On Bayesian prediction under nonparametric transformation models with doubly censored data

CHEN SHOUZHENG

MPhil

The Hong Kong Polytechnic University

2024

The Hong Kong Polytechnic University

Department of Applied Mathematics

On Bayesian prediction under nonparametric transformation models with doubly censored data

Chen Shouzheng

A thesis submitted in partial fulfilment of the requirements for the degree of Master of Philosophy

June 2024

CERTIFICATE OF ORIGINALITY

I hereby declare that this thesis is my own work and that, to the best of my knowledge and belief, it reproduces no material previously published or written, nor material that has been accepted for the award of any other degree or diploma, except where due acknowledgement has been made in the text.

__________________________ (Signed)

Chen Shouzheng ______________ (Name of student)

Dedication

Dedicated to my parents for their unconditional, unreserved, and unwavering love.

Abstract

This MPhil thesis extends, in two main ways, a previous framework on the predictions of right censored survival outcomes under nonparametric transformation models. The first extension is regarding censoring scheme complexity. Specifically, we extend from right censoring to double censoring, which contains both left and right censoring. The previous work relied on proposing a weakly informative prior for the transformation function to mitigate model unidentifiability, and we heavily modify this prior so that nonparametric transformation models can be implemented under both random and fixed double censoring. By comparing our predictions results to two leading methods, we demonstrate that the proposed approach is computationally effective under double censoring and successfully utilizes a robust and flexible nonparametric transformation model. The second extension is regarding model complexity. Specifically, we extend from nonparametric transformation models, which contain two unidentified infinite-dimensional parameters to a certain model with more than two such parameters. For this more sophisticated model, we attempt to adjust priors to control posterior Markov Chain Monte Carlo (MCMC) mixing. Numerical illustrations show that carefully chosen priors can indeed mitigate poor mixing under the more complex model. Subsequently, we consider how *weak* a prior can be to still allow for well-mixed MCMC

chains. Driven by this question, we explore the concept of informativeness further and speculate to quantify it through a mathematical definition. This definition may contribute to a criterion for identifying priors that are sufficiently *informative*, as to potentially help address poor mixing generally for models with more than two unidentified infinite-dimensional parameters.

Acknowledgements

Standing at yet another turning point in my life, I realize now more than ever how fortunate I am to even be here. I have had many people along the way who are willing to support me, guide me, and share with me what they have learned about the world, and for these people, I will be eternally grateful.

I would like to express my deep gratitude to my parents and family for providing me to explore other cultures and broaden my perspectives. I honestly cannot imagine my life without them.

However, with regards to my studies and my journey toward an MPhil degree in the past two years, the person I would like to thank the most is, without a doubt, my chief supervisor, Dr. Catherine C. Liu. I am not exaggerating when I say Dr. Liu devotes the majority of her time to the success of her students. She cares greatly about her students on an academic and also a personal level. Over the past two years, she has spent night and day on educating me to be a better learner, a better researcher, and a better personal for society. Through her, I have gradually come to realize the steps needed to grow into an independent scholar and researcher. I witness in awe her work ethic and what she has accomplished and can only aspire to possess a fraction of her qualities in the future.

Another special acknowledgement must go out to Dr. Chong Zhong for his insights and guidance on my research. I sincerely appreciate his patience despite me being a novice, and indeed I have learned a great deal from our discussions. My research progress in the these past years is much beholden to his generous help. I also wish to express my gratitude for my co-supervisor, Dr. James Lee, for teaching me concepts about survival analysis and encouraging me to explore the research direction I have chosen.

Last but not least, I would like to thank my friends, here and back home. They have always been there for me when I am down, leading a listening ear when I need, and sharing their candidates thoughts for my personal development. I will cherish their friendships and make every effort to be there for them.

I have been so fortunately to have these people in my life to lift me up and lead by examples. One day, I hope to be for others what they are for me.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Motivation

This MPhil thesis is motivated by Part II, Chapter three of Dr. Chong Zhong's PhD thesis (Zhong, 2023), which succeeded in predicting right censored survival outcome under nonparametric transformation models (referred to as "Chapter three" below for simplicity). Given the effectiveness of his methods, we are inspired to consider two points of further development. The first line of thought is whether we may extend his methodology to predict doubly censored survival outcomes, a more general and complex censoring scheme; and the second is to examine root causes behind the successful implementation of his methodology.

Why: The first reason for conducting our work is that nonparametric transformation models are more flexible and robust than parametric and semiparametric models, yet there has not been any success in extant literature to implement such models under the double censoring scheme (refer to Section 1, Chapter 2 for specific discussions). The second reason is although it was demonstrated in Chapter three that weakly informative priors can facilitate computations, any further exploration of how and when weakly informative priors may be of help was to be desired.

Challenges: The first critical challenge revolves around the fact that double censoring consists of both right and left censoring, which means when attempting to apply nonparametric transformation models to doubly censored data, one must devise ways to analyze left censored data while maintaining integrity of the original method. Consequently, one major contribution in Chapter three that lends viability to his methodology, the *Quantile-knot I-splines prior*, needs to be reconstructed as it cannot account for changes in observed data when more than one type of censoring is present. Extra precaution was given with the elicitation and any modification of this prior, since it is the prior for the transformation function, the most integral part of any transformation model, and can have the greatest impact on posterior inference.

The other apparent challenge is the ambiguous concept of *weakly informative priors*. The class of informative priors refers to prior distributions that are constructed by incorporating prior knowledge from experience, expert opinion, or similar studies (Reich and Ghosh, 2019). In mathematical expressions, let $p(\theta|m)$ be a prior distribution of a model m with the parameter θ, then such a prior distribution may be called "informative" if θ is biased towards particular values. Weakly informative priors, as a subclass of informative priors, can therefore be considered as prior distributions with mild prior knowledge. However, it is worth noting that although definitions such as "(the prior) is set up so that the information it does provide is intentionally weaker than whatever actual prior knowledge is available" (Gelman, 2006) has been offered, no exact definition in mathematical terminologies has been provided for the class of weakly informative priors. In other words, it remains unclear at what point the influence of weakly informative priors, exerted by pre-existing information or beliefs, becomes insignificant on posterior distributions.

1.2 Our work in this thesis

The first objective of this thesis is to implement nonparametric transformation models to analyze doubly censored data under the Bayesian framework. It should be noted that we consider two censoring schemes within double censoring which are both important. Random censoring is commonly discussed in survival analysis, as it is theoretically interesting and convenient. On the other hand, fixed censoring may be more challenging to analyze and compute, as less information is available. Fixed censoring, as the name implies, is a censoring scheme when all time-to-events are censored at one identical time; and in the context of double censoring, it means that all left censored observations will have one same numerical value, and all right censored observations another.

The second objective of this thesis revolves around the poor mixing issue, a serious problem in Bayesian computation caused by unidentifiability. It is well known that the posterior of an unidentified model can have multiple modes, which are difficult to be fully explored through MCMC sampling (Brooks et al., 2011). If the MCMC cannot satisfactorily explore the multi-modal posterior, poorly-mixed chains occurs. A direct consequence of poor mixing is an erroneous approximation of the posterior prediction distributions (PPD), yielding unreliable predictions. Regarding multi-modal posterior, researchers either developed feasible *samplers* for the multimodal posterior (Neal (1996); Kou et al. (2006); Pompe et al. (2020); Jacob et al. (2022); among others), or conducted *post-process* on *posterior* samples to obtain the posterior of identified parameters through general samplers (McCulloch and Rossi (1994); Gelfand and Sahu (1999); Burgette et al. (2021); among others). We consider an alternative approach where the posterior will not be directly modified but rather indirectly influenced through careful prior elicitation. A certain class of priors that we believe could mitigate poor mixing is suggested and applied to an extension of nonparametric transformation models. Subsequently, we attempt to define such priors more clearly in mathematical expressions.

1.3 Organization of thesis

In the next Chapter, we will discuss more specifically the background, application, and performance of a certain class of nonparametric transformation models in the context of doubly censored survival analysis. Methodologically, we go into the details of developing a novel type of weakly informative priors, which is equipped to analyze doubly censored time-to-events under two distinct yet similarly crucial censoring schemes.

We propose an extension of nonparametric transformation models in Chapter 3 to include more than two unidentified infinite-dimensional parameters. A feasible approach to mitigate poor mixing under this new class of models is discussed and illustrated. As a theoretical exploration, we go more in depth on the concept of *weakly informative priors* and attempt to offer a mathematical definition that quantifies the degree of *informativeness* needed to ensure discernibility. It has been suggested in literature that priors with a sufficient level of informativeness can mitigate poor mixing (Branscum et al., 2008), and although weakly informative priors could contain the necessary informativeness, the term is simply too broad and lacks mathematical rigor. Therefore, major benefit could be reaped from enumerating informativeness levels and constructing thresholds for weakly informative priors to be weaker than informative priors yet still powerful enough.

In Chapter 4 of this MPhil thesis, we summary our work so far on utilizing the Bayesian framework to analyze doubly censored survival data, most notably, actualizing nonparametric transformation models and contributing a concrete definition of *moderately weak informative priors*. Future directions of research following our work in Chapters 2 and 3 are also discussed in length.

Chapter 2

Extension to double censoring

2.1 Background

Survival analysis, a long-explored area of research related to medicine, social sciences, and statistics, primarily concerns how long it will take for particular events of interest to take place (time-to-event). However, such seemingly simplistic goals are often muddled by the inability to directly observe events of interest. In situations where some (or even all) of the events cannot be observed, the missing of key information leads to the collected data being censored. The most frequently encountered type of censoring in practice is right censoring, where subjects of interest are no longer followed after a certain time point, making events that occur afterward impossible to observe. Similarly, another common type of censoring is left censoring, where events occur before subjects of interest are followed. Therefore, it becomes apparent that a more complex and more challenging censoring scheme could arise from the combination of the two types of censoring mentioned above. This more complicated type of censoring is called ”double censoring”, which refers to when collected observations can be divided into three groups: exact observations of events, left censored observations, and right censored observations. Concisely, double censoring occurs when the event of interest, for each subject in the data, can only be observed within a certain time frame. It thus follows logically that for such doubly censored data, useful information can be highly limited, since, for unobserved events, it is only known whether the time-to-events are smaller than some left censoring endpoints or greater than some right censoring endpoints. Thus, it is quite obvious that making accurate estimations and predictions regarding doubly censored data will be extremely difficult. Additionally, to avoid any confusion, it should be noted that double censoring has also been used in literature to describe another type of censoring scheme in survival analysis. Under this second definition, the elapsed time between two related events, which they themselves may be either right or interval censored, becomes the event of interest (Sun, 2006). We will focus solely on the first definition of double censoring in this thesis.

In real life, doubly censored data often arise in biomedical and health science fields such as pharmacology and epidemiology. One such example results from a randomized AIDs clinical trial was conducted in 1997. The study aimed to compare HIV-infected children's responses to three different treatments. One major endpoint in the study was the plasma HIV-1 RNA level, and researchers relied on its values to accurately measure the treatment efficacy. Plasma HIV-1 RNA level, given by the NucliSens assay, is highly unreliable below 400 or above 75,000 per mililiter of plasma; therefore, it can be treated as a doubly censored variable only observable between 400 and 75,000.

Survival analysis, as a whole, has been fairly well explored with quite an abundance of extant literature. However, as previous research largely focused on the most common types of censoring, especially right censoring, there still exist significant gaps in certain sub-fields. In particular, in the case of double censoring, current statistical methods are still underdeveloped and often inefficient. The concept of double censoring was first introduced by Gehan in 1965, where he studied and extended the two-sample Wilcoxon test to doubly censored data. According to Gehan (1965), ”doubly censored data” refer to group(s) of observations where each observation has a probability to be either right censored, left censored, or uncensored. A few researchers adopted this definition and continued to either expand upon Gehan’s initial work or explore other intriguing directions. For example, Mantel (1967) and Hughes (2000) also considered the problem of two-sample comparison and further modified Gehan’s proposed test. Others, such as Turnbull (1974), proposed a ”self-consistent” procedure to obtain the nonparametric maximum likelihood estimator (NPMLE) in 1974 and focused on estimating the survival function of doubly censored data. Similarly, Chang (1990) studied the weak convergence of the NPMLE, Gu and Zhang (1993) studied the asymptotic properties of the NPMLE, Mykland and Ren (1996) and Zhang and Jamshidian (2004) developed algorithms algorithms to compute the NPMLE. One common feature of the above mentioned earlier research on double censored data is that they do not consider the situation where covariates are involved, which may add to the degree of complexity even further.

When there exist covariates within the doubly censored data, new techniques are often needed to correctly estimate the properties and parameters. An effective way to analyze doubly censored data with covariates is to apply transformation models. Transformation models help explain the relationship between a function of the time-to-event and the covariates in the data. The most commonly used transformation models in survival analysis are the semiparametric models, particularly the popular proportional hazards model, the proportional odds model, and the accelerated failure times model. The popularity of these models is largely attributed to their superior flexibility over the simpler and more restricted parametric models. Such research prominence is also evident in the context of double censoring. For example, Cai and Cheng (2004) extended a class of semiparametric transformation models to study the effects of covariates on failure time. Similarly, Li et al. (2018) and Choi and Huang (2021) considered the nonparametric maximum likelihood estimation of such semiparametric transformation models. Yet more flexible is another class of models, the nonparametric transformation models, which given their robustness to model misspecifications, can perform especially well in practical settings. However, two major challenges are encountered when attempting to apply such nonparametric transformation models, namely infinite-dimensional parameters and model unidentifiability. Specifically, model unindentifiability refers to when different sets of parameters in the model can generate an identical likelihood. Existing approaches to solve model unidentifiability are mainly divided into two schools of thought: making the model identifiable by imposing constraints (Chen (2002); Ye and Duan (1997); Chiappori et al. (2015); among others) or circumventing unidentifiability by making strong a priori assumptions (Ding and Nan (2011); Zeng and Lin (2007); Zhou and Hanson (2018); among others). It should be noted that nonparametric transformation models have not yet been utilized to analyze doubly censored data. The reluctance can be understood since previously mentioned approaches to tackle model unidentifiability either sacrifies computational feasibility or consistency. Therefore, an important issue to be addressed is how to solve the problem of model unidentifiability effectively, so that nonparametric transformation models can be applied on doubly censored data.

The remainder of this chapter is organized as follows. We introduce our data structure and model in Section 2. Our proposed innovative priors will be explained in detail in Section 3. Posterior inference and estimation will be explored in Section 4. Simulation results will be presented in Section 5. We also apply our method to real data in Section 6. A discussion will be given in Section 7.

2.2 Data, model, and assumptions

2.2.1 Data structure

Here we describe the typical data structure of doubly censored data in survival analysis. Consider a study that involves n independent subjects. For subject i, let T_i denote the time-to-event and Z_i be the p-dimensional vector of time-invariant covariates. The time-to-event T_i can only be observed between L_i and R_i, and if not observed, it is either left censored at L_i or right censored at R_i.

Define $\delta_{i1} = I(T_i \leq L_i)$, $\delta_{i2} = I(L_i < T_i \leq R_i)$, $\delta_{i3} = I(R_i < T_i)$, where $I(\cdot)$ is the indicator function. Then it follows $\delta_{i1} + \delta_{i2} + \delta_{i3} = 1$. The observed data are of the form $\{(\tilde{T}_i, L_i, R_i, Z_i, \delta_{i1}, \delta_{i2}, \delta_{i3}); i = 1, \ldots, n\}$, where $\tilde{T}_i = \max\{L_i, \min(R_i, T_i)\}$ is the observed time-to-event for subject i.

Here we assume that $L_i = 0$ if $\delta_{i3} = 1$ and $R_i = \infty$ if $\delta_{i1} = 1$, since such information is generally not available, for better data organization. Furthermore, T_i and (L_i, R_i) are assumed to be conditionally independent given Z_i (noninformative censoring) as common practice.

2.2.2 Nonparametric transformation models

We consider a class of linear transformation models, which relate the time-to-event to the relative risk in a multiplicative way.

$$H(T) = \xi \exp(\beta^{\mathrm{T}} Z), \tag{2.1}$$

where $H(\cdot)$ is a strictly increasing transformation function that is positive on $\mathbb{R}^+$, β is the p-dimensional vector of regression coefficients coupling Z, and ξ is the model error with distribution function F_ξ. The above transformation model is considered a nonparametric transformation model when the functional forms of both $H(\cdot)$ and F_ξ are unknown. As mentioned earlier, when model (2.1) is nonparametric, model unidentifiability will be encountered, which means that different sets of (H, β, F_ξ) can generate an identical likelihood function. Mathematically, suppose model (2.1) holds for a special triplet solution (H_0, β_0, ξ_0), then model (2.1) also holds on the set $\mathcal{C}\{(H, \beta, \xi)\} = \{(c_1 H_0^{c_2}, c_2\beta_0, c_1\xi_0^{c_2})\}$ for any pair of positive constants $(c_1, c_2) \in \mathbb{R}_+^2$. Consequently, the joint posterior has uncountable many modes on $\mathcal{C}\{(H, \beta, \xi)\}$. For the rest of this chapter, model (2.1) will be treated as a nonparametric transformation model and referred to as the NTM.

The NTM is obtained by applying an exponential transformation on a class of linear transformation models with additive relative risk.

$$h(T) = \beta^{\mathrm{T}} Z + \epsilon, \tag{2.2}$$

where $h(\cdot) = \log(H(\cdot))$ and $\epsilon = \log(\xi)$. This transformation is necessary since the transformation function $h(\cdot)$ in model (2.2) is sign-varying on $\mathbb{R}^+$, which leads to insoluble problems regarding prior elicitation and posterior sampling. After the transformation, $H(\cdot)$ is strictly positive on $\mathbb{R}^+$, thus allowing the NTM to avoid the above problems.

2.2.3 Assumptions

We will now state some generate assumptions for doubly censored data and the NTM.

(A1) The transformation function $H(\cdot)$ is differentiable.

(A2) The model error ξ is continuous.

(A3) The covariate Z is independent of ξ.

(A1) is required due to the $H'(\cdot)$ in the likelihood function. (A2) is mild. (A3) is general for transformation models.

2.3 Likelihood and priors

2.3.1 Likelihood function

Given observed data $\{(\tilde{T}_i, L_i, R_i, Z_i, \delta_{i1}, \delta_{i2}, \delta_{i3}); i = 1, \ldots, n\}$, we can construct the likelihood function as

$$\begin{aligned} &\mathcal{L}\left(H, f_\xi, S_\xi, \beta \mid \tilde{T}, Z, \delta_1, \delta_2, \delta_3\right) \\ &= \prod_{i=1}^{n} \left[F_\xi\left\{H\left(\tilde{T}_i\right) e^{-\beta^{\mathrm{T}} Z_i}\right\}\right]^{\delta_{i1}} \\ &\times \left[f_\xi\left\{H\left(\tilde{T}_i\right) e^{-\beta^{\mathrm{T}} Z_i}\right\} H'\left(\tilde{T}_i\right) e^{-\beta^{\mathrm{T}} Z_i}\right]^{\delta_{i2}} \\ &\times \left[S_\xi\left\{H\left(\tilde{T}_i\right) e^{-\beta^{\mathrm{T}} Z_i}\right\}\right]^{\delta_{i3}}, \end{aligned} \tag{2.3}$$

where $S_\xi = 1 - F_\xi$ is the tail distribution of ξ.

2.3.2 Pseudo-Quantile I-splines prior

Regarding the transformation function of the NTM and its derivative, we rely on a type of I-spline priors to capture the relevant information. To construct such priors, we first take $\tau = \max(\tilde{T})$ to be the largest observed time-to-event in the collected data, then $D = (0, \tau]$ is the interval that contains all observed time-to-events. Note that $H(\cdot)$ is differentiable on D, thus we can model $H(\cdot)$ and $H'(\cdot)$ by

$$H(t) = \sum_{j=1}^{K} \alpha_j B_j(t), H'(t) = \sum_{j=1}^{K} \alpha_j B_j'(t), \tag{2.4}$$

where $\{\alpha_j\}_{j=1}^K$ are positive coefficients, $\{B_j\}_{j=1}^K$ are I-spline basis functions on D, and $\{B_j'\}_{j=1}^K$ are the corresponding derivatives.

The number of I-spline basis functions $K = N+r$, where N is the total number of interior knots and r is the order of smoothness with $(r-1)$th order derivative existing. Also, the intercept $H(0) = 0$ is set.

Then it becomes our primary task to specify the exact number of interior knots and pinpoint their locations, and one logical way to approach this task is to base the selection of interior knots on empirical quantiles of the collected data. In doing so, we can effectively utilize any useful knowledge inherent to the distribution of the real and observed time-to-events.

Let $\hat{F}_X(t) = n^{-1} \sum_{i=1}^n I(X_i \leq t)$ be the empirical CDF of some X and $\hat{Q}_X(p) = \inf\{t : p \leq \hat{F}_X(t)\}$ be the corresponding empirical quantile function, where X can be equivalently replaced by T, $\tilde{T}$, and other similar random variables. Note that since the real time-to-events T cannot always be observed, $\hat{F}_T(\cdot)$

and $\hat{Q}_T(\cdot)$ can only be constructed based on a part of the collected data where $\delta_{i2} = 1$ (i.e., when real time-to-events are observed). We first consider knot selection via empirical functions under the random censoring setting, which is very often the assumed setting in related literature.

Random Censoring Knot Selection

Define $\tilde{T}_L = \{\tilde{T}_i \in \tilde{T} : \delta_{i3} = 0\}$ and $\tilde{T}_R = \{\tilde{T}_i \in \tilde{T} : \delta_{i1} = 0\}$. Let N_I be the initial number of knots. The interior knots selection procedure can be described as follows.

Step 1: Choose N_I empirical quantiles of real time-to-events as interior knots, where each knot $t_j = \hat{Q}_T\{j/(N_I - 1)\}$ and $j = 0, \ldots, N_I - 1$, such that $0 < t_0 < \cdots < t_{N_I} - 1 \leq \tau$.

Step 2: For $j = 0, \ldots, N_I - 1$, if $|\ \hat{F}_T(t_j) - \hat{F}_{\tilde{T}_L}(t_j)\ | \geq 0.05$, interpolate a new knot $t_j^* = \hat{Q}_{\tilde{T}_L}(j/(N_I - 1))$.

Step 3: For $j = 0, \ldots, N_I - 1$, if $|\ \hat{F}_T(t_j) - \hat{F}_{\tilde{T}_R}(t_j)\ | \geq 0.05$, then interpolate another new knot $t_j^{**} = \hat{Q}_{\tilde{T}_R}(j/(N_I - 1))$.

Step 4: Sort all the chosen and interpolated knots $\{t_0, \ldots, t_j, t_j^*, t_j^{**}, \ldots, t_{N_I - 1}\}$ from smallest to largest in numerical value and output the sorted series as the finally selected interior knots.

It is worth noting that only real time-to-events can provide information about H', therefore the initial interval knots are chosen by equally spaced empirical quantiles of T. To mitigate the lack of information when the percentage of left or right censored observations is high, extra interior knots are generated as needed.

The problem of interior knot selection becomes much more complex and difficult under fixed censoring. In such circumstance, the empirical distributions of observed time-to-events would heavily gravitate toward the fixed censoring points, making interpolation of additional interior knots infeasible. Therefore, in cases of high censoring, attempts have to be made to extract some information from the unobserved time-to-events. Thus, we propose a novel method for effective interior knots selection that synthesizes pseudo data to mimic the distribution of unobserved time-to-events. This innovative method then leads to a new type of prior for the transformation function, which we name to be the ”Pseudo-Quantile I-splines prior” (PQI prior).

Fixed Censoring Knot Selection

Fixed censoring occurs when $L_{i_1} = L$ for $i_1 = 1, \dots, n_1$ and $R_{i_2} = R$ for $i_2 = 1, \dots, n_2$, where n_1 and n_2 are the numbers of left censored and right censored observations, respectively. In this scenario, define $n_3 = n - n_1 - n_2$, $\delta_{i1} = I(T_i \leq L)$, $\delta_{i2} = I(L < T_i \leq R)$, and $\delta_{i2} = I(T_i > R)$.

The specification procedure can be described as follows.

For $k = 1, \dots, K$ (steps 1-3),

Step 1 (pseudo left censored data generation):

Generate pseudo observations $(\mathcal{T}_{L_{k1}}, \dots, \mathcal{T}_{L_{ki_1}}, \dots, \mathcal{T}_{L_{kn_1}})$ from some distribution (e.g. weibull, gamma) such that all $\mathcal{T}_{L_{ki_1}} < L$.

Step 2 (pseudo right censored data generation):

Generate pseudo observations $(\mathcal{T}_{R_{k1}}, \dots, \mathcal{T}_{R_{ki_2}}, \dots, \mathcal{T}_{R_{kn_2}})$ from the same distribution such that all $\mathcal{T}_{R_{ki_2}} > R$.

Step 3 (pseudo quantile computation):

Let $\mathcal{T}_k = (\mathcal{T}_{L_{k1}}, \dots, \mathcal{T}_{L_{kn_1}}) \cup (T_1, \dots, T_{n_3}) \cup (\mathcal{T}_{R_{k1}}, \dots, \mathcal{T}_{R_{kn_2}})$. Compute $\hat{F}_{\mathcal{T}_k}(t) = n^{-1} \sum_{i=1}^{n} I(\mathcal{T}_{ki} \leq t)$ and $\hat{Q}_{\mathcal{T}_k}(p) = \inf\{t : p \leq \hat{F}_{\mathcal{T}_k}(t)\}$.

Step 4 (quantile averaging):

Compute $\hat{Q}_{\mathcal{T}}(p) = K^{-1} \sum_{k=1}^{K} \hat{Q}_{\mathcal{T}_k}$. Choose N averaged empirical quantiles of the combined time-to-events as interior knots, where each knot $t_j = \hat{Q}_{\mathcal{T}}(j/(N-1))$ and $j = 0, \dots, N-1$. Output this series $\{t_0, \dots, t_j, \dots, t_{N-1}\}$ as the finally selected interior knots.

To ensure the pseudo data can accurately reflect any valuable knowledge hidden in the unobserved time-to-events, data are generated iteratively to improve their representativeness. They are also combined with the real time-to-events that are observed for completeness. The averaged empirical quantiles should closely imitate the true quantiles of the real time-to-events (observed and unobserved), thus the selected interior knots should provide reliable and sufficient information. Any pre-existing knowledge about the potential distribution of the real time-to-events could help facilitate the selection process and refine the results.

2.3.3 DPM prior

To characterize model error in the NTM, we choose the common DPM models as priors for f_ξ and S_ξ. Taking the truncated stick-breaking approach, such priors can be constructed as

$$f_\xi(\cdot) = \sum_{l=1}^{L} p_l f_w(\psi_l, \nu_l), S_\xi(\cdot) = 1 - \sum_{l=1}^{L} p_l F_w(\psi_l, \nu_l), \tag{2.5}$$

where $f_w(\psi, \nu)$ and $F_w(\psi, \nu)$ are the PDF and the CDF of the Weibull distribution, respectively. We select the Weibull distributions as a common choice to allow for a nonincreasing hazard rate.

The stick-breaking weights p_l and the hyperparameters (ψ_l, ν_l) are generated as follows

$$p_l = q_k \prod_{k=1}^{L-1}(1-q_k), q_k \sim \text{Beta}(1,c), (\psi_l, \nu_l) \sim \text{Gamma}(1,1). \tag{2.6}$$

2.4 Posterior inference

2.4.1 Posterior prediction and nonparametric estimation

Given the prior settings, the nonparametric parts of (2.1), specifically, the functionals H and S_ξ can be represented by elements in (α, p, ψ, ν), where $\alpha = \{\alpha_j\}_{j=1}^K$, $p = \{p_l\}_{l=1}^L$, $\psi = \{\psi_l\}_{l=1}^L$, and $\nu = \{\nu_l\}_{l=1}^L$. Let $\Theta = (\beta, \alpha, p, \psi, \nu)$ contain all such unknown parameters; the estimators of (H, β, S_ξ) can then be obtained through the posterior distribution of Θ.

First set the priors for parameters in Θ as (recall ((2.6))

$$\begin{aligned} &\alpha_j \sim \exp(\eta), p(\beta) \propto 1, \\ &p_L = 1 - \sum_{l=1}^{L-1} p_l, \\ &G_0(\psi_l, \nu_l) = \text{Gamma}(1,1) \times \text{Gamma}(1,1), \end{aligned} \tag{2.7}$$

where $p(\cdot)$ is a prior density and G_0 is the base measure for the DPM prior. The posterior density of Θ can then be represented as

$$\pi(\Theta \mid \tilde{T}, Z, \delta_1, \delta_2, \delta_3) \propto \mathcal{L}(\Theta \mid \tilde{T}, Z, \delta_1, \delta_2, \delta_3) p(\beta) p(\alpha) p(p) \prod_{l=1}^{L} G_0(\psi_l, \nu_l). \tag{2.8}$$

In the above prior setting, the hyperparameter η can be dependent on other hyperparameters or fixed to some constant based on existing knowledge. It is, however, recommended that the mass parameter of the Beta distribution be fixed as $c = 1$ and the base measure G_0 also be fixed as above.

It should be noted that the prior choice for β is the improper uniform prior. Such choice simplifies the posterior form and accelerates MCMC sampling. Under mild conditions, the posterior in ((2.8)) is still guaranteed to be proper. The NUTS (No-U-Turn Sampler) from Stan (Carpenter et al., 2017) is implemented to achieve posterior sampling. After sufficient sampling procedures, the posterior predictive survival probability of any future time-to-event T_0 can be obtained given some vector of covariates Z_0.

For such prediction of a future time-to-event, denote the corresponding conditional posterior predictive survival probability as $S_{T_0|Z_0}(t)$. Mathematically, $S_{T_0|Z_0}(t)$ can be calculated through

$$\begin{aligned} S_{T_0|Z_0}(t) &= \int S_{T_0|Z_0}(t \mid \Theta)\pi(\Theta \mid \tilde{T}, Z, \delta_1, \delta_2, \delta_3) d\Theta \\ &= \int S_\xi\{H(t)\exp(-\beta^{\mathrm{T}} Z_0)\}, \end{aligned} \tag{2.9}$$

where $S_{T_0|Z_0}(t \mid \Theta)$ is the conditional posterior predictive survival probability given Θ, and $S_{T_0|Z_0}(t \mid \Theta)$ can uniquely determine $S_{T_0|Z_0}(t)$ if the posterior $\pi(\Theta \mid \tilde{T}, Z, \delta_1, \delta_2, \delta_3)$ is proper.

Note that the integral in ((2.9)) can be approximated by averaging over drawn posterior samples. Denote the drawn samples of β, H, and S_ξ by $\beta^{(p)}$, $H^{(p)}$, and $S_\xi^{(p)}$, $p = 1, \ldots, N$, respectively. Then the estimations of the conditional survival probability and conditional cumulative hazard can be given as

$$\hat{S}_{T_0|Z_0}(t) = N^{-1}\sum_{p=1}^{N} S_\xi^{(p)}\{H^{(p)}(t)\exp(\beta^{(p)^{\mathrm{T}}}Z)\},$$
$$\hat{\Lambda}_{T_0|Z_0}(t) = -\log(\hat{S}_{T_0|Z_0}(t)). \tag{2.10}$$

2.4.2 Posterior projection and parametric estimation

Recall that the joint posterior in ((2.8)) can be obtained from the prior settings in ((2.6)) and ((2.7)), thus making the set of parameters (H, β, S_ξ) jointly estimable. However, it is still important to marginally estimate each parameter, especially the parametric component β and the relative risk $\exp(-\hat{\beta}^{\mathrm{T}}Z)$. However, as the marginal posterior of β lacks interpretability, it is more meaningful to obtain the marginal posterior of an identified equivalence of β. Through the process of normalization, we denote by β^* the identified unit vector $\beta/\|\beta\|_2$ with $\|\beta^*\|_2 = 1$, and we now focus on obtaining a Bayes estimator of β^*.

Note that the parameter space of β^* is the same as the Stiefel manifold $\mathrm{St}(1, p)$ in $\mathbb{R}^p$, thus we utilize a posterior projection technique to estimate β^*. Hypothetically, consider some set $\mathcal{A}$, the metric projection operator $m_{\mathcal{A}} : \mathbb{R}^p \to \mathcal{A}$ of such set is

$$m_{\mathcal{A}}(\boldsymbol{x}) = \{\boldsymbol{x}^* \in \mathcal{A} : ||\boldsymbol{x} - \boldsymbol{x}^*||_2 = \inf_{\boldsymbol{v}\in\mathcal{A}} ||\boldsymbol{x} - \boldsymbol{v}||_2\}.$$

Thus, the metric projection of the vector $\beta \in \mathbb{R}^p$ into $\mathrm{St}(1, p)$ is uniquely determined by $m_{\mathrm{St}(1,p)}(\beta) = \beta/||\beta||_2$ (Absil and Malick, 2012), and the estimation of β^* is given by the mean or median of the projected posterior.

2.5 Simulations

In this section, we present the results of our simulation studies. These studies were conducted to assess the performance of our proposed methods under both random and fixed censoring schemes. We compare the proposed method with competitors the **R** package **spBayesSurv** (Zhou and Hanson, 2018), a Bayesian approach that can be applied to doubly censored data; the algorithm developed by Li et al. (2018), a frequentist method specifically works on doubly censored data.

Simulated survival times are generated following model (2.1). Under each case within both censoring schemes, we generate 100 Monte Carlo replicates, each with sample size $n = 200$. The vector of regression coefficients is set as $\boldsymbol{\beta} = (\beta_1, \beta_2, \beta_3)^{\mathrm{T}} = (\sqrt{3}/3, \sqrt{3}/3, \sqrt{3}/3)^{\mathrm{T}}$ such that $||\boldsymbol{\beta}|| = 1$. For covariates $\boldsymbol{Z} = (z_1, z_2, z_3)$, set $z_1 \sim \text{Bin}(1, 0.5)$, $z_2 \sim N(0, 1)$, and $z_3 \sim N(0, 1)$.

Under the random censoring scheme, the performance of our method is assessed under four different cases, where the true model is either the PH model, the PO model, the AFT model, or none of these three models.

Case 1: Non-PH/PO/AFT:

$$\epsilon \sim 0.5N(-0.5, 0.5^2) + 0.5N(1.5, 1^2),$$
$$h(t) = log((0.8t + t^{1/2} + 0.825)(0.5\phi_{1,0.3}(t) + 0.5\phi_{3,0.3}(t) - c_1)),$$
$$L_i \sim U(0, 1), R_i \sim U(8/3, 4).$$

4.0% left censored, 64.2% observed, 31.8% right censored.

Case 2: PH model:

$$\epsilon \sim \text{EV}(0,1),$$
$$h(t) = log((0.8t + t^{1/2} + 0.825)(0.5\phi_{0.5,0.2}(t) + 0.5\phi_{2.5,0.3}(t) - c_2)),$$
$$L_i \sim U(0,1), R_i \sim U(8/3,4).$$

26.2% left censored, 50.1% observed, 23.7% right censored.

Case 3: PO model:

$$\epsilon \sim \text{Logistic}(0,1),$$
$$h(t) = log((0.8t + t^{1/2} + 0.825)(0.5\phi_{0.5,0.2}(t) + 0.5\phi_{2.5,0.3}(t) - c_3)),$$
$$L_i \sim U(0,1), R_i \sim U(4/3,2).$$

31.8% left censored, 37.5% observed, 30.7% right censored.

Case 4: AFT model:

$$\epsilon \sim N(0,1),$$
$$h(t) = log(t),$$
$$L_i \sim U(0,1), R_i \sim U(4/3,2).$$

21.7% left censored, 34.7% observed, 43.6% right censored.

Similarly, under the fixed censoring scheme, the performance of our method is assessed under the same four different cases. The differences between the two censoring schemes are marked by the simulated left and right censoring times.

Case 1: Non-PH/PO/AFT:

$$\epsilon \sim 0.5N(-0.5, 0.5^2) + 0.5N(1.5, 1^2),$$
$$h(t) = log((0.8t + t^{1/2} + 0.825)(0.5\phi_{1,0.3}(t) + 0.5\phi_{3,0.3}(t) - c_1)),$$
$$L_i = 1, R_i = 6.$$

23.3% left censored, 55.5% observed, 21.2% right censored.

Case 2: PH model:

$$\epsilon \sim \text{EV}(0, 1),$$
$$h(t) = log((0.8t + t^{1/2} + 0.825)(0.5\phi_{0.5,0.2}(t) + 0.5\phi_{2.5,0.3}(t) - c_2)),$$
$$L_i = 0.5, R_i = 2.$$

34.5% left censored, 38.5% observed, 27.0% right censored.

Case 3: PO model:

$$\epsilon \sim \text{Logistic}(0, 1),$$
$$h(t) = log((0.8t + t^{1/2} + 0.825)(0.5\phi_{0.5,0.2}(t) + 0.5\phi_{2.5,0.3}(t) - c_3)),$$
$$L_i = 0.5, R_i = 3.$$

29.0% left censored, 46.3% observed, 24.7% right censored.

Case 4: AFT model:

$$\epsilon \sim N(0,1),$$

$$h(t) = log(t),$$

$$L_i = 0.5, R_i = 2.$$

22.6% left censored, 39.4% observed, 38.0% right censored.

The comparison results to spBayesSurv and Li2018 under the different censoring schemes and cases are shown in the tables below. As presented, we focus on six statistics, namely, the mean, the average, the average of posterior standard error (PSD), the square root of the mean squared error (RMSE), the standard error (SDE), and the coverage probability of the 95% credible interval (CP).

Table 2.1: Simulation Results of Random Censoring under Case 1

	Proposed method			spBayesSurv			Li2018 (r=3.5)		
	β_1	β_2	β_3	β_1	β_2	β_3	β_1	β_2	β_3
Mean	0.605	0.557	0.537	0.430	0.420	0.411	0.618	0.583	0.562
Bias	**0.028**	-0.020	-0.041	-0.147	-0.157	-0.166	-0.041	**-0.006**	**0.015**
PSD	0.086	0.065	0.065	0.180	0.094	0.094	0.187	0.092	0.092
RMSE	0.098	0.072	0.079	0.233	0.189	0.195	0.181	0.090	0.091
SDE	0.095	0.075	0.069	0.181	0.105	0.103	0.177	0.090	0.091
CP	0.88	0.94	0.89	0.84	0.56	0.55	0.95	0.97	0.94

Table 2.2: Simulation Results of Random Censoring under Case 2

	Proposed method			spBayesSurv			Li2018 (r=0)		
	β_1	β_2	β_3	β_1	β_2	β_3	β_1	β_2	β_3
Mean	0.588	0.569	0.552	0.685	0.695	0.668	0.573	0.596	0.577
Bias	0.011	**-0.008**	-0.026	0.108	0.118	0.091	**0.005**	-0.018	**0.001**
PSD	0.118	0.084	0.082	0.240	0.134	0.132	0.183	0.100	0.099
RMSE	0.101	0.081	0.099	0.268	0.197	0.164	0.155	0.112	0.104
SDE	0.101	0.081	0.099	0.246	0.159	0.138	0.155	0.111	0.105
CP	0.96	0.93	0.87	0.92	0.86	0.90	0.99	0.93	0.90

Table 2.3: Simulation Results of Random Censoring under Case 3

	Proposed method			spBayesSurv			Li2018 (r=1)		
	β_1	β_2	β_3	β_1	β_2	β_3	β_1	β_2	β_3
Mean	0.576	0.552	0.559	0.493	0.524	0.538	0.484	0.511	0.513
Bias	**-0.001**	**-0.025**	**-0.018**	-0.084	-0.053	-0.039	0.093	0.066	0.064
PSD	0.182	0.123	0.122	0.304	0.161	0.159	0.234	0.117	0.115
RMSE	0.158	0.118	0.115	0.296	0.164	0.175	0.240	0.118	0.129
SDE	0.159	0.116	0.114	0.285	0.156	0.172	0.222	0.099	0.113
CP	0.97	0.96	0.95	0.98	0.95	0.92	0.95	0.95	0.89

Table 2.4: Simulation Results of Random Censoring under Case 4

	Proposed method			spBayesSurv			Li2018 (r=0)		
	β_1	β_2	β_3	β_1	β_2	β_3	β_1	β_2	β_3
Mean	0.621	0.545	0.543	0.402	0.394	0.394	0.650	0.655	0.654
Bias	**0.044**	**-0.032**	**-0.035**	-0.176	-0.183	-0.183	-0.072	-0.078	-0.077
PSD	0.105	0.080	0.079	0.156	0.087	0.086	0.211	0.114	0.113
RMSE	0.112	0.088	0.081	0.235	0.203	0.202	0.206	0.139	0.130
SDE	0.104	0.082	0.073	0.156	0.088	0.087	0.193	0.115	0.106
CP	0.92	0.93	0.94	0.77	0.47	0.51	0.94	0.90	0.90

Table 2.5: Simulation Results of Fixed Censoring under Case 1

	Proposed method			spBayesSurv			Li2018 (r=1.5)		
	β_1	β_2	β_3	β_1	β_2	β_3	β_1	β_2	β_3
Mean	0.596	0.560	0.555	0.378	0.408	0.413	0.621	0.585	0.582
Bias	**0.018**	-0.017	-0.022	-0.199	-0.168	-0.164	-0.044	**-0.007**	**-0.005**
PSD	0.112	0.079	0.079	0.300	0.154	0.156	0.217	0.101	0.100
RMSE	0.109	0.079	0.080	0.417	0.233	0.230	0.213	0.103	0.087
SDE	0.108	0.078	0.077	0.369	0.162	0.162	0.209	0.103	0.087
CP	0.96	0.97	0.95	0.80	0.75	0.79	0.96	0.97	0.98

Table 2.6: Simulation Results of Fixed Censoring under Case 2

	Proposed method			spBayesSurv			Li2018 (r=0)		
	β_1	β_2	β_3	β_1	β_2	β_3	β_1	β_2	β_3
Mean	0.583	0.559	0.560	0.601	0.602	0.597	0.583	0.569	0.570
Bias	0.006	-0.018	-0.018	0.024	0.024	0.019	**-0.005**	**0.008**	**0.007**
PSD	0.127	0.091	0.089	0.234	0.133	0.132	0.176	0.098	0.095
RMSE	0.128	0.095	0.096	0.258	0.138	0.145	0.187	0.094	0.101
SDE	0.129	0.094	0.095	0.254	0.137	0.144	0.188	0.094	0.101
CP	0.90	0.93	0.94	0.92	0.92	0.93	0.95	0.95	0.94

Table 2.7: Simulation Results of Fixed Censoring under Case 3

	Proposed method			spBayesSurv			Li2018 (r=1.5)		
	β_1	β_2	β_3	β_1	β_2	β_3	β_1	β_2	β_3
Mean	0.589	0.525	0.544	0.414	0.438	0.442	0.533	0.527	0.534
Bias	**0.012**	-0.053	**-0.033**	-0.163	-0.139	-0.135	0.044	**0.051**	0.043
PSD	0.216	0.150	0.146	0.312	0.165	0.164	0.247	0.120	0.118
RMSE	0.185	0.148	0.159	0.335	0.226	0.212	0.221	0.128	0.135
SDE	0.186	0.148	0.159	0.294	0.179	0.164	0.217	0.118	0.128
CP	0.93	0.93	0.95	0.95	0.82	0.88	0.95	0.92	0.92

Table 2.8: Simulation Results of Fixed Censoring under Case 4

	Proposed method			spBayesSurv			Li2018 (r=5)		
	β_1	β_2	β_3	β_1	β_2	β_3	β_1	β_2	β_3
Mean	0.633	0.544	0.529	0.295	0.288	0.290	0.662	0.618	0.612
Bias	**0.055**	**-0.034**	-0.048	-0.282	-0.290	-0.287	-0.084	-0.041	**-0.034**
PSD	0.108	0.084	0.084	0.123	0.069	0.070	0.222	0.111	0.111
RMSE	0.118	0.085	0.096	0.305	0.297	0.296	0.219	0.110	0.112
SDE	0.105	0.079	0.083	0.117	0.064	0.071	0.203	0.103	0.107
CP	0.89	0.91	0.92	0.41	0.01	0.04	0.95	0.95	0.96

Under the random censoring scheme, our method generally performs the best in case 1 while significantly outperforming spBayesSurv. This can be expected since spBayesSurv is specifically designed to handle estimation under cases 2 through 4, yet our results in these cases are still comparable to spBayesSurv, indicating that the proposed method can be applied in a broader spectrum of situations while maintaining sufficient power. The simulation results of spBayesSurv and Li2018 under the fixed censoring scheme are not supposed to carry too much weight, as these methods did not take such censoring scheme into consideration, but our results are quite promising regardless, which lends credence to the claim that our method can be applied to analyze fixed doubly censored time-to-events. It is also worth noting that the proposed method is more flexible than Li2018 as Li2018 takes a semiparametric approach which assumes the form of the hazard function, making their results sensitive to the selection of an r value.

2.6 Real Data

In this section, we apply our proposed method in a practical scenario by examining data from the randomized AIDs clinical trial conducted in 1997 (recall from the introduction section). As stated previously, one major objective of this study was to examine treatment effects across different treatment groups through the plasma HIV-1 RNA level, which correspond to doubly censored time-to-events. We label the treatment group which receives a combination of 3 drugs as the ”trt $=1$” group and the treatment group which receives a combination of only 2 drugs as the ”trt $=0$” group. A remark should be made that this dataset was actually conducted under a fixed censoring scenario due to limitations of measuring techniques, resulting in the fact that all baseline log(RNA) levels could only be observed between -2.60 and 5.88. This lends further reason to our focus on an dataset that can be argued to be out-dated. We acknowledge that this dataset may not be practically relevant in contemporary times, yet it nonetheless holds methodological value given it contains fixed double censored observations. In addition, the dataset is still being analyzed in recent literature (e.g., Li et al. (2018),Choi and Huang (2021)) while newer datasets have not yet surfaced.

Below we present the analysis results of the proposed method along with the results from spBayesSurv and Li2018 methods, mainly for demonstration purposes. A visual aid is also provided for better distinction of treatments effects between the two treatment groups.

Table 2.9: Results of AIDS Study Analysis

	Proposed method			
	trt	baseRNA		
Est	1.057	0.254		
SD	0.323	0.003		
95% CI	(0.504, 1.770)	(-0.032, 0.564)		
	spBayes PH		spBayes PO	
	trt	baseRNA	trt	baseRNA
Est	1.058	0.245	1.437	0.382
SD	0.258	0.194	0.330	0.264
95% CI	(0.571, 1.583)	(-0.138, 0.623)	(0.803, 2.097)	(-0.141, 0.893)
	Li2018 PH r=0		Li2018 PO r=1	
	trt	baseRNA	trt	baseRNA
Est	0.982	0.067	1.291	0.145
SD	0.272	0.180	0.308	0.261
95% CI	(0.449, 1.516)	(-0.285, 0.419)	(0.688, 1.894)	(-0.365, 0.656)

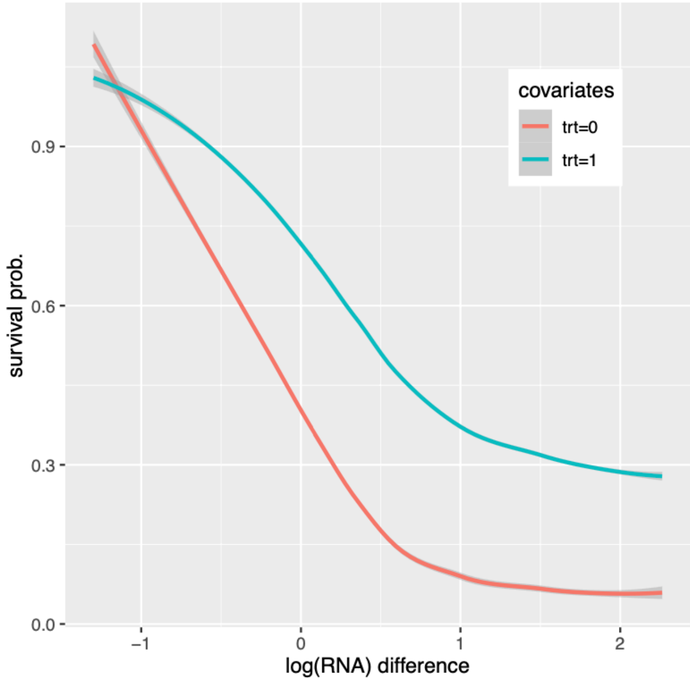


Figure 2.1: Predicted survival functions for the two treatment groups

2.7 Discussion

In this chapter, we have proposed an innovative approach to analyze doubly censored time-to-event data and demonstrated its superior accuracy and flexibility over alternative methods. Namely, we bring up a new type of weakly informative prior, the pseudo-quantile I-splines priors, that allows for nonparametric estimation and prediction of doubly censored time-to-event data under both random and fixed censoring schemes. We illustrate the effectiveness of this innovative prior by comparing simulation results in several scenarios with two leading methods with respect to our context. For fixed censoring specifically, in addition to outperforming these methods considerably in some cases and displaying comparable results otherwise, our results are quite close to the assigned true values. This lends credibility to the statement that our method is not just the first to target estimation and prediction of doubly censored time-to-events under fixed censoring, but also a valid method that deserves practical considerations. Subsequently, more attention should be brought to fixed censoring as a whole, since professionals who encounter such type of data can now be enabled with our proposed method or any future modification of it. We understand the complexity of fixed censoring and the intricacy around when only minimal information can be drawn from a substantial portion of observations. Despite these challenges, we believe our approach of pseudo data substitution has its merit, as the generated data may eventually mimic the true distribution of observations with minimally available information.

Chapter 3

Extension in model

3.1 Background

Our choice of an weakly informative prior in the last chapter was motivated by its potential to mitigate unidentifiability, as demonstrated many times in literature with unidentified parametric models (McCulloch and Rossi (1994); Gutiérrez et al. (2014); McElreath (2020); Cole (2020); among others). In this chapter, we further explore the poor mixing problem induced by unidentifiability as well as prior informativeness. As stated previously, the PPD of some future observation can, in theory, be estimated despite model unidentifiability. However, in practice, such estimations are not always reliable, as unidentified infinite-dimensional parameters can cause sampled MCMC chains to mix poorly, which may ultimately result in non-convergent sampled chains of the PPD (Vehtari et al., 2021). Given that model unidentifiability can interfere with posterior predictions through poor mixing has been established, it remains imperative to differentiate models with one unidentified infinite-dimensional parameter and models with multiple such parameters. Extant research has found that, for models with only a single unidentified infinite-dimensional parameter, poorly-mixed MCMC chains do not necessarily generate non-convergent sampled PPDs. Hence, poor mixing may not be a critical concern for this class of models such as Bayesian mixture models (Celeux et al. (2000); Geweke (2007)) and Bayesian neural networks (Izmailov et al. (2021); Papamarkou et al. (2022)), as accurate predictions can still be achieved. On the contrary, strong evidence (Sparapani et al. (2021); Kim and Rockova (2023)) suggests that poor mixing can dampen prediction accuracy significantly for models with multiple unidentified infinite-dimensional parameters (of which nonparametric transformation models belong). Consequently, how to address poor mixing under such models with more than one unidentified infinite-dimensional parameters is a vital question to improving posterior prediction precision, yet currently it stands unsolved.

As illustrated in the two previous chapters, we have proposed and demonstrated the viability of utilizing weakly informative priors to assist MCMC sampling and ensure reliable posterior predictions. However, it should be noted that the NTM (2.1) only contains two unidentified infinite-dimensional parameters, and it is unclear whether our approach would still work under models with more than two such parameters. Therefore, we extend nonparametric transformation models under right censoring to a more complex model to examine whether the issue of poor mixing can be similarly mitigated. In addition, we hope to uncover the underlying mechanism behind weakly informative priors' effectiveness in combating poor mixing. Our goal is to ultimately offer some guidelines on how to appropriately elicit priors for models with multiple infinite-dimensional parameters.

The philosophy behind establishing this kind of a criterion arises from a well-known standard to judge whether MCMC chains are sufficiently mixed (Vehtari et al., 2021). In essence, when the ratio of between-chain variance over within-chain variance for MCMC chains can be small enough, poor mixing will not be a hindrance. Based on this realization, we speculate that as long as the within-chain MCMC variance is sufficiently large, poor mixing can be effectively mitigated. Another crucial insight is that as more data become available, the posterior variances of unidentified parameters can never vanish (Amewou-Atisso et al., 2003). This understanding allows us to draw from the Bayes formula that, for unidentified parameters, prior variances will dominate posterior variances.

Consequently, we propose to further categorize the class of weakly informative priors by their variance and distinguish a new class of *moderately weak informative priors* that can help with poor mixing. It should be noted that our approach is based entirely upon prior manipulation and is distinctly different from bypassing poor mixing by adjusting posterior samples (Yao et al., 2022).

In the next section, we apply moderately weak informative priors to an extension of nonparametric transformation models and present numerical illustrations of their effectiveness. Conceptual explorations and proposed definitions are elaborated in Section 3.

3.2 Extension: nonparametric additive model

Consider the following nonparametric additive model (NAM) with transformed survival outcomes.

$$h(T) = \sum_{j=1}^{p} g_j(X_j) + \epsilon, \tag{3.1}$$

where h is a monotone and smooth function and g_j are unknown smooth regression functions. Obviously, the infinite dimensional parameters $(h, g_1, \ldots, g_p, f_\epsilon)$ in NAM (3.1) are unidentified. The conditions to identify NAM (3.1) are complicated, referred to Chiappori et al. (2015) and references therein. To address the prediction under NAM (3.1), we similarly transfer it to its equivalent inference model

$$H(T) = \xi \exp\left(\sum_{j=1}^{p} g_j(X_j)\right). \tag{3.2}$$

We model the nonnegative H by the Quantile-knot I-splines prior (Zhong, 2023) and model f_ξ by DPM model (2.5). For the regression functions g_j, we model it through the following Karhunen-Loeve expansion

$$g_j(t) = \sum_{m=1}^{M} w_{jm}\phi_m(t), \tag{3.3}$$

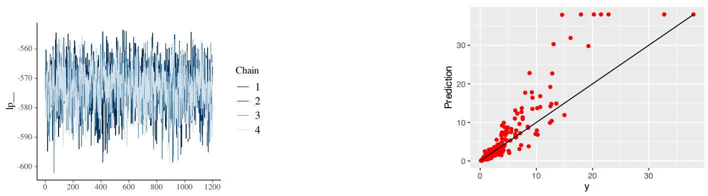
where $\{\phi_m\}_{m=1}^{M}$ are the orthornomal basis functions and $w_{jm} \in \mathbb{R}$ are the coefficients. We set $M = 10$ and $\{\phi_m\}_{m=1}^{M}$ as the Fourier basis functions in the following example.

As a numerical illustration, we consider the following data-generating model

$$\log(T) = X_1 + \sin(X_2) + \epsilon, \quad \epsilon \sim N(0,1), \tag{3.4}$$

where $X_1, X_2 \sim U(0,1)$ are all continuous variables.

We empirically assign $w_{jm} \sim N(0,1)$ as the moderately weak informative priors for g_j and adopt the same setting of moderately weak informative priors for (H, f_ξ) in the previous chapter. Through the trace plot of posterior likelihood in Figure 3.1(a), we find the MCMC chains are well-mixed. The prediction check in Figure 3.1(b) shows that the prediction performance is satisfactory.



(a) **(b)**

Figure 3.1: (a) MCMC trace plot of posterior likelihood; (b) prediction check plot: true (x axis), prediction (y axis).

3.3 Adequate informativeness and moderately weak informative priors

Consider a nonparametric model $\mathcal{M}$ with two unidentified infinite-dimensional parameters $\theta_1(t)$ and $\theta_2(t)$, and assume that $\mathcal{M}$ becomes identified once either $\theta_1(t)$ or $\theta_2(t)$ is specified. Write the corresponding weakly informative priors of the two parameters as $\pi(\theta_1)$ and $\pi(\theta_2)$, and let $\widetilde{\boldsymbol{\Theta}}$ denote the collection of all hyperparameters of these two priors. Now assume there exists some function $g_{\theta_j(t)}$ such that

$$\mathbb{V}\{\theta_j(t)|\mathcal{D}\} = g_{\theta_j(t)}(\widetilde{\boldsymbol{\Theta}}) + O(n^{-1}),\ j = 1, 2, \tag{3.5}$$

where $\mathbb{V}$ is the variance operator and $\mathbb{V}\{\theta_j(t)|\mathcal{D}\}$ is the posterior variance. Let $\mathbb{V}\{\theta_j^{\mathcal{S}}(t)|\mathcal{D}\}$ be the within-chain MCMC variance under a sampler $\mathcal{S}$, then we say priors $\pi(\theta_1)$ and $\pi(\theta_2)$ are *moderately weak informative priors* if $\mathbb{V}\{\theta_j^{\mathcal{S}}(t)|\mathcal{D}\} \geq g_{\theta_j(t)}(\widetilde{\boldsymbol{\Theta}})$, where $g_{\theta_j(t)}(\widetilde{\boldsymbol{\Theta}})$ is defined as the *adequate informativeness level*.

Here we attempt to apply the definitions to nonparametric transformation models and establish a criterion. Suppose one draws $M > 1$ parallel MCMC chains. For $m = 1, \ldots, M$, let $\mathbb{V}(H^{(m)}(s_j)|\mathcal{D})$ be the within-chain variance of the mth MCMC chain of $H(s_j)$. Suppose $\pi(\alpha) = \exp(\eta)$ and $\pi(\psi) = \exp(\zeta)$.

We say the priors for (H, f_ξ) are moderately weak informative priors and have adequate informativeness to achieve MCMC mixing if, for knot s_{j_0},

$$M^{-1} \sum_{m=1}^{M} \mathbb{V}\left\{H^{(m)}(s_{j_0})|\mathcal{D}\right\} \geq g_{j_0}(\eta, \zeta), \tag{3.6}$$

where

$$s_{j_0} = \max\left\{\hat{Q}_T(q/N_I), \hat{Q}_{\widetilde{T}}(q/N_I)\right\},\ \ q = \min_{q=0,\ldots,N_I-1}\left\{1 - \frac{q}{N_I} < e^{-1}\right\}.$$

It is worth recognizing that the above definitions are not meant to be airtight. Rather, they serve as mere recommendations for prior elicitation. This is not to say that such guidelines, if followed, could not help save redundant time and effort, but to express that there is nonetheless room left to improve upon what we have proposed here. It would not be unfair to say that our work is still at a preliminary stage, given that more theoretical support is needed to further justify our definitions. We do believe further explorations in theory and proof are worthy of a separate work and may be beyond our capabilities at the moment. To echo one examiner of the thesis, here we attempt to outline a potential route of theoretical proofs under the Bayesian paradigm.

Route of theoretical proofs: Recall that we use the notation $\widetilde{\mathbf{\Theta}}$ to specify all hyperparameters outside the two priors elicitated for the two nonparametric parameters. In our conception, we conjecture that the adequate informativeness level may be expressed as a function of hyperparameters $\tilde{\mathbf{\Theta}}$ in (3.5). Specifically, under NTM (2.1), the hyperparameters $\tilde{\mathbf{\Theta}}$ is presented by (η, ζ, ρ), where η is the hyperparameter of the exponential prior for the I-spline coefficients, ζ, and ρ are the hyperparameters of the exponential priors for (ψ, ν) in the Weibull kernel.

In NTM (2.1), we consider $\mathbb{V}\{H(t)|\mathcal{D}\}$, the marginal posterior variance of H with respect to every t. Our thought is that, based on the total variance formula, one can decompose $\mathbb{V}\{H(t)|\mathcal{D}\}$ as the sum of two parts,

Part 1 The expectation of local variance $\mathbb{E}_{\psi,\nu}\mathbb{V}\{H(t)|\mathcal{D}, \psi, \nu\}$.

Part 2 The variance of local expectation $\mathbb{V}_{\psi,\nu}\mathbb{E}\{H(t)|\mathcal{D}, \psi, \nu\}$.

Note that once (ψ, ν) is specified, $H(t)$ is identified. In this sense, the first term of (expected) local variance will vanish in the semiparametric rate of n^{-1} (the square of $n^{-1/2}$), given the Bernstein-von Mises result holds locally. Therefore, the first step of proof is to establish the Bernstein-von Mises theorem given the model error f_ξ specified. This step obtains the $O(n^{-1})$ residual term in (3.5).

Suppose we have shown the Bernstein-von Mises theorem locally in the first step. In the next step, we are in a position to formulate the second part, the variance of local expectation. Based on the Bernstein-von Mises theorem, we immediately have that

$$\mathbb{E}\{H(t)|\mathcal{D},\psi,\nu\} = H_0(t)|\psi,\nu,$$

where $H_0(t)$ is the "ground truth" of $H(t)$ given ψ and ν specified. Therefore, the second part of $\mathbb{V}\{H(t)|\mathcal{D}\}$ is indeed the variance of "true" values of $H_0(t)$. This variance can be accomplished by formulating the association between $H_0(t)$ and (ψ,ν). One possible way is to formulate $\mathbb{V}_\psi\{H_0(t)|\psi\}|\nu$ first and repeatedly employ the total variance formula. Finally, by integrating out (ψ,ν), we obtain the form of $g_{H(t)}(\tilde{\mathbf{\Theta}})$ in (3.5).

Chapter 4

Conclusion

This chapter serves to summarize the research conducted in previous chapters and highlight our main contributions. Potential extensions and directions to be explored in the future will also be discussed in detail.

Our first major contribution lies in the successful application of nonparametric transformation models in doubly censored survival analysis. Methodologically, we have managed to construct a weakly informative prior, the *Pseudo-Quantile I-splines prior*, which enables computations under the Bayesian framework to nullify unidentifiability with such nonparamteric transformation models. In particular, the proposed prior is also capable of handling fixed censoring, a rarely studied yet important censoring scheme. Additionally, although the prior is formed for the double censoring scheme specifically, it can be trivially extended to simpler censoring schemes, and perhaps with more careful modifications, to other complex ones.

The next major contribution revolves around poor mixing for MCMC and unreliable Bayesian predictions under models with multiple unidentified infinite-dimensional parameters. Conceptually, we propose a definition of *adequate informativeness level* to quantify the degree of impact, in terms of variance, that a certain prior can have on the corresponding posterior distribution. The introduction of this definition further breaks down the entirety of informative priors into more nuanced categories, and we name the newly formed class as *moderately weak informative priors*. In terms of applications, we extend nonparametric transformation models to include more than unidentified infinite-dimensional parameters and demonstrate the capability of moderately weak informative priors to mitigate poor mixing under such models.

We hope that our work in chapter 2 can inspire interest from more fellow researchers to continue developing more effective estimation and prediction tools for doubly censored time-to-event data, especially under fixed censoring. It should be noted that the proposed method is only applicable with time-independent covariates, and more complicated analysis is required to understand how our framework functions when time-dependent covariates are involved, as an entirely different approach may be necessary. To suggest some other potential research directions, we hope the proposed method can be generalized to interval censoring. This will not be an effortless task due to the absence of exactly observed time-to-events, although this may not cause as much of an issue for fixed interval censoring, since contextual information can likely be obtained given all events occur within one specific time interval. One final thought comes down to prediction in a hypothetical situation where *all* time-to-events are fixed doubly censored, leaving only two distinct values in the entire dataset. It is certainly up to debate how much practical worth can arise from solving this problem, but there is little doubt that it remains intriguing theoretically, and we aspire for a meaningful solution in the future.

Regarding poor mixing and informativeness, as we have suggested in chapter 3, there are many intriguing research problems left to be solved within the preliminary framework we have established. Although we have not yet the time to venture deeper into these problems ourselves, here we dare to compose some of the questions we believe are crucial for a more comprehensive perspective.

One natural line of inquiry is whether a stricter threshold can be obtained as the adequate informativeness level, such that priors which contain at least this amount of information (if not strictly more) can guarantee the nonoccurence of poor mixing. It is entirely possible that such a tighter bound could exist, but in case the current criterion can be refined, it may nevertheless be worthwhile to investigate the resulting empirical differences in performance, as to determine the practicality and necessity of a more restricted definition. Another idea that is perhaps more abstract comes down to if the presently defined moderately weak informative priors can be conceptually distinguished from weakly informative priors in other ways, and as a follow-up, it may even be meaningful to further differentiate within this class of moderately weak informative priors. Subsequently, a mathematical distinction between informative priors and moderately weak informative priors should be considered. Although this may not be immediately relevant regarding poor mixing, it could provide valuable insight into the characteristics of moderately weak informative priors.

Computationally, we are curious to discover if poor mixing becomes an obstacle when moderately weak informative priors and weakly informative priors, informative priors, or noninformative priors are elicited simultaneously, meaning when one or some, but not all, of the unidentified parameters have moderately weak informative priors. The results findings may shed additional light on the power of moderately weak informative priors and their interaction effects with other types of priors. We expect sufficiently accurate posterior predictions in situations where moderately weak informative priors are combined with weakly informative or informative priors, but the difference in accuracy may still be significant enough when compared to only applying moderately weak informative priors.

Theoretically, it remains to be proved the mathematical rigor of our proposed definitions. Although we acknowledge that any concrete proof is beyond the scope of this work, we have proposed a Bayesian framework regarding how we envision a set of proofs may be obtained. We suspect the same set of proofs can be extended trivially across different models with multiple unidentified infinite-dimensional parameters, and it may be beneficial to focus on one particular type of model initially. It may be a good starting point to consider the NTM and NAM we have proposed, as the proof should follow logically from two infinite-dimensional parameters to more than two.

It is to our regret that we do not have time, at this moment, to further pursue research regarding the issue of poor mixing under models with multiple unidentified infinite-dimensional parameters, and we hope our future work could attempt to answer the many questions posed here. In the meantime, we welcome and encourage keen researchers to extend the work in this thesis and formulate yet more problems.

Bibliography

Absil, P.-A. and Malick, J. (2012). Projection-like retractions on matrix manifolds. *SIAM Journal on Optimization*, 22(1):135–158.

Amewou-Atisso, M., Ghosal, S., Ghosh, J. K., and Ramamoorthi, R. (2003). Posterior consistency for semi-parametric regression problems. *Bernoulli*, 9(2):291–312.

Branscum, A. J., Johnson, W. O., Hanson, T. E., and Gardner, I. A. (2008). Bayesian semiparametric roc curve estimation and disease diagnosis. *Statistics in Medicine*, 27(13):2474–2496.

Brooks, S., Gelman, A., Jones, G., and Meng, X.-L. (2011). *Handbook of markov chain monte carlo*. CRC press.

Burgette, L. F., Puelz, D., and Hahn, P. R. (2021). A symmetric prior for multinomial probit models. *Bayesian Analysis*, 16(3):1–18.

Cai, T. and Cheng, S. (2004). Semiparametric regression analysis for doubly censored data. *Biometrika*, 91(2):277–290.

Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M. A., Guo, J., Li, P., and Riddell, A. (2017). Stan: a probabilistic programming language. *Journal of Statistical Software*, 76(1):1–32.

Celeux, G., Hurn, M., and Robert, C. P. (2000). Computational and inferential difficulties with mixture posterior distributions. *Journal of the American Statistical Association*, 95(451):957–970.

Chang, M. N. (1990). Weak convergence of a self-consistent estimator of the survival function with doubly censored data. *The Annals of Statistics*, 18(1):391–404.

Chen, S. (2002). Rank estimation of transformation models. *Econometrica*, 70(4):1683–1697.

Chiappori, P.-A., Komunjer, I., and Kristensen, D. (2015). Nonparametric identification and estimation of transformation models. *Journal of Econometrics*, 188:22–39.

Choi, S. and Huang, X. (2021). Efficient inferences for linear transformation models with doubly censored data. *Communications in Statistics - Theory and Methods*, 50:2188–2200.

Cole, D. J. (2020). *Parameter Redundancy and Identifiability*. Chapman and Hall/CRC.

Ding, Y. and Nan, B. (2011). A sieve m-theorem for bundled parameters in semiparametric models, with application to the efficient estimation in a linear model for censored data. *The Annals of Statistics*, 39:3032–3061.

Gehan, E. A. (1965). A generalized two-sample wilcoxon test for doubly censored data. *Biometrika*, 52:650–653.

Gelfand, A. E. and Sahu, S. K. (1999). Identifiability, improper priors, and Gibbs sampling for generalized linear models. *Journal of the American Statistical Association*, 94(445):247–253.

Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models (comment on article by browne and draper). *Bayesian Analysis*, 1(3):515–534.

Geweke, J. (2007). Interpretation and inference in mixture models: Simple mcmc works. *Computational Statistics & Data Analysis*, 51(7):3529–3550.

Gu, M. G. and Zhang, C. H. (1993). Asymptotic properties of self-consistent estimators based on doubly censored data. *The Annals of Statistics*, 21(2):611–624.

Gutiérrez, L., Gutiérrez-Peña, E., and Mena, R. H. (2014). Bayesian nonparametric classification for spectroscopy data. *Computational Statistics & Data Analysis*, 78:56–68.

Hughes, M. D. (2000). Analysis and design issues for studies using censored biomarker measurements with an example of viral load measurements in hiv clinical trials. *Statistics in medicine*, 19:3171–3191.

Izmailov, P., Vikram, S., Hoffman, M. D., and Wilson, A. G. G. (2021). What are bayesian neural network posteriors really like? In *International conference on machine learning*, pages 4629–4640. PMLR.

Jacob, V. G., Torben, S., and Singh, S. S. (2022). Gradient-based markov chain monte carlo for bayesian inference with non-differentiable priors. *Journal of the American Statistical Association*, 117(540):2182–2193.

Kim, J. and Rockova, V. (2023). On mixing rates for Bayesian CART. *arXiv preprint arXiv:2306.00126*.

Kou, S., Zhou, Q., and Wong, W. H. (2006). Equi-energy sampler with applications in statistical inference and statistical mechanics. *The Annals of Statistics*, 34(4):1581.

Li, S., Hu, T., Wang, P., and Sun, J. (2018). A class of semiparametric transformation models for doubly censored failure time data. *Scandinavian Journal of Statistics*, 45:682–698.

Mantel, N. (1967). Ranking procedures for arbitrarily restricted observation. *Biometrics*, 23:65–78.

McCulloch, R. and Rossi, P. E. (1994). An exact likelihood analysis of the multinomial probit model. *Journal of Econometrics*, 64(1-2):207–240.

McElreath, R. (2020). *Statistical Rethinking: A Bayesian Course with Examples in R and Stan*. Chapman and Hall/CRC.

Mykland, P. A. and Ren, J. (1996). Algorithms for computing self-consistent and maximum likelihood estimators with doubly censored data. *The Annals of Statistics*, 24(4):1740–1764.

Neal, R. M. (1996). Sampling from multimodal distributions using tempered transitions. *Statistics and computing*, 6:353–366.

Papamarkou, T., Hinkle, J., Young, M. T., and Womble, D. (2022). Challenges in markov chain monte carlo for bayesian neural networks. *Statistical Science*, 37(3):425–442.

Pompe, E., Holmes, C., and Łatuszyński, K. (2020). A framework for adaptive mcmc targeting multimodal distributions. *The Annals of Statistics*, 48(5):2930–2952.

Reich, B. J. and Ghosh, S. K. (2019). *Bayesian Statistical Methods*. CRC press.

Sparapani, R., Spanbauer, C., and McCulloch, R. (2021). Nonparametric machine learning and efficient computation with Bayesian additive regression trees: the BART R package. *Journal of Statistical Software*, 97:1–66.

Sun, J. (2006). *The Statistical Analysis of Interval-censored Failure Time Data*. Springer.

Turnbull, B. W. (1974). Nonparametric estimation of a survivorship function with doubly censored data. *Journal of the American Statistical Association*, 69:169–173.

Vehtari, A., Gelman, A., Simpson, D., Carpenter, B., and Bürkner, P.-C. (2021). Rank-normalization, folding, and localization: An improved R for assessing convergence of MCMC. *Bayesian Analysis*, 1(1):1–28.

Yao, Y., Vehtari, A., and Gelman, A. (2022). Stacking for non-mixing bayesian computations: The curse and blessing of multimodal posteriors. *Journal of Machine Learning Research*, 23(79):1–45.

Ye, J. and Duan, N. (1997). Nonparametric $n^{-1/2}$-consistent estimation for the general transformation models. *The Annals of Statistics*, 25(6):2682–2717.

Zeng, D. and Lin, D. Y. (2007). Efficient estimation for the accelerated failure time model. *Journal of the American Statistical Association*, 102:1387–1396.

Zhang, Y. and Jamshidian, M. (2004). On algorithms for the nonparametric maximum likelihood estimator of the failure function with censored data. *Journal of Computational and Graphical Statistics*, 13:123–140.

Zhong, C. (2023). *Nonparametric Bayesian statistics harnessing the forces of data in change-point detection and survival analysis*. PhD thesis, The Hong Kong Polytechnic University.

Zhou, H. and Hanson, T. (2018). A unified framework for fitting bayesian semiparametric models to arbitrarily censored survival data, including spatially referenced data. *Journal of the American Statistical Association*, 113:571–581.