic distribution for its proficiency in accurately modeling the density and survival functions as described in

Chapter 3. Under our joint modeling framework, both survival and gap times share the same formulation, which is expressed as:

$$\begin{aligned} f(t_i \mid G_z, \epsilon_z, \boldsymbol{x}_i) &= \sum_{l=1}^{\infty} \left\{ V_{zl} \prod_{r<l} (1 - V_{zr}) \right\} f_{\mathrm{LL}}(t_i \mid \theta_l^* \exp(\boldsymbol{x}_i' \boldsymbol{\beta}_l^*)/\epsilon_{iz}, \phi_l^*) \\ f(w_{ij} \mid H_x, \xi_z, \boldsymbol{x}_i) &= \sum_{l=1}^{\infty} \left\{ \pi_{zl} \prod_{r<l} (1 - \pi_{zr}) \right\} f_{\mathrm{LL}}(w_{ij} \mid \lambda_l^* \exp(\boldsymbol{x}_i' \boldsymbol{\gamma}_l^*)/\xi_{iz}, \eta_l^*), \end{aligned}$$

for $i = 1, \ldots, n$, and for $j = 1, \ldots, N_i$, where $f_{\mathrm{LL}}(a, b)$ denotes the log-logistic density with scale parameter a and shape parameter b. Atoms $(\theta_l^*, \boldsymbol{\beta}_l^*, \phi_l^*)$ and $(\lambda_l^*, \boldsymbol{\gamma}_l^*, \eta_l^*)$ are i.i.d. arising from G_0 and H_0, mixture weights are constructed through bivariate latent variables $(V_{ctr,l}, V_{trt,l}) \overset{i.i.d.}{\sim}$ Bi-Be$(1 - \alpha_0, \alpha_0, \alpha_0, \alpha - \alpha_0)$ and $(\pi_{ctr,l}, \pi_{trt,l}) \overset{i.i.d.}{\sim}$ Bi-Be$(1 - \zeta_0, \zeta_0, \zeta_0, \zeta - \zeta_0)$. This model formulation incorporates dependent atoms as well as dependent weights to allow extra dependency through a log-linear component to the kernel distribution.

4.3.2 Prior specification

In the context of DDP models, the selection of the baseline distribution is pivotal, balancing computational efficiency and the model's adaptability. For G_0, we opt for a composite baseline distribution encompassing $\mathrm{N}(\log(\theta^*) \mid \mu_\theta, \sigma_\theta^2)$, $\mathrm{N}(\boldsymbol{\beta}^* \mid \mu_\beta, \Sigma_\beta)$, and $\mathrm{Ga}(\phi^{*2} \mid a_\phi, b_\phi)$. A similar approach is adopted for H_0, utilizing $\mathrm{N}(\log(\lambda^*) \mid \mu_\lambda, \sigma_\lambda^2)$, $\mathrm{N}(\boldsymbol{\gamma}^* \mid \mu_\gamma, \Sigma_\gamma)$, and $\mathrm{Ga}(\eta^{*2} \mid a_\eta, b_\eta)$. This selection offers extensive support and incorporates hyperpriors for the parameters, facilitating swift parameter updates via the Gibbs sampler. In practice, parameters such as σ_θ, Σ_β, a_ϕ, σ_λ, Σ_γ, and a_η can be fixed, with priors assigned as $\mu_\theta \sim \mathrm{N}(s_\theta, S_\theta)$, $\mu_\beta \sim \mathrm{N}(s_\beta, S_\beta)$, $b_\phi \sim$ inv-$\mathrm{Ga}(r_\phi, R_\phi)$, $\mu_\lambda \sim \mathrm{N}(s_\lambda, S_\lambda)$,

$\mu_\gamma \sim \mathrm{N}(s_\gamma, S_\gamma)$, and $b_\eta \sim \text{inv-Ga}(r_\eta, R_\eta)$. This setup results in a marginal prior base measure of $\log(\theta^*)$ following a normal distribution, $\mathrm{N}(s_\theta, S_\theta + \sigma_\theta^2)$, and the marginal prior base measure of ϕ^{*2} is proportional to $(\phi^{*2})^{a_\phi - 1}(\phi^{*2} + R_\phi)^{-(a_\phi + r_\phi)}$ with mean value equal to $a_\phi R_\phi/(r_\phi - 1)$. When $R_\phi = 1$, the distribution is a Beta prime. For practical implementation, the hyperparameter r_ϕ can be set as the smallest integer value (3) that yields an inverse-gamma distribution with finite mean and finite but largest variance. The selection of values for a_ϕ, r_ϕ and R_ϕ allows incorporation of prior information about the density function's shape and dispersion. Specifically, smaller values of ϕ^{*2} correspond to more widely dispersed mixture components. This approach is similarly applied to μ_λ, μ_γ, and b_η.

For the bivariate Beta prior applied to the paired latent stick-breaking weights, $(V_{ctr,l}, V_{trt,l})$, which follows Bi-Be$(1 - \alpha_0, \alpha_0, \alpha_0, \alpha - \alpha_0)$ with $0 < \alpha_0 < 1$ and $\alpha > \alpha_0$, we adopt a joint prior distribution for α_0 and α as $p(\alpha_0, \alpha) = p(\alpha_0)p(\alpha \mid \alpha_0)$. Here, α_0 is subject to a Beta prior $\mathrm{Be}(a_0, b_0)$, and conditional on α_0, α has a Pareto prior $\mathrm{Pa}(c_0, \alpha_0)$, characterized by shape parameter c_0 and scale parameter α_0. For the gap times baseline distribution H_0, a similar structure is applied: $\zeta_0 \sim \mathrm{Be}(a_1, b_1)$ and $\zeta \mid \zeta_0 \sim \mathrm{Pa}(c_1, \zeta_0)$. We then assign independent inverse-Wishart priors for Σ_{ez}, denoted as $\Sigma_{ez} \overset{i.i.d.}{\sim} \text{inv-Wish}(c_e, C_e)$ for each group. These priors are characterized by a mean of $C_e/(c_e - 3)$. Given a set of hyperparameters, we visualize prior estimates of marginal and conditional density and survival functions to validate values of hyperparameters.

4.3.3 Posterior computation

We obtain samples from the posterior distribution of the DDP mixture model using the blocked Gibbs sampler (Ishwaran & James 2001). Our Markov chain Monte Carlo (MCMC) posterior simulation method utilizes a truncated approximation to the mixing distribution, denoted by G_{zB} and H_{zB} for survival and gap times, with truncation levels B_G and B_H, respectively. These approximations are defined as:

$$\begin{aligned} G_{zB} &\approx \sum_{l=1}^{B_G} p_{zl}\delta_{(\theta_l^*,\boldsymbol{\beta}_l^*,\phi_l^*)}, \quad (\theta_l^*,\boldsymbol{\beta}_l^*,\phi_l^*) \overset{i.i.d.}{\sim} G_0, \\ H_{zB} &\approx \sum_{l=1}^{B_H} \omega_{zl}\delta_{(\lambda_l^*,\boldsymbol{\gamma}_l^*,\eta_l^*)}, \quad (\lambda_l^*,\boldsymbol{\gamma}_l^*,\eta_l^*) \overset{i.i.d.}{\sim} H_0, \end{aligned}$$

with the construction of mixture weights p_{zl} and ω_{zl} designed to reflect the dependency across treatment groups. By introducing configuration variables $L_{iz} \in \{1,\ldots,B_G\}$ for survival times, $U_{ijz} \in \{1,\ldots,B_H\}$ for gap times, where $i = 1,\ldots,n$, $j = 1,\ldots,N_i$, $z \in \{ctr, trt\}$, the joint model can be expressed in a hierarchical representation,

$$\begin{aligned} T_i \mid L_{iz},\boldsymbol{\theta}^*,\boldsymbol{\beta}^*,\boldsymbol{\phi}^*,\epsilon_{iz},\boldsymbol{x}_i &\overset{ind.}{\sim} \mathrm{LL}(\theta_{L_{iz}}^* \exp(\boldsymbol{\beta}_{L_{iz}}^*\boldsymbol{x}_i)/\epsilon_{iz}, \phi_{L_{iz}}^*), \\ L_{iz} \mid p_{lz} &\overset{ind.}{\sim} p_{lz}\delta_l(L_{iz}), \text{ for } l = 1,\ldots,B_G, \\ W_{ij} \mid U_{ijz},\boldsymbol{\lambda}^*,\boldsymbol{\gamma}^*,\boldsymbol{\eta}^*,\xi_{iz},\boldsymbol{x}_i &\overset{i.i.d.}{\sim} \mathrm{LL}(\lambda_{U_{ijz}}^* \exp(\boldsymbol{\gamma}_{U_{ijz}}^*\boldsymbol{x}_i)/\xi_{iz}, \eta_{U_{ijz}}^*), \\ U_{ijz} \mid \omega_{lz} &\overset{ind.}{\sim} \omega_{lz}\delta_l(U_{ijz}), \text{ for } l = 1,\ldots,B_H, \end{aligned}$$

where survival and gap times, and covariates are rearranged to corresponding groups. Thus, the likelihood function of augmented model can be written as

$$\begin{aligned} L &= \prod_{i=1}^{n} \left\{ \prod_{j=1}^{N_i} f_{\mathrm{LL}}(w_{ij} \mid \lambda^*_{U_{ijz}} \exp(\boldsymbol{\gamma}^*_{U_{ijz}} \boldsymbol{x}_i)/\xi_{iz}, \eta^*_{U_{ijz}}) \right\} \\ &\quad \times S_{\mathrm{LL}}(\tau_i - y_{i,N_i} \mid \lambda^*_{U_{i,N_i,z}} \exp(\boldsymbol{\gamma}^*_{U_{i,N_i,z}} \boldsymbol{x}_i)/\xi_{iz}, \eta^*_{U_{i,N_i,z}}) \\ &\quad \times \left\{ f_{\mathrm{LL}}(\tau_i \mid \theta^*_{L_{iz}} \exp(\boldsymbol{\beta}^*_{L_{iz}} \boldsymbol{x}_i)/\epsilon_{iz}, \phi^*_{L_{iz}} \right\}^{\nu_i} \\ &\quad \times \left\{ S_{\mathrm{LL}}(\tau_i \mid \theta^*_{L_{iz}} \exp(\boldsymbol{\beta}^*_{L_{iz}} \boldsymbol{x}_i)/\epsilon_{iz}, \phi^*_{L_{iz}}) \right\}^{1-\nu_i}. \end{aligned}$$

The detailed MCMC algorithm is elaborated in the Supplementary Materials.

Given the conditional independence of survival and gap times, marginal density, survival functions are readily derived by marginalizing over the random effects distribution. This process is facilitated at any time t and for covariates $\boldsymbol{x}$, through the following expressions:

$$\begin{aligned} f_S(t \mid \Theta_S, \boldsymbol{x}) &= \int f_S(t \mid G_B, \epsilon_z, \boldsymbol{x}) f_{\mathrm{LN}}(\epsilon_z \mid 0, \Sigma_{ez,1,1}) d\epsilon_z, \\ S_S(t \mid \Theta_S, \boldsymbol{x}) &= \int S_S(t \mid G_B, \epsilon_z, \boldsymbol{x}) f_{\mathrm{LN}}(\epsilon_z \mid 0, \Sigma_{ez,1,1}) d\epsilon_z, \\ f_R(w \mid \Theta_R, \boldsymbol{x}) &= \int f_R(t \mid H_B, \xi_z, \boldsymbol{x}) f_{\mathrm{LN}}(\xi_z \mid 0, \Sigma_{ez,2,2}) d\xi_z, \\ S_R(w \mid \Theta_R, \boldsymbol{x}) &= \int S_R(t \mid H_B, \xi_z, \boldsymbol{x}) f_{\mathrm{LN}}(\xi_z \mid 0, \Sigma_{ez,2,2}) d\xi_z, \end{aligned}$$

where Θ_S and Θ_R represent the parameters corresponding to survival times and recurrent events. Although direct integration does not yield closed-form expressions, we can

obtain posterior estimates through Monte Carlo approximations:

$$
\begin{aligned}
f_S(t \mid \Theta_S^{(b)}, \boldsymbol{x}) &\approx \frac{1}{n_r} \sum_{i_r=1}^{n_r} \sum_{l=1}^{B_G} p_{zl}^{(b)} f_{\mathrm{LL}}(t \mid \theta_l^{(b)} \exp(\boldsymbol{x}'\boldsymbol{\beta}_l^{(b)})/\epsilon_z^{i_r,(b)}, \phi_l^{(b)}), \\
S_S(t \mid \Theta_S^{(b)}, \boldsymbol{x}) &\approx \frac{1}{n_r} \sum_{i_r=1}^{n_r} \sum_{l=1}^{B_G} p_{zl}^{(b)} S_{\mathrm{LL}}(t \mid \theta_l^{(b)} \exp(\boldsymbol{x}'\boldsymbol{\beta}_l^{(b)})/\epsilon_z^{i_r,(b)}, \phi_l^{(b)}), \\
f_R(w \mid \Theta_R^{(b)}, \boldsymbol{x}) &\approx \frac{1}{n_r} \sum_{i_r=1}^{n_r} \sum_{l=1}^{B_H} \omega_{zl}^{(b)} f_{\mathrm{LL}}(w \mid \lambda_l^{(b)} \exp(\boldsymbol{x}'\boldsymbol{\gamma}_l^{(b)})/\xi_z^{i_r,(b)}, \eta_l^{(b)}), \\
S_R(w \mid \Theta_R^{(b)}, \boldsymbol{z}) &\approx \frac{1}{n_r} \sum_{i_r=1}^{n_r} \sum_{l=1}^{B_H} \omega_{zl}^{(b)} S_{\mathrm{LL}}(t \mid \lambda_l^{(b)} \exp(\boldsymbol{x}'\boldsymbol{\gamma}_l^{(b)})/\xi_z^{i_r,(b)}, \eta_l^{(b)}),
\end{aligned}
$$

where $(\epsilon_z^{i_r,(b)}, \xi_z^{i_r,(b)})' \overset{ind.}{\sim} \mathrm{LL}((0,0)', \Sigma_{ez}^{(b)})$, for $i_r = 1, \ldots, n_r$, and superscript (b) represents b^{th} posterior sample from the MCMC. This approach enables the acquisition of point-wise functional estimates over a grid of time.

In exploring the interplay between recurrent events and survival times, our focus lies on the conditional survival probability, given no recurrent event occurrence up to time t_0, $Pr(T \geq t \mid T > t_0, N = 0, \boldsymbol{x}, \Theta_S, \Theta_R)$, which implies that the first gap time exceeds t_0. Under our joint model, the conditional survival probability is formulated as:

$$
\Pr(T \geq t \mid T > t_0, W_1 > t_0, \Theta_S, \Theta_R, \boldsymbol{x}) = \frac{\mathrm{E}_e\left\{S_S(t \mid \epsilon_z, \Theta_S, \boldsymbol{x}) S_R(t_0 \mid \xi_z, \Theta_R, \boldsymbol{x})\right\}}{\mathrm{E}_e\left\{S_S(t_0 \mid \epsilon_z, \Theta_S, \boldsymbol{x}) S_S(t_0 \mid \xi_z, \Theta_R, \boldsymbol{x})\right\}},
$$

where the expectation is with respect to the bivariate random effects model. Its posterior estimates can be approximated by:

$$
\frac{\frac{1}{n_r}\sum_{i_r=1}^{n_r}\left\{\sum_{l=1}^{B_G} \hat{p}_{zl}^{(b)} S_{\mathrm{LL}}(t \mid \hat{\theta}_l^{(b)} \exp(\boldsymbol{x}'\hat{\boldsymbol{\beta}}_l^{(b)})/\epsilon_z^{i_r,(b)}, \hat{\phi}_l^{(b)}) \sum_{l=1}^{B_H} \hat{\omega}_{zl}^{(b)} S_{\mathrm{LL}}(t_0 \mid \hat{\lambda}_l^{(b)} \exp(\boldsymbol{x}'\hat{\boldsymbol{\gamma}}_l^{(b)})/\xi_z^{i_r,(b)}, \hat{\eta}_l^{(b)})\right\}}{\frac{1}{n_r}\sum_{i_r=1}^{n_r}\left\{\sum_{l=1}^{B_G} \hat{p}_{zl}^{(b)} S_{\mathrm{LL}}(t_0 \mid \hat{\theta}_l^{(b)} \exp(\boldsymbol{x}'\hat{\boldsymbol{\beta}}_l^{(b)})/\epsilon_o^{i_r,(b)}, \hat{\phi}_l^{(b)}) \sum_{l=1}^{B_H} \hat{\omega}_{zl}^{(b)} S_{\mathrm{LL}}(t_0 \mid \hat{\lambda}_l^{(b)} \exp(\boldsymbol{x}'\hat{\boldsymbol{\gamma}}_l^{(b)})/\xi_z^{i_r,(b)}, \hat{\eta}_l^{(b)})\right\}}.
$$

The posterior functional estimates can provide valuable insights into subjects' conditions under various scenarios. The posterior marginal density estimates can reveal the

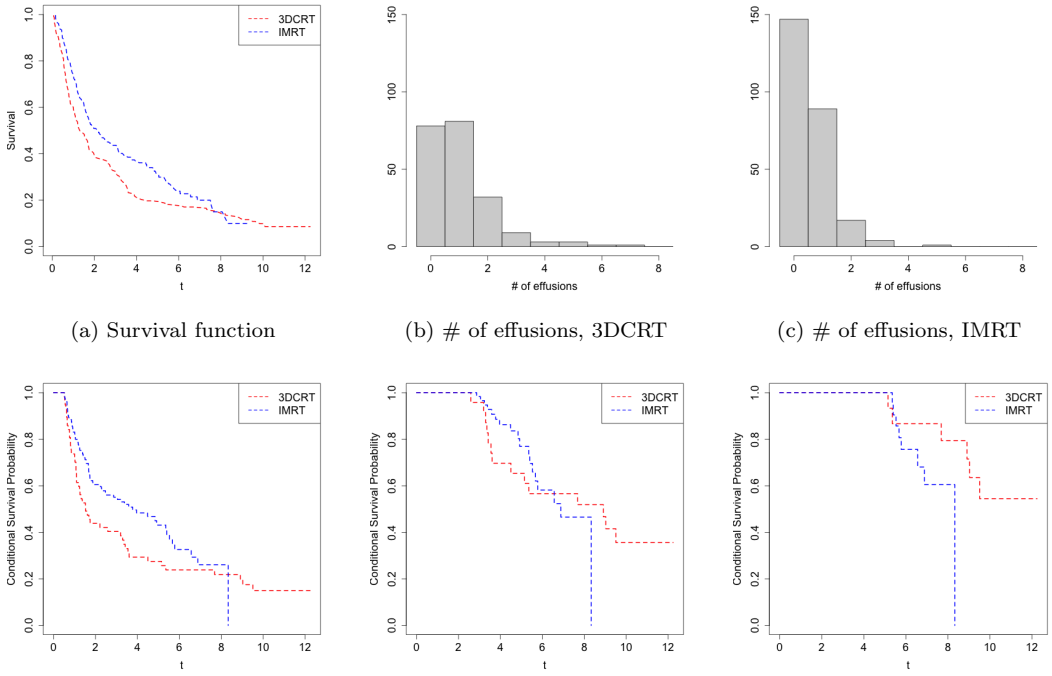
distribution's shape, highlighting features such as multi-modality, skewness, and heavy tails. Additionally, the posterior marginal and conditional survival probabilities can be utilized to compare the effectiveness of control versus treatment groups.

The fully specified stochastic Bayesian joint models for survival times and recurrent events can be used to predict future outcomes. For a hypothetical patient given a covariate vector $\boldsymbol{x}_0$, which contains the group assignment z_0, we can jointly predict a survival time t' and realization of recurrent events $y'_1, \ldots, y'_{N'}$ upto t'. For b^{th} posterior sample, we can generate predictive survival and gap times as follows:

- Draw $(\epsilon_{z_0}^{(b)}, \xi_{z_0}^{(b)})$ from $\text{LN}((0,0)', \Sigma_{ez_0}^{(b)})$.
- Conditional on $\epsilon_{z_0}^{(b)}$, generate a predictive survival time t' from the truncated mixture survival model, $\sum_{l=1}^{B_G} p_{z_0 l}^{(b)} \text{LL}(t \mid \theta_l^{(b)} \exp(\boldsymbol{x}_0 \boldsymbol{\beta}_l^{(b)})/\epsilon_{z_0}^{(b)}, \phi_l^{(b)})$.
- Conditional on $\xi_{z_0}^{(b)}$, take $N'+1$ independent draws $\{w'_j : j = 1, \ldots, N'+1\}$ from its posterior distribution, $\sum_{l=1}^{B_H} \omega_{z_0 l}^{(b)} \text{LL}(w \mid \lambda_l^{(b)} \exp(\boldsymbol{x}_0 \boldsymbol{\gamma}_l^{(b)})/\xi_{z_0}^{(b)}, \hat{\eta}_l^{(b)})$, such that $\sum_{j=1}^{N'} w'_j < t' < \sum_{j=1}^{N'+1} w'_j$.

We demonstrate the utility of our model through out-of-sample predictions. The model performance is evaluated by comparing observed and predicted survival times and number of recurrent events. Additionally, we calculate conditional predictive probabilities for prediction matching observation in leave-one-out cross-validation. Finally, we compare the predictive accuracy of several other joint models with our proposed model.

4.4 Esophageal cancer data example



(a) Survival function (b) # of effusions, 3DCRT (c) # of effusions, IMRT

(d) $t_0 = 0.5$ (e) $t_0 = 2.5$ (f) $t_0 = 5$

Figure 4.1: EC Data. Panel (a) displays Kaplan-Meier survival curves for patients. Panels (b)-(c) illustrate histograms of effusions counts for each treatment group. And panels (d)-(f) present empirical estimates of the conditional survival probabilities given no occurrence of effusion upto t_0, $\Pr(T > t \mid T > t_0, N(t_0) = 0)$, , where t_0 =0.5, 2.5, and 5.0. We depict 3DCRT group in red and IMRT group in blue.

We applied the proposed methodology to a dataset of esophageal cancer (EC), consisting 466 patients who underwent either three-dimensional conformal radiotherapy (3DCRT) or intensity-modulated radiotherapy (IMRT). The dataset records from the initiation of treatment to death or administrative censoring, with 29 censored observa-

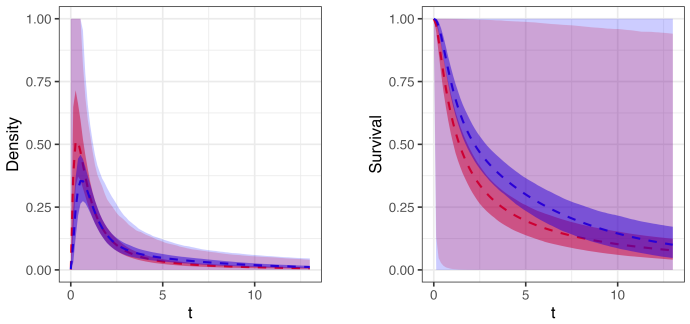
tions in the 3DCRT group and 84 in the IMRT group. Some patients also experienced recurrent effusions around heart or lungs. In addition to treatment type, the dataset includes prognostic covariates such as age, body mass index (BMI), performance status (KPS score), tumor histology, and cancer stage. Continuous variables, age and BMI, were standardized to have a mean of 0 and variance of 1. Binary covariates were coded as follows: KPS score = 1 for good, histology = 1 for adenocarcinoma, and cancer stage = 1 for advanced stages (3-4). Treatment types were coded with 3DCRT as 0 (control) and IMRT as 1 (treatment), since the IMRT is a more modern approach. This dataset has been studied by He et al. (2016) utilizing a semi-competing risks model to investigate time-to-effusion and survival post-effusion. Subsequent enhancements by Chapple et al. (2017) introduced Bayesian variable selection, and Lee et al. (2019) extended the analysis to accommodate all effusion events per patient via a semi-parametric Bayesian joint model with a Poisson process assumption on recurrent events. The previous studies revealed that IMRT decreased modalities relative to 3DCRT. However, the models for survival times in these studies are based on either PH or AFT assumptions.

Kaplan-Meier survival curves depicted in Figure 4.1(a) suggest that the IMRT group generally exhibits better survival than the 3DCRT group throughout the observation period. Intriguingly, the conditional survival probabilities illustrated in subfigures (d) and (e) reveal a dynamic interplay between the two treatments. With an increase in the reference time t_0, the point at which conditional survival probabilities for the two groups intersect shifts to earlier times as t_0 progresses from six months to five years. Additionally, Panels (b) and (c) highlight the infrequent occurrence of effusions, peaking

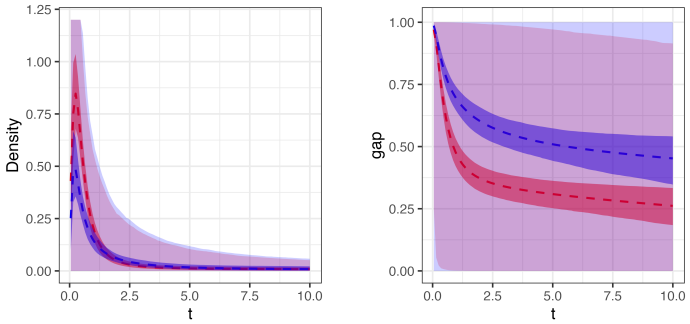
at seven incidents, with a significant proportion of patients experiencing no effusions at all.

4.4.1 Analysis with only treatment variable

We implement the joint model on the dataset, incorporating solely the treatment assignment covariate. For the prior hyperparameters, we set the values as follows: $a_0 = b_0 = a_1 = b_1 = 2$, $c_0 = c_1 = 4$, indicating a balanced initial belief in the distribution of the treatment effects. The scale parameters σ_θ^2 and σ_λ^2 are assigned a value of 2, reflecting moderate variability in the log-transformed scale parameters for survival and gap times, respectively. The mean parameters for the log-transformed scale, s_θ and s_λ, are set to 0.45 and 0, with corresponding variances $S_\theta = S_\lambda = 1$, suggesting a central tendency with a reasonable spread. For parameters β^* and γ^*, we assume a mean of 0 ($s_\beta = s_\gamma = 0$) with variances $\Sigma_\beta = \Sigma_\gamma = 2$ and $S_\beta = S_\gamma = 2$, capturing the uncertainty in the treatment effect size. The shape parameters for the log-logistic distribution, a_ϕ and a_η, are set to 3, with the inverse-Gamma parameters for the scale, $r_\phi = r_\eta = 6$ and $R_\phi = R_\eta = 30$, chosen to reflect a belief in the variability of the shape parameters. Lastly, for the covariance matrix of the random effects, Σ_{eo}, we employ an inverse-Wishart prior with $c_e = 18$ and C_e set to a diagonal matrix with elements $(4, 4)$, resulting prior predictive random effects remained within a plausible range. This hyperparameter configuration in the prior models allows for substantial uncertainty across different functionals. We present their pointwise interval estimates based on the prior models in Figures 4.2 and 4.3.



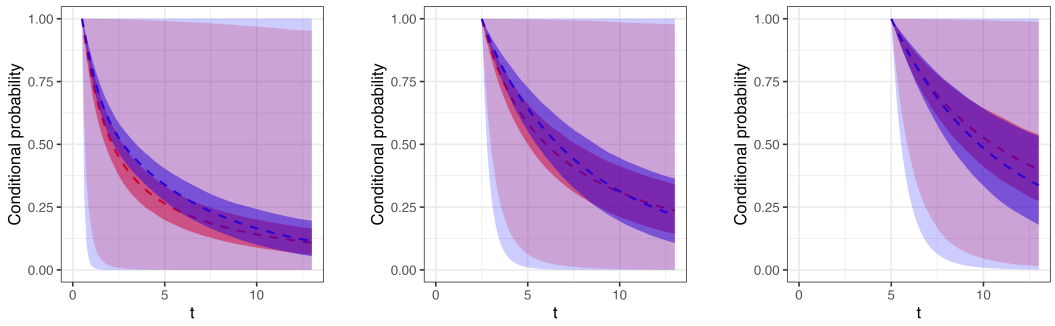
(a) $f(t)$ (b) $S(t)$



(c) $f(w)$ (d) $S(w)$

Figure 4.2: EC Data with Treatment Assignments Only. This figure presents posterior mean estimates (dashed lines) and 95% pointwise credible intervals for posterior (dark shaded regions) and prior (light shaded regions) estimates. Panels (a) and (b) display the survival times density and survival function, respectively, while panels (c) and (d) show the gap times density and survival function. All estimates are derived after marginalizing over the posterior and prior random effects models. The 3DCRT group is represented in red and the IMRT group in blue.

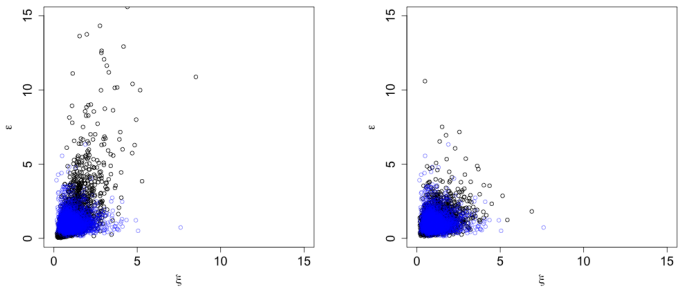
We examined convergence and mixing of the MCMC algorithms using standard diagnostic techniques. We ran the MCMC chain for 100,000 iterations, then discarded the first 25% posterior samples and keep every 38th iteration for posterior inference.



(1) $t_0 = 0.5$ (2) $t_0 = 2.5$ (3) $t_0 = 5$

Figure 4.3: EC Data with Treatment Assignments Only. This figure presents posterior mean estimates (dashed lines) and 95% pointwise credible intervals for posterior (dark shaded regions) and prior (light shaded regions) estimates. Panels (a)-(c) depict conditional survival probability given no occurrence of effusion up to t_0 years, progressively displaying time points at $t_0 = 0.5$, 2.5, and 5 years from left to right. All estimates are derived after marginalizing over the posterior and prior random effects models. The 3DCRT group is represented in red and the IMRT group in blue.

The posterior estimates for survival and gap times density, along with survival functions presented in Figure 4.2, suggest that IMRT is associated with higher survival probabilities and less frequent effusions. This includes longer gap time between effusions in the lungs and heart and longer survival times. However, in Figure 4.3, the posterior mean estimates of conditional survival probabilities, given no occurrence of effusions up to t_0 years, demonstrate a gradually reversing stochastic order between the 3DCRT and IMRT groups as t_0 increases. A similar trend is observed by using Kaplan-Meier curve, as shown in Figure 4.1 (d)-(e). Notably, our posterior estimates provide smoothed point



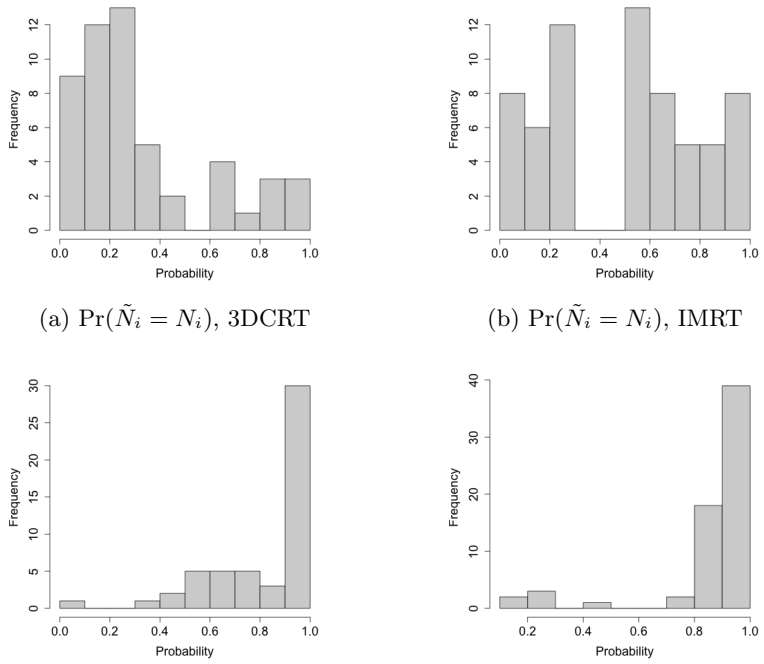
(a) 3DRT (b) IMRT

Figure 4.4: EC Data with Treatment Assignments Only. This figure presents generated random effects from posterior (black) and prior (blue) random effects models. Panel (a) focuses on the 3DCRT group, while panel (b) presents for the IMRT group.

estimates along with quantification of uncertainty, enhancing their interpretative value.

The posterior predictive random effects for the 3DCRT and IMRT groups are illustrated in Figure 4.4. These effects show varying levels of association between survival and gap times across the groups, highlighting the importance of employing distinct random effects models for each treatment group.

Model Assessment To evaluate the model's accuracy, we conducted out-of-sample prediction by randomly selecting 25% of the dataset (117 patients) as the test set, with the remaining 349 patients used for training. The model focused solely on the treatment covariate, employing the same hyperparameters outlined in the previous part. The model was used to predict recurrent effusion events up to the last observed time (survival or censored time) for each patient in the test set, as outlined in Section 4.3.3. We evaluated the accuracy of effusion predictions by comparing the predicted number of



(a) $\Pr(\tilde{N}_i = N_i)$, 3DCRT

(b) $\Pr(\tilde{N}_i = N_i)$, IMRT

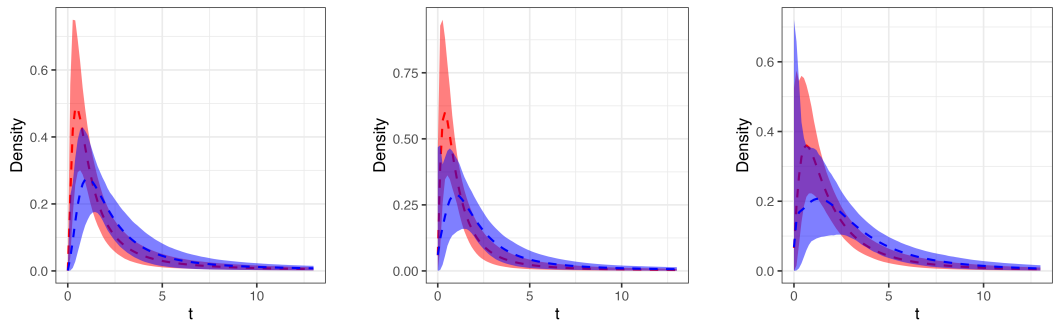
(c) $\Pr(N_i - 1 \leq \tilde{N}_i \leq N_i + 1)$, 3DCRT

(d) $\Pr(N_i - 1 \leq \tilde{N}_i \leq N_i + 1)$, IMRT

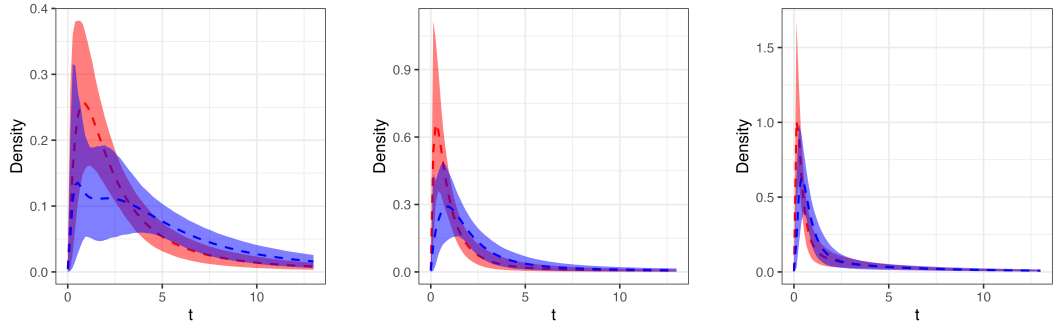
Figure 4.5: Out-of-Sample Prediction Accuracy for the EC Data with Treatment Assignments only. The top row presents histograms of the probabilities for exact effusion count matches ($\Pr(\tilde{N}_i = N_i)$), while the bottom row depicts the probabilities of predicted effusion counts falling within one of the actual observed counts ($\Pr(N_i - 1 \leq \tilde{N}_i \leq N_i + 1)$).

effusions, $\tilde{N}_i$, against the actual observed counts. Our metrics included the probability of exact matches, $\Pr(\tilde{N}_i = N_i \mid x_i)$, and the probability of predictions within one count of actual observations, $\Pr(N_i - 1 \leq \tilde{N}_i \leq N_i + 1 \mid x_i)$. Figure 4.5 showcases the model's proficient capability in forecasting effusion events, where the predicted counts are generally within a one-event margin of the actual observations.

4.4.2 Analysis with additional covariates



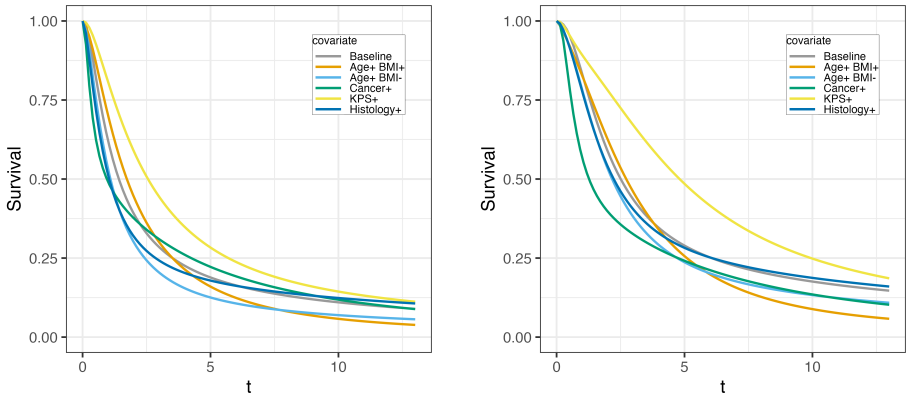
(a) Baseline (b) Age+, BMI- (c) Age+, BMI+



(d) KPS+ (e) Histology+ (f) Cancer+

Figure 4.6: EC Data with All Covariates. This figure presents posterior mean estimates (dashed lines) and 95% pointwise credible intervals for the density estimates after marginalizing over the posterior random effect model given different values of $\boldsymbol{x}_0^*$. The configurations in panels (a)-(f) are as follows: $(0,0,0,0,0)$ serving as the baseline, $(1,-1,0,0,0)$ representing older age and lower BMI, $(1,1,0,0,0)$ for older age and higher BMI, $(0,0,1,0,0)$ indicating good performance status, $(0,0,0,1,0)$ for a history of adenocarcinoma, and $(0,0,0,0,1)$ reflecting cancer stages 3-4. The 3DCRT group is represented in red and the IMRT group in blue.

The analysis was extended to include previously discussed prognostic covari-

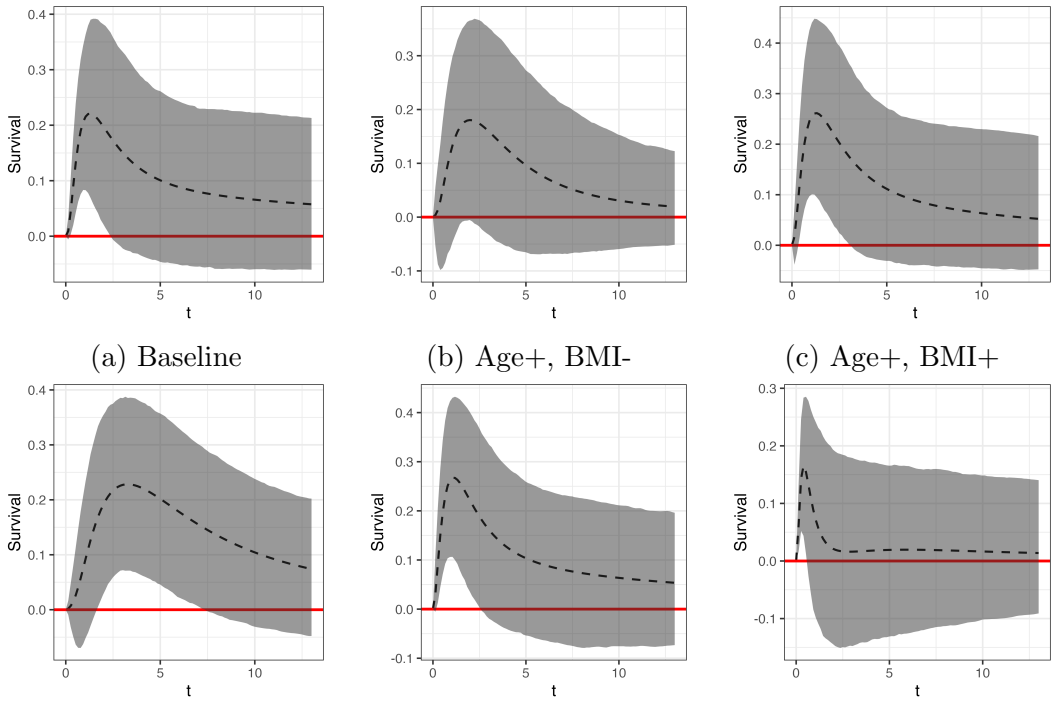


(a) 3DCRT (b) IMRT

Figure 4.7: EC Data with All Covariates. This figure displays posterior mean estimates for survival functions with the left panel showcasing the 3DCRT group and the right panel displaying the IMRT group. Each color within the panels represents a specific fixed covariate.

ates: age, BMI, performance status, tumor histology, and cancer stage, all alongside treatment types. These covariates, standardized and coded appropriately as discussed, were incorporated into the model to evaluate their impact on both survival times and effusion events.

The same hyperparameters are applied here for the following parameters as detailed in the previous section, a_0, b_0, a_1, b_1, c_0, c_1, σ_θ^2, σ_λ^2, s_θ, s_λ, S_θ, S_λ, a_ϕ, a_η, r_ϕ, r_η, R_ϕ, R_η, c_e and C_e. The adjustments is is needed for accommodating the increased dimensionality of the covariates. We set $\Sigma_\beta = 2\mathrm{I}_6$, $\Sigma_\gamma = 2\mathrm{I}_6$, where I_6 denotes an identity matrix. This adjustment ensures that the covariates are appropriately scaled. Additionally, the mean vectors for these coefficients are set to zero, $s_\beta = \mathbf{0}_6$ and $s_\gamma = \mathbf{0}_6$.



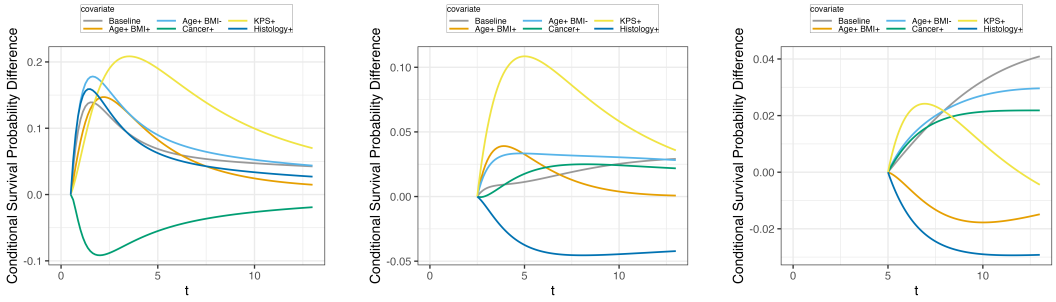
(a) Baseline (b) Age+, BMI- (c) Age+, BMI+

(d) KPS+ (e) Histology+ (f) Cancer+

Figure 4.8: EC Data with All Covariates. This figure displays posterior mean (dashed lines) and 95% pointwise credible intervals illustrating the differences in survival probabilities between IMRT and 3DCRT groups. The red solid lines serves as a reference indicating no difference between the groups.

The scale matrices for the prior distributions of these coefficients are also adjusted to I_6, ensuring that $S_\beta = 2\mathrm{I}_6$ and $S_\gamma = 2\mathrm{I}_6$, which matches the dimensionality and maintains consistency across the model parameters.

The posterior inferences for marginal density and survival functions across various covariates are displayed in Figure 4.6. We selected specific values for the covariate vector components, including age, BMI, KPS score, histology, and cancer stage. The



(1) $t_0 = 0.5$ (2) $t_0 = 2.5$ (3) $t_0 = 5$

Figure 4.9: EC Data with All Covariates. This figure displays the posterior mean estimates of the differences in conditional survival probability up to t_0 years between IMRT and 3DCRT groups. Each color within the panels represents a specific fixed covariate.

binary covariates—KPS score, histology, and cancer stage—are each set to 1. The continuous covariates—age and BMI—are standardized and fixed at 1 for high values and -1 for lower values. Across these covariates, the density of the 3DCRT group exhibits greater concentration in the early years compared to the IMRT group. Notably, subfigures (c) and (d) underscore the IMRT group's nonstandard density shapes, displaying a small bump at early times. Subfigure (f) reveals a sharp mode in the 3DCRT group's density for small values of t.

The posterior mean estimates of survival function for both the 3DCRT and IMRT groups, presented in Figure 4.7, vary across different covariates. Patients with a good performance status (KPS+) in each group show markedly higher survival probabilities, whereas those in advanced cancer stages (3-4) have the lowest survival probabilities early on. The differences in their posterior estimates of survival functions between the

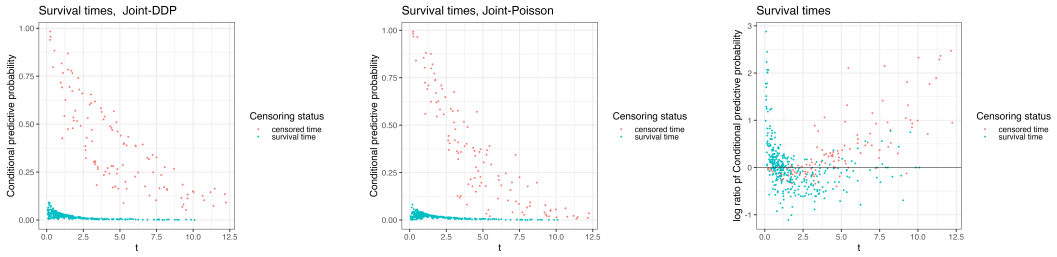
IMRT and 3DCRT groups, expressed as $S_S(t \mid \Theta_S, (trt, \boldsymbol{x}_0^*)) - S_S(t \mid \Theta_S, (ctr, \boldsymbol{x}_0^*))$ where $\boldsymbol{x}_0^*$ represents covariates excluding treatment assignments, are further detailed in Figure 4.8. This figure includes posterior point estimates and 95% pointwise credible intervals highlighting significant early discrepancies across most covariates. These differences lessen and become more uncertain as time progresses.

Figure 4.9 explores the differences in posterior estimates of conditional survival probabilities between the two treatment groups, given no occurrence of effusions up to t_0 years, formulated as $Pr(T \geq t \mid T > t_0, N = 0, \Theta_S, \Theta_R, ctr, \boldsymbol{x}_0^*) - Pr(T \geq t \mid T > t_0, N = 0, \Theta_S, \Theta_R, trt, \boldsymbol{x}_0^*)$. Early on, at $t_0 = 0.5$ years, the IMRT group shows lower conditional survival probabilities for patients in advanced cancer stages (3-4), but higher for other covariates. As t_0 increases, the conditional survival probabilities for the IMRT group consistently exceed those of the 3DCRT group. This suggests that, despite initial disadvantages in later-stage cancer, IMRT may pose higher mortality risks among older patients when effusions are absent.

	Survival times	Effusion counts	Joint
joint-DDP	-1610.9244	-493.3734	-2074.0173
joint-parametric-renewal	-1628.1498	-513.6613	-2122.3151
joint-parametric-Poisson	-1672.2402	-473.1304	-2164.2198

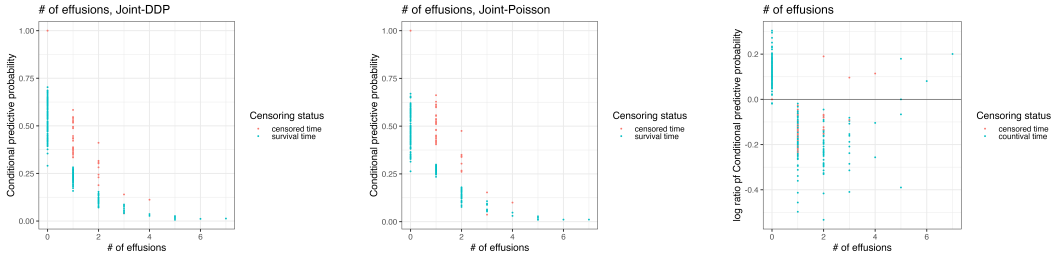
Table 4.1: Sum of log conditional predictive probabilities under different models for the EC data.

Comparison with other models We compare our joint model against two parametric alternatives, 1) a parametric version of our proposed model using log-logistic median



(a) Joint-DDP (b) Joint-Poisson (c) log ratio

Figure 4.10: EC Data with All Covariates Leave-one-out Cross-validation. This figure illustrates the conditional predictive probabilities for survival times. Panels (a) and (b) display results from our joint-DDP and a joint-parametric-Poisson models, respectively. Panel (c) presents the log-transformed ratio of conditional predictive survival probabilities between the joint-DDP and joint-parametric-Poisson models.



(a) Joint-DDP (b) Joint-Poisson (c) log ratio

Figure 4.11: EC Data with All Covariates Leave-one-out Cross-validation. This figure illustrates the conditional predictive probabilities for the number of effusions. Panels (a) and (b) display results from our joint-DDP and joint-parametric-Poisson models, respectively. Panel (c) presents the log-transformed ratio of conditional predictive probabilities of the number of effusions between the joint-DDP and joint-parametric-Poisson models.

regression for both survival and gap times, referred as joint-parametric-renewal model, and 2) a parametric model that integrates a Weibull AFT model for survival times and a non-homogeneous Poisson process with a Weibull intensity function for modeling recurrent events, referred as joint-parametric-Poisson model. Both models are conditioned on the random effects generated from a bivariate lognormal distribution.

The models are compared through leave-one-out cross-validation method. For subject i, we fit the model on the leave-one-out dataset $\mathcal{D}_{-i}$ and collect B posterior samples, then generate predictive survival times $\{\tilde{t}_b : b = 1, \ldots, B\}$ and gap times $\{\tilde{w}_{bj} : j = 1, \ldots, \tilde{N}_b, b = 1, \ldots, B\}$ jointly, where $\tilde{N}_b$ is the number of predictive effusion events using b^{th} saved posterior sample, as described in 4.3.3. The parametric-joint-renewal model employs the same prediction algorithm. For the parametric-joint-Poisson model, a slight modification is implemented: predictive survival times are generated along with the counts of effusion events instead of gap times. The prior specification for each parametric joint is given in the Supplementary material.

We calculate three types of cross-validated posterior predictive probabilities for model comparison with respect to predicting survival times, count of effusions, and both jointly. Three probabilities are defined as follows:

- For survival times, given the covariate vector $\boldsymbol{x}_i$, the cross-validated posterior predictive probability (CPP) is calculated depending on whether the observation

is censored:

$$\text{CPP}_i^S = \begin{cases} \Pr(t_i - 1/24 \leq \tilde{t} \leq t_i + 1/24 \mid \boldsymbol{x}_i, \mathcal{D}_{-i}), \text{ if } \nu_i = 1, \\ \Pr(t_i \leq \tilde{t} \mid \boldsymbol{x}_i, \mathcal{D}_{-i}), \text{ if } \nu_i = 0. \end{cases}$$

This measures the probability that the predicted survival time $\tilde{t}$ is within two weeks of the observed survival time or after the censored time.

- For recurrent events, we compute the conditional predictive probabilities of the count of effusions for subject i:

$$\text{CPP}_i^R = \begin{cases} \Pr(N_i = \tilde{N}_i \mid \boldsymbol{x}_i, \mathcal{D}_{-i}), \text{ if } \nu_i = 1, \\ \Pr(N_i \leq \tilde{N}_i \mid \boldsymbol{x}_i, \mathcal{D}_{-i}), \text{ if } \nu_i = 0. \end{cases}$$

This measures the probability that the count of predicted effusions equals the observed effusion counts if the death is observed, and the probability that the predicted count exceeds the observed count when the observation is censored.

- Joint probability for survival and recurrent events:

$$\text{CPP}_i = \begin{cases} \Pr(t_i - 1/24 \leq \tilde{t} \leq t_i + 1/24, N_i = \tilde{N}_i \mid \boldsymbol{x}_i, \mathcal{D}_{-i}), \text{ if } \nu_i = 1, \\ \Pr(t_i \leq \tilde{t}, N_i \leq \tilde{N}_i \mid \boldsymbol{x}_i, \mathcal{D}_{-i}), \text{ if } \nu_i = 0. \end{cases}$$

This assesses the joint accuracy of the model in forecasting both the timing of survival and the count of recurrent events.

A higher conditional predictive probability suggests a better model fit. Figures 4.10 and 4.11 show the values for our DDP-based joint model (joint-DDP) with a

renewal process assumption (panel (a)) and a parametric joint model based on a Poisson process assumption (joint-Poisson) (panel (b)), along with their log-scaled ratios (panel (c)). Overall, our model demonstrates superior predictive performance for patients with survival times ranging from 0 to 2 years and beyond 5 years, particularly excelling for patients with censored survival times. Regarding the predicted number of effusions, our model generally performs better for patients who did not experience any effusion events.

Similar to the log-pseudo marginal likelihood statistics (LPML), we compute the sum of the log conditional predictive probabilities as a summary score for model comparison. Table 4.1 reveals that models with a renewal process assumption generally perform better in predicting survival times but are less effective at predicting the count of effusions. Both renewal-process-based models yield lower scores for joint predictions. Among these, the proposed joint model with a Bayesian nonparametric regression approach outperforms the parametric counterpart across all three assessment types.

4.5 Summary

We have developed a comprehensive joint modeling framework for recurrent events with informative terminations. This framework utilizes a conditional renewal-process model for recurrent events and models survival and gap times using DDP mixtures, incorporating bivariate random effects. A particular focus is placed on control-treatment groups by making the DDP mixture weights dependent on treatment assignments and integrating a log-linear model with all covariates into the mixture atoms.

This approach significantly enhances the model's flexibility in representing complex functional shapes, including densities, survival functions, and conditional survival probabilities.

The model is demonstrated using a real data example with a binary covariate indicating treatment assignment. In this example, we presented marginal posterior estimates for density, survival functions, as well as conditional survival probability given no occurrence of recurrent events. The model is assessed through out-of-sample predictions, using 75% of the data for training and 25% for testing. We compared the predicted and observed number of recurrent events. Furthermore, we applied the model on the same dataset, including additional covariates, both binary and continuous. The proposed model is compared against two parametric joint models: one based on a renewal process and another on a Poisson process. Model performance is evaluated using leave-one-out prediction techniques, focusing on the conditional predictive probability. Overall, both renewal-process-based models perform more effectively in joint prediction tasks. Moreover, the proposed model, utilizing a Bayesian nonparametric regression approach, outperforms its parametric counterparts in predicting survival times, the number of recurrent events, and joint prediction tasks.

Chapter 5

Conclusion

This dissertation contributes to the field of survival analysis through the development of Bayesian nonparametric models. These model enhance flexibility, improve computational efficiency, and better handle complex data structures.

We first introduced a parsimonious Erlang mixture model, designed as a robust methodological tool for nonparametric Bayesian survival analysis. The model is built on a basis representation using Erlang densities with a common scale parameter. The mixture weights are derived from increments of a random distribution function, flexibly modeled with a Dirichlet process prior. By extending this model to accommodate binary covariates in a control-treatment setting through a common-weights dependent Dirichlet process prior, we achieved a balance between model flexibility and computational efficiency. This approach was demonstrated using both synthetic and real data examples, illustrating its practical applicability in various survival analysis scenarios.

We further developed a Bayesian nonparametric method that integrates a

Dirichlet process mixture model with a log-logistic kernel. The model's computational efficiency is enhanced by employing MCMC techniques and the Pólya-gamma data augmentation method for posterior computations. An extension of the core model is to incorporate covariates via a density regression framework. This extension allows for local adjustments in mixture weights, treating covariates as random variables. Additionally, the model integrates covariates into the log-linear model for the scale parameter of the log-logistic kernel. We illustrated the model on both synthetic and real data examples.

The final component of this dissertation is the development of a comprehensive joint modeling framework for survival times and gap times of recurrent events. This framework utilizes the DDP mixture models with log-logistic kernel for survival and gap times. The model is particularly focused on control-treatment groups, making the DDP mixture weights dependent on treatment assignments and integrating a log-linear model with covariates into the mixture atoms. We connected the Bayesian nonparametric mixtures with survival and gap times via bivariate random effects. This approach enhances the model's flexibility in representing complex functional shapes, including densities, survival functions, and conditional survival probabilities, particularly when there are no recurrent events.

The effectiveness of this model was demonstrated using a real dataset, comparing the outcomes of patients receiving two different treatments. The model's performance was validated through out-of-sample predictions, splitting the dataset into training and testing sets for modeling fitting and model assessment, respectively. It was compared against two parametric joint models based on renewal and Poisson processes,

respectively. Evaluations using leave-one-out prediction showed that the Bayesian nonparametric regression approach consistently outperformed its parametric counterparts, demonstrating superior performance in joint prediction tasks.

Overall, Bayesian nonparametric models are highly suitable for survival analysis and the joint analysis of survival times and gap times of recurrent events. These model can accommodate complex dependencies and provide robust performance across various applications. The methodologies developed here not only enhance current analytical capabilities but also pave the way for future research in survival analysis and related fields. Additionally, exploring fully nonparametric Bayesian model to jointly analyze survival times and recurrent events would be an interesting direction for future research.

Bibliography

Antoniadis, A., Grégoire, G. & Nason, G. (1999), 'Density and hazard rate estimation for right-censored data by using wavelet methods', *Journal of the Royal Statistical Society: Series B (Methodological)* **61**, 63–84.

Antoniak, C. (1974), 'Mixtures of Dirichlet processes with applications to Bayesian nonparametric problems', *The Annals of Statistics* **2**, 1152–1174.

Ayala, D., Jofré, L., Gutiérrez, L. & Mena, R. H. (2022), 'On a dirichlet process mixture representation of phase-type distributions', *Bayesian Analysis* **17**(3), 765–790.

Barcella, W., Iorio, M. & Baio, G. (2017), 'A comparative review of variable selection techniques for covariate dependent dirichlet process mixture models', *Canadian journal of statistics* **45**(3), 254–273.

Barrientos, A. F., Jara, A. & Quintana, F. A. (2015), 'Bayesian density estimation for compositional data using random Bernstein polynomials', *Journal of Statistical Planning and Inference* **166**, 116–125.

Barrientos, A. F., Jara, A. & Quintana, F. A. (2017), 'Fully nonparametric regression

for bounded data using dependent Bernstein polynomials', *Journal of the American Statistical Association* **112**, 806–825.

Barry, D. & Hartigan, J. (1993), 'A bayesian analysis for change point problems', *Journal of the American Statistical Association* **88**(421), 309–319.

Bennett, S. (1983), 'Log-Logistic Regression Models for Survival Data', *Journal of the Royal Statistical Society. Series C (Applied Statistics)* **32**(2), 165–171.

Blackwell, D. & MacQueen, J. B. (1973*a*), 'Ferguson distributions via Pólya urn schemes', *Annals of Statistics* **1**, 353–355.

Blackwell, D. & MacQueen, J. B. (1973*b*), 'Ferguson distributions via polya urn schemes', *Ann. Statist.* **1**(2), 353–355.

Boom, W., DeIorio, M. & Tallarita, M. (2022), 'Bayesian inference on the number of recurrent events: A joint model of recurrence and survival', *Statistical Methods in Medical Research* **31**(1), 139–153.

Bush, C. A. & MacEachern, S. N. (1996), 'A semiparametric Bayesian model for randomised block designs', *Biometrika* **83**, 275–285.

Butzer, P. (1954), 'On the extensions of Bernstein polynomials to the infinite interval', *Proceedings of the American Mathematical Society* **5**, 547–553.

Canale, A. & Dunson, D. (2016), 'Multiscale bernstein polynomials for densities', *Statistica Sinica* **26**, 1175–1195.

Chang, I. S., Hsiung, C. A., Wu, Y. & Yang, C. (2005), 'Bayesian survival analysis using Bernstein polynomials', *Scandinavian Journal of Statistics* **32**, 447–466.

Chapple, A. G., Vannucci, M., Thall, P. F. & Lin, S. (2017), 'Bayesian Variable Selection for a Semi-competing Risks Model with Three Hazard Functions', *Computational Statistics and Data Analysis* **112**, 170–185.

Collett, D. (2003), *Modelling survival data in medical research*, second edition edn, Chapman & Hall.

Cook, R. J. & Lawless, J. F. (2007), *The statistical analysis of recurrent events*, Statistics for biology and health, Springer, New York.

Cox, D. R. & Oakes, D. (1984), *Analysis of survival data*, Chapman and Hall.

De Iorio, M., Johnson, W. O., Müller, P. & L., R. G. (2009), 'Bayesian nonparametric nonproportional hazards survival modeling', *Biometrics* **65**(3), 762–771.

De Iorio, M., Johnson, W. O., Müller, P. & Rosner, G. L. (2009), 'Bayesian nonparametric nonproportional hazards survival modeling', *Biometrics* **65**, 762–771.

De Iorio, M., Müller, P., Rosner, G. L. & MacEachern, S. N. (2004), 'An ANOVA model for dependent random measures', *Journal of the American Statistical Association* **99**, 205–215.

De Iorio, M., MÃŒller, P., Rosner, G. L. & MacEachern, S. N. (2004), 'An anova model for dependent random measures', *Journal of the American Statistical Association* **99**(465), 205–215.

DeYoreo, M. & Kottas, A. (2018), 'Bayesian nonparametric modeling for multivariate ordinal regression', *Journal of computational and graphical statistics* **27**(1), 71–84.

Di Lucca, M. A., Guglielmi, A., Müller, P. & Quintana, F. A. (2013), 'A Simple Class of Bayesian Nonparametric Autoregression Models', *Bayesian analysis* **8**(1), 63–87.

Doksum, K. (1974), 'Tailfree and neutral random probabilities and their posterior distributions', *The Annals of Probability* **2**, 183–201.

Doss, H. (1994), 'Bayesian nonparametric estimation for incomplete data via successive substitution sampling', **22**, 1763–1786.

Doss, H. & Huffer, F. W. (2003), 'Monte Carlo methods for Bayesian analysis of survival data using mixtures of Dirichlet process priors', *Journal of Computational and Graphical Statistics* **12**, 282–307.

Duan, J. A., Guindani, M. & Gelfand, A. E. (2007), 'Generalized Spatial Dirichlet Process Models', *Biometrika* **94**(4), 809–825.

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

Escobar, M. D. & West, M. (1995), 'Bayesian density estimation and inference using mixtures', *Journal of the American Statistical Association* **90**, 577–588.

Ferguson, T. S. (1973), 'A Bayesian analysis of some nonparametric problems', *The Annals of Statistics* **1**, 209–230.

Ferguson, T. S. & Phadia, E. G. (1979), 'Bayesian nonparametric estimation based on censored data', *The Annals of Statistics* **7**, 163–186.

Fronczyk, K. & Kottas, A. (2014), 'A Bayesian nonparametric modeling framework for developmental toxicity studies', *Journal of the American Statistical Association* **109**, 873–888.

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* pp. 289–305.

Gelfand, A. E., Kottas, A. & MacEachern, S. N. (2005), 'Bayesian nonparametric spatial modeling with Dirichlet process mixing', *Journal of the American Statistical Association* **100**, 1021–1035.

Ghosal, S., Ghosh, J. K. & Van Der Vaart, A. (2000), 'Convergence rates of posterior distributions', *Annals of Statistics* **28**, 500–531.

Ghosal, S. & Vaart, A. W. v. d. (2016), *Fundamentals of nonparametric Bayesian inference*, Cambridge series in statistical and probabilistic mathematics ; 44, Cambridge University Press, Cambridge.

Ghosh, D. & Lin, D. Y. (2000), 'Nonparametric analysis of recurrent events and death', *Biometrics* **56**(2), 554–562.

Gui, W., Huang, R. & Lin, X. S. (2018), 'Fitting the Erlang mixture model to data

via a GEM-CMM algorithm', *Journal of Computational and Applied Mathematics* **343**, 189–205.

Hanson, T. E. (2006), 'Modeling censored lifetime data using a mixture of gammas baseline', *Bayesian Analysis* **1**, 575–594.

Hanson, T., Johnson, W. & Laud, P. (2009), 'Semiparametric inference for survival models with step process covariates', *Canadian Journal of Statistics* **37**, 60–79.

Hanson, T. & Johnson, W. O. (2002), 'Modeling regression error with a mixture of Pólya trees', *Journal of the American Statistical Association* **97**, 1020–1033.

Hartigan, J. A. (1990), 'Partition models', *Communications in Statistics - Theory and Methods* **19**(8), 2745–2756.

Haupt, G. & Mansmann, U. (1995), 'Survcart: S and c code for classification and regression trees analysis with survival data'.

He, L., Chapple, A., Liao, Z., Komaki, R., Thall, P. F. & H., L. S. (2016), 'Bayesian Regression Analyses of Radiation Modality Effects on Pericardial and Pleural Effusion and Survival in Esophageal Cancer', *Radiother Oncology* **121**, 70 – 74.

Hjort, N. (1990), 'Nonparametric Bayes estimators based on Beta processes in models for life history data', *The Annals of Statistics* **18**, 1259–1294.

Huang, C. & Wang, M. (2004), 'Joint modeling and estimation for recurrent event processes and failure time data', *Journal of the American Statistical Association* **99**(468), 1153–1165.

Huang, X. & Liu, L. (2007), 'A Joint Frailty Model for Survival and Gap Times Between Recurrent Events', *Biometrics* **63**(2), 389–397.

Huang, Y. & Wang, M. (2003), 'Frequency of recurrent events at failure time: Modeling and inference', *Journal of the American Statistical Association* **98**(463), 663–670.

Ibrahim, J. G., Chen, M. & Sinha, D. (2001), *Bayesian Survival Analysis*, Springer, New York, NY.

Ishwaran, H. & James, L. F. (2001), 'Gibbs sampling methods for stick-breaking priors', *Journal of the American Statistical Association* **96**(453), 161–173.

Ishwaran, H. & Zarepour, M. (2000), 'Markov chain monte carlo in approximate dirichlet and beta two-parameter process hierarchical models', *Biometrika* **87**(2), 371–390.

Jara, A., Hanson, T., Quintana, F. A., Müller, P. & Rosner, G. L. (2011), 'DPpackage: Bayesian Semi- and Nonparametric Modeling in R', *Journal of Statistical Software* **40**(5), 1 – 30.

Kalbfleisch, J. (1978), 'Nonparametric Bayesian Analysis of Survival Time Data', *Journal of the Royal Statistical Society: Series B (Methodological)* **40**, 214–221.

Kalbfleisch, J. D., Schaubel, D. E., Ye, Y. & Gong, Q. (2013), 'Estimating Function Approach to the Analysis of Recurrent and Terminal Events', *Biometrics* **69**(2), 366–374.

Kim, H. & Kottas, A. (2022), 'Erlang mixture modeling for Poisson process intensities', *Statistics and Computing* **32**, 3.

Klein, J. P. & Moeschberger, M. L. (1997), *Survival Analysis: Techniques for Censored and Truncated Data*, Statistics for Biology and Health, Springer.

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

Koul, H., Susarla, V. & V., R. J. (1981*a*), 'Regression analysis with randomly right-censored data', *The Annals of Statistics* **9**, 1276 – 1288.

Koul, H., Susarla, V. & V., R. J. (1981*b*), 'Regression analysis with randomly right-censored data', *The Annals of Statistics* **9**(6), 1276 – 1288.

Lancaster, T. & Intrator, O. (1998), 'Panel data with survival: Hospitalization of hiv-positive patients', *Journal of the American Statistical Association* **93**(441), 46–53.

Lavine, M. (1992), 'Some aspects of pólya tree distributions for statistical modelling', *The Annals of Statistics* **20**, 1222–1235.

Lee, J., Thall, P. F. & Lin, S. H. (2019), 'Bayesian semiparametric joint regression analysis of recurrent adverse events and survival in esophageal cancer patients.', *The annals of applied statistics* **13**(1), 221–247.

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.

Liu, L., Wolfe, R. A. & Huang, X. (2004), 'Shared Frailty Models for Recurrent Events and a Terminal Event', *Biometrics* **60**(3), 747–756.

Loader, C. (2022), 'locfit: Local regression, likelihood and density estimation'. R package version 1.5-9.6.

Loprinzi, C. L., Laurie, J. A., Wieand, H. S., Krook, J. E., Novotny, P. J., Kugler, J. W., Bartel, J., Law, M., Bateman, M. & Klatt, N. E. (1994), 'Prospective evaluation of prognostic variables from patient-completed questionnaires. North Central Cancer Treatment Group.', *Journal of Clinical Oncology* **12**, 601–607.

MacEachern, S. & Muller, P. (1998), 'Estimating Mixture of Dirichlet Process Models', *Journal of computational and graphical statistics* **7**(2), 223–238.

MacEachern, S. N. (1994), 'Estimating normal means with a conjugate style Dirichlet process prior', *Communications in Statistics - Simulation and Computation* **23**, 727–741.

MacEachern, S. N. (2000), 'Dependent Dirichlet processes', *Technical Report, Ohio State University* .

Mitra, R. & Müller, P., eds (2015), *Nonparametric Bayesian Inference in Biostatistics*, Springer, Cham, Switzerland.

Muliere, P. & Walker, S. (1997), 'A Bayesian nonparametric approach to survival analysis using Pólya trees', *Scandinavian Journal of Statistics* **24**, 331–340.

Müller, P., Quintana, F. A., Jara, A. & Hanson, T. (2015), *Bayesian nonparametric data analysis*, Springer Series in Statistics, 1st ed. 2015. edn, Springer International Publishing, Cham.

Müller, P., Erkanli, A. & West, M. (1996), 'Bayesian curve fitting using multivariate normal mixtures', *Biometrika* **83**(1), 67–79.

Müller, P., Quintana, F. & Rosner, G. L. (2011), 'A Product Partition Model with Regression on Covariates', *Journal of computational and graphical statistics* **20**(1), 260–278.

Neal, M. (2000), 'Markov Chain sampling methods for Dirichlet process mixture models', *Journal of Computational and Graphical Statistics* **9**, 249–265.

Neal, R. M. (2003), 'Slice sampling', *The Annals of Statistics* **31**(3), 705–741.

Nieto-Barajas, L. E., Müller, P., Ji, Y., Lu, Y. & Mills, G. B. (2012), 'A time-series ddp for functional proteomics profiles', *Biometrics* **68**(3), 859–868.

Olkin, I. & Trikalinos, T. A. (2015), 'Constructions for a bivariate Beta distribution', *Statistics and Probability Letters* .

Ouyang, B., Sinha, D., Slate, E. H. & Van Bakel, B. (2013), 'Bayesian analysis of recurrent event with dependent termination: an application to a heart transplant study', *Statistics in Medicine* **32**(15), 2629–2642.

Park, J. & Dunson, D. B. (2010), 'Bayesian generalized product partition model', *Statistica Sinica* **20**(3).

Paulon, G., De Iorio, M., Guglielmi, A. & Ieva, F. (2020), 'Joint modeling of recurrent events and survival: a bayesian non-parametric approach.', *Biostatistics (Oxford, England)* **21**(1), 1–14.

Petrone, S. (1999*a*), 'Bayesian density estimation using Bernstein polynomials', *The Canadian Journal of Statistics* **27**(1), 105–126.

Petrone, S. (1999*b*), 'Random Bernstein polynomials', *Scandinavian Journal of Statistics* **26**(3), 373–393.

Petrone, S. & Wasserman, L. (2002), 'Consistency of Bernstein polynomial posteriors', *Journal of the Royal Statistical Society, Series B* **64**(1), 79–100.

Phadia, E. G. (2013), *Prior Processes and Their Applications*, Springer, Berlin Heidelberg.

Pitman, J. (1996), 'Some Developments of the Blackwell-Macqueen URN Scheme', *Lecture Notes-Monograph Series* **30**, 245–267.

Polson, N. G., Scott, J. G. & Windle, J. (2013), 'Bayesian Inference for Logistic Models using Pólya-Gamma Latent Variables', *Journal of the American Statistical Association* **108**(504), 1339–1349.

Poynor, V. & Kottas, A. (2019), 'Nonparametric Bayesian inference for mean residual life functions in survival analysis.', *Biostatistics* **20**, 240–255.

Poynor, V. & Kottas, A. (2024), 'Bayesian nonparametric mean residual life regression', *arXiv.org* .

Quintana, F. A. & Iglesias, P. L. (2003), 'Bayesian Clustering and Product Partition Models', *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **65**(2), 557–574.

Quintana, F. A., Müller, P., Jara, A. & MacEachern, S. N. (2022), 'The Dependent Dirichlet Process and Related Models', *Statistical Science* **37**(1), 24 – 41.

Quintana, F. A., Müller, P., Jara, A. & MacEachern, S. N. (2022), 'The dependent Dirichlet process and related models', *Statistical Science* **37**, 24–41.

Reich, B. J. & Fuentes, M. (2007), 'A Multivariate Semiparametric Bayesian Spatial Modeling Framework for Hurricane Surface Wind Fields', *The annals of applied statistics* **1**(1), 249–264.

Richardson, R., Kottas, A. & Sansó, B. (2020), 'Spatiotemporal modelling using integro-difference equations with bivariate stable kernels', *Journal of the Royal Statistical Society, Series B* **82**, 1371–1392.

Riva-Palacio, A., Leisen, F. & Griffin, J. (2021), 'Survival regression models with dependent Bayesian nonparametric priors', *Journal of the American Statistical Association* **117**, 1530–1539.

Roberts, G. O. & Rosenthal, J. S. (2009*a*), 'Examples of adaptive MCMC', *Journal of Computational and Graphical Statistics* **18**, 349–367.

Roberts, G. O. & Rosenthal, J. S. (2009*b*), 'Examples of Adaptive MCMC', *Journal of Computational and Graphical Statistics* **18**(2), 349–367.

Rodriguez, A. & ter Horst, E. (2008), 'Bayesian Dynamic Density Estimation', *Bayesian analysis* **3**(2), 339–365.

Rondeau, V., Mathoulin-Pelissier, S., Jacqmin-Gadda, H., Brouste, V. & Soubeyran, P. (2007), 'Joint Frailty Models for Recurring Events and Death Using Maximum Penalized Likelihood Estimation: Application on Cancer Events', *Biostatistics (Oxford, England)* **8**(4), 708–721.

Royston, P. & Parmar, M. K. B. (2002), 'Flexible parametric proportional-hazards and proportional-odds models for censored survival data, with application to prognostic modelling and estimation of treatment effects', *Statistics in medicine* **21**(15), 2175–2197.

Rubio, F. J. & Hong, Y. (2016), 'Survival and lifetime data analysis with a flexible class of distributions', *Journal of Applied Statistics* **43**, 1794–1813.

Schomoor, C., Olschewski, M. & Schumacher, M. (1996), 'Randomized and Non-randomized patients in clinicial trials: experieneces with comprehensive cohort studies', *Statistics in medicine* **15**(3), 263–271.

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

Shi, Y., Laud, P. & J., N. (2021), 'A dependent Dirichlet process model for survival data with competing risks.', *Lifetime Data Anal* pp. 156—-176.

Sinha, D. & Maiti, T. (2004), 'A bayesian approach for the analysis of panel-count data with dependent termination', *Biometrics* **60**, 34–40.

Sinha, D., Maiti, T., Ibrahim, J. G. & Ouyang, B. (2008), 'Current methods for recurrent

events data with dependent termination: A bayesian perspective', *Journal of the American Statistical Association* **103**(482), 866–878.

Sparapani, R. A., Rein, L. E., Tarima, S. S., Jackson, T. A. & Meurer, J. R. (2020), 'Non-parametric recurrent events analysis with bart and an application to the hospital admissions of patients with diabetes', *Biostatistics (Oxford, England)* **21**(1), 69–85.

Stafanski, A. L. (1991), 'A normal scale mixture representation of the logistic distribution', *Statistics and Probability Letters* **11**(1), 69–70.

Susarla, V. & Van Ryzin, J. (1976), 'Nonparametric Bayesian estimation of survival curves from incomplete observations', *Journal of the American Statistical Association* **71**, 897–902.

Taddy, M. A. & Kottas, A. (2010), 'A bayesian nonparametric approach to inference for quantile regression', *Journal of business & economic statistics* **28**(3), 357–369.

Tawiah, R., McLachlan, G. J. & Ng, S. K. (2019), 'A Bivariate Joint Frailty Model with Mixture Framework for Survival Analysis of Recurrent Events with Dependent Censoring and Cure Fraction', *Biometrics* **76**, 753–766.

Tijms, H. C. (1994), *Stochastic Models: An Algorithmic Approach*, Chichester: Wiley.

Venturini, S., Dominici, F. & Parmigiani, G. (2008), 'Gamma shape mixtures for heavy-tailed distributions', *The Annals of Applied Statistics* **2**, 756–776.

Wade, S., Dunson, D. B., Petrone, S. & Trippa, L. (2014), 'Improving prediction

from dirichlet process mixtures via enrichment', *Journal of machine learning research* **15**, 1041–1071.

Walker, S. & Muliere, P. (1997), 'Beta-Stacy processes and a generalization of Pólya-urn scheme', *The Annals of Statistics* **25**, 1762–1780.

Wen, S., Huang, X., Frankowski, R. F., Cormier, J. N. & Pisters, P. (2016), 'A Bayesian Multivariate Joint Frailty Model for Disease Recurrences and Survival: A Bayesian Multivariate Joint Frailty Model', *Statistics in medicine* **35**(26), 4794–4812.

Xiao, S., Kottas, A., Sansó, B. & Kim, H. (2021), 'Nonparametric Bayesian modeling and estimation for renewal processes', *Technometrics* **63**, 100–115.

Xu, G., Chiou, S. H., Huang, C.-Y., Wang, M.-C. & Yan, J. (2017), 'Joint scale-change models for recurrent events and failure time', *Journal of the American Statistical Association* **112**(518), 794–805.

Xu, Z., Sinha, D. & R., B. J. (2021), 'Joint analysis of recurrence and termination: A bayesian latent class approach', *Statistical methods in medical research* **30**(2), 508–522.

Ye, Y., Kalbfleisch, J. D. & Schaubel, D. E. (2007), 'Semiparametric Analysis of Correlated Recurrent and Terminal Events', *Biometrics* **63**(1), 78–87.

Ying, Z., Jung, S. & Wei, L. (1995), 'Survival analysis with median regression models', *Journal of the American Statistical Association* **90**, 178–184.

Yu, Z. & Liu, L. (2011), 'A Joint Model of Recurrent Events and a Terminal Event with a Nonparametric Covariate Function', *Statistics in medicine* **30**(22), 2683–2695.

Zhang, Y. (2013), 'Epidemiology of esophageal cancer', *World journal of gastroenterology : WJG* **19**(34), 5598–5606.

Zheng, Y., Zhu, J. & Roy, A. (2010), 'Nonparametric Bayesian inference for the spectral density function of a random field', *Biometrika* **97**(1), 238–245.

Appendix A

Supplementary Material: Bayesian Nonparametric Erlang Mixture Modeling for Survival Analysis

A.1 MCMC algorithm for the DP-based Erlang mixture model

In this section, we provide details of posterior simulation for the DP-based Erlang mixture model in Section 2.2.1. Recall that we have the augmented model using

latent variables ϕ_i,

$$
\begin{aligned}
t_i \mid \phi_i, \theta, M &\overset{ind.}{\sim} \sum_{m=1}^{M} \mathbb{1}_{B_m}(\phi_i)\mathrm{Er}(t \mid m, \theta), \\
(\phi_1, \ldots, \phi_n) \mid \alpha, \zeta &\sim \mathrm{Exp}(\phi_1 \mid \zeta) \prod_{i=2}^{n} \left\{ \frac{\alpha}{\alpha + i - 1}\mathrm{Exp}(\phi_i \mid \zeta) + \frac{1}{\alpha + i - 1} \sum_{j=1}^{i-1} \delta_{\phi_j}(\phi_i) \right\}, \\
\zeta &\sim \text{inv-Ga}(a_\zeta, b_\zeta), \\
\theta &\sim \mathrm{Ga}(a_\theta, b_\theta), \\
M \mid \theta &\sim \mathrm{Unif}(\lceil M_1/\theta \rceil, \ldots, \lceil M_2/\theta \rceil), \\
\alpha &\sim \mathrm{Ga}(a_\alpha, b_\alpha),
\end{aligned}
$$

where $B_m = ((m-1)\theta, m\theta]$ for $m = 1, \ldots, M-1$, and $B_M = ((M-1)\theta, \infty)$. Here, $\mathrm{Er}(t \mid a, b)$ denotes the density of the gamma distribution with shape parameter a and scale parameter b evaluated at t, and $\mathrm{Exp}(\phi \mid a)$ the density of the exponential distribution with mean parameter a evaluated at ϕ. The likelihood function under the augmented model can be written as

$$
L(M, \theta, \boldsymbol{\phi}; \mathcal{D}) = \prod_{i=1}^{n} \sum_{m=1}^{M} \left\{ \mathbb{1}_{B_m}(\phi_i)\mathrm{Er}(y_i \mid m, \theta) \right\}^{\nu_i} \left\{ \mathbb{1}_{B_m}(\phi_i) S_{\mathrm{Er}}(y_i \mid m, \theta) \right\}^{1-\nu_i}, \quad \text{(A.1)}
$$

where $S_{\mathrm{Er}}(y_i \mid m, \theta) = \int_{y_i}^{\infty} \mathrm{Er}(u \mid m, \theta) du$, $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_n)$, and $\mathcal{D} = \{(y_i, \nu_i), i = 1, \ldots, n\}$. The joint posterior distribution of the random parameters, $\boldsymbol{\phi}, \theta, M, \zeta$, and α is

$$
\begin{aligned}
p(\boldsymbol{\phi}, \theta, M, \zeta, \alpha \mid \mathcal{D}) \propto & \prod_{i=1}^{n} \sum_{m=1}^{M} \left\{ \mathbb{1}_{B_m}(\phi_i)\mathrm{Er}(y_i \mid m, \theta) \right\}^{\nu_i} \left\{ \mathbb{1}_{B_m}(\phi_i) S_{\mathrm{Er}}(y_i \mid m, \theta) \right\}^{1-\nu_i} \\
& \times p(\boldsymbol{\phi} \mid \alpha, \zeta) p(\zeta) p(\theta) p(M \mid \theta) p(\alpha).
\end{aligned}
$$

We use a Metropolis-within-Gibbs algorithm for posterior simulation if direct sampling is not available. The parameters in the proposal distributions for Metropolis-Hastings

update are automatically tuned by adaptive Metropolis-Hastings algorithms in Roberts & Rosenthal (2009*a*) for fast convergence and improved mixing. We checked mixing and convergence of the Markov chain and did not find any evidence indicating those issues. The full conditionals are given below.

1. M and θ

   - Sample M from the following categorical distribution,

$$\begin{aligned} p(M = j_M \mid -) &= \frac{L(M = j_M, \theta, \boldsymbol{\phi}; \mathcal{D})}{\sum_{i_M=\left\lceil \frac{M_1}{\theta} \right\rceil}^{\left\lceil \frac{M_2}{\theta} \right\rceil} L(M = i_M, \theta, \boldsymbol{\phi}; \mathcal{D})}, \\ \text{where } j_M &= \left\lceil \frac{M_1}{\theta} \right\rceil, \ldots, \left\lceil \frac{M_2}{\theta} \right\rceil, \end{aligned}$$

     where $L(j_M, \theta, \boldsymbol{\phi}; \mathcal{D})$ is the likelihood function of the augmented model in (A.1) evaluated with $M = j_M$ and the current values of $\boldsymbol{\phi}$ and θ.

   - The full conditional of θ is

$$p(\theta \mid -) \propto \text{Ga}(\theta \mid a_\theta, b_\theta) L(M, \theta, \boldsymbol{\phi}; \mathcal{D}).$$

     We update θ using a random walk Metropolis-Hasting algorithm.

   - We also jointly update (M, θ) via a Metropolis-Hasting algorithm. Given the current values $(M^{(t-1)}, \theta^{(t-1)})$ at iteration t, we first generate a proposal, $\theta^\star$ of θ; $\log(\theta^*) \sim \text{N}(\log(\theta^{(t-1)}), \epsilon)$, where ϵ is an adaptive step size, and generate

$M^\star$ from

$$\begin{aligned} q(M^\star = j_M \mid M^{(t-1)}, \theta^\star) &= \frac{\{(j_M - M^{(t-1)})^2 + 1\}^{-1}}{\sum_{i_M=\left\lceil \frac{M_1}{\theta} \right\rceil}^{\left\lceil \frac{M_2}{\theta} \right\rceil} \{(i_M - M^{(t-1)})^2 + 1\}^{-1}}, \\ \text{where } j_M &= \left\lceil \frac{M_1}{\theta^*} \right\rceil, \ldots, \left\lceil \frac{M_2}{\theta^*} \right\rceil. \end{aligned}$$

We then accept $(\theta^\star, M^\star)$ with probability $\min(1, r^\star)$, where

$$r^\star = \frac{\theta^\star p(\theta^\star) p(M^\star \mid \theta^\star) L(M^\star, \theta^\star, \boldsymbol{\phi}; \mathcal{D}) q(M^{(t-1)} \mid \theta^{(t-1)}, M^\star)}{\theta^{(t-1)} p(\theta^{(t-1)}) p(M^{(t-1)} \mid \theta^{(t-1)}) L(M^{(t-1)}, \theta^{(t-1)}, \boldsymbol{\phi}; \mathcal{D}) q(M^\star \mid \theta^\star, M^{(t-1)})}.$$

2. ζ

Let $\boldsymbol{\phi}^\star = (\phi_1^\star, \ldots, \phi_{n^*}^\star)$ the set of all distinct values in $(\phi_1, \ldots, \phi_n)$ and $n^\star$ the number of elements in $\boldsymbol{\phi}^\star$. The full conditional of ζ is

$$\text{inv-Ga}\left(a_\zeta + n^\star, b_\zeta + \sum_{j=1}^{n^\star} \phi_j^\star\right).$$

3. α

We use the augmentation method in (Escobar & West 1995) to update α. We first introduce an auxiliary variable η, $\eta \mid \alpha, n \sim \text{Be}(\alpha + 1, n)$, and sample α from a mixture of two gamma distributions;

$$\begin{aligned} \alpha \mid - \;\sim\; & \frac{a_\alpha + n^\star - 1}{n(b_\alpha^{-1} - \log(\eta)) + a_\alpha + n^\star - 1} \text{Ga}(a_\alpha + n^\star, (b_\alpha^{-1} - \log(\eta))^{-1}) \\ & + \frac{n(b_\alpha^{-1} - \log(\eta))}{n(b_\alpha^{-1} - \log(\eta)) + a_\alpha + n^\star - 1} \text{Ga}(a_\alpha + n^\star - 1, (b_\alpha^{-1} - \log(\eta))^{-1}). \end{aligned}$$

4. $\boldsymbol{\phi}$

Let $\boldsymbol{\phi}_i^{\star -} = (\phi_1^{\star -}, \ldots, \phi_{n^{\star -}}^{\star -})$ be the set of distinct values in $\boldsymbol{\phi}_{-i}$, where $\boldsymbol{\phi}_{-i} = (\phi_1, \ldots, \phi_{i-1}, \phi_{i+1}, \ldots, \phi_n)$ and $n^{\star -}$ is the number of elements in $\boldsymbol{\phi}_i^{\star -}$. Let n_j^- be

the number of elements in $\boldsymbol{\phi}_{-i}$ that equal $\phi_j^{\star -}$. The full conditional of ϕ_i is

$$\begin{aligned}\phi_i \mid \boldsymbol{\phi}_{-i}, y_i, \alpha, \zeta, \theta, M \quad \sim \quad & \frac{\alpha q_0}{\alpha q_0 + \sum_{j=1}^{n^{*-}} n_j^- q_j} h(\phi_i \mid y_i, \theta, M, \zeta) \\ & + \sum_{j=1}^{n^{\star -}} \frac{n_j^- q_j}{\alpha q_0 + \sum_{k=1}^{n^{\star -}} n_k^- q_k} \delta_{\phi_j^{*-}}(\phi_i),\end{aligned}$$

where

$$\begin{aligned}q_0 \quad = \quad & \sum_{m=1}^{M-1} \{G_{\text{Exp}}(m\theta \mid \zeta) - G_{\text{Exp}}((m-1)\theta \mid \zeta)\} \{\text{Er}(y_i \mid m, \theta)\}^{\nu_i} \{S_{\text{Er}}(y_i \mid m, \theta)\}^{1-\nu_i} \\ & + \{1 - G_{\text{Exp}}((M-1)\theta \mid \zeta)\} \{\text{Er}(y_i \mid m, \theta)\}^{\nu_i} \{S_{\text{Er}}(y_i \mid m, \theta)\}^{1-\nu_i}, \\ q_j \quad = \quad & \sum_{m=1}^{M-1} \mathbb{1}_{((m-1)\theta, m\theta]}(\phi_j^{*-}) \{\text{Er}(y_i \mid m, \theta)\}^{\nu_i} \{S_{\text{Er}}(y_i \mid m, \theta)\}^{1-\nu_i} \\ & + \mathbb{1}_{((M-1)\theta, \infty)}(\phi_j^{*-}) \{\text{Er}(y_i \mid m, \theta)\}^{\nu_i} \{S_{\text{Er}}(y_i \mid m, \theta)\}^{1-\nu_i},\end{aligned}$$

with $G_{\text{Exp}}(\cdot \mid \zeta)$ denoting the exponential distribution function with mean ζ, and

$$h(\phi_i \mid y_i, \theta, M, \zeta) \quad = \quad \sum_{m=1}^{M} \Omega_m \text{T-Exp}_m(\phi_i \mid \zeta),$$

with

$$\begin{aligned}\Omega_m \quad = \quad & \{\text{Er}(y_i \mid m, \theta)\}^{\nu_i} \{S_{\text{Er}}(y_i \mid m, \theta)\}^{1-\nu_i} \\ & \times (G_{\text{Exp}}(m\theta \mid \zeta) - G_{\text{Exp}}((m-1)\theta \mid \zeta)) q_0^{-1}, m = 1, \ldots, M-1, \\ \Omega_M \quad = \quad & \{\text{Er}(y_i \mid M, \theta)\}^{\nu_i} \{S_{\text{Er}}(y_i \mid M, \theta)\}^{1-\nu_i} \\ & \times (1 - G_{\text{Exp}}((M-1)\theta \mid \zeta)) q_0^{-1}.\end{aligned}$$

Here, $h(\phi_i \mid y_i, \theta, M, \zeta)$ is a mixture of truncated exponential distributions, and $\text{T-Exp}_m(\phi \mid \zeta)$ is the density function of the truncated exponential distribution with mean parameter ζ with the support $((m-1)\theta, m\theta]$. ϕ_i is equal to $\phi_j^{\star -}$ with probability $n_j^- q_j / A$, where $A = \alpha q_0 + \sum_{h=1}^{n^{\star -}} n_h^- q_h$; or it is drawn from $h(\phi_i \mid$

$t_i, \theta, M, \zeta)$. The inverse-cdf sampling method can be used to draw a sample from $h(\phi_i \mid t_i, \theta, M, \zeta)$.

A.2 MCMC algorithm for the DDP mixture model

We present here the posterior simulation details for the model developed in Section 2.2.2. The augmented model using latent variables $\boldsymbol{\varphi}_i = (\varphi_{Ci}, \varphi_{Ti})$ is written as

$$
\begin{aligned}
t_i \mid M_{x_i}, \theta_{x_i}, \varphi_{x_i} &\overset{ind.}{\sim} \sum_{m=1}^{M_{x_i}} \mathbb{1}_{B_{x_i m}}(\varphi_{x_i,i}) \mathrm{Er}(t \mid m, \theta_{x_i}),\ i = 1, \ldots, n, \text{ and } x_i \in \{C, T\}, \\
(\boldsymbol{\varphi}_1, \ldots, \boldsymbol{\varphi}_n) \mid \alpha, \boldsymbol{\mu} &\sim \mathrm{LN}_2(\boldsymbol{\varphi}_1 \mid \boldsymbol{\mu}, \Sigma) \prod_{i=2}^{n} \left\{ \frac{\alpha}{\alpha + i - 1} \mathrm{LN}_2(\boldsymbol{\varphi}_i \mid \boldsymbol{\mu}, \Sigma) + \frac{1}{\alpha + i - 1} \sum_{j=1}^{i-1} \delta_{\boldsymbol{\varphi}_j}(\boldsymbol{\varphi}_i) \right\}, \\
\theta_x &\overset{ind.}{\sim} \mathrm{Ga}(a_{x\theta}, b_{x\theta}), \\
M_x \mid \theta_x &\overset{ind.}{\sim} \mathrm{Unif}(\lceil M_{x1}/\theta_x \rceil, \ldots, \lceil M_{x2}/\theta_x \rceil), \\
\alpha &\sim \mathrm{Ga}(a_\alpha, b_\alpha), \\
\boldsymbol{\mu} &\sim \mathrm{N}_2(\bar{\boldsymbol{\mu}}, \Sigma_0), \\
\Sigma &\sim \text{inv-Wish}(r, R),
\end{aligned}
$$

where $B_{x_i m} = ((m-1)\theta_{x_i}, m\theta_{x_i}]$ for $m = 1, \ldots, M_{x_i} - 1$, and $B_{x_i M_{x_i}} = ((M_{x_i} - 1)\theta_{x_i}, \infty)$. The likelihood function for the augmented model for observation i is

$$
L_i(M_{x_i}, \theta_{x_i}, \varphi_{x_i,i}; \mathcal{D}) = \left[\sum_{m=1}^{M_{x_i}} \mathbb{1}_{B_{x_i m}}(\varphi_{x_i,i}) \mathrm{Er}(y_i \mid m, \theta_{x_i}) \right]^{\nu_i} \left[\sum_{m=1}^{M_{x_i}} \mathbb{1}_{B_{x_i m}}(\varphi_{x_i,i}) S_{\mathrm{Er}}(y_i \mid m, \theta_{x_i}) \right]^{1-\nu_i}.
$$

where $\mathcal{D} = \{(y_i, \nu_i, x_i), i = 1, \ldots, n\}$ denotes data. Similar to the algorithm in Appendix A, we use an adaptive Metropolis-within-Gibbs algorithm in Roberts & Rosenthal (2009*a*) for the Metropolis-Hastings updates. Mixing and convergence of Markov chain are checked, and no evidence indicating those issues is found. The full conditionals

are given below.

1. $\boldsymbol{M} = (M_C, M_T)$

   Sample M_C from the following categorical distribution,

   $$p(M_C = j_M \mid -) = \frac{L_C(j_M, \theta_C, \boldsymbol{\varphi}; \mathcal{D})}{\sum_{i_M=\left\lceil \frac{M_{C1}}{\theta_C} \right\rceil}^{\left\lceil \frac{M_{C2}}{\theta_C} \right\rceil} L_C(i_M, \theta_C, \boldsymbol{\varphi}; \mathcal{D})}, \qquad j_M = \left\lceil \frac{M_{C1}}{\theta_C} \right\rceil, \ldots, \left\lceil \frac{M_{C2}}{\theta_C} \right\rceil,$$

   where $L_C(j_M, \theta_C, \boldsymbol{\varphi}; \mathcal{D}) = \prod_{i:x_i=C} L_i(j_M, \theta_C, \varphi_{Ci}; \mathcal{D})$. We then draw M_T in a similar way.

2. $\boldsymbol{\theta} = (\theta_C, \theta_T)$

   The full conditional of $\boldsymbol{\theta}$ is

   $$p(\boldsymbol{\theta} \mid -) \propto \mathrm{Ga}(\theta_C \mid a_{C\theta}, b_{C\theta}) \mathrm{Ga}(\theta_T \mid a_{T\theta}, b_{T\theta}) \prod_{i=1}^{n} L_i(M_{x_i}, \theta_{x_i}, \varphi_{x_i,i}; \mathcal{D}).$$

   We use the algorithm in Roberts & Rosenthal (2009*a*) to sample $\boldsymbol{\theta}$. Let $\boldsymbol{\theta}^{(t-1)} = (\theta_C^{(t-1)}, \theta_T^{(t-1)})$ the current values of $\boldsymbol{\theta}$. A proposal of $\boldsymbol{\theta}$ is generated from

   $$\log(\boldsymbol{\theta}^\star) \sim 0.95\mathrm{N}(\log(\boldsymbol{\theta}^{(t-1)}), 2.38^2/2\Sigma_n) + 0.05\mathrm{N}(\log(\boldsymbol{\theta}^{(t-1)}), 0.01/2I_2),$$

   where Σ_n is the empirical covariance matrix of $\log(\boldsymbol{\theta})$ based on the run so far. Then we accept $\boldsymbol{\theta}^\star$ with probability $\min(1, r^\star)$, where

   $$r^\star = \frac{\theta_C^\star \theta_T^\star \mathrm{Ga}(\theta_C^\star | a_{C\theta}, b_{C\theta}) \mathrm{Ga}(\theta_T^\star \mid a_{T\theta}, b_{T\theta}) \prod_{i=1}^n L_i(M_{x_i}, \theta_{x_i}^\star, \varphi_{x_i,i}; \mathcal{D})}{\theta_C^{(t-1)} \theta_T^{(t-1)} \mathrm{Ga}(\theta_C^{(t-1)} \mid a_{C\theta}, b_{C\theta}) \mathrm{Ga}(\theta_T^{(t-1)} \mid a_{T\theta}, b_{T\theta}) \prod_{i=1}^n L_i(M_{x_i}, \theta_{x_i}^{(t-1)}, \varphi_{x_i,i}; \mathcal{D})}.$$

3. $\boldsymbol{\mu}$

   Let $\boldsymbol{\varphi}^\star = (\boldsymbol{\varphi}_1^\star, \ldots, \boldsymbol{\varphi}_{n^\star}^\star)$ be the set of distinct values in $\boldsymbol{\varphi}$, where $n^\star$ is the number of elements in $\boldsymbol{\varphi}^\star$. The full conditional of $\boldsymbol{\mu}$ is

   $$\mathrm{N}_2(\boldsymbol{\mu}_1, \Sigma_1),$$

where

$$\Sigma_1 = \left[\Sigma_0^{-1} + n^\star \Sigma^{-1}\right]^{-1} \quad \text{and} \quad \boldsymbol{\mu}_1 = \Sigma_1 \left[\Sigma_0^{-1}\bar{\boldsymbol{\mu}} + \Sigma^{-1}\sum_{i=1}^{n^\star} \log(\boldsymbol{\varphi}_i^\star)\right].$$

4. Σ

The full conditional of Σ is,

$$\text{inv-Wishart}(r^*, R^*),$$

where

$$r^* = r + n^* \text{ and } R^* = R + \sum_{i=1}^{n^*} (\log(\varphi_i^*) - \boldsymbol{\mu})(\log(\varphi_i^*) - \boldsymbol{\mu})'.$$

5. α

We use the augmentation method in Escobar & West (1995) to update α. We first introduce an auxiliary variable η, $\eta \mid \alpha, n \sim \text{Be}(\alpha + 1, n)$, and sample α from a mixture of two gamma distributions;

$$\begin{aligned}\alpha \mid - \quad \sim \quad & \frac{a_\alpha + n^\star - 1}{n(b_\alpha^{-1} - \log(\eta)) + a_\alpha + n^\star - 1} \text{Ga}(a_\alpha + n^\star, (b_\alpha^{-1} - \log(\eta))^{-1}) \\ & + \frac{n(b_\alpha^{-1} - \log(\eta))}{n(b_\alpha^{-1} - \log(\eta)) + a_\alpha + n^\star - 1} \text{Ga}(a_\alpha + n^\star - 1, (b_\alpha^{-1} - \log(\eta))^{-1}).\end{aligned}$$

6. $\boldsymbol{\varphi}$

Let $\boldsymbol{\varphi}_i^{\star -} = (\boldsymbol{\varphi}_1^{\star -}, \ldots, \boldsymbol{\varphi}_{n^{\star -}}^{\star -})$ be the set of distinct values in $\boldsymbol{\varphi}_{-i} = (\boldsymbol{\varphi}_1, \ldots, \boldsymbol{\varphi}_{i-1}, \boldsymbol{\varphi}_{i+1}, \ldots, \boldsymbol{\varphi}_n)$, where $n^{\star -}$ is the number of elements in $\boldsymbol{\varphi}_i^{\star -}$. Let n_j^- be number of elements in $\boldsymbol{\varphi}_{-i}$ that is equal to $\boldsymbol{\varphi}_j^{\star -}$. The full conditional of $\boldsymbol{\varphi}_i$ is

$$\begin{aligned}\boldsymbol{\varphi}_i \mid \boldsymbol{\varphi}_{-i}, \boldsymbol{\mu}, \Sigma, \boldsymbol{\theta}, \boldsymbol{M}, \mathcal{D} \quad \sim \quad & \frac{\alpha q_0}{\alpha q_0 + \sum_{j=1}^{n^{*-}} n_j^- q_j} h(\boldsymbol{\varphi}_i \mid y_i, \boldsymbol{\mu}, \Sigma, \boldsymbol{\theta}, \boldsymbol{M}) \\ & + \sum_{j=1}^{n^{\star -}} \frac{n_j^- q_j}{\alpha q_0 + \sum_{j=1}^{n^{*-}} n_j^- q_j} \delta_{\boldsymbol{\varphi}_j^{*-}}(\boldsymbol{\varphi}_i),\end{aligned}$$

where, for $x_i = C$,

$$
\begin{aligned}
q_0 &= \sum_{m=1}^{M_C-1} \{\mathrm{Er}(y_i \mid m, \theta_C)\}^{\nu_i} \{S_{\mathrm{Er}}(y_i \mid m, \theta_C)\}^{1-\nu_i} \\
&\quad \times \{G_{\mathrm{LN}}(m\theta_C \mid \mu_{C|T}, \Sigma_{C|T}) - (G_{\mathrm{LN}}((m-1)\theta_C \mid \mu_{C|T}, \Sigma_{C|T})\} \\
&\quad + \{\mathrm{Er}(y_i \mid M_C, \theta_C)\}^{\nu_i} \{S_{\mathrm{Er}}(y_i \mid M_C, \theta_C)\}^{1-\nu_i} \{1 - G_{\mathrm{LN}}((M_C-1)\theta_C \mid \mu_{C|T}, \Sigma_{C|T})\}, \\
q_j &= \sum_{m=1}^{M_C-1} \mathbb{1}_{((m-1)\theta_C, m\theta_C]}(\varphi_{Ci}^{\star -}) \{\mathrm{Er}(y_i|m, \theta_C)\}^{\nu_i} \{S_{\mathrm{Er}}(y_i|m, \theta_C)\}^{1-\nu_i} \\
&\quad + \mathbb{1}_{((M_C-1)\theta_C, \infty)}(\varphi_{Ci}^{\star -}) \{\mathrm{Er}(y_i|M_C, \theta_C)\}^{\nu_i} \{S_{\mathrm{Er}}(y_i|M_C, \theta_C)\}^{1-\nu_i}, \\
\mu_{C|T} &= \mu_1 + \Sigma_{12}/\Sigma_{22}(\varphi_{Ti} - \mu_2), \\
\Sigma_{C|T} &= \Sigma_{11} - \Sigma_{12}\Sigma_{21}/\Sigma_{22}
\end{aligned}
$$

with $G_{\mathrm{LN}}(\cdot \mid \mu_{C|T}, \Sigma_{C|T})$ denoting a lognormal distribution function with mean $\mu_{C|T}$ and variance $\Sigma_{C|T}$, and

$$
h(\boldsymbol{\varphi}_i \mid y_i, \boldsymbol{\mu}, \Sigma, \boldsymbol{\theta}, \boldsymbol{M}) = \mathrm{LN}(\varphi_{Ti} \mid \mu_1, \Sigma_{11}) \times \sum_{m=1}^{M_C} \Omega_m \mathrm{T\text{-}LN}_m(\varphi_{Ci} \mid \mu_{C|T}, \Sigma_{C|T})
$$

with

$$
\begin{aligned}
\Omega_m &= \{\mathrm{Er}(y_i \mid m, \theta_C)\}^{\nu_i} \{S_{\mathrm{Er}}(y_i \mid m, \theta_C)\}^{1-\nu_i} \\
&\quad \times \{G_{\mathrm{LN}}(m\theta_C \mid \mu_{C|T}, \Sigma_{C|T}) - G_{\mathrm{LN}}((m-1)\theta_C \mid \mu_{C|T}, \Sigma_{C|T})\} q_0^{-1}, \\
&\quad \text{for } m = 1, \ldots, M_C - 1, \\
\Omega_{M_C} &= \{\mathrm{Er}(y_i \mid M_C, \theta_C)\}^{\nu_i} \{S_{\mathrm{Er}}(y_i \mid M_C, \theta_C)\}^{1-\nu_i} \\
&\quad \times \{1 - G_{\mathrm{LN}}((M_C-1)\theta_C \mid \mu_{C|T}, \Sigma_{C|T})\} q_0^{-1}.
\end{aligned}
$$

Similar to the algorithm of updating φ_i for the DP-based Erlang mixture model, we let $\boldsymbol{\varphi}_i = \boldsymbol{\varphi}_j^{\star -}$ with probability $n_j^- q_j / A$, where $A = \alpha q_0 + \sum_{h=1}^{n^{\star -}} n_h^- q_h$, or draw a new $\boldsymbol{\varphi}_i$ from $h(\boldsymbol{\varphi}_i \mid y_i, \boldsymbol{\mu}, \Sigma, \boldsymbol{\theta}, \boldsymbol{M})$ with probability $\alpha q_0 / A$. To draw a sample from $h(\boldsymbol{\varphi}_i \mid y_i, \boldsymbol{\mu}, \Sigma, \boldsymbol{\theta}, \boldsymbol{M})$, we first draw φ_{Ti} from $\mathrm{LN}(\mu_1, \Sigma_{11})$ and then, conditional on φ_{Ti}, draw φ_{Ci} from a mixture of truncated lognormal distributions using

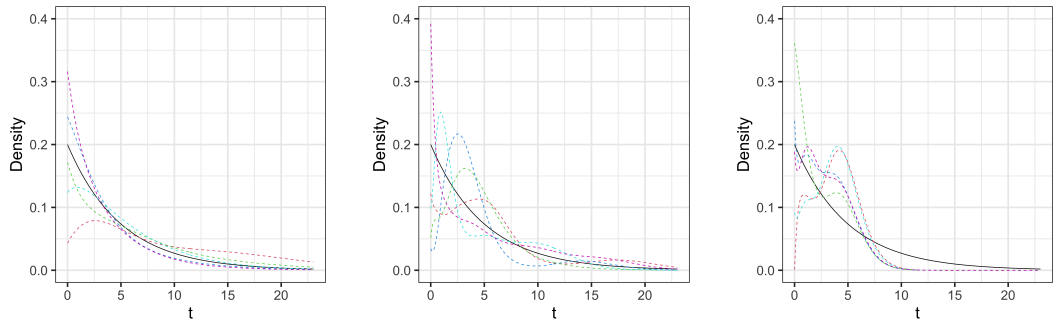
an inverse-cdf sampling method, where each component, T-LN$_m$ is a lognormal distribution with support of $((m-1)\theta_C, m\theta_C]$. The same method is applied for the observations with $x_i = T$ by simply switching C with T.

A.3 Prior realizations from DP-based Erlang mixture model

To illustrate the impact of (M, θ) on $f(t \mid M, \theta, \boldsymbol{\omega})$ in (1) of the main text, five prior realizations of the Erlang mixture density are shown by varying the values of (M, θ) in Figure A.1. Three combinations of (M, θ) are considered, and in all cases $\alpha = 10$, and an Exp(5) distribution for G_0 are used. For panels (a) and (b), $M\theta = 20$. The resulting density realizations have similar effective support, although the ones in panel (b) involve more variable shapes, as expected since the value of θ is smaller than that in panel (a). For panel (c), $M\theta = 5$, resulting in noticeably smaller effective support for the realized densities relative to panels (a) and (b).

A.4 Dependence structure under DDP-based Erlang mixture model

Here, we examine the dependence structure between the control and treatment time to event distributions under the DDP-based Erlang mixture model. In this



(a) $(M, \theta) = (10, 2)$ (b) $(M, \theta) = (40, 0.5)$ (c) $(M, \theta) = (10, 0.5)$

Figure A.1: Prior realizations for $f(t \mid M, \theta, G)$, under $(M, \theta) = (10, 2)$, $(40, 0.5)$, and $(10, 0.5)$. In all cases, $\alpha = 10$ and $G_0 = \text{Exp}(5)$. The black dashed line denotes the density of G_0.

respect, we follow the standard approach for DDP prior models, that is, we study the covariance (given prior hyperparameters) induced by the DDP prior across the relevant random distributions. In our context (refer to Section 2.2 of the main text), the random distributions are $F_x(\cdot)$, for $x \in \{C, T\}$, with associated densities,

$$f_x(t) \equiv f(t \mid M_x, \theta_x, G_x) = \sum_{m=1}^{M_x} \omega_{xm} \, \text{Er}(t \mid m, \theta_x), \quad t \in \mathbb{R}^+,$$

where $\omega_{xm} = G_x(B_{xm})$, with $B_{xm} = ((m-1)\theta_x, m\theta_x]$, for $m = 1, \ldots, M_x - 1$, and $B_{xM_x} = ((M_x - 1)\theta_x, \infty)$. Moreover, recall that

$$G_x = \sum_{\ell=1}^{\infty} p_\ell \, \delta_{\varphi^\star_{x\ell}}, \quad x \in \{C, T\}$$

with $p_1 = v_1$, $p_\ell = v_\ell \prod_{r=1}^{\ell-1}(1 - v_r)$, for $\ell \geq 2$, where $v_\ell \mid \alpha \overset{i.i.d.}{\sim} \text{Beta}(1, \alpha)$, $\boldsymbol{\varphi}^\star_\ell = (\varphi^\star_{C\ell}, \varphi^\star_{T\ell}) \mid \boldsymbol{\mu}, \Sigma \overset{i.i.d.}{\sim} G_0 = \text{LN}_2(\boldsymbol{\mu}, \Sigma)$, and $\{v_\ell\}$ and $\{\boldsymbol{\varphi}^\star_\ell\}$ are independent sequences of random variables.

Denote by $F_{\mathrm{Er}}(\cdot \mid m, \theta_x)$ the Erlang distribution with parameters m and θ_x, and by $\boldsymbol{\psi}$ the vector that comprises parameters $(\theta_C, \theta_T, M_C, M_T)$ as well as the DDP prior hyperparameters $(\alpha, \boldsymbol{\mu}, \Sigma)$. Then, for any (measurable) set $A \subset \mathbb{R}^+$, we can write the associated random probabilities as

$$F_x(A) \;=\; \sum_{m=1}^{M_x} G_x(B_{xm})\, F_{\mathrm{Er}}(A \mid m, \theta_x)$$

and, therefore, the (conditional) prior expectation for the control-treatment random probability is given by

$$\mathbb{E}(F_x(A) \mid \boldsymbol{\psi}) \;=\; \sum_{m=1}^{M_x} \mathbb{E}(G_x(B_{xm}) \mid \boldsymbol{\psi})\, F_{\mathrm{Er}}(A \mid m, \theta_x) \;=\; \sum_{m=1}^{M_x} G_{0x}(B_{xm})\, F_{\mathrm{Er}}(A \mid m, \theta_x). \quad \text{(A.2)}$$

Here, we have used the fact that, for each $x \in \{C, T\}$, G_x follows marginally a $\mathrm{DP}(\alpha, G_{0x})$ prior, where G_{0C} and G_{0T} are the marginals of the DDP centering distribution G_0 associated with the control and treatment group, respectively. Note that the prior expectation in (A.2) does not depend on the total mass parameter α.

To develop the expression for the covariance between $F_C(A)$ and $F_T(A)$, we also need

$$\mathbb{E}(F_C(A)F_T(A) \mid \boldsymbol{\psi}) \;=\; \sum_{m=1}^{M_C}\sum_{k=1}^{M_T} \mathbb{E}(G_C(B_{Cm})\, G_T(B_{Tk}) \mid \boldsymbol{\psi})\, F_{\mathrm{Er}}(A \mid m, \theta_C)\, F_{\mathrm{Er}}(A \mid k, \theta_T). \quad \text{(A.3)}$$

The product of the mixture weights can be written as

$$\begin{aligned} G_C(B_{Cm})\, G_T(B_{Tk}) \;&=\; \left\{\sum_{\ell=1}^{\infty} p_\ell\, \mathbb{1}(\varphi^\star_{C\ell} \in B_{Cm})\right\} \left\{\sum_{s=1}^{\infty} p_s\, \mathbb{1}(\varphi^\star_{Ts} \in B_{Tk})\right\} \\ &=\; \sum_{\ell=1}^{\infty} p_\ell^2\, \mathbb{1}\{(\varphi^\star_{C\ell}, \varphi^\star_{T\ell}) \in B_{Cm} \times B_{Tk}\} \\ &\quad +\; \sum_{\ell=1}^{\infty}\sum_{\substack{s=1 \\ \ell \neq s}}^{\infty} p_\ell\, p_s\, \mathbb{1}(\varphi^\star_{C\ell} \in B_{Cm})\, \mathbb{1}(\varphi^\star_{Ts} \in B_{Tk}). \end{aligned}$$

Hence, taking the expectation (conditional on $\boldsymbol{\psi}$) and using the DDP prior definition, we obtain

$$\begin{aligned}\mathbb{E}(G_C(B_{Cm})\,G_T(B_{Tk}) \mid \boldsymbol{\psi}) &= G_0(B_{Cm}\times B_{Tk})\sum_{\ell=1}^{\infty}\mathbb{E}(p_\ell^2 \mid \alpha) \\ &\quad + G_{0C}(B_{Cm})\,G_{0T}(B_{Tk})\sum_{\ell=1}^{\infty}\sum_{\substack{s=1\\ \ell\neq s}}^{\infty}\mathbb{E}(p_\ell\,p_s \mid \alpha).\end{aligned}$$

Now, we can use the definition of the DDP weights p_ℓ through the underlying i.i.d. Beta$(1,\alpha)$ random variables to obtain $\mathbb{E}(p_\ell^2 \mid \alpha) = 2\alpha^{\ell-1}/\{(\alpha+1)(\alpha+2)^\ell\}$, and thus $\sum_{\ell=1}^{\infty}\mathbb{E}(p_\ell^2 \mid \alpha) = (\alpha+1)^{-1}$. Moreover, for $\ell < s$,

$$\mathbb{E}(p_\ell\,p_s \mid \alpha) = \frac{\alpha^{s-1}}{(\alpha+1)^{s-\ell+1}\,(\alpha+2)^\ell}$$

and, since the v_ℓ are i.i.d., the expression is the same for $\ell > s$. Therefore,

$$\begin{aligned}\mathbb{E}(G_C(B_{Cm})\,G_T(B_{Tk}) \mid \boldsymbol{\psi}) &= 2\,G_{0C}(B_{Cm})\,G_{0T}(B_{Tk})\sum_{\ell=1}^{\infty}\sum_{\substack{s=1\\ \ell<s}}^{\infty}\frac{\alpha^{s-1}}{(\alpha+1)^{s-\ell+1}\,(\alpha+2)^\ell} \\ &\quad + \frac{1}{\alpha+1}\,G_0(B_{Cm}\times B_{Tk}) \qquad (A.4)\end{aligned}$$

Finally, the expression for $\mathbb{C}\mathrm{o}\mathbb{v}(F_C(A), F_T(A) \mid \boldsymbol{\psi})$ can be obtained by combining (A.4) with (A.3), and using also (A.2). An interesting special case for set A is (t,∞), resulting in the covariance between the control and treatment group survival functions at any specified time point $t \in \mathbb{R}^+$. The derivation can be easily extended to different sets for the control and treatment group distributions. Although the covariance expression is complicated, its various components can be used to study the effect of the DDP prior common weights (through the total mass parameter) as well as of the parameters of the centering distribution G_0 for the dependent atoms.

A.5 Additional results for Simulation 3: Comparison with a Linear-DDP Model

We use Simulation Example 3 in Section 3.3 of the main text and compare the DDP-based Erlang mixture model with the linear-DDP (LDDP) model with log-normal kernels in De Iorio, Johnson, Müller & Rosner (2009). The LDDP model assumes

$$\begin{aligned} p(t_i \mid \boldsymbol{x}_i, G) &= \frac{1}{t_i} \int \mathrm{N}(\log(t_i) \mid \boldsymbol{x}_i\boldsymbol{\beta}, \sigma^2) dG(\boldsymbol{\beta}, \sigma^2), \\ G \mid \alpha, G_0 &\sim DP(\alpha, G_0). \end{aligned}$$

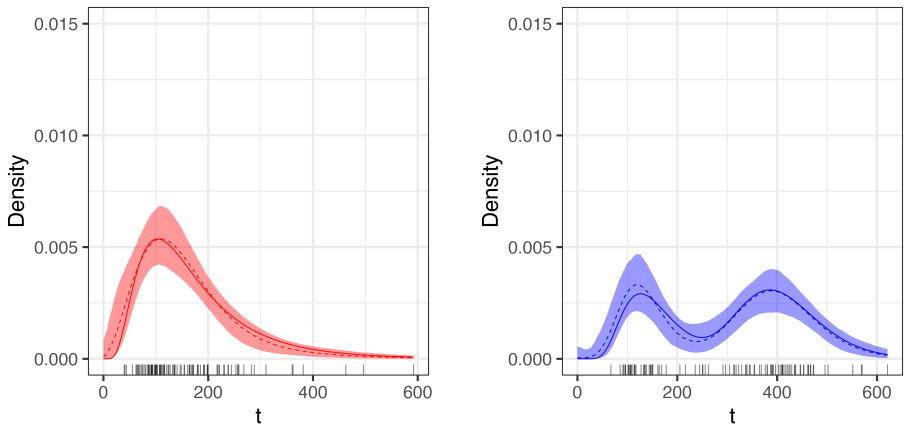
A conditional conjugate base measure is assumed; $G_0 = \mathrm{N}(\boldsymbol{\beta} \mid \mu_b, S_b)\mathrm{Ga}(\sigma^{-2} \mid \tau_1/2, \tau_2/2)$. Conjugate hyperpriors are assumed for μ_b, S_b and τ_2; $\mu_b \sim \mathrm{N}(m_0, S_0)$, $S_b \sim$ inv-Wishart(v, ψ) and $\tau_2 \sim \mathrm{Ga}(\tau_{s1}, \tau_{s2})$. A gamma prior is assumed for α; $\alpha \sim \mathrm{Ga}(a_\alpha, b_\alpha)$. The hyperparameters are specified as follows; $m_0 = (5, 0.5)$, $S_0 = 0.5\mathrm{I}_2$, $v = 4$, $\Psi^{-1} = 2\mathrm{I}_2$, $\tau_1 = 6$, $\tau_{s1} = 6$, $\tau_{s2} = 2$, $a_\alpha = 3$, and $b_\alpha = 2$. We use a R package *DPpackage* in Jara et al. (2011) to fit the LDDP model.

Posterior inference results for survival functionals under the LDDP model are illustrated in Figures A.2 and A.3. Panels (a) and (b) of Figure A.2 show the posterior estimates under the DDP-based Erlang mixture model for easy comparison. From panels (c) and (d), the posterior estimates of the density functions under the LDDP model capture the overall shapes of the true density functions, but the estimate for the control group has misfit for $t < 300$. Figure A.3 shows the estimates

of the survival functions and hazard functions under the LDDP model. Comparing the inferences under the LDDP model to those under the DDP-based Erlang mixture model, the DDP-based Erlang mixture model yields better estimates for all survival functionals with improved accuracy for this simulation example.

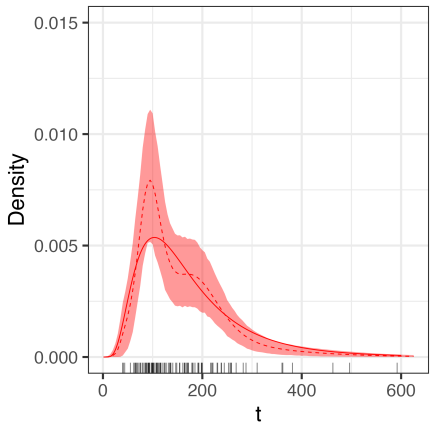
A.6 Additional results for Simulation 3: Sensitivity Analysis

In this section, we show sensitivity analysis for DDP-based Erlang mixture model by varying the values of some fixed hyperparameters using Simulation 3 in Section 3.3 of the main text. We first set $\alpha \sim \text{Ga}(2,1)$, $\theta_x \overset{ind.}{\sim} \text{Ga}(2,10)$, $M_x \mid \theta_x \overset{ind.}{\sim} \text{Unif}(\lceil 1000 \rceil \theta_x, \ldots, \lceil 4000 \rceil \theta_x)$. Then let $\boldsymbol{\mu} \sim \text{N}_2((5.0, 5.5)', 0.5\text{I}_2)$, and $\Sigma \sim$ inv-Wishart$(8, 0.5\text{I}_2)$. Figure A.4 illustrates the posterior inferences. In addition, we assume $\theta_x \overset{ind.}{\sim} \text{Ga}(2,50)$, $\boldsymbol{\mu} \sim \text{N}_2((5.0, 5.5)', 10\text{I}_2)$, and $\Sigma \sim$ inv-Wishart$(8, 50\text{I}_2)$, while keeping the other fixed hyperparameter values the same. The posterior inferences are shown in Figure A.5. Comparison of the inferences in Figures A.4 and A.5 to those in Figure 5 of the main text shows that the posterior inferences are not sensitive to specification of $a_{x\theta}$, $b_{x\theta}$, Σ_0, r and R.

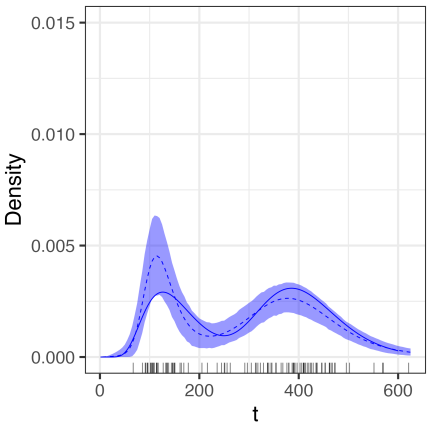


(a) DDP-Erlang (control)

(b) DDP-Erlang (treatment)

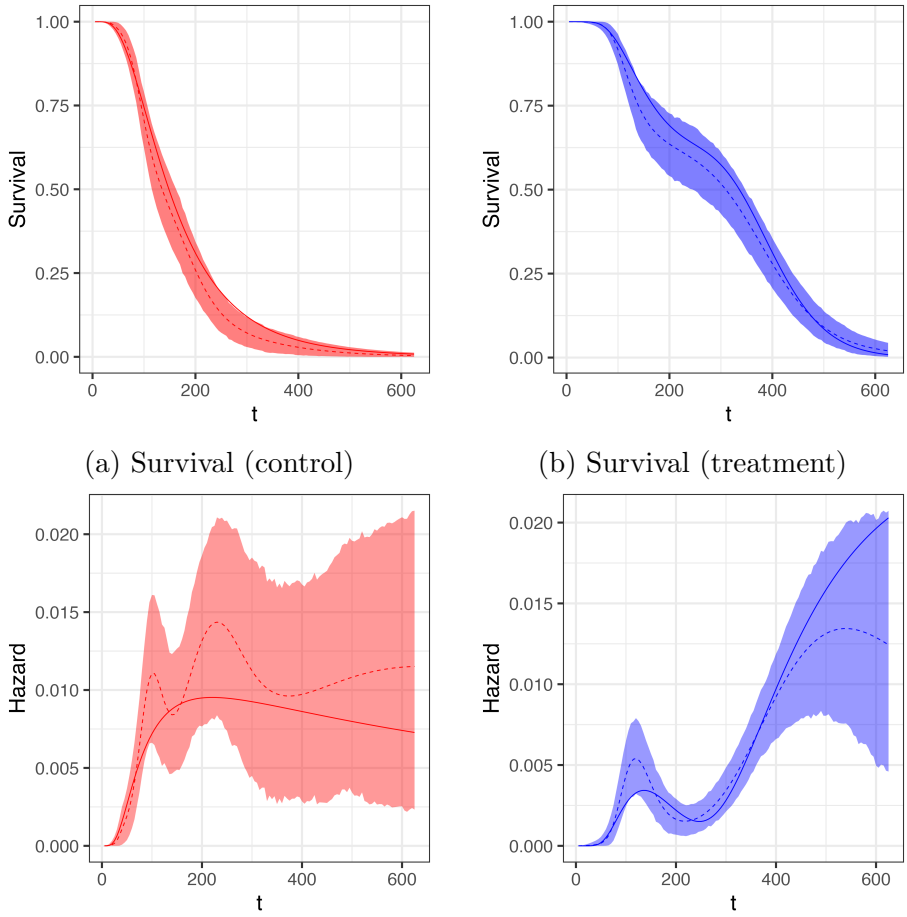


(c) LDDP (control)



(d) LDDP (treatment)

Figure A.2: LDDP model in Simulation Example 3. Each panel plots the density estimates for control (left panels) and treatment (right panels) groups. The results in panels (a) and (b) are estimated by the DDP-based Erlang mixture model, and the results in panels (c) and (d) are estimated by the LDDP model. In each panel, the dashed lines denote the posterior mean estimates, the solid line represents the true underlying function, and the shaded regions indicate the 95% credible intervals. Red and blue colors are used for the control and treatment groups, respectively.



(a) Survival (control)

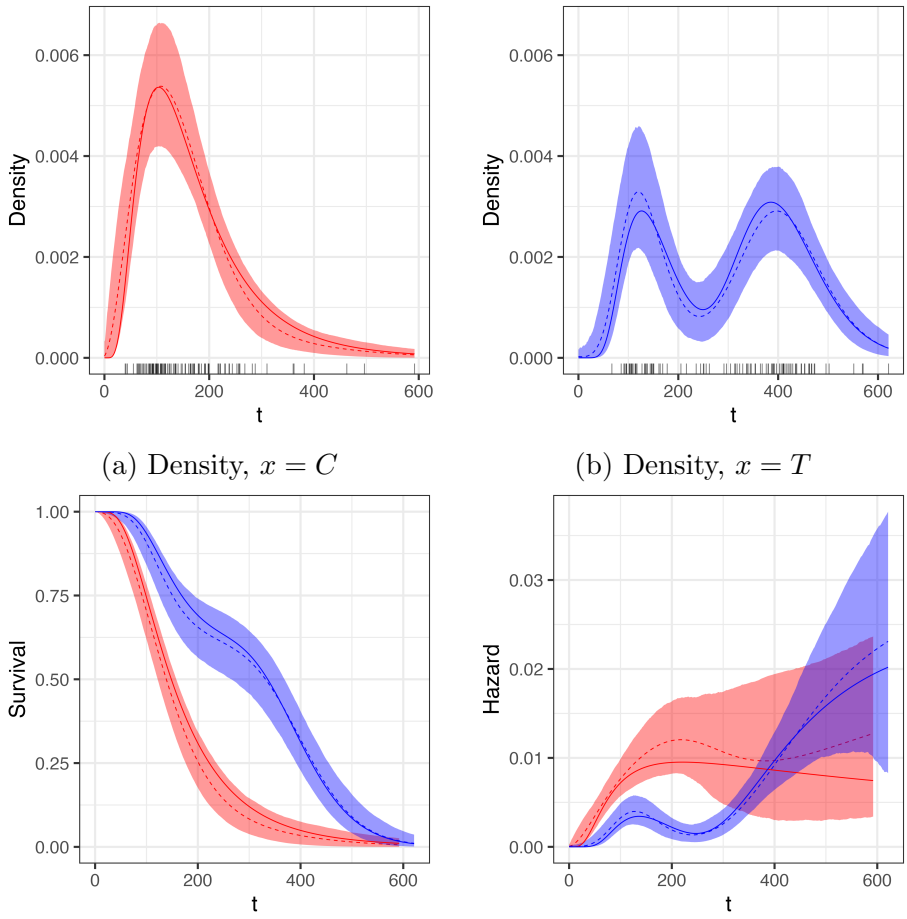
(b) Survival (treatment)

(c) Hazard (control)

(d) Hazard (treatment)

Figure A.3: LDDP model in Simulation Example 3. Panels (a) and (b) plot the estimates for the control and treatment group survival functions, respectively. Panels (c) and (d) plot the hazard function estimates for control and treatment group, respectively. In each panel, the dashed lines denote the posterior mean estimates, the solid line represents the true underlying function, and the shaded regions indicate the 95% credible intervals. Red and blue colors are used for the control and treatment groups, respectively.

!ht



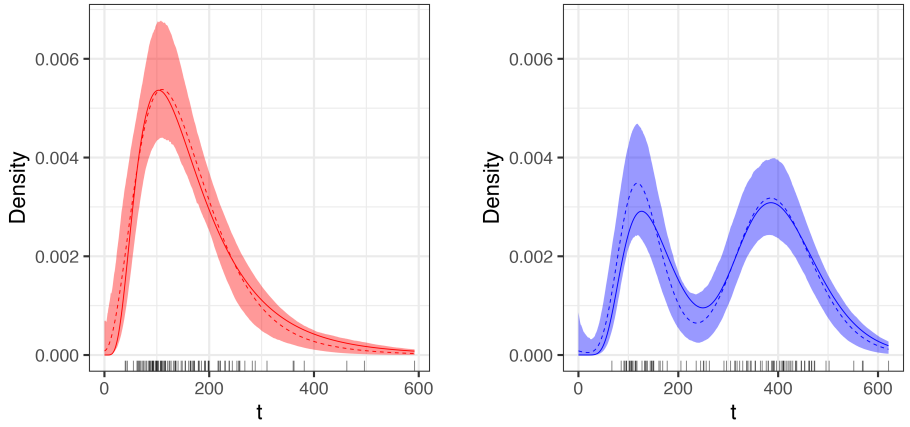
(a) Density, $x = C$

(b) Density, $x = T$

(c) Survival

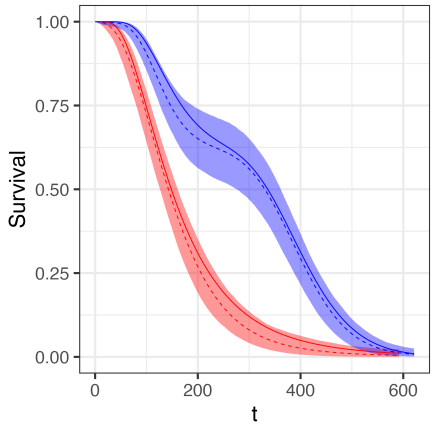
(d) Hazard

Figure A.4: Simulation Example 3 for Sensitivity Analysis - I. $\theta_x \overset{ind.}{\sim} \text{Ga}(2, 10)$, $\boldsymbol{\mu} \sim \text{N}_2((5.0, 5.5)', 0.5\text{I}_2)$, and $\Sigma \sim \text{inv-Wishart}(8, 0.5\text{I}_2)$ are used. Panels (a) and (b) plot the estimates for the control and treatment group density, respectively (the rug plots show the corresponding survival times). Panels (c) and (d) compare the estimates for the survival and hazard function, respectively. In each panel, the dashed lines denote the posterior mean estimates, the solid line the true underlying function, and the shaded regions indicate the 95% credible intervals. Red and blue colors are used for the control and treatment groups, respectively.

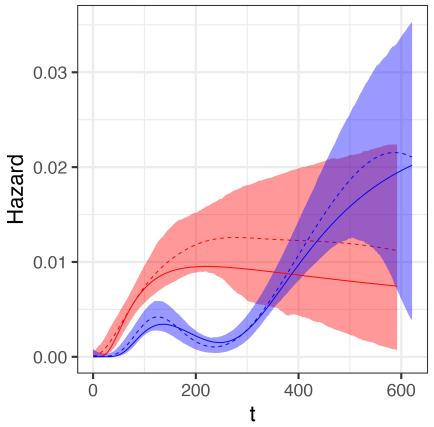


(a) Density, $x = C$

(b) Density, $x = T$



(c) Survival



(d) Hazard

Figure A.5: Simulation Example 3 for Sensitivity Analysis - II. $\theta_x \overset{ind.}{\sim} \text{Ga}(2, 50)$, $\boldsymbol{\mu} \sim \text{N}_2((5.0, 5.5)', 10\text{I}_2)$, and $\Sigma \sim \text{inv-Wishart}(8, 50\text{I}_2)$ are used. Panels (a) and (b) plot the estimates for the control and treatment group density, respectively (the rug plots show the corresponding survival times). Panels (c) and (d) compare the estimates for the survival and hazard function, respectively. In each panel, the dashed lines denote the posterior mean estimates, the solid line the true underlying function, and the shaded regions indicate the 95% credible intervals. Red and blue colors are used for the control and treatment groups, respectively.

Appendix B

Supplementary material to Bayesian Nonparametric Survival Regression Model using Log-logistic kernel

B.1 Proof of lemma 1

To start, the log-logistic density function, $f_{\text{LL}}(t \mid \theta, \phi)$, is given by:

$$f_{\text{LL}}(t \mid \theta, \phi) = \frac{\phi}{t} \frac{\exp\{\phi(\log(t) - \log(\theta))\}}{[1 + \exp\{\phi(\log(t) - \log(\theta))\}]^2},$$

which can be equivalently expressed in terms of the Polya-Gamma density, $f_{\text{PG}}(\cdot \mid a, b)$, as per equation (7) from Polson et al. (2013), yielding:

$$f_{\text{LL}}(t \mid \theta, \phi) = \frac{\phi}{4t} \int_0^\infty \exp\left\{-\frac{(\log(t) - \log(\theta))^2}{2(u\phi^2)^{-1}}\right\} f_{\text{PG}}(u \mid 2, 0) du.$$

In this expression, we identify the kernel of the log-normal distribution within the integral. By further completing the log-normal density representation, we obtain:

$$f_{\text{LL}}(t \mid \theta, \phi) = \frac{\sqrt{2\pi}}{4} \int_0^\infty f_{\text{LN}}(t \mid \log(\theta), (u\phi^2)^{-1}) u^{-1/2} f_{\text{PG}}(u \mid 2, 0) du.$$

Moreover, the density representation of $f_{\text{PG}}(u \mid 2, 0)$ is given by:

$$f_{\text{PG}}(u \mid 2, 0) = \sum_{n=1}^{\infty} (-1)^n \frac{4(n+1)^2}{\sqrt{2\pi u^3}} \exp\left\{ -\frac{(n+1)^2}{2u} \right\}.$$

which leads the density representation:

$$f_{\text{LL}}(t \mid \theta, \phi) = \int_0^\infty f_{\text{LN}}(t \mid \log(\theta), (u\phi^2)^{-1}) \sum_{n=1}^{\infty} (-1)^n \frac{(n+1)^2}{u^2} \exp\left\{ -\frac{(n+1)^2}{2u} \right\} du.$$

Finally, recognizing an inverse Gamma kernel within the summation, we obtain:

$$f_{\text{LL}}(t \mid \theta, \phi) = \int_0^\infty f_{\text{LN}}(t \mid \log(\theta), (u\phi^2)^{-1}) \sum_{n=1}^{\infty} 2(-1)^n f_{\text{inv-Ga}}\left(u \mid 1, (n+1)^2/2\right) du.$$

B.2 MCMC details - DPM-LL

We present here the posterior simulation details for the DPM-LL model. Considering configuration variable L_i for each subject, where L_i is a membership indicator.

The augmented model is written as

$$
\begin{aligned}
t_i \mid L_i, \boldsymbol{\theta^*}, \boldsymbol{\phi^*} &\overset{ind.}{\sim} \mathrm{LL}(\theta^*_{L_i}, \phi^*_{L_i}) \text{ for } i = 1, \ldots, n, \\
P(L_i = l \mid L_1, \ldots, L_{i-1}, \alpha) &= \begin{cases} \frac{n_{i,l}}{\alpha+i-1} \text{ if } \exists L_j = l \text{ for } j < i \\ \frac{\alpha}{\alpha+i-1} \text{ otherwise} \end{cases}, \\
\log(\theta^*_l) \mid \mu_\theta, \sigma^2_\theta &\overset{i.i.d.}{\sim} \mathrm{N}(\mu_\theta, \sigma^2_\theta) \text{ for } l = 1, \ldots, k, \\
\phi^2_l \mid b_\phi &\overset{i.i.d.}{\sim} \mathrm{Ga}(a_\phi, b_\phi) \text{ for } l = 1, \ldots, k, \\
\mu_\theta &\sim \mathrm{N}(s_\theta, S_\theta), \\
b_\phi &\sim \text{inv-Ga}(r_\phi, R_\phi), \\
\alpha &\sim \mathrm{Ga}(a_\alpha, b_\alpha),
\end{aligned}
$$

where $n_{i,l}$ represents the number of elements in the set $\{L_j = l : j < i\}$. The likelihood function for the augmented model for all observations is

$$
L(\boldsymbol{\theta^*}, \boldsymbol{\phi^*}, \boldsymbol{L}; \mathcal{D}) = \prod_{i=1}^{n} \{f_{\mathrm{LL}}(y_i \mid \theta^*_{L_i}, \phi^*_{L_i})\}^{\nu_i} \{S_{\mathrm{LL}}(y_i \mid \theta^*_{L_i}, \phi^*_{L_i})\}^{1-\nu_i},
$$

where $\mathcal{D} = \{(t_i, \nu_i) : i = 1, \ldots, n\}$ denotes data. The detailed posterior computation algorithm based on the Algorithm 8 in Neal (2000) is presented as follows.

Updating membership indicators The algorithm keeps track of four sets of variables, cluster membership indicators $\boldsymbol{L}$, number of clusters k, cluster sizes $(n_1, \ldots, n_k)$, and distinct mixing parameters (θ^*_l, ϕ^*_l) for $l = 1, \ldots, k$.

For i in $1, \ldots .n$, each cluster membership indicator L_i is updated iteratively depending on all other parameters. Let superscript '-' denote notations after removing L_i. Assume $L_i = l'$,

- if cluster l' is not a singleton in $\boldsymbol{L}$, let $\boldsymbol{L}_{-i} = (L_1^-, \ldots, L_{i-1}^-, L_{i+1}^-, \ldots, L_n^-)$ where $L_j^- = L_j$ for $j \neq i$ and L_i^- is left blank. And let $k^- = k$. For $l = 1, \ldots, k^-$, let $n_l^- = \begin{cases} n_l, & \text{if } l \neq l' \\ n_l - 1, & \text{if } l = l' \end{cases}$ and let $(\theta_l^{*-}, \phi_l^{*-}) = (\theta_l^*, \phi_l^*)$.
- If cluster l' is a singleton, i.e. $n_{l'} = 1$, let $k^- = k - 1$, then we let $L_j^- = \begin{cases} L_j, & \text{if } L_j < l' \\ L_j - 1, & \text{if } L_j > l' \end{cases}$ for $j \neq i$. Moreover, for $l = 1, \ldots, k^-$, if $l < l'$, let $n_l^- = n_l$ and $(\theta_l^{*-}, \phi_l^{*-}) = (\theta_l^*, \phi_l^*)$; otherwise, $n_l^- = n_{l+1}$ and $(\theta_l^{*-}, \phi_l^{*-}) = (\theta_{l+1}^*, \phi_{l+1}^*)$.

Then draw m auxiliary variables from the baseline distribution function independently, $(\theta_l^{*-}, \phi_l^{*-}) \overset{i.i.d.}{\sim} G_0$, for $l = k^- + 1, \ldots, k^- + m$. A new value for L_i is drawn from $\{1, \ldots, k^- + m\}$ with the following probabilities,

$$P(L_i = l \mid \boldsymbol{L_{-i}}, -) \propto \begin{cases} n_l^- \left\{ f_{\text{LL}}(y_i \mid \theta_l^{*-}, \phi_l^{*-}) \right\}^{\nu_i} \left\{ S_{\text{LL}}(y_i \mid \theta_l^{*-}, \phi_l^{*-}) \right\}^{1-\nu_i} & \text{for } 1 \leq l \leq k^- \\ \frac{\alpha}{m} \left\{ f_{\text{LL}}(y_i \mid \theta_l^{*-}, \phi_l^{*-}) \right\}^{\nu_i} \left\{ S_{\text{LL}}(y_i \mid \theta_l^{*-}, \phi_l^{*-}) \right\}^{1-\nu_i} & \text{for } k^- + 1 \leq l \leq k^- + m \end{cases}.$$

- If L_i is joining in an existing cluster, $L_i \leq k^-$, then let $k = k^-$, and for $l = 1, \ldots, k$, let $(\theta_l^*, \phi_l^*) = (\theta_l^{*-}, \phi_l^{*-})$, and $n_l = \begin{cases} n_l^-, & \text{if } l \neq L_i \\ n_l^- + 1, & \text{o.w.} \end{cases}$, and for $j = 1, \ldots, n$, let $L_j = L_j^-$ for $j \neq i$.
- If L_i is a new cluster, $L_i \geq k^- + 1$, then let $k = k^- + 1$, for $l = 1, \ldots, k-1$, let $(\theta_l^*, \phi_l^*) = (\theta_l^{*-}, \phi_l^{*-})$, $n_l = n_l^-$, for $j = 1, \ldots, n$, let $L_j = L_j^-$ for $j \neq i$. Finally, let $n_k = 1$, $(\theta_k^*, \phi_k^*) = (\theta_{L_i}^{*-}, \phi_{L_i}^{*-})$, and $L_i = k$.

Updating distinct mixing parameters Given all other parameters, the posterior full conditional of (θ_l^*, ϕ_l^*) is upto a proportion,

$$p((\theta_l^*, \phi_l^*) \mid -) \propto N(\log(\theta_l^*) \mid \mu_\theta, \sigma_\theta^2)\mathrm{Ga}(\phi_l^{*2} \mid a_\phi, b_\phi) \prod_{i:L_i=l}^{n} \{f_{\mathrm{LL}}(y_i \mid \theta_l^*, \phi_l^*)\}^{\nu_i} \{S_{\mathrm{LL}}(y_i \mid \theta_l^*, \phi_l^*)\}^{1-\nu_i}.$$

A efficient sampling algorithm can be derived by using a Pólya-Gamma (PG) data augmentation technique introduced by Polson et al. (2013). We reorganize the above expression can be rewritten in the following way

$$\begin{aligned} p((\theta_l^*, \phi_l^*) \mid -) &\propto N(\log(\theta_l^*)|\mu_\theta,\sigma_\theta^2)\mathrm{Ga}(\phi_l^{*2}|a_\phi,b_\phi) \prod_{i:L_i=l}^{n} \left(\frac{\phi_l^*}{t_i}\right)^{\nu_i} \frac{[\exp\{\phi_l^*(\log(t_i)-\log(\theta_l^*))\}]^{\nu_i}}{\left(1+\exp(\phi_l^*(\log(t_i)-\log(\theta_l^*)))\right)^{\nu_i+1}} \\ &\propto N(\log(\theta_l^*) \mid \mu_\beta, \sigma_\beta^2)\mathrm{Ga}(\phi_l^{*2} \mid a_\phi, b_\phi) \\ &\quad \prod_{i:L_i=l}^{n} \left(\frac{\phi_l^*}{t_i}\right)^{\nu_i} \int \exp\left(0.5(\nu_i-1)\psi_{il}\right) \exp(-0.5\epsilon_i\psi_{il}^2) p(\epsilon_i) d\epsilon_i, \end{aligned}$$

where $\psi_{il} = \phi_l^*(\log(y_i) - \log(\theta_l^*))$ and $p(\epsilon_i)$ is a density function of $\mathrm{PG}(1+\nu_i, 0)$ evaluated at ϵ_i. By introducing latent variables $(\epsilon_1, \ldots, \epsilon_n)$ for each observation, we first sample latent variables $(\epsilon_1, \ldots, \epsilon_n)$ from $\mathrm{PG}(1+\nu_i, \phi_{L_i}^*(\log(y_i) - \log(\theta_{L_i})))$ independently. Then conditional on $(\epsilon_1, \ldots, \epsilon_n)$, draw (θ_l^*, ϕ_l^*), for $l = 1, \ldots, k$ from

$$\log(\theta_l^*) \mid \boldsymbol{\epsilon}, - \quad \sim \quad \mathrm{N}(\tilde{\mu}_{\theta l}, \tilde{\sigma}_{\theta l}^2),$$

$$\text{where } \tilde{\sigma}_{\theta l}^2 = \left(\frac{1}{\sigma_{\tilde{\theta}}^2} + \phi_l^{*2} \sum_{i:L_i=l} \epsilon_i\right)^{-1}, \quad \text{and} \quad 5\tilde{\mu}_{\theta l} = \tilde{\sigma}_{\theta l}^2 \left(\frac{\mu_\theta}{\sigma_{\tilde{\theta}}^2} + \phi_l^{*2} \sum_{i:L_i=l} \epsilon_i \log(y_i) + 0.5\phi_l^* \sum_{i:L_i=l} (1-\nu_i)\right),$$

and

$$
\begin{aligned}
p(\phi_l^* \mid -) &\propto \text{Ga}(\tilde{a}_\phi, \tilde{b}_\phi) \times A, \\
\text{where } \tilde{a}_\phi &= 0.5 \sum_{i:L_i=l} \nu_i + a_\phi, \tilde{b}_\phi = \left(0.5 \sum_{i:L_i=l} \epsilon_i (\log(y_i) - \log(\theta_l^*))^2 + \frac{1}{b_\phi} \right)^{-1}, \\
\text{and } A &= \exp\left(0.5 \sum_{i:L_i=l} (\nu_i - 1) \phi_l^* (\log(y_i) - \log(\theta_l^*)) \right).
\end{aligned}
$$

If $\nu_i = 1$ for all i such that $L_i = l$, draw ϕ_l^* from $\text{Ga}(\tilde{a}_\phi, \tilde{b}_\phi)$ directly, otherwise, a Metropolis Random-Walk algorithm can be applied.

Updating hyperparameters

$$
\begin{aligned}
p(b_\phi, \mu_\theta, \sigma_\theta^2 \mid -) \propto\ & \text{N}(\mu_\theta \mid s_\theta, S_\theta) \text{inv-Ga}(\sigma_\theta^2 \mid c_\theta, C_\theta) \text{inv-Ga}(b_\phi \mid r_\phi, R_\phi) \\
& \prod_{l=1}^{k} \text{N}(\log(\theta_l^*) \mid \mu_\theta, \sigma_\theta^2) \text{Ga}(\phi_l^{*2} \mid a_\phi, b_\phi).
\end{aligned}
$$

We thus have

$$
\begin{aligned}
\mu_\theta \mid - &\sim \text{N}\left(\left(\frac{1}{S_\theta} + \frac{k}{\sigma_\theta^2} \right)^{-1} \left(\frac{s_\theta}{S_\theta} + \frac{\sum_{l=1}^k \log(\theta_l^*)}{\sigma_\theta^2} \right), \left(\frac{1}{S_\theta} + \frac{k}{\sigma_\theta^2} \right)^{-1} \right), \\
\sigma_\theta^2 \mid - &\sim \text{inv-Ga}\left(c_\theta + \frac{k}{2}, C_\theta + \frac{1}{2} \sum_{l=1}^{k} (\log(\theta_l^*) - \mu_\theta)^2 \right), \\
\text{and } b_\phi \mid - &\sim \text{inv-Ga}\left(r_\phi + k a_\phi, R_\phi + \sum_{l=1}^{k} \phi_l^{*2} \right).
\end{aligned}
$$

Updating DP total mass parameter The full conditional of α is only related to its prior, the number of observations, and the number of distinct clusters. Upto a proportionality, it is written as

$$
p(\alpha \mid -) \propto \text{Ga}(\alpha \mid a_\alpha, b_\alpha) \alpha^k \Gamma(\alpha) / \Gamma(\alpha + n).
$$

Using the data augmentation method proposed by Escobar & West (1995), we could update it in the following way,

$$\begin{aligned}
\eta|\alpha,- &\sim \text{Be}(\alpha+1,n)\\
\alpha|\eta,k,- &\sim \frac{a_\alpha+k-1}{n(\frac{1}{b_\alpha}-\log(\eta))+a_\alpha+k-1}\text{Ga}(a_\alpha+k,\frac{1}{\frac{1}{b_\alpha}-\log(\eta)})\\
&\quad+\frac{n(\frac{1}{b_\alpha}-\log(\eta))}{n(\frac{1}{b_\alpha}-\log(\eta))+a_\alpha+k-1}\text{Ga}(a_\alpha+k-1,\frac{1}{\frac{1}{b_\alpha}-\log(\eta)}).
\end{aligned}$$

Posterior functional estimation To evaluate different functional estimations such as density and survival functions, we use a truncation approximation through the following representation (Pitman 1996),

$$G \mid (k,\boldsymbol{L},\boldsymbol{\theta^*},\boldsymbol{\phi^*}),\alpha,(\mu_\theta,\sigma_\theta^2,b_\phi) \stackrel{\mathcal{D}}{=} q_{k+1}G^* + \sum_{l=1}^{k} q_l \delta_{(\theta_l^*,\phi_l^*)},$$

where $G^* \sim DP(\alpha, G_0(\mu_\theta,\sigma_\theta^2,b_\phi))$ and $(q_1,\ldots,q_k,q_{k+1}) \sim \text{Dir}(n_1,\ldots,n_k,\alpha)$. Using DP's constructive definition (Sethuraman 1994) for G^*, we obtain

$$G^* = \sum_{l=1}^{\infty} p_l \delta_{(\tilde{\theta}_l,\tilde{\phi}_l)}(\cdot),$$

where $(\tilde{\theta}_l,\tilde{\phi}_l) \overset{i.i.d.}{\sim} G_0(\mu_\theta,\sigma_\theta^2,b_\phi)$, $v_l \overset{i.i.d.}{\sim} \text{Be}(1,\alpha)$, and $p_1 = v_1, p_l = v_l\prod_{r=1}^{l-1}(1-v_l)$ for $l>1$. With an appropriate truncation level N, we can obtain an approximated G^*, through

$$G_N^* = \sum_{l=1}^{N} p_l \delta_{(\tilde{\theta}_l,\tilde{\phi}_l)}(\cdot),$$

where $p_N = 1-\sum_{l=1}^{N-1} p_l$. The approximation level is related to both N and α, for a fixed α, $E(\sum_{l=1}^{N} p_l \mid \alpha) = 1-\left(\frac{\alpha}{\alpha+1}\right)^N$. If $\alpha=1$, the approximation level is then 0.999

for $N = 10$. Then the posterior G is approximated by

$$q_{k+1} \sum_{l=1}^{N} p_l \delta_{(\tilde{\theta}_l, \tilde{\phi}_l)}(\cdot) + \sum_{r=1}^{k} q_r \delta_{(\theta_r^*, \phi_r^*)}(\cdot).$$

In each MCMC iteration b, we can obtain a set of parameters $\Theta^{(b)} = (\alpha^{(b)}, \boldsymbol{\theta}^{*(b)}, \boldsymbol{\phi}^{*(b)}, b_\phi^{(b)}, \mu_\theta^{(b)}, \sigma_\theta^{2(b)}, \boldsymbol{L}^{(b)}, k^{(b)}, \boldsymbol{n}^{(b)})$. Based on $\Theta^{(b)}$, we first draw a weight vector $(q_1^{(b)}, \ldots, q_{k+1}^{(b)}) \sim \text{Dir}(n_1^{(b)}, \ldots, n_k^{(b)}, \alpha^{(b)})$, and then obtain an approximated $G_N^{*(b)} = \{(p_l^{(b)}, \tilde{\theta}_l^{(b)}, \tilde{\phi}_l^{(b)}) : l = 1, \ldots, N\}$, where $(\tilde{\theta}_l^{(b)}, \tilde{\phi}_l^{(b)}) \overset{i.i.d.}{\sim} G_0((\theta^*, \phi^*) \mid b_\phi^{(b)}, \mu_\theta^{(b)}, \sigma_\theta^{2(b)})$ and $p_1^{(b)} = v_1^{(b)}$, $p_l^{(b)} = v_l^{(b)} \prod_{r=1}^{l-2}(1 - v_l^{(b)})$ for $l > 1$ with $v_l^{(b)} \overset{i.i.d.}{\sim} \text{Be}(1, \alpha)$ for $l < N$, and $p_N^{(b)} = 1 - \sum_{l=1}^{N-1} p_l^{(b)}$. Evatually, density and survival functions can be evaluated in the following way,

$$\begin{aligned} f(\cdot \mid \Theta^{(b)}) &\approx q_{k+1}^{(b)} \sum_{l=1}^{N} p_l^{(b)} f_{\text{LL}}(\cdot \mid \tilde{\theta}_l^{(b)}, \tilde{\phi}_l^{(b)}) + \sum_{r=1}^{k} q_r^{(b)} f_{\text{LL}}(\cdot \mid \theta_r^{*(b)}, \phi_r^{*(b)}) \\ \text{and } S(\cdot \mid \Theta^{(b)}) &\approx q_{k+1}^{(b)} \sum_{l=1}^{N} p_l^{(b)} S_{\text{LL}}(\cdot \mid \tilde{\theta}_l^{(b)}, \tilde{\phi}_l^{(b)}) + \sum_{r=1}^{k} q_r^{(b)} S_{\text{LL}}(\cdot \mid \theta_r^{*(b)}, \phi_r^{*(b)}). \end{aligned}$$

Thus the posterior functionals can be evaluated on a grid of t.

B.3 MCMC details - Density Regression model

Here, we illustrate a generic method, although the specific computational details may vary across different scenarios, where different types of covariates are included. The algorithm is a straightforward extension of the algorithm described in Appendix B.

The augmented model is written as

$$
\begin{aligned}
t_i \mid L_i, x_i, \boldsymbol{\theta^*}, \boldsymbol{\beta^*}, \boldsymbol{\phi^*} &\overset{ind.}{\sim} \text{LL}(\theta^*_{L_i} \exp(\boldsymbol{x}_i' \boldsymbol{\beta}^*_{L_i}), \phi^*_{L_i}), \\
\boldsymbol{x} \mid L_i, \boldsymbol{\Omega^*} &\overset{ind.}{\sim} f(\boldsymbol{x} \mid \boldsymbol{\Omega}^*_{L_i}), \\
P(L_1, \ldots, L_n \mid \alpha) &= \alpha^k \frac{\prod_{l=1}^k \Gamma(n_l)}{\prod_{i=1}^n (\alpha + i - 1)}, \\
\log(\theta^*_l) \mid \mu^*_\theta, \sigma^2_\theta &\overset{i.i.d.}{\sim} \text{N}(\mu_\theta, \sigma^2_\theta), \\
\phi^{*2}_l \mid b_\phi &\overset{i.i.d.}{\sim} \text{inv-Ga}(a_\phi, b_\phi), \\
\boldsymbol{\beta}^*_l \mid \mu_\beta, \sigma^2_\beta &\overset{i.i.d.}{\sim} \text{N}(\mu_\beta \mathbf{1}_p, \sigma^2_\beta I_p), \\
\boldsymbol{\Omega}^*_l &\overset{i.i.d.}{\sim} \prod_{j=1}^p G_0^{(j)}(\boldsymbol{\Omega}^*_{l,j}), \\
\alpha &\sim \text{Ga}(a_\alpha, b_\alpha).
\end{aligned}
$$

For the rest parameters, we assume the following hyper priors, $b_\phi \sim \text{inv-Ga}(r_\phi, R_\phi)$, $\mu_\theta \sim \text{N}(s_\theta, S_\theta)$, $\sigma^2_\theta \sim \text{inv-Ga}(c_\theta, C_\theta)$, $\mu_\beta \sim \text{N}(s_\beta, S_\beta)$, and $\sigma^2_\beta \sim \text{inv-Ga}(c_\beta, C_\beta)$. In subsequent sections, we will outline the updating algorithm for each parameter.

Update distinct mixing parameter For each $l \in \{1, \ldots, k\}$, the full conditional of $(\theta^*_l, \beta^*_l, \phi^*_l, \boldsymbol{\Omega}^*_l)$ upto a proportionality is

$$
\begin{aligned}
p(\theta^*_l, \beta^*_l, \phi^*_l, \boldsymbol{\Omega}^*_l) \propto\ & \text{N}(\log(\theta^*_l) \mid \mu_\theta, \sigma^2_\theta) \text{Ga}(\phi^{*2}_l \mid a_\phi, b_\phi) \text{N}(\beta^*_l \mid \mu_\beta \mathbf{1}_p, \sigma^2_\beta I_p) g_0(\boldsymbol{\Omega}^*_l) \\
& \prod_{i: L_i = l} \left(f_{\text{LL}}(y_i \mid \theta^*_l \exp(\boldsymbol{x}' \boldsymbol{\beta}^*_l), \phi^*_l) \right)^{\nu_i} \left(S_{\text{LL}}(y_i \mid \theta^*_l \exp(\boldsymbol{x}_i \boldsymbol{\beta}^*_l), \phi^*_l) \right)^{1 - \nu_i}.
\end{aligned}
$$

Using the Pólya-Gamma data augmentation trick the same way as in previous sections, we design the following sampling mechanisam,

- (a) sample $\boldsymbol{\Omega}^*_l$ for $l = 1, \ldots, k$ from its posterior full conditional, the actual sam-

pling function depends on the choice of the kernel function and its prior base measure.

- (b) Sample Pólya-Gamma latent variables,

$$\epsilon_i \mid - \overset{ind.}{\sim} \text{PG}(1+\nu_i, \phi^*_{L_i}(\log(y_i) - \log(\theta^*_{L_i}) - \boldsymbol{x}_i'\boldsymbol{\beta}^*_{L_i})).$$

- (c) Sample $\log(\theta^*_l)$ from $\text{N}(\tilde{\mu}_{\theta l}, \tilde{\sigma}^2_{\theta l})$, where

$$\tilde{\sigma}^2_{\theta l} = \left(\frac{1}{\sigma^2_\theta} + \phi^*_l \sum_{i:L_i=l} \epsilon_i \right)^{-1}$$

,

$$\tilde{\mu}_{\theta l} = \tilde{\sigma}^2_{\theta l} \left(\frac{\mu_\theta}{\sigma^2_\theta} + \phi^{*2}_l \sum_{i:L_i=l} \epsilon_i (\log(y_i) - \boldsymbol{x}_i'\boldsymbol{\beta}^*_l) + 0.5\phi^*_l \sum_{i:L_i=l} (1-\nu_i) \right).$$

- (d) Sample $\boldsymbol{\beta}^*_l$ from $\text{N}(\tilde{\mu}_{\beta l}, \tilde{\Sigma}_{\beta l})$, where

$$\tilde{\Sigma}_{\beta l} = \left(\frac{1}{\sigma^2_\beta} I_2 + \phi^{*2}_l \sum_{i:L_i=l} \epsilon_i \boldsymbol{x}_i \boldsymbol{x}_i' \right)^{-1}$$

and

$$\tilde{\mu}_{\beta l} = \tilde{\Sigma}_{\beta l} \left(\frac{\mu_\beta}{\sigma^2_\beta} \mathbf{1}_2 + \phi^{*2}_l \sum_{i:L_i=l} \epsilon_i (\log(y_i) - \log(\theta^*_l)) \boldsymbol{x}_i + 0.5\phi^*_l \sum_{i:L_i=l} (1-\nu_i) \boldsymbol{x}_i \right).$$

- (e) For $l = 1, \ldots, k$, the full conditional for ϕ^*_l is proportional to $\text{Ga}(\tilde{a}_\phi, \tilde{b}_\phi) \times A$, where

$$\tilde{a}_\phi = 0.5 \sum_{i:L_i=l} \nu_i + a_\phi, \; \tilde{b}_\phi = \left(0.5 \sum_{i:L_i=l} \epsilon_i (\log(y_i) - \log(\theta^*_l) - \boldsymbol{x}_i'\boldsymbol{\beta}^*_l)^2 + \frac{1}{b_\phi} \right)^{-1},$$

and

$$A = \exp\left(0.5 \sum_{i:L_i=l} (\nu_i - 1) \phi^*_l (\log(y_i) - \log(\theta^*_l) - \boldsymbol{x}_i'\boldsymbol{\beta}^*_l) \right).$$

Updating membership indicators The updating precedure is adpated from the previous algorithm directly. We keep tracking membership indicators $\boldsymbol{L}$, number of clusters k, cluster sizes $(n_1, \ldots, n_k)$, and all distinct mixing parameters $\Psi_l^* = (\theta_l^*, \theta_l^*, \boldsymbol{\Omega}_l^*)$ for $l = 1, \ldots, k$. Recall that the superscript '-' denotes notations with L_i removed. For $i = 1, \ldots, n$, assume $L_i = l'$,

- if cluster l' is not a singleton in $\boldsymbol{L}$, let $k^- = k$, and $\boldsymbol{L}^- = (L_1^-, \ldots, L_n^-)$, where $L_j^- = L_j$ for $j \neq i$ and L_i^- is left blank. For $l = 1, \ldots, k^-$, let $n_l^- = \begin{cases} n_l, & \text{if } l \neq l' \\ n_l - 1, & \text{if } l = l' \end{cases}$ and let $\Psi_l^{*-} = \Psi_l^*$.
- If cluster l' is a singleton, i.e. $n_{l'} = 1$, $k^- = k-1$, and let $L_j^- = \begin{cases} L_j, & \text{if } L_j < l' \\ L_j - 1, & \text{if } L_j > l' \end{cases}$ for $j \neq i$. Moreover, for $l = 1, \ldots, k^-$, if $l < l'$, let $n_l^- = n_l$ and $\Psi_l^{*-} = \Psi_l^*$; otherwise, $n_l^- = n_{l+1}$ and $\Psi_l^{*-} = \Psi_{l+1}^*$.

Then draw m auxiliary variables from the baseline distribution function independently, $\Psi_l^{*-} \overset{i.i.d.}{\sim} G_0$, for $l = k^- + 1, \ldots, k^- + m$. A new value for L_i is drawn from $\{1, \ldots, k^- + m\}$ with the following probabilities, $P(L_i = l \mid \boldsymbol{L_{-i}}, -) \propto$

- $n_l^- \big\{ f_{\mathrm{LL}}(y_i \mid \theta_l^{*-}, \phi_l^{*-}) \big\}^{\nu_i} \big\{ S_{\mathrm{LL}}(y_i \mid \theta_l^{*-}, \phi_l^{*-}) \big\}^{1-\nu_i} f(\boldsymbol{x}_i \mid \boldsymbol{\Omega}_l^{*-})$ for $1 \leq l \leq k^-$,
- $\alpha/m \big\{ f_{\mathrm{LL}}(y_i \mid \theta_l^{*-}, \phi_l^{*-}) \big\}^{\nu_i} \big\{ S_{\mathrm{LL}}(y_i \mid \theta_l^{*-}, \phi_l^{*-}) \big\}^{1-\nu_i} f(\boldsymbol{x}_i \mid \boldsymbol{\Omega}_l^*)$ for $k^- + 1 \leq l \leq k^- + m$,

where $f(\boldsymbol{x}_i \mid \boldsymbol{\Omega}_l^*)$ is the joint distribution of covariates.

- If L_i is joining in an existing cluster, $L_i \leq k^-$, then let $k = k^-$, and for $l = 1, \ldots, k$, let $(\theta_l^*, \phi_l^*, \boldsymbol{\Omega}_l^*) = (\theta_l^{*-}, \phi_l^{*-}, \boldsymbol{\Omega}^{*-})$, and $n_l = \begin{cases} n_l^-, & \text{if } l \neq L_i \\ n_l^- + 1, & \text{o.w.} \end{cases}$, and for $j = 1, \ldots, n$, let $L_j = L_j^-$ for $j \neq i$.

- If L_i is a new cluster, $L_i \geq k^- + 1$, then let $k = k^- + 1$, for $l = 1, \ldots, k-1$, let $\Psi_l^* = \Psi_l^{*-}$, $n_l = n_l^-$, for $j = 1, \ldots, n$, let $L_j = L_j^-$ for $j \neq i$. Finally, let $n_k = 1$, $\Psi_k^* = \Psi_{L_i}^{*-}$, and $L_i = k$.

Updating hyperparameters

$$\begin{aligned} p(b_\phi, \mu_\theta, \sigma_\theta^2 \mid -) \quad \propto \quad & \mathrm{N}(\mu_\theta \mid s_\theta, S_\theta)\text{inv-Ga}(\sigma_\theta^2 \mid c_\theta, C_\theta)\text{inv-Ga}(b_\phi \mid r_\phi, R_\phi) \\ & \prod_{l=1}^{k} \mathrm{N}(\log(\theta_l^*) \mid \mu_\theta, \sigma_\theta^2)\mathrm{Ga}(\phi_l^{*2} \mid a_\phi, b_\phi) \end{aligned}$$

We thus have

$$\begin{aligned} \mu_\theta \mid - \quad &\sim \quad \mathrm{N}\left(\left(\frac{1}{S_\theta} + \frac{k}{\sigma_\theta^2}\right)^{-1}\left(\frac{s_\theta}{S_\theta} + \frac{\sum_{l=1}^k \log(\theta_l^*)}{\sigma_\theta^2}\right), \left(\frac{1}{S_\theta} + \frac{k}{\sigma_\theta^2}\right)^{-1}\right), \\ \mu_\beta \mid - \quad &\sim \quad \mathrm{N}\left(\left(\frac{1}{S_\beta} + \frac{2k}{\sigma_\beta^2}\right)^{-1}\left(\frac{s_\beta}{S_\beta} + \frac{\sum_{l=1}^k \sum_{j=1}^p \beta_{jl}^*}{\sigma_\beta^2}\right), \left(\frac{1}{S_\beta} + \frac{2k}{\sigma_\beta^2}\right)^{-1}\right), \\ \text{and } b_\phi \mid - \quad &\sim \quad \text{inv-Ga}\left(r_\phi + ka_\phi, R_\phi + \sum_{l=1}^k \phi_l^{*2}\right) \end{aligned}$$

Functional estimations It is the same method obtaining the posterior mixing distribution G,

$$G \mid k, \boldsymbol{L}, \boldsymbol{\Psi^*}, \alpha, (\mu_\theta, \sigma_\theta^2, b_\phi) \stackrel{\mathcal{D}}{=} q_{k+1} G^* + \sum_{l=1}^{k} q_l \delta_{(\Psi_l^*)},$$

where $G^* \sim DP(\alpha, G_0(\mu_\theta, \sigma_\theta^2, b_\phi))$ and $(q_1, \ldots, q_k, q_{k+1}) \sim \text{Dir}(n_1, \ldots, n_k, \alpha)$. Using DP's constructive definition (Sethuraman 1994) for G^*, we obtain

$$G^* = \sum_{l=1}^{\infty} p_l \delta_{\tilde{\Psi}_l}(\cdot),$$

where $\tilde{\Psi}_l \overset{i.i.d.}{\sim} G_0(\mu_\theta, \sigma_\theta^2, b_\phi)$, $v_l \overset{i.i.d.}{\sim} \text{Be}(1, \alpha)$, and $p_1 = v_1, p_l = v_l \prod_{r=1}^{l-1}(1 - v_l)$ for $l > 1$. With an appropriate truncation level N, we can obtain an approximated G^*, through

$$G_N^* = \sum_{l=1}^{N} p_l \delta_{\tilde{\Psi}_l}(\cdot),$$

where $p_N = 1 - \sum_{l=1}^{N-1} p_l$. The approximation level is related to both N and α, for a fixed α, $E(\sum_{l=1}^{N} p_l \mid \alpha) = 1 - \left(\frac{\alpha}{\alpha+1}\right)^N$. Then the posterior G is approximated by

$$q_{k+1} \sum_{l=1}^{N} p_l \delta_{\tilde{\Psi}_l}(\cdot) + \sum_{r=1}^{k} q_r \delta_{\Psi_r^*}(\cdot).$$

In each MCMC iteration b, we can obtain a set of parameters $\Theta^{(b)} = (\boldsymbol{\Psi}^{*(b)}, b_\phi^{(b)}, \mu_\theta^{(b)}, \sigma_\theta^{2(b)}, \boldsymbol{L}^{(b)}, k^{(b)}, \boldsymbol{n}^{(b)})$. Based on $\Theta^{(b)}$, we first draw a weight vector $(q_1^{(b)}, \ldots, q_{k+1}^{(b)}) \sim \text{Dir}(n_1^{(b)}, \ldots, n_k^{(b)}, \alpha)$, then obtain an approximated $G_b^* = \{(p_l^{(b)}, \tilde{\Psi}_l^{(b)}) : l = 1, \ldots, N\}$, where $\tilde{\Psi}_l^{(b)} \overset{i.i.d.}{\sim} G_0(\Psi^* \mid b_\phi^{(b)}, \mu_\theta^{(b)}, \sigma_\theta^{2(b)})$ and $p_1^{(b)} = v_1^{(b)}$, $p_l^{(b)} = v_l^{(b)} \prod_{r=1}^{l-2}(1 - v_l^{(b)})$ for $l > 1$ with $v_l^{(b)} \overset{i.i.d.}{\sim} \text{Be}(1, \alpha)$ for $l < N$, and $p_N^{(b)} = 1 - \sum_{l=1}^{N-1} p_l^{(b)}$. Evatually, for given covariates $x = x_0$ and $z = z_0$, we can estimate the conditional density and survival

functions on grids of time,

$$
\begin{aligned}
f(\cdot \mid x_0, H^{(b)}) &\approx q_{k+1}^{(b)} \sum_{l=1}^{N} \tilde{p}_l^{(b)} f_{\mathrm{LL}}(\cdot|\tilde{\theta}_l^{(b)} \exp(\boldsymbol{x}_0'\tilde{\boldsymbol{\beta}}_l^{(b)}), \tilde{\phi}_l^{(b)}) + \sum_{r=1}^{k} \tilde{q}_r^{(b)} f_{\mathrm{LL}}(\cdot|\theta_r^{*(b)} \exp(\boldsymbol{x}_0'\boldsymbol{\beta}_r^{*(b)}), \phi_r^{*(b)}), \\
\text{where } \tilde{p}_l^{(b)} &= \frac{p_l^{(b)} f(\boldsymbol{x}_0 \mid \tilde{\boldsymbol{\Omega}}_l^{(b)})}{q_{k+1}^{(b)} \sum_{l=1}^{N} p_l^{(b)} f(\boldsymbol{x}_0 \mid \tilde{\boldsymbol{\Omega}}_l^{(b)}) + \sum_{r=1}^{k} q_r^{(b)} f(\boldsymbol{x}_0 \mid \tilde{\boldsymbol{\Omega}}_r^{*(b)})} \\
\text{and } \tilde{q}_r^{(b)} &= \frac{q_r^{(b)} f(\boldsymbol{x}_0 \mid \tilde{\boldsymbol{\Omega}}_l^{(b)})}{q_{k+1}^{(b)} \sum_{l=1}^{N} p_l^{(b)} f(\boldsymbol{x}_0 \mid \tilde{\boldsymbol{\Omega}}_l^{(b)}) + \sum_{r=1}^{k} q_r^{(b)} f(\boldsymbol{x}_0 \mid \tilde{\boldsymbol{\Omega}}_r^{*(b)})} \\
S(\cdot \mid x_0, H^{(b)}) &\approx q_{k+1}^{(b)} \sum_{l=1}^{N} \tilde{p}_l^{(b)} S_{\mathrm{LL}}(\cdot|\tilde{\theta}_l^{(b)} \exp(\boldsymbol{x}_0'\tilde{\boldsymbol{\beta}}_l^{(b)}), \tilde{\phi}_l^{(b)}) + \sum_{r=1}^{k} \tilde{q}_r^{(b)} S_{\mathrm{LL}}(\cdot|\theta_r^{*(b)} \exp(\boldsymbol{x}_0'\boldsymbol{\beta}_r^{*(b)}), \phi_r^{*(b)}), \\
h(\cdot \mid x_0, H^{(b)}) &= \frac{f(\cdot \mid x_0, H^{(b)})}{S(\cdot \mid x_0, H^{(b)})}.
\end{aligned}
$$

B.4 Additional results with censored data

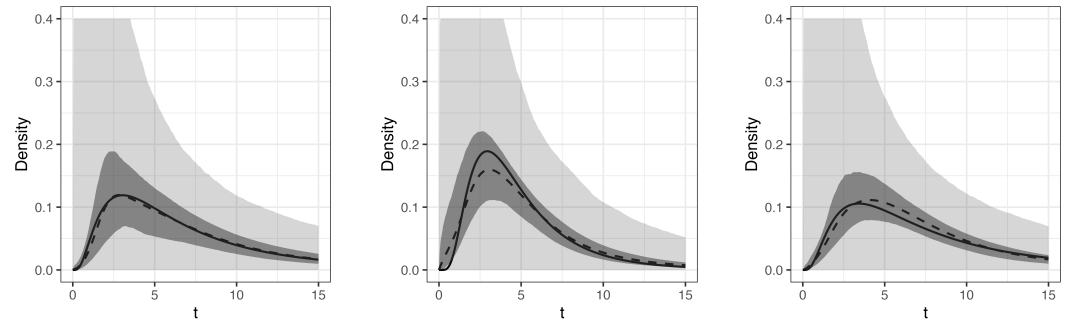
In Section 3.2, we demonstrated the application of the DPM-LLx model to a dataset featuring both a binary and a continuous covariate in a complex scenario. Additionally, we evaluated the model's performance in handling censored data by implementation a random censoring mechanism, as outlined in Section 3.1. Utilizing the synthetic dataset from Section 3.2, we crafted two datasets featuring distinct proportions of censored observations, namely, 24.2% and 44.2%. The prior hyperparameters were set consistently across analyses, and the ensuing posterior inferences are depicted in panels (d)-(i) of Figures B.1, B.2, and B.3. The prior 95% pointwise credible intervals are represented by light gray shaded areas, encompassing a broad spectrum of values. Dashed lines indicate posterior mean estimates, while posterior 95% pointwise credible intervals are marked by dark gray shaded regions. These findings demonstrate the model's proficiency in accurately capturing the underlying functions (depicted as solid

lines) across varying covariate values, even in scenarios of light to moderate censoring.

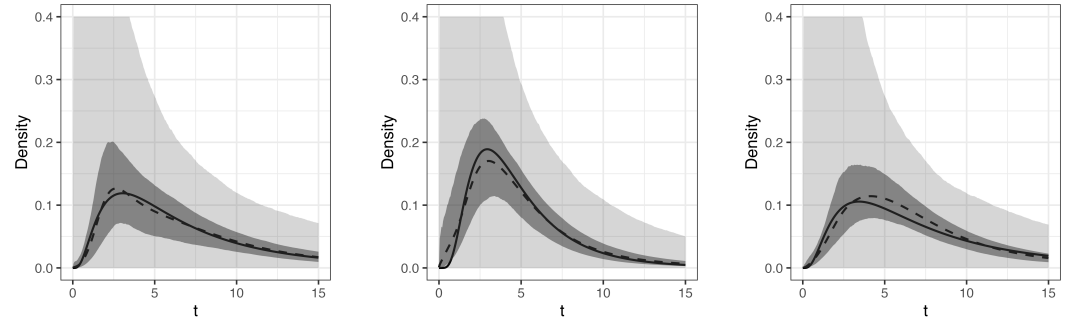
B.5 Additional results of the small cells lung data anlaysis

In Section 3.5, we analyzed the small cells lung dataset with the DMP-LLx model. We illustrate the additional results for survival and hazard estimates in Figures B.6 and B.8. The dashed lines show the posterior mean estimates with 95% posterior and prior credible intervals are represented by dark and light shaded regions, respectively.

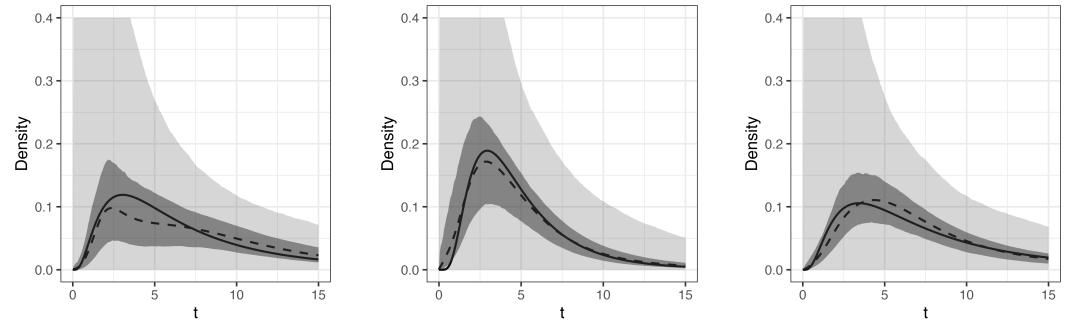
In addition, we performed prior sensitivity analysis by change hyperparameters. Let $b_\phi \sim$ inv-Ga$(5, 20)$, and $\mu_\theta \sim \mathrm{N}(5.5, 1)$, and $\mu_\beta \sim \mathrm{N}((0, 0)', \mathrm{I}_2)$, and $\alpha \sim$ Ga$(5, 1)$. For other parameters, we let a_ϕ, σ^2_θ, σ^2_β, s_μ, S_μ, and τ^2 be fixed at 5, 2, $1.5\mathrm{I}_2$, and 0.6. The posterior inferences for the density, survival, and hazard functions, across a range of covariate values, are showcased in Figures B.5, B.7, and B.9. The results indicate that the inferences remain relatively stable, even with moderate adjustments to the prior hyperparameters.



(a) $z = 1, x = -1$ (b) $z = 0, x = 0$ (c) $z = 1, x = 0$

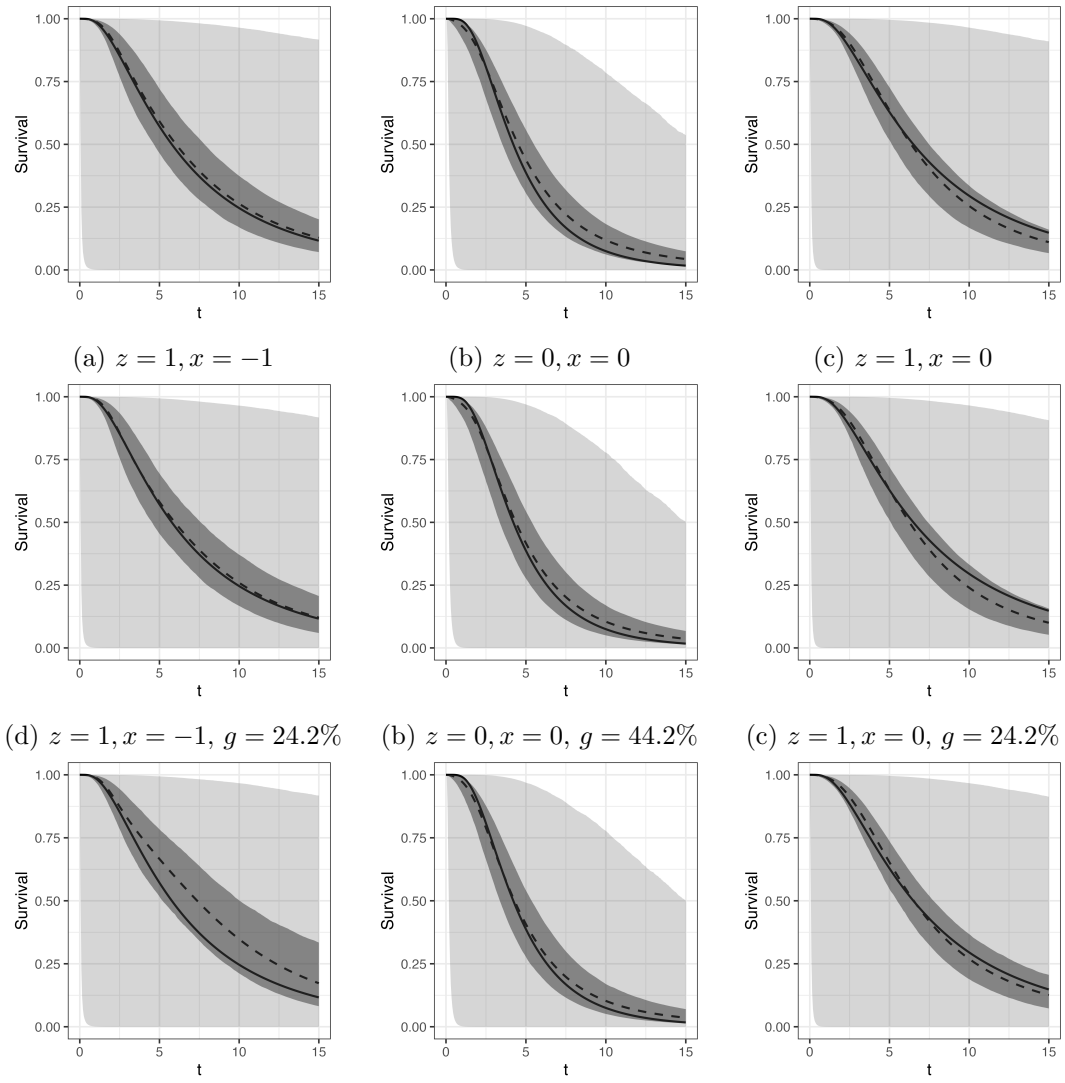


(d) $z = 1, x = -1,\ g = 24.2\%$ (b) $z = 0, x = 0,\ g = 24.2\%$ (c) $z = 1, x = 0,\ g = 24.2\%$



(g) $z = 1, x = -1,\ g = 44.2\%$ (h) $z = 0, x = 0,\ g = 44.2\%$ (i) $z = 1, x = 0,\ g = 44.2\%$

Figure B.1: Simulation 2. The three rows have density estimates on datasets with varying proportion of censored survival times (0%, 24.2%, and 44.2%). The three columns present density estimates with different values of covariates (z, x). In each panel, posterior mean estimates (dashed lines), 95% pointwise posterior credible interval estimates (dark shaded regions), 95% pointwise prior credible interval estimates (light shaded regions), and the simulation truth (solid lines) are shown.

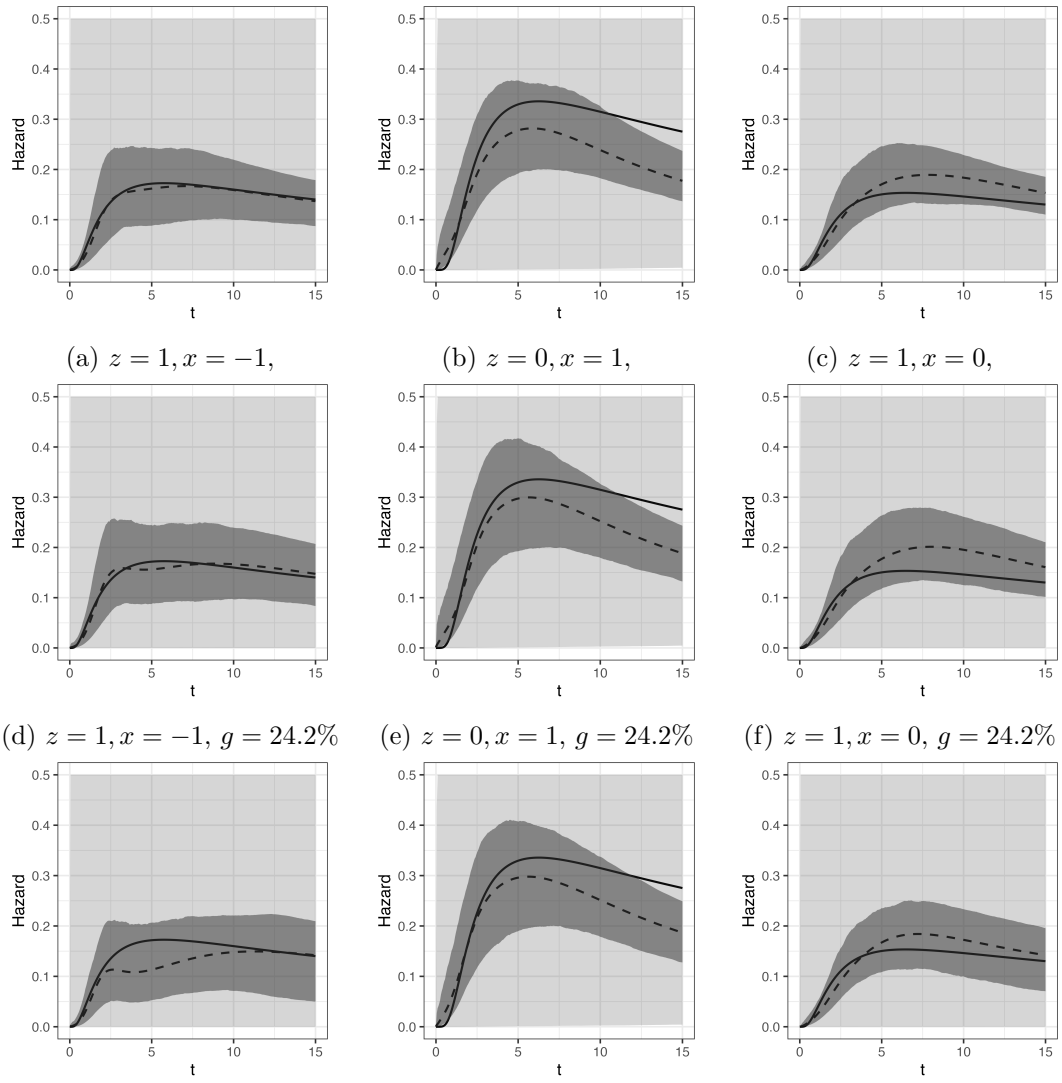


(a) $z = 1, x = -1$ (b) $z = 0, x = 0$ (c) $z = 1, x = 0$

(d) $z = 1, x = -1,\ g = 24.2\%$ (b) $z = 0, x = 0,\ g = 44.2\%$ (c) $z = 1, x = 0,\ g = 24.2\%$

(g) $z = 1, x = -1,\ g = 44.2\%$ (h) $z = 0, x = 1,\ g = 44.2\%$ (i) $z = 1, x = 0,\ g = 44.2\%$

Figure B.2: Simulation 2. The three rows have survival function estimates on datasets with varying proportion of censored survival times (0%, 24.2%, and 44.2%). The three columns present survival function estimates with different values of covariates (z, x). In each panel, posterior mean estimates (dashed lines), 95% pointwise posterior credible interval estimates (dark shaded regions), 95% pointwise prior credible interval estimates (light shaded regions), and the simulation truth (solid lines) are shown.

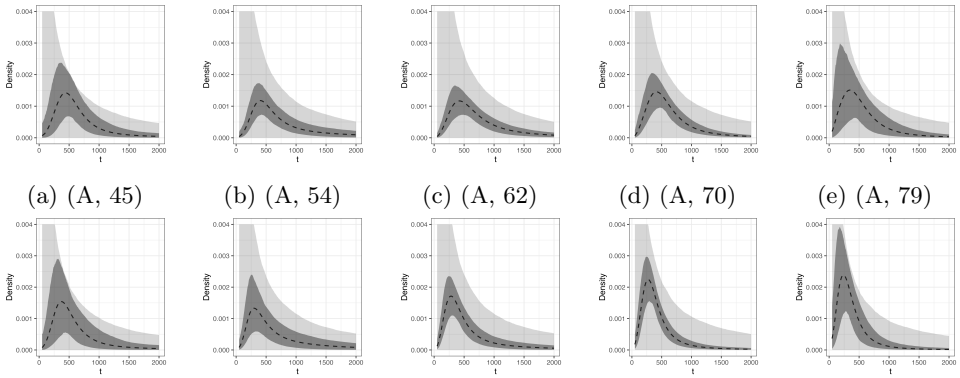


(a) $z = 1, x = -1,$ (b) $z = 0, x = 1,$ (c) $z = 1, x = 0,$

(d) $z = 1, x = -1,\ g = 24.2\%$ (e) $z = 0, x = 1,\ g = 24.2\%$ (f) $z = 1, x = 0,\ g = 24.2\%$

(g) $z = 1, x = -1,\ g = 44.2\%$ (h) $z = 0, x = 1,\ g = 44.2\%$ (i) $z = 1, x = 0,\ g = 44.2\%$

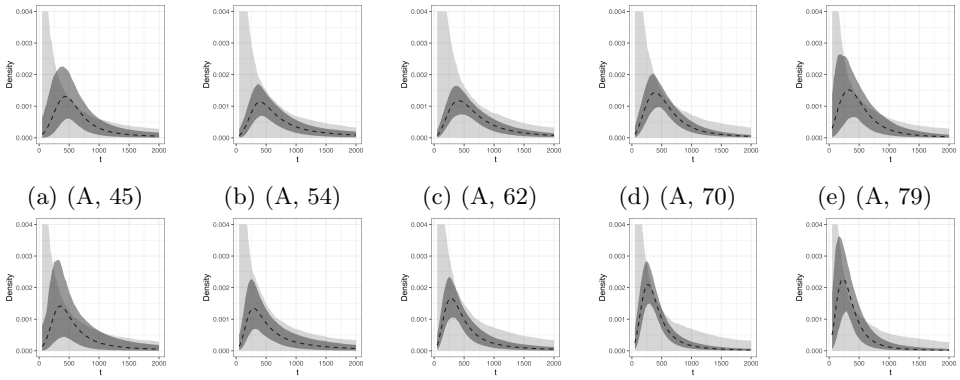
Figure B.3: Simulation 2. The three rows have hazard estimates on datasets with varying proportion of censored survival times (0%, 24.2%, and 44.2%). The three columns present hazard estimates with different values of covariates (z, x). In each panel, posterior mean estimates (dashed lines), 95% pointwise posterior credible interval estimates (dark shaded regions), 95% pointwise prior credible interval estimates (light shaded regions), and the simulation truth (solid lines) are shown.



(a) (A, 45) (b) (A, 54) (c) (A, 62) (d) (A, 70) (e) (A, 79)

(f) (B, 45) (g) (B, 54) (h) (B, 62) (i) (B, 70) (j) (B, 79)

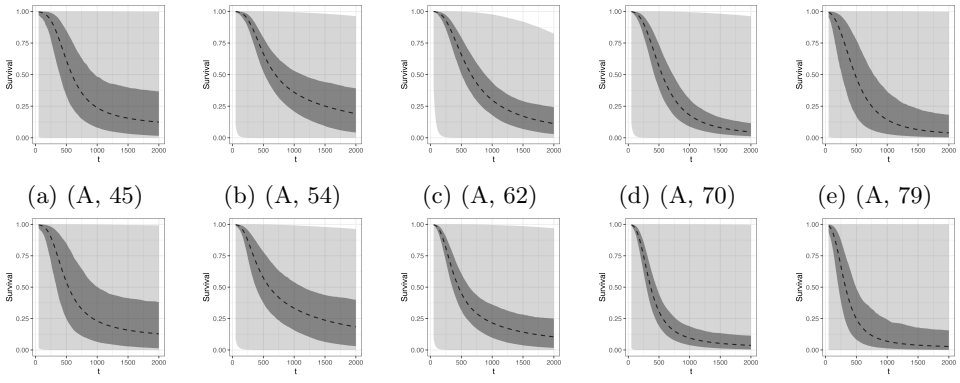
Figure B.4: Small cells lung cancer data. Each row presents density function estimates with different treatment Arm (A and B). Each column corresponds to different values of age (45, 54, 62, 70, 79).



(a) (A, 45) (b) (A, 54) (c) (A, 62) (d) (A, 70) (e) (A, 79)

(f) (B, 45) (g) (B, 54) (h) (B, 62) (i) (B, 70) (j) (B, 79)

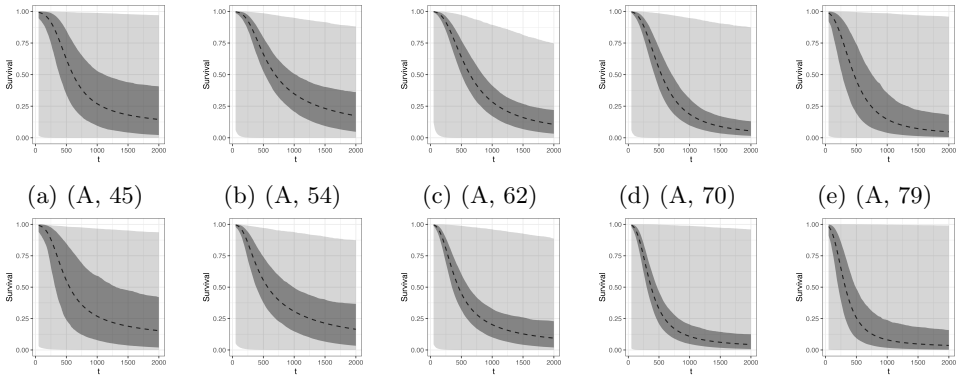
Figure B.5: Small cells lung cancer data. Sensitivity analysis. Each row presents density function estimates with different treatment Arm (A and B).



(a) (A, 45) (b) (A, 54) (c) (A, 62) (d) (A, 70) (e) (A, 79)

(f) (B, 45) (g) (B, 54) (h) (B, 62) (i) (B, 70) (j) (B, 79)

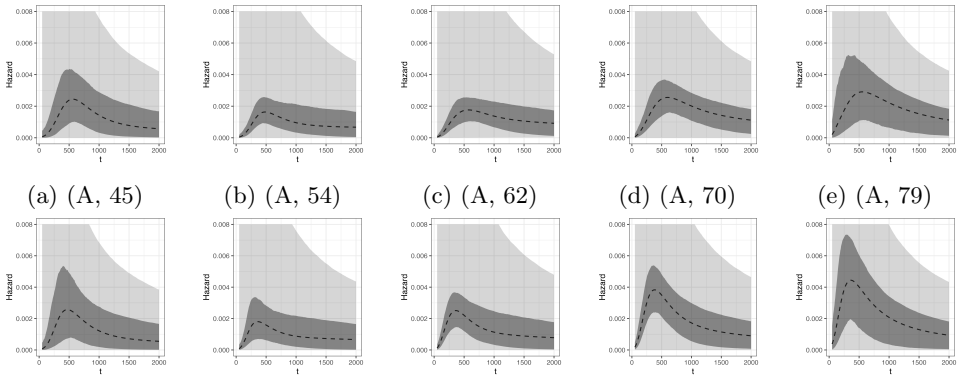
Figure B.6: Small cells lung cancer data. Each row presents survival function estimates with different treatment Arm (A and B). Each column corresponds to different values of age (45, 54, 62, 70, 79).



(a) (A, 45) (b) (A, 54) (c) (A, 62) (d) (A, 70) (e) (A, 79)

(f) (B, 45) (g) (B, 54) (h) (B, 62) (i) (B, 70) (j) (B, 79)

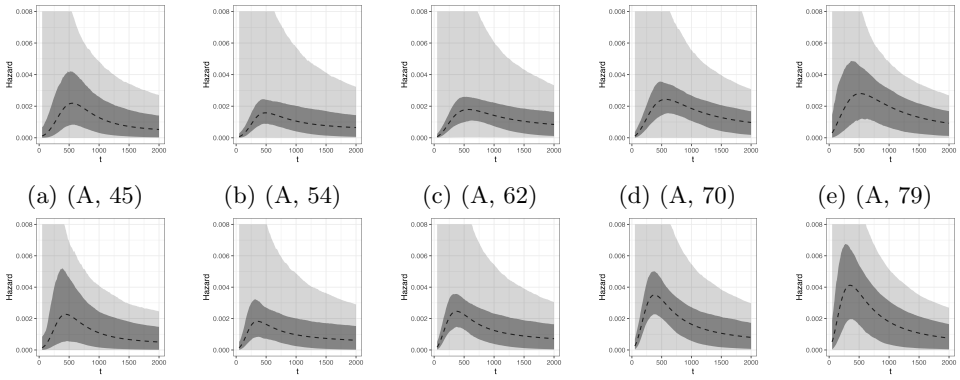
Figure B.7: Small cells lung cancer data. Sensitivity analysis. Each row presents survival function estimates with different treatment Arm (A and B). Each column corresponds to different values of age (45, 54, 62, 70, 79).



(a) (A, 45) (b) (A, 54) (c) (A, 62) (d) (A, 70) (e) (A, 79)

(f) (B, 45) (g) (B, 54) (h) (B, 62) (i) (B, 70) (j) (B, 79)

Figure B.8: Small cells lung cancer data. Each row presents hazard estimates with different treatment Arm (A and B). Each column corresponds to different values of age (45, 54, 62, 70, 79).



(a) (A, 45) (b) (A, 54) (c) (A, 62) (d) (A, 70) (e) (A, 79)

(f) (B, 45) (g) (B, 54) (h) (B, 62) (i) (B, 70) (j) (B, 79)

Figure B.9: Small cells lung cancer data. Sensitivity analysis. Each row presents hazard estimates with different treatment Arm (A and B). Each column corresponds to different values of age (45, 54, 62, 70, 79).

B.6 Model Comparison with parametric Weibull AFT model

In addition, we compare the DPM-LLx model with a parametric AFT Weibull model using the small cell lung cancer dataset. The Weibull density function, conditional on covariate $\boldsymbol{x}$, is given by

$$f(t \mid \boldsymbol{x}) = \phi_w t^{\phi-1} \theta_w^{-\phi_w} \exp(-\phi_w \boldsymbol{\beta}_w \boldsymbol{x}') \exp\left\{-t^{-\phi_w} \theta_w^{-\phi_w} \exp(-\phi_w \boldsymbol{\beta} \boldsymbol{x}')\right\},$$

where the scale parameter is $\theta_w \exp(\boldsymbol{x}'\boldsymbol{\beta}_w)$ and the shape parameter is ϕ_w. Notably, the Weibull AFT model can also be expressed within a PH framework. Thus, the hazards ratio given covariate vectors $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ is $\exp(-\phi_w \boldsymbol{\beta}_w (\boldsymbol{x}_1 - \boldsymbol{x}_2)')$.

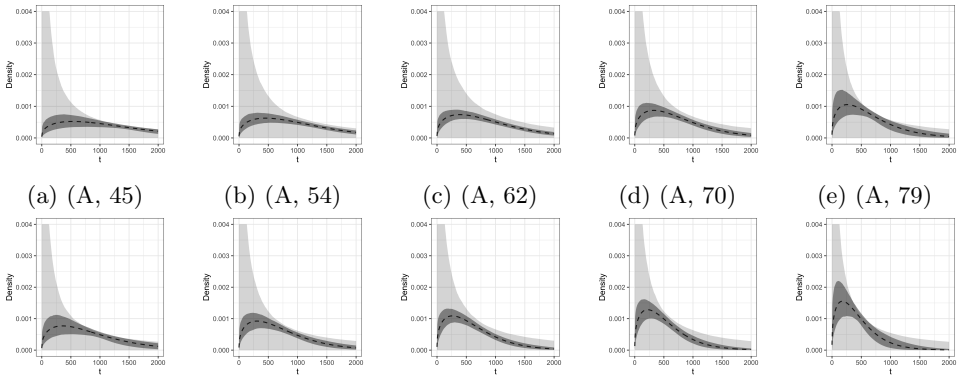
The prior distributions for the parameters are specified as follows: $\theta_w \sim \text{Ga}(a_\theta, b_\theta)$, $\boldsymbol{\beta}_w \sim \text{N}(\mu_\beta, \Sigma_\beta)$, and $\phi_w \sim \text{Ga}(a_\phi, b_\phi)$. For this dataset, the hyperparameters are set as follows: $a_\theta = 1$ and $b_\theta = 1000$, $a_\phi = 1$, and $b_\phi = 1$, μ_β is fixed at $(0, 0)'$, and Σ_β is taken as I_2.

The hyperparameters for the parametric survival regression model are chosen to ensure that the prior uncertainty is sufficiently broad over range of observed survival times. The 95% prior interval estimates for density and survival function are demonstrated by the light gray regions in Figures B.10 and B.11, respectively. Alongside the prior interval estimates, we demonstrate the posterior point estimates as solid lines and posterior 95% pointwise credible interval in dark gray regions.

Furthermore, the posterior mean estimates of survival and hazard functions given different values of covariates are displayed in Figure B.12. The parametric assumptions of the model result in hazard functions that are monotonically increasing.

The analysis indicates that as age increases, the survival probability decreases while the hazard increases. However, unlike the DPM-LLx model, the parametric model imposes constraints that prevent the crossing of survival curves for different covariate values. This limitation reduces the model's flexibility in capturing more complex relationships within the data.

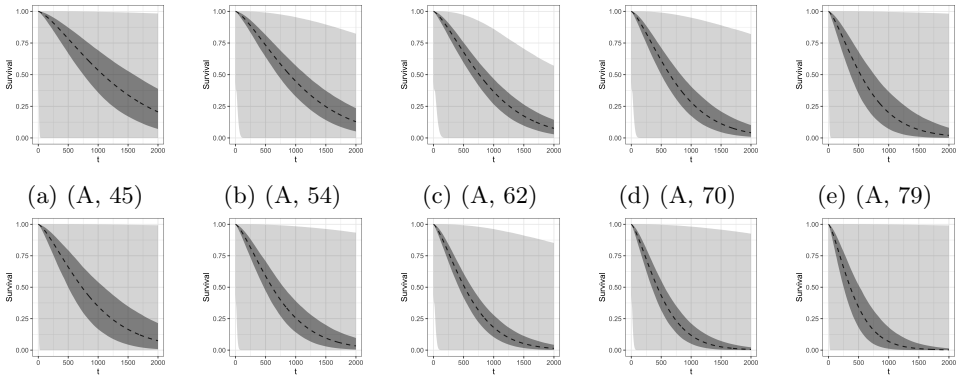
The log-pseudo marginal likelihood statistic (LPML) is a metric that measures model performance based on cross-validated posterior predictive probability. It is defined as the sum of the log conditional predictive ordinates (CPO). Specifically, the CPO is calculated as $f(t_i \mid t_{-i}, \boldsymbol{x}_{-i})$ for observed survival times and $S(t_i \mid t_{-i}, \boldsymbol{x}_{-i})$ for censored time, where subscript $-i$ represents all subjects except the ith one. When comparing the AFT Weibull model and the DPM-LLx model using LPML scores, we observe a difference in their predictive performance. The LPML score for the AFT Weibull model is -746.39, whereas the DPM-LLx model achieves a higher LPML score of -735.56. This difference indicates that the DPM-LLx model provides a better fit to the data, offering more accurate predictions and demonstrating greater flexibility in capturing the underlying structure of the data compared to the parametric AFT Weibull model.



(a) (A, 45) (b) (A, 54) (c) (A, 62) (d) (A, 70) (e) (A, 79)

(f) (B, 45) (g) (B, 54) (h) (B, 62) (i) (B, 70) (j) (B, 79)

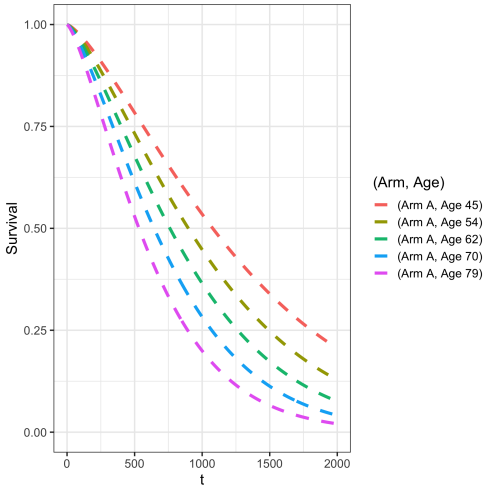
Figure B.10: Small cells lung cancer data using parametric Weibull AFT model. Each row presents density function estimates with a different treatment arm (A and B). Each column corresponds to different values of age (45, 54, 62, 70, 79).



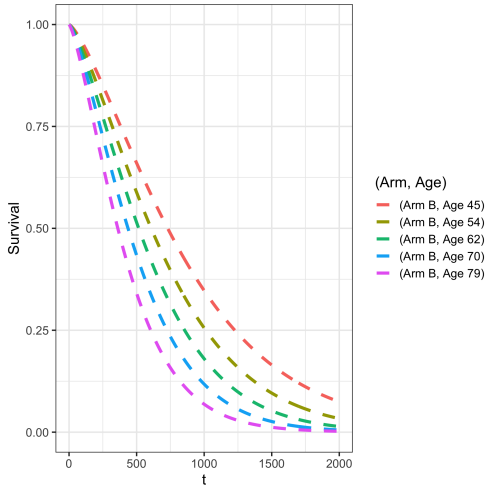
(a) (A, 45) (b) (A, 54) (c) (A, 62) (d) (A, 70) (e) (A, 79)

(f) (B, 45) (g) (B, 54) (h) (B, 62) (i) (B, 70) (j) (B, 79)

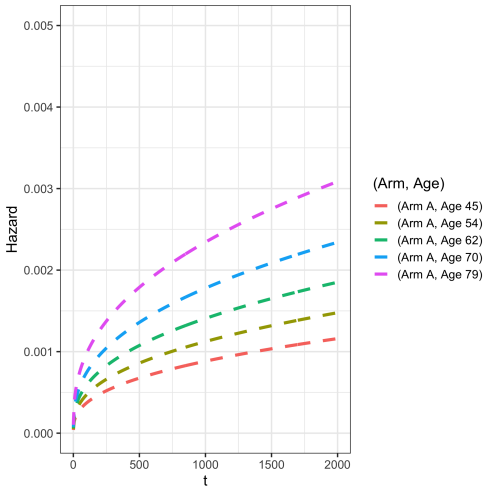
Figure B.11: Small cells lung cancer data using parametric Weibull AFT model. Each row presents survival function estimates with different treatment Arm (A and B). Each column corresponds to different values of age (45, 54, 62, 70, 79).



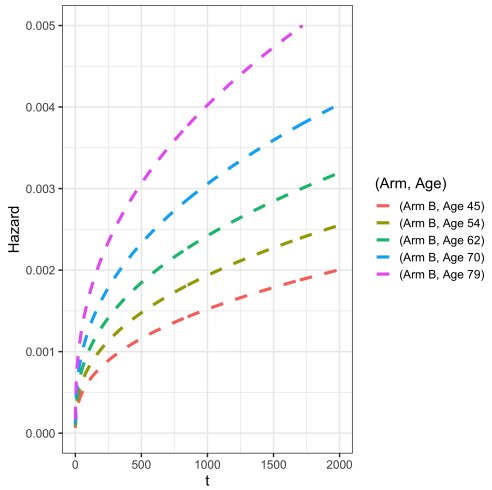
(a) Survival, Arm A



(b) Survival, Arm B



(a) Hazard, Arm A



(b) Hazard, Arm B

Figure B.12: Small cells lung cancer data using parametric Weibull AFT model. The first row shows the posterior mean estimates of the conditional survival function, while the second panel displays posterior estimates of the conditional hazard function. The left and right columns are corresponding to Arm A and Arm B, respectively. Each color corresponds to a distinct age of patients.

Appendix C

Supplementary material - Bayesian Nonparametric Joint Model for recurrent events and survival data

C.1 MCMC details - Blocked Gibbs sampler

The posterior computation is based on the Blocked Gibbs sampler, where the mixing distributions G_z and H_z are approximated by its truncated version at level of B_G and B_H, G_{zB} and H_{zB}, as $G_{zB} = \sum_{l=1}^{B_G} p_{zl}\delta_{(\theta_l,\boldsymbol{\beta}_l,\phi_l)}(\cdot)$ and $H_{zB} = \sum_{l=1}^{B_H} \omega_{zl}\delta_{(\lambda_l,\boldsymbol{\gamma}_l,\eta_l)}(\cdot)$.

The mixing parameters are denoted by $(\theta_l^*, \boldsymbol{\beta}_l^*, \phi_l^*)$ and $(\lambda_l^*, \boldsymbol{\gamma}_l^*, \eta_l^*)$ for $l = 1, \ldots, N$, respectively. For each survival and gap times, we introduce an indicator variable as in the previous section. $\boldsymbol{L_z} = \{L_{zi} : i = 1, \ldots, n_z\}$, for $z = ctr, trt$. Thus,

for $j = 1, \ldots, N_{iz}$, $i = 1, \ldots, n_z$, $z = ctr, trt$, the full hierarchical model is written as,

$$
\begin{aligned}
T_{iz} \mid L_{iz}, \boldsymbol{\theta}^*, \boldsymbol{\beta}^*, \boldsymbol{\phi}^*, \epsilon_{is}, x_{is} &\overset{ind.}{\sim} \text{LL}(\theta^*_{L_{iz}} \exp(\boldsymbol{\beta}^*_{L_{iz}} x_{is})/\epsilon_{is}, \phi^*_{L_{iz}}), \\
L_{iz} \mid p_{lz} &\overset{ind.}{\sim} p_{lz}\delta_l(L_{iz}), \text{ for } l = 1, \ldots, B_G \\
p_{lz} &= V_{lz} \prod_{r=1}^{l-1}(1 - V_{ro}), \\
(V_{lC}, V_{lT}) \mid \alpha_0, \alpha &\overset{i.i.d.}{\sim} \text{Bi-Be}(1 - \alpha_0, \alpha_0, \alpha_0, \alpha - \alpha_0), \\
(\log(\theta^*_l), \boldsymbol{\beta}^*_l, \phi^{*2}_l) \mid \mu_\theta, \sigma^2_\theta, \mu_\beta, \Sigma_\beta, a_\phi, b_\phi &\overset{i.i.d.}{\sim} \text{N}(\log(\theta^*_l) \mid \mu_\theta, \sigma^2_\theta)\text{N}(\boldsymbol{\beta}^*_l \mid \mu_\beta, \Sigma_\beta)\text{inv-Ga}(\phi^{*2}_l \mid a_\phi, b_\phi), \\
\alpha_0 \mid a_0, b_0 &\sim \text{Be}(a_0, b_0) \\
\alpha \mid \alpha_0, c_0 &\sim \text{Pa}(c_0, \alpha_0) \\
(\mu_\theta, \mu_\beta, b_\phi) \mid s_\theta, S_\theta, s_\beta, S_\beta, r_\phi, R_\phi &\sim \text{N}(s_\theta, S_\theta)\text{N}(s_\beta, S_\beta)\text{inv-Ga}(r_\phi, R_\phi)
\end{aligned}
$$

$$
\begin{aligned}
W_{ijz} \mid U_{ijz}, \boldsymbol{\lambda}^*, \boldsymbol{\gamma}^*, \boldsymbol{\eta}^*, \xi_{iz}, z_{iz} &\overset{i.i.d.}{\sim} \text{LL}(\lambda^*_{U_{ijz}} \exp(\boldsymbol{\gamma}^*_{U_{ijz}} z_{iz})/\xi_{iz}, \eta^*_{U_{ijz}}), \text{ for } l = 1, \ldots, B_H \\
U_{ijz} \mid \omega_{lz} &\overset{ind.}{\sim} \omega_{lz}\delta_l(U_{ijz}), \\
\omega_{lz} &= \pi_{lz} \prod_{r=1}^{l-1}(1 - \pi_{ro}), \\
(\pi_{lC}, \pi_{lT}) \mid \zeta_0, \zeta &\overset{i.i.d.}{\sim} \text{Bi-Be}(1 - \zeta_0, \zeta_0, \zeta_0, \zeta - \zeta_0), \\
(\log(\lambda^*_l), \boldsymbol{\gamma}^*_l, \eta^{*2}_l) \mid \mu_\lambda, \sigma^2_\lambda, \mu_\gamma, \Sigma_\gamma, a_\eta, b_\eta &\overset{i.i.d.}{\sim} \text{N}(\log(\lambda^*_l) \mid \mu_\lambda, \sigma^2_\lambda)\text{N}(\boldsymbol{\gamma}^*_l \mid \mu_\gamma, \Sigma_\gamma)\text{inv-Ga}(\eta^{*2}_l \mid a_\eta, b_\eta), \\
\zeta_0 \mid a_1, b_1 &\sim \text{Be}(a_1, b_1) \\
\zeta \mid \zeta_0, c_0 &\sim \text{Pa}(c_1, \zeta_0) \\
(\mu_\lambda, \mu_\gamma, b_\eta) \mid s_\lambda, S_\lambda, s_\gamma, S_\gamma, r_\eta, R_\eta &\sim \text{N}(s_\lambda, S_\lambda)\text{N}(s_\gamma, S_\gamma)\text{inv-Ga}(r_\eta, R_\eta)
\end{aligned}
$$

$$
\begin{aligned}
(\epsilon_{iz}, \xi_{iz})' \mid \Sigma_{ez} &\overset{i.i.d.}{\sim} \text{LN}((\epsilon_{iz}, \xi_{iz})' \mid (0, 0)', \Sigma_{ez}), \\
\Sigma_{ez} \mid c_e, C_e &\sim \text{inv-Wish}(c_e, C_e).
\end{aligned}
$$

Update survival model indicators

The posterior computation algorithm starts from the survival times. The latent indicator functions $\{L_{iz} : i = 1, \ldots, n_z, z = ctr, trt\}$ are updated

$$
p(L_{iz} = l \mid -) \propto p_{lz} \left\{ f_{\text{LL}}(\tau_{iz} \mid \theta^{*-}_l/\epsilon_{iz}, \phi^{*-}_l) \right\}^{\nu_{iz}} \left\{ S_{\text{LL}}(\tau_{iz} \mid \theta^{*-}_l/\epsilon_{iz}, \phi^{*-}_l) \right\}^{1-\nu_{iz}}.
$$

Update survival model mixture weights

The mixture weights are updated through the latent constructive variables. For $l = 1, \ldots, BG - 1$,

$$\begin{aligned} p(u_{11l}, u_{10l}, u_{01l}, u_{00l} \mid -) &\propto u_{11l}^{1-\alpha_0-1} u_{10l}^{\alpha_0-1} u_{01l}^{\alpha_0-1} u_{00l}^{\alpha-\alpha_0-1} (u_{11l} + u_{10l})^{n_{lC}} (1 - u_{11l} - u_{10l})^{\sum_{r=1}^{B} n_{rC}} \\ &\quad \times (u_{11l} + u_{01l})^{n_{lT}} (1 - u_{11l} - u_{01l})^{\sum_{r=1}^{B} n_{rT}} \\ &\propto u_{11l}^{1-\alpha_0-1} u_{10l}^{\alpha_0-1} u_{01l}^{\alpha_0-1} u_{00l}^{\alpha-\alpha_0-1} (u_{11l} + u_{10l})^{n_{lC}} (u_{01l} + u_{00l})^{\sum_{r=1}^{B} n_{rC}} \\ &\quad \times (u_{11l} + u_{01l})^{n_{lT}} (u_{10l} + u_{00l})^{\sum_{r=1}^{B} n_{rT}}, \end{aligned}$$

where n_{lz} represents the cardinality of the set $\{L_{iz} = l : i = 1, \ldots, n_z, z = ctr, trt\}$. Here, a Metropolis-hasting algorithm is applied here to sample $(u_{11l}, u_{10l}, u_{01l}, u_{00l})$. Let the current state be $\boldsymbol{u}_l^{(t)} = (u_{11l}^{(t)}, u_{10l}^{(t)}, u_{01l}^{(t)}, u_{00l}^{(t)})$, and the proposing state be $\boldsymbol{u}_l^* = (u_{11l}^*, u_{10l}^*, u_{01l}^*, u_{00l}^*)$. Let Q denote the proposal distribution, $Q(\boldsymbol{u}^* \mid \boldsymbol{u}_l^{(t)}) = \mathrm{Dir}(\boldsymbol{u}^* \mid a^* \boldsymbol{u}_l^{(t)})$, with $\mathrm{E}(\boldsymbol{u}_l^* \mid \boldsymbol{u}_l^{(t)}) = \boldsymbol{u}_l^{(t)}$, and a^* controls the variance, in practice, we recommend to set it at 100. The acceptance rate is

$$\min\left(1, \frac{Q(\boldsymbol{u}_l^{(t)} \mid \boldsymbol{u}_l^*) P(\boldsymbol{u}_l^*)}{Q(\boldsymbol{u}_l^* \mid \boldsymbol{u}_l^{(t)}) P(\boldsymbol{u}_l^{(t)})}\right),$$

where

$$\begin{aligned} P(\boldsymbol{u}_l^*) &= u_{11l}^{1-\alpha_0-1} u_{10l}^{\alpha_0-1} u_{01l}^{\alpha_0-1} u_{00l}^{\alpha-\alpha_0-1} \\ &\quad \times (u_{11l} + u_{10l})^{n_{lC}} (u_{01l} + u_{00l})^{\sum_{r=l+1}^{B_G} n_{rC}} (u_{11l} + u_{01l})^{n_{lT}} (u_{10l} + u_{00l})^{\sum_{r=l+1}^{B_G} n_{rT}}. \end{aligned}$$

The related hyperparameters α and α_0 are updated as

$$p(\alpha, \alpha_0 \mid -) \propto p(\alpha \mid \alpha_0) p(\alpha_0) \prod_{l=1}^{B_G} u_{11l}^{-\alpha_0} u_{10l}^{\alpha_0-1} u_{01l}^{\alpha_0-1} u_{00l}^{\alpha-\alpha_0-1}.$$

There is no closed form, thus Metropolis-hasting algorithm is applied here. Conditional on $\alpha_0^{(t)}$, α_0^* is proposed from $\mathrm{Be}(\alpha_0^{(t)}, 1 - \alpha_0^{(t)})$ with mean $\alpha_0^{(t)}$. $\alpha^* \mid \alpha_0^*$ is proposed

from $\mathrm{Pa}(2, \alpha_0^*)$. The proposing distribution is written as $Q(\alpha_0^*, \alpha^* \mid \alpha_0^{(t)}, \alpha^{(t)}) = \mathrm{Be}(\alpha_0^* \mid \alpha_0^{(t)}, 1-\alpha^{(t)})\mathrm{Pa}(\alpha^* \mid 2, \alpha_0^*)$.

Update mixing parameters in survival model The mixing parameters $(\theta_l^*, \boldsymbol{\beta}_l^*, \phi_l)$ for $l = 1, \ldots, B_G$, are updated in the following way. For each l, if $l \notin \{L_{zi} : i = 1, \ldots, n_z, z = ctr, trt\}$, draw each parameter from their prior base measure. Otherwise, we sample them in the following way.

$$
\begin{aligned}
p(\theta_l^*, \boldsymbol{\beta}_l^*, \phi_l^* \mid -) \quad &\propto \quad \mathrm{N}(\log(\theta_l^*) \mid \mu_\theta, \sigma_\theta^2)\mathrm{N}(\boldsymbol{\beta}_l^* \mid \mu_\beta, \Sigma_\beta)\mathrm{Ga}(\phi_l^{*2} \mid a_\phi, b_\phi) \\
&\quad \times \prod_{z=ctr,trt} \prod_{i:L_{iz}=l} \left(\frac{\phi_l^*}{\tau_{iz}}\right)^{\nu_{iz}} \frac{[\exp\{\phi_l^*(\log(\tau_{iz}) + \log(\epsilon_{iz}) - \log(\theta_l^*))\}]^{\nu_{iz}}}{[1 + \exp\{\phi_l^*(\log(\tau_{iz}) + \log(\epsilon_{iz}) - \log(\theta_l^*))\}]^{\nu_{iz}+1}} \\
&\propto \quad \mathrm{N}(\log(\theta_l^*) \mid \mu_\theta, \sigma_\theta^2)\mathrm{N}(\boldsymbol{\beta}_l^* \mid \mu_\beta, \Sigma_\beta)\mathrm{Ga}(\phi_l^{*2} \mid a_\phi, b_\phi) \\
&\quad \times \prod_{z=ctr,trt} \prod_{i:L_{iz}=l} \left(\frac{\phi_l^*}{\tau_{iz}}\right)^{\nu_{iz}} \int \exp\left(0.5(\nu_{iz}-1)\psi_{zil}\right) \exp(-0.5 u_{iz}\psi_{il}^2) p(u_{iz}) du_{iz},
\end{aligned}
$$

where $\psi_{zil} = \phi_l^*(\log(\tau_{iz}) - x_{iz}'\boldsymbol{\beta}_l^* + \log(\epsilon_{zi}) + \log(u_{zi}) - \log(\theta_l^*))$, and $p(u_{iz})$ is a density function of $\mathrm{PG}(1+\nu_{iz}, 0)$ evaluated at u_{iz}. We first sample latent variables $\{u_{iz} : i = 1, \ldots, n_z, z = ctr, trt\}$ from $\mathrm{PG}(1+\nu_{iz}, \phi_{L_{iz}}^*(\log(\tau_{iz}) - \boldsymbol{\beta}_l^* x_{iz} + \log(\epsilon_{iz}) - \log(\theta_{L_{iz}}^*)))$ independently. Then conditional on $\{u_{iz} : i = 1, \ldots, n_z, z = ctr, trt\}$, draw $(\theta_l^*, \boldsymbol{\beta}_l^*, \phi_l^*)$, for $l = 1, \ldots, B_G$ from

$$
\begin{aligned}
\log(\theta_l^*) \mid \boldsymbol{u}, - \quad &\overset{ind.}{\sim} \quad \mathrm{N}(\tilde{\mu}_{\theta l}, \tilde{\sigma}_{\theta l}^2), \\
\text{where } \tilde{\sigma}_{\theta l}^2 \quad &= \quad \left(\frac{1}{\sigma_\theta^2} + \phi_l^{*2} \sum_{z=ctr,trt} \sum_{i:L_{iz}=l} u_{iz}\right)^{-1}, \\
\text{and } \tilde{\mu}_{\theta l} \quad &= \quad \tilde{\sigma}_{\theta l}^2 \Bigg(\frac{\mu_\theta}{\sigma_\theta^2} + \phi_l^{*2} \sum_{z=ctr,trt} \sum_{i:L_{iz}} u_{iz}(\log(\tau_{iz}) - \boldsymbol{\beta}_{L_{iz}}^* x_{iz} + \log(\epsilon_{iz})) \\
&\qquad +0.5\phi_l^* \sum_{z=ctr,trt} \sum_{i:L_{iz}} (1-\nu_{iz})\Bigg),
\end{aligned}
$$

and

$$\begin{aligned}
\boldsymbol{\beta}_l^* \mid \boldsymbol{u}, - &\overset{ind.}{\sim} \mathrm{N}(\tilde{\mu}_{\beta l}, \tilde{\Sigma}_{\beta l}), \\
\text{where } \tilde{\Sigma}_\mu &= \left(\Sigma_\beta^{-1} + \phi_l^{*2} \sum_{z=ctr,trt} \sum_{i:L_{iz}} u_{iz} x_{iz} x_{iz}' \right)^{-1} \\
\text{and } \tilde{\mu}_{\beta l} &= \tilde{\Sigma}_{\beta l} \Bigg(\mu_\beta \Sigma_\beta^{-1} + \phi_l^{*2} \sum_{z=ctr,trt} \sum_{i:L_{iz}} u_{iz} (\log(\tau_{iz}) + \log(\epsilon_{iz}) - \log(\theta_l^*)) x_{iz} \\
&\quad + 0.5 \phi_l^* \sum_{z=ctr,trt} \sum_{i:L_{iz}} (1 - \nu_{iz}) x_{iz} \Bigg),
\end{aligned}$$

and

$$\begin{aligned}
p(\phi_l^* \mid -) &\propto \mathrm{Ga}(\tilde{a}_\phi, \tilde{b}_\phi) \times A, \\
\text{where } \tilde{a}_\phi &= 0.5 \sum_{z=ctr,trt} \sum_{i:L_{iz}} \nu_{iz} + a_\phi, \\
\tilde{b}_\phi &= \left(0.5 \sum_{z=ctr,trt} \sum_{i:L_{iz}} u_{iz} (\log(\tau_{iz}) - \boldsymbol{\beta}_l^* x_{iz} + \log(\epsilon_{iz}) - \log(\theta_l^*))^2 + \frac{1}{b_\phi} \right)^{-1}, \\
\text{and } A &= \exp\left(0.5 \sum_{z=ctr,trt} \sum_{i:L_{iz}} (\nu_{iz} - 1) \phi_l^* (\log(\tau_{iz}) - \boldsymbol{\beta}_l^* x_{iz} + \log(\epsilon_{iz}) - \log(\theta_l^*)) \right).
\end{aligned}$$

If $\nu_{iz} = 1$ for all i such that $L_{iz} = l$, draw ϕ_l^* from $\mathrm{Ga}(\tilde{a}_\phi, \tilde{b}_\phi)$ directly, otherwise, a Metropolis Random-Walk algorithm can be applied.

Updating hyperparameters in survival model

$$\begin{aligned}
p(b_\phi, \mu_\theta, \mu_\beta \mid -) &\propto \mathrm{N}(\mu_\theta \mid s_\theta, S_\theta) \text{inv-Ga}(b_\phi \mid r_\phi, R_\phi) \mathrm{N}(\mu_\beta \mid s_\beta, S_\beta) \\
&\quad \times \prod_{l=1}^{B_G} \mathrm{N}(\log(\theta_l^*) \mid \mu_\theta, \sigma_\theta^2) \mathrm{Ga}(\phi_l^{*2} \mid a_\phi, b_\phi) \mathrm{N}(\boldsymbol{\beta}_l^* \mid \mu_\beta, \Sigma_\beta).
\end{aligned}$$

We thus have

$$\begin{aligned}
\mu_\theta \mid - &\sim \mathrm{N}\left(\left(\frac{1}{S_\theta}+\frac{k}{\sigma_\theta^2}\right)^{-1}\left(\frac{s_\theta}{S_\theta}+\frac{\sum_{l=1}^{B_G}\log(\theta_l^*)}{\sigma_\theta^2}\right),\left(\frac{1}{S_\theta}+\frac{B_G}{\sigma_\theta^2}\right)^{-1}\right),\\
b_\phi \mid - &\sim \text{inv-Ga}\left(r_\phi+ka_\phi, R_\phi+\sum_{l=1}^{B_G}\phi_l^{*2}\right),\\
\text{and } \mu_\beta \mid - &\sim \mathrm{N}\left(\left(S_\beta^{-1}+B_G\Sigma_\beta^{-1}\right)^{-1}\left(S_\beta^{-1}s_\beta+\Sigma_\beta^{-1}\sum_{l=1}^{B_G}\boldsymbol{\beta}_l^*\right),\left(S_\beta^{-1}+B_G\Sigma_\beta^{-1}\right)^{-1}\right).
\end{aligned}$$

Updating gap membership indicators Let the indicators for gap times be $\{U_{ijz} : j = 1,\ldots,N_{iz}, i = 1,\ldots,n_z, o = C,T\}$, and their equivalent notations $\{\tilde{U}_{hz} : h = 1,\ldots,N_z+n_z, z = ctr, trt\}$. Each cluster membership indicator $\tilde{U}_{hz}$ is updated

$$\begin{aligned}
p(\tilde{U}_{hz}=l \mid -) \propto\ & \omega_{lz}\{f_{\mathrm{LL}}(w_{ijz} \mid \lambda^*_{\tilde{U}_{hz}}\exp(\boldsymbol{\gamma}^*_{\tilde{U}_{hz}}z_{iz})/\epsilon_i, \eta^*_{\tilde{U}_{hz}})\}^{\iota_{hz}}\\
&\times\{S_{\mathrm{LL}}(w_{ijz} \mid \lambda^*_{\tilde{U}_{hz}}\exp(\boldsymbol{\gamma}^*_{\tilde{U}_{hz}}z_{iz})/\epsilon_i, \eta^*_{\tilde{U}_{hz}})\}^{1-\iota_{hz}},
\end{aligned}$$

where $w_{ijz} = \tau_{iz} - y_{i,N_{iz},z}$ for $j > N_i$. The stick-breaking weights are updated through the bivariate Beta latent variables in the same method for survival times.

Updating distinct mixing parameters in gap times model Given all other parameters, the posterior full conditional of (θ_l^*, ϕ_l^*) is upto a proportion,

$$\begin{aligned}
p(\lambda_l^*, \boldsymbol{\gamma}_l^*, \eta_l^* \mid -) \propto\ & N(\log(\lambda_l^*) \mid \mu_\lambda, \sigma_\lambda^2)\mathrm{N}(\boldsymbol{\gamma}_l^* \mid \mu_\gamma, \Sigma_\gamma)\mathrm{Ga}(\eta_l^{*2} \mid a_\eta, b_\eta)\\
&\times\sum_{z:C,T}\prod_{h:\tilde{U}_{hz}=l}^{N_z+n_z}\{f_{\mathrm{LL}}(w_{ijz} \mid \lambda_l^*\exp(\boldsymbol{\gamma}_l^* z_{iz})/\xi_{iz}, \eta_l^*)\}^{\iota_{hz}}\\
&\times\{S_{\mathrm{LL}}(w_{ijz} \mid \lambda_l^*\exp(\boldsymbol{\gamma}_l^* z_{iz})/\xi_i, \eta_l^*)\}^{1-\iota_{hz}},
\end{aligned}$$

where ι_{hz} is an indicator if $j = N_{iz}$. A similar Pólya-Gamma augmentation algorithm

is applied here. $p(\lambda_l^*, \boldsymbol{\gamma}_l^*, \eta_l^* \mid -)$ is proportional to

$$
\begin{aligned}
& \mathrm{N}(\log(\lambda_l^*) \mid \mu_\lambda, \sigma_\lambda^2)\mathrm{N}(\boldsymbol{\gamma}_l^* \mid \mu_\gamma, \Sigma_\gamma)\mathrm{Ga}(\eta_l^{*2} \mid a_\eta, b_\eta) \\
& \times \sum_{z=ctr,trt} \prod_{h:\tilde{U}_{hz}=l} \left(\frac{\lambda_l^*}{w_{ijz}}\right)^{\iota_{hz}} \frac{[\exp\{\eta_l^*(\log(w_{ijz}) - \boldsymbol{\gamma}_l^* z_{iz} + \log(\xi_{iz}) - \log(\lambda_l^*)))\}]^{\iota_{hz}}}{[1 + \exp\{\eta_l^*(\log(w_{ijz}) - \boldsymbol{\gamma}_l^* z_{iz} + \log(\xi_{iz}) - \log(\lambda_l^*)))\}]^{\iota_{hz}+1}} \\
\propto\ & \mathrm{N}(\log(\eta_l^*) \mid \mu_\lambda, \sigma_\lambda^2)\mathrm{N}(\log(\lambda_l^*) \mid \mu_\lambda, \sigma_\lambda^2)\mathrm{Ga}(\eta_l^{*2} \mid a_\eta, b_\eta) \\
& \times \prod_{h:\tilde{U}_{hz}=l} \left(\frac{\lambda_l^*}{w_{ijz}}\right)^{\iota_{hz}} \int \exp(-0.5(\iota_{hz}-1)\varphi_{hlo}) \exp(-0.5\varsigma_{hz}\varphi_{hlo}^2) p(\varsigma_{hz}) d\varsigma_{hz},
\end{aligned}
$$

where $\varphi_{hlo} = \eta_l^*(\log(w_{ijz}) + \log(\xi_{iz}) + \log(\varsigma_{hz}) - \log(\lambda_l^*) - \boldsymbol{\gamma}_l^* z_{iz}))$, and $p(\varsigma_{hz})$ is a density function of $\mathrm{PG}(1 + \iota_{hz}, 0)$ evaluated at ς_{hz}. By introducing latent variables $(\varsigma_{1o}, \ldots, \varsigma_{N_z+n_z,o})$ for each gap time, we then first sample latent variables $(\varsigma_{1o}, \ldots, \varsigma_{N_z+n_z,o})$ from $\mathrm{PG}(1 + \iota_{hz}, \eta_{\tilde{U}_{hz}}^*(\log(w_{ijz}) + \log(\xi_{zi}) - \log(\lambda_{\tilde{U}_{hz}}^*) - \boldsymbol{\gamma}_{\tilde{U}_{hz}}^* z_{iz}))$ independently. Then the conditional on $(\varsigma_{1o}, \ldots, \varsigma_{N_z+n_z,o})$, draw $(\lambda_l^*, \boldsymbol{\gamma}_l^*, \eta_l^*)$, for $l = 1, \ldots, B_H$ from

$$
\begin{aligned}
\log(\lambda_l^*) \mid \boldsymbol{\varsigma}, - \ &\sim\ \mathrm{N}(\tilde{\mu}_{\lambda l}, \tilde{\eta}_{\lambda l}^2) \\
\text{where } \tilde{\sigma}_{\lambda l}^2 \ &=\ \left(\frac{1}{\sigma_\lambda^2} + \eta_l^{*2} \sum_{z=ctr,trt} \sum_{h:\tilde{U}_{hz}=l} \varsigma_{hz}\right)^{-1}, \\
\text{and } \tilde{\mu}_{\lambda l} \ &=\ \tilde{\sigma}_{\lambda l}^2 \Bigg(\frac{\mu_\lambda}{\sigma_\lambda^2} + \eta_l^{*2} \sum_{z=ctr,trt} \sum_{h:\tilde{U}_{hz}=l} \varsigma_{hz}(\log(w_{ijz}) - \boldsymbol{\gamma}_l^* z_{iz} + \log(\xi_{iz})) \\
& \qquad +0.5\eta_l^* \sum_{z=ctr,trt} \sum_{h:\tilde{U}_{hz}=l} (1 - \iota_{hz}))\Bigg),
\end{aligned}
$$

and

$$
\begin{aligned}
\boldsymbol{\gamma}_l^* \mid \boldsymbol{\varsigma}, - \ &\overset{ind.}{\sim}\ \mathrm{N}(\tilde{\mu}_{\gamma l}, \tilde{\Sigma}_{\gamma l}), \\
\text{where } \tilde{\Sigma}_\gamma \ &=\ \left(\Sigma_\gamma^{-1} + \eta_l^{*2} \sum_{z=ctr,trt} \sum_{h:\tilde{U}_{hz}} \varsigma_{hz} z_{iz} z_{iz}'\right)^{-1} \\
\text{and } \tilde{\mu}_{\gamma l} \ &=\ \tilde{\Sigma}_{\gamma l} \Bigg(\mu_\gamma \Sigma_\gamma^{-1} + \eta_l^{*2} \sum_{z=ctr,trt} \sum_{h:\tilde{U}_{hz}} u_{iz}(\log(w_{ijz}) - \boldsymbol{\gamma}_l^* z_{iz} + \log(\xi_{iz}) - \log(\lambda_l^*)) z_{iz} \\
& \qquad +0.5\eta_l^* \sum_{z=ctr,trt} \sum_{h:\tilde{U}_{hz}} (1 - \iota_{hz}) z_{iz}\Bigg),
\end{aligned}
$$

and

$$
\begin{aligned}
p(\eta_l^* \mid -) &\propto \text{Ga}(\tilde{a}_\eta, \tilde{b}_\eta) \times A_\eta \\
\text{where } \tilde{a}_\eta &= 0.5 \sum_{z=ctr,trt} \sum_{h:\tilde{U}_{hz}=l} \iota_{hz} + a_\eta, \\
\tilde{b}_\eta &= \left(0.5 \sum_{z=ctr,trt} \sum_{h:\tilde{U}_{hz}=l} \varsigma_{hz} (\log(w_{ijz}) + \log(\xi_{iz}) - \log(\lambda_l^*))^2 + \frac{1}{b_\eta} \right)^{-1}, \\
\text{and } A_\eta &= \exp\left(0.5 \eta_l^* \sum_{z=ctr,trt} \sum_{h:\tilde{U}_{hz}=l} (\iota_{hz} - 1) \eta_l^* (\log(w_{ijz}) + \log(\xi_{iz}) - \log(\lambda_l^*) - \boldsymbol{\gamma}_l^* z_{iz}) \right).
\end{aligned}
$$

If $\iota_{hz} = 1$ for all h and all o such that $\tilde{U}_{hz} = l$, draw η_l^* from $\text{Ga}(\tilde{a}_\eta, \tilde{b}_\eta)$ directly, otherwise, a Metropolis random-walk algorithm is used.

Updating hyperparameters in gap times model

$$
\begin{aligned}
p(b_\eta, \mu_\lambda, \mu_\gamma \mid -) \propto\ & \text{N}(\mu_\lambda \mid s_\lambda, S_\lambda) \text{N}(\mu_\gamma \mid s_\gamma, S_\gamma \text{inv-Ga}(b_\eta \mid r_\eta, R_\eta) \\
& \times \prod_{l=1}^{B_H} \text{N}(\log(\lambda_l^*) \mid \mu_\lambda, \sigma_\lambda^2) \text{N}(\boldsymbol{\gamma}_l^* \mid \mu_\gamma, \Sigma_\gamma) \text{Ga}(\eta_l^{*2} \mid a_\eta, b_\eta).
\end{aligned}
$$

We thus have

$$
\begin{aligned}
\mu_\lambda \mid - &\sim \text{N}\left(\left(\frac{1}{S_\lambda} + \frac{B_H}{\sigma_\lambda^2} \right)^{-1} \left(\frac{s_\lambda}{S_\lambda} + \frac{\sum_{l=1}^{B_H} \log(\lambda_l^*)}{\sigma_\lambda^2} \right), \left(\frac{1}{S_\lambda} + \frac{B_H}{\sigma_\lambda^2} \right)^{-1} \right), \\
\text{and} \mu_\gamma \mid - &\sim \text{N}\left(\left(S_\gamma^{-1} + B_H \Sigma_\gamma^{-1} \right)^{-1} \left(S_\gamma^{-1} s_\gamma + \Sigma_\gamma^{-1} \sum_{l=1}^{B_H} \boldsymbol{\gamma}_l^* \right), \left(S_\gamma^{-1} + B_H \Sigma_\gamma^{-1} \right)^{-1} \right) \\
\text{and } b_\eta \mid - &\sim \text{inv-Ga}\left(r_\eta + B_H a_\eta, R_\eta + \sum_{l=1}^{B_H} \eta_l^{*2} \right).
\end{aligned}
$$

Updating random effects The posterior distribution of random effect $(\epsilon_{iz}, \xi_{iz})$ is propor-

tional to

$$
\begin{aligned}
p(\epsilon_{iz}, \xi_{iz} \mid -) \quad \propto \quad & \left\{ f_{\mathrm{LL}}(\tau_{iz} \mid \theta^*_{L_{iz}} \exp(\boldsymbol{\beta}_{L_{iz}} x_{iz})/\epsilon_{iz}, \phi^*_{L_{iz}}) \right\}^{\nu_{iz}} \\
& \times \left\{ S_{\mathrm{LL}}(\tau_{iz} \mid \theta^*_{L_{iz}} \exp(\boldsymbol{\beta}_{L_{iz}} x_{iz})/\epsilon_{iz}, \phi^*_{L_{iz}}) \right\}^{1-\nu_{iz}} \\
& \times \prod_{j=1}^{N_{iz}} f_{\mathrm{LL}}(w_{ijz} \mid \lambda^*_{U_{ijz}} \exp(\boldsymbol{\gamma}_{U_{ijz}} z_{iz})/\xi_{iz}, \phi^*_{U_{ijz}}) \\
& \times S_{\mathrm{LL}}(w_{ijz} \mid \lambda^*_{U_{ijz}} \exp(\boldsymbol{\gamma}_{U_{ijz}} z_{iz})/\xi_{iz}, \phi^*_{U_{ijz}}) \\
& \times f_{\mathrm{LN}}((\epsilon_{iz}, \xi_{iz})' \mid (0,0), \Sigma_{ez}).
\end{aligned}
$$

Updating hyperparameter in random effects model We sample Σ_e from

$$
\text{inv-Wish}\left(c_e + n_z, C_e + \sum_{i=1}^{n_z} (\log(\epsilon_{iz}), \log(\xi_{iz}))(\log(\epsilon_{iz}), \log(\xi_{iz}))' \right).
$$

C.2 Sensitivity analysis

In Section 4.4, we analyzed the Esophageal cancer dataset with only treatment assignment. Here, we provide the sensitivity analysis to demonstrate the robustness of the model. The hyperparameters are adjusted as follows:

- Sensitivity 1. In the first scenario, we set $a_0 = b_0 = c_0 = a_1 = b_1 = c_1 = 3$, with scale parameters $\sigma^2_\theta = \sigma^2_\lambda = 1$, mean parameters $s_\theta = 0.45$, $s_\lambda = -0.1$, and their corresponding variances $S_\theta = S_\lambda = 0.5$. The variance for regression coefficients $\Sigma_\beta = \Sigma_\gamma = 1$ is accompanied by zero means and unit variances for the beta distributions. The shape parameters for the log-logistic distribution are set as $a_\phi = a_\eta = 4$, with inverse-Gamma parameters $r_\phi = r_\eta = 3$ and $R_\phi = R_\eta = 20$.

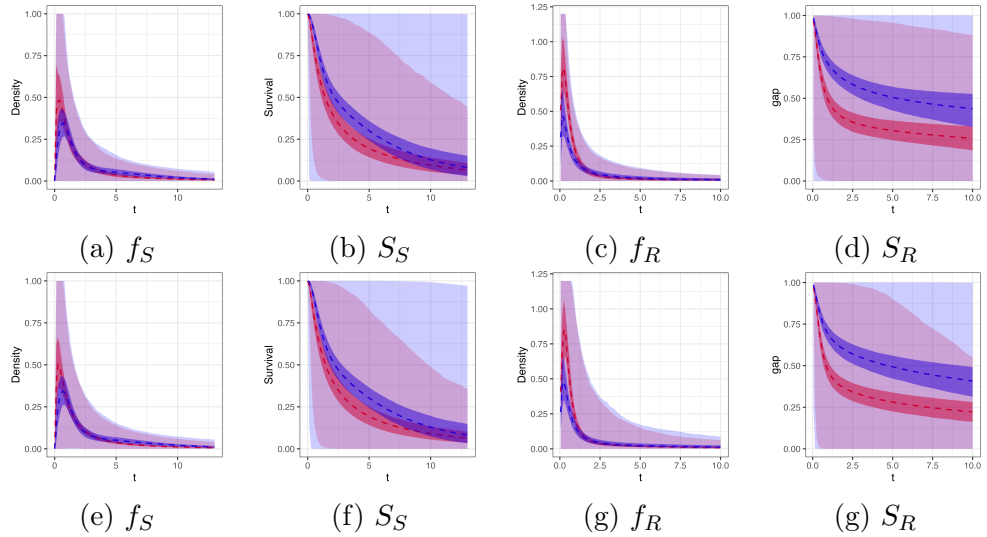
The covariance matrix Σ_{ez} is defined by an inverse-Wishart prior with $c_e = 18$ and $C_e = ((3,0)', (0,0.5)')$.

- Sensitivity 2. In the second setup, the hyperparameters are symmetrically set to 4 ($a_0 = b_0 = c_0 = a_1 = b_1 = c_1 = 4$), with scale parameters reduced to $\sigma_\theta^2 = \sigma_\lambda^2 = 0.75$. The mean parameters $s_\theta = 0.45$, $s_\lambda = 0$, and their variances $S_\theta = S_\lambda = 0.75$ are adjusted accordingly. Variance for regression coefficients $\Sigma_\beta = \Sigma_\gamma = 0.75$ is set with zero means. The shape parameters for the log-logistic distribution are $a_\phi = a_\eta = 3$, with inverse-Gamma parameters $r_\phi = r_\eta = 2$ and $R_\phi = R_\eta = 25$. The covariance matrix Σ_{ez} follows an inverse-Wishart prior with $c_e = 18$ and $C_e = ((3.5,0)', (0,0.7)')$.

These setups aim to explore the model's behavior under varying prior beliefs and parameter configurations.

C.3 Prior specification for parametric joint models and additional model comparison results

In Section 4.4, we compared proposed the proposed model with two parametric joint models. The first one is based on the renewal process assumption of the recurrent events while the second one makes a Poisson process assumption. Here, we give details on their prior specifications.

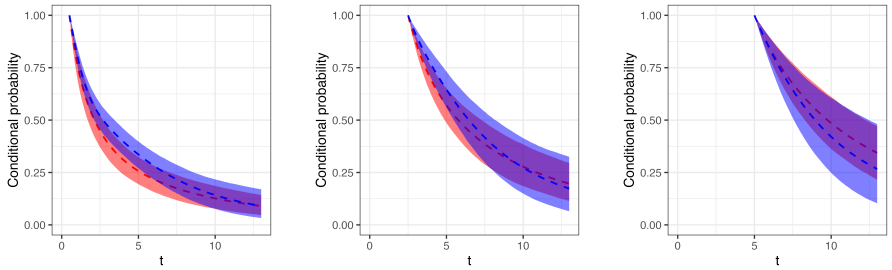


(a) f_S (b) S_S (c) f_R (d) S_R

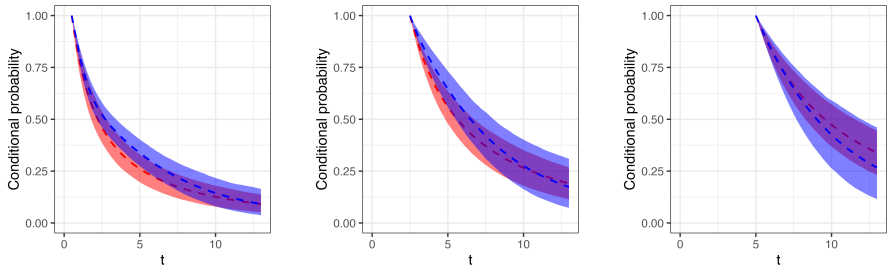
(e) f_S (f) S_S (g) f_R (g) S_R

Figure C.1: The sensitivity analysis of the EC dataset is illustrated in this figure, presenting posterior mean estimates (dashed lines) and 95% pointwise credible intervals for posterior (dark shaded regions) and prior (light shaded regions) estimates. The first and second columns display the density and survival function for survival times, respectively, while the third and fourth columns show the density and survival function for gap times. The figures in the first and second rows are based on priors from sensitivity analyses 1 and 2, respectively. All estimates are derived after marginalizing over the posterior and prior random effects models. The 3DCRT group is represented in red and the IMRT group in blue.

Parametric joint model with renewal process assumption For the joint model with a renewal process recurrent events assumption, the model is formulated in the following way. Conditional on random effects, both models for survival and gap times are univariate log-logistic distributions. For $i = 1, \ldots, n$, the hierarchical representation



(1) $t_0 = 0.5$ (2) $t_0 = 1$ (3) $t_0 = 2$

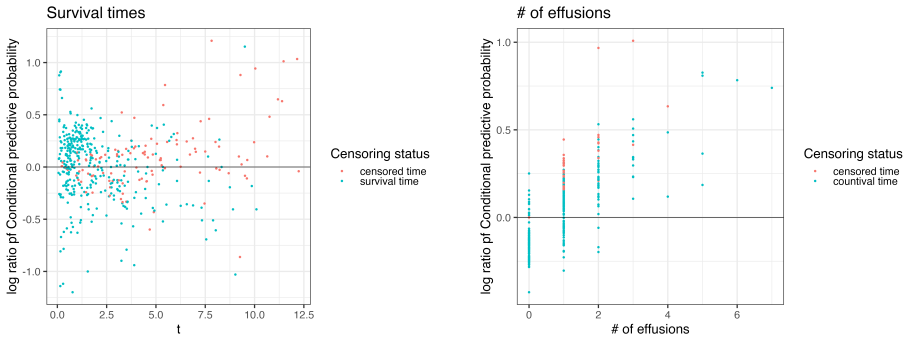


(7) $t_0 = 0.5$ (8) $t_0 = 1$ (9) $t_0 = 2$

Figure C.2: The sensitivity analysis of the EC dataset is illustrated in this figure, presenting posterior mean estimates (dashed lines) and 95% pointwise credible intervals for posterior (dark shaded regions) and prior (light shaded regions) estimates. From left to right, each column displays the conditional survival probability given no occurrence of effusions with $t_0 = 0.5$, 1, 2, respectively. All estimates are derived after marginalizing over the posterior and prior random effects models. The 3DCRT group is represented in red and the IMRT group in blue.

of the model is written as:

$$
\begin{aligned}
T_i \mid \epsilon_i \boldsymbol{x}_i, \theta, \boldsymbol{\beta}, \phi &\overset{ind.}{\sim} \mathrm{LL}(\theta \exp(\boldsymbol{\beta}\boldsymbol{x}_i)/\epsilon_i, \phi), \\
W_{ij} \mid \xi_i, \boldsymbol{x}_i, \lambda, \boldsymbol{\gamma}, \eta &\overset{i.i.d.}{\sim} \mathrm{LL}(\lambda \exp(\boldsymbol{\gamma}\boldsymbol{x}_i)/\xi_i, \eta), \text{ for } j = 1, \ldots, N_i, \\
(\epsilon_i, \xi_i) \mid \Sigma_e &\overset{i.i.d.}{\sim} \mathrm{LN}((0,0)', \Sigma_e)
\end{aligned}
$$



(a) survival times (b) number of effusions

Figure C.3: EC data with All Covariates Leave-one-out Cross-validation. This figure presents the log-transformed ratio of conditional predictive probabilities of survival times and the number of effusions between the joint-DDP and joint-Poisson models in panels (a) and (b), respectively.

We assign prior distributions to each parameter as $\theta \sim \text{LN}(s_\theta, S_\theta)$, $\boldsymbol{\beta} \sim \text{N}(s_\beta, S_\beta)$, $\phi \sim \text{Ga}(a_\phi, b_\phi)$, $\lambda \sim \text{LN}(s_\lambda, S_\lambda)$, $\eta \sim \text{Ga}(a_\eta, b_\eta)$, and $\Sigma_e \sim \text{inv-Wish}(c_e, C_e)$.

Applying the model to the Esophageal cancer dataset, we specify the prior hyperparameter values are following, $s_\theta = 0$, $S_\theta = 100$, $s_\beta = \mathbf{0}_6$, $S_\beta = 100\text{I}_6$, $a_\phi = 1$, $B_\phi = 1$, $s_\lambda = 0$, $S_\lambda = 100\text{I}_6$, $a_\eta = 1$, $b_\eta = 1$, and $c_e = 18$, and $C_e = 4\text{I}_2$.

Parametric joint model with Poisson process assumption Conditional on bivariate random effects, the joint model can be constructed in the following hierarchical way, for $i = 1, \ldots, n$,

$$T_i \mid \epsilon_i, \boldsymbol{x}_i, \boldsymbol{\beta}, \phi \overset{ind.}{\sim} \text{Weib}(\boldsymbol{\beta}\boldsymbol{x}_i/\epsilon_i, \phi),$$

where the survival function of the Weibull distribution is $\exp(-t_i\epsilon_i/\exp(\boldsymbol{\beta}\boldsymbol{x}_i))^\phi)$. We use a parametric Poisson process intensity function as $\xi_i\eta y_{ij}^{(}\eta - 1)\exp(\boldsymbol{\beta}\boldsymbol{x}_i))$, for $j =$

$1, \ldots, N_i$. Thus, the likelihood function for all subjects under the parametric joint model with Poisson process assumption is then,

$$\begin{aligned} L &= \prod_{i=1}^{n} \left(\prod_{j=1}^{N_i} \xi_i \exp(\boldsymbol{\gamma x}_i) \eta y_{ij}^{\eta-1} \right) \exp(-\xi_i \lambda \exp(\boldsymbol{\gamma x}_i) t_i^{\eta}) \\ &\quad \{ f_{\text{Weib}}(t_i \mid \theta \exp(\boldsymbol{\beta x}_i)/\epsilon_i), phi) \}^{\nu_i} \{ S_{\text{Weib}}(t_i \mid \theta \exp(\boldsymbol{\beta x}_i)/\epsilon_i), phi) \}^{1-\nu_i}, \end{aligned}$$

where f_{Weib} and S_{Weib} are density and survival functions of a Weibull distribution, respectively.

The prior distributions for each parameter are follows, $\theta \sim \text{LN}(s_\theta, S_\theta)$, $\boldsymbol{\beta} \sim \text{N}(s_\beta, S_\beta)$, $\phi \sim \text{Ga}(a_\phi, b_\phi)$, $\lambda \sim \text{LN}(a_\eta, b_\eta)$, and $\Sigma_e \sim \text{inv-Wish}(c_e, C_e)$. The prior hyperparameters are set as $a_\phi = 2$, $b_\phi = 1$, $s_\theta = \mathbf{0}_7$, $S_\theta = 100$, $s_\beta = \mathbf{0}_7$, $S_\beta = 100\text{I}_7$, $a_\eta = 2$, $b_\eta = 1$, $s_\lambda = 0$, $S_\gamma = 100\text{I}_7$, $c_e = 18$, and $C_e = 4\text{I}_2$.

Figure C.3 illustrates log-transformed ratio of conditional predictive probabilities between joint-DDP and joint-parametric-renewal models for survival times and count of effusion events in panels (a) and (b), respectively. Overall, the joint-DDP model outperformed the joint-parametric-renewal model for both survival times and effusion counts predictions. Specifically, the joint-DDP model has higher CPP^{S} values for censored times and higher values of CPP^{R} values for instances with more than one effusion event.

C.4 Additional results with survival regression model

In Section 4.2, we presented a joint modeling framework that integrates survival times with the gap times of recurrent events, utilizing bivariate random effects to

account for the dependence between these two processes. Here, we focus specifically on the submodel for survival times. This involves removing the components related to recurrent events and random effects, resulting in a standalone model that we refer to as the survival-DDP model. By comparing the performance of the survival-DDP model with the full joint-DDP model, we aim to underscore the distinct advantages that joint modeling offers in the context of survival analysis, particularly in its ability to leverage additional information from recurrent events to improve predictive accuracy.

The hyperparameters are set to be the same with the corresponding ones in the joint model. Specifically, we let $a_\phi = 3$, $b_\phi = 6$, $R_\phi = 30$, $\sigma_\theta^2 = 2$, $s_\theta = 0.45$, $S_\theta = 1$, $\Sigma_\beta = 2\mathrm{I}_5$, $s_\beta = \mathbf{0}_5$, $S_\beta = 2\mathrm{I}_5$, along with $a_0 = 2$, $b_0 = 2$, and $c_0 = 2$.

As discussed in Section 4.4.2, we computed the sum of the log conditional predictive probabilities for survival times to assess model performance. The results for the various joint models are presented in Table 4.1. A larger sum indicates better model performance. The survival-DDP model achieves a value of -1621.07, outperforming the parametric joint models, which scored -1628.1498 and -1672.2402, respectively, in predicting survival times. However, our joint-DDP model, with a score of -1610.9244, demonstrated the best performance. This result highlights the joint model's ability to effectively incorporate information from the recurrent events data, leading to enhanced predictive power.