UNIVERSITY OF CALIFORNIA,
IRVINE

A New Class of Bayesian Semi-Parametric Joint Longitudinal-Survival Models for Biomarker Discovery

DISSERTATION

submitted in partial satisfaction of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

in Statistics

by

Sepehr Akhavan Masouleh

Dissertation Committee:
Professor Daniel Gillen, Chair
Professor Babak Shahbaba
Professor Wesley Johnson

2016

DEDICATION

To my beloved parents, Soudabeh and Nasser, and my beloved brother, Soheil.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

I would like to express my deepest appreciation to my advisor and my committee chair, Professor Daniel Gillen. Dr. Gillen has provided me with utmost support and guidance over the past 7 years and has always encouraged me to push myself beyond my current limits.

I would like to thank my committee members, Professor Babak Shahbaba and Professor Wesley Johnson. Dr. Shahbaba, my co-advisor, has provided me with invaluable instruction, advice, and feedback throughout my PhD. My deepest appreciation goes to Dr. Johnson for teaching me the fundamentals of Bayesian non-parametrics.

I would like to offer a special thank you to Dr. Alexander Vandenberg-Rodes for generously sharing his deep knowledge on Gaussian processes and for his invaluable collaboration.

I would like to thank my advancement committee, Professor Yaming Yu and Professor Kam Kalantarzadeh. I would also like to thank Dr. Padhraic Smyth for giving me the opportunity to serve as an R programming instructor for the UCI data science initiative, Dr. Pathik Wadhwa for giving me the opportunity to collaborate with the Department of Psychiatry and Human Behavior, Dr. Elani Streja for giving me the opportunity to access real data on patients with kidney disease, and Dr. Tracy Holsclaw for her valuable collaboration.

I would like to thank Dr. Hashem Mahlooji who first introduced me to the field of Statistics and encouraged me to pursue professional and graduate degree in this field. I would also like to send a warm appreciation to Mr. Andrew Coulson who has been always a great mentor to me.

I would like to thank all the members of the Department of Statistics, including all the faculty members, students, the department manager Rosemary Busta and Lisa Stieler.

Finally, I would like to thank my parents, Nasser and Soudabeh Akhavan, and my brother Soheil. Without their support and their dedications, this dream wouldn't come true.

CURRICULUM VITAE

Sepehr Akhavan Masouleh

EDUCATION

Doctor of Philosophy in Statistics **2016**
University of California Irvine *Irvine, CA*

Master of Science in Statistics **2011**
University of California Irvine *Irvine, CA*

Bachelor of Science in Industrial Engineering **2009**
Sharif University of Technology *Tehran, Iran*

RESEARCH EXPERIENCE

Trainee Researcher **2012–2016**
Statistics Dept., University of California, Irvine *Irvine, California*

Graduate Research Assistant **2010–2011**
Epidemiology Dept., University of California, Irvine *Irvine, California*

TEACHING EXPERIENCE

Teaching Assistant **2012–2016**
Statistics Dept., University of California, Irvine *Irvine, California*

R Workshop Instructor **2014–2016**
Data Science Initiative, University of California, Irvine *Irvine, California*

ABSTRACT OF THE DISSERTATION

A New Class of Bayesian Semi-Parametric Joint Longitudinal-Survival Models for Biomarker Discovery

By

Sepehr Akhavan Masouleh

Doctor of Philosophy in Statistics

University of California, Irvine, 2016

Professor Daniel Gillen, Chair

In studying the progression of a disease and to better predict time to death (survival data), investigators often collect repeated measures over time (longitudinal data) and are interested in testing the association between risk factors, including collected repeated measures, and time to death. One such example is testing the association between the biomarker serum albumin that is measured repeatedly on end-stage renal disease (ESRD) patients. A modeling framework that is capable of modeling longitudinal and survival outcomes simultaneously is called a joint longitudinal-survival model.

Joint longitudinal-survival models have received a great deal of attention over the past years where many different joint models have been proposed. Joint models commonly make parametric assumptions on either the functional form of the repeated measures or on the distribution of survival times. In this dissertation we are interested in joint models that are robust to common parametric and semi-parameteric survival and longitudinal assumptions. We propose a flexible Bayesian joint longitudinal-survival framework that avoids common parametric and semi-parameteric assumptions. More specifically, our modeling framework incorporates a flexible longitudinal component by utilizing Gaussian process (GP) technique. This technique avoids any explicit functional assumption on the trajectory of the repeated

measures. Our modeling framework also uses Dirichlet process (DP) prior to avoid explicit distributional assumptions on survival times.

We further extend our framework to modeling multiple longitudinal processes simultaneously. We propose a multivariate joint longitudinal-survival technique to jointly model the association between multiple longitudinal processes with survival outcomes. Our proposed technique is capable of taking correlation between longitudinal processes into account. This is particularly useful when observed measures from different longitudinal processes are taken at different frequencies. That means, some longitudinal processes are observed less frequently compared to other longitudinal processes. By jointly modeling these processes, one can take the correlation between the processes into account, and hence, better estimate the trajectory of the processes including those less frequent ones.

Our proposed joint modeling frameworks use Dirichlet process techniques. Therefore, understanding parameter estimation in these models is vital. Using synthetic longitudinal and survival data, we compare parameter estimation under DPM models as opposed to commonly used parametric techniques. We are particularly interested in evaluation of the performance of the model in parameter estimation when a population consists of sub-populations with latent features that are different across subgroups. We propose a Dirichlet process mixture survival model that is capable of detecting latent subpopulations characterized by differing baseline risks for mortality. Our proposed technique is particularly useful when interest lies in estimation of the conditional effect of covariates as opposed to estimates that are marginalized across all subpopulations.

Throughout, our work is motivated by data on patients with end stage renal disease (ESRD), a condition where the kidneys are no longer capable of cleaning blood sufficiently enough to sustain life. In this context a modeling framework capable of finding mortality-related biomarkers, which are measured longitudinally over time, can significantly help physicians and practitioners to lower mortality among these patients.

Chapter 1

Motivation

This thesis is focused on developing flexible joint longitudinal-survival models. Our work has been motivated by data on end-stage renal disease (ESRD) patients that are obtained from the United States Renal Data System (USRDS). In this chapter, we motivate our work from both clinical (Section 1.1) and statistical (Section 1.2) point of view. We first start with the clinical point of view, where we introduce end stage renal disease and the common treatments for ESRD patients. In the same section, we also introduce our scientific questions of interest. Next, we describe the statistical point of view, where we provide a brief description on common statistical approaches to address our scientific questions of interest.

1.1 Clinical Motivation

According to the United States Renal Data System's 2011 annual report, one in 10 American adults suffer from chronic kidney disease, a condition where kidneys gradually become dysfunctional. End stage renal disease (ESRD) is the final stage of chronic kidney disease where kidneys are completely incapable of their main task of filtering blood from toxins and

removing the excess fluid from the body. In this situation, the built-up toxins in the body can damage other organs and can cause death. According to the 2015 USRDS annual report, currently, over 660,000 persons in the United States are being treated for end stage renal disease (ESRD).

Kidney transplantation is an ideal treatment for most ESRD patients. According to the data from United Network for Organ Sharing (UNOS), currently there are more than ninety-nine thousand individuals waitlisted for a kidney transplantation in the US. When there is no access to a viable kidney transplant, dialysis is the most common treatment for these patients.

Dialysis is a technique to remove toxic wastes and extra fluid from the body when kidneys are not functioning. There are two main types of dialysis, hemodialysis and peritoneal dialysis. Each patient's characteristic defines which dialysis type will suit the patient better (McDonald et al. (2009)).

Hemodialysis is a process where blood gets filtered out of toxic wastes outside of the patient's body and through a dialysis machine. A catheter is placed inside a patient's vein and is used to remove the blood from the patient's body and return a cleaned blood back into the body. Dialysis machine includes filters to clean the blood (Figure 1.1).

Peritoneal dialysis is a process where abdomen lining membrane (peritoneum) is used as a filter to clean blood from toxic wastes and to remove excessive fluids out of the body. In this method, a sterile solution (dialysate) is inserted into the abdominal area through a catheter. This solution remains to absorb wastes and is drained out after few hours (Figure 1.2).

Despite improvements in dialysis techniques, dialysis patients experience high mortality rate. According to the United States Renal Data System, these patients experience a 20%-25% mortality rate after one year of treatment, and with a five-year mortality rate of 65%. These patients also experience high hospitalization rate and low quality of life rate (Kalantar-Zadeh et al. (2001), CARLSON et al. (1984), Habach et al. (1995)). In 2002, a multi-

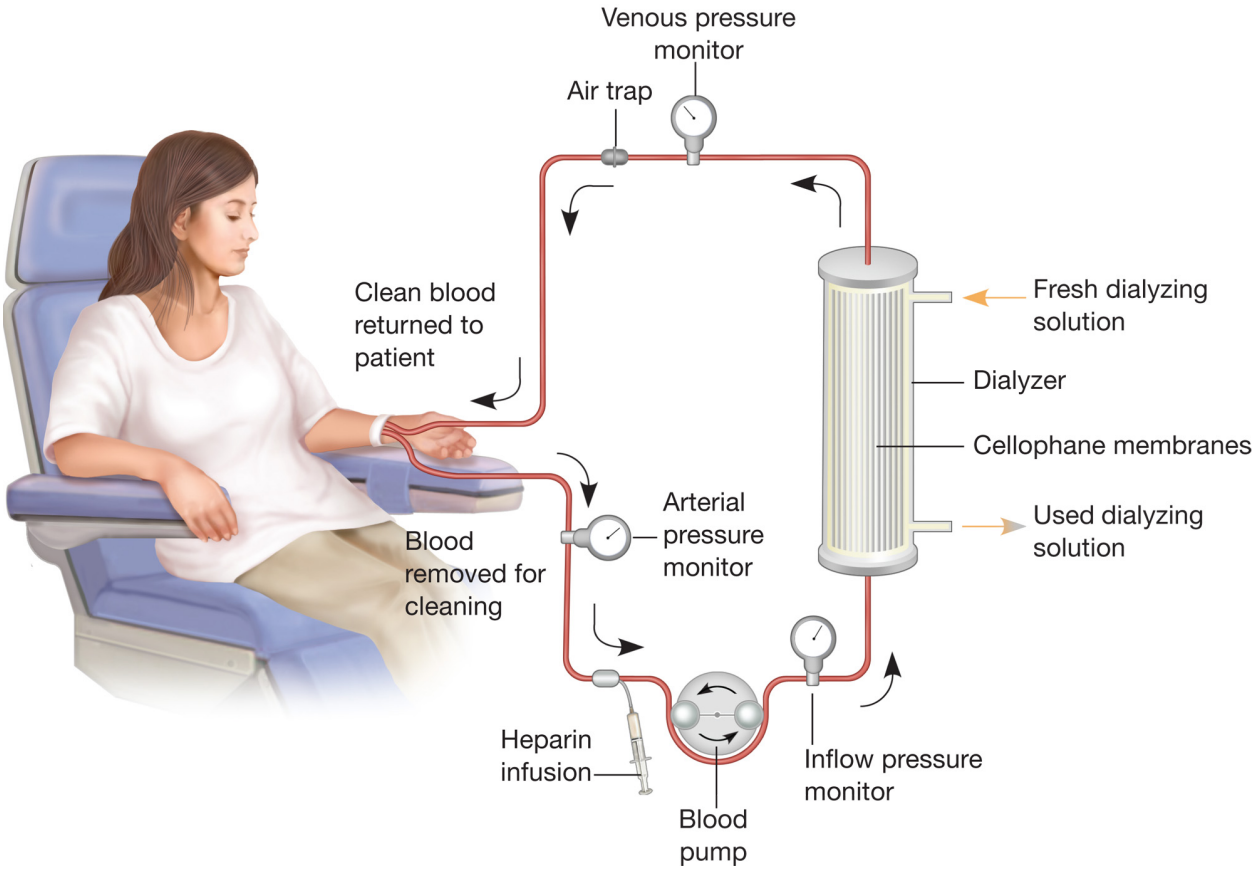


Figure 1.1: Hemodialysis process where blood gets filtered out of toxins outside of the body (Circle (2016)).

center randomized clinical trial known as the HEMO study failed to find any significant improvement in survival rate among high dose maintenance hemodialysis patients or among patients with high-flux membranes (Eknoyan et al. (2002), Kalantar-Zadeh et al. (2003)). In this situation, finding influential factors on poor dialysis outcome is of interest.

Many epidemiologic studies have consistently shown protein-energy malnutrition (PEM) as a risk factor for clinical outcomes among dialysis patients (Kopple (1997), Kopple et al. (1999)). Kalantar-Zadeh et al. (2003) defined PEM as "the state of decreased body pools of protein with or without fat depletion or a state of diminished functional capacity, caused at least partly by inadequate nutrient intake relative to nutrient demand and/or which is improved by nutritional repletion." Prevalence of PEM among dialysis patients depends on multiple factors, including the dialysis modality, and has been reported in multiple studies from low

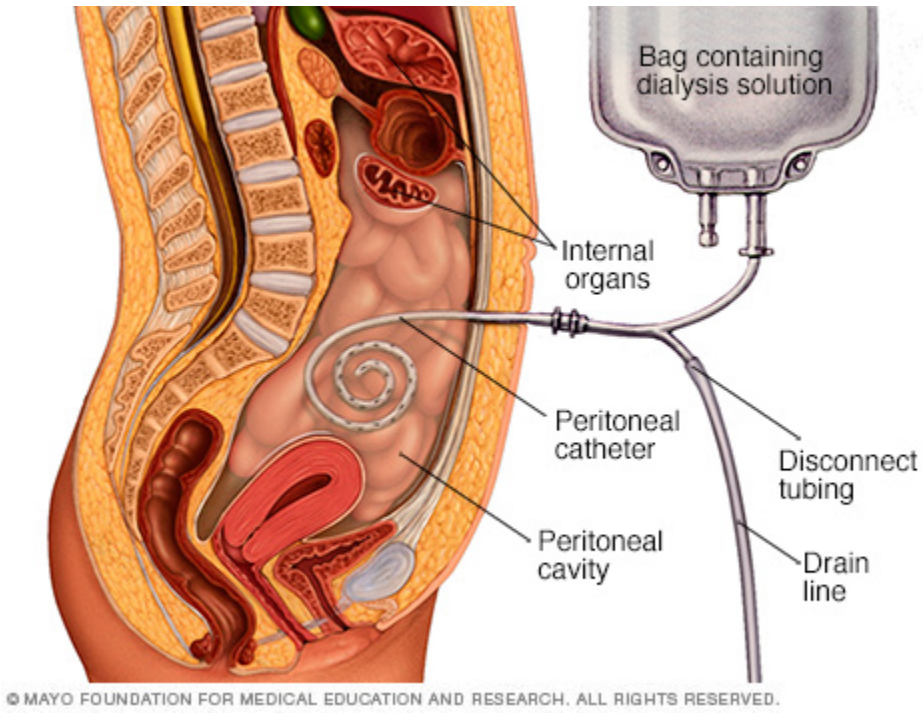


Figure 1.2: Peritoneal dialysis where abdomen peritoneum acts as a filter to clean blood (Clinic (2016))

18% to high 75% (Kalantar-Zadeh and Kopple (2001), Kalantar-Zadeh et al. (2003)).

While it's known that PEM is a risk factor for mortality among dialysis patients, however, what causes PEM is not clearly known. Some of the factors that may cause PEM include nutrient losses during the dialysis process, dietary restrictions for dialysis patients in order to minimize toxic wastes in their bodies, inadequate nutrient intake, hypercatabolism caused by inadequate protein intake during the dialysis process, hypercatabolism that can be caused by comorbid illnesses including cardiovascular diseases and diabetic complications, and nutrient loss with blood losses (Kalantar-Zadeh et al. (2003), Vendrely et al. (2003), Chung et al. (2003)).

Kalantar-Zadeh et al. (2003) reports on frequently studied indicators of malnutrition that are associated with mortality. Some of these indicators include: decreased in protein intake, reduced body mass index (BMI), decreased total body nitrogen and potassium levels, and decreased in serum albumin. In particular, Fung et al. (2002) showed that baseline serum

albumin, as an index of PEM, and slope of albumin over time were independent risk factors of mortality among ESRD patients.

Albumin is a water-soluble serum protein that is synthesized in liver. It primarily functions as a protein carrier to organs and is essential in proper distribution of body fluids (Hawkins and Dugaiczyk (1982)) as well as in maintenance of colloid osmotic pressure. Serum albumin can be measured longitudinally over time as a biomarker that might be related to clinical outcomes among hemodialysis patients. Motivated by testing the scientific question that serum biomarkers are associated with mortality among ESRD patients, and by using the data from the United States Renal Data System (USRDS), we are interested in addressing the following scientific questions:

- **1) testing the association between a biomarker level and time to death**: Fung et al. (2002) found baseline serum albumin and slope of albumin are risk factors of death among ERDS patients. It's of interest to test whether serum albumin at the time of death is also a risk factor of mortality. This finding can significantly help practitioners and doctors to lower mortality among these patients by monitoring their serum albumin values. Moreover, a methodology that can provide a general framework to detect other biomarkers that are related to mortality among these patients is also of interest.

- **2) testing the association between history of changes in the biomarker and time to death:** Between two patients both with the same albumin level but one with an increasing slope and the other with a decreasing slope, risk of mortality might be different. Hence, the rate of changes in the biomarker can be another important risk factor. It's of interest to propose a methodology capable of testing the association between rate of changes in a biomarker and risk of death.

- **3) testing the association between summary measures of the trajectory of**

the biomarker and time to death: While Fung et al. (2002) focused on the linear trend of albumin, higher within-subject volatility of albumin over time, as an indication of instability of the albumin level, might also be a risk factor of mortality. A measure of volatility is an example of a summary measure of the trajectory of a biomarker. It's of interest to propose a methodology capable of deriving new summary measures from the trajectory of a biomarker and testing the association between the summary measure and survival outcomes.

- **4) Simultaneously modeling the association between multiple biomarkers and mortality:** Often times, when multiple longitudinal biomarkers are measured over time, some biomarkers are measured less frequently. A modeling framework capable of modeling multiple biomarkers simultaneously can help model the trajectory of the less-frequent biomarkers more accurately, and hence, can better test the association between those biomarkers and survival outcomes. It's of interest to propose a joint modeling framework of longitudinal and survival data that is capable of modeling multiple longitudinal biomarkers simultaneously.

1.2 Statistical Motivation

In order to address our scientific questions of interest in Section 1.1, we are interested in evaluating the effect of longitudinally measured biomarkers (e.x. serum albumin) on mortality. Our primary goal is to develop a modeling framework that allows characterizing the trajectory of longitudinal covariates, in order to test which covariates are predictive of mortality, while adjusting for other related covariates in the model.

Testing the relationship between longitudinal covariates and survival outcomes has received a great deal of attention over the past 30 years. The question was first motivated by AIDS clinical trials where studies were attempting to find a proper surrogate biomarker for clinical

progression of the disease, so that efficacious treatments could be found in a shorter time (Wang and Taylor (2001a)).

One might consider incorporating longitudinal measures as time-varying covariates inside of a separate survival model. While subjects fail on a continuous basis, longitudinal measures are typically measured at discrete times (ex. monthly lab visits). When there is no measurement at the event time, one may choose to use the last observation carried forward (LOCF) method. This approach ignores biological variability of biomarkers overtime. Moreover, when longitudinal biomarkers are measured with error, including them directly as a traditional time-varying covariate in a survival model may lead to biased regression estimates (Prentice (1982)).

Alternatively, one could apply a two-stage method, where the first stage consists of modeling the longitudinal components via a mixed-effects model, and in the second stage, the modeled values or their summaries (e.g., first-order trends) are included in a survival model (Dafni and Tsiatis (1998), Tsiatis et al. (1995)). However, this approach fails to account for uncertainty in the estimated longitudinal summary measures. Moreover, when longitudinal covariates are missing at the event time, there is a potential for biased regression estimates in this approach. Finally, by separating data into two pieces, longitudinal data and survival data, we may lose statistical precision as opposed to a method that uses all the data at once.

To overcome these issues, several joint longitudinal-survival models have been proposed both from the Bayesian standpoint (Faucett and Thomas (1996), Wang and Taylor (2001a), Brown and Ibrahim (2003), citehanson2011predictive), as well as the frequentist standpoint (Wulfsohn and Tsiatis (1997a), Law et al. (2002a), Song et al. (2002)). All these models account for uncertainty in longitudinal measures by modeling them simultaneously with the survival outcome. However, most existing joint models still rely on multiple restrictive parametric and semi-parametric assumptions and they generally focus only on associating the first moment of the distribution of the longitudinal covariates with survival outcomes.

We propose a flexible joint longitudinal-survival modeling framework that avoids any explicit functional assumptions on the trajectory of the longitudinal measures. Our model also avoids explicit distributional assumption on the survival times where we allow for subject-specific baseline hazard in modeling a survival outcome. Our models are motivated by data on end-stage renal disease (ESRD) patients obtained from the United States Renal Data System (USRDS). Specifically, our interest lies in quantifying the association between the longitudinally measured serum albumin and time-to-death using a joint survival-longitudinal modeling approach.

Flexibility in our joint models is achieved in the longitudinal component by using a Gaussian process prior with a parameter that captures within-subject volatility in longitudinally sampled albumin. The survival component of our proposed models quantifies the association between longitudinally measured albumin and the risk of mortality using a Dirichlet process mixture of Weibull distributions. Estimation for the longitudinal and survival parameters is carried out simultaneously via a Bayesian parameter posterior sampling approach.

The remainder of the thesis is organized as follows. In Chapter 2, we provide a detailed background on Bayesian non-parametrics including GP and DP models, background on survival and longitudinal models. We conclude Chapter 2 by providing a detailed literature review on existing joint longitudinal-survival models. In Chapter 3, we propose a semi-parametric survival model that is capable of estimating conditional subgroup parameters when the data include latent population subgroups. In Chapter 4, we present the details of our proposed joint modeling framework. Simulation studies will present and we conclude by applying our modeling framework to the motivating USRDS data. In Chapter 5, we extend our modeling framework to a general case of modeling more than one biomarker simultaneously. Finally, in Chapter 6, we conclude the thesis with a brief summary and directions for future research.

Chapter 2

Background

In this chapter, we start with a brief background on longitudinal (Section 2.1) and survival models (Section 2.2). We continue by introducing Gaussian Processes (Section 2.3) and Dirichlet Processes (Section 2.4). Finally, we will conclude with a detailed literature review on existing joint longitudinal-survival models and explaining how our joint modeling framework can contribute to the literature (Section 2.5).

2.1 Longitudinal Analysis

Data that include repeated measures on the same individuals, with those measurements usually ordered by time, are called longitudinal data. Studies that deal with longitudinal data with the primary goal of characterizing within-subject changes in the outcome over time are called longitudinal studies. In the study of changes in body fatness in girls after menarche, a cross-sectional study comparing two groups of girls one pre-menarche with an average age of 10 and the other post-menarche with an average age of 15, is not capable of characterizing changes in body-fatness as girls age from 10 to 15, whereas a longitudinal

study following each girl from 10 to 15 is capable of estimating within-subject aging effect on body-fatness (Fitzmaurice et al. (2012)).

In longitudinal studies, repeated measures taken over time on each unit tend to be positively correlated compared to the measures on other units. This within-unit correlation violates a key assumption in many statistical models where the independence of measures is required. Longitudinal models are typically very similar to common statistical models except that these models directly account for this within-unit correlation in order to provide valid statistical inference.

Longitudinal data can be considered as a special case of clustered data where each unit is a cluster and measurements on that unit are members of that cluster. Unlike longitudinal data where measurements are taken over time, measurements in clustered data need not have a temporal order. For example, in a study designed to test the effect of a math intervention on students' standardized test scores, students are clustered by classrooms where all students from the same classroom share the same teacher. Because of this, within-cluster members tend to be positively correlated. Similar to longitudinal data, this within-cluster correlation should be taken into account for valid inference and the general methodology for doing so is similar to the methods used for analyzing longitudinal data. Hence, all these methods are considered as methods on modeling clustered data with longitudinal data as a special case.

2.1.1 Linear Models for Correlated Data

Consider n randomly selected and independent units each with m_i number of within unit outcomes and covariate measurements. Outcome measurement for the i^{th} subject is then an $m_i \times 1$ vector $\boldsymbol{Y}_i$, with elements $\{Y_{i1}, Y_{i2}, \ldots, Y_{im_i}\}$. Corresponding to each Y_{ij}, where $j \in \{1, \ldots, m_i\}$, there is an associated covariate vector $\boldsymbol{X}_{ij} = (X_{ij}^1, \ldots, X_{ij}^P)$, where $\boldsymbol{X}_{ij}$ is a

$P \times 1$ covariate vector. We then define a linear model of the the form,

$$E[\boldsymbol{Y}_{ij}|\boldsymbol{X}_{ij}] = \beta_0 + \beta_1 X_{ij}^1 + \beta_2 X_{ij}^2 + \cdots + \beta_P X_{ij}^P,$$
$$Var[\boldsymbol{Y}_i] = \boldsymbol{\Sigma}_i,$$

where $\boldsymbol{\Sigma_i}$ is an $m_i \times m_i$ covariance matrix. Ordinary least square estimate of $\boldsymbol{\beta}$, where $\boldsymbol{\beta}$ is a $P \times 1$ vector of the coefficients $\beta_1, \ldots, \beta_P$, is in-efficient. If $\boldsymbol{\Sigma_i}$ is known, one can use a weighted least square to estimate $\boldsymbol{\beta}$, where the estimated $\hat{\boldsymbol{\beta}}_W$ has the form

$$\hat{\boldsymbol{\beta}}_W = \left(\Sigma_{i=1}^N \boldsymbol{X}_i^T \boldsymbol{W}_i \boldsymbol{X_i}\right)^{-1} \left(\boldsymbol{X}_i^T \boldsymbol{W}_i \boldsymbol{Y}_i\right),$$

where $\boldsymbol{W}_i$ is a weight matrix and is equal to $\boldsymbol{\Sigma}_i^{-1}$ and $\boldsymbol{X_i}$ is the design matrix for subject i. In general, $\boldsymbol{\Sigma}_i$ is not known. When $Var[\boldsymbol{Y}_i] = \boldsymbol{\Sigma}_i$ is equal for all units, iteratively re-weighted least square can be used. The algorithm works as follows:

Linear Model for Correlated Data Using Iteratively Weighted Least Square

---

Start with an initial OLS estimate: $\boldsymbol{\beta}_{temp} \leftarrow \hat{\boldsymbol{\beta}}^{OLS}$

Repeat the steps below until convergence:

1. Obtain $\hat{\boldsymbol{\Sigma}} = \frac{1}{n} \sum_{i=1}^n (\boldsymbol{Y}_i - \boldsymbol{\beta}_{temp}^T \boldsymbol{X}_i)(\boldsymbol{Y}_i - \boldsymbol{\beta}_{temp}^T \boldsymbol{X}_i)^T$
2. $\boldsymbol{W}_i \leftarrow \hat{\boldsymbol{\Sigma}}^{-1}$
3. $\boldsymbol{\beta}_{temp} \leftarrow \left(\Sigma_{i=1}^N \boldsymbol{X}_i^T \boldsymbol{W}_i \boldsymbol{X}_i\right)^{-1} \left(\boldsymbol{X}_i^T \boldsymbol{W}_i \boldsymbol{Y}_i\right)$

When the number of within unit measurements is large, however, IRWLS requires estimating a large number of parameters for $\boldsymbol{\Sigma}_i$ and this requires a large sample size in terms of number of units. In this case, one common approach is to assume a structure on $\boldsymbol{\Sigma}_i$ which leads to an estimation of fewer number of parameters. Common covariance structures include compound symmetric and autoregressive correlation.

2.1.2 Linear Mixed Models

Laird and Ware (1982) introduced linear mixed model as a model that includes both shared covariate effect across all clusters as well as cluster-specific covariate effects. They introduced a model of this form,

$$\boldsymbol{Y}_i = \boldsymbol{\beta}^T \boldsymbol{X}_i + \boldsymbol{b_i}^T \boldsymbol{Z}_i + \boldsymbol{\epsilon}_i, \tag{2.1}$$

where $\boldsymbol{Y}_i$ is a response vector of size $m_i \times 1$, with m_i as the number of measurements on the i^{th} cluster. $\boldsymbol{X}_i$ is a design matrix of size $m_i \times p$, where p is the number of covariates in the model. $\boldsymbol{\beta}$ is a coefficient vector of size $p \times 1$ and indicates a shared covariate effect across all clusters. Among the covariates, q of them may have a cluster-specific effect. $\boldsymbol{Z}_i$ is a design matrix of size $m_i \times q$ that includes covariates with a cluster specific effect. $\boldsymbol{b}_i$ is the cluster effect coefficient vector for the i^{th} cluster that is of size $q \times 1$. In the context of mixed effects models, $\boldsymbol{\beta}$ coefficient vector is called fixed effect as the effect is fixed across clusters, whereas $\boldsymbol{b_i}$ coefficient vector is called random effect. Random effects are typically assumed to be Normally distributed. Finally, $\boldsymbol{\epsilon}_i$ is a vector of measurement errors of size $m_i \times 1$ that is assumed to be independent from the random effects $\boldsymbol{b}_i$ and is Normally distributed. This can be summarized as

$$\boldsymbol{b}_i \sim N(0, \boldsymbol{\Sigma}_b),$$
$$\boldsymbol{\epsilon}_i \sim N(0, \boldsymbol{\Sigma}_{\epsilon_i}).$$

With this model specification, $\boldsymbol{Y}_i$ is distributed according to a multivariate Normal distribution with mean,

$$E[\boldsymbol{Y}_i] = \boldsymbol{\beta}^T \boldsymbol{X}_i,$$

and variance,

$$\begin{aligned} Var[\boldsymbol{Y}_i] &= \boldsymbol{\Sigma}_{Y_i}(\alpha) \\ &= \boldsymbol{Z}_i \boldsymbol{\Sigma}_b \boldsymbol{Z}_i{}^T + \boldsymbol{\Sigma}_{\epsilon_i}, \end{aligned}$$

where parameters inside $\boldsymbol{\Sigma}_b$ and $\boldsymbol{\Sigma}_{\epsilon_i}$ are represented by α. α and $\boldsymbol{\beta}$ can be estimated via the maximum likelihood method.

2.1.3 Generalized Linear Models for Correlated Data

In Section 2.1.1, a semi-parameteric approach to fit a linear model for correlated data introduced. In Section 2.1.2, with parametric assumptions on random effect, a similar model was fit using the maximum likelihood method. Similar techniques can be used to model correlated data under a more general case where outcome is distributed according to a distribution from exponential family distribution, that is a distribution that can be written as

$$P(Y_{ij}|\theta) = h(Y_{ij}) exp\big(\eta(\theta) T(Y_{ij}) - A(\theta)\big), \tag{2.2}$$

where θ is a vector of parameters, $\eta(.)$ and $A(.)$ are functions that they only depend on parameters θ, $h(.)$ and $T(.)$ is a function that only depends on the data Y_{ij}. These models are known as generalized linear models with count and binary outcome data as some examples of these models.

Analogous to linear models for correlated data (Section 2.1.1), we introduce two approaches to model correlated data using the generalized linear models, one an approach to fit the model using IRWLS (GEE) and another a parameteric approach that use the maximum likelihood technique to fit the model (GLMM).

Generalized Estimating Equation Models (GEE)

One can extend the linear model introduced in Section 2.1.1 to a generalized linear model where the outcome is distributed according to a distribution from an exponential family of distributions with mean,

$$E[\boldsymbol{Y}_i|\boldsymbol{X}_i] = \boldsymbol{\mu}_i,$$

and variance,

$$Var[\boldsymbol{Y}_i|\boldsymbol{X}_i] = \boldsymbol{\Sigma}(\alpha, \beta),$$

where $\boldsymbol{Y}_i$ and $\boldsymbol{X}_i$ are the outcome vector and the design matrix for thei^{th} cluster, respectively. $Var[\boldsymbol{Y}_i|\boldsymbol{X}_i]$ may depend on some parameter α and some coefficient vector $\boldsymbol{\beta}$. coefficient vector $\boldsymbol{\beta}$, design matrix $\boldsymbol{X}_i$, and mean vector $\boldsymbol{\mu}_i$ are linked via the function $g(\cdot)$, where

$$g(\mu_{ij}) = \boldsymbol{\beta}^T \boldsymbol{X}_{ij}.$$

Parameter estimation is done using IRWLS. In order to estimate fewer parameters, one can specify a covariance structure, where

$$\begin{aligned} Var[\boldsymbol{Y}_i] &= \boldsymbol{\Sigma}(\alpha, \beta) \\ &= \boldsymbol{S}_i(\boldsymbol{\mu}_i)^{1/2} \boldsymbol{R}_i(\alpha) \boldsymbol{S}_i(\boldsymbol{\mu}_i)^{1/2}. \end{aligned}$$

with $\boldsymbol{S}_i(\boldsymbol{\mu}_i) = diag(\boldsymbol{V}_i)$, where $diag(.)$ is a function that returns a diagonal matrix with diagonal elements $V_{ij} = Var[Y_{ij}|\boldsymbol{X}_i]$. In exponential family, variance is often a function of

mean, so is $\boldsymbol{S}_i$ in our notation. $\boldsymbol{R}_i(\alpha)$ is a correlation matrix, where

$$\boldsymbol{R}_{ij} = Corr[Y_{ij}, Y_{ik}|\boldsymbol{X}_i].$$

$\boldsymbol{R}_i(\alpha)$ is commonly specified according to a compound symmetry or autoregressive correlation structure. One can show that the estimate of $\boldsymbol{\beta}$, when α is known, is an estimator of the estimating equation

$$U(\boldsymbol{\beta}) = \sum_{i=1}^{n} \boldsymbol{D}_i^T(\beta)\boldsymbol{V}_i^{-1}(\alpha, \boldsymbol{\beta})[\boldsymbol{Y}_i - \boldsymbol{\mu}_i(\boldsymbol{\beta})],$$

where $\boldsymbol{D}_i$ is a matrix of the partial derivatives of $\boldsymbol{\mu}_i$ with respect to the $\boldsymbol{\beta}$ coefficients with

$$\boldsymbol{D}_i(j,k) = \frac{\partial \mu_{ij}}{\partial \beta_k}.$$

Estimation of α is done using moment estimators (Liang and Zeger (1986)) that depend on residuals r_{ij} defined as

$$r_{ij} = \frac{Y_{ij} - \hat{\mu_{ij}}}{\sqrt{\hat{V_{ij}}}}.$$

Depending on the assumed covariance structure, one can estimate the parameter α using the standardized residuals r_{ij}. In particular, for a compound symmetric covariance structure, that is an $\boldsymbol{R}$ matrix structured as

$$\boldsymbol{R}(\alpha) = \begin{pmatrix} 1 & \alpha & \cdots & \alpha \\ \alpha & 1 & \cdots & \alpha \\ \vdots & \vdots & \ddots & \vdots \\ \alpha & \alpha & \cdots & 1 \end{pmatrix}$$

, α is estimated as

$$\alpha = \frac{\sum_{i=1}^{n}}{\sum_{j \neq k} r_{ij} r_{ik}} \sum_{i=1}^{n} m_i(m_i - 1) - p,$$

where p is the number of $\boldsymbol{\beta}$ coefficients. In summary, α and $\boldsymbol{\beta}$ in a GEE model are estimated iteratively as follows:

Parameter Estimation in Generalized Linear Models for Correlated Data

---

Start with an initial $\boldsymbol{\beta}$ that is estimated using an independent working correlation structure and set: $\boldsymbol{\beta}_{temp} \leftarrow \boldsymbol{\beta}^0$

2. Given $\boldsymbol{\beta}_{temp}$, calculate the moment estimator of α

3. Given an estimated α, estimate a new $\boldsymbol{\beta}$ as follows:

$$\boldsymbol{\beta}_{temp} \rightarrow \boldsymbol{\beta}_{temp} + \left(\boldsymbol{\Sigma}_{i=1}^{n} \boldsymbol{D}_i^T \boldsymbol{V}_i \boldsymbol{D}_i\right)^{-1} \boldsymbol{\Sigma}_{i=1}^{n} \boldsymbol{D}_i^T \boldsymbol{V}_i^{-1} [\boldsymbol{Y}_i - \boldsymbol{\mu}_i]$$

4. Repeat Step2 and Step3 until convergence.

Generalized Linear Mixed Models (GLMM)

One can extend the model introduced in Section 2.1.2 to a generalized linear model where outcomes are distributed according to a distribution that belongs to the exponential family of distributions (equation (2.2)) and with

$$E[Y_{ij} | \boldsymbol{X}_{ij}, \boldsymbol{b}_i] = \mu_{ij},$$

where Y_{ij} and $\boldsymbol{X}_{ij}$ are outcome and vector of covariates for the j^{th} measurement on the i^{th} cluster, respectively. $\boldsymbol{b}_i$ is a vector of random effects associated with the i^{th} cluster and of size $q \times 1$. The mean μ_{ij}^b, vector of fixed covariates $\boldsymbol{X}_{ij}$, fixed effect coefficients $\boldsymbol{\beta}$, vector of random covariates $\boldsymbol{Z}_{ij}$, and random effect coefficients $\boldsymbol{b}_i$ are linked via a link function $g(\cdot)$,

where

$$g(\mu_{ij}) = \boldsymbol{\beta}^T \boldsymbol{X}_{ij} + \boldsymbol{b}_i^T \boldsymbol{Z}_{ij}. \tag{2.3}$$

Typically, $\boldsymbol{b}_i$'s, where $i \in \{1, \ldots, n\}$, are assumed to be distributed according to a multivariate Normal distribution

$$\boldsymbol{b}_i \sim N_q(0, \boldsymbol{\Sigma}_b).$$

Further, within each cluster i, and conditional on the cluster random effects $\boldsymbol{b_i}$, Y_{ij}'s are assumed to be conditionally independent, where $j \in \{1, \ldots, m_i\}$. $\boldsymbol{\beta}$ and $\boldsymbol{\Sigma}_b$ parameter estimation can be done using the maximum likelihood method, where the likelihood of the data is

$$L(\boldsymbol{\beta}, \boldsymbol{\Sigma}_b) = \prod_{i=1}^{n} f(\boldsymbol{Y}_i | \boldsymbol{X}_i),$$

with n as the number of clusters, $\boldsymbol{Y}_i$ a vector of outcomes in cluster i, and $\boldsymbol{X}_i$ the fixed effect design matrix in cluster i. In order to get the marginal likelihood $f(\boldsymbol{Y}_i | \boldsymbol{X}_i)$, one can integrate out random effects $\boldsymbol{b}_i$ from the full likelihood $f(\boldsymbol{Y}_i | \boldsymbol{X}_i, \boldsymbol{b}_i)$ as

$$\begin{aligned} f(\boldsymbol{Y}_i | \boldsymbol{X}_i) &= \int_b f(\boldsymbol{Y}_i | \boldsymbol{X}_i, \boldsymbol{b}_i) f(\boldsymbol{b}_i | \boldsymbol{X}_i) d\boldsymbol{b}_i \\ &= \int_b \prod_{j=1}^{n} f(Y_{ij} | \boldsymbol{X}_i, b_i) f(\boldsymbol{b}_i | \boldsymbol{X}_i) d\boldsymbol{b}_i. \end{aligned} \tag{2.4}$$

When the Y_{ij}'s are not normally distributed, evaluation of the integral in equation (2.4) is rather difficult. By approximating the data, Zeger et al. (1988) and Breslow and Clayton (1993) proposed parameter estimation methods that approximate the maximum likelihood method. Other possible estimation methods include a numerical maximization of the log likelihood using Gauss-Hermite Quadrature technique (Liu and Pierce (1994)), parameter

estimation using the EM algorithm, or parameter estimation by approximating the integral in equation (2.4) using Monte-Carlo techniques.

In particular, Breslow and Clayton (1993) proposed a penalized quasi-likelihood (PQL) method where they used a first-order Taylor expansion to approximate data with pseudo data. They, then, estimated GLMM parameters by using the pseudo data and through an iterative approach. They considered a GLMM of the form

$$\begin{aligned} Y_{ij} &= \mu_{ij} + \epsilon_{ij} \\ &= g^{-1}(\boldsymbol{\beta}^T \boldsymbol{X}_{ij} + \boldsymbol{b}_i^T \boldsymbol{Z}_{ij}) + \epsilon_{ij}, \end{aligned}$$

where $g(.)$ is the canonical link function, $\boldsymbol{X}_{ij}$ and $\boldsymbol{Z}_{ij}$ are fixed effect and random effect vectors for the j^{th} measurement on the i^{th} cluster, respectively, and ϵ_{ij} is some measurement error with a variance of $Var[Y_{ij}|\boldsymbol{b}_i] = V(\mu_{ij})$. Breslow and Clayton (1993) considered a first order Taylor expansion around a current $\boldsymbol{\beta}^*$ and $\boldsymbol{b}_i^*$ estimates as

$$\begin{aligned} Y_{ij} \approx g^{-1}(\boldsymbol{\beta}^{*T} \boldsymbol{X}_{ij} + \boldsymbol{b}_i^{*T} \boldsymbol{Z}_{ij}) + \dot{g}^{-1}(\boldsymbol{\beta}^{*T} \boldsymbol{X}_{ij} + \boldsymbol{b}_i^{*T} \boldsymbol{Z}_{ij})(\beta - \beta^*)^T \boldsymbol{X}_{ij} \\ + \dot{g}^{-1}(\boldsymbol{\beta}^{*T} \boldsymbol{X}_{ij} + \boldsymbol{b}_i^{*T} \boldsymbol{Z}_{ij})(\boldsymbol{b}_i - \boldsymbol{b}_i^*)^T \boldsymbol{Z}_{ij} + \epsilon_{ij}. \end{aligned}$$

One can equivalently write

$$\begin{aligned} Y_{ij} &\approx \mu_{ij}^* + V(\mu_{ij}^*)(\boldsymbol{\beta} - \boldsymbol{\beta}^*)^T \boldsymbol{X}_{ij} && (2.5) \\ &\quad + V(\mu_{ij}^*)(\boldsymbol{b} - \boldsymbol{b}_i^*)^T \boldsymbol{Z}_{ij} + \epsilon_{ij}. && (2.6) \end{aligned}$$

By re-arranging equation (2.5) and separating unknown parameters $\boldsymbol{\beta}$ and $\boldsymbol{b}$ from the current estimates $\boldsymbol{\beta}^*$ and $\boldsymbol{b}_i^*$,

$$\beta^T \boldsymbol{X}_{ij} + \boldsymbol{b_i}^T \boldsymbol{Z}_{ij} + V({\mu_{ij}}^*)^{-1}\epsilon_{ij} \approx V({\mu_{ij}}^*)^{-1}(Y_{ij} - \mu_{ij}^*) + \boldsymbol{\beta}^{*T} \boldsymbol{X}_{ij} + \boldsymbol{b}_i^{*T} \boldsymbol{Z}_i. \quad (2.7)$$

Breslow and Clayton (1993) called the right hand-side of the equation (2.7) as a pseudo response, that is represented by Y_{ij}^*, and showed that GLMM parameters could be estimated by iteratively fitting the linear mixed model

$$Y_{ij}^* = \boldsymbol{\beta}^T \boldsymbol{X}_{ij} + \boldsymbol{b}_i{}^T \boldsymbol{Z}_{ij} + V(\mu_{ij}{}^*)^{-1}\epsilon_{ij}.$$

Alternatively, one can estimate parameters in a GLMM model by approximating the integral in equation (2.4) using numerical approximations. Gaussian quadrature method is an example of such approximation. Consider the likelihood for subject i in equation (2.4). In this equation, $f(\boldsymbol{b}_i|\boldsymbol{X}_i)$ is the density of a multivariate Normal distribution. Using Gaussian quadrature approximation, one can approximate this integral as

$$\begin{aligned} f(\boldsymbol{Y}_i|\boldsymbol{X}_i) &= \int_b f(\boldsymbol{Y}_i|\boldsymbol{X}_i, \boldsymbol{b}_i) f(\boldsymbol{b}_i|\boldsymbol{X}_i) d\boldsymbol{b}_i \\ &\approx \sum_{q=1}^{Q} W_q f(\boldsymbol{Y}_i|\boldsymbol{X}_i, \boldsymbol{b}_i^{(q)}), \end{aligned} \tag{2.8}$$

where Q is the order of approximation, W_q and $\boldsymbol{b}_i^{(q)}$ are weights and nodes that are solutions to the Q^{th} order Hermite polynomial.

2.1.4 Bayesian Longitudinal Models

In this section, we introduce Bayesian linear mixed effect and generalized linear mixed effect models to analyze correlated data.

Bayesian Linear Mixed Effect Models

Consider the linear mixed effect model introduced in equation (2.1). To fit this model under the Bayesian framework, where fixed effect coefficient vector $\boldsymbol{\beta}$, random effect coefficient vector $\boldsymbol{b}_i$ where $i \in \{1, \ldots, n\}$ with n as the number of clusters, and measurement error σ^2 are the parameters, one can consider a hierarchical model with the likelihood

$$L = \prod_{i=1}^{n} f(\boldsymbol{Y}_i | \boldsymbol{X}_i, \boldsymbol{Z}_i, \boldsymbol{\beta}, \boldsymbol{b}_i, \sigma^2), \tag{2.9}$$

where $\boldsymbol{Y}_i$ is a vector of outcome measures on cluster i, $\boldsymbol{X}_i$ and $\boldsymbol{Z}_i$ are fixed effect and random effect design matrices, respectively. Typically, a multivariate Normal distribution is assumed on random effect coefficients as

$$\boldsymbol{b}_i \sim N(0, \boldsymbol{\Sigma_b}), \tag{2.10}$$

where $i \in \{1, \ldots, \}$, and with independent priors on the fixed effect coefficients $\boldsymbol{\beta}$ and the measurement error σ^2 as

$$f(\boldsymbol{\beta}, \sigma^2) = f(\boldsymbol{\beta}) f(\sigma^2). \tag{2.11}$$

In the model setting above, while typically $\boldsymbol{\Sigma}_b$ is fixed, some may treat it as a hyper-parameter with a proper prior $f(\boldsymbol{\Sigma}_b)$, commonly an inverse-Wishart distribution. Common prior for $\boldsymbol{\beta}$ is the multivariate Normal distribution. σ^2 is commonly assumed to be independent of $\boldsymbol{\beta}$ with a prior inverse-gamma distribution. With the specified likelihood and the priors, the

posterior distribution of the parameters is of the form

$$f(\boldsymbol{\beta}, \boldsymbol{b}_1, \ldots, \boldsymbol{b}_n, \sigma^2, \boldsymbol{\Sigma}_b | \boldsymbol{X}_1, \ldots, \boldsymbol{X}_n, \boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_n, \boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_n) \propto \prod_{i=1}^{n} f(\boldsymbol{Y}_i | \boldsymbol{X}_i, \boldsymbol{Z}_i, \boldsymbol{\beta}, \boldsymbol{b}_i, \sigma^2) \times f(\boldsymbol{b}_i | \boldsymbol{\Sigma}_b) \times f(\boldsymbol{\beta}) f(\sigma^2) \times f(\boldsymbol{\Sigma}_b) \quad (2.12)$$

Bayesian Generalized Linear Mixed Effect Models

Consider the generalized linear mixed effect model (GLMM) introduced in equation (2.3). Bayesian GLMM includes a likelihood of the form

$$L = \prod_{i=1}^{n} f(\boldsymbol{Y}_i | \boldsymbol{X}_i, \boldsymbol{Z}_i, \boldsymbol{\beta}, \boldsymbol{b}_i),$$

where the likelihood terms $f(\boldsymbol{Y}_i | \boldsymbol{X}_i, \boldsymbol{Z}_i, \boldsymbol{\beta}, \boldsymbol{b}_i)$ are distributed according to an exponential family distribution (equation (2.2)). $\boldsymbol{Y}_i$ is a vector of outcome measures on cluster i, $\boldsymbol{X}_i$ and $\boldsymbol{Z}_i$ are fixed effect and random effect design matrices, $\boldsymbol{\beta}$ is a vector of fixed effect coefficients, and $\boldsymbol{b}_i$ is a vector of random effect coefficients for cluster i. Typically, assumed random effect coefficients $\boldsymbol{b}_i$ are distributed according to a multivariate Normal distribution

$$\boldsymbol{b}_i \sim N(0, \boldsymbol{\Sigma_b}), \quad (2.13)$$

where $i \in \{1, \ldots, n\}$, and with a $\boldsymbol{\Sigma_b}$ with a typical inverse-Wishart prior. Fixed effect coefficients $\boldsymbol{\beta}$ and measurement error σ^2 are assumed to have independent priors

$$f(\boldsymbol{\beta}, \sigma^2) = f(\boldsymbol{\beta}) f(\sigma^2). \quad (2.14)$$

With the specified likelihood and priors, the joint posterior distribution for parameters will be the same as in equation (2.12).

2.2 Survival Analysis

2.2.1 Introduction to Survival Analysis

Time-to-event analyses are generally referred to as survival analysis, where interest is in estimating the length of time to some event of interest and to identify factors that may affect this length of time. Using survival terminology, we refer to the event of interest as "death", even though death is not the event of interest in all time-to-event analyses. Also, for simplicity, we refer to the time till the event of interest as "survival" time. One example of survival analysis is modeling time-to-death among end-stage renal disease (ESRD) patients as a function of patient's age, BMI, and serum albumin.

One complication of survival data is that not all subjects experience "death" during the study follow-up. Also, some may leave the study or may die to an unrelated cause. While these subjects' event times are un-observed, their event times are beyond the last time point when those subjects were observed in the study. In survival terminology, these subjects are known as "right censored" subjects. Right censoring is considered as a type of missingness and should be accounted for in the analysis when they occur. While conventional statistical methods, including regression models, can be used to analyze time-to-event data in the absence of censoring, however, when censoring occurs, these models typically fail to account for censored data. Survival analysis, on the other hand, takes censoring into account.

To formally introduce survival analysis, let T and C be continuous non-negative time-to-event outcome, also known as survival time, and censoring time, respectively. Observed

time Y, is either the event time T, if the event of interest observed, or is the censoring time C, if the subject gets censored. When an event-time is observed, we set the event-indicator $\delta = 1$, otherwise $\delta = 0$. It's typically assumed survival time, T, and censoring time, C, are independent. Observed time Y can be mathematically written as:

$$Y = \begin{cases} T, & \text{if } \delta = 1 \\ C, & \text{if } \delta = 0 \end{cases}$$

For a non-negative random variable T, four functions below characterize the distribution of this random variable:

1. Probability Density Function (pdf):

$f(t)$ is the probability density function that the event occurs at time t and can be written as

$$f(t) = \lim_{\Delta t \to 0^+} \frac{1}{\Delta t} Pr[t \leq T \leq t + \Delta t]. \tag{2.15}$$

2. Survival Function:

Survival Function $S(t)$ is the probability that a subject survives up to time t. As the definition implies, survival function is also equal to the complement of a cumulative distribution $F(t)$ up until time t. Survival function can be written as

$$S(t) = Pr[T > t] = 1 - Pr[T \leq t] = 1 - F(t) = 1 - \int_0^t f(z)dz. \tag{2.16}$$

3. Hazard Rate:

Hazard rate $\lambda(t)$ is the chance that a subject who has survived up to time t, experiences the event of interest at the next instant in time. Hazard rate can be formally written as

$$\lambda(t) = \lim_{dt \to 0^+} \frac{Pr[t \leq T \leq t + dt | T \geq t]}{dt}. \tag{2.17}$$

4. Cumulative Hazard Function:

Cumulative hazard function $\Lambda(t)$ is simply an integration over hazard rate up until time t as

$$\lambda(t) = \int_0^t \lambda(z) dz. \tag{2.18}$$

The four functions introduced in equation (2.15), equation (2.16), equation (2.17), and equation (2.18) are related as when one function is known, the other three can be derived. In specific, using the definition of the hazard rate (equation 2.17) and by extending the conditional probability in the numerator, one can write

$$\lambda(t) = \lim_{dt \to 0^+} \frac{1}{dt} \frac{Pr[t \leq T \leq t + dt]}{Pr[T \geq t]} = \lim_{dt \to 0^+} \frac{1}{dt} \frac{f(t)dt}{S(t)} = \frac{f(t)}{S(t)}. \tag{2.19}$$

Given that the derivative of the survival function $S(t)$ is equal to $-f(t)$ one can write $\lambda(t) = \frac{f(t)}{S(t)} = -\frac{d}{dt} log(S(t))$. Consequently, we can re-write the survival function as

$$S(t) = exp\{-\int_0^t \lambda(z) dz\} = exp\{-\Lambda(t)\}. \tag{2.20}$$

equation (2.19) and equation (2.20) show that similar to the survival function, the hazard function also provides equivalent information on the distribution of the survival time, T.

Further, accounting for right-censoring is easier with the hazard function as this function at each time t, is a conditional function that only needs a risk set at that time, where a risk at time t includes censored subjects at time t as well. Formally, risk set at time t is defined as a set of all subject who have not failed until time t. As an example, one may consider a simple survival distribution where the risk of death over time is constant. This means, a constant hazard of the form $\lambda(t) = \lambda$ where λ is a parameter that is constant for all values of t. Using E.2.20, the corresponding survival function is of the form $S(t) = exp\{-\int_0^t \lambda dz\} = exp(-\lambda t)$. Hence, the probability density function of survival times under this setting is $f(t) = -\frac{d}{dt}S(t) = \lambda exp\{-\lambda t\}$, that is an exponential distribution with the mean survival time of $1/\lambda$.

With survival and a hazard rate of $\lambda(t)$, one can form a likelihood function. Consider survival data on n subjects, some of whom may have been censored. If subject i dies at time t_i (i.e. survival time t_i with event indicator $\delta_i = 1$), her/his contribution to the likelihood is $f(t_i)$ which instead, can be written as $S(t_i)\lambda(t_i)$. On the other hand, If subject i gets censored (i.e. survival time t_i with event indicator $\delta_i = 0$), she/he is still alive at t_i and hence, her/his contribution to the likelihood is $S(t_i)$. The corresponding likelihood is then

$$\begin{aligned} L &= \prod_{i=1}^{n} L_i \\ &= \prod_{i=1}^{n} \lambda(t_i)^{\delta_i} S(t_i). \end{aligned} \tag{2.21}$$

Using equation (2.20), one can write the log likelihood as

$$\begin{aligned} log(L) &= \sum_{i=1}^{n} log(L_i) \\ &= \sum_{i=1}^{n} \delta_i log(\lambda(t_i)) + log(S(t_i)) \\ &= \sum_{i=1}^{n} \{\delta_i log\big(\lambda(t_i)\big) - \Lambda(t_i)\}. \end{aligned} \tag{2.22}$$

2.2.2 Cox Proportional Hazards Model

In survival analysis, often researchers are interested in testing the association between a set of covariates and survival times. Focused on the hazard function, Cox (1972) introduced a family of models to model survival data. In particular, he introduced the proportional hazards model where he used the multiplicative hazard function

$$\lambda(t|\boldsymbol{Z}) = \lambda_0(t)exp\{\boldsymbol{\beta}^T\boldsymbol{Z}\}, \tag{2.23}$$

where $exp(.)$ is the exponential function with $expx = e^x$. Based on equation (2.23), the hazard function at any particular time point can be decomposed into a time-dependent baseline hazard $\lambda_0(t)$ and a time-independent covariate effect $exp\{\boldsymbol{\beta}^T\boldsymbol{Z}\}$, where $\boldsymbol{Z}$ is a vector of covariates, $\lambda_0(t)$ is baseline hazard at time t, $\boldsymbol{\beta}$ is a vector of coefficients, and $exp\{\boldsymbol{\beta}^T\boldsymbol{Z}\}$ serves as the relative risk which can increase or decrease the risk proportionately depending on the covariate values. The hazard is a proportional hazard in the sense that the increase or decrease in the hazard rate, $\lambda(t)$, compared to the baseline hazard, $\lambda_0(t)$, is the same at all values of t.

Baseline hazard, $\lambda_0(t)$, is usually unknown and difficult to specify without a strong parameteric assumption. Without a baseline hazard specification, survival probability density is unknown, and hence, forming a full likelihood in order to estimate $\boldsymbol{\beta}$ coefficients is not possible. Cox (1972) proposed a semi-parameteric model where $\boldsymbol{\beta}$ coefficients are estimated without any baseline hazard specification. He proposed using a partial likelihood where at each event time, covariate values for the failed subject are compared with the covariate values of all subjects in the risk set, who have not failed yet. The contribution of the i^{th} subject

to the partial likelihood, L_i^P, can be written as

$$L_i^P = Pr\{\text{Subject with covariate vector } \boldsymbol{Z}_i \text{ failed at } t_i | \text{ Some subject from the risk set } R(t_i) \text{ failed at } t_i\},$$

where $R(t_i)$ is the risk set at time t_i that includes all subjects who have not died by this time yet. Using the definition of conditional probability, one can write L_i^P as

$$\begin{aligned} L_i^P &= \frac{\lambda_0(t_i) exp(\boldsymbol{\beta}^T \boldsymbol{Z}_i)}{\sum_{k \in R(t_i)} \lambda_0(t_i) exp(\boldsymbol{\beta}^T \boldsymbol{Z}_k)} \\ &= \frac{exp(\boldsymbol{\beta}^T \boldsymbol{Z}_i)}{\sum_{k \in R(t_i)} exp(\boldsymbol{\beta}^T \boldsymbol{Z}_k)}. \end{aligned}$$

The partial likelihood for all subjects, L^P, is the multiplication of the individual likelihood contribution as

$$L^P = \prod_{\text{failure times i}} \frac{exp(\boldsymbol{\beta}^T \boldsymbol{Z}_i)}{\sum_{k \in R(t_i)} exp(\boldsymbol{\beta}^T \boldsymbol{Z}_k)}. \tag{2.24}$$

Maximization of the partial likelihood can be done by solving the partial likelihood score equation and by using the Newton-Raphson technique. Plausibility of coefficient estimation using the partial likelihood has been justified by authors including Cox (1975) and Andersen and Gill (1982).

Proportional hazard models are not limited to only fixed-at-baseline covariates. One can include time-varying covariates, $\boldsymbol{Z}(t)$, into the model, where now $\boldsymbol{Z}(t)$ is a vector of covariates, some of which may change over time. Consider a multiplicative hazard model of the form

$$\lambda(t|\boldsymbol{Z}(.)) = \lambda_0(t) exp\{\boldsymbol{\beta}^T \boldsymbol{Z}(t)\}. \tag{2.25}$$

The corresponding partial likelihood for this model is of the form

$$L^P = \prod_{\text{failure times i}} \frac{exp(\boldsymbol{\beta}^T \boldsymbol{Z}_i(t_i))}{\sum_{k \in R(t_i)} exp(\boldsymbol{\beta}^T \boldsymbol{Z}_k(t_i))}, \tag{2.26}$$

where $\boldsymbol{Z}_k(t_i)$ indicates the value of the covariate vector for subject k at time t_i. Definition of the risk set $R(t_i)$ is the same as before. While fixed-at-baseline covariate values remain the same for each subject throughout the study, however, time-varying covariates change over time. This means in order to evaluate the likelihood, one needs to compute time-varying covariates for all subjects in the risk set at each failure time point, t_i, where $i \in \{1, \ldots, k\}$ with k as the number of failure time points.

2.2.3 Weibull Model - A Parametric Proportional Hazards Model

One may consider survival times t to be distributed according to a Weibull distribution with a shape parameter, τ, and a scale parameter, $\gamma = exp(\lambda)$. According to Ibrahim et al. (2005), we consider a Weibull distribution with the parameterization

$$\begin{aligned} t|\tau, \lambda &\sim Weibull(\tau, \lambda) \\ f(t|\tau, \lambda) &= \tau t^{\tau-1} exp(\lambda - exp(\lambda) t^{\tau}). \end{aligned} \tag{2.27}$$

Using the Weibull density in equation (2.27), one can derive the survival function

$$S(t|\tau, \lambda) = exp\{-exp(\lambda) t^{\tau}\},$$

and the hazard function

$$\begin{aligned}\lambda(t) &= \frac{f(t)}{S(t)} \\ &= \tau t^{\tau-1} exp(\lambda). \end{aligned} \tag{2.28}$$

equation (2.28) is similar to the proportional hazard equation (equation 2.23), where it can be decomposed into a time-dependent baseline hazard of the form $\tau t^{\tau-1}$ and a time-independent multiplier of the form $exp(\lambda)$. In that sense, Weibull model is also a proportional hazards model. One can introduce covariates into the model using the scale parameter, where $\lambda = \boldsymbol{\beta}^T \boldsymbol{Z}$.

Considering the equation (2.28), for a shape parameter $\tau = 1$, the model reduces to an exponential survival distribution with a constant baseline hazard. A shape parameter greater than 1 leads to a model with an increasing baseline risk over time. Conversely, a shape parameter less than 1, leads to a model with a decreasing baseline risk over times.

By taking log of the equation (2.28), one can observe that the log of the risk over time changes linearly with time. This indicates that a log-linear changes in the baseline hazard over time in the Weibull models.

2.2.4 Bayesian Survival Analysis

While it's possible to approach survival analysis from the frequentist perspective, there are quite some advantages considering survival analysis from the Bayesian perspective, some of which are reviewed in this section.

Fitting a Bayesian survival model is generally easier than a frequentist model, specially when there exists a complex censoring mechanism. In this case, instead of maximizing a

complex likelihood, one can use posterior sampling computation with MCMC methods to fit the model. Further, while frequentist methods rely on asymptotics and require a minimum sample size, an exact inference is possible with Bayesian models through posterior sampling computation. This matter becomes more obvious in estimating variance in frequentist models as opposed to variance estimation under the Bayesian paradigm. Additionally, in clinical trials it is often of interest to incorporate information from prior studies. While the Bayesian perspective provides a natural way to incorporate prior information, this is not possible with the frequentist approach. Finally, while model selection under the frequentist approach is mainly limited to nested models and is often complex, the Bayesian paradigm provides easier model selection techniques both for nested as well as non-nested models (Ibrahim et al. (2005)).

Bayesian Weibull Model

Consider $\{t_1, t_2, \ldots, t_n\}$ as n independent survival times that are distributed identically according to a Weibull distribution with a parameterization introduced in the equation (2.27). Using the equation (2.21), one can write the full likelihood for a censored sample and under standard censoring assumptions

$$\begin{aligned} L(\alpha, \lambda) &= \prod_{i=1}^{n} (\tau {t_i}^{\tau-1} exp\{\lambda\})^{\delta_i} (exp\{-exp(\lambda) t^{\tau}\}) \\ &= \tau^{\sum_{i=1}^{n} \delta_i} exp\Big\{\lambda \sum_{i=1}^{n} \delta_i + \sum_{i=1}^{n} \big(\delta_i(\tau - 1) log(t_i) - exp(\lambda) {t_i}^{\tau}\big)\Big\}. \end{aligned}$$

In terms of the prior specification on the parameters, when both α and λ are unknown, there is no conjugate prior. A typical prior for α is a Gamma prior, and a typical prior for λ is a Normal prior. With these priors, the posterior is not in a closed form, however, by using MCMC methods, one can easily sample from the joint posterior distribution of the parameters.

Using the parameterization introduced in equation (2.27), one can introduce covariates, where $\lambda = \boldsymbol{\beta}^T \boldsymbol{Z}$. The Normal distribution is a typical prior on $\boldsymbol{\beta}$ coefficients. Under this setting, one can write the likelihood as

$$L(\alpha, \boldsymbol{\beta}) = \tau^{\sum_{i=1}^{n} \delta_i} exp\big\{(\boldsymbol{\beta}^T Z) \sum_{i=1}^{n} \delta_i + \sum_{i=1}^{n} \big(\delta_i(\tau - 1) log(t_i) - exp(\boldsymbol{\beta}^T \boldsymbol{Z}) {t_i}^{\tau}\big)\big\}$$

Semi-Parametric Bayesian Survival Models Using a Dirichlet Process Prior

With today's efficient computational algorithms, including MCMC methods, non-parameteric and semi-parameteric Bayesian models, with fewer parameteric assumptions compared to the parameteric models, are becoming popular. Focusing on the introduced Bayesian Weibull model as a proportional hazards model, one may be interested in a model with fewer parameteric distributional assumptions to model real data better. One popular solution is to use the non-parameteric Dirichlet process prior technique to limit explicit distributional assumptions (Section 2.4). In this section, we briefly introduce semi-parameteric survival models by using the Dirichlet process technique. A detailed introduction to the Dirichlet process models will be provided in Section 2.4.

Survival models with Dirichlet process prior were first studied by Susarla and Van Ryzin (1976) and Ferguson and Phadia (1979), where by using the Dirichlet process prior, they proposed a fully non-parameteric method of estimation for survival curves with right censored data. Under the squared error loss, Susarla and Van Ryzin (1976) derived an estimator for survival function of the form

$$\hat{S}(t) = \frac{\alpha(1 - G_0(t)) + N^+(t)}{\alpha + n} \times \prod_{j=k+1}^{l} \big(\frac{\alpha(1 - G_0(t_{(j)})) + N(t_{(j)})}{\alpha(1 - G_0(t_{(j)})) + N(y_{(j)}) - \lambda_j}\big),$$

where $\{t_1, t_2, \ldots, t_n\}$ are survival times with $\{t_1, t_2, \ldots, t_k\}$ as the observed event times and the remaining $\{t_{k+1}, t_{k+2}, \ldots, t_n\}$ as the censoring times. Among the censoring times, suppose

there are $m-k$ unique time points of the form $\{t_{(k+1)}, t_{(k+2)}, \ldots, t_{(m)}\}$. and suppose λ_j counts the number of censored observations at each unique censoring time $t_{(j)}$ where $j \in \{k+1, k+2, \ldots, m\}$. $N(t)$ is the number of observations in the risk set with an observed time (censoring or event time) greater than or equal to t. Similarly, $N(t^+)$ is the number of observations in the risk set with an observed time greater than t. Finally, α is the concentration parameter and G_0 is the base measure in the Dirichlet process prior.

A survival model with a Dirichlet process mixture (DPM) prior can be considered as a compromise between a completely parametric survival model and a completely non-parameteric model. Some authors, including Doss (1994), Doss and Narasimhan (1998), and Doss and Huffer (2000), proposed using the Dirichlet process mixture prior on modeling the cumulative probability of survival times. Their proposed methods are rather difficult when one needs to introduce covariates into the model. Alternatively, one can assume a parameteric distribution for the survival times with an unknown parameter vector θ. In order to add more flexibility to the model, one can assume that θ is distributed according to an unknown distribution G, where G itself has a Dirichlet process prior. This model can be written as

$$
\begin{aligned}
t|\theta &\sim F(.|\theta) \\
\theta &\sim G \\
G &\sim DP(\alpha, G_0).
\end{aligned}
$$

This model specification induces a marginal distribution on survival times t that is an infinite-mixture of the parameteric distribution F that is mixed over the θ parameter vector (Sethuraman (1994)). Dirichlet process (DP) and Dirichlet process mixture models are introduced in detail in Section 2.4.

Frailty Models

In survival studies, survival times may depend on covariates. Not all covariates are measurable, as some may be unknown or if known, not possible to measure. Vaupel et al. (1979) used the word frailty (also known as individual's heterogeneity) to refer to these unobservable risk factors.

Among latent risk factors, some may cluster a population into some sub-populations where members of a sub-population share the same level of a latent risk factor that is different from other sub-populations. Hence, frailty models are useful tools in modeling the within-cluster association that exists between subjects' survival times within each cluster.

Frailty models have been approached both from a parameteric as well as a semi-parameteric point of view. With the advances in computational algorithms, the Bayesian paradigm in frailty models has got a lot of attention in the literature. In particular, Sahu et al. (1997) approached frailty models using a fully parameteric approach with a baseline hazard specified according to a Weibull distribution. Clayton (1991), Sinha and Dey (1997), and Aslanidou et al. (1998) approached frailty models from a Bayesian semi-parameteric standpoint. Clayton (1991) proposed a frailty model with a Gamma process prior. Sinha and Dey (1997) and Aslanidou et al. (1998) proposed a frailty model with piece-wise exponential baseline hazards. Gustafson (1997) and Sargent (1998) considered a proportional hazards model with frailty using the Cox's partial likelihood.

The most common frailty model is an extension of the Cox proportional hazards model and is commonly known as the shared frailty model. This model considers n clusters each with m_i within cluster subjects. Survival time and covariate vectors for the j^{th} subject in the i^{th} cluster are denoted by t_{ij} and $\boldsymbol{X}_{ij}$, respectively. w_i is considered as the subject i's frailty term that is latent. Mirroring the notation from Ibrahim et al. (2005), a frailty model is of

the form

$$\lambda(t|w_i, \boldsymbol{X}_{ij}) = \lambda_0(t) w_i exp(\boldsymbol{\beta}^T \boldsymbol{X}_{ij}), \tag{2.29}$$

where $\lambda_0(t)$ is an unknown baseline hazard that is shared across all subjects. $\boldsymbol{X}_{ij}$ is the covariate for the j^{th} subject from the i^{th} cluster, that is of size $p \times 1$. $\boldsymbol{\beta}$ is a vector of coefficients that is also of the size $p \times 1$. This model can be approached from the parameteric perspective, where a parameteric distribution is assumed on the frailty terms w_i. Alternatively, by using semi-parameteric techniques, one can relax explicit distributional assumptions on w_i (Vaupel et al. (1979)).

2.3 Gaussian Processes

Gaussian process (GP) can be thought as a Bayesian non-parameteric technique that is widely used to define a prior distribution over functions and has been used in statistics literature for a long time (O'Hagan and Kingman (1978) ,Wahba (1990), Rasmussen (2006), Neal (2012)). This technique is particularly useful in non-parametric Bayesian regression models where instead of an explicit functional assumption, one may use a Gaussian process prior to relax any functional assumption. Like functions, a prior distribution on functions is also infinite-dimensional. Such a prior can be defined using Gaussian process (O'Hagan and Kingman (1978)). Gaussian process has been known in spatial statistics for a long time where GP regression is known as "Kriging". In general, Gaussian processes can be considered as a general-purpose regression with no explicit functional assumption and hence, with a great flexibility.

Rasmussen (2006) defines Gaussian process as a collection of random variables, any finite number of which are distributed according to a multivariate Gaussian distribution. Let

$\boldsymbol{F} = (f(t_1), f(t_2), \ldots, f(t_N))$ be an N-dimensional random vector of function values evaluated at N input points $t_i \in \mathcal{T}$, where $i \in \{1, \ldots, N\}$. A random function f is distributed according to a Gaussian process if for any finite subset $\{t_1, t_2, \ldots, t_N\} \subset \mathcal{T}$, $\boldsymbol{F}$ is distributed according to a multivariate Gaussian distribution.

Gaussian processes are fully specified by a mean function, $\mu(t)$, and a covariance function, $C(t, t')$, where t and t' are two values from the input space. This can be formally written as

$$f(t) \sim GP(\mu(t), C(t, t')),$$

where realizations of GP are random functions $f(t)$.

A GP, as a collection of random variables, on a finite subset is a multivariate Gaussian distribution that satisfies all properties of a multivariate Gaussian distribution (Rasmussen (2006)). In particular, consider $\boldsymbol{t}_1$ and $\boldsymbol{t}_2$ as two finite input vectors and assume $f(t)$ is distributed according to a Gaussian process

$$f(t) \sim GP(\mu(t), C(t, t')).$$

Since $\boldsymbol{t}_1$ and $\boldsymbol{t}_2$ are finite vectors, $(\boldsymbol{t}_1, \boldsymbol{t}_2)$, as the combination of two finite vectors, is still finite. Hence, by the definition of the Gaussian process that is evaluated on a finite subset, one can write

$$\begin{pmatrix} \boldsymbol{t}_1 \\ \boldsymbol{t}_2 \end{pmatrix} \sim N(\begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix}, \begin{pmatrix} \boldsymbol{\Sigma}_{1,1} & \boldsymbol{\Sigma}_{1,2} \\ \boldsymbol{\Sigma}_{2,1} & \boldsymbol{\Sigma}_{2,2} \end{pmatrix}).$$

Based on the consistency requirement, $\boldsymbol{t}_1 \sim N(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{1,1})$ and $\boldsymbol{t}_2 \sim N(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_{2,2})$.

Gaussian process models are particularly useful in the regression context where the interest lies in avoiding explicit functional assumptions between covariates and the outcome in a

regression model. In a prediction problem in the regression context, one may be interested in predicting an output y^* for a new input t^*, given the observed data $\boldsymbol{D} = \{(t_i, y_i), i = 1, 2, \ldots, N\}$, where $y_i = f(t_i) + \epsilon_i$, with an unknown function f. By using a Gaussian process prior, one can consider a pool of plausible functions as opposed to any explicit functional assumption on the function f. This means, assuming $f(t) \sim GP(\mu(t), C(t, t'))$. In this situation, in order to predict Y^*, one can write

$$P(Y^*|t^*, \boldsymbol{D}) = \int P(Y^*|t^*, \boldsymbol{D}, f) dP(f|\boldsymbol{D})$$

In application, Gaussian processes are often used as a mean zero stochastic process that is added to a parameteric mean trend in order to add additional flexibility in modeling data. A mean-zero Gaussian process leaves the covariance function specification as the main modeling choice. Smoothness and differentiability in Gaussian process models are controlled by the covariance function. A variety of different covariance functions have been proposed in the literature, with squared exponential and Matérn covariance functions as two such examples.

The squared exponential covariance function is of the form

$$Cov(Y_i, Y_j) = \kappa^2 exp\{-\rho^2 (t_i - t_j)^2\}, \tag{2.30}$$

where the covariance between outcome Y_i and Y_j is a function of the input space $\mathcal{T}$. The parameter κ^2 controls the height of oscillation and the parameter ρ^2 controls the correlation length. The squared exponential covariance function specified above is infinitely differentiable. One can extend the squared exponential covariance function from a single input space to a multivariate input space as

$$Cov(Y_i, Y_j) = \kappa^2 exp\{\sum_{u=1}^{p} \rho^2 (x_{iu} - x_{ju})^2\}, \tag{2.31}$$

Another common Gaussian process covariance function is Matérn covariance function that is given by

$$Cov(Y_i, Y_j) = \frac{2^{1-\nu}}{\Gamma(\nu)}\big(\frac{\sqrt{2\nu}d}{l}K_\nu\big(\frac{\sqrt{2\nu}d}{l}\big)\big), \tag{2.32}$$

where ν and l are positive parameters, d is the distance between t_i and t_j, and K_ν is a modified Bessel function (Abramowitz et al. (1966)). As pointed out by Rasmussen and Williams (2006), by specifying $\nu = p + 1/2$, where p is a positive integer, the Matérn covariance function is simplified to the form

$$Cov(Y_i, Y_j) = exp\big(-\frac{\sqrt{2\nu}r}{l}\big)\frac{\Gamma(p+1)}{\Gamma(2p+1)}\sum_{i=0}^{p}\frac{(p+i)!}{i!(p-i)!}\big(\frac{\sqrt{8\nu}r}{l}\big)^{p-i}. \tag{2.33}$$

Covariance function specification, even with scientific knowledge and rich data, is by no means a trivial problem (Flaxman et al. (2015)). While in some applications, maximum likelihood estimations of the parameters of the covariance function is used, we emphasize treating these parameters as hyper-parameters that are sampled as part of an MCMC process.

As an example of a Gaussian process prior, we randomly sampled 15 functions from a Gaussian process with a squared exponential covariance function, where $\kappa^2 = 1$, $\rho^2 = 0.5$, and with an input space $\mathcal{T} = (-5, 5)$ (Figure 2.1). Using observed data, out of all plausible functional forms under the specified GP prior, only those functions that are consistent with the observed data are selected. In case of no measurement error, consistent functions are functions that pass through all observed points (Figure 2.1).

2.4 Dirichlet Processes

In statistical models, typically parameteric distributions with finite number of parameters are used to model data. When the parameteric distribution has too many parameters,

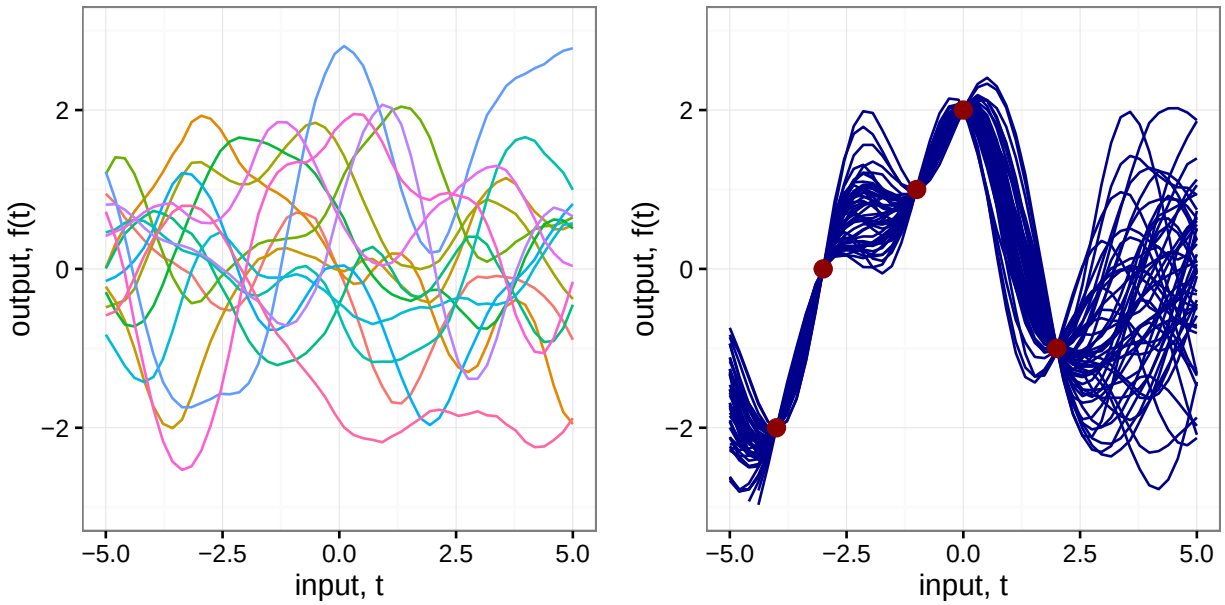


Figure 2.1: The picture to the left includes 15 randomly sampled functions from a Gaussian process with $\kappa^2 = 1$ and $\rho^2 = 0.5$. The picture to the right, includes samples from the posterior of the Gaussian process after observing data. With a noise-free measurements, posterior samples are functions that perfectly pass through all the data points.

compared to the amount of data observed, a model may suffer from over-fitting. Conversely, the model may suffer from under-fitting when there are not enough parameters to model the data. Therefore, a proper model selection technique, which is often not an easy task, is critical in parameteric models. Alternatively, Bayesian non-parameteric techniques with an unbounded number of parameters, where the posterior samples of the parameters used to model data, can avoid both under-fitting and over-fitting. The Dirichlet process (DP) is a Bayesian non-parametric technique that is often used to create flexible models that allow for a broad class of distributions. The DP is also often used in density estimation without any explicit parameteric distributional assumptions. As an example, one may consider modeling the distribution of a random variable X with observed values $x_1, x_2, \ldots, x_n$, where observed values belong to an unknown distribution F. While a typical modeling choice is to consider a parameteric family of distributions for F, however, by using the Dirichlet process priors, one can use a prior on all possible probability distributions for F (Teh (2011)).

Formally, a DP is a probability measure. In particular, a DP prior is defined as a distribution over probability measures such that marginals on a finite partition are distributed according to a Dirichlet distribution (Ferguson (1973)). G is distributed according to a DP ($G \sim DP(\alpha, G_0)$) with a base distribution G_0 and a concentration parameter α, if and only if for every finite partition of a sample space $\mathcal{X}$ that is of the form $(A_1, A_2, \ldots, A_k)$, we can write

$$\big(G(A_1), G(A_2), \ldots, G(A_n)\big) \sim Dirichlet\big(\alpha G_0(A_1), \alpha G_0(A_2), \ldots, \alpha G_0(A_k)\big).$$

Further, for any measurable subset A from the sample space $\mathcal{X}$, A and it's complement, A^c, form a finite partition of the sample space, and hence, one can write

$$\big(G(A), G(A^c)\big) \sim Dirichlet\big(\alpha G(A), \alpha G(A^c)\big),$$

where a Dirichlet distribution of this form is usually known as the Beta distribution. Using properties of the Dirichlet distribution, one can write the mean and the variance as

$$E[G(A)] = G_0(A), \tag{2.34}$$

$$Var[G(A)] = \frac{G_0(A)(1 - G_0(A))}{\alpha + 1}. \tag{2.35}$$

As the variance formula in equation (2.35) implies, a larger concentration parameter, α, indicates samples from the DP that are more concentrated around the mean, which is the base distribution G_0.

The Dirichlet process was first introduced by Ferguson (1973) as a generalization of the Dirichlet distribution to infinite-dimensional space, where he proposed a Dirichlet process construction using a normalized Gamma process. Using the Kolmogrov consistency theorem,

Ferguson (1973) proved that the Dirichlet process exists as a probability distribution on the space of all probability distributions where it follows all the probability axioms. Blackwell and MacQueen (1973) proved the existence of the Dirichlet process using de Finitti's theorem. Later, Sethuraman (1994) proved the existence of the Dirichlet process by a direct construction using his stick-breaking construction.

A Dirichlet process prior is a conjugate prior. This means, posterior of the Dirichlet process is also distributed DP (Sethuraman (1994)). Consider a random variable X and n iid sampled values of the form $x_1, x_2, \ldots, x_n$ from the sample space $\mathcal{X}$. Random variable X is distributed according to an unknown distribution G ($X|G \sim G$), where G has a Dirichlet process prior ($G \sim DP(\alpha, G_0)$). The posterior distribution of G, given the observed data, can be written as

$$G|x_1, x_2, \ldots, x_n \sim DP(\alpha + n, \frac{\alpha}{\alpha + n} G_0 + \frac{1}{\alpha + n} \sum_{i=1}^{n} \delta_{x_i}). \tag{2.36}$$

As equation (2.36) shows, the posterior distribution of the Dirichlet process is distributed DP with an updated concentration parameter and an updated base distribution. The new base distribution is a mixture of the prior base distribution, G_0, and a discrete distribution of sampling one of the existing observed x_i values, where δ_{x_i} is a selection indicator that is equal to 1 if x_i is selected.

Using the equation (2.36), one can derive the predictive distribution for a future sampled

X_{n+1}, given the observed $x_1, x_2, \ldots, x_n$ values as

$$\begin{aligned} P(X_{n+1}|x_1, x_2, \ldots, x_n) &= \int P(X_{n+1}|G, x_1, x_2, \ldots, x_n) dG \\ &= E(G|x_1, x_2, \ldots, x_n) \\ &= \frac{\alpha}{\alpha + n} G_0 + \frac{1}{\alpha + n} \sum_{i=1}^{n} \delta_{x_i}. \end{aligned} \tag{2.37}$$

The formula in equation (2.37) represents the Blackwell-MacQueen urn scheme (Blackwell and MacQueen (1973)), where a new value, X_{n+1}, is either a new sample from the base distribution, G_0, or is one of the existing samples (x_1, or x_2, ..., or x_n). Using the Blackwell-MacQueen urn scheme, one can consider realizations of the Dirichlet process as a probability distribution with a sequence of sampled values that are drawn according to equation (2.37). The sequence is as follows. The first sample is drawn directly from G_0. The second sample, is either a new draw from the base distribution G_0 with the probability $\frac{\alpha}{\alpha+1}$, or is the previously sampled X_1 with the probability $\frac{1}{\alpha+1}$. Following the same pattern in the sequence, the i^{th} sample is either a sample drawn from G_0 with the probability $\frac{\alpha}{\alpha+i-1}$, or is one of the existing samples each sampled with the probability $\frac{n_k}{\alpha+i-1}$, where n_k is the number of times that X_k has been sampled so far.

Alternatively, the marginalized DP can be also represented as the Chinese restaurant process which works as follows. The first customer can choose any dish she/he likes. The next customer, either with the probability $\frac{1}{1+\alpha}$ can choose the same dish as the first customer and can sit on the same table, or with the probability $\frac{\alpha}{1+\alpha}$ she/he can choose a new dish on a different table. The sequence continues with the same pattern where when there are n customers currently in the restaurant eating their dishes, a new customer can either choose one of the existing tables with the probability $\frac{n_i}{n+\alpha}$, or can choose a new table with the probability $\frac{\alpha}{n+\alpha}$, where n_i is the number of customers sitting at table i and n is the number of customers in the restaurant except the new cluster who just came in.

Using a direct construction, Sethuraman (1994) proposed the stick-breaking construction to prove the existence of the Dirichlet process. In order to construct the Dirichlet process G with the concentration parameter α and the base distribution G_0 of the form

$$G \sim DP(\alpha, G_0),$$

he started with a stick of the length 1. He then cut the stick to pieces of the length π_k, where $k \in \{1, 2, \dots\}$. The length π_k is calculated according to

$$\pi_k = \beta_k \prod_{i=1}^{i=k-1} (1 - \beta_i), \tag{2.38}$$

where β_k 's are iid samples from a Beta distribution of the form

$$\beta_k \overset{iid}{\sim} Beta(1, \alpha). \tag{2.39}$$

Also, he assumed X_i's are iid samples from the base distribution G_0. He then proved that G, which is distributed according to the Dirichlet process, can be constructed as

$$G = \sum_{k=1}^{\infty} \pi_k \delta_{X_k}. \tag{2.40}$$

As equation (2.40) shows, G that is distributed according to the Dirichlet process, can be constructed as an infinite mixture of point masses, and hence, G is discrete almost surely. This indicates that the realizations of the Dirichlet process, even with a continuous base distribution G_0, are discrete almost surely.

Considering the Chinese restaurant process representation of the marginalized Dirichlet process, one can easily understand the clustering aspect of the Dirichlet process by considering tables as clusters. A new customer can either choose a new dish on a new table and form a new cluster, or can join one of the existing clusters. The clustering aspect of the Dirichlet

process is also obvious from the Blackwell-MacQueen urn scheme as well as the stick-breaking construction. Using equation (2.37), the i^{th} sample, independently from other samples, can form a new cluster with the probability $P_i = \frac{\alpha}{\alpha+i-1}$. Let C be the number of clusters in a sample of size n, where $C = \sum_{i=1}^{n} C_i$, with C_i an indicator variable that is 1 if the i^{th} sample forms a new cluster. Antoniak (1974) show that the expected number of cluster in a sample of size n is

$$\begin{aligned} E(C|\text{ n samples}) &= \sum_{i=1}^{n} E(C_i) \\ &= \sum_{i=1}^{n} P_i \\ &\approx \alpha log(1+\frac{n}{\alpha}),\ \text{for } n, \alpha \gg 0, \end{aligned} \tag{2.41}$$

and the variance of the number of clusters is of the form

$$\begin{aligned} Var(C|\text{ n samples}) &= \sum_{i=1}^{n} Var(C_i) \\ &= \sum_{i=1}^{n} (P_i)(1-P_i) \\ &\approx \alpha log(1+\frac{n}{\alpha}),\ \text{for } n > \alpha \gg 0. \end{aligned} \tag{2.42}$$

Density estimation is a common application of the DP (Lo et al. (1984), Neal (1992), Escobar (1994), Escobar and West (1995a)). However, sampled distributions from the DP are discrete almost surely Sethuraman (1994). In order to estimate continuous distributions, one can convolve the Dirichlet process with a smooth distribution. This means, instead of putting a Dirichlet process prior directly on the unknown distribution of the sampled data, one can assume that the observed data are distributed according to a smooth parameteric distribution F where a parameter of that distribution is distributed according to an unknown distribution G, where G is distributed according to the DP. This new model is known as Dirichlet process

mixture (DPM) model and is of the form

$$
\begin{aligned}
&X_i|\theta_i \overset{ind}{\sim} F(.|\theta_i),\\
&\theta_i|G \overset{iid}{\sim} G,\\
&G \sim DP(\alpha, G_0).
\end{aligned}
\tag{2.43}
$$

Using the stick-breaking construction of the Dirichlet process, and by integrating out G, in the Dirichlet process mixture model in equation (2.43), one can write

$$
\begin{aligned}
F_x(.) &= \int F(.|\theta, G) dG(\theta)\\
&= \sum_{k=1}^{\infty} \pi_k F(.|\theta_k^*),
\end{aligned}
\tag{2.44}
$$

where θ_k^*, with $k \in \{1, 2, \dots\}$, are independent and identically distributed random variables according to the base distribution G_0 and π_k's are constructed according to the stick-breaking construction setup introduced earlier in equation (2.38) and equation (2.39). Based on the formulation in equation (2.44), one can see that the Dirichlet process mixture model is essentially an infinite mixture model. In application, however, since π_i's exponentially decrease, the practical number of mixing components will become finite. Unlike the finite mixture models where the number of mixing groups are fixed a priori, the number of mixing components in the Dirichlet process mixture models is inferred from the data. This is particularly useful as determining the number of mixing components in the finite mixture models is a technically difficult task (Neal (2000)). The Dirichlet process mixture models initially introduced by Ferguson (1973) and Antoniak (1974) and later become more practical by Escobar (1994), Escobar and West (1995b), MacEachern and Müller (1998), and Neal (2000).

Using the Dirichlet process mixture model introduced in equation (2.43), one can form a predictive density for a new random variable X_{n+1}, given the observed $X_1, X_2, \dots, X_n$ data

as

$$f(X_{n+1}|X_1, X_2, \ldots, X_n) = \int f(X_{n+1}|\theta_{n+1}) \tag{2.45}$$

$$\times f(\theta_{n+1}|\theta_1, \ldots, \theta_n, X_1, \ldots, X_n) \tag{2.46}$$

$$\times f(\theta_1, \ldots, \theta_n|X_1, X_2, \ldots, X_n) \tag{2.47}$$

$$\times d\theta_1 d\theta_2 \ldots d\theta_n. \tag{2.48}$$

In order to sample from the predictive density $f(X_{n+1}|X_1, X_2, \ldots, X_n)$, one should be able to first sample from the density $f(\theta_1, \ldots, \theta_n|X_1, X_2, \ldots, X_n)$ in equation (2.47), which is rather a difficult task. Next, given sampled $\theta_1, \theta_2, \ldots, \theta_n$, one should sample from the density $f(\theta_{n+1}|\theta_1, \ldots, \theta_n, X_1, \ldots, X_n)$ in equation (2.46). The density $f(\theta_{n+1}|\theta_1, \ldots, \theta_n, X_1, \ldots, X_n)$ is equivalent to the density $f(\theta_{n+1}|\theta_1, \ldots, \theta_n)$, where sampling using the density $f(\theta_{n+1}|\theta_1, \ldots, \theta_n)$ is easy using the Blackwell-MacQueen urn scheme. Next, given the sampled θ_{n+1}, one should sample from the density $f(X_{n+1}|\theta_{n+1})$, which is easy as this density is a parameteric density.

Sampling from the density $f(\theta_1, \ldots, \theta_n|X_1, \ldots, X_n)$ (equation (2.47)), when G_0 is not a conjugate prior for the likelihood $F(.|\theta)$, is rather difficult. When G_0 is a conjugate prior for F, Escobar and West (1995a) proposed an MCMC algorithm with the state of the Markov chain as $(\theta_1, \theta_2, \ldots, \theta_n)$, where at each iteration of the MCMC, they proposed a sequential draw for $i \in \{1, 2, \ldots, n\}$ from $\theta_i|\theta_{(-i)}, X_i$, where $\theta_{(-i)}$ represents all θ values except the i^{th} one. Escobar and West (1995a) showed that $\theta_i|\theta_{(-i)}, X_i$ is distributed as

$$\theta_i|\theta_{(-i)}, X_i \sim r_i H_i + \sum_{j \neq i} q_{i,j}\delta(\theta_j), \tag{2.49}$$

where, H_i is the posterior distribution for θ_i with G_0 as the prior distribution and a likelihood that is based on the single observation X_i using the distribution function F. Other elements

in the equation 2.49 are defined as

$$q_{i,j} = bF(x_i, \theta_i),$$
$$r_i = b\alpha \int F(X_i, \theta) dG_0(\theta),$$

where b is such that $\sum_{j \neq i} q_{i,j} + r_i = 1$.

Escobar (1994) and Escobar and West (1995a) made Dirichlet process mixture models applicable by providing a posterior sampling algorithm. Their algorithm, however, only works when the base distribution, G_0, is a conjugate prior. Further, their algorithm is rather inefficient in sampling from the posterior when posterior samples suffer from the sticky-cluster problem, where a parameter value gets stuck in a cluster. Neal (2000) proposed a new sampling algorithm using the Metropolis-Hasting algorithm combined with partial Gibbs sampling that is capable of sampling from the posterior of the Dirichlet process mixture model, even when G_0 is not a conjugate prior. For more efficient sampling, Neal (2000) proposed indexing sample values so when a new value is sampled for a parameter, all observations with an index to that recently-updated parameter value are simultaneously updated. Neal (2000) presented 8 algorithms to sample from the posterior of the Dirichlet process mixture model. His algorithm 8 is his most efficient algorithm to handle non-conjugate priors. We shall explain this algorithm in more detail as follows.

Neal (2000) proposed an efficient algorithm to sample from the posterior of the Dirichlet process mixture model (equation (2.43)) with a non-conjugate base distribution, G_0. In particular, Neal (2000) defined the state of the Markov chain to include θ^* and C, where θ^* is a set of unique parameter values from $(\theta_1, \theta_2, \ldots, \theta_n)$, and C is a set of indices of the form $C = \{c_i, i \in \{1, 2, \ldots, n\}\}$, where $\theta_i = \theta^*_{c_i}$. In every iteration of the MCMC, a new set of auxiliary θ^* parameters of the form $\theta^*_j, j \in \{1, \ldots, M\}$, are introduced, where M is a pre-specified number of auxiliary parameters. When c_i's are updated, along with the unique

θ^* values, the introduced auxiliary parameters are also likely to be chosen. Neal (2000) outlines the details of his algorithm 8 as follows,

---

Neal's Algorithm 8 - Sampling from Posterior of DPM

State of the Markov-Chain: $C = (c_1, c_2, \ldots, c_n)$, and $\theta^* = (\theta_1^*, \theta_2^*, \ldots, \theta_k^*)$
for $i \in \{1, \ldots, n\}$ **do**
  $k^- \leftarrow$ number of distinct $c_j, j \neq i$
  $h \leftarrow k^- + m$, where m is the number of auxiliary parameters
  if $c_i = c_j, j \neq i$ **then**
    Draw m independent samples from G_0 as auxiliary parameters
  else
    Draw $m-1$ independent samples from G_0 and use $\theta^*_{c_i}$ as the last auxiliary parameter
  end if
  Sample a value for c_i from $\{1, 2, \ldots, h\}$ with a probability as follows where $n_{-i,c}$ is the number of $c_j = c, j \neq i$, and b is a normalizing constant:
  if $1 \geq c \leq k^-$ **then**
    $P(c_i = c | c_{(-i)}, X_i, \theta_1^*, \theta_2^*, \ldots, \theta_n^*) = b \frac{n_{-i,c}}{n-1+\alpha} F(X_i, \theta_c^*)$
  else
    $P(c_i = c | c_{(-i)}, X_i, \theta_1^*, \theta_2^*, \ldots, \theta_n^*) = b \frac{\alpha/m}{n-1+\alpha} F(X_i, \theta_c^*)$
  end if
  Throw away any θ^* that is not associated with any $c_k, k \in \{1, 2, \ldots, n\}$
end for
Perform remixing: An additional one-step posterior sampling of each θ_i^* with G_0 prior and a likelihood with all subjects who share the same θ_i^* value.

The concentration parameter α inside a Dirichlet process can be treated as a hyper-parameter that also gets sampled during the MCMC process. One can choose a prior on α and learn this parameter from data. This is particularly useful as α is directly related to the expected number of clusters, and is better learned from the data to avoid any over-fitting or under-fitting. Any prior on α, induces a prior on number of clusters, k. Consider a Dirichlet process setting of the form

$$\begin{aligned} \theta | G &\sim G, \\ G &\sim DP(\alpha, G_0), \\ \alpha &\sim P(\alpha). \end{aligned}$$

Escobar and West (1995b) proposed an algorithm to sample from the posterior of α, when α has a Gamma prior of the form $Gamma(a, b)$, where a is the shape parameter and b is the scale parameter . For any α value and at each iteration of the MCMC, the number of unique θ values represents the number of clusters and is denoted by k. Given a value for α and a value for k, a new value for α can be drawn as follows

---

Posterior Sampling of α - Escobar and West Algorithm

Sample a random value for η where $\eta \sim Beta(\alpha + 1, n)$

$\pi_\eta \leftarrow \frac{\alpha+k-1}{(\alpha+k-1)+(n\times(b-log(\eta)))}$

sample a new alpha from a mixture of two gamma distributions:

$$\alpha|\eta, k \sim \pi_\eta \Gamma(a + k, b - log(\eta)) + (1 - \pi_\eta)\Gamma(a + k - 1, b - log(\eta)).$$

2.5 Joint Longitudinal-Survival Models

Jointly modeling longitudinal and survival data has received a great deal of attention in the literature. These models were first motivated by clinical studies in AIDS where researchers were interested in testing the association between longitudinally measured biomarkers, including CD4 counts, and survival outcomes.

By jointly modeling longitudinal and survival data, one can account for the relationship between the survival event times and the implicit censoring in the longitudinal measurements (Hanson et al. (2011a)). Joint models are also capable of accounting for measurement error in the longitudinal covariates. Moreover, these models correctly treat longitudinal covariates as random variables. Finally, by using both longitudinal and survival data at once, joint models lead to a higher statistical efficiency (Hogan and Laird (1997), Tsiatis and Davidian (2004), Ibrahim et al. (2005), Hanson et al. (2011a)) as opposed to separate longitudinal and survival models in which data are split into survival and longitudinal data.

There are two main approaches to jointly modeling longitudinal and survival data. One approach is to jointly model survival and longitudinal data with a focus on the survival data as the primary outcome and longitudinal measures as covariates inside the survival model. In this approach, the joint longitudinal-survival likelihood $L_{Y,X}$ is of the form

$$L_{Y,X} = L_X L_{Y|X},$$

where X denotes the longitudinal measurements, Y denotes the survival outcome, L_X is the marginal longitudinal likelihood, and $L_{Y|X}$ is the conditional likelihood of the survival data conditioned upon the longitudinal data. In the second approach, one can jointly model the longitudinal-survival data by focusing on the longitudinal data as the primary outcome and the survival data as some information on time to loss of followup in the longitudinal measures (Ibrahim et al. (2005)). The joint likelihood in the second approach is of the form

$$L_{Y,X} = L_Y L_{X|Y},$$

where L_Y is the marginal survival likelihood, and $L_{X|Y}$ is the conditional longitudinal likelihood that is conditioned on the survival data. Several authors including Schluchter (1992), De Gruttola and Tu (1994), and Schluchter et al. (2001) used the second approach to generalize random effects models in the presence of informative censoring. Our primary interest in this thesis, however, is in the first approach where survival data are the primary outcome.

Authors have tackled joint longitudinal-survival models from both the frequentist standpoint as well as the Bayesian standpoint. Some examples of frequentist joint models are found in De Gruttola and Tu (1994), Wulfsohn and Tsiatis (1997b), Tsiatis and Davidian (2001) and Law et al. (2002b). Other authors including Faucett and Thomas (1996), Wang and Taylor (2001b), R Brown and G Ibrahim (2003), and Hanson et al. (2011a) approached modeling joint longitudinal-survival data from the Bayesian standpoint. Hogan and Laird (1997),

Tsiatis and Davidian (2004), and Ibrahim et al. (2005) have provided a complete review on the existing joint longitudinal-survival models.

2.5.1 Frequentist Joint Longitudinal-Survival Models

De Gruttola and Tu (1994) proposed a frequentist joint longitudinal-survival model where they assume ad Normally distributed random effects. Their likelihood specification is based on the assumption that given the random effects, the longitudinal and survival outcomes are independent. Maximum likelihood estimation is obtained by using the EM algorithm. In their model, they assumed that the monotonic transformations of the survival times are Normally distributed, which begs the question that whether the model is still robust under the non-Normal data.

Wulfsohn and Tsiatis (1997b) proposed a joint longitudinal-survival model where they assumed a proportional hazards model for the survival data conditional on the longitudinal measures. They used a random effects model for the trajectory of the longitudinal measures. Maximum likelihood estimation was then done by using the EM algorithm. In general, their parameteric assumptions of the random effects in their longitudinal model may not hold in real data setting.

Tsiatis and Davidian (2001) proposed a joint longitudinal-survival model where they used a proportional hazards model for the survival data conditional on the longitudinal measures. For longitudinal covariates, they assumed a subject-specific linear model. They then proposed an estimator to estimate the effect of the longitudinal covariates on the survival outcome, where the estimator does not depend on the Normality assumption of the random effects in the longitudinal model. While their proposed model is not limited to Normally distributed random effects, however, large sample property of their estimator is still an open problem.

Law et al. (2002b) proposed a joint longitudinal-survival cure model where they assumed a fraction of patients are immune to the disease throughout the followup. They used a time-dependent Cox proportional hazards model with longitudinal measures as covariates. They modeled longitudinal covariates using a hierarchical non-linear mixed effects model. Parameter estimations was done using a Monte Carlo EM algorithm. One limitation of their proposed method is that the common distributional assumptions of the mixed effects model need not be true in all settings.

2.5.2 Bayesian Joint Longitudinal-Survival Models

Compared to the frequentist approach, the Bayesian approach in modeling joint longitudinal-survival data may be more advantageous as parameter estimation is typically easier using Bayesian computation as opposed to the EM algorithm. In addition, estimates need no asymptotic approximations and prior knowledge of the domain of interest can be easily incorporated into the analysis using a Bayesian framework Ibrahim et al. (2005). On the other hand, under the Bayesian approach, the hazard function, as the key component of the likelihood, needs to be fully specified. A typical approach is to discretize time and provide piece-wise approximations.

Faucett and Thomas (1996) approached modeling longitudinal-survival data using the same model as the one Wulfsohn and Tsiatis (1997b) had proposed, except from the Bayesian perspective, where they proposed a joint model with a proportional hazards survival and a random effects longitudinal model. They used rejection sampling to sample from the posterior density. In order to achieve the same results as the frequentist approach proposed by Wulfsohn and Tsiatis (1997b), Faucett and Thomas (1996) used flat priors.

Wang and Taylor (2001b) took a different approach where they used the same proportional hazards model as Faucett and Thomas (1996), however, they made the longitudinal

component of the model more flexible by adding an integrated Ornstein-Uhlenbeck (IOU) stochastic process to the trajectory function. This stochastic process adds more flexibility to the longitudinal component of the model. Brownian motion and random effects models are two special cases of their proposed longitudinal model. While IOU process adds flexibility to the model, it however, introduces a great number of parameters making the model computationally complex.

Brown and Ibrahim (2003) proposed a Bayesian joint longitudinal-survival model where they relaxed any distributional assumption on the longitudinal component by using the Dirichlet Process prior on parameters of the longitudinal model. They used a proportional hazards model for the survival component conditional on the longitudinal measurements. Their model is particularly useful where patient's response is diverse and can not be easily modeled with known parameteric distributions. Cancer vaccine trials is one such example.

In proposing joint longitudinal-survival models, while most authors had focused on proposing a more flexible longitudinal component, Hanson et al. (2011a) focused more on proposing a more flexible survival component. By using mixture of finite Polya trees, they proposed a flexible proportional hazards survival model and a flexible accelerated failure time model.

In order to add flexibility to the longitudinal component of the joint models, while most authors have focused on relaxing distributional assumptions on the longitudinal parameters, it's also desirable to relax functional assumptions on the modeled longitudinal trajectories. Further, while most authors have focused on modeling the the first moment of the longitudinal trajectories, a flexible longitudinal component capable of modeling the second moment of the longitudinal trajectories could be of scientific interest.

On the survival component of the joint models, while other authors have proposed semi-parameteric and non-parametric survival components with no explicit distributional assumption on survival times, however, we are seeking a method that is also capable of clustering

subjects who share the same baseline hazard trend in order to gain more efficiency in modeling subject-specific baseline hazards.

Finally, we are interested in a joint modeling framework that is capable of modeling multiple longitudinal biomarkers simultaneously and relating these biomarkers to survival outcomes jointly. When there exist multiple longitudinal biomarkers, simultaneously modeling these biomarkers by taking the correlation between these longitudinal measures into account can lead to a more statistical efficiency in estimating trajectories of these biomarkers as opposed to modeling each longitudinal biomarker separately and independently from other biomarkers. In particular, when there exists biomarkers that are measured less frequently compared to other biomarkers, simultaneously modeling longitudinal biomarkers can help estimate the trajectory of those less-frequent biomarkers with higher precision.

In the next chapter, using Dirichlet process mixture models, we propose logistic and survival models that are capable of detecting sub-population latent effects. While marginalizing over the unknown sub-population latent effects can cause estimates to shrink, we will show that our proposed models, by detecting sub-population latent effects, are capable of estimating true conditional estimands.

Chapter 3

Non-Collapsibility in Dirichlet Process Mixture Models

In this chapter, we consider coefficient estimation under the Dirichlet process mixture models. We introduce marginal and conditional covariate effects in a typical regression model. We introduce the non-collapsibility concept and propose a Dirichlet process mixture model to estimate conditional covariate effects in non-collapsible models. Section 3.1 provides an introduction to collapsible and non-collapsible models. Section 3.2 introduces our proposed Dirichlet process mixture models. In Section 3.3, using synthetic data, we evaluate the performance of our proposed Dirichlet process mixture models in estimating conditional coefficients in non-collapsible models. In Section 3.4, we provide sensitivity analysis on parameters that may change the performance of our proposed Dirichlet process mixture models. In Section 3.5, we apply our proposed models to real data on access failure in hemodialysis patients. Finally, we conclude with a discussion in Section 3.6.

3.1 Introduction

Statistically, non-collapsibility represents the setting where the marginal measure of association between two random variables X and Y, differs from the conditional measure of association between these two random variables, after conditioning upon the levels of a third random variable Z, where Z is not a confounder, that is, Z is associated with one random variable but not the other (Greenland et al. (1999)). In this situation, a careful attention is required to properly interpret a conditional association as opposed to a marginal association. Further, one should note that in the absence of confounding, both the marginal association and the conditional association, despite being different, are unbiased. Hence, a clear distinction between confounding and non-collapsibility is required.

Similarly, non-collapsibility exists in a regression setting when the marginal association between a predictor variable, X, and a response variable, Y, differs from the conditional association in a separate regression model where a third variable Z is adjusted in the model, where Z is not a confounder, that is, Z is only associated with the response variable.

In general, one needs to consider the relative importance of estimating the marginal association between the two random variables X and Y, as opposed to the conditional association that conditioned upon a third random variable Z. When Z is observed, it is possible to heuristically compare the difference between the marginal and the conditional associations by simply comparing the adjusted and unadjusted estimated associations. However, when Z is latent, analysts generally default to estimating a marginal association without thought to the relative merits of the two estimands.

In longitudinal studies, non-collapsibility has garnered some attention when comparing the estimates from the generalized linear mixed model with the estimates from the generalized estimating equation model, where the former provides conditional estimates that are conditioned upon the subject-specific random effects, and the latter provides estimates that

are marginalized over all subjects. Longitudinal data can be considered as a special case of the repeated measure data with measurements indexed by time. We shall use the words "longitudinal data" and "repeated measure data" interchangeably.

As a simple case, one may consider n subjects, each with l_i within subject measurements with Y_{ij} and t_{ij} as the outcome and the covariate for the j^{th} measurement on the i^{th} subject, respectively. One can write a generalized linear mixed effect model with random intercepts of the form

$$E[Y_{ij}|t_{ij}, \beta_{0i}] = \mu_{ij},$$

where the mean μ_{ij} and the covariate t_{ij} and the subject-specific random intercept β_{0i} are linked using a link function $g(.)$, where

$$g(\mu_{ij}) = \beta_{0i} + \beta_0 + \beta_1 t_{ij}. \tag{3.1}$$

In this model, β_0 and β_1 are intercept and slope that are shared across all subjects. In a typical mixed effects model, β_{0i}, where $i \in \{1, \ldots, n\}$, are assumed to be independent and Normally distributed. Under this model setting, conditioned upon the subject-specific random intercepts, β_{0i}, β_1 represents the conditional association between the random variable t and the outcome, Y.

Alternatively, one may consider a model of the form

$$E[Y_{ij}|t_{ij}] = \eta_{ij},$$

where the mean η_{ij} is related to the covariate $t_i j$ through a link function $g(.)$, where

$$g(\eta_{ij}) = \gamma_0 + \gamma_1 t_{ij}. \tag{3.2}$$

In this model, γ_0 is the intercept and γ_1 is the slope where both are shared across all subjects. Under this model setting, γ_1 represents the marginal association between the covariate t and the outcome, Y.

Generally, even with random intercepts with no confounding effect, the conditional covariate effect β_1 (equation (3.1)) and the marginal covariate effect γ_1 (equation (3.2)) need not be equal. Several authors including Gail et al. (1984), Gail (1986) showed that with non-confounding subject-specific random intercept, β_{0i}, β_1 is guaranteed to be collapsible, if $g(.)$ is either the identity link or the log link. That means with the identity or the log link and in the absence of confounding, equality of the conditional covariate effect β_1 and the marginal covariate effect γ_1 is guaranteed. Hence, we are primarily interested in studying non-collapsibility in logistic and proportional hazards models.

To show the non-collapsibility effect in the logistic regression model, we generated synthetic data, where we considered three different groups with different intercepts of $\beta_{01} = -2$, $\beta_{02} = 0$, $\beta_{03} = 2$. Independently of the intercepts, we generated covariate X, where X is simulated from the standard Normal $N(\mu = 0, \sigma = 1)$. Using the a logistic link and with a true coefficient values of $\beta_1 = 2$, we generated binary outcomes. We then fit a conditional model of the form

$$logit[E(Y_{ij}|X_{ij}, \beta_{0i})] = \beta_{0i} + \beta_0 + \beta_1 X_{ij},$$

where Y_{ij} is a binary outcome for the j^{th} measurement on the i^{th} cluster, X_{ij} is the covariate value corresponding to the outcome Y_{ij}, and β_{0i} is the true value of the cluster-specific intercept that is directly adjusted in the model. We also fit a marginal model of the form

$$logit[E(Y_{ij}|X_{ij})] = \gamma_0 + \gamma_1 X_{ij}.$$

After fitting the conditional and the marginal models above, we plot the results, where the

x-axis is the covariate values and the y-axis is the predicted probability of $Y = 1$. In this plot, the red curve shows the predicted values from the marginal model and the three black curves show the predicted values from the conditional each corresponding to a sub-group. As Figure 3.1 shows, the marginal slope that is averaged across sub-groups (γ_1) is smaller than the stratum-specific slope (β_1). This plots clearly shows non-collapsibility in logit link.

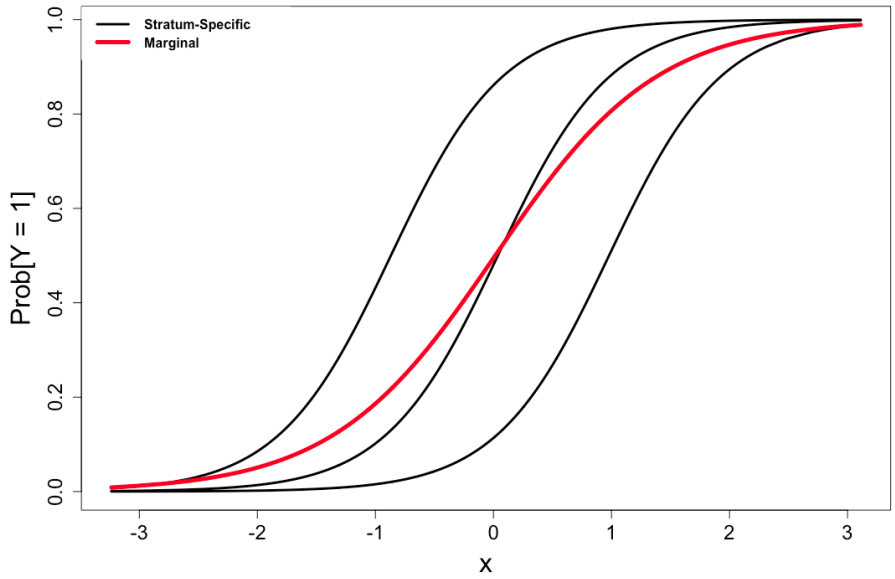


Figure 3.1: Graphical representation of non-collapsibility in logistic regression using synthetic data. Synthetic binary data were generated with three sub-groups with different intercept of $\beta_{01} = -2$, $\beta_{02} = 0$, $\beta_{03} = 2$. Independently of the intercepts, covariate X was simulated from the standard Normal $N(\mu = 0, \sigma = 1)$. This figure shows that the marginal slope (in red) is smaller than the stratum-specific slope (in black).

As Figure 3.1 shows, the marginal coefficient estimand γ_1 is shrunk towards the null hypothesis of no covariate effect compared to the conditional coefficient estimand β_1. When random intercepts are latent, even under the conditional generalized linear mixed effects model (equation (3.1)), the coefficient estimate $\hat{\beta}_1$ may shrink towards 0 compared to the true conditional estimand and that is when the distribution of the random intercepts are mis-specified. One such example is a random intercept model with true random intercepts distributed according to a bi-modal distribution. In this situation, coefficient estimates under a model that assumes random intercepts are distributed Gaussian, may still attenuate

towards 0 compared to the true conditional coefficients.

Similar to the logistic regression models, proportional hazards models are also non-collapsible. Let T_{ij} denote the j^{th} survival time for the i^{th} cluster. One example of such repeated measure survival data is the survival data on access failure among hemodialysis patients where each patient may have multiple access failures. Let X_{ij} be the covariate corresponding to the T_{ij} survival outcome. One can write a multiplicative hazard function of the form

$$h(T_{ij}|X_{ij}, \beta_{0i}) = h_0(T_{ij})exp\{\beta_{0i} + \beta_1 X_{ij}\},$$

where $h(T_{ij}|X_{ij}, \beta_{0i})$ is the hazard at time T_{ij}, $h_0(T_{ij})$ is the baseline hazard at time T_{ij}, $exp\{\beta_{0i}\}$ is the frailty term including latent cluster-specific baseline hazard multipliers, and β_1 is the log relative risk of the effect of the covariate X_{ij} on the risk of "death". Under this model setting, β_1 is the conditional relative risk of the covariate X_{ij} that is conditioned on the frailty them $exp\{\beta_{0i}\}$.

Alternatively, a marginal proportional hazards model is of the form

$$h(T_{ij}|X_{ij}) = h_0(T_{ij})exp\{\gamma_1 X_{ij}\},$$

where $h(T_{ij}|X_{ij})$ and $h_0(T_{ij})$ are hazard and baseline hazard at time T_{ij}, respectively. γ_1 represents the marginal log relative risk of the effect of every one unit changes in the covariate X_{ij} on the risk of "death". As we shall show with synthetic data, proportional hazards models are non-collapsible where the marginal parameter γ_1 shrinks toward 0 compared to the conditional parameter β_1.

A method that can uncover the underlying latent subgroups, can provide the merits of estimating the conditional associations that is conditioned upon population subgroups. The Dirichlet process and the Dirichlet process mixture models are becoming popular as these

non-parametric models impose minimal distributional assumptions. As evidenced by the stick-breaking representation of the Dirichlet process (Sethuraman (1994)), these models cluster units of analysis into sub-clusters based on the distributional similarities of those units. In this situation, a critical question is when population subgroups are latent, are DP and DPM models capable to uncover the conditional associations?

In this chapter, we explore non-collapsibility in longitudinal data when there exists latent subject-specific random intercepts. For non-collapsible logistic regression and proportional hazards models, we propose Dirichlet process mixture models that are capable of detecting latent random intercepts. Using simulation studies, we compare our proposed models with the common statistical models to analyze longitudinal data. Finally, we use our proposed models to analyze data on hemodialysis patients in order to find risk factors associated with access failure among these patients.

3.2 Methodology

With the focus on logistic regression and the proportional hazards models, and in the context of modeling correlated longitudinal data where repeated measures on sampling units are collected over time, we propose Dirichlet process mixture models capable of estimating conditional covariate effects when there exists latent sub-population effects. In Section 3.2.1, we introduce our proposed Bayesian logistic model and in Section 3.2.2 we introduce our proposed Bayesian proportional hazards model.

3.2.1 A Bayesian Hierarchical Logistic Regression with Dirichlet Process Mixture Priors

The logistic link is non-collapsible. This means, when there exists latent population subgroup effects in the form of random intercepts, failure to adjust for these subgroup effects leads to coefficient estimates that are shrunk toward 0 compared to the true conditional estimands from a separate model with those latent random intercepts taken into account. Generalized linear mixed effects models are capable of modeling random intercepts where they typically assume random intercepts to be distributed according to a Gaussian distribution, however, distributional mis-specification of the random intercepts may still cause coefficient estimates to shrink. A model capable of detecting subgroup random intercepts, that is also robust to distributional mis-specification of random intercepts, can provide the merits of estimating the conditional coefficient estimates.

We propose a hierarchical Bayesian model that is capable of detecting latent subgroup effects that are in the form of latent random intercepts. The models is capable of estimating conditional parameters. Using a Dirichlet process mixture prior, our proposed model is robust to distributional mis-specification of the random intercepts. In our proposed model, we consider the binary data Y_{ij} to be distributed according to

$$Y_{ij}|\beta_{0i}, \beta_0, \beta_1, X_{ij} \sim Bernoulli(p_i = \beta_{0i} + \beta_0 + \beta_1 X_{ij}),$$

where $i \in \{1, \ldots, n\}$ and $j \in \{1, \ldots, l_i\}$ with n as the number of subjects and l_i as the number of measurements on the i^{th} subject, X_{ij} is the corresponding covariate to the outcome Y_{ij}, β_{0i} is the subject-specific intercept for subject i, and β_0 and β_1 are the intercept and the slope that are shared across all subjects, respectively. We consider Gaussian priors on the

shared intercept β_0 and the shared slope β_1 of the form

$$\beta_0 \sim N(0, \sigma_{\beta_0}),$$
$$\beta_1 \sim N(0, \sigma_{\beta_1}).$$

We propose using the Dirichlet process mixture prior on the random intercepts β_{0i}, where $i \in \{1, \ldots, n\}$ with n as the number of subjects in the data. Using the Dirichlet process mixture prior, as opposed to an explicit distributional assumption, will make the model robust to distributional mis-specification. Further, DPM prior will allow subjects to cluster based on the distributional similarities of their latent random intercepts, hence, provides higher precision in estimating those latent subject effects. We specify a Dirichlet process mixture prior on β_{0i} as

$$\begin{aligned} \beta_{0i} &\sim N(\mu_i, \sigma_{\beta_{0i}}), \\ \mu_i | G &\sim G, \\ G &\sim DP(\alpha, G_0 = N(0, \sigma_0)). \end{aligned} \tag{3.3}$$

The Dirichlet process mixture prior above induces a prior on β_{0i} that is essentially an infinite mixture of Normal distributions that are mixed over the mean parameter. We shall refer to this model as a Mean-DPM model. Alternatively, one may set a Dirichlet process prior that induces an infinite Gaussian mixture prior that are mixed over the standard deviation parameter. Such a prior can be specified as

$$\begin{aligned} \beta_{0i} &\sim N(0, \sigma^{(i)}_{\beta_0^{(i)}}), \\ \sigma^{(i)}_{\beta_0^{(i)}} | G &\sim G, \\ G &\sim DP(\alpha, G_0 = log - Normal(\mu_{G_0}, \sigma_{G_0})). \end{aligned} \tag{3.4}$$

We shall refer to this model as Sigma-DPM model.

3.2.2 A Bayesian Hierarchical Proportional Hazards Model with Dirichlet Process Mixture Priors

Similar to the logistic regression models, proportional hazards models are also non-collapsible. When there exists differential subject-specific baseline hazard risk, even in the absence of confounding in the baseline hazard risks, failure to adjust for these subject-specific baseline risks in a proportional hazards model leads to coefficient estimates that are shrunk toward 0 compared to the true conditional estimands from a separate model with those latent baseline risks taken into account. In this situation, a proportional hazards model that is capable of detecting subject-specific baseline hazards, can provide the merits of estimating the conditional coefficient estimates.

We propose a hierarchical Bayesian proportional hazards model that is capable of detecting the differential subject-specific baseline hazard risk across subjects. Our proposed model uses a Dirichlet process mixture prior on the latent subject-specific baseline hazards. The Dirichlet process mixture prior allows clustering subjects based on the distributional similarities of their baseline hazards. Further, by using the Dirichlet process mixture prior, we avoid any explicit distributional assumption on the latent subject-specific baseline hazards. In our proposed model, we consider survival times T_{ij}, where $i \in \{1, \ldots, n\}$ and $j \in \{1, \ldots, l_i\}$ with n as the number of subjects and l_i as the number of measurements on the i^{th} subject, to be distributed according to a Weibull distribution of the form

$$T_{ij}|\tau, \beta_{0i}, \beta_0, \beta_1, X_{ij} \sim Weibull(\tau, \theta_i),$$
$$log(\theta_i) = \beta_{0i} + \beta_0 + \beta_1 X_{ij},$$

where X_{ij} is the covariate value corresponding to the T_{ij} survival time, $exp\{\beta_{0i}\}$ is a subject-specific baseline hazard, β_0 is a shared intercept across all subjects, β_1 is a shared slope across all subjects that represents the log relative risk of every one unit increase in the covariate X_{ij}, τ is the shape parameter, and θ_i is a subject-specific scale parameter. In the model specification above, we introduced covariates into the model through the scale parameter and using the equation $log(\theta_i) = \beta_{0i} + \beta_0 + \beta_1 X_{ij}$. For our proposed model, we consider Gaussian priors on β_0 and β_1 parameters as

$$\beta_0 \sim N(0, \sigma_{\beta_0}),$$
$$\beta_1 \sim N(0, \sigma_{\beta_1}),$$

where σ_{β_0} and σ_{β_1} are fixed numbers. We also assume a log-Normal prior on the shape parameter, τ, as

$$\tau \sim log - Normal(\mu_\tau, \sigma_\tau),$$

with μ_τ and σ_τ as fixed numbers.

We use a Dirichlet process mixture prior for the subject-specific β_{0i} parameters as

$$\begin{aligned} \beta_{0i} &\sim N(\mu_i, \sigma_{\beta_{0i}}) \\ \mu_i | G &\sim G, \\ G &\sim DP(\alpha, G_0 = N(0, \sigma_0)). \end{aligned} \tag{3.5}$$

The Dirichlet process mixture prior above is essentially an infinite mixture of Normal distributions that are mixed over the mean parameter. We shall refer to this model with the Mean-DPM proportional hazards model. Alternatively, we propose a Dirichlet process mixture model that induces an infinite Normal distributions mixed overt the standard deviation

parameter. This Dirichlet process prior can be written as

$$\begin{aligned} \beta_{0i} &\sim N(0, \sigma^{(i)}_{\beta_{0i}}), \\ \sigma^{(i)}_{\beta_{0i}} | G &\sim G, \\ G &\sim DP(\alpha, G_0 = log - Normal(\mu_{G_0}, \sigma_{G_0})). \end{aligned} \tag{3.6}$$

We shall refer to this new model with the Dirichlet process mixture prior above with Sigma-DPM proportional hazards model.

3.3 Simulation Studies

Using simulation studies, we investigate non-collapsibility in logistic regression and proportional hazards models. We consider three simulation scenarios: one when subject-specific intercepts are sampled independently from the standard Normal $N(\mu = 0, \sigma = 1)$, another when subject-specific intercepts are sampled from a mixture distribution of the form

$$\beta_{0i} \overset{iid}{\sim} \theta_i N(\mu = -1.5, \sigma = 1) + (1 - \theta_i) N(\mu = 1.5, \sigma = 1),$$

where $\theta_i \sim Bernoulli(p = 0.5)$ with $i \in \{1, \ldots, n\}$ where n is the number of subjects. Finally, in the third scenario subject-specific intercepts are sampled from a mixture distribution of the form

$$\beta_{0i} \overset{iid}{\sim} \theta_i N(\mu = 0, \sigma = 1) + (1 - \theta_i) N(\mu = 0, \sigma = \sqrt{5}),$$

where $\theta_i \sim Bernoulli(p = 0.5)$ for $i \in \{1, \ldots, n\}$.

We compare parameter estimation between our proposed models and some common statistical models used to analyze repeated measure binary data and survival data. For every

simulation scenario, we run 1,000 simulations each with 300 subjects and 12 within-subject measurements per subject.

3.3.1 Logistic Regression Models

Unlike linear and log links, logistic link is not collapsible. In this section, using synthetic data we compare parameter estimation under our proposed Mean-DPM and Sigma-DPM Bayesian hierarchical logistic regressions we the following common statistical models to analyze repeated measure binary data:

- Generalized linear model with a logit link (GLM): We fit a frequentist GLM model with the logit link. This technique ignores the correlation between within-subject measurements. Further, this model does not account for any subject-specific effect. Due to the ignorance of within subject correlations in this model, standard error for the estimated coefficients tend to underestimate the true standard error once the within subject correlation is taken into account.

- Generalized estimating equation (GEE): Instead of a simple generalized linear model with the logit link where all within-subject measurements are treated as independent measures, one can use the generalized estimating equation framework to account for the correlation between within-subject measurements. Despite accounting for the correlation between measurements taken on the same subject, GEE does not consider any subject-specific random effect.

- Generalized linear mixed effects model (GLMM): We also fit the frequentist generalized linear mixed effects model with subject-specific random intercepts to model binary data. GLMM is capable of taking the correlation in with-subject measurements into account. Further, GLMM is also capable of estimating subject-specific random intercepts with the assumption that the random intercepts are Normally distributed.

- Bayesian logistic regression: We also consider a Bayesian logistic regression model with a likelihood of the form

$$Y_{ij}|\beta_0, \beta_1, X_{ij} \sim Bernoulli(p_i = \beta_0 + \beta_1 X_{ij}),$$

where Y_{ij} is the outcome of the j^{th} measurement on the i^{th} subject, X_{ij} is the measured covariate corresponding to Y_{ij} outcome, and β_0 and β_1 are intercept and slope. We assume priors of the form

$$\beta_0 \sim N(0, \sigma_{\beta_0}),$$
$$\beta_1 \sim N(0, \sigma_{\beta_1}),$$

where σ_{β_0} and σ_{β_1} are fixed numbers. Detailed programming code for this model is in Section A.1.1.

- Hierarchical Bayesian logistic regression model: Analogous to the the GLMM model to analyze binary data, one can setup a Bayesian hierarchical model with a likelihood of the form

$$Y_{ij}|\beta_{0i}, \beta_0, \beta_1, X_{ij} \sim Bernoulli(p_i = \beta_0 + \beta_1 X_{ij}),$$

where Y_{ij} is the outcome of the j^{th} measurement on the i^{th} subject, X_{ij} is the measured covariate corresponding to Y_{ij} outcome, β_{0i} is the subject-specific random intercepts where $i \in \{1, \ldots, n\}$ with n as the number of subjects in the data, and β_0 and β_1 are intercept and slope. We assume Gaussian priors on subject-specific random intercepts β_{0i} of the form

$$\beta_{0i} \sim N(0, \sigma_{\beta_{0i}}),$$

where $i \in \{1, \ldots, n\}$. Also, Gaussian priors are assumed on coefficients β_0, and β_1 of the form

$$\beta_0 \sim N(0, \sigma_{\beta_0}),$$
$$\beta_1 \sim N(0, \sigma_{\beta_1}),$$

where σ_{β_0} and σ_{β_1} are fixed numbers. Detailed programming code for this model is in Section A.1.1.

Figure 3.2 shows the histogram of the posterior median of μ_i, where $i \in \{1, \ldots, n\}$ from the proposed Mean-DPM hierarchical Bayesian logistic model, where μ_i is the subject-specific prior mean on the random intercept of subject i (equation (3.3)). Under each simulation scenario, we simulated a single dataset with 300 subjects each with 12 within-subject measurements and applied our proposed Mean-DPM model. The plot to the left shows a histogram of the posterior median of μ_i when data are simulated with random intercept β_{0i} sampled from the standard Normal $N(\mu = 0, \sigma = 1)$. As the histogram shows, most the posterior medians are close to zero. The histogram in the model shows the distribution of the posterior median μ_i when data are simulated with random intercepts sampled from mixture of two Normal distributions of the form $\theta_i N(\mu_1 = -1.5, \sigma = 1) + (1 - \theta_i) N(\mu_1 = 1.5, \sigma = 1)$, where θ is distributed Bernoulli with parameter $p = 0.5$. As the histogram in the middle shows, posterior medians are bi-modal where modes are around the true values of -1.5 and 1.5. Finally, the histogram to the right shows the posterior median of μ_i when data are simulated with random intercepts sampled from mixture of two Normal distributions of the form $\theta_i N(\mu = 0, \sigma_1 = 1) + (1 - \theta_i) N(\mu = 0, \sigma_2 = \sqrt{5})$. Due to the differences in the standard deviations, one may expect the histogram to be spread more widely compare to the first scenario, nonetheless, posterior medians are still centered around the true mean of 0.

Figure 3.3 shows the histogram of the posterior median of σ_i, where $i \in \{1, \ldots, n\}$ from the

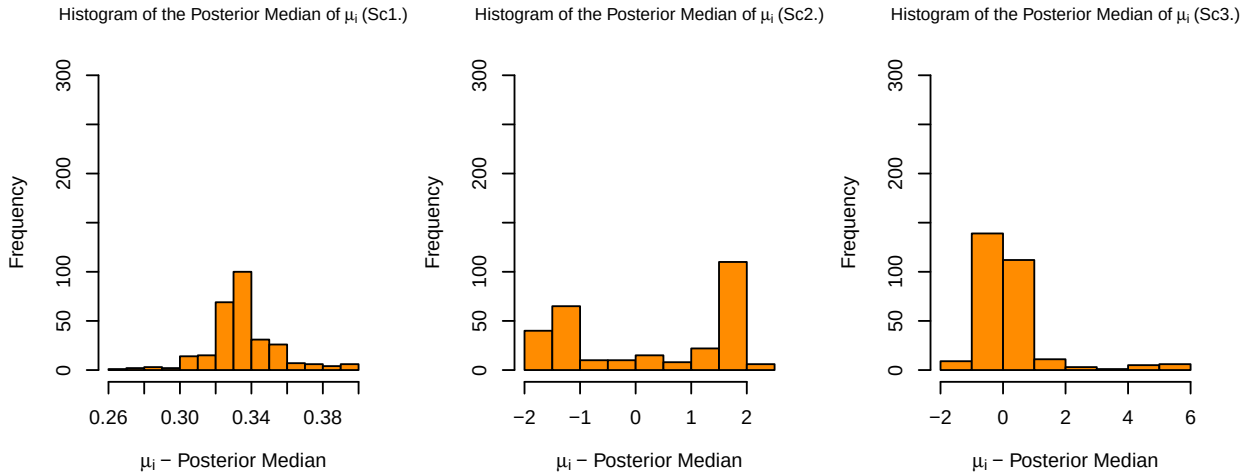


Figure 3.2: Histogram of the posterior median of μ_i's from the proposed Mean-DPM hierarchical Bayesian logistic model, where μ_i is the subject-specific prior mean on the random intercept of subject "i". The plot to the left is the histogram of the posterior median of the sampled μ_i from the model when it runs under the first simulation scenario where all random intercepts are sampled from the standard Normal distribution. The plot in the middle shows the histogram of the posterior medians of μ_i's under the second scenario where random intercepts are sampled from a mixture of two Normal distributions of $N(\mu_1 = -1.5, \sigma = 1)$ and $N(\mu_1 = 1.5, \sigma = 1)$ that are equally weighted. The plot to the right is the histogram of the posterior medians under the third simulation scenario where the random intercepts are simulated from the mixture of two Normal distributions of $N(\mu = 0, \sigma_1 = 1)$ and $N(\mu = 0, \sigma_2 = \sqrt{5})$. Results, under each simulation, are from one single simulated data with $N = 300$ subjects and $l_i = 12$ within subject measurements.

proposed Sigma-DPM hierarchical Bayesian logistic model, where σ_i is the subject-specific prior standard deviation on the random intercept of subject i (equation (3.4)). Under each simulation scenario, we simulated a single dataset with 300 subjects each with 12 within-subject measurements and applied our proposed Sigma-DPM model. The plot to the left shows a histogram of the posterior median of σ_i when data are simulated with random intercept β_{0i} sampled from the standard Normal $N(\mu = 0, \sigma = 1)$. As the histogram shows, most the posterior medians are close to 1. The histogram in the model shows the distribution of the posterior median σ_i when data are simulated with random intercepts sampled from mixture of two Normal distributions of the form $\theta_i N(\mu_1 = -1.5, \sigma = 1) + (1 - \theta_i)N(\mu_1 = 1.5, \sigma = 1)$, where θ is distributed Bernoulli with parameter $p = 0.5$. As the histogram

in the middle shows, posterior medians are uniformly distributed from 3.18 to 3.28. This results make sense as now the data is widely spread with two distinct mean with a distance of 3. Our Sigma-DPM model with prior mean 0 on random intercepts has to have a larger standard deviation to provide a prior to cover all plausible subject-specific random intercepts β_{0i}. Finally, the histogram to the right shows the posterior median of σ_i when data are simulated with random intercepts sampled from mixture of two Normal distributions of the form $\theta_i N(\mu = 0, \sigma_1 = 1) + (1 - \theta_i)N(\mu = 0, \sigma_2 = \sqrt{5})$. It seems that in this case, the model converged to a standard deviation that is close $\sigma_2 = \sqrt{5}$. This makes sense since a when a random intercept β_{0i} is plausible under the prior $N(0, \sigma_1)$, it's also plausible under a prior with larger standard deviation. Hence, posterior medians converged to a large standard deviation that is plausible according to the random intercepts sampled from $N(0, \sigma_2 = \sqrt{5})$.

While Figure 3.2 and Figure 3.3 show the performance of our proposed models in estimating prior mean and prior standard deviation of the random intercepts, β_{0i}, however, the main interest is on evaluating the performance of the model on estimating the actual random intercepts. Figure 3.4 provides a grid of scatter plots each shows the relation between the true random intercept value and the posterior median or the estimate of random intercepts. As one can see in the plot, when random intercepts are Normally distributed according to the standard Normal $N(\mu = 0, \sigma = 1)$ distribution, in terms of estimating the latent random intercepts, our proposed Mean-DPM and Sigma-DPM models work equally well as the GLMM model and the hierarchical Bayesian logistic model with explicit Normal assumption on the random intercepts. When the reference distribution of the sampled random intercepts is not Normal, our proposed Mean-DPM and Sigma-DPM models that are robust to distributional mis-specification of the random intercepts, outperform the GLMM and the hierarchical Bayesian logistic regression in terms of estimating the latent random intercepts.

As tables (3.1), (3.2), and (3.3) show, coefficient estimates under marginal Bayesian model and marginal frequentist GLM and GEE shrank toward the 0 compared to the true condi-

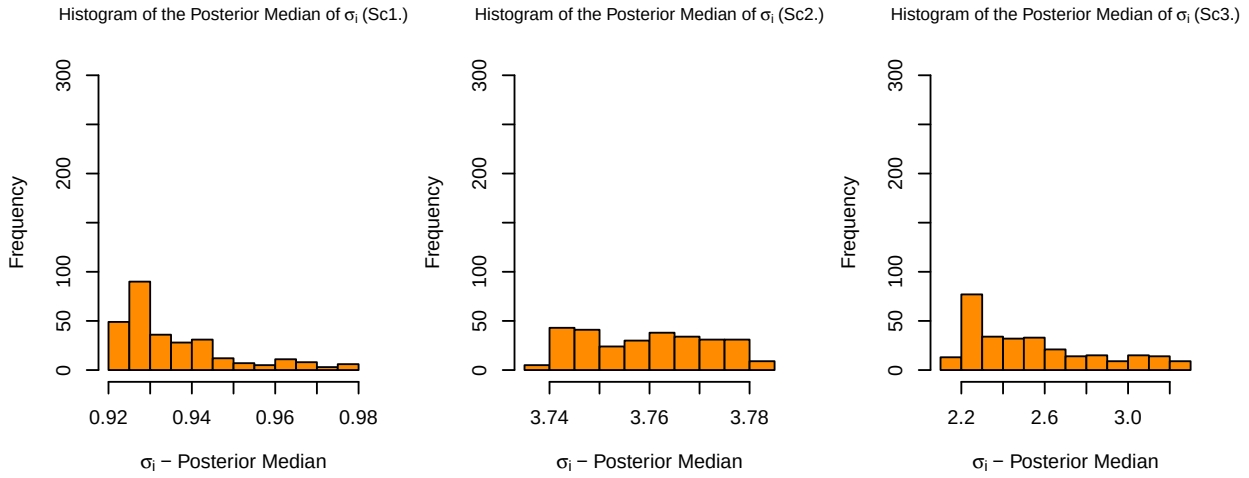


Figure 3.3: Histogram of the posterior median of σ_i's from the proposed Sigma-DPM hierarchical Bayesian logistic model, where σ_i is the subject-specific prior standard deviation on the random intercept of subject "i". The plot to the left is the histogram of the posterior median of the sampled σ_i from the model when it runs under the first simulation scenario where all random intercepts are sampled from the standard Normal distribution. The plot in the middle shows the histogram of the posterior medians of σ_i's under the second scenario where random intercepts are sampled from a mixture of two Normal distributions of $N(\mu_1 = -1.5, \sigma = 1)$ and $N(\mu_1 = 1.5, \sigma = 1)$ that are equally weighted. The plot to the right is the histogram of the posterior medians under the third simulation scenario where the random intercepts are simulated from the mixture of two Normal distributions of $N(\mu = 0, \sigma_1 = 1)$ and $N(\mu = 0, \sigma_2 = \sqrt{5})$. Results, under each simulation, are from one single simulated data with $N = 300$ subjects and $l_i = 12$ within subject measurements.

tional value. The fact that in table (3.1) coefficient estimates under both GLM and GEE are the same is not surprising as we are using balanced data with the canonical link. By taking sub-group intercepts into account, coefficient estimates from the generalized linear mixed effect model and the hierarchical Bayesian model with Normal prior on the random intercepts are closer to the true conditional estimand compared to the marginal models. However, the coefficient estimate under these models still shrink toward no 0. The amount of shrinkage is larger under the second and the third scenarios when the distribution of random intercepts is mis-specified. Our proposed Dirichlet process mixture models, however, are capable of detecting sub-group intercepts and are robust to distributional mis-specification of the random intercepts. Coefficient estimates from our proposed models lead to the minimum mean

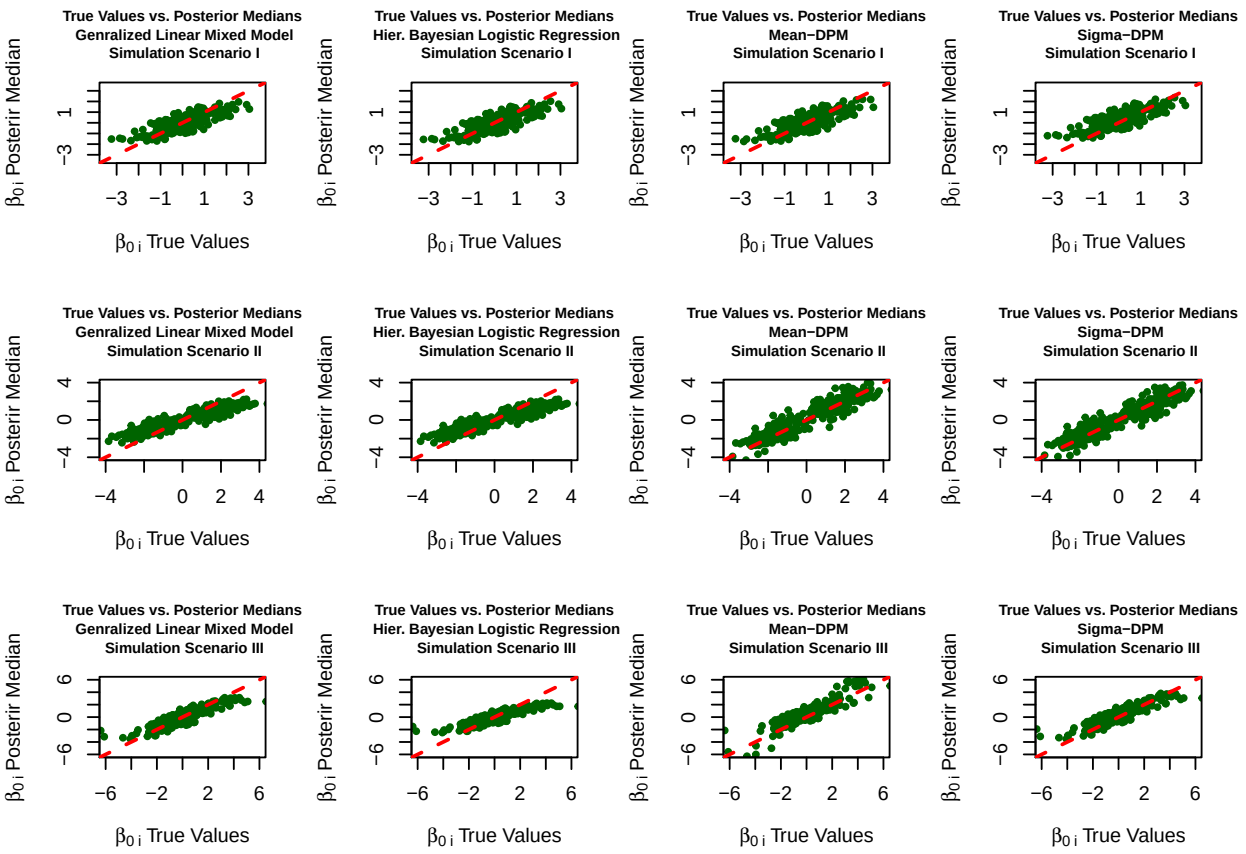


Figure 3.4: A grid of scatter plots that shows the relation between the true values of the subject-specific random intercepts, β_{0i}, and the posterior median (or estimated) random intercepts from the GLMM model, the hierarchical Bayesian logistic model, our proposed Mean-DPM hierarchical Bayesian logistic model, and the proposed Sigma-DPM hierarchical Bayesian logistic model. The red dashed line in every plot represents the 45 degree line and the results are from a single simulated data under each simulation scenario. The first row represents the scatter plots from data simulated under the first scenario where subject-specific random intercepts are sampled from the standard Normal $N(\mu = 0, \sigma = 1)$. The second row represents scatter plots resulted from data simulated under the second simulation scenario where random intercepts are sampled from an equally weighted mixture of two Normal distributions of the form $N(\mu_1 = -1.5, \sigma = 1)$ and $N(\mu_1 = 1.5, \sigma = 1)$. Finally, the last row of plots represents results from data simulated under the third simulation scenario where random intercepts are sampled from an equally weighted mixture of two Normals of the form $N(\mu = 0, \sigma_1 = 1)$ and $N(\mu_1 = 0, \sigma_2 = \sqrt{5})$.The first column of scatter plots from left represents results from fitting the generalized linear mixed effect model, the second column represents the results from a hierarchical Bayesian logistic regression, third column represents the results from fitting our proposed Mean-DPM hierarchical Bayesian logistic model, and finally the last column to the right represents results from our proposed Sigma-DPM hierarchical Bayesian logistic model.

squared error (MSE) in estimating the true conditional coefficient value [1].

	$\beta_x = 1.000$	SD	MSE
GLM	0.845	0.031	0.025
GEE	0.845	0.031	0.025
Bayesian Logistic Reg.	0.847	0.031	0.025
GLMM	0.951	0.032	0.004
Hierarchical Bayes Logistic Reg.	0.947	0.035	0.005
Mean-DPM Hierarchical Logistic Reg.	1.001	0.036	0.001
Sigma-DPM Hierarchical Logistic Reg.	1.003	0.037	0.001

Table 3.1: Binary data generated with random intercepts that are distributed according to the standard Normal distribution $N(\mu = 0, \sigma = 1)$. Results are from $1,000$ different simulated data each with $N = 300$ subjects and $l_i = 12$ within subject measurements.

	$\beta_x = 1.000$	SD	MSE
GLM	0.626	0.027	0.141
GEE	0.627	0.275	0.140
Bayesian Logistic Reg.	0.626	0.027	0.141
GLMM	0.938	0.034	0.005
Hierarchical Bayes Logistic Reg.	0.931	0.033	0.006
Mean-DPM Hierarchical Logistic Reg.	1.006	0.042	0.002
Sigma-DPM Hierarchical Logistic Reg.	0.978	0.042	0.002

Table 3.2: Binary data generated with random intercepts that are distributed according to a mixture distribution of the form $\theta_i N(\mu = -1.5, \sigma = 1) + (1 - \theta_i)N(\mu = 1.5, \sigma = 1)$, where θ_i are distributed $Bernoulli$ with parameter $p = 0.5$. Results are from $1,000$ different simulated data each with $N = 300$ subjects and $l_i = 12$ within subject measurements.

3.3.2 Proportional Hazards Survival Models

To explore non-collapsibility in proportional hazards models and to compare coefficient estimation under our proposed Mean-DPM and Sigma-DPM models with common proportional hazards models, we consider the following proportional hazards models:

- The frequentist Cox model: We fit the frequentist Cox proportional hazards model. This model assumes an overall baseline hazards for all subjects. Using the partial

[1] posterior samples traceplots for the Bayesian logistic models are provided in the Appendix B.1.1

	$\beta_x = 1.000$	SD	MSE
GLM	0.702	0.028	0.090
GEE	0.701	0.030	0.090
Bayesian Logistic Reg.	0.700	0.028	0.091
GLMM	0.946	0.033	0.004
Hierarchical Bayes Logistic Reg.	0.935	0.034	0.005
Mean-DPM Hierarchical Logistic Reg.	0.998	0.041	0.001
Sigma-DPM Hierarchical Logistic Reg.	0.994	0.040	0.001

Table 3.3: Binary data generated with random intercepts that are distributed according to a mixture distribution of the form $\theta_i N(\mu = 0, \sigma = 1) + (1 - \theta_i) N(\mu = 0, \sigma = \sqrt{5})$, where θ_i are distributed *Bernoulli* with parameter $p = 0.5$. Results are from $1,000$ different simulated data each with $N = 300$ subjects and $l_i = 12$ within subject measurements.

likelihood techniques, Cox model does not need any baseline hazard specification as that measure gets canceled out during the estimation process. The Cox frequentist model does not take the differential baseline hazards across subjects into account. In fitting the Cox model, we take the within subject correlation between multiple within-subject measurements into account using the approach proposed by Lee et al. (1992) where we first estimate model coefficients using the independent covariance matrix and then we use a robust sandwich covariance matrix to account for within subject correlation between measurements.

- Weibull accelerated failure time model (AFT): AFT models describe survival times as a function of predictor variables. Generally, Weibull AFT models are of the form

$$log(T_{ij}) = \beta_0 + \beta_1 X_{ij} + \epsilon,$$

where T_{ij} is the survival time for the j^{th} measurement on the i^{th} subject, X_{ij} is the corresponding covariate to the outcome T_{ij}, and a random error ϵ such that $T_{ij}$is distributed according to a Weibull distribution with shape parameter τ and scale parameter $exp(\lambda)$. When there exists multiple measurements per subject, failure to account for the correlation between within subject measurements leads to incorrect estimated standard error of coefficients. In order to account for this intra class correlations, we

take the the approach proposed by Lee et al. (1992) where first coefficients in the model are estimated using an independent covariance structure between within subject measurements and then a robust sandwich covariance matrix is used to account for the within cluster correlations.

- Bayesian marginal proportional hazards model: We consider a Bayesian proportional hazard model with a likelihood of the form

$$T_{ij}|\tau, \beta_0, \beta_1, X_{ij} \sim Weibull(\tau, \lambda_i = \beta_0 + \beta_1 X_{ij}),$$

where T_{ij} and X_{ij} are the survival times and the measured covariate on the j^{th} measurement on the i^{th} subject, τ is the shape parameter, β_0 and β_1 are the intercepts and the slope with β_1 as the log relative risk of death per every one unit change in X_{ij}. Similar to the previously introduced Weibull distribution for survival times, λ_i is the log of the subject-specific scale parameter. We specify a log-Normal prior on the shape parameter τ that is of the form

$$\tau \sim log - Normal(\mu_\tau, \sigma_\tau),$$

where μ_τ and σ_τ are fixed numbers. Also, β_0 and β_1 are assumed to have Gaussian priors of the form

$$\beta_0 \sim N(0, \sigma_{\beta_0}),$$
$$\beta_1 \sim N(0, \sigma_{\beta_1}).$$

Detailed programming code for this model is in Section A.1.2.

- Hierarchical Bayesian proportional hazards model: In order to account for the differ-

ential baseline hazard across subjects, one can consider a likelihood of the form

$$T_{ij}|\tau, \beta_{0i}, \beta_0, \beta_1, X_{ij} \sim Weibull(\tau, \lambda_i = \beta_{0i} + \beta_0 + \beta_1 X_{ij}),$$

where β_{0i} can be considered as the subject-specific log baseline hazard. For this model, we assume similar to priors as the one specified for the "Bayesian marginal proportional hazards model". Additionally, we assume β_{0i}, where $i \in \{1, \ldots, n\}$, to have a Gaussian prior of the form:

$$\beta_{0i} \sim N(0, \sigma_{\beta_{0i}}),$$

where $\sigma_{\beta_{0i}}$ is a fixed number. To see the programming code for this model, you may refer to the appendix in Section A.1.2.

Figure 3.5 shows the histogram of the posterior median of μ_i, where $i \in \{1, \ldots, n\}$ from the proposed Mean-DPM hierarchical Bayesian proportional hazard model, where μ_i is the subject-specific prior mean on the subject-specific log baseline hazard of subject i, which we represent it with β_{0i} and for the sake consistency, we shall refer to it as the subject-specific random intercept (equation (3.5)). Under each simulation scenario, we simulated a single dataset with 300 subjects each with 12 within-subject measurements and applied our proposed Mean-DPM model. The plot to the left shows a histogram of the posterior median of μ_i when data are simulated with random intercept β_{0i} sampled from the standard Normal $N(\mu = 0, \sigma = 1)$. As the histogram shows, most the posterior medians are close to zero. The histogram in the model shows the distribution of the posterior median μ_i when data are simulated with random intercepts sampled from mixture of two Normal distributions of the form $\theta_i N(\mu_1 = -1.5, \sigma = 1) + (1 - \theta_i)N(\mu_1 = 1.5, \sigma = 1)$, where θ is distributed Bernoulli with parameter $p = 0.5$. As the histogram in the middle shows, posterior medians are bi-modal where modes are around the true values of -1.5 and 1.5. Finally, the histogram to

the right shows the posterior median of μ_i when data are simulated with random intercepts sampled from mixture of two Normal distributions of the form $\theta_i N(\mu = 0, \sigma_1 = 1) + (1 - \theta_i) N(\mu = 0, \sigma_2 = \sqrt{5})$. Due the the differences in the standard deviations, one may expect the histogram to be spread more widely compare to the first scenario, nonetheless, posterior medians are still centered around the true mean of 0.

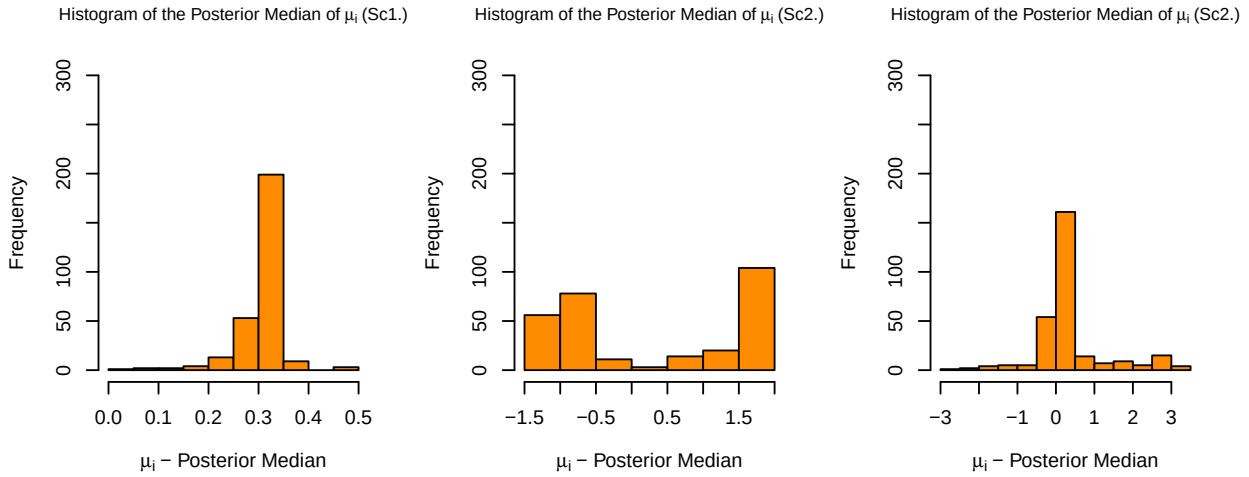


Figure 3.5: Histogram of the posterior median of μ_i's from the proposed Mean-DPM hierarchical Bayesian proportional hazard model, where μ_i is the subject-specific prior mean on the random intercept of subject i. The plot to the left is the histogram of the posterior median of the sampled μ_i from the model when it runs under the first simulation scenario where all random intercepts are sampled from the standard Normal distribution. The plot in the middle shows the histogram of the posterior medians of μ_i's under the second scenario where random intercepts are sampled from a mixture of two Normal distributions of $N(\mu_1 = -1.5, \sigma = 1)$ and $N(\mu_1 = 1.5, \sigma = 1)$ that are equally weighted. The plot to the right is the histogram of the posterior medians under the third simulation scenario where the random intercepts are simulated from the mixture of two Normal distributions of $N(\mu = 0, \sigma_1 = 1)$ and $N(\mu = 0, \sigma_2 = \sqrt{5})$. Results, under each simulation, are from one single simulated data with $N = 300$ subjects and $l_i = 12$ within subject measurements.

Figure 3.6 shows the histogram of the posterior median of σ_i, where $i \in \{1, \ldots, n\}$ from the proposed Sigma-DPM hierarchical Bayesian proportional hazard model, where σ_i is the subject-specific prior standard deviation on the random intercept of subject i (equation (3.6)). Under each simulation scenario, we simulated a single dataset with 300 subjects each with 12 within-subject measurements and applied our proposed Sigma-DPM model. The

plot to the left shows a histogram of the posterior median of σ_i when data are simulated with random intercept β_{0i} sampled from the standard Normal $N(\mu = 0, \sigma = 1)$. As the histogram shows, most the posterior medians are close to 1. The histogram in the model shows the distribution of the posterior median σ_i when data are simulated with random intercepts sampled from mixture of two Normal distributions of the form $\theta_i N(\mu_1 = -1.5, \sigma = 1) + (1 - \theta_i)N(\mu_1 = 1.5, \sigma = 1)$, where θ is distributed Bernoulli with parameter $p = 0.5$. As the histogram in the middle shows, posterior medians are uniformly distributed from 3.18 to 3.28. This results make sense as now the data is widely spread with two distinct mean with a distance of 3. Our Sigma-DPM model with prior mean 0 on random intercepts has to have a larger standard deviation to provide a prior to cover all plausible subject-specific random intercepts β_{0i}. Finally, the histogram to the right shows the posterior median of σ_i when data are simulated with random intercepts sampled from mixture of two Normal distributions of the form $\theta_i N(\mu = 0, \sigma_1 = 1) + (1 - \theta_i)N(\mu = 0, \sigma_2 = \sqrt{5})$. It seems that in this case, the model converged to a standard deviation that is close $\sigma_2 = \sqrt{5}$. This makes sense since a when a random intercept β_{0i} is plausible under the prior $N(0, \sigma_1)$, it's also plausible under a prior with larger standard deviation. Hence, posterior medians converged to a large standard deviation that is plausible according to the random intercepts sampled from $N(0, \sigma_2 = \sqrt{5})$.

Based on Figure 3.5 and Figure 3.6, our proposed models show good performance when estimating the prior mean and prior standard deviation of the random intercepts, β_{0i}, however, the main interest is on evaluating the performance of the proposed model on estimating the actual random intercepts. In Figure 3.7, we provide a grid of scatter plots each shows the relation between the true random intercept value and the posterior median estimates of those random intercepts. As Figure 3.7 shows, when random intercepts are distributed according to the standard Normal $N(\mu = 0, \sigma = 1)$ distribution, in terms of estimating the latent random intercepts, our proposed Mean-DPM and Sigma-DPM models work equally well as the the hierarchical Bayesian proportional hazard model with explicit Normal assumption on

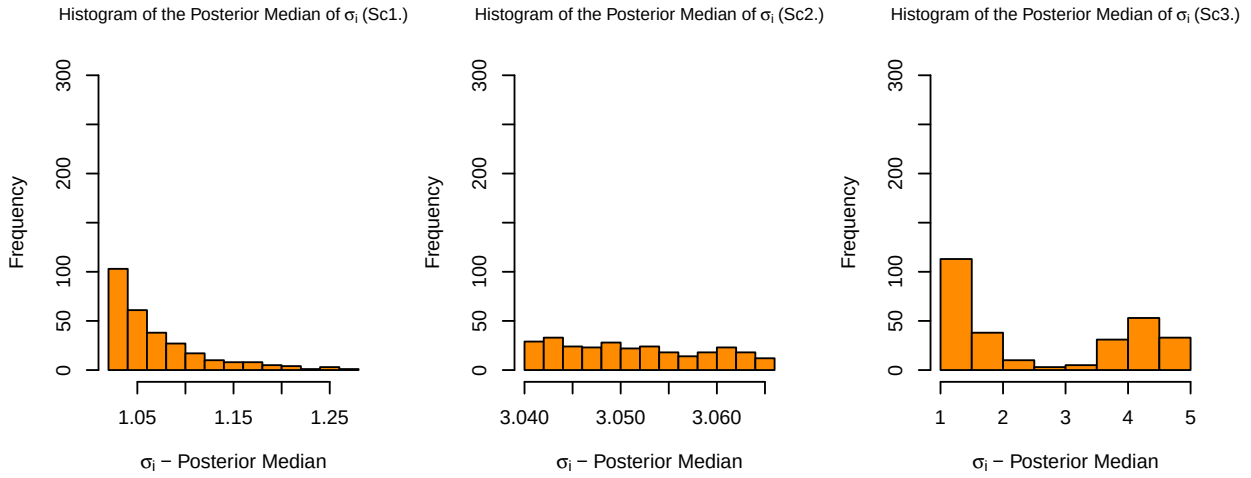


Figure 3.6: Histogram of the posterior median of σ_i's from the proposed Sigma-DPM hierarchical Bayesian proportional hazard model, where σ_i is the subject-specific prior standard deviation on the random intercept of subject i. The plot to the left is the histogram of the posterior median of the sampled σ_i from the model when it runs under the first simulation scenario where all random intercepts are sampled from the standard Normal distribution. The plot in the middle shows the histogram of the posterior medians of σ_i's under the second scenario where random intercepts are sampled from a mixture of two Normal distributions of $N(\mu_1 = -1.5, \sigma = 1)$ and $N(\mu_1 = 1.5, \sigma = 1)$ that are equally weighted. The plot to the right is the histogram of the posterior medians under the third simulation scenario where the random intercepts are simulated from the mixture of two Normal distributions of $N(\mu = 0, \sigma_1 = 1)$ and $N(\mu = 0, \sigma_2 = \sqrt{5})$. Results, under each simulation, are from one single simulated data with $N = 300$ subjects and $l_i = 12$ within subject measurements.

the random intercepts. When the reference distribution of the sampled random intercepts is not Normal, our proposed Mean-DPM and Sigma-DPM models that are robust to distributional mis-specification of the random intercepts, outperform the hierarchical Bayesian proportional hazard model in terms of estimating the latent random intercepts β_{0i}.

Tables 3.4, 3.5, and 3.6 show the results for the proportional hazards models. Coefficient estimates under the Cox model, the Bayesian marginal model, and the Weibull AFT model, all examples of marginal models, are smaller compared to the true conditional estimand and the marginal coefficient estimate under these models shrink toward 0.

By taking the differential subject-specific baseline hazard into account, the hierarchical Bayes

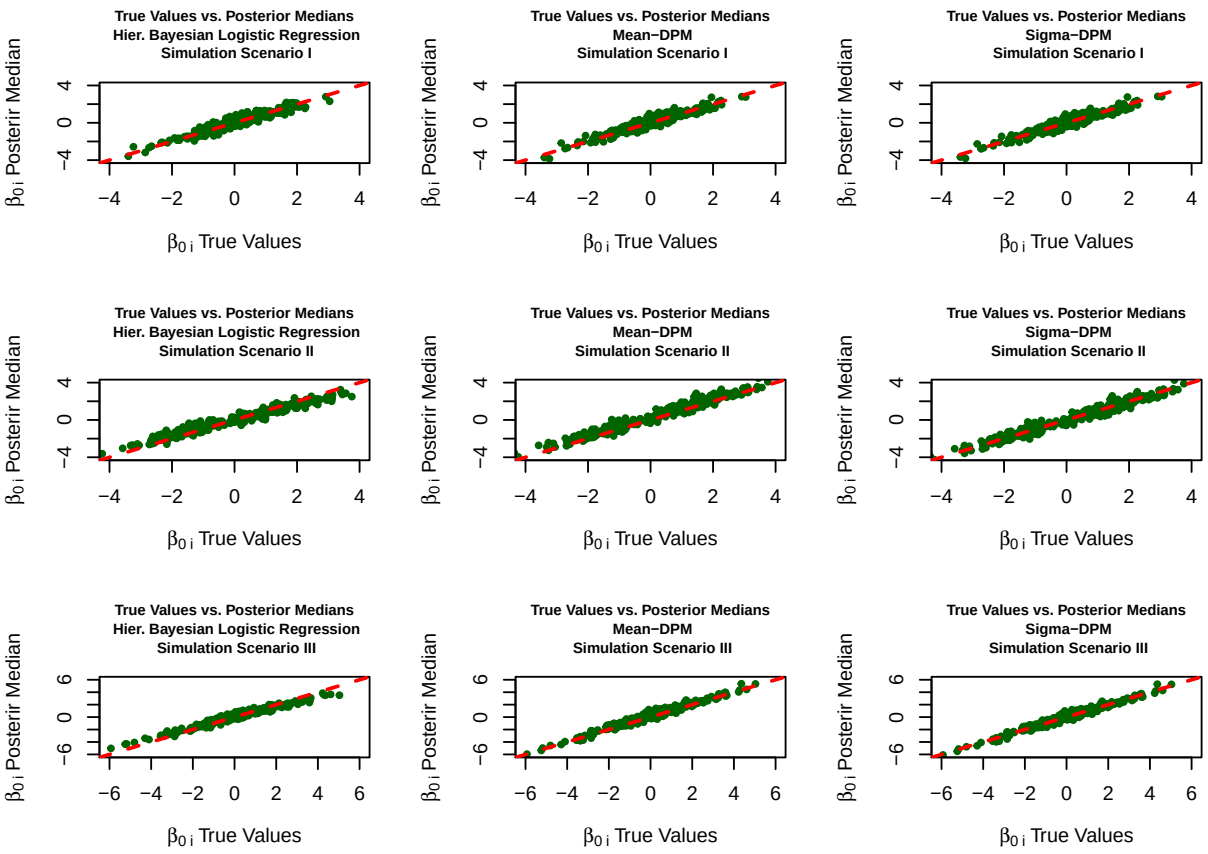


Figure 3.7: A grid of scatter plots that shows the relation between the true values of the subject-specific random intercepts, β_{0i}, and the posterior median of random intercepts from the hierarchical Bayesian proportional hazard model, our proposed Mean-DPM hierarchical Bayesian proportional hazard model, and the proposed Sigma-DPM hierarchical Bayesian proportional hazard model. The red dashed line in every plot represents the 45 degree line and the results are from a single simulated data under each simulation scenario. The first row represents the scatter plots from data simulated under the first scenario where subject-specific random intercepts are sampled from the standard Normal $N(\mu = 0, \sigma = 1)$. The second row represents scatter plots resulted from data simulated under the second simulation scenario where random intercepts are sampled from an equally weighted mixture of two Normal distributions of the form $N(\mu_1 = -1.5, \sigma = 1)$ and $N(\mu_1 = 1.5, \sigma = 1)$. Finally, the last row of plots represents results from data simulated under the third simulation scenario where random intercepts are sampled from an equally weighted mixture of two Normals of the form $N(\mu = 0, \sigma_1 = 1)$ and $N(\mu_1 = 0, \sigma_2 = \sqrt{5})$.The first column of scatter plots from left represents results from fitting the hierarchical Bayesian proportional hazard regression, the second column represents the results from fitting our proposed Mean-DPM hierarchical Bayesian logistic model, and finally the last column to the right represents results from our proposed Sigma-DPM hierarchical Bayesian proportional hazard model.

model with the Normal prior on random intercepts β_{0i} is capable of estimating the true conditional estimand when the random intercepts are truly Normally distributed (Table 3.4). However, the model is not robust to distributional mis-specification as under the second and the third scenarios, the coefficient estimate of β_1 shrank toward 0 (Table 3.5 and Table 3.6).

Finally, our proposed Mean-DPM and Sigma-DPM proportional hazards models assume no explicit distributional assumption on the random intercepts, are capable of detecting subject-specific random intercepts, and are robust to distributional mis-specification of the random intercepts. Hence, our proposed DPM proportional hazard models can estimate the true conditional estimand[2].

	$\beta_x = 1.000$	SD	MSE
Frequentist Cox Model	0.661	0.089	0.123
Weibull_AFT	0.709	0.096	0.095
Bayesian Marginal Proportional Hazard Model	0.700	0.038	0.100
Hierarchical Bayesian Proportional Hazard Model	1.015	0.122	0.014
Mean-DPM Proportional Hazard Model	0.995	0.124	0.015
Sigma-DPM Proportional Hazard Model	0.999	0.122	0.016

Table 3.4: Time-to-event data generated with differential subject-specific log baseline hazards induced by subject-specific random intercepts that are distributed according to a standard Normal distribution $N(\mu = 0, \sigma = 1)$. Results are from $1,000$ different simulated data each with $N = 300$ subjects and $l_i = 12$ within subject measurements.

3.4 Sensitivity Analysis

Using synthetic data, we showed that our proposed Mean-DPM and Sigma-DPM are capable of estimating latent cluster-specific intercepts and are robust to distributional mis-specification. Based on the simulation results presented in Section 3.3, in terms of MSE of

[2]posterior samples traceplots for Bayesian survival models are presented in Appendix B.1.2

	$\beta_x = 1.000$	SD	MSE
Frequentist Cox Model	0.471	0.101	0.290
Weibull_AFT	0.507	0.107	0.255
Bayesian Marginal Proportional Hazard Model	0.506	0.038	0.257
Hierarchical Bayesian Proportional Hazard Model	0.898	0.122	0.047
Mean-DPM Proportional Hazard Model	1.002	0.170	0.029
Sigma-DPM Proportional Hazard Model	1.000	0.209	0.033

Table 3.5: Time-to-event data generated with differential subject-specific log baseline hazards induced by subject-specific random intercepts that are distributed according to a mixture distribution of the form $\theta_i N(\mu = -1.5, \sigma = 1) + (1 - \theta_i)N(\mu = 1.5, \sigma = 1)$, where θ_i are distributed *Bernoulli* with parameter $p = 0.5$. Results are from $1,000$ different simulated data each with $N = 300$ subjects and $l_i = 12$ within subject measurements.

	$\beta_x = 1.000$	SD	MSE
Frequentist Cox Model	0.460	0.107	0.303
Weibull_AFT	0.481	0.109	0.292
Bayesian Marginal Proportional Hazard Model	0.483	0.038	0.290
Hierarchical Bayesian Proportional Hazard Model	0.924	0.121	0.037
Mean-DPM Proportional Hazard Model	1.014	0.184	0.029
Sigma-DPM Proportional Hazard Model	0.997	0.206	0.046

Table 3.6: Time-to-event data generated with differential subject-specific log baseline hazards induced by subject-specific random intercepts that are distributed according to a mixture distribution of the form $\theta_i N(\mu = 0, \sigma = 1) + (1 - \theta_i)N(\mu = 0, \sigma = \sqrt{5})$, where θ_i are distributed *Bernoulli* with parameter $p = 0.5$. Results are from $1,000$ different simulated data each with $N = 300$ subjects and $l_i = 12$ within subject measurements.

estimating conditional coefficients, our proposed models outperform common frequentist and Bayesian models to analyze repeated measure binary data and survival data. In this section, we are interested in testing the sensitivity of our proposed Mean-DPM and Sigma-DPM proportional hazards models with respect to the three main parameters of the number of within unit measurements (l_i), the difference in mean parameter μ_1 and μ_2 when random intercepts are simulated from the mixture of two Normal distributions of the form $N(\mu_1, \sigma)$ and $N(\mu_2, \sigma)$, and the ratio between the two parameters σ_1 and σ_2 when random intercepts are simulated from the mixture of two Normal distributions of the form $N(0, \sigma_1)$ and $N(0, \sigma_2)$.

3.4.1 Sensitivity to l_i

In this section, we test the sensitivity of our proposed Mean-DPM proportional hazards and Sigma-DPM proportional hazards models with respect to the number of within subject measurements l_i and under the case where the distribution of the random random intercepts is mis-specified. We generate synthetic repeated measure binary and survival data under two scenarios - one when subject-specific intercepts are sampled from a mixture distribution of the form $\beta_{0i} \overset{iid}{\sim} \theta_i N(\mu = -1.5, \sigma = 1) + (1-\theta_i)N(\mu = 1.5, \sigma = 1)$, and another when subject-specific intercepts are sampled from a mixture distribution of the form $\beta_{0i} \overset{iid}{\sim} \theta_i N(\mu = 0, \sigma = 1) + (1-\theta_i)N(\mu = 0, \sigma = \sqrt{5})$, where $\theta_i \sim Bernoulli(p = 0.5)$ with $i \in \{1, \ldots, n\}$ and n as the number of subjects. By changing the number of within subject measurements l_i, we test the sensitivity of our proposed models.

Figure 3.8 provides a histogram of posterior medians of the prior mean μ_i on the random intercepts β_{0i}. The results are from our proposed Mean-DPM hierarchical Bayesian proportional hazard model that is run on a single dataset that is generated under the simulation scenario where random intercepts β_{0i}'s are sampled from an equally weighted mixture of two Normal distributions with means $\mu_1 = 1.5$ or $\mu_2 = 1.5$ and with the standard deviation of 1. As one can see, as the number of within subject measurements l_i increases, our proposed Mean-DPM can better estimate the prior mean μ_i's with the true values that are either -1.5 or 1.5.

Similarly, Figure 3.9 provides a histogram of posterior medians of the prior standard deviation σ_i on the random intercepts β_{0i}. The results are from our proposed Sigma-DPM hierarchical Bayesian proportional hazard model that is run on a single dataset generated under the simulation scenario where random intercepts β_{0i}'s are sampled from an equally weighted mixture of two Normal distributions both with mean $\mu = 0$ and with the standard deviation of $\sigma_1 = 1$ and $\sigma_2 = \sqrt{5}$. As one can see, as the number of within subject

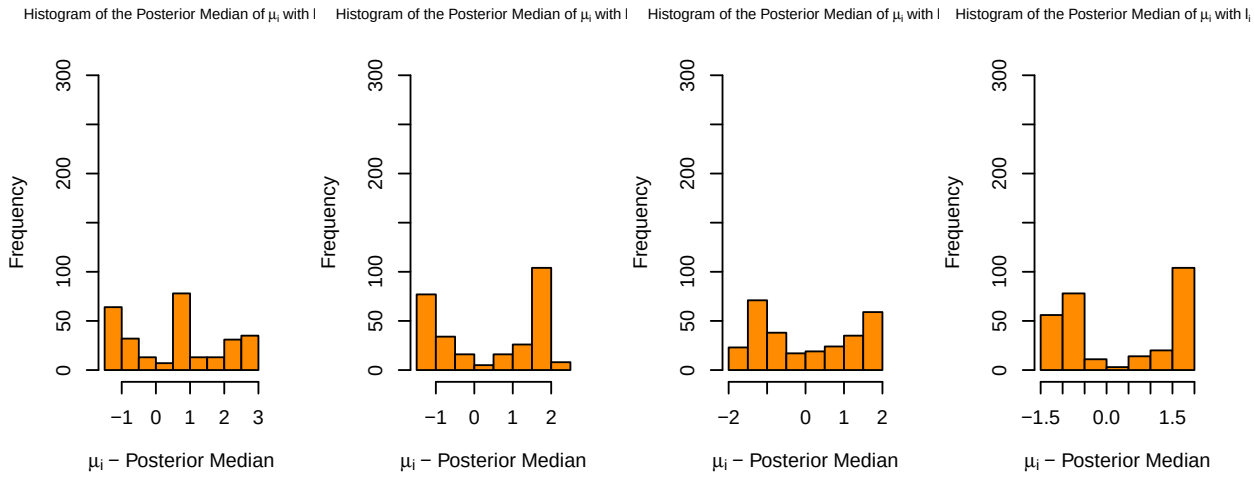


Figure 3.8: Histogram of the posterior median of μ_i's from the proposed Mean-DPM hierarchical Bayesian proportional hazard model, where μ_i is the subject-specific prior mean on the random intercept of subject i. All plot are based on a simulation scenario where random intercepts are sampled from a mixture of two Normal distributions of $N(\mu_1 = -1.5, \sigma = 1)$ and $N(\mu_1 = 1.5, \sigma = 1)$ that are equally weighted. Moving from left to right, the first plots shows posterior median of μ_i's with $l_i = 1$ within subject measurement, the next plot shows the results with $l_i = 3$, the next plot shows the results under data with $l_i = 6$ within subject measurements, and finally, the last plot to the right shows the results with $l_i = 12$ within subject measurements.

measurements l_i increases, our proposed Sigma-DPM can better estimate the prior standard deviations σ_i's with the true values that are either 1 or $\sqrt{5}$.

As Figure3.8 and Figure3.9 show, using our proposed Mean-DPM and Sigma-DPM hierarchical Bayesian proportional hazard model, the larger within subject number of measurements, l_i, are, the more accurate the posterior medians of prior means μ_i and prior standard deviations σ_i will be. μ_i and σ_i are the hyper-parameters that are parameters of prior distributions on the random intercepts β_{0i}.

Figure 3.10 includes scatterplots that show the relation between the true β_{0i} values and the posterior medians from our proposed Mean-DPM and Sigma-DPM proportional hazard models on simulated data with the true subject-specific random intercepts β_{0i} sampled from a mixture of two Normal distributions of the form $\theta_i N(\mu = -1.5, \sigma = 1) + (1 - \theta_i)N(\mu =$

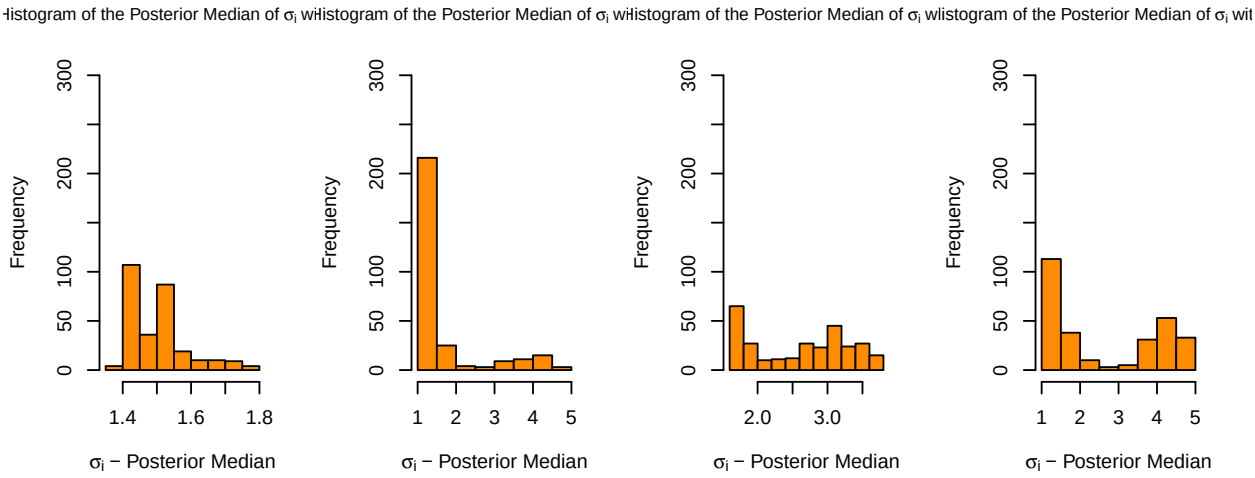


Figure 3.9: Histogram of the posterior median of σ_i's from the proposed Sigma-DPM hierarchical Bayesian proportional hazard model, where σ_i is the subject-specific prior standard deviation on the random intercept of subject i. All plot are based on a simulation scenario where random intercepts are sampled from a mixture of two Normal distributions of $N(\mu_1 = 0, \sigma = 1)$ and $N(\mu_1 = 0, \sigma = \sqrt{5})$ that are equally weighted. Moving from left to right, the first plots shows posterior median of σ_i's with $l_i = 1$ within subject measurement, the next plot shows the results with $l_i = 3$, the next plot shows the results under data with $l_i = 6$ within subject measurements, and finally, the last plot to the right shows the results with $l_i = 12$ within subject measurements.

$1.5, \sigma = 1)$, where θ_i is distributed Bernoulli with the parameter $p = 0.5$. As one can infer from the plots in this figure, as the number of within-subject measurements increase, posterior medians of the random intercepts provide a more accurate estimate of the true β_{0i}.

Similarly, Figure 3.11 includes similar scatterplots that show the relation between the true β_{0i} values and the posterior medians from our proposed Mean-DPM and Sigma-DPM proportional hazard models on data simulated with the true subject-specific random intercepts β_{0i} sampled from a mixture of two Normal distributions of the form $\theta_i N(\mu = 0, \sigma = 1) + (1 - \theta_i) N(\mu = 0, \sigma = \sqrt{5})$, where θ_i is distributed Bernoulli with the parameter $p = 0.5$. From the plots in the figure, one can clearly realize that as the number of within-subject measurements increase, posterior medians of the random intercepts provide a more accurate estimate of the true β_{0i}.

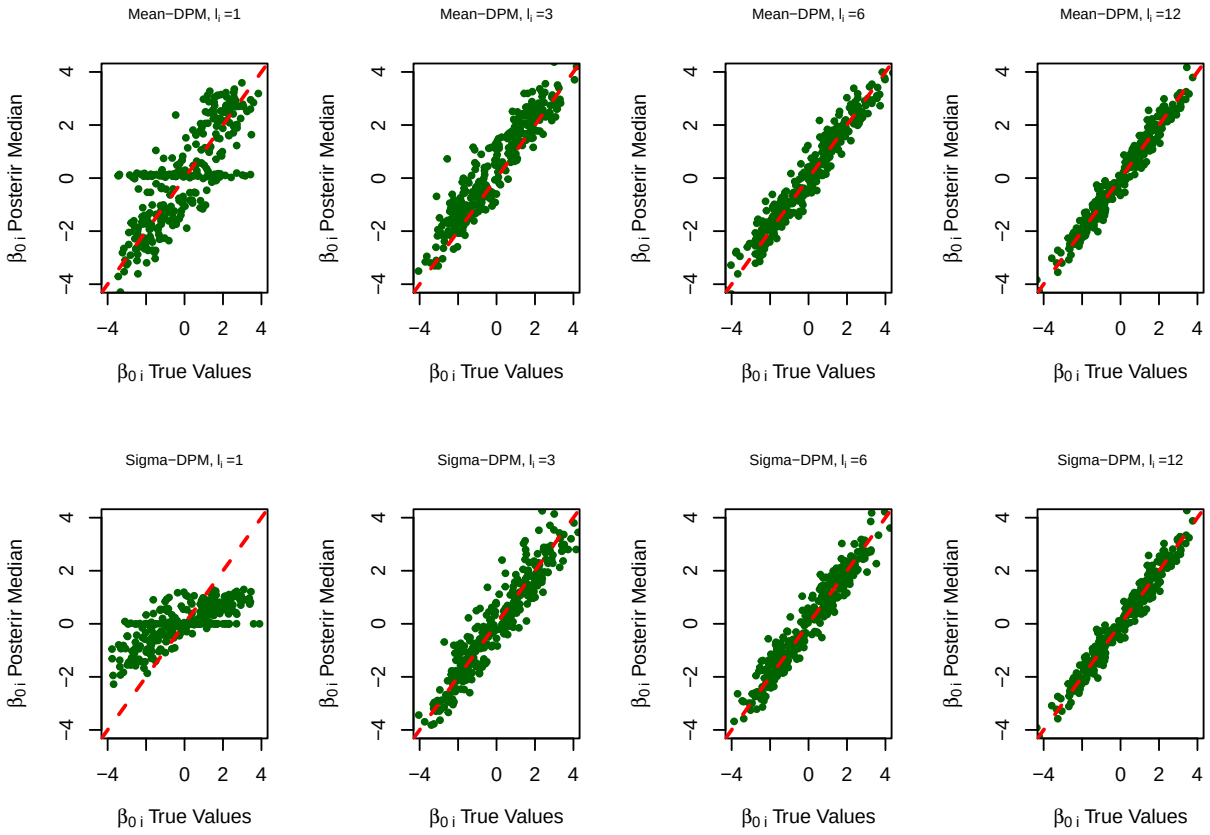


Figure 3.10: A grid of scatter plots that shows the relation between the true values of the subject-specific random intercepts, β_{0i}, and the posterior median of random intercepts from our proposed Mean-DPM and Sigma-DPM hierarchical Bayesian proportional hazard models. The red dashed line in every plot represents the 45 degree line and the results are from a single simulated under the simulation scenario where random intercepts β_{0i} are simulated from an equally weighted mixture of two Normal distributions one with mean $\mu_1 = -1.5$ and the other with mean $\mu_2 = 1.5$, where both distributions have the standard deviation of $\sigma = 1$. The first row represents the results from our proposed Mean-DPM and the second row represents results from our proposed Sigma-DPM model. On each row, from left to right, the scatter plots represents the results from a simulated data with $l_i = 1$, $l_i = 3$, $l_i = 6$, and $l_i = 12$ within subject measurements.

Table 3.7 provides results on the sensitivity of our models under the first simulation scenario and table 3.8 provides the result on the sensitivity of our models under the second simulation scenario. As the results in Table 3.7 and Table 3.8 show, with larger number of within subject measurements l_i, our proposed models can better estimate the latent random intercepts, and hence, lead to a smaller error in estimating the true conditional coefficient estimate.

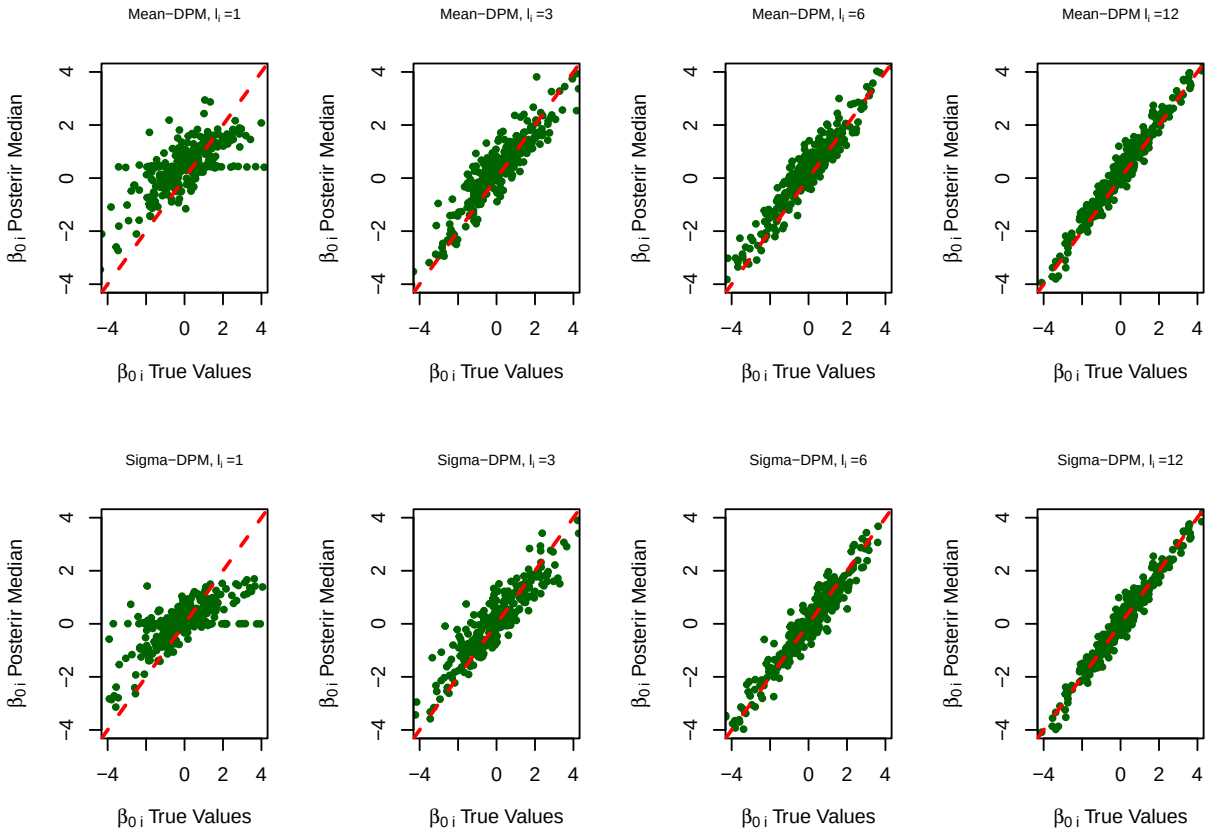


Figure 3.11: A grid of scatter plots that shows the relation between the true values of the subject-specific random intercepts, β_{0i}, and the posterior median of random intercepts from our proposed Mean-DPM and Sigma-DPM hierarchical Bayesian proportional hazard models. The red dashed line in every plot represents the 45 degree line and the results are from a single simulated under the simulation scenario where random intercepts β_{0i} are simulated from an equally weighted mixture of two Normal distributions both with mean $\mu = 0$ but one with the standard deviation $\sigma_1 = 1$ and another with the standard deviation of $\sigma_2 = \sqrt{5}$. The first row represents the results from our proposed Mean-DPM and the second row represents results from our proposed Sigma-DPM model. On each row, from left to right, the scatter plots represents the results from a simulated data with $l_i = 1$, $l_i = 3$, $l_i = 6$, and $l_i = 12$ within subject measurements.

3.4.2 Sensitivity to $|\mu_2 - \mu_1|$

In this section, we test the sensitivity of our proposed Mean-DPM and Sigma-DPM proportional hazards models with respect to the distance between the mean parameters μ_1 and μ_2, where μ_1 and μ_2 are the mean parameters of two Normal distributions that are used

	Mean-DPM			Sigma-DPM		
l_i	$\beta_x = 1.000$	SD	MSE	$\beta_x = 1.000$	SD	MSE
1	0.998	0.223	0.0291	0.754	0.228	0.104
3	1.014	0.200	0.0283	0.939	0.218	0.044
6	1.010	0.182	0.033	0.958	0.211	0.039
12	1.002	0.170	0.029	1.000	0.209	0.033

Table 3.7: To test the sensitivity of our proposed proportional hazards models with respect to the number of within subject measurements l_i, time-to-event data generated with differential subject-specific log baseline hazards induced by subject-specific random intercepts that are distributed according to a mixture distribution of the form $\theta_i N(\mu = -1.5, \sigma = 1) + (1 - \theta_i)N(\mu = 1.5, \sigma = 1)$, where θ_i are distributed $Bernoulli$ with parameter $p = 0.5$. Results are from $1,000$ different simulated data each with $N = 300$ subjects and l_i within subject measurements.

	Mean-DPM			Sigma-DPM		
l_i	$\beta_x = 1.000$	SD	MSE	$\beta_x = 1.000$	SD	MSE
1	0.939	0.321	0.077	0.803	0.237	0.080
3	0.984	0.201	0.031	0.947	0.201	0.039
6	0.987	0.190	0.039	0.995	0.210	0.047
12	1.014	0.184	0.046	0.997	0.206	0.046

Table 3.8: To test the sensitivity of our proposed proportional hazards models with respect to the number of within subject measurements l_i, time-to-event data were generated with differential subject-specific log baseline hazards induced by the subject-specific random intercept. The random intercepts are distributed according to a mixture distribution of the form $\theta_i N(\mu = 0, \sigma = 1) + (1 - \theta_i)N(\mu = 1.5, \sigma = \sqrt{5})$, where θ_i are distributed $Bernoulli$ with parameter $p = 0.5$. Results are from $1,000$ different simulated data each with $N = 300$ subjects and l_i within subject measurements.

to simulate subject-specific random intercepts. Subject-specific random intercepts are sampled from a mixture of two Normal distributions of the form $\beta_{0i} \overset{iid}{\sim} \theta_i N(\mu = -1.5, \sigma = 1) + (1 - \theta_i)N(\mu = 1.5, \sigma = 1)$, where θ_i is distributed Bernoulli with the parameter $P = 0.5$. In this section, we evaluate the sensitivity of our proposed Mean-DPM and Sigma-DPM proportional hazards models with respect to the distance between the means μ_1 and μ_2. In particular, we consider five cases where the distance is half of the standard deviation shared between both components, σ, or is equal to the σ, or is two times bigger than the σ, or three times bigger, or four times bigger (Table 3.9).

Figure 3.12 provides a histogram of posterior medians of the prior mean μ_i on the random

intercepts β_{0i}. The results are from our proposed Mean-DPM hierarchical Bayesian proportional hazard model that is run on a single dataset that is generated under the simulation scenario where random intercepts β_{0i}'s are sampled from an equally weighted mixture of two Normal distributions with means μ_1 or μ_2 and with the standard deviation of $\sigma = 1$. In order to test the sensitivity of our models with respect to the distance between μ_1 and μ_2, we consider 5 cases based on the distance between μ_1 and μ_2. Those cases are when the distance between the means is half of the standard deviation σ, equal to σ, twice of the σ, three times of the σ, or four times of the σ.

Figure 3.13 includes scatterplots that show the relation between the true β_{0i} values and the posterior medians from our proposed Mean-DPM and Sigma-DPM proportional hazard models on simulated data with the true subject-specific random intercepts β_{0i} sampled from a mixture of two Normal distributions of the form $\theta_i N(\mu_1, \sigma = 1) + (1 - \theta_i)N(\mu_2, \sigma = 1)$, where θ_i is distributed Bernoulli with the parameter $p = 0.5$. To test the sensitivity of our proposed models with respect to the distance between μ_1 and μ_2, we consider 5 different cases. Those cases are when the distance between the means are $\sigma/2$, σ, 2σ, 3σ, and 4σ.

As the results in Table 3.9 show, our proposed models are very robust in terms of the distance between the mean parameters μ_1 and μ_2. One may consider this fact that when μ_1 and μ_2 are far apart, the Dirichlet process mixture prior can easily differentiate random intercepts that are sampled from the Normal distribution with the mean μ_1 from random intercepts sampled from the Normal distribution with the mean μ_2. On the other hand, when μ_1 and μ_2 are very close, a Normal prior with an incorrectly specified mean can still cover the random intercepts that are sampled from the correct Normal distribution. Hence, our proposed models are not sensitive to the distance between the means of the Normal distributions they are sampled from.

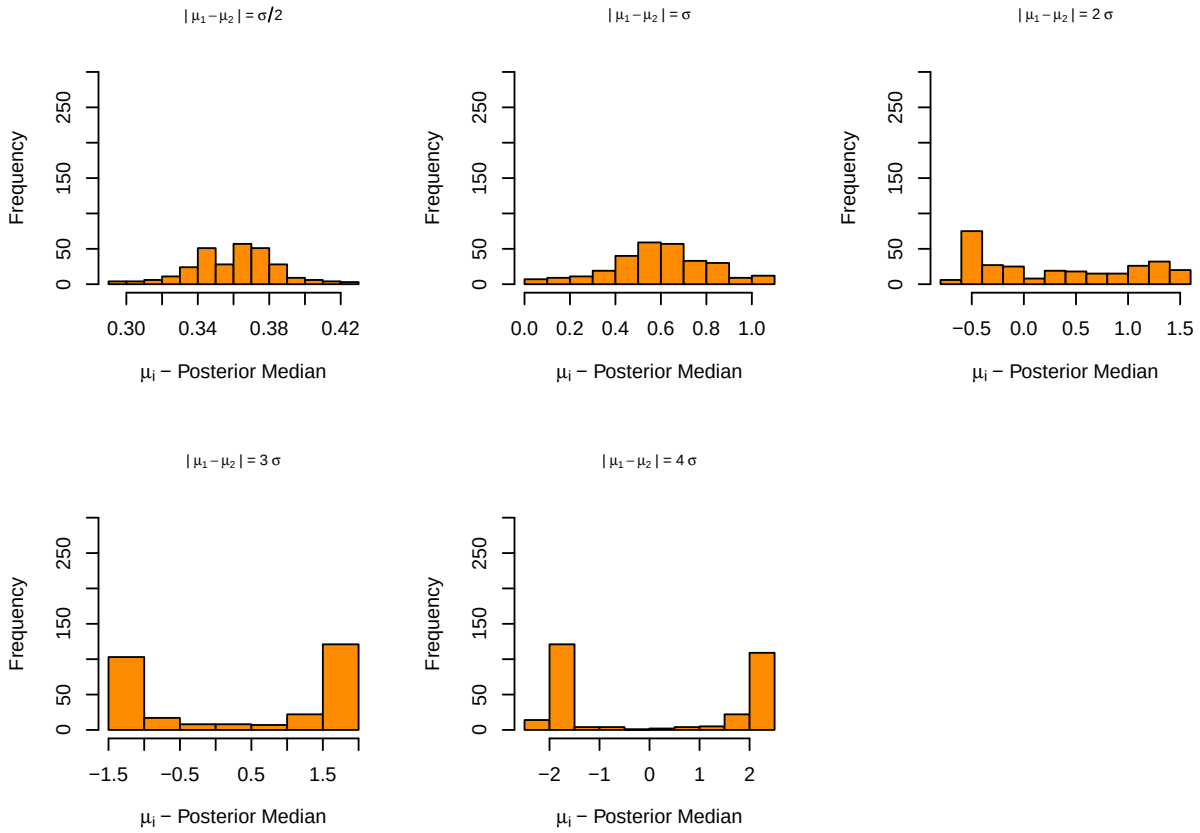


Figure 3.12: Histogram of the posterior median of μ_i's from the proposed Mean-DPM hierarchical Bayesian proportional hazard model, where μ_i is the subject-specific prior mean on the random intercept of subject i. All plot are based on a simulation scenario where random intercepts are sampled from a mixture of two Normal distributions of $N(\mu_1, \sigma = 1)$ and $N(\mu_2, \sigma = 1)$ that are equally weighted with $N = 300$ subjects each with $l_i = 12$ within subject measurements. Moving from the left to right, the first plots shows posterior median of μ_i's when $\mu_1 = -0.25$ and $\mu_2 = 0.25$ (a distance of $\sigma/2$), the next plot shows the results when $\mu_1 = -0.5$ and $\mu_2 = 0.5$ (a distance of σ), the next plot is corresponding to the true $\mu_1 = -1.0$ and $\mu_2 = 1.0$ (a distance of 2σ), the next plot is corresponding to the true $\mu_1 = -1.5$ and $\mu_2 = 1.5$ (a distance of 3σ), the next plot is corresponding to the true $\mu_1 = -2$ and $\mu_2 = 2$ (a distance of 4σ).

3.4.3 Sensitivity to $\frac{\sigma_2}{\sigma_1}$

In this section, we test the sensitivity of our proposed Mean-DPM and Sigma-DPM proportional hazards models with respect to the relative ratio of the standard deviations σ_1 and σ_2 when the subject-specific random intercepts are sampled from a mixture of two Normal

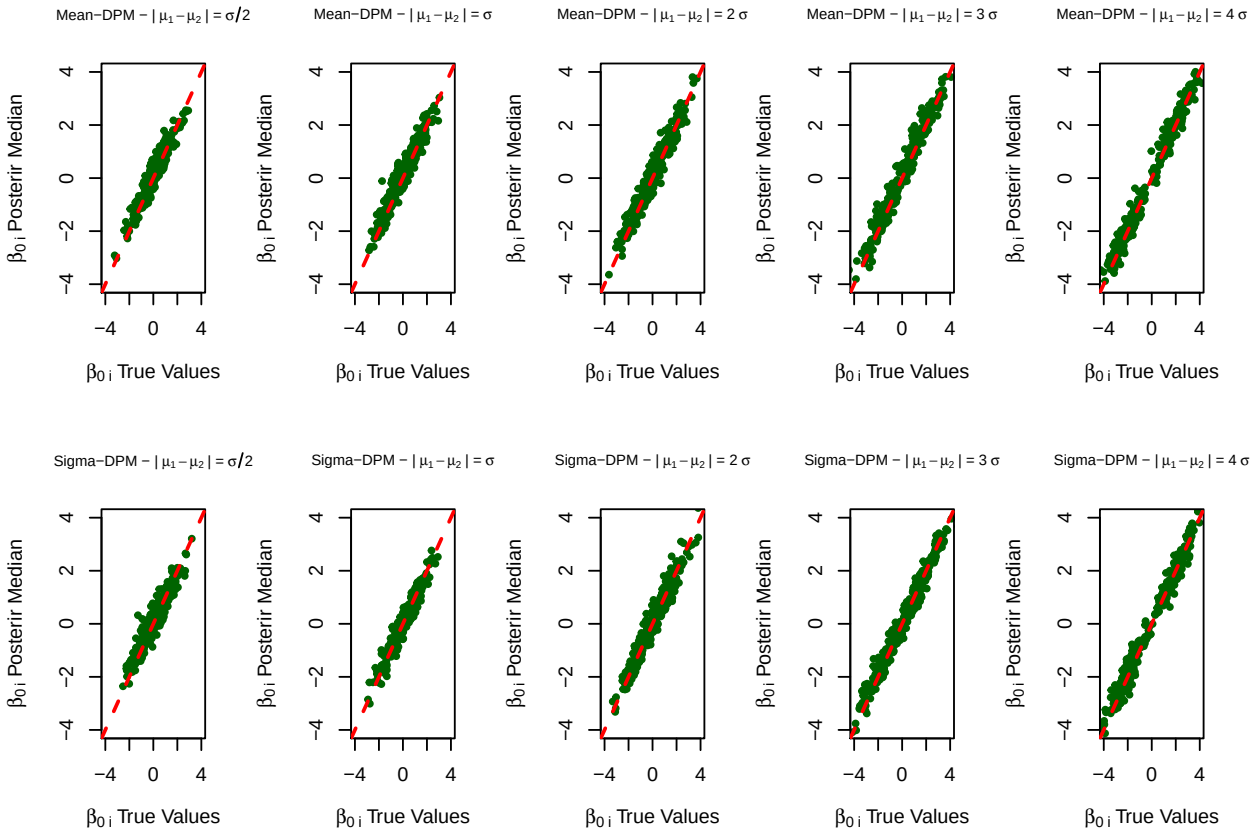


Figure 3.13: A grid of scatter plots that shows the relation between the true values of the subject-specific random intercepts, β_{0i}, and the posterior median of random intercepts from our proposed Mean-DPM and Sigma-DPM hierarchical Bayesian proportional hazard models. The red dashed line in every plot represents the 45 degree line and the results are from a single simulated under the simulation scenario where random intercepts β_{0i} are simulated from an equally weighted mixture of two Normal distributions of $N(\mu_1, \sigma = 1)$ and $N(\mu_2, \sigma = 1)$. The first row represents the results from our proposed Mean-DPM and the second row represents results from our proposed Sigma-DPM model. On each row, from the left to the right, the scatter plots represents the results from a simulated data under the 5 cases of $\mu_1 = -0.25$ and $\mu_2 = 0.25$ (a distance of $\sigma/2$), $\mu_1 = -0.5$ and $\mu_2 = 0.5$ (a distance of σ), $\mu_1 = -1.0$ and $\mu_2 = 1.0$ (a distance of 2σ), $\mu_1 = -1.5$ and $\mu_2 = 1.5$ (a distance of 3σ), and $\mu_1 = -2$ and $\mu_2 = 2$ (a distance of 4σ).

distributions of the form $\beta_{0i} \overset{iid}{\sim} \theta_i N(\mu = 0, \sigma_1) + (1 - \theta_i) N(\mu = 0, \sigma_2)$, where θ_i is distributed Bernoulli with the parameter $P = 0.5$. In this section, we evaluate the sensitivity of our proposed Mean-DPM and Sigma-DPM proportional hazards models with respect to the relative ratio of σ_1 and σ_2 that is of the from σ_2/σ_1. In particular, we consider four cases where the ratio 1.5, or the ratio is 2.0, or 3.0, or 5.0. As the results in Table 3.10 show, our proposed

	Mean-DPM			Sigma-DPM		
$\lvert\mu_1 - \mu_2\rvert$	$\beta_x = 1.000$	SD	MSE	$\beta_x = 1.000$	SD	MSE
$\sigma/2$	0.996	0.125	0.016	1.021	0.216	0.043
σ	1.000	0.132	0.020	1.019	0.231	0.049
2σ	1.021	0.164	0.027	0.987	0.285	0.076
3σ	1.002	0.170	0.029	1.000	0.209	0.033
4σ	0.998	0.152	0.021	1.001	0.211	0.034

Table 3.9: To test the sensitivity of our proposed proportional hazards models with respect to the distance between μ_1 and μ_2, time-to-event data were generated with differential subject-specific log baseline hazards induced by the subject-specific random intercept. The random intercepts are distributed according to a mixture distribution of the form $\theta_i N(\mu_1, \sigma = 1) + (1-\theta_i)N(\mu_2, \sigma = 1)$, where θ_i are distributed *Bernoulli* with parameter $p = 0.5$. Results are from $1,000$ different simulated data each with $N = 300$ subjects and $l_i = 12$ within subject measurements.

models are robust to the changes in the ratio between the standard deviations of the mixture components.

	Mean-DPM			Sigma-DPM		
σ_2/σ_1	$\beta_x = 1.000$	SD	MSE	$\beta_x = 1.000$	SD	MSE
1.5	0.990	0.255	0.060	1.001	0.257	0.069
2.0	0.988	0.302	0.100	1.000	0.304	0.090
3.0	0.960	0.351	0.113	0.989	0.338	0.118
5.0	1.027	0.368	0.129	0.976	0.365	0.129

Table 3.10: To test the sensitivity of our proposed proportional hazards models with respect to the ratio of σ_1 and σ_2, time-to-event data were generated with differential subject-specific log baseline hazards induced by the subject-specific random intercept. The random intercepts are distributed according to a mixture distribution of the form $\theta_i N(\mu = 0, \sigma_1) + (1-\theta_i)N(\mu = 0, \sigma_2)$, where θ_i are distributed *Bernoulli* with parameter $p = 0.5$. Results are from $1,000$ different simulated data each with $N = 300$ subjects and $l_i = 12$ within subject measurements.

Figure 3.14 provides a histogram of posterior medians of the prior standard deviation σ_i on the random intercepts β_{0i}. The results are from our proposed Sigma-DPM hierarchical Bayesian proportional hazard model that is run on a single dataset that is generated under the simulation scenario where random intercepts β_{0i}'s are sampled from an equally weighted mixture of two Normal distributions of the form $N(\mu = 0, \sigma_1)$ and $N(\mu = 0, \sigma_2)$. In order to test the sensitivity of our models with respect to the relative ratio of σ_2 and σ_1 (σ_2/σ_1),

we consider 4 cases. Those cases are when the relative ratio of σ_2/σ_1 is either 1.5, or 2.0, or 3.0, or 5.0.

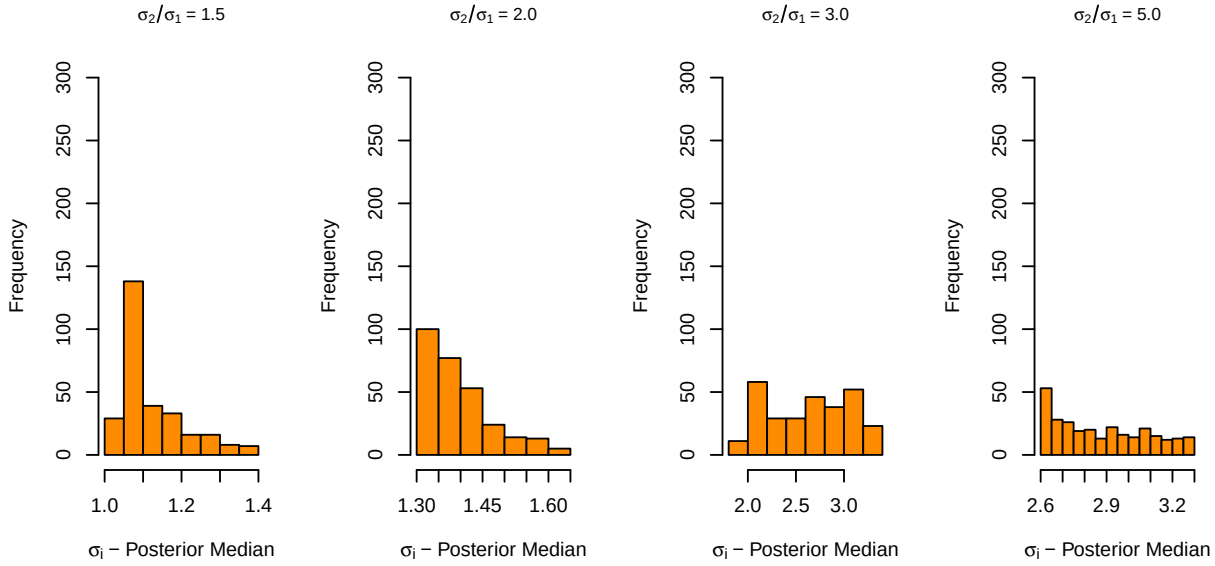


Figure 3.14: Histogram of the posterior median of σ_i's from the proposed Sigma-DPM hierarchical Bayesian proportional hazard model, where σ_i is the subject-specific prior standard deviation on the random intercept of subject i. All plot are based on a simulation scenario where random intercepts are sampled from a mixture of two Normal distributions of $N(\mu = 0, \sigma_1)$ and $N(\mu = 0, \sigma_2)$ that are equally weighted with $N = 300$ subjects each with $l_i = 12$ within subject measurements. Moving from the left to right, the first plots shows posterior median of σ_i's when $\sigma_1 = 1$ and $\sigma_2 = 1.5$ (a relative ratio of 1.5), the next plot shows the results when $\sigma_1 = 1$ and $\sigma_2 = 2.0$ (a relative ratio of 2.0), the next plot is corresponding to the true $\sigma_1 = 1$ and $\sigma_2 = 3.0$ (a relative ratio of 3.0), and the last plot to the right is corresponding to the true $\sigma_1 = 1.0$ and $\sigma_2 = 5.0$ (a relative ratio of 5.0).

Figure 3.15 includes scatterplots that show the relation between the true β_{0i} values and the posterior medians from our proposed Mean-DPM and Sigma-DPM proportional hazard models under the four simulation scenarios and on simulated data with the true subject-specific random intercepts β_{0i} sampled from a mixture of two Normal distributions of the form $\theta_i N(\mu = 0, \sigma_1) + (1 - \theta_i)N(\mu = 0, \sigma_2)$, where θ_i is distributed Bernoulli with the parameter $p = 0.5$

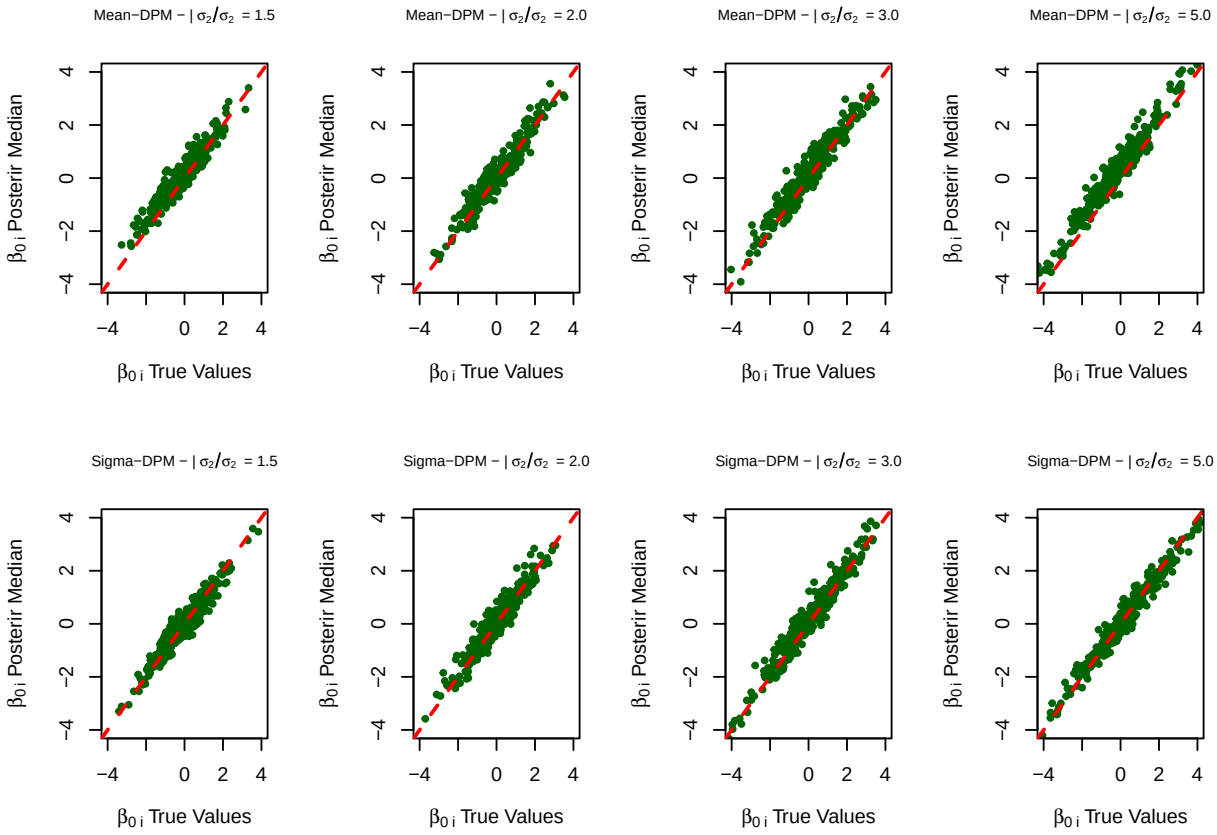


Figure 3.15: A grid of scatter plots that shows the relation between the true values of the subject-specific random intercepts, β_{0i}, and the posterior median of random intercepts from our proposed Mean-DPM and Sigma-DPM hierarchical Bayesian proportional hazard models. The red dashed line in every plot represents the 45 degree line and the results are from a single simulated under the simulation scenario where random intercepts β_{0i} are simulated from an equally weighted mixture of two Normal distributions of $N(\mu = 0, \sigma_1)$ and $N(\mu = 0, \sigma_2)$. The first row represents the results from our proposed Mean-DPM and the second row represents results from our proposed Sigma-DPM model. On each row, from the left to the right, the scatter plots represents the results from a simulated data under the 4 cases of $\sigma_1 = 1.0$ and $\sigma_2 = 1.5$ (a relative ratio of 1.5), $\sigma_1 = 1.0$ and $\sigma_2 = 2.0$ (a relative ratio of 2.0), $\sigma_1 = 1.0$ and $\sigma_2 = 3.0$ (a relative ratio of 3.0), and $\sigma_1 = 1.0$ and $\sigma_2 = 5.0$ (a relative ratio of 5.0).

3.5 Application of the Proposed Models to Compare Durability of Different Dialysis Access Types Among Hemodialysis Patients

End stage renal disease (ESRD) is a condition where kidneys are not capable of filtering blood from toxins. Standard care for ESRD patients are either kidney transplantation or hemodialysis. Hemodialysis is a technique that removes blood from the body through access needles and cleans the blood out of toxins using a dialysis machine.

ESRD patients who are treated with hemodialysis, typically undergo this treatment three to four sessions a week each session three to four hours. Given the frequency of the treatment, it's unfeasible to insert a new access at every treatment session as repeatedly inserting a new access may result in irreparable damage to the patient's vein. As an alternative to a temporary access, a permanent access may be surgically placed in patient's body. Permanent accesses are in two main types of prosthetic graft and autogenous arteriovenous fistula (AVF). Prosthetic graft can be easily placed in patient's body. Similarly, AVF access can be placed as a standard attachment to a vein. When veins are hard to find, which is common among diabetic patients, AVF access is placed in the patient's body using a venous transplantation.

Although permanent access technology has improved over time, yet access failure has remained a major issue among the hemodialysis patients. It's of interest to compare durability of different access types among hemodialysis patients. To do so, observational data were collected on 1,255 hemodialysis patients from clinics across the United States. Patients were asked to participate in the study at the time that they had their first permanent access placement. They were then followed over time prospectively and the time to failure from the time of access placement was recorded. Since patients must always have an access in order to do hemodialysis, if an access fails, the access is replaced with another access which may

be of a different type than the previous access. Our data include an overall of 1,647 access records from an overall of 1,255 subjects. Some subjects may have multiple access failures during the study. In particular, over the study followup, 76.7% of subjects had no access failure, 18% of subjects had one access failure, 4% had two access failures, and 1.3% had three or more access failures.

Table 3.11 shows the result of analyzing the association between the access type and time to failure of the access. We started by analyzing the data using the Cox proportional hazards model. Note that in the case of multiple access failures per subject in the data, within subject measurements are correlated. In this case, the within subject correlation should be taken into account and standard errors of the estimated coefficients should be taken into account. To do so, we considered the approach proposed by Lee et al. (1992) in which first the coefficients in the Cox model are estimated using maximizing the partial likelihood under an independent covariance assumption and then a robust sandwich covariance matrix is used to account for within-cluster correlations. This method is available in R programming language using the 'cluster()' function that is used inside the 'coxph()' function in order to fit a Cox model.

Next, we analyzed the data with our proposed Mean-DPM and Sigma-DPM proportional hazards model. While Cox model is not capable of taking the latent population subgroups into account and hence, coefficient estimates from this model are marginalized over all population subgroups, our proposed Mean-DPM and Sigma-DPM models, however, by accounting for the differential subject-specific baseline hazards, are capable of estimating the conditional coefficient estimates that are conditioned on subject-specific baseline hazards.

Our proposed Mean-DPM and Sigma-DPM models suggest that different access types, after adjusting for other potential risk factors in the model, are different in terms of the risk of failure. In particular, compared to the graft access, both venous and the standard fistula method have higher risk of failure under. The fitted Cox model, however, finds the standard

Covariates	No. of Cases	No. of Failure	Cox Model Relative Risk (95% CI)	P-value	Mean-DPM Model Relative Risk (95% CR)	Sigma-DPM Model Relative Risk (95% CR)
Access Type						
graft	1,140	271	1.0		1.0	1.0
standard fistula	367	81	0.91 (0.70,1.18)	0.471	1.11 (0.80,2.13)	1.09 (0.73, 2.18)
venous transposition fistula	140	40	1.43 (1.02,2.00)	0.039	1.46 (0.96,2.19)	1.45 (0.93,2.17)
Age	1,647	392	1.00 (0.99-1.01)	0.513	1.00 (0.94,1.01)	1.01 (0.92,1.04)
Female	1,647	392	1.11 (0.90-1.38)	0.328	1.12 (0.85,1.52)	1.14 (0.82,1.57)
Race						
Caucasian	987	218	1.0		1.0	1.0
African American	550	152	1.22 (0.98,1.52)	0.071	1.24 (0.71,1.65)	1.24 (0.70,1.65)
other	110	22	0.79 (0.50,1.22)	0.288	0.81 (0.43,1.40)	0.82 (0.39,1.38)
BMI	1,647	392	0.99 (0.98-1.01)	0.287	0.99 (0.93,1.01)	0.98 (0.91,1.03)
Smoking						
never smoked	900	219	1.0		1.0	1.0
former smoker	517	116	0.98 (0.78,1.24)	0.89	0.98 (0.66,1.30)	0.99 (0.64,1.32)
current smoker	230	57	1.10 (0.81,1.50)	0.541	1.08 (0.44,1.61)	1.07 (0.42,1.65)
Serum Calcium (mg/dL)	1,647	392	0.97 (0.87-1.08)	0.595	0.99 (0.19,1.10)	0.99 (0.18,1.13)
Serum Phosphorus (mg/dL)	1,647	392	1.02 (0.96-1.07)	0.524	1.00 (0.71,1.09)	1.00 (0.68,1.10)
Hematocrit (g/dL)	1,647	392	0.99 (0.97-1.01)	0.317	0.99 (0.91,1.01)	0.99 (0.90,1.04)
Serum Albumin (g/dL)	1,647	392	1.01 (0.84-1.21)	0.909	0.88 (0.36,1.24)	0.83 (0.36,1.29)
Diabetes	1,647	392	1.24 (1.01-1.54)	0.041	1.18 (0.5,1.58)	1.16 (0.52 ,1.51)

Table 3.11: In order to compare durability of different hemodialysis access types, observational data on 1,255 hemodialysis patients were analyzed using the Cox proportional hazards model, our proposed Mean-DPM proportional hazards model, and our proposed Sigma-DPM hazards model.

fistula access to have lower risk of failure compared to the graft method. This difference might be an indication of the attenuation in the marginal coefficient estimates under the Cox model, compared to the conditional coefficient estimate suggested by our proposed Mean-DPM model.

3.6 Discussion

A model with different marginal and conditional coefficient estimands is a non-collapsible model. Examples of such models include the logistic regression and the proportional hazard models. In this chapter and in the context of analyzing repeated measure data, we proposed hierarchical Bayesian models with the Dirichlet process mixture priors and we show our proposed models are capable of detecting latent subgroup effects and hence, are capable to estimate the true conditional parameters where a population consists of sub-populations with latent sub-population effects. In particular, we considered hierarchical Bayesian logistic regression and hierarchical Bayesian proportional hazards models with the Dirichlet process mixture prior on latent subgroup intercepts. We compared coefficient estimates under our proposed models with the coefficient estimates under common logistic regression and proportional hazards models. Further, we showed that our proposed models are robust to distributional mis-specification of the latent subgroup effects. Further, the sensitivity of our proposed models were tested in terms of their sensitivity to the number of within-cluster measurements as well as the distribution parameters of the latent cluster-specific intercepts.

Using simulation studies, we compared coefficient estimation under our proposed Dirichlet process mixture models with common statistical longitudinal models. In particular, we compared our proposed Dirichlet process logistic regression models with the generalized linear model with the logit link, the generalized estimating equation with the logit link, the generalized linear mixed effects model with the logit link, Bayesian logistic regression, and Bayesian hierarchical logistic regression. We also compared our proposed proportional hazards models with the frequentist Cox model, the Weibull accelerated failure time model, a marginal Bayesian proportional hazards model, and a hierarchical Bayesian model. We learned that among all these models, our proposed Dirichlet process mixture models lead to the minimum mean squared errors in estimating the conditional coefficient estimands. Furthermore, while other candidate models may depend on explicit distributional assump-

tions over the latent sub-group random intercepts, our proposed Dirichlet process mixture models are robust to distributional mis-specification. Using sensitivity analysis, we showed that our proposed Dirichlet process mixture models are robust in terms of the number of within-cluster measurements. We also showed that when cluster-specific random intercepts simulated from a mixture of two normal distributions, our proposed models are robust regardless of the distributional overlap of the mixing components. More generally and with the support of the simulation studies presented in this chapter, in analyses aiming to characterize conditional effect of covariates using the proportional hazards or the logistic models, our proposed Dirichlet process mixture models will serve the best in terms of mean square error of estimating conditional estimands compared to other candidate models that were considered in this chapter.

Despite the capability of our proposed methods in estimating conditional estimands in repeated measure data with latent sub-group random intercepts, our proposed methods, however, are computationally demanding. Our proposed Dirichlet mixture models, on average, and for a dataset with 300 subjects each with 12 within subject measurements and using, takes 3 hours to fit using a 2.53 GHz intel Core 2 Duo processor and 4 GB 1067 MHz DDR3 RAM. In future, instead of using MCMC posterior sampling, one may use the variational methods in Dirichlet process mixture models to gain more computational efficiency and more scalability as the number of subjects and the number of within-subject measurements increase.

Robustness to distributional mis-specifications and the capability of estimating conditional covariate effects even under the non-collapsible models have made Dirichlet process mixture models an interesting modeling choice. In the next chapter, we will introduce a broad joint longitudinal-survival modeling framework, where as part of the joint model, we propose using a Dirichlet process mixture survival model in order to better estimate conditional covariate effects.

Chapter 4

Flexible Joint Longitudinal-Survival Models for Quantifying the Association Between Longitudinal Biomarkers and Survival Outcomes

In this chapter we propose flexible joint longitudinal-survival models in order to test the association between a longitudinally collected biomarker and a time-to-event endpoint. Our proposed models are robust to common parametric and semi-parametric assumptions in that they avoid explicit distributional assumptions on longitudinal measures and allow for subject-specific baseline hazard in the survival component. Fully joint estimation is performed to account for the uncertainty in the estimated time-dependent biomarker covariate in the survival model.

We start with a brief introduction in Section 4.1. We then introduce the methodology in Section 4.2. In Section 4.3, simulation studies are presented to assess the operating

characteristics of the proposed models. All models are then applied to the data on end-stage renal disease patients that was obtained from the United States Renal Data System in Section 4.4. We finally conclude with a discussion in Section 4.5.

4.1 Introduction

Survival analysis often involves evaluating the effects of longitudinally measured biomarkers on mortality. When longitudinal measures are sparsely collected, incomplete, or prone to measurement error, including them directly as a traditional time-varying covariate in a survival model may lead to biased regression estimates (Prentice (1982)). Alternatively, one could apply a two-stage method, where the first stage consists of modeling the longitudinal components via a mixed-effects model, and in the second stage, the modeled values or their summaries (e.g., first-order trends) are included in a survival model (Dafni and Tsiatis (1998), Tsiatis et al. (1995)). However, this approach fails to account for uncertainty in the estimated longitudinal summary measures. To overcome these issues, several joint longitudinal-survival models have been proposed (Prentice (1982); Bycott and Taylor (1998); Hanson et al. (2011b); Wang and Taylor (2001a); Faucett and Thomas (1996); Brown and Ibrahim (2003); Wulfsohn and Tsiatis (1997a); Song et al. (2002), and Law et al. (2002a)), where all these models account for uncertainty in longitudinal measures by modeling them simultaneously with the survival outcome. However, most existing joint models still rely on multiple restrictive parametric and semi-parametric assumptions and generally focus only on associating the first moment of the distribution of the longitudinal covariate with survival.

In this chapter we propose three flexible joint longitudinal-survival models that avoid simple distributional assumptions on longitudinal measures and allow for subject-specific baseline hazard in modeling a survival outcome. Our models are motivated by data on end-stage renal disease (ESRD) patients obtained from the United States Renal Data System (US-

RDS). Specifically, our interest lies in quantifying the association between the longitudinally measured serum albumin (a leading index of protein-energy malnutrition (PEM)) and time-to-death using a joint survival-longitudinal modeling approach.

Flexibility in our joint model is achieved in the longitudinal component via the use of a Gaussian process prior with a parameter that captures within-subject volatility in the longitudinally sampled albumin. The survival component of our proposed models quantifies the association between the longitudinally measured albumin and the risk of mortality using a Dirichlet process mixture of Weibull distributions. The clustering mechanism of the Dirichlet process provides a platform for borrowing information when estimating subject-specific baseline hazards in the survival component. Estimation for the longitudinal and survival parameters is carried out simultaneously via Bayesian parameter posterior sampling approach.

4.2 Methodology

In this section, we provide the details of our proposed joint models for a longitudinal covariate, $\boldsymbol{X}$, and a survival outcome, $\boldsymbol{Y}$. Throughout this section, we consider n independent subjects where l_i longitudinal measurements, X_{ij}, are obtained for subject i at time points $t_{ij},\ j = 1, \ldots, l_i$. Also, associated with each subject, there is an observed survival time, $Y_i \equiv \min\{T_i, C_i\}$ and event indicator $\delta_i \equiv 1_{[Y_i = E_i]}$, where T_i and C_i denote the true event and censoring time for subject i, respectively. Further, we make the common assumption that C_i is independent of T_i for all i, $i = 1, \ldots, n$.

4.2.1 The Joint Model

Being interested in estimating the effect of longitudinal measures on survival outcomes, in specifying the joint model likelihood, we took a similar approach as Brown and Ibrahim (2003), where we define the contribution of each subject to the joint model likelihood as the multiplication of the likelihood function of the longitudinal measures for that subject and her/his time-to-event likelihood that is conditioned on her/his longitudinal measures. Let $f_L^{(i)}$, $f_{S|L}^{(i)}$, and $f_{L,S}^{(i)}$ denote the longitudinal likelihood contribution, the conditional survival likelihood contribution, and the joint likelihood contribution for subject i. One can write the joint longitudinal-survival likelihood function as

$$f_{L,S} \;=\; \prod_{i=1}^{n} f_{L,S}^{(i)} = \prod_{i=1}^{n} \big(f_L^{(i)} \times f_{S|L}^{(i)}\big). \tag{4.1}$$

4.2.2 Longitudinal Component

We motivate the development of the Gaussian process model for the longitudinal biomarker by first considering the following simple linear model for estimating the trend in the biomarker for a single subject i with an $l_i \times 1$ vector of measure biomarkers of $\boldsymbol{X_i}$ which is of the form

$$\boldsymbol{X_i} = \begin{pmatrix} X_i(t_{i1}) \\ X_i(t_{i2}) \\ \vdots \\ X_i(t_{il_i}) \end{pmatrix},$$

where

$$\boldsymbol{X_i}|\beta_{i0}^{(L)} \sim N(\boldsymbol{\beta_{i0}^{(L)}}, \boldsymbol{\Sigma_i}).$$

with $\beta_{i0}^{(L)}$ as the subject-specific intercept, $\boldsymbol{\beta_{i0}^{(L)}}$ is vector of repeated $\beta_{i0}^{(L)}$ value that is of size $l_i \times 1$, and $\boldsymbol{\Sigma_i} = \sigma^2 I_{l_i \times l_i}$.

By adding a stochastic component that is indexed by time in the model, one can extend the model to capture non-linear patterns over time. Specifically, we consider a stochastic vector, $\boldsymbol{W}$, that is a realization from a Gaussian process prior, $W(t)$ with mean zero and covariance function $C(t, t')$. Thus for subject i, $\boldsymbol{W_i} \sim N_{l_i}(\boldsymbol{0}, \boldsymbol{C_{l_i \times l_i}})$, where $\boldsymbol{W_i} = (W_{t_{i1}}, \ldots, W_{t_{il_i}})'$ and the (j, j') element of $\boldsymbol{C_{l_i \times l_i}}$ is given by $C(t_{ij}, t_{ij'})$,$j, j' \in \{1, \ldots, l_i\}$. We characterize the covariance function, $\boldsymbol{C_{l_i \times l_i}}$, using the following squared exponential form

$$\boldsymbol{C_{l_i \times l_i}}(j, j') = {\kappa_i}^2 e^{-\rho^2 (t_{ij} - t_{ij'})^2}.$$

In this setting, the hyperparameter ρ^2 controls the correlation length, and κ^2 controls the height of oscillations (Banerjee et al. (2014)), and t_{ij} and $t_{ij'}$ are two different time points. For notational simplicity, we define $\boldsymbol{K_i} = e^{-\rho^2 (t_{ij} - t_{ij'})^2}$; $j, j' \in \{1, \ldots, l_i\}$, and re-write our longitudinal model as

$$\boldsymbol{X_i}|\beta_{i0}^{(L)}, {\kappa_i}^2, \rho^2, \sigma^2 \sim N(\boldsymbol{\beta_{i0}^{(L)}}, {\kappa_i}^2 \boldsymbol{K_i} + \sigma^2 I_{l_i \times l_i}),$$

where σ^2 is assumed to be common across all subjects. The correlation length parameter ρ^2 controls the maximum distance in time between two time-dependent measurements to be still correlated. This distance for GP models is often called the practical range. Diggle et al. (2007) defined the practical range for GP as the distance in time between two time-dependent measurements where the correlation between those two measurements is 0.05. With the squared exponential covariance function, that practical range distance is of the form $\sqrt{3/\rho^2}$. At a $\rho^2 = 0.1$, the practical range distance is 5.7 months which is a reasonable range for the real data on end-stage renal disease patients that was obtained from the USRDS. Hence, we fix ρ^2 to 0.1, where this value was obtained from the real data on end stage renal disease patients data. By defining our model in this way, subject-specific parameter κ_i^2 will have the role of capturing within-subject volatility of the longitudinal measures. In the context of the motivating USRDS example, κ_i^2 can be of primary scientific interest as it reflects the within-subject volatility (Figure 4.1) in serum albumin over time, which is hypothesized to be negatively correlated with longer survival time (Holsclaw et al, 2014).

We specify the longitudinal component of our joint model to have a likelihood of the form

$$\boldsymbol{X_i}|\boldsymbol{W_i}, \beta_{i0}^{(L)}, {\kappa_i}^2, \rho^2, \sigma^2 \quad \sim \quad N(\boldsymbol{\beta_{i0}^{(L)}} + \boldsymbol{W_i}, \sigma^2 I_{l_i \times l_i}), \tag{4.2}$$

where $\boldsymbol{X_i}$ is a vector of longitudinal measures on subject i, $\boldsymbol{W_i}$ is a Gaussian process stochastic vector, $\beta_{i0}^{(L)}$ is subject specific intercept for subject i, ${\kappa_i}^2$ is a subject-specific measure of volatility for subject i, ρ^2 is a fixed correlation length, σ^2 is a measurement error that is shared across all subjects, and finally $I_{l_i \times l_i}$ represents the identity matrix of size l_i where l_i is the number of longitudinal measures on subject i. The Gaussian process stochastic vector

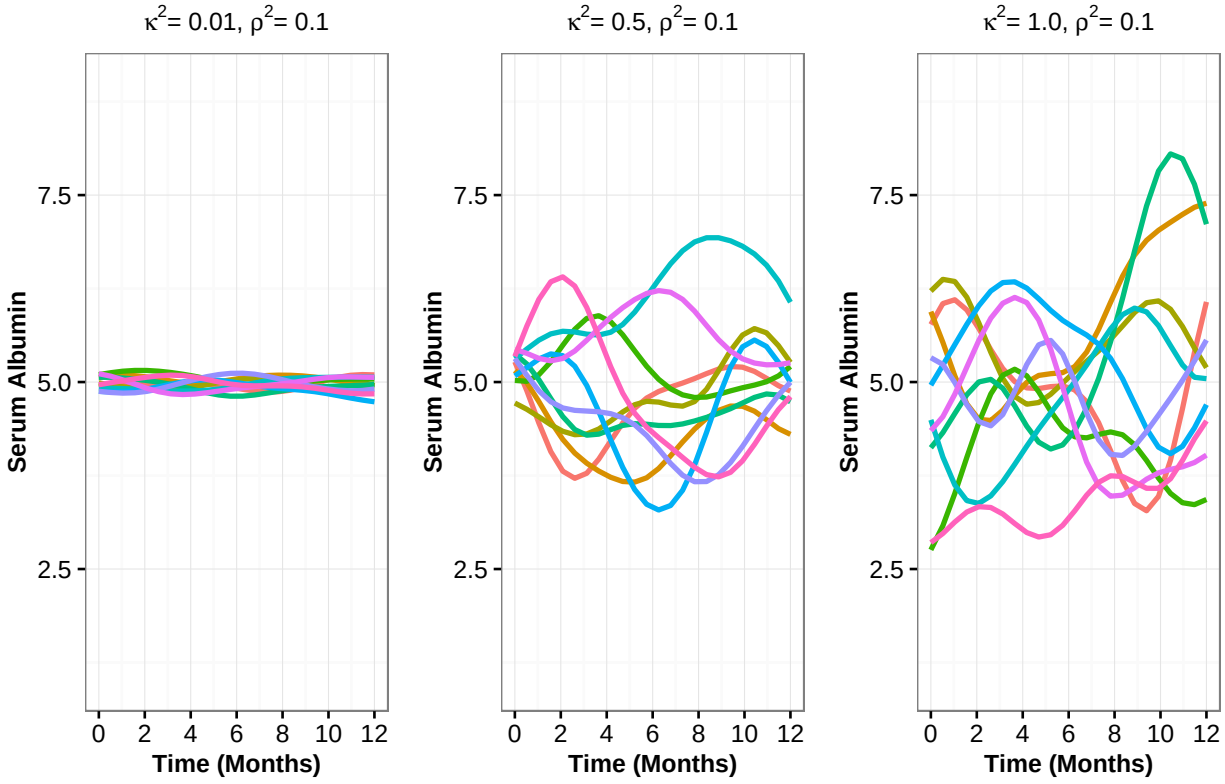


Figure 4.1: With a fixed correlation length parameter ρ^2, κ^2 parameter captures volatility in Gaussian process models with the squared exponential covariance function. In each plot, ten random realizations of the Gaussian process were selected, where the plot to the left has a κ^2 parameter of 0.01, the plot in the middle has a κ^2 value of 0.5, and the plot to the right has a κ^2 value of 1.0. In all plots, correlation length ρ^2 is fixed to 0.1 .

$\boldsymbol{W_i}$ is distributed Gaussian process as

$$\boldsymbol{W_i}|{\kappa_i}^2, \boldsymbol{t_i} \overset{ind}{\sim} GP_{m_i}(\vec{0}, {\kappa_i}^2 \boldsymbol{K_i}), \tag{4.3}$$

where $\boldsymbol{t_i}$ is a vector of the time points at which longitudinal measures on subject i were collected and $\boldsymbol{K_i} = e^{-\rho^2(t_{ij}-t_{ij'})^2}$, with t_{ij} and $t_{ij'}$ are the j^{th} and j'^{th} element of the time vector $\boldsymbol{t_i}$. We assume a Normal prior on the subject-specific random intercepts $\beta_{i0}^{(L)}$ that is of the form

$$\beta_{i0}^{(L)} \overset{i.i.d}{\sim} N(\mu_{\beta_0^{(L)}}, \sigma^2_{\beta_0^L}), \tag{4.4}$$

where $\mu_{\beta_0^{(L)}}$ and $\sigma^2_{\beta_0^L}$ are prior mean and prior variance respectively. $\boldsymbol{K_i}$, where $i \in \{1, \ldots, n\}$ with n as the number of subjects in the study, are assumed to have a log-Normal prior with the prior mean μ_{κ^2} and the prior variance σ_{κ^2} that is of the form

$${\kappa_i}^2 \overset{i.i.d}{\sim} log - Normal(\mu_{\kappa^2}, \sigma_{\kappa^2}). \tag{4.5}$$

The correlation length ρ^2 is assumed to be fixed and known in our model. Finally, the measurement error σ^2 is assumed to have a log-Normal prior of the form

$$\sigma^2 \sim log - Normal(\mu_{\sigma^2}, \sigma_{\sigma^2}), \tag{4.6}$$

where μ_{σ^2} and σ_{σ^2} are the prior mean and the prior variance respectively.

4.2.3 Survival Component

In order to quantify the association between a longitudinal biomarker and a time-to-event outcome, we define our survival component by using a multiplicative hazard model with the general form of

$$\lambda(T_i|\boldsymbol{Z_i}^{(s)}, \boldsymbol{Z_i}^{(L)}) = \lambda_0(T_i)exp\{\boldsymbol{\zeta}^{(s)}\boldsymbol{Z_i}^{(s)} + \boldsymbol{\zeta}^{(L)}\boldsymbol{Z_i}^{(L)}(\boldsymbol{t})\}, \tag{4.7}$$

where $\boldsymbol{Z_i}^{(s)}$ is a vector of baseline covariates, $\boldsymbol{Z_i}^{(L)}$ is a vector of longitudinal covariates from the longitudinal component of the model, $\lambda_0(T_i)$ denotes the baseline hazard function, and $\boldsymbol{\zeta}^{(S)}$ and $\boldsymbol{\zeta}^{(L)}$ are regression coefficients for the baseline survival covariates and the longitudinal covariates, respectively.

We consider a Weibull distribution for the survival component to allow for log-linear changes in the baseline hazard function over time. Thus we assume

$$T_i \sim Weibull(\tau, \lambda_i), \tag{4.8}$$

where T_i is the survival time, τ is the shape parameter of the Weibull distribution, and $exp\{\lambda_i\}$ is the the scale parameter of the Weibull distribution. One can write the density function for the Weibull distribution above for the random variable T_i as

$$f(T_i|\tau, \lambda_i) = \tau {T_i}^{\tau-1} exp\big(\lambda_i - exp(\lambda_i){T_i}^{\tau}\big). \tag{4.9}$$

In this case, the Weibull distribution is available in closed form providing greater computational efficiency. Under this parameterization, covariates can be incorporated into the model by defining $\lambda_i = \boldsymbol{\zeta^{(s)}} \boldsymbol{Z_i}^{(s)} + \boldsymbol{\zeta^{(L)}} \boldsymbol{Z_i}^{(L)}$. In particular, we specify our model as

$$T_i|\tau, \boldsymbol{\zeta^{(s)}}, \boldsymbol{\zeta^{(L)}}, \boldsymbol{Z_i}^{(s)}, \boldsymbol{Z_i}^{(L)} \sim Weibull(\tau, \lambda_i = \beta_{i0}^{(s)} + \boldsymbol{\zeta^{(s)}} \boldsymbol{Z_i}^{(s)} + \boldsymbol{\zeta^{(L)}} \boldsymbol{Z_i}^{(L)}), \quad (4.10)$$

where τ is a common shape parameter shared across all subjects. $\beta_{i0}^{(s)}$ is a subject specific coefficient in the model which allows for a subject-specific baseline hazard. $\boldsymbol{Z_i}^{(s)}$ and $\boldsymbol{\zeta^{(s)}}$ are baseline covariates and their corresponding coefficients, respectively. Finally, $\boldsymbol{Z_i}^{(L)}$ and $\boldsymbol{\zeta^{(L)}}$ are coefficients linking the longitudinal parameters of interest to the hazard for mortality.

In order to avoid an explicit distributional assumption for the survival times, we specify our survival model as an infinite mixture of Weibull distributions that is mixed on the $\beta_{i0}^{(s)}$ parameter. In particular, we use the Dirichlet process mixture of Weibull distributions that is defined as

$$\begin{aligned} \beta_{i0}^{(s)}|\mu_i, \sigma^2_{\beta_0^{(s)}} &\sim N(\mu_i, \sigma^2_{\beta_0^{(s)}}), && (4.11) \\ \mu_i|G &\sim G, && (4.12) \\ G &\sim DP\big(\alpha^{(S)}, G_0\big), && (4.13) \end{aligned}$$

where $\sigma^2_{\beta_0^{(s)}}$ is a fixed parameter, μ_i is a subject-specific mean parameter from a distribution G with a DP prior, $\alpha^{(S)}$ is the concentration parameter of the DP and G_0 is the base distribution. By using the Dirichlet process prior on the distribution of $\beta_{i0}^{(s)}$, we allow patients with similar baseline hazards to cluster together which subsequently provides a stronger likelihood to estimate the baseline hazards. For other covariates in the model, we assume a

multivariate normal prior of the form

$$\left(\boldsymbol{\zeta}^{(\boldsymbol{s})}, \boldsymbol{\zeta}^{(\boldsymbol{L})}\right) \sim MVN(\mathbf{0}, {\sigma_0}^2 I),$$

where ${\sigma_0}^2$ is a prior variance and I is an identity matrix.

The shared scale parameter τ is considered to have a Log-Normal prior of the form

$$\tau \sim Log - Normal(a_\tau, b_\tau),$$

where a_τ and b_τ are fixed prior location and prior scale parameters, respectively.

Finally, we assume that information about the concentration parameter of the Dirichlet process can be specified with the prior

$$\alpha^{(S)} \sim \Gamma(a_\alpha^{(S)}, b_\alpha^{(S)}),$$

where $a_\alpha^{(S)}$ and $b_\alpha^{(S)}$ are fixed prior shape and prior scale parameters, respectively.

4.2.4 Linking Summary Measures of the Biomarker to Survival Times

The proposed modeling framework easily allows for associating multiple summaries of the longitudinal biomarker with the time-to-event outcome. Here we consider three models that incorporate various summary measures of the longitudinal trajectory that are easily and

flexibly estimated using the GP model presented in Section 4.2.2:

Model I: directly modeling longitudinal outcome at each event time t as a covariate in the survival model:

$$Z_i{}^{(L)} = X^i(t)$$

Model II: modeling both the value of the longitudinal covariate and also the average rate at which the biomarker changes for each subject. We define this average rate as a weighted area under the derivative curve of the biomarker trajectory

$$\boldsymbol{Z_i}^{(\boldsymbol{L})} = \left(X_i(t), {X'_{AUC}}^{\tau_0-\tau_1}\right)$$

$$\text{where, } {X'_{AUC}}^{\tau_0-\tau_1} = \int_{\tau_0}^{\tau_1} Q(u)X'(u)du$$

where ${X'_{AUC}}^{\tau_0-\tau_1}$ is a time-dependent covariate that is a weighted average of the derivative of the biomarker trajectory, that is denoted by $X'(u)$ from τ_0 to τ_1 where τ_1 is the time of death for each subject. This average area under the derivative curve can be a weighted average with weights $Q(u)$.

Model III: modeling summary measures of the longitudinal trajectory. Motivated by the scientific question of interest, in this paper we consider **random intercepts** and **subject-specific volatility** as summary measures of interest:

$$\boldsymbol{Z_i}^{(\boldsymbol{L})} = \left(\beta_{0i}^{(L)}, \kappa_i^2\right)$$

Below, we shall explain the three models above in more detail.

Model I: a Survival model with Longitudinal Biomarker at event time as a covariate

This model quantifies the association between a longitudinal biomarker of interest and the time-to-event outcome by directly adjusting for the biomarker measured values in the survival component. While usually biomarkers are measured on a discrete lab-visit basis, the event of interest happens on a continuous basis. While common frequentist models use the so-called last-observation-carried forward (LOCF) technique where the biomarker value at each even time is assumed to be the same as the last measured value for that biomarker, our joint flexible longitudinal-survival model provides a proper imputation method for the biomarker values at each individual's event time. In particular, in each iteration of the MCMC, given the sampled parameters for each individual and by using the flexible Gaussian process prior, there exists posterior trajectories of biomarker for that individual. Our method, then, considers the posterior mean of those trajectories as the proposed trajectory for that individual's biomarker values over time at that iteration. The trajectory, then, can be used to impute time-dependent biomarker covariate value inside the survival component. To be more specific, consider the longitudinal biomarker $\boldsymbol{X_i}$ of the form

$$\boldsymbol{X_i}|\beta_{i0}^{(L)}, {\kappa_i}^2, \rho^2, \sigma^2 \quad \sim \quad N(\boldsymbol{\beta_{i0}^{(L)}}, {\kappa_i}^2 \boldsymbol{K_i} + \sigma^2 I_{l_i \times \boldsymbol{l_i}}),$$

where $\beta_{i0}^{(L)}$ is subject-specific random intercept for subject i, $\boldsymbol{\beta_{i0}^{(L)}}$ is a vector of repeated subject-specific intercept $\beta_{i0}^{(L)}$ that is of size $l_i \times 1$, ${\kappa_i}^2$ is subject-specific measure of volatility in the longitudinal biomarker for individual i, ρ^2 is a fixed measure of correlation length, σ^2 is the measurement error shared across all subjects, $\boldsymbol{K_i}$ is a an $l_i \times l_i$ matrix with it's jj' element as $\boldsymbol{K_{i}}_{jj'} = e^{-\rho^2(t_{ij}-t_{ij'})^2}$ where l_i is the number of longitudinal biomarker measures on subject i, and $I_{l_i \times l_i}$ is the identity matrix.

For a new time-point t^*, predicted albumin biomarker for individual i is X^* and can be written as

$$X^*|\boldsymbol{X_i}, \boldsymbol{t}, t^* \quad \sim \quad N(\mu^*, \Sigma^*),$$

where the conditional posterior mean μ^* is

$$\mu^* = \beta_{i0}^L + K(t^*, \boldsymbol{t}){K_X}^{-1}(\boldsymbol{X_i} - \beta_{i0}^{(L)}), \tag{4.14}$$

and the conditional posterior variance Σ^* is

$$\Sigma^* = K(t^*, t^*) - K(t^*, \boldsymbol{t}){K_X}^{-1}K(t^*, \boldsymbol{t})', \tag{4.15}$$

where $K(t^*, \boldsymbol{t})$ is defined as

$$K(t^*, \boldsymbol{t}) = {\kappa_i}^2 e^{-\rho^2(t^* - \boldsymbol{t})^2}, \tag{4.16}$$

and K_X^{-1} is defined as

$$K_X^{-1} = (K(\boldsymbol{t}, \boldsymbol{t}) + \sigma^2 I_{l_i \times l_i})^{-1}. \tag{4.17}$$

In order to relate the biomarker value at each time point t to the risk of the event of interest

at that time point, "death", we form the survival component of the model as

$$T_i|\tau, \boldsymbol{\zeta^{(s)}}, \zeta_{X_i} \sim Weibull(\tau, \lambda_i = \beta_{i0}^{(s)} + \boldsymbol{\zeta^{(s)}}\boldsymbol{Z_i}^{(s)} + \zeta_{X_i}X_i(t)),$$

where T_i is the survival time, τ is the shape parameter of the Weibull distribution, $\boldsymbol{\zeta^{(s)}}$ is a vector of coefficients relating baseline survival covariates to the risk of the occurrence of the event of interest, ζ_{X_i} is the coefficient that relates the biomarker value at time t and the risk of "death" at that time point, λ_i is the log of the scale parameter in the Weibull distribution, $\beta_{i0}^{(s)}$ is the subject-specific baseline hazard for subject i, $\boldsymbol{Z_i}^{(s)}$ is a vector of survival coefficients, and $X_i(t)$ is the biomarker value at time t.

Model II: A Survival model with covariates of Biomarker Value and the Derivative of its Trajectory at event Time

In order to get more precision in quantifying the association between the biomarker value at time t and the risk of death, we can extend our proposed model-I by including a measure of the average slope of the biomarker over time. In particular, we define this average slope from time τ_0 to τ_1 as the area under the derivative of the trajectory curve of the biomarker from τ_0 to τ_1. More generally, this area under the curve can be a weighted sum where weights are chosen according to the scientific question of interest. One may hypothesize that the area under the derivative curve that are closer to the event time should be weighted higher compared to the areas that are farther away from time point t. In general, we define a weighted area under the derivative curve of the form

$${X'_{AUC}}^{\tau_0-\tau_1} = \int_{\tau_0}^{\tau_1} Q(u)X'(u)du,$$

where τ_0 and τ_1 are arbitrary time points chosen according to the scientific question of interest, $Q(t)$ is a weight, and $X'(t)$ represents the derivative of the biomarker over time. In particular, we consider two weighted approaches, where one assumes an equal weight of the form

$$Q(t) = \frac{1}{\tau_1 - \tau_0},$$

and another weight of the form

$$Q(t) = \begin{cases} 1, & \text{if } t = T_i \\ 0, & \text{otherwise.} \end{cases}$$

Under the first weighting scheme, X'_{AUC} will be the area under the derivative with equal weights, whereas the second weighting scheme leads to the pointwise derivative value at the the event time. Under this model, the survival component of our joint model will now include two longitudinal covariates, one the biomarker value $X_i(t)$, and another the average derivative of the biomarker trajectory, X'_{AUC}.

The derivative of the Gaussian process is still a Gaussian process with the same hyper parameters ρ^2 and κ^2. Therefore, using the same idea of modeling the trajectory of the biomarker, we can also model the derivative of that trajectory. In our model-I, we proposed using the posterior mean of all plausible biomarker trajectories as the proposed trajectory for each subject in order to impute biomarker values at any time point t inside the survival component of the model. Similarly, we propose using the posterior mean of all plausible derivative trajectories for each subject in order to compute the average derivative up until time t. Given the fact that differentiation is a linear operation, one can easily compute the posterior mean of the derivative curve by simply switching the order of the differentiation

and the expectation as

$$\begin{aligned} E(X_i'(t)) &= E(\frac{\partial X_i(t)}{\partial t}) \\ &= \frac{\partial\big(E(X_i(t))\big)}{\partial t}. \end{aligned}$$

Hence, by using Formula (4.14) and by taking the derivative of the posterior mean trajectory of the biomarker with respect to time t^*, the posterior mean of the derivative of the biomarker trajectory is of the form

$$\frac{\partial\big(E(X_i(t^*))\big)}{\partial t^*} = -2\rho^2(t^* - \boldsymbol{t})'\big(K(t^*, \boldsymbol{t}){K_X}^{-1}(\boldsymbol{X_i} - \beta_{i0}^{(L)})\big), \tag{4.18}$$

where $E\big(X_i(t^*)\big)$ denotes the posterior mean of the biomarker trajectory as a function of time t^*, ρ^2 is the correlation length, t^* is the time-point at which we desire to impute the biomarker value and the average derivative of the biomarker trajectory, $\beta_{i0}^{(L)}$ is subject-specific random intercept, $K(t^*, \boldsymbol{t})$ is defined as

$$K(t^*, \boldsymbol{t}) = {\kappa_i}^2 e^{-\rho^2(t^* - \boldsymbol{t})^2}, \tag{4.19}$$

and K_X^{-1} is defined as

$$K_X^{-1} = (K(\boldsymbol{t}, \boldsymbol{t}) + \sigma^2 I_{l_i \times l_i})^{-1}. \tag{4.20}$$

Given the the biomarker value $X_i(t)$ and the average derivative value $X'_{AUC,i}(t)$, the survival

component of our proposed joint model is of the from

$$T_i|\tau, \boldsymbol{\zeta^{(s)}}, \zeta_{X_i}, \zeta_{X_i'} \sim Weibull(\tau, \lambda_i = \beta_{i0}^{(s)} + \boldsymbol{\zeta^{(s)}}\boldsymbol{Z_i}^{(s)} + \zeta_{X_i} X_i(t) + \zeta_{X_i'} X'_{AUC,i}(t)),$$

where T_i is the survival time, τ is the shape parameter of the Weibull distribution, $\boldsymbol{\zeta^{(s)}}$ is a vector of coefficients relating baseline survival covariates to the risk of the occurrence of the event of interest, ζ_{X_i} is the coefficient that relates the biomarker value at time t and the risk of "death" at that time point, $\zeta_{X_i'}$ is the coefficient that relates the average derivative of the biomarker trajectory up until t and the risk of "death" at that time point, λ_i is the log of the scale parameter in the Weibull distribution, $\beta_{i0}^{(s)}$ is the subject-specific baseline hazard for subject i, and $\boldsymbol{Z_i}^{(s)}$ is a vector of survival coefficients.

Model III: A Survival Model with Summary Measures of the Longitudinal Curve as Covariates

One may choose to characterize longitudinal trajectories with summary measures instead of using the actual biomarker value. In specific, the longitudinal model we proposed provides a natural parameter for describing the within-subject volatility. Given the nature of our proposed longitudinal model, one can summarize the longitudinal trajectory of biomarker by using $\beta_{0i}^{(L)}$ as a measure of subject-specific intercept of longitudinal biomarker as well as κ_i^2 as a measure of volatility of those trajectories. The survival component of the model is then of the form

$$T_i|\tau, \beta_{i0}^{(s)}, \boldsymbol{\zeta^{(s)}}, \zeta_{\beta_{0i}}{}^{(L)}, \zeta_{\kappa_i{}^2}{}^{(L)} \sim Weibull(\tau, \lambda_i = \beta_{i0}^{(s)} + \boldsymbol{\zeta^{(s)}}\boldsymbol{Z_i}^{(s)} + \zeta_{\beta_{0i}}{}^{(L)}\beta_{0i}{}^{(L)} + \zeta_{\kappa_i{}^2}{}^{(L)}\kappa_i{}^2),$$

where T_i is the survival time, τ is the shape parameter of the Weibull distribution, $\boldsymbol{\zeta^{(s)}}$ is a vector of coefficients relating baseline survival covariates to the risk of the occurrence of the event of interest, ${\zeta_{\beta_{0i}}}^{(L)}$ is the coefficient that relates the subject specific random intercept ${\beta_{0i}}^{(L)}$ and the risk of "death", ${\zeta_{\kappa_i}}^{2(L)}$ is the coefficient that relates the subject-specific measure of volatility of the biomarker measure and the risk of "death", λ_i is the log of the scale parameter in the Weibull distribution, $\beta_{i0}^{(s)}$ is the subject-specific baseline hazard for subject i, and $\boldsymbol{Z_i}^{(s)}$ is a vector of survival coefficients.

4.2.5 The Posterior Distribution

Consider the joint longitudinal-survival likelihood function, $f_{\boldsymbol{L},\boldsymbol{S}}$, introduced in equation 4.1. Let $\boldsymbol{\omega}$ be a vector of all model parameters with the joint prior distribution $\pi(\boldsymbol{\omega})$. The posterior distribution of the parameter vector $\boldsymbol{\omega}$ can be written as

$$\pi(\boldsymbol{\omega}|\boldsymbol{X},\boldsymbol{Y}) \quad \propto \quad f_{L,S} \times \pi(\boldsymbol{\omega}), \tag{4.21}$$

where $\boldsymbol{X}$ and $\boldsymbol{Y}$ denote longitudinal and time-to-event data respectively, and $f_{L,S}$ is the joint model likelihood function (equation 4.1).

The posterior distribution of the parameters in our proposed joint model is not available in closed form. Hence, samples from the posterior distribution of the model parameters are obtained via Markov Chain Monte Carlo (MCMC) methods. We use a hybrid sampling technique where in each iteration of the MCMC, we first sample subject-specific frailty terms in the survival model using Neal's algorithm 8. Then given the sampled frailty terms, we use the Hamiltonian Monte Carlo (Neal (2011)) to draw samples from the posterior distribution. Prior distributions on parameters of the joint model were explained in details in Sections

4.2.2 and 4.2.3, and we assume independence among model parameters in the prior (ie. $\pi(\boldsymbol{\omega})$ is the product of the prior components specified previously). We provide further detail on less standard techniques for sampling from the posterior distribution when using a GP prior and issues in evaluating the survival portion of the likelihood function when time-varying covariates are incorporated into the model.

Evaluation of the Longitudinal Likelihood

The longitudinal component of our model uses the Gaussian process technique. Gaussian process models are typically computationally challenging to fit because in each iteration of the MCMC the evaluation of the log-posterior probability becomes computationally challenging as the number of measurements increases. In particular, consider our proposed longitudinal model introduced in Section 4.2.2 where

$$\begin{aligned} \boldsymbol{X_i}|\boldsymbol{W_i}, \beta_{i0}^{(L)}, {\kappa_i}^2, \rho^2, \sigma^2 &\sim N(\boldsymbol{\beta_{i0}^{(L)}} + \boldsymbol{W_i}, \sigma^2 I_{l_i \times l_i}), \\ \boldsymbol{W_i}|{\kappa_i}^2, \boldsymbol{t_i} &\sim GP_{m_i}(\vec{0}, {\kappa_i}^2 \boldsymbol{K_i}), \end{aligned}$$

with $\boldsymbol{X_i}$ denoting a vector of longitudinal measures on subject i, $\boldsymbol{W}_i$ a Gaussian process stochastic vector, $\beta_{i0}^{(L)}$ a subject specific intercept for subject i, ${\kappa_i}^2$ a subject-specific measure of volatility for subject i, ρ^2 a fixed correlation length, σ^2 a measurement error that is shared across all subjects, $I_{l_i \times l_i}$ denoting the identity matrix of size l_i with l_i the number of longitudinal measures on subject i, $\boldsymbol{t_i}$ a vector of the time points at which longitudinal measures on subject i were collected, and $\boldsymbol{K_i} = e^{-\rho^2(t_{ij} - t_{ij'})^2}$, where t_{ij} and $t_{ij'}$ are the j^{th} and j'^{th} element of the time vector $\boldsymbol{t_i}$.

In order to sample from the posterior distribution of ${\kappa_i}^2$ and σ^2 parameters, one can consider

a marginal distribution of the following form

$$\boldsymbol{X_i}|\beta_{i0}^{(L)}, {\kappa_i}^2, \rho^2, \sigma^2 \quad \sim \quad N(\beta_{i0}^{(L)}, {\kappa_i}^2\boldsymbol{K_i} + \sigma^2 I_{l_i \times l_i}) \tag{4.22}$$

The marginal distribution above has log-density of the form

$$\begin{aligned} log(f(\boldsymbol{X_i}|\beta_{i0}^{(L)}, {\kappa_i}^2, \rho^2, \sigma^2)) &= \text{constant} \\ &- \frac{1}{2} log|{\kappa_i}^2\boldsymbol{K_i} + \sigma^2 I_{l_i \times l_i}| \\ &- \frac{1}{2}(\boldsymbol{X_i} - \beta_{i0}^{(L)})^T({\kappa_i}^2\boldsymbol{K_i} + \sigma^2 I_{l_i \times l_i})^{-1}(\boldsymbol{X_i} - \beta_{i0}^{(L)}), \end{aligned} \tag{4.23}$$

that is the log contribution of subject i to the longitudinal likelihood (i.e. $log(f_L^{(i)})$).

sampling from the posterior distribution of ${\kappa_i}^2$ and σ^2 requires evaluation of the log-density in equation (4.23) that involves evaluation of the determinant and the computation of the inverse of the covariance matrix at each iteration of the MCMC. This process requires $O({l_i}^2)$ memory space and a computation time of $O({l_i}^3)$ per subject, with l_i as the number of within subject measurements.

In our model setting, we defined $\boldsymbol{K_i} = e^{-\rho^2(t_{ij} - t_{ij'})}$ with a fixed ρ^2 parameter. This means K_i can be pre-computed before starting posterior sampling using MCMC. Furthermore, we propose using the eigenvalue decomposition technique for a faster log-posterior probability computation. Our proposed method was motivated by Flaxman et al. (2015) and is as follows.

Consider the covariance matrix ${\kappa_i}^2\boldsymbol{K_i} + \sigma^2 I_{l_i \times l_i}$ in the marginal log-density in equation (4.23). Our goal is now to propose a method that makes computation of the inverse and the

determinant of this covariance function as efficient as possible. As shown earlier, $\boldsymbol{K_i}$ can be pre-computed before starting the MCMC process as it does not involve any parameter. Consider the eigenvalue decomposition of $\boldsymbol{K_i} = Q\Lambda Q^T$, where Λ is a diagonal matrix with the eigenvalues of $\boldsymbol{K_i}$ as the diagonal elements, and Q is the corresponding matrix of eigenvectors. ${\kappa_i}^2$ is a scalar parameter that is sampled in each iteration of the MCMC. Multiplication of ${\kappa_i}^2$ times the matrix $\boldsymbol{K_i}$ implies the eigenvalues of this matrix will be ${\kappa_i}^2$ times bigger where the eigenvectors remain the same. Hence, we can conclude that the eigenvalue decomposition of the matrix ${\kappa_i}^2\boldsymbol{K_i}$ is of the form ${\kappa_i}^2\boldsymbol{K_i} = Q({\kappa_i}^2\Lambda)Q^T$, where Q and Λ are elements of the eigenvalue decomposition of the pre-computed matrix $\boldsymbol{K_i}$. Given the pre-computed eigenvalue decomposition of the matrix $\boldsymbol{K_i}$, at each iteration of the MCMC, the determinant of the covariance function of the marginal log-density in equation (4.23) can be computed as

$$\begin{aligned} log|{\kappa_i}^2\boldsymbol{K_i} + \sigma^2 I_{l_i\times l_i}| &= log|Q({\kappa_i}^2\Lambda)Q^T + \sigma^2 I_{l_i\times l_i}| \\ &= log|Q({\kappa_i}^2\Lambda + \sigma^2 I_{l_i\times l_i})Q^T| \\ &= log\big(\prod_{k=1}^{l_i}({\kappa_i}^2\lambda_{ik} + \sigma^2)\big) \\ &= \sum_{k=1}^{l_i} log\big({\kappa_i}^2\lambda_{ik} + \sigma^2\big). \end{aligned} \tag{4.24}$$

In equation (4.24), λ_{ik}'s are pre-computed eigenvalues of the matrix $\boldsymbol{K_i}$ whereas κ_i and σ^2 are parameters sampled at each iteration of the MCMC.

Similarly and by using the same trick, we can compute the term $(\boldsymbol{X_i} - \beta_{i0}^{(L)})^T({\kappa_i}^2\boldsymbol{K_i} +$

$\sigma^2 I_{l_i \times l_i})(\boldsymbol{X_i} - \beta_{i0}^{(L)})$ in a more computationally efficient as

$$\begin{aligned}(\boldsymbol{X_i} - \beta_{i0}^{(L)})^T(\kappa_i{}^2\boldsymbol{K_i} + \sigma^2 I_{l_i \times l_i})^{-1}(\boldsymbol{X_i} - \beta_{i0}^{(L)}) &= (\boldsymbol{X_i} - \beta_{i0}^{(L)})^T\big(Q(\kappa_i{}^2\Lambda)Q^T + \sigma^2 I_{l_i \times l_i}\big)^{-1}(\boldsymbol{X_i} - \beta_{i0}^{(L)}) \\ &= (\boldsymbol{X_i} - \beta_{i0}^{(L)})^T\big(Q(\kappa_i{}^2\Lambda + \sigma^2 I_{l_i \times l_i})Q^T\big)^{-1}(\boldsymbol{X_i} - \beta_{i0}^{(L)}) \\ &= (\boldsymbol{X_i} - \beta_{i0}^{(L)})^T\big(Q(\kappa_i{}^2\Lambda + \sigma^2 I_{l_i \times l_i})^{-1}Q^T\big)(\boldsymbol{X_i} - \beta_{i0}^{(L)}).\end{aligned} \tag{4.25}$$

In equation (4.25), $\boldsymbol{X_i}$ is the data matrix and is fixed, Q and Λ are pre-computed eigenvector and diagonal eigenvalue matrices corresponding to the eigenvalue decomposition of the matrix $\boldsymbol{K_i}$. Finally, by utilizing an eigenvalue decomposition, instead of evaluating the term $(\kappa_i{}^2\boldsymbol{K_i} + \sigma^2 I_{l_i \times l_i})^{-1}$, one can simply evaluate $\big(Q(\kappa_i{}^2\Lambda + \sigma^2 I_{l_i \times l_i})^{-1}Q^T\big)$, where the term $(\kappa_i{}^2\Lambda + \sigma^2 I_{l_i \times l_i})^{-1}$ in the middle is simply the inverse of a diagonal matrix.

Evaluation of the Survival Likelihood

Here we consider evaluation of the survival component of the decomposed joint likelihood. Consider the survival time for subject i that is denoted by t_i and is distributed according to a Weibull distribution with shape parameter τ and scale parameter $exp(\lambda_i)$, where $\lambda_i = \boldsymbol{\zeta^{(S)}}\boldsymbol{Z}_i^{(S)} + \boldsymbol{\zeta^{(L)}}\boldsymbol{Z}_i^{(L)}(t)$, where $\boldsymbol{Z}_i^{(S)}$ and $\boldsymbol{Z}_i^{(L)}(t)$ are vectors of covariates for subject i, with potentially time-varying covariates, corresponding to the survival and the longitudinal covariates respectively, and $\boldsymbol{\zeta^{(S)}}$ and $\boldsymbol{\zeta^{(L)}}$ are vectors of survival and longitudinal coefficients respectively. One can write the hazard function $h_i(t)$ as

$$h_i(t) = \tau t^{\tau-1} exp(\lambda_i - exp(\lambda_i)t^\tau). \tag{4.26}$$

The survival function $S_i(t)$ can be written as

$$S_i(t) = exp\{-\int_0^t h_i(w)dw\}.$$

Consider survival data on n subjects, some of whom may have been censored. Let event indicator δ_i that is 1 if the event is observed, and 0 otherwise. The survival likelihood contribution of subject i can be written in terms of the the hazard function $h_i(t)$ and the survival function $S_i(t)$ as

$$\begin{aligned} f_{S|L}^{(i)} &= h_i(t_i)^{\delta_i} S_i(t_i) \\ &= h_i(t_i)^{\delta_i} e^{-\int_0^{t_i} h_i(w)dw}. \end{aligned}$$

The overall survival log-likelihood can be written as

$$\begin{aligned} log(L) &= \sum_{i=1}^{n} log(f_{S|L}^{(i)}) \\ &= \sum_{i=1}^{n} \big(\delta_i log(h_i(t_i)) - \int_0^{t_i} h_i(w)dw\big). \end{aligned}$$

The hazard function in the equation (4.26) includes some time-varying covariates which often makes the integral of the hazard function non-tractable. In this case, one can estimate the integral using the rectangular integration as follows:

Integration of Survival Hazard with Time-Varying Covariates

1. Set a fixed number of rectangles m and set $A = 0$
2. Divide $(0, t_i)$ interval into m equal pieces each of length $L = t_i/m$

for $i \in \{1, \ldots, m\}$ **do**
  $t_{mid} \leftarrow L/2 + (i-1) * L$
  $A_{temp} \leftarrow L * h_i(t_{mid})$
  $A \leftarrow A + A_{temp}$
end for

4.3 Simulation Study

In this section, we evaluate our proposed models using a simulation study. We simulated 200 datasets that resembled the real data on end stage renal disease patients that was obtained from the United States Renal Data System (USRDS). To this end, we first simulated longitudinal trajectories with κ^2's which are sampled from a uniform distribution from 0 to 1. We fixed $\rho^2 = 0.1$ for all subjects. The subject-specific intercepts for albumin trajectories were randomly sampled from the Normal distribution $N(\mu = 5.0, \sigma^2 = 0.5)$. We simulate 9 to 12 longitudinal albumin values per subject. Using the simulated albumin trajectories, we generated survival times from the Weibull distribution in equation (5.17) that is of the following form for each of the proposed models

- **Model I**:

$$T_i|\tau, \beta_1, X_i(t) \sim Weibull(\tau, \lambda_i = \beta_{i0}^{(s)} + \beta_1 X_i(t)), \tag{4.27}$$

- **Model II**:

$$T_i|\tau, \beta_1, X_i(t) \sim Weibull(\tau, \lambda_i = \beta_{i0}^{(s)} + \beta_1 X_i(t) + \beta_2 X'_{AUC,i}(t)), \tag{4.28}$$

- **Model III**:

$$T_i|\tau, \beta_1, X_i(t) \sim Weibull(\tau, \lambda_i = \beta_{i0}^{(s)} + \beta_1 Gender + \beta_2 {\beta_0}^{(L)} + \beta_3 {\kappa_i}^2). \tag{4.29}$$

The true values of the coefficients are set as follows

- model I: $\beta_1 = -0.5$,

- model II: $\beta_1 = 0.5$,
- model III: $\beta_1 = 0.5$,

where in all simulations, $\beta_{i0}^{(s)}$ are simulated from a mixture of two Normal distributions of the form $\theta_i N(\mu = -1.5, \sigma = 1) + (1 - \theta_i) N(\mu = 1.5, \sigma = 1)$, where θ_i is distributed Bernoulli with parameter $p = 0.5$.

Finally, the censoring times were sampled from a uniform distribution and independently from the simulated event times with an overall censoring rate of 20%.

All results are from 200 simulated datasets of size $n = 300$ subjects each. For each dataset, we fit our proposed joint models with 10,000 draws where the first 5,000 considered as a burn-in period. Relatively diffuse priors were considered for all parameters. Details of the priors used in the simulations ass well as the results are as follow.

4.3.1 Model I - Simulation Results

In order to compare our proposed joint longitudinal-survival model that is capable of flexibly modeling longitudinal trajectories with simpler models with explicit functional assumptions on the longitudinal trajectories, we simulated longitudinal data once from quadratic polynomial longitudinal trajectory curves and another time from random non-linear curves. We then fit our joint model with a Gaussian Process longitudinal component as well as a joint model with the explicit assumption that the longitudinal trajectories are from a quadratic polynomial curve. As a comparison model, we also fit a two-stage Cox model where in stage one longitudinal data are modeled using our proposed Gaussian process longitudinal model and in the second stage, given the posterior mean parameters from the longitudinal fit, a Cox proportional hazard will fit the survival data.

In particular, we generate synthetic longitudinal and survival data on 300 subjects, each with 9 to 12 within subject longitudinal albumin measures. Under the scenario where the longitudinal data are generated from quadratic polynomial longitudinal trajectories, we consider quadratic polynomial curves of the form

$$X_{ij} = \beta_{0i}^{(L)} + \beta_{1i}t + \beta_2 t^2 + \epsilon_{ij},$$

where the true value of β_{0i} are simulated from the Normal distribution $N(\mu = 5, \sigma = 1)$, β_{1i} are simulated from the Normal distribution $N(\mu = -0.5, \sigma = 0.1)$, β_2 is set to be equal to -0.1, and finally ϵ_{ij} is the measurement error that is independent across measures and across subjects and are simulated from the Normal distribution $N(\mu = 0, \sigma = 0.1)$.

Under the second scenario, longitudinal albumin values are generated from random non-linear curves. In particular, we generate random non-linear albumin trajectories that are realizations of a Gaussian process that are centered around the subject-specific random intercepts $\beta_{0i}^{(L)}$ that are generated from the Normal distribution $N(\mu = 5, \sigma = 1)$. We consider a Gaussian process with the squared exponential covariance function with the correlation length of $\rho^2 = 0.1$ and the subject-specific measures of volatility κ_i^2 that are generated from the uniform distribution $U(0, 1)$.

For each simulation scenario, once longitudinal measures are generated, we generate survival data where survival times are distributed according to the Weibull distribution $Weibull(\tau, \lambda_i)$, where the shape parameter τ is set to 1.5 and λ_i, which is the log of the scale parameter of the Weibull distribution, is set to $\beta_{i0}^{(S)} + \beta_1 X_i(t)$, where $\beta_{i0}^{(S)}$ are generated from an equally weighted mixture of two Normal distributions of $N(\mu = -1.5, \sigma = 1)$ and $N(\mu = 1.5, \sigma = 1)$, β_1 is fixed to -0.5, and $X_i(t)$ is the longitudinal value for subject i at time t that is already simulated in the longitudinal step of the data simulation.

Our proposed joint longitudinal-survival model assumes the Normal prior $N(\mu = 5, \sigma = 2)$

on the random intercepts β_0^i, the log-Normal prior $log-Normal(-1,2)$ on κ_i^2, the log-Normal prior $log-Normal(-1,1)$ on σ^2, the log-Normal prior $log-Normal(0,1)$ on τ, the Normal prior $N(\mu=0,\sigma=5)$ on the survival shared intercept β_0, the Normal prior $N(\mu=0,\sigma=5)$ on the survival coefficient β_1, the Gamma prior $\Gamma(3,3)$ on the concentration parameter of the Dirichlet distribution, and the Normal prior $N(\mu=0,\sigma=5)$ as the base distribution of the Dirichlet distribution.

As the results in Table 4.1 show, when data are simulated with a longitudinal trajectories that are quadratic polynomial curves, the joint polynomial model performs better in terms of estimating the albumin coefficient in the survival model with a smaller mean squared error compared to our proposed joint longitudinal-survival. In real world, however, the true functional forms of the trajectories of the biomarkers are not known. Under a general case where the biomarker trajectories can be any random non-linear curve (scenario 2), our proposed joint model outperforms the joint polynomial model. Further, our joint modeling framework that is capable of estimating differential subject-specific log baseline hazards provides significantly better coefficient estimates compared to the proportional hazard Cox model. Estimates under the Cox model are marginalized over all subjects and due to the non-collapsibility aspect of this model (Struthers and Kalbfliesch (1986), Martinussen and Vansteelandt (2013)), coefficient estimates shrink toward 0.

4.3.2 Model II - Simulation Results

In model II, not only do we adjust for the albumin value at time Y_i, but we also adjust for a weighted average slope of albumin from time $\tau_1 = 0$ up until the time $\tau_2 = Y_i$, where Y_i is either the event time for subject i or is the time that the subject got censored. This new model differentiates between the risk of death for a patient whose albumin value is improving compared to another patient with the same albumin level whose albumin is deteriorating. In

Covariate of Interest	True Conditional Estimand	Two-Stage Cox			Joint Polynomial Model			Joint Model		
		Mean	SD	MSE	Mean	SD	MSE	Mean	SD	MSE
Scenario 1										
Albumin(t)	-0.5	-0.273	0.056	0.119	-0.495	0.019	0.003	-0.441	0.105	0.012
Scenario 2										
Albumin(t)	-0.5	-0.258	0.080	0.125	-0.380	0.080	0.034	-0.462	0.110	0.010

Table 4.1: Model I Simulation results - joint longitudinal-survival data were generated under the simulation scenarios of one when longitudinal measures are sampled from the quadratic polynomial trajectories (scenario 1) and another scenario when longitudinal measures are sampled from random non-linear curves (scenario 2). Under each scenario, we fit three models of a joint longitudinal-survival model with the assumption that longitudinal trajectories are quadratic polynomial (Joint Polynomial Model), our proposed joint longitudinal-survival with a flexible Gaussian process longitudinal component (Joint Model), and a two-stage Cox proportional model with longitudinal trajectories with parameters that set to the posterior mean of a Gaussian process longitudinal model that is fit separately.

particular, we consider weighted average slope of albumin once under the weighting scheme of the form

$$Q(t) = \frac{1}{\tau_1 - \tau_0},$$

and another time under the weighting scheme of

$$Q(t) = \begin{cases} 1, & \text{if } t = T_i \\ 0, & \text{otherwise.} \end{cases}$$

The first weighting scheme leads to the area under the derivative curve. The second weighting scheme will result in a point-wise derivative of albumin at time Y_i.

We generate synthetic data for 300 subjects each with 9 to 12 longitudinal measurements where longitudinal albumin values are generated from a Gaussian process that is centered around the subject-specific random intercepts $\beta_{0i}^{(L)}$ which are generated from the Normal distribution $N(\mu = 5, \sigma = 1)$. We consider a Gaussian process with the squared exponential covariance function with the correlation length of $\rho^2 = 0.1$ and the subject-specific

measures of volatility κ_i^2 that are generated from the uniform distribution $U(0,1)$. Once longitudinal measures are generated, we generate survival data where survival times are distributed according to the Weibull distribution $Weibull(\tau, \lambda_i)$, where the shape parameter τ is set to 1.5 and λ_i, which is the log of the scale parameter in Weibull distribution, is set to $\beta_{i0}^{(S)} + \beta_1 X_i(t) + \beta_2 X'_{AUC,i}(t)$, where $\beta_{i0}^{(S)}$ are generated from an equally weighted mixture of two Normal distributions of $N(\mu = -1.5, \sigma = 1)$ and $N(\mu = 1.5, \sigma = 1)$, β_1 is fixed to 0.3, β_2 is fixed to 0.5, $X_i(t)$ is the longitudinal value for subject i at time t and $X'_{AUC,i}(t)$ is the average slope of albumin.

Our proposed joint longitudinal-survival model assumes the Normal prior $N(\mu = 5, \sigma = 2)$ on the random intercepts β_0^i, the log-Normal prior $log-Normal(-1, 2)$ on κ_i^2, the log-Normal prior $log-Normal(-1, 1)$ on σ^2, the log-Normal prior $log-Normal(0, 1)$ on τ, the Normal prior $N(\mu = 0, \sigma = 5)$ on the survival shared intercept β_0, the Normal prior $N(\mu = 0, \sigma = 5)$ on the survival coefficient β_1, the Normal prior $N(\mu = 0, \sigma = 5)$ on the survival coefficient β_2, the Gamma prior $\Gamma(3, 3)$ on the concentration parameter of the Dirichlet distribution, and the Normal prior $N(\mu = 0, \sigma = 5)$ as the base distribution of the Dirichlet distribution.

We fit our proposed joint longitudinal-survival model. As a comparison, we also fit a two-stage Cox model where the longitudinal curve of albumin and its derivative curve are estimated using hyper-parameters set as the posterior median of a Bayesian Gaussian Process model. As we can see from table 4.2, our joint model provides closer estimates to the coefficient values with a smaller mean squared error compared with the two-stage Cox model. Our proposed model is capable of detecting differential subject-specific baseline hazards whereas the Cox model is not capable of differentiating between subjects and provides estimates that are marginalized across all subjects. Further, the simulation results show the capability of our method in detecting the true underlying longitudinal curves and the ability of our method on properly estimating the average derivative of those curves.

Covariate of Interest	True Conditional Estimand	Two-Stage Cox			Joint Model		
		Mean	SD	MSE	Mean	SD	MSE
Case 1 - Uniform Weights							
Albumin(t)	0.3	0.191	0.099	0.022	0.303	0.109	0.008
Area under the derivative curve(t)	0.5	0.346	0.179	0.053	0.449	0.188	0.030
Case 2 - Point-Wise Weights							
Albumin(t)	0.3	0.142	0.095	0.033	0.261	0.104	0.009
$\frac{d(Albumin(t))}{dt}$	0.5	0.412	0.123	0.022	0.477	0.152	0.013

Table 4.2: Model II simulation results - joint longitudinal-survival data were generated for 300 subjects each with 9 to 12 within subject measurements where longitudinal albumin values are generated from a Gaussian process that is centered around the subject-specific random intercepts $\beta_{0i}^{(L)}$ which are generated from the Normal distribution $N(\mu = 5, \sigma = 1)$. We consider a Gaussian process with the squared exponential covariance function with the correlation length of $\rho^2 = 0.1$ and the subject-specific measures of volatility κ_i^2 that are generated from the uniform distribution $U(0, 1)$. Once longitudinal measures are generated, we generate survival data where survival times are distributed according to the Weibull distribution $Weibull(\tau, \lambda_i)$, where the shape parameter τ is set to 1.5 and λ_i, which is the log of the scale parameter in Weibull distribution, is set to $\beta_{i0}^{(S)} + \beta_1 X_i(t) + \beta_2 X'_{AUC,i}(t)$, where $\beta_{i0}^{(S)}$ are generated from an equally weighted mixture of two Normal distributions of $N(\mu = -1.5, \sigma = 1)$ and $N(\mu = 1.5, \sigma = 1)$, β_1 is fixed to 0.3, β_2 is fixed to 0.5, $X_i(t)$ is the longitudinal value for subject i at time t and $X'_{AUC,i}(t)$ is the average slope of albumin. We fit our proposed joint longitudinal-survival model as well as a two-stage Cox proportional hazard model as a the comparison model.

4.3.3 Model III - Simulation Results

In model III, we test the association between the summary measures of the longitudinal biomarker trajectories and the survival outcomes. In particular, we consider the relation between the summary measures of subject-specific random intercept $\beta_{i0}^{(L)}$ and subject-specific measure of volatility κ_i^2 and survival times.

We generate synthetic data for $N = 300$ subjects each with 9 to 12 longitudinal measurements where longitudinal albumin values are generated from a Gaussian process that is centered around the subject-specific random intercepts $\beta_{0i}^{(L)}$ which are generated from the Normal distribution $N(\mu = 5, \sigma = 1)$. We consider a Gaussian process with the squared exponential covariance function with the correlation length of $\rho^2 = 0.1$ and the subject-specific

measures of volatility κ_i^2 that are generated from the uniform distribution $U(0,1)$. Once longitudinal measures are generated, we generate survival data where survival times are distributed according to the Weibull distribution $Weibull(\tau, \lambda_i)$, where the shape parameter τ is set to 1.5 and λ_i, which is the log of the scale parameter in Weibull distribution, is set to $\beta_{i0}^{(S)} + \beta_1 Age + \beta_2 {\beta_{i0}}^{(L)} + \beta_3 \kappa_i^{2(L)}$, where $\beta_{i0}^{(S)}$ are generated from an equally weighted mixture of two Normal distributions of $N(\mu = -1.5, \sigma = 1)$ and $N(\mu = 1.5, \sigma = 1)$, β_1 is fixed to 0.5, β_2 is fixed to -0.3, β_3 is fixed to 0.7, Age is a standardized covariate that is generated from the Normal distribution $N(\mu = 0, \sigma = 1)$, ${\beta_{i0}}^{(L)}$ is subject-specific random intercepts of the longitudinal trajectories, and $\kappa_i^{2(L)}$ are subject specific measure of volatility of the longitudinal trajectories.

Our proposed joint longitudinal-survival model assumes the Normal prior $N(\mu = 5, \sigma = 2)$ on the random intercepts β_0^i, the log-Normal prior $log-Normal(-1, 2)$ on κ_i^2, the log-Normal prior $log-Normal(-1, 1)$ on σ^2, the log-Normal prior $log-Normal(0, 1)$ on τ, the Normal prior $N(\mu = 0, \sigma = 5)$ on the survival shared intercept β_0, the Normal prior $N(\mu = 0, \sigma = 5)$ on the survival coefficient β_1, the Normal prior $N(\mu = 0, \sigma = 5)$ on the survival coefficient β_2, the Gamma prior $\Gamma(3, 3)$ on the concentration parameter of the Dirichlet distribution, and the Normal prior $N(\mu = 0, \sigma = 5)$ as the base distribution of the Dirichlet distribution.

We fit our proposed joint survival-longitudinal model (model III) as well as a two-stage Cox proportional hazard model as a comparison model. The two-stage Cox model is a simple Cox proportional hazard model with covariate $\beta_{0i}^{(L)}$ and $\kappa_i^{2(L)}$ that are posterior medians from a separate longitudinal Gaussian process model. As the results in Table 4.3 show, our proposed joint model provides closer estimates to the true coefficients that also have significantly smaller mean squared error compared to the two-stage Cox model. Our proposed joint model is capable of detecting the differential subject-specific baseline hazards. Unlike our model, Cox model is blind to the subject-specific baseline hazards and hence, provides coefficient estimates that are marginalized over all subjects. These marginalized estimates

from the Cox model shrink toward 0 as the Cox model with a multiplicative hazard function is non-collapsible.

As one can see in the joint model results in Table 4.3, the coefficient estimate for $\kappa^{2(L)}$ is not as close to the true coefficient value compared with other coefficient estimates. This is due to the fact that only 9 to 12 longitudinal measures per subject, there exists many plausible $\kappa_i^{2(L)}$ values that flexibly characterize the trajectory of the measured albumin values. This additional variability in plausible $\kappa_i^{2(L)}$ values has caused the coefficient estimate to shrink toward 0. Larger number of within subject longitudinal measures will provide more precision in estimating the true underlying $\kappa_i^{2(L)}$ and will lead to a coefficient estimate closer to the true value. In order to confirm this fact, we simulated additional data once with 36 within subject measures and another time with 72 within subject measures. Table 4.4 shows the results of fitting our proposed joint longitudinal-survival model to datasets that include subjects with 9 to 12 within subject measurements, to datasets with subjects with 36 within subject measurements, and to datasets with subjects with 72 within subject measurements. As the results show, with larger number of within subject measurements, coefficient estimate for $\kappa_i^{2(L)}$ is closer to the true value. This is due to the fact that with larger number of within subject albumin measurements, there exists a stronger likelihood to estimate the subject-specific volatility measures κ_i^2, and hence, there is less uncertainty about the estimated value of volatility measures.

Covariate of Interest	True Conditional Estimand	Two-Stage Cox			Joint Model		
		Mean	SD	MSE	Mean	SD	MSE
Age (scaled)	0.5	0.262	0.124	0.070	0.492	0.149	0.013
Baseline Albumin ($\beta_{0i}{}^{(L)}$)	-0.3	-0.141	0.118	0.040	-0.284	0.116	0.008
$\kappa_i^{2(L)}$	0.7	0.414	0.212	0.127	0.595	0.271	0.042

Table 4.3: Model III simulation results - joint longitudinal-survival data were generated for 300 subjects each with 9 to 12 longitudinal measurements where longitudinal albumin values are generated from a Gaussian process that is centered around the subject-specific random intercepts $\beta_{0i}^{(L)}$ which are generated from the Normal distribution $N(\mu = 5, \sigma = 1)$. We consider a Gaussian process with the squared exponential covariance function with the correlation length of $\rho^2 = 0.1$ and the subject-specific measures of volatility κ_i^2 that are generated from the uniform distribution $U(0, 1)$. Once longitudinal measures are generated, we generate survival data where survival times are distributed according to the Weibull distribution $Weibull(\tau, \lambda_i)$, where the shape parameter τ is set to 1.5 and λ_i, which is the log of the scale parameter in Weibull distribution, is set to $\beta_{i0}^{(S)} + \beta_1 Age + \beta_2 \beta_{i0}{}^{(L)} + \beta_3 \kappa_i^{2(L)}$, where $\beta_{i0}^{(S)}$ are generated from an equally weighted mixture of two Normal distributions of $N(\mu = -1.5, \sigma = 1)$ and $N(\mu = 1.5, \sigma = 1)$, β_1 is fixed to 0.5, β_2 is fixed to -0.3, β_3 is fixed to 0.7, Age is a standardized covariate that is generated from the Normal distribution $N(\mu = 0, \sigma = 1)$, $\beta_{i0}{}^{(L)}$ are subject-specific random intercepts of the longitudinal trajectories, and $\kappa_i^{2(L)}$ are subject specific measures of volatility of the longitudinal trajectories. We fit our proposed joint longitudinal-survival model as well as a two-stage Cox proportional hazard model as a the comparison model.

4.4 Application of the Proposed Joint Longitudinal-Survival Models to Data from the United States Renal Data System

In this section, we apply our proposed joint longitudinal-survival models to data on $n = 1,112$ end stage renal disease patients participating in the Dialysis Morbidity and Mortality Studies (DMMS) nutritional study that is obtained from the United States Renal Data System. For every participating patient in the study, up to 12 albumin measurements were taken uniformly over two years of followup. The presented analyses are restricted to only the patients who had at least nine albumin measurements in order to provide sufficient data for modeling the trajectory and the volatility of albumin. The censoring rate in the data is at

Covariate of Interest	True Conditional Estimand	Joint Model ($l_i = 12$)			Joint Model ($l_i = 36$)			Joint Model ($l_i = 72$)		
		Mean	SD	MSE	Mean	SD	MSE	Mean	SD	MSE
Age (scaled)	0.5	0.492	0.149	0.013	0.493	0.144	0.015	0.495	0.145	0.016
${\beta_{i0}}^{(L)}$	-0.3	-0.284	0.116	0.008	-0.308	0.116	0.006	-0.295	0.115	0.007
$\kappa_i^{2(L)}$	0.7	0.595	0.271	0.042	0.639	0.284	0.043	0.651	0.293	0.039

Table 4.4: Model III simulation results with datasets with $l_i = 36$ and $l_i = 72$ within subject measurements. In order to test the sensitivity of the $\kappa^{2(L)}$ coefficient estimate to the number of within subject measurements, l_i, we simulated joint longitudinal-survival data once when each subject has 36 within subject measurements and another time when each subject has 72 within subject measurements. Under each scenario, we simulated 200 datasets each with 300 subjects. Other simulation parameters remained the same as the simulation parameters used in Table 4.3. This means, we simulated longitudinal data from Gaussian process that is centered around the subject-specific random intercepts $\beta_{0i}^{(L)}$ which are generated from the Normal distribution $N(\mu = 5, \sigma = 1)$. We consider a Gaussian process with the squared exponential covariance function with the correlation length of $\rho^2 = 0.1$ and the subject-specific measures of volatility κ_i^2 that are generated from the uniform distribution $U(0, 1)$. Once longitudinal measures are generated, we generate survival data where survival times are distributed according to the Weibull distribution $Weibull(\tau, \lambda_i)$, where the shape parameter τ is set to 1.5 and λ_i, which is the log of the scale parameter in Weibull distribution, is set to $\beta_{i0}^{(S)} + \beta_1 Age + \beta_2 {\beta_{i0}}^{(L)} + \beta_3 \kappa_i^{2(L)}$, where $\beta_{i0}^{(S)}$ are generated from an equally weighted mixture of two Normal distributions of $N(\mu = -1.5, \sigma = 1)$ and $N(\mu = 1.5, \sigma = 1)$, β_1 is fixed to 0.5, β_2 is fixed to -0.3, β_3 is fixed to 0.7, Age is a standardized covariate that is generated from the Normal distribution $N(\mu = 0, \sigma = 1)$, ${\beta_{i0}}^{(L)}$ are subject-specific random intercepts of the longitudinal trajectories, and $\kappa_i^{2(L)}$ are subject specific measures of volatility of the longitudinal trajectories.

43% over a maximal follow-up time of 4.5 years.

Using the same data, Fung et al. (2002) showed that both baseline albumin level and the slope of albumin over time are significant predictors of mortality among ESRD patients. While our models are capable of replicating Fung et al's findings, our models are also capable of:

- model 1: testing the association between albumin value at the time of death and the risk of death
- model 2: testing the association between albumin value and an average derivative of albumin up until time t and the risk of death.
- model 3: Testing the association between risk of mortality and the two summary

measures of the baseline and the volatility of albumin measures

In order to adjust for other potential confounding factors, our proposed models also include patient's age, gender, race, smoking status, diabetes, an indicator of whether the patient appeared malnourished at baseline, BMI at baseline, baseline cholesterol, and baseline systolic blood pressure. The adjusted covariates are consistent with those originally presented in Fung et al. (2002).

Table 4.5 and Table 4.6 provide the results of fitting our proposed model I, Model II, and Model III to the USRDS data. All joint models were run for 10,000 posterior samples where the initial 5,000 samples are discarded as burn-in samples.

4.4.1 Model I - Application to USRDS data

We fit our proposed joint model I to the data. As a comparison model, we also fit as last-observation carried forward (LOCF) Cox model. Table 4.5 shows the estimated coefficients from both models. Between the two models, the estimated relative risk associated with all time-invariant baseline survival covariates are similar between the two models. However, the relative risk associated with every one unit decrement in serum albumin is much larger under our proposed joint model compared to the last-observation carried forward Cox model. This is quite expected as our model is capable of estimating subject-specific albumin trajectories over time and is capable of accurately testing the association between albumin value at time of death and risk of death. Unlike our model,the LOCF Cox model uses the most recent albumin measure which in reality might be quite different than the albumin value at the time of death. In both models, albumin is identified as a significant risk factor of mortality. In particular, based on the results from our proposed joint Model I, it is estimated that every 1 g/dL decrement in albumin is associated with a 4.5 times higher risk of death.

Covariates	No. of Cases	No. of Deaths	LOCF Cox Model Relative Risk (95% CI)	P-Value	Joint Model Relative Risk (95% CR)
Age (10y)	1,112	630	1.44 (1.35-1.53)	<.001	1.45 (1.36,1.55)
Sex					
Men	560	312	1.0		1.0
Women	552	318	0.96 (0.81,1.13)	0.60	0.97 (0.82,1.16)
Race					
White	542	350	1.0		1.0
Black	482	243	0.81 (0.68,0.96)	0.01	0.79 (0.67,0.94)
Other	88	37	0.52 (0.37,0.74)	<.001	0.49 (0.34,0.69)
Smoking					
Nonsmoker	645	337	1.0		1.0
Former	307	197	1.17 (0.98,1.41)	0.09	1.20 (0.99,1.44)
Current	160	96	1.52 (1.19,1.94)	<.001	1.53 (1.21,1.95)
Diabetes					
No	716	363	1.0		1.0
Yes	396	267	1.66 (1.40,1.97)	<.001	1.69 (1.43,2.00)
Undernourished					
No	958	517	1.0		1.0
Yes	154	113	1.39 (1.12,1.72)	0.003	1.35 (1.08,1.66)
BMI (per-5 kg/m^2 decrement)	1,112	630	1.08 (1.00,1.17)	0.07	1.08 (1.00,1.17)
Cholesterol (per 20 mg/dL)	1,112	630	0.97 (0.93,1.00)	0.08	0.96 (0.93,1.00)
Systolic blood pressure (per 10mm Hg)	1,112	630	0.98 (0.95,1.02)	0.38	0.98 (0.95,1.02)
Serum albumin(t) (1-g/dL decrement)	1,112	630	2.48 (2.00,3.07)	<0.001	4.54 (3.03,5.55)

Table 4.5: Estimated Relative Risk and corresponding 95% credible region from our proposed joint model where we adjust for time-dependent albumin value that is imputed from the longitudinal component of the model. We also fit a last-observation carried forward Cox proportional hazards model with last albumin value carried forward where we report coefficients estimates, 95% confidence interval, and p-value for the estimated coefficients. In both models, we adjust for potential confounding factors as reported by Fung et al. (2002).

4.4.2 Model II - Application to USRDS data

Other than the albumin value at the time of death, the average slope of albumin over time might also be a risk factor of mortality in end-stage renal disease patients. In our proposed joint Model II, we also adjust for the area under the derivative curve of the albumin trajectory from the time that the followup starts until the survival time which is either the time of death or the censoring time. Table 4.6 shows the results from our proposed model. Based on the

results, every one g/dL decrement in albumin is associated with 3.95 times higher risk of death. Also, higher average slope of albumin, that is every 1 g/dL/month increase in the average slope, is associated with 2.3 times higher risk of death. This is consistent with Fung et al. (2002) results on the association between the slope of albumin and the risk of death. Our proposed method is also capable of adjusting for the local effect of the slope of albumin. For instance, instead of averaging the slope of the follow-up time, one may only integrate over the 6 months prior to the time of death.

4.4.3 Model III - Application to USRDS data

Fung et al. (2002) showed that the baseline albumin and the slope of albumin over time are two independent risk factors of mortality among the end-stage renal disease patients. It is quite natural to hypothesize that the volatility of albumin could also be a risk factor of mortality among these patients. In our proposed joint longitudinal-survival Model III, we consider two summary measures of the trajectories of the longitudinal albumin values, one the baseline albumin measures ($\beta_{0i}{}^{(L)}$), and another the subject-specific volatility measure of albumin ($\kappa_i^{2(L)}$). Table 4.6 also shows the results from our proposed Model III. The results from our model confirms that the baseline serum albumin is a risk factor of mortality. Further, the results from our model indicate that the volatility of albumin is also a significant risk factor of mortality, where every one unit increase in κ^2, which indicates a higher volatility, is associated with 1.2 times higher risk of death.

Figure 4.2 shows albumin trajectories of 10 randomly sampled individuals. Hollow circles are the actual albumin measures for each subject. Also, plots show the posterior Gaussian process trajectory fit along with it's 95% credible region.

Covariates	No. of Cases	No. of Deaths	Joint Model Relative Risk (95% CI)
Model 2			
Serum Albumin(t) (1-g/dL decrement)	1,112	630	3.95 (3.18,4.71)
Average Derivative of Serum Albumin[1] (1-g/dL/month decrement)	1,112	630	2.33 (1.40,3.73)
Model 3			
Baseline Albumin($\beta_{0i}{}^{(L)}$) (1-g/dL decrement)	1,112	630	5.54 (4.19,6.94)
$\kappa_i^{2(L)}$ (increase in volatility)[2]	1,112	630	1.23 (1.02,1.41)

1 : One may only consider the local effect of average serum albumin slope by computing the area under the derivative from 6 months prior to death up until the time of death.

2 : In a similar model, we adjusted for κ^2 values as a categorical variable with a cut point equal to the posterior mean of all κ^2 values (0.1) and we got a similar estimate relative risk (1.21).

Table 4.6: Model II and Model III results that show the estimated relative risk and corresponding confidence intervals from our proposed joint model II and model III. Potential confounding factors, as reported by Fung et al. (2002), were also adjusted in the model but have been removed from the tables for brevity. Our proposed Model II is capable of testing the association between albumin values at the time of death as well as the average derivative of the subject-specific albumin trajectories from the time the follow up time starts up until the death or the censoring time. Our proposed Model III tests the association risk of mortality and two albumin trajectory summary measures of the subject-specific random intercepts ($\beta_{0i}{}^{(L)}$) and the subject specific volatility measures ($\kappa_i^{2(L)}$).

4.5 Discussion

Monitoring the health of patients often involves recording risk factors over time. In such situations, it is essential to evaluate the association between those longitudinal measurements and survival outcome. To this end, joint longitudinal-survival models provide an efficient inferential framework.

We proposed a joint longitudinal-survival framework that avoids some of the restrictive assumptions commonly used in the existing models. Further, our methods propose a stronger link between longitudinal and survival data through an introduction of new ways of adjusting for the biomarker value at time t, adjusting for the average derivative of the biomarker over time, and moving beyond the first-order trend and accounting for volatility of biomarker measures over time.

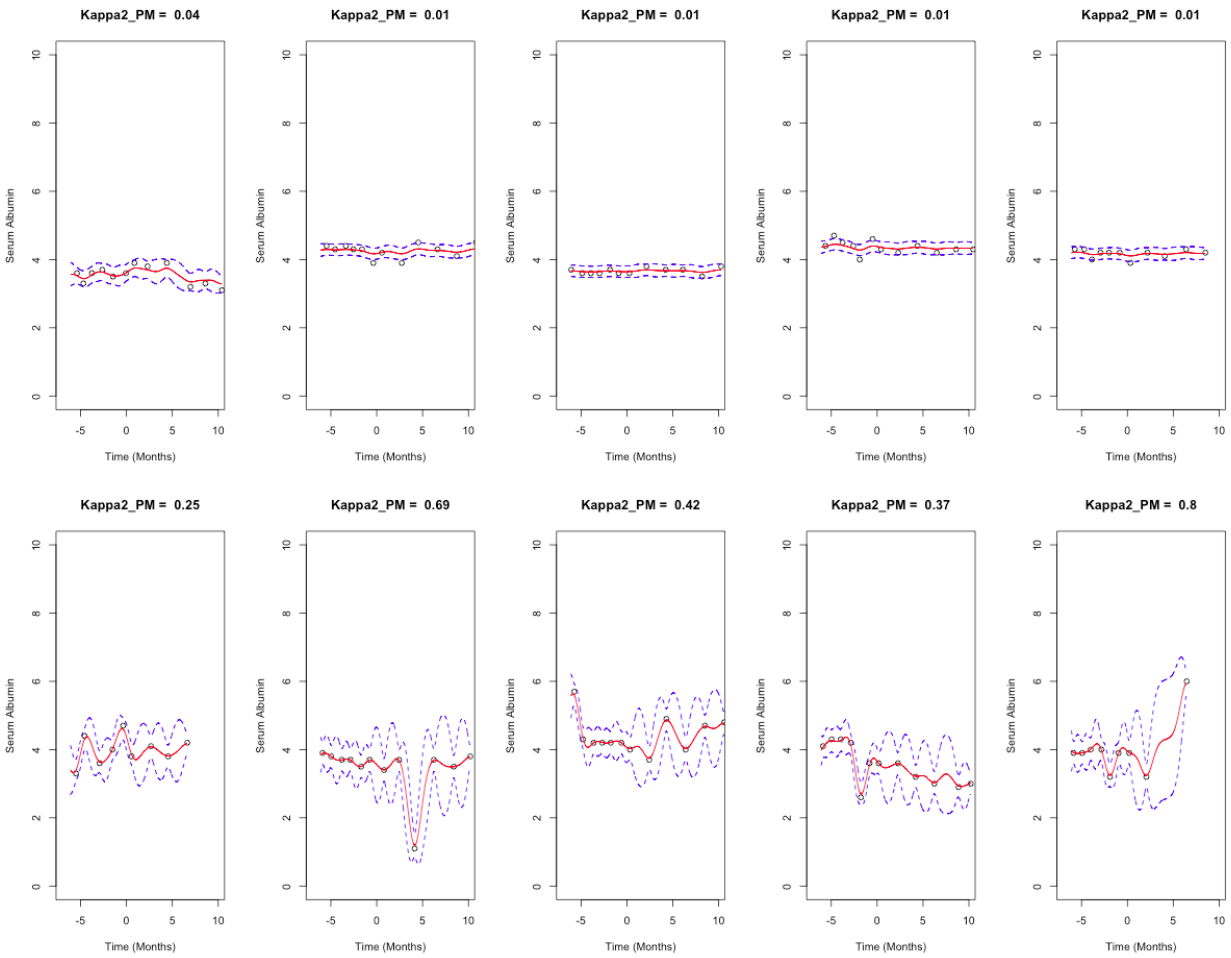


Figure 4.2: Actual longitudinal albumin trajectories of 10 randomly selected individuals with end-stage renal disease that were selected from the USRDS data. Hollow circles are the actual measured albumin values, red lines are the posterior median fitted curves from our proposed Model III, and the dashed blue lines are the corresponding 95% posterior prediction intervals for the fitted trajectories. The title of each plot shows the posterior median of the volatility measure κ^2 for the subject whose albumin measures are shown in the plot.

A two-stage approach to associating biomarker volatility with the survival outcome has been proposed by Holsclaw et al. (2014). Our work here extends this approach by simultaneously estimating longitudinal and survival parameters, thus accounting for uncertainty in the longitudinal measures. Our proposed models can also be considered as an extension of the joint model proposed by Brown and Ibrahim (2003) in that we use the same idea of dividing the joint likelihood into a marginal longitudinal likelihood and conditional survival likelihood. However, instead of fitting quadratic trajectories, we use a flexible longitudinal model based

on the Gaussian processes. Further, for the survival outcome, instead of assuming a piecewise exponential model, we use a flexible survival model by incorporating the Dirichlet process mixture of Weibull distributions. Our proposed modeling framework is capable of modeling additional summary measures of longitudinally measured biomarkers and relating them to the survival outcome in a time-dependent fashion.

Our proposed models, despite their flexibility and novelty, have some limitations. By using th Bayesian non-parameteric Dirichlet process and the Gaussian process techniques, while we provide a flexible modeling framework that avoids common distributional assumptions, however, these techniques are generally not scalable when the number of subjects and the number of within subject measurements increase. Furthermore, the survival component of our model still relies on the proportional hazard assumption. In future, our modeling framework can be extended to include a more general non-proportional hazard survival models that can also include time-dependent coefficients inside the survival model. By using some alternatives to the common MCMC techniques, including parallel-MCMC methods and variational methods, our method can become more computationally efficient and scalable for larger datasets.

Often times in monitoring the health of patients, multiple longitudinal risk factors are measured. One can use our introduced modeling framework in this chapter in order to build a joint longitudinal-survival model with multiple longitudinal processes each process modeled independently from other longitudinal processes. In reality, however, one expects that patients longitudinal risk factors to be correlated. A methodology that is capable of modeling multiple biomarkers simultaneously by taking the correlation between biomarkers into account can be beneficial specially when there exists differential densities between different longitudinal processes. In the next chapter, we introduce a joint modeling framework that is capable of modeling multiple longitudinal processes simultaneously by taking the correlation between those processes into account.

Chapter 5

A Flexible Joint Longitudinal-Survival Model for a Simultaneous Modeling of Multiple Longitudinal Biomarkers

One natural extension to the flexible joint longitudinal-survival model introduced in Chapter 4 is to build a model that is capable of modeling multiple longitudinal biomarkers simultaneously. Using the flexible joint modeling framework introduced in Chapter 4, one can include multiple longitudinal biomarkers each modeled separately. However, by simultaneously modeling biomarkers and by taking the correlation between biomarkers into account, one can gain more efficiency in modeling the trajectory of those biomarkers and hence, can better establish the relation between the longitudinal biomarkers and the survival outcomes.

We start with a brief introduction in Section 5.1. Our proposed methodology is introduced in Section 5.2. In Section 5.3, Simulation studies are presented to assess the operating characteristics of the proposed model. In Section 5.4, we present the result of applying the proposed method to real data on dialysis patients from DaVita database. We finally conclude

with a discussion in Section 5.5.

5.1 Introduction

Often times in clinical trials with survival end points, multiple longitudinal biomarkers are collected during the study follow up. Collected longitudinal measures on each subject are correlated as they are taken on the same subject as well as they are temporally dependent. An example is blood test biomarkers collected on end-stage renal disease patients on a monthly basis. While one can model the trajectory of each biomarker independently from other collected biomarkers, however, by simultaneously modeling the trajectory of all biomarkers, one can take the correlation between the different biomarkers into account, and hence, can gain more precision in modeling the biomarkers trajectories. When some biomarkers are measured less frequently compared to the other biomarkers, simultaneously modeling biomarkers can be particularly useful as one can gain more precision in estimating less frequent biomarkers by taking the correlation between all biomarkers into account and by borrowing information from the higher frequency biomarkers to better predict the lower frequency ones. Finally, a joint longitudinal-survival model with a flexible longitudinal component, which is capable of modeling the trajectory of multiple biomarkers simultaneously, can be used to test the association between the survival outcomes and the longitudinal biomarkers. We shall use the word "joint" when we model longitudinal-survival data simultaneously. We reserve the word "simultaneous" to refer to modeling multiple longitudinal measures at a same time rather than modeling each biomarker trajectory independently from others.

Simultaneously modeling longitudinal biomarkers can be considered in the context of a multivariate temporal process model that can be used both for inferential purposes as well as prediction and interpolation purposes. In order to develop such model, specification of a valid cross-covariance function is necessary. A valid cross-covariance function is required

to lead to a valid positive-definite covariance matrix for any number of time points and at any choice of these time points Gelfand and Banerjee (2010). A common cross-covariance function is to use separable cross-covariance function construction that shall be explained in more detail in Section 5.2.

The question of describing the correlation between multiple longitudinal measures has been addressed from a different perspective in geostatistics literature with a focus on spatial data. Bernardo et al. (1998) and Berger et al. (2003) described the kernel convolution technique for creating stationary and non-stationary spatial processes. Majumdar and Gelfand (2007) proposed producing cross-covariance functions by using convolution of covariance functions. Multivariate models in geostatistics started with Matheron (1973) and by introduction of some new concepts including cross-variogram and co-Kriging. Gelfand and Banerjee (2010) defined co-Kriging as a spatial prediction method that uses both the information of the process being considered as well as the information from other related processes. Co-Kriging is commonly addressed in the context of the linear models as these models are easily interpreted. These models that are commonly known as linear model of co-regionalization (LMC) have been considered widely in the literature including Grzebyk and Wackernagel (1994), Schmidt and Gelfand (2003), and Ver Hoef et al. (2004).

Following our idea from Chapter 4 on developing a flexible longitudinal model using Gaussian Process models, we are interested in extending our method to a flexible Gaussian process model capable of modeling multiple biomarkers simultaneously. In Sec 5.2, we will show how a Gaussian process model can be extended to a multivariate Gaussian process to model multiple longitudinal biomarkers simultaneously. Using our proposed multivariate Gaussian process, we then build a joint longitudinal-survival model capable of relating longitudinal biomarkers to survival outcome.

5.2 Methodology

5.2.1 Multivariate Gaussian Process

Consider the function f that relates an input space $\mathcal{T}$ to an output space $\mathcal{X}$. As an alternative to an explicit functional assumption on f, one can assume a Gaussian process prior on f . One may consider time as the input space and the space of a longitudinal measure X as the output space and may use Gaussian process prior as a prior on all plausible functions f relating time t to the longitudinal measure X at time t.

$$X(t) = f(t) + \epsilon$$
$$f \sim GP(\mu(t), C_f(t, t'))$$

In the setting above, f is considered as a univariate function with one output at each time point t. In general, however, f can be a multivariate function with a vector of outputs at each time t. A multivariate function $\boldsymbol{f}$ will require a multivariate Gaussian process prior. One can consider a more general frame work of the following form

$$\boldsymbol{X(t)} = \boldsymbol{f}(t) + \epsilon,$$
$$\boldsymbol{f}(t) \sim GP(\boldsymbol{\mu_f}(t), \boldsymbol{C_f(t, t')}),$$

where $\boldsymbol{X}(t)$ is a vector of outputs at time t, $\boldsymbol{f}$ is a multivariate function with a multivariate Gaussian process prior with a mean vector function $\boldsymbol{\mu_f}(t)$ and a cross-covariance matrix function $\boldsymbol{C_f}(t, t')$.

Consider a multivariate longitudinal vector $\boldsymbol{X}(t)$ at time t with the dimension $q \times 1$. Without loss of generality and for simplicity of the notations, we assume a multivariate longitudinal random vector with mean zero, $E[\boldsymbol{X}(t)] = 0$. A cross-covariance function between two

generic time points t and t' is a matrix function with dimension $q \times q$ where the $(i,j)^{th}$ element of this matrix is defined as

$$\begin{aligned} C_{ij}(t,t') &= Cov\big(X_i(t), X_j(t')\big) \\ &= E[X_i(t)X_j(t')] - E[X_i(t)]E[X_j(t')] \\ &= E[X_i(t)X_j(t')], \end{aligned} \tag{5.1}$$

where $E[X_i(t)]$ and $E[X_j(t')]$ are 0. Equation (5.1) indicates that the cross-covariance matrix function $\boldsymbol{C}(t,t')$ can be defined as

$$\boldsymbol{C}(t,t') = E[\boldsymbol{X}(t)\boldsymbol{X}(t')^T]. \tag{5.2}$$

Consider n arbitrary time points $\{t_1, t_2, \ldots, t_n\}$. At each time point t_i, $\boldsymbol{X}(t_i)$ is a $q \times 1$ vector. Concatenating n such output vectors, one can define an $nq \times 1$ vector $\boldsymbol{X}$, where $\boldsymbol{X} = [\boldsymbol{X}(t_1), \boldsymbol{X}(t_2), \ldots, \boldsymbol{X}(t_n)]$. Random vector $\boldsymbol{X}$ is mean-zero with a covariance matrix $\boldsymbol{\Sigma_X}$ with the dimension $nq \times nq$. $\boldsymbol{\Sigma_X}$ is a block matrix where each block is the cross-covariance matrix corresponding to time t_i and t_j for all q outputs at each time point.

As a covariance matrix, $\boldsymbol{\Sigma_X}$ has to be symmetric and positive definite. As Gelfand and Banerjee (2010) showed, this requires the covariance matrix function $\boldsymbol{C}(t,t')$ to satisfy the two following conditions of

1. $\boldsymbol{C}(t,t') = \boldsymbol{C}^T(t,t')$,
2. for any integer value n and for any arbitrary collection of "n" time points:
   $\sum_{i=1}^{n}\sum_{j=1}^{n} \boldsymbol{X}^T(t_i)C(t_i,t_j)\boldsymbol{X}(t_j) > 0$.

A multivariate process $\boldsymbol{X}(t)$ is stationary if the cross-covariance matrix function depends only on the time difference between t and t', where we can write $\boldsymbol{C}(t,t') = \boldsymbol{C}(t'-t)$. Multivariate

process $\boldsymbol{X}(t)$ is called isotropic if the cross-covariance matrix function depends only on the absolute difference between t and t', where we can write $\boldsymbol{C}(t,t') = \boldsymbol{C}(|t'-t|)$. Yadrenko (1987) showed that under isotropic condition, a covariance function $C_{ij}(h)$ will form a valid cross-covariance matrix function if and only if the $C_{ij}(h)$ function for a positive-definite measure $F(.)$ has a cross-spectral representation of the form

$$C_{ij}(h) = \int exp(2\pi i t^T h) d(F_{ij}(t)).$$

Despite the existence of many methods proposed in the literature, when cross-covariance functions are unknown, specification of a valid cross-covariance function based on the observed data is a very difficult task. One common approach to specify a valid cross-covariance function is to use a class of covariance functions known as separable cross-covariance structures that shall be introduced in the next section.

5.2.2 Separable Cross-Covariance Functions

Consider $\boldsymbol{X}(t)$ as a vector of q longitudinal variables all measured at time t. The cross-covariance matrix function $\boldsymbol{C}(t,t')$ is then a $q \times q$ matrix where it's $(i,j)^{th}$ element is equal to $C_{ij}(t,t')$ that was defined in equation (5.1). One can assume that the covariance between the q longitudinal variables at each time-point remains the same at all time points t and can be specified with a positive definite covariance matrix $\boldsymbol{R}$. The correlation between measured values at time t and measured values at time t' can then be expressed using a univariate correlation function $\rho(t,t')$. Given this specification, one can write

$$\boldsymbol{C}_(t,t') = \rho(t,t')\boldsymbol{R}, \tag{5.3}$$

where $\boldsymbol{R}$ is of size $q \times q$ and represents the covariance matrix between the q elements of $\boldsymbol{X}(t)$ that remain the same at all time points t, and $\rho(t, t')$ represents the correlation between measures at time t and measures at time t'.

Now consider n time points $\{t_1, t_2, \ldots, t_n\}$ where at each time point t_i, we observe a stack of q longitudinal random variables $\boldsymbol{X}(t_i)$. Consider $\boldsymbol{X} = [\boldsymbol{X}(t_1), \boldsymbol{X}(t_2), \ldots, \boldsymbol{X}(t_n)]^T$ as the vertical stack of n longitudinal vectors, each element of that vector is an observed $\boldsymbol{X}(t_i)$ at a time point t_i. $Cov(\boldsymbol{X})$ can be defined as

$$Cov(\boldsymbol{X}) = \boldsymbol{R} \otimes \boldsymbol{S}, \tag{5.4}$$

where the $(i, j)^{th}$ element of $\boldsymbol{S}$ is represented with S_{ij} and is equal to $S_{ij} = \rho(t_i, t_j)$, the notation $\otimes$ represents the Kronecker product, and $\boldsymbol{R}$ is the non-temporal covariance function between the q longitudinal variables that is assumed to remain the same across time points t_i's.

By using a separable cross-covariance function and given that $\boldsymbol{R}$ matrix is positive definite by definition, with a positive definite matrix $\boldsymbol{S}$, it's guaranteed that $Cov(\boldsymbol{X})$ will be positive definite. Furthermore, by using a separable cross-covariance structure, the determinant of the cross-covariance function, $|Cov(\boldsymbol{X})|$, and the inverse of the cross-covariance function,$Cov(\boldsymbol{X})^{-1}$, will become computational convenient to deal with. In particular, by using the properties of the Kronecker product, one can show that the determinant of the cross-covariance can be written as $|Cov(\boldsymbol{X})| = |\boldsymbol{R}|^q |\boldsymbol{S}|^n$, where $|\boldsymbol{R}|$ and $|\boldsymbol{S}|$ are the determinant of the $\boldsymbol{R}$ matrix and the determinant of the $\boldsymbol{S}$, respectively. Also, the inverse of the cross-covariance function can be written as $Cov(\boldsymbol{X})^{-1} = \boldsymbol{R}^{-1} \otimes \boldsymbol{S}^{-1}$. We shall use the class of separable cross-covariance functions to setup our proposed model.

While the class of separable covariance functions are guaranteed to provide a valid cross covariance function, however, a limitation of a separable cross-covariance function is that

it assumes that the covariance structure between longitudinal outputs at time $t$remains the same as t changes. In reality, this assumption might not necessarily hold.

5.2.3 A Joint Longitudinal-Survival Model with Multiple Longitudinal Biomarkers

In this section we introduce our proposed joint longitudinal-survival model that is capable of simultaneously modeling multiple longitudinal biomarkers and jointly relating them to the survival outcomes. We first start by introducing the likelihood specification of joint models. We then continue with introducing the longitudinal component of the model and the survival component. Our proposed method can work for any number of longitudinal biomarkers, however, for the sake of simplicity of the notations, we consider only two longitudinal biomarkers. We refer to the first longitudinal biomarker with $\boldsymbol{X}^{(1)}$ and to the second biomarker with $\boldsymbol{X}^{(2)}$. We denote survival outcome with Y. n refers to how many subjects are being followed up in the study. $l_i^{(1)}$ and $l_i^{(2)}$ refer to the number of longitudinal biomarker 1 measures and biomarker 2 measures obtained for subject i at time points $t_{ij}^{(1)}, j = 1, 2, \ldots, l_i^{(1)}$ and $t_{ik}^{(2)}, k = 1, 2, \ldots, l_i^{(2)}$, respectively. Also, associated with each subject, there is an observed survival time, $Y_i \equiv \min\{T_i, C_i\}$ and event indicator $\delta_i \equiv 1_{[Y_i=E_i]}$, where T_i and C_i denote the true event time and the censoring time for subject i, respectively. Further, we make the common assumption that C_i is independent of T_i for all i, $i = 1, \ldots, n$.

5.2.4 The Joint Model

Similar to Section 4.2.1, we define the contribution of each subject to the joint model likelihood as the multiplication of the likelihood function of the longitudinal measures for that subject and her/his time-to-event likelihood that is conditioned on her/his longitudinal mea-

sures. Let $f_L^{(i)}$, $f_{S|L}^{(i)}$, and $f_{L,S}^{(i)}$ denote the longitudinal likelihood contribution, the conditional survival likelihood contribution, and the joint likelihood contribution for subject i. One can write the joint longitudinal-survival likelihood function as

$$f_{L,S} \;=\; \prod_{i=1}^{n} f_{L,S}^{(i)} = \prod_{i=1}^{n} \big(f_L^{(i)} \times f_{S|L}^{(i)}\big). \tag{5.5}$$

We now explain the components specification of the model and we will conclude this section by explaining the posterior distribution of the joint model.

Longitudinal Model

We motivate the development of the multivariate Gaussian process model of two longitudinal biomarkers by first considering the following simple model for a single subject

$$\begin{bmatrix} \boldsymbol{X}_i^{(1)} \\ \boldsymbol{X}_i^{(2)} \end{bmatrix} \mid \begin{bmatrix} \boldsymbol{\beta}_{i0}^{(1)} \\ \boldsymbol{\beta}_{i0}^{(2)} \end{bmatrix} \sim N\Big(\begin{bmatrix} \boldsymbol{\beta}_{i0}^{(1)} \\ \boldsymbol{\beta}_{i0}^{(2)} \end{bmatrix}, \boldsymbol{\Sigma}_{\boldsymbol{\epsilon}} = \begin{bmatrix} \boldsymbol{\Sigma}^{(1)} & 0 \\ 0 & \boldsymbol{\Sigma}^{(2)} \end{bmatrix} \Big). \tag{5.6}$$

For simplicity of the notation, we assume that both biomarkers are measured simultaneously, that means at each time point t, we get to obtain measures on both biomarkers. Our method is not limited to this assumption and once readers are introduced with the model, we shall extend the notation to a model with no such assumption. We define l_i as the number of longitudinal measures per biomarker. These measures are obtained at an arbitrary time points $t_{i1}, t_{i2}, \ldots, t_{il_i}$. In the equation (5.6), $\boldsymbol{X}_i^{(1)}$ and $\boldsymbol{X}_i^{(2)}$ are vectors of longitudinal biomarker 1 measurements and longitudinal biomarker 2 measurements, each of size $l_i \times 1$ respectively. We shall stack $\boldsymbol{X}_i^{(1)}$ and $\boldsymbol{X}_i^{(2)}$ together into a column vector $\boldsymbol{X}_i$ that is of size

$2l_i \times 1$. $\boldsymbol{\beta}_{i0}^{(1)}$ is a vector of repeated random intercepts corresponding to biomarker 1 and $\boldsymbol{\beta}_{i0}^{(2)}$ is a vector of repeated random intercepts corresponding to biomarker 2, each of size $l_i \times 1$. Also, we shall stack $\boldsymbol{\beta}_{i0}^{(1)}$ and $\boldsymbol{\beta}_{i0}^{(2)}$ into a column vector $\boldsymbol{\beta}_{i0}^{(L)}$ of size $2l_i \times 1$. The model in equation (5.6) assumes biomarker 1 and biomarker 2 are independent and each biomarker has its own measurement error matrix. We consider $\Sigma^{(1)} = \sigma_1^2 I_{l_i \times l_i}$ and $\Sigma^{(2)} = \sigma_2^2 I_{l_i \times l_i}$, where σ_1^2 and σ_2^2 are shared parameters across all subjects.

By adding a stochastic component that is indexed by time to the model in equation (5.6), we can relax the independence assumption between biomarkers and we can also extend the model to capture non-linear patterns over time. Specifically, we consider a stochastic vector, $\boldsymbol{W}$, that is a realization from a multivariate Gaussian process prior, $\boldsymbol{W}(t)$, that is mean zero and has a separable cross-covariance function. Thus for subject i, $\boldsymbol{W_i} \sim N_{2l_i}(\mathbf{0}, C^i_{2l_i \times 2l_i})$, where $\boldsymbol{W_i}$ is a column vector that includes a stack of $\boldsymbol{W_i^{(1)}}$ and $\boldsymbol{W_i^{(2)}}$, where $\boldsymbol{W_i^{(1)}} = (W_{t_{i1}}, W_{t_{i2}}, \ldots, W_{t_{il_i}})$ and $\boldsymbol{W_i^{(2)}} = (W_{t_{i1}}, W_{t_{i2}}, \ldots, W_{t_{il_i}})$.

We characterize the covariance function, $Cov(\boldsymbol{W_i})$, using a separable cross-covariance structure as

$$Cov(\boldsymbol{W_i}) = R \otimes S_i, \tag{5.7}$$

where R is a 2 by 2 matrix that characterizes the the covariance between the two biomarkers and is assumed to be time-invariant, and S_i is a temporal covariance matrix. The R matrix is shared across all subjects with elements R_{11} and R_{22} characterizing marginal variances of the first and the second biomarker processes respectively, and with the R_{12} element characterizing the covariance between the two processes. In specific, one can decompose the covariance

matrix R into the following form

$$R = \begin{bmatrix} \tau_1 & 0 \\ 0 & \tau_2 \end{bmatrix} \Omega \begin{bmatrix} \tau_1 & 0 \\ 0 & \tau_2 \end{bmatrix}, \tag{5.8}$$

where τ_1 and τ_2 are square-root of the within biomarker 1 and biomarker 2 variances and Ω is a correlation matrix. Commonly, τ_1 is set to equal 1 where in that case, it's assumed that S_i will also capture the within biomarker 1 variability and τ_2 is treated as a parameter indicating the relative biomarker variability between biomarker 2 and biomarker 1. Hence, we define

$$\tau = \begin{bmatrix} 1 & 0 \\ 0 & \tau_2 \end{bmatrix}.$$

We re-write the equation (5.9) as

$$R = \tau\Omega\tau. \tag{5.9}$$

S_i is the covariance matrix characterizing how longitudinal measures change over time. Under the separable cross-covariance structure, we assume both biomarkers share the same covariance structure for changes in their values over time. We characterize S_i as an $l_i \times l_i$ matrix with elements $S_i(j, j')$ that is defined as

$$S_i(j, j') = {\kappa_i}^2 e^{-\rho^2(t_{ij} - t_{ij'})^2}, \tag{5.10}$$

where the hyperparameter ρ^2 controls the correlation length, and κ^2 controls the height of oscillations (Banerjee et al. 2004), and t_{ij} and $t_{ij'}$ are two different time points. For notational simplicity, we define $\boldsymbol{K_i} = e^{-\rho^2(t_{ij} - t_{ij'})^2}$; $j, j' \in \{1, \ldots, l_i\}$. We can extend the

longitudinal model in equation (5.6) to the flexible model below

$$\boldsymbol{X}_i|\boldsymbol{\beta}_{i0}^{(L)}, \boldsymbol{W_i}, \sigma_1^2, \sigma_2^2 \ \sim N(\boldsymbol{\beta}_{i0}^{(L)} + \boldsymbol{W_i}, \boldsymbol{\Sigma}_\epsilon), \tag{5.11}$$

where $\boldsymbol{W_i}$ is a stochastic vectors sampled from a Gaussian process prior of the form

$$\boldsymbol{W_i}|{\kappa_i}^2, \rho^2 \ \sim GP(\boldsymbol{0}, \boldsymbol{R} \otimes \boldsymbol{S}_i). \tag{5.12}$$

In the model defined by equation (5.11), σ_1^2 and σ_2^2 are assumed to be common across all subjects. Also, we assume the correlation length ρ^2 is fixed and hence, the subject-specific parameter ${\kappa_i}^2$ will have the role of capturing the within-subject volatility of the longitudinal biomarkers. Finally, the longitudinal component of our proposed joint model can be written as

$$\begin{aligned}
\boldsymbol{X_i}|\boldsymbol{W_i}, \beta_{i0}^{(1)}, \beta_{i0}^{(2)}, {\kappa_i}^2, \rho^2, \sigma_1^2, \sigma_2^2 &\sim N(\boldsymbol{\beta}_{i0}^{(L)} + \boldsymbol{W_i}, \Sigma_\epsilon) \\
\beta_{i0}^{(1)} &\sim N(\mu_{\beta_0^{(1)}}, \sigma^2_{\beta_0^{(1)}}) \\
\beta_{i0}^{(2)} &\sim N(\mu_{\beta_0^{(2)}}, \sigma^2_{\beta_0^{(2)}}) \\
\sigma_1^2 &\sim log - Normal(\mu_{\sigma_1^2}, \sigma_{\sigma_1^2}) \\
\sigma_2^2 &\sim log - Normal(\mu_{\sigma_2^2}, \sigma_{\sigma_2^2}) \\
\boldsymbol{W_i}|{\kappa_i}^2, \rho^2, \boldsymbol{t_i} &\sim GP(\boldsymbol{0}, R \otimes S_i) \\
{\kappa_i}^2 &\sim log - Normal(\mu_{\kappa^2}, \sigma_{\kappa^2}) \\
\tau_2 &\sim Cauchy(0, \lambda_0) \\
\Omega &\sim LKJcorr(\nu_0),
\end{aligned} \tag{5.13}$$

where $\beta_{i0}^{(1)}$ and $\beta_{i0}^{(2)}$ are random intercepts associated with biomarker 1 and biomarker 2, respectively. $\boldsymbol{\beta}_{i0}^{(L)}$ is a column vectors of size $2l_i \times 1$ that is a stack of $\beta_{i0}^{(1)}$ and $\beta_{i0}^{(2)}$ each repeated l_i times. Finally, $\boldsymbol{R}$ matrix was decomposed based on the equation (5.9).

Survival Model

Our goal is to quantify the association between the longitudinal biomarkers of interest and the time-to-event outcomes by directly adjusting for biomarkers measured values in a survival component of our proposed joint model. While usually biomarkers are measured on a discrete lab-visit basis (ex. every month), the event of interest happens on a continuous basis. While common frequentist models use the so-called "last-observation-carried" forward, by jointly modeling longitudinal-survival data, one can properly impute biomarker measures at each individual's event time. In particular and from the Bayesian modeling perspective, in each MCMC iteration, given the sampled parameters for each individual and by using the flexible multivariate Gaussian process in the longitudinal component of the model, there exists posterior trajectories of biomarkers for each individual. Our method, then, considers the posterior mean of those trajectories as the proposed trajectory for each individual's biomarker values over time at that iteration. The posterior mean trajectories of our biomarkers of interest, then, can be used to impute time-dependent biomarker covariates inside the survival component of the model.

In order to quantify the association between two longitudinal biomarkers, which are modeled simultaneously using the longitudinal component of the model, and the time-to-event outcomes, we define our survival component by using a multiplicative hazard model with the general form of

$$\lambda(T_i|\boldsymbol{Z_i}^{(s)}, \boldsymbol{Z_i}^{(L)}) = \lambda_0(T_i)exp\{\boldsymbol{\zeta}^{(s)}\boldsymbol{Z_i}^{(s)} + \boldsymbol{\zeta}^{(L)}\boldsymbol{Z_i}^{(L)}(\boldsymbol{t})\},$$

where T_i is the event time for subject i, $\lambda_0(T_i)$ denotes a baseline hazard function, $\boldsymbol{Z_i}^{(s)}$ is a vector of baseline covariates, $\boldsymbol{Z_i}^{(L)}$ are longitudinal covariates from the longitudinal component of the model, and $\boldsymbol{\zeta}^{(S)}$ and $\boldsymbol{\zeta}^{(L)}$ are regression coefficients interpretable as the log relative risk of "death" per every unit increase of their corresponding covariates.

Similar to the model in Chapter 4, we consider a Weibull distribution for the survival component to allow for log-linear changes in the baseline hazard function over time. Thus we assume

$$T_i \sim Weibull(\nu, \lambda_i), \tag{5.14}$$

that means

$$f(T_i|\nu, \lambda_i) = \nu {T_i}^{\nu-1} exp\big(\lambda_i - exp(\lambda_i){T_i}^{\nu}\big). \tag{5.15}$$

Weibull distribution is available in closed form and can be evaluated computationally efficiently. Under this parameterization of the Weibull distribution, covariates can be incorporated into the model by defining $\lambda_i = \boldsymbol{\zeta}^{(s)}\boldsymbol{Z_i}^{(s)} + \boldsymbol{\zeta}^{(L)}\boldsymbol{Z_i}^{(L)}$, where $\boldsymbol{\zeta}^{(s)}$ are coefficients associated with baseline survival covariates $\boldsymbol{Z_i}^{(s)}$, and $\boldsymbol{\zeta}^{(L)}$ are longitudinal coefficients associated with longitudinal covariates $\boldsymbol{Z_i}^{(L)}$.

In particular, we are interested in a model that directly includes the two longitudinal biomarker values at time t as a covariate inside the survival model. Hence, we define our model as

$$T_i|\nu, \boldsymbol{\zeta}^{(s)}, \zeta^{(X^{(1)})}, \zeta^{(X^{(2)})}, \boldsymbol{Z_i}^{(s)}, X_i^{(1)}(T_i), X_i^{(2)}(T_i) \sim Weibull(\nu, \lambda_i), \tag{5.16}$$

with

$$\lambda_i = \beta_{i0}^{(s)} + \boldsymbol{\zeta^{(s)}} \boldsymbol{Z_i^{(s)}} + \zeta^{(X^{(1)})} X_i^{(1)}(T_i) + \zeta^{(X^{(2)})} X_i^{(2)}(T_i),$$

where ν is a common shape parameter shared with all subjects. $\beta_{i0}^{(s)}$ is a subject specific coefficient in the model which allows subject-specific baseline hazard. $\boldsymbol{Z_i^{(s)}}$ and $\boldsymbol{\zeta^{(s)}}$ are baseline covariates and their corresponding coefficients, respectively. Finally, $\zeta^{(X^{(1)})}$ and $\zeta^{(X^{(2)})}$ are coefficients linking the longitudinal biomarker1 value $X_i^{(1)}(T_i)$ at time T_i and longitudinal biomarker2 value $X_i^{(2)}(T_i)$ at time T_i to the hazard for mortality, respectively.

In order to fit a fully joint longitudinal-survival model, at each iteration of the MCMC and for a time-point t^*, predicted biomarker1 and biomarker2 values for individual i are of the following form

$$\boldsymbol{X}^* | \boldsymbol{X_i}, \boldsymbol{t}, t^* \quad \sim \quad N_2(\boldsymbol{\mu}^*, \boldsymbol{\Sigma}^*),$$

where $\boldsymbol{X}$ is a 2×1 vector where its first element is the predicted value of the first biomarker and its second element is the predicted value for the second biomarker. $\boldsymbol{X_i}$ is a column vectors of size $2l_i \times 1$ that represents the observed biomarker values, where its first l_i elements are the observed biomarker1 values and the remaining elements are the observed biomarker2 values. $\boldsymbol{t}$ is column vector of size l_i that includes all time points at which values of the biomarkers were observed. t^* is the time at which by using the posterior trajectory of biomarkers at each MCMC iteration, we want to impute a predicted value per biomarker. Given our proposed longitudinal component setup, $\boldsymbol{X}$ is distributed bivariate Normal with mean $\boldsymbol{\mu}^*$ and with

covariance matrix $\boldsymbol{\Sigma}^*$ that are of the following forms

$$\boldsymbol{\mu^*} = \boldsymbol{\beta^L_{i0}} + \boldsymbol{K}(t^*, \boldsymbol{t})\boldsymbol{K}_X{}^{-1}(\boldsymbol{X_i} - \boldsymbol{\beta^{(L)}}_{i0}), \tag{5.17}$$

$$\boldsymbol{\Sigma}^* = \boldsymbol{K}(t^*, t^*) - \boldsymbol{K}(t^*, \boldsymbol{t})\boldsymbol{K}_X{}^{-1}\boldsymbol{K}(t^*, \boldsymbol{t})', \tag{5.18}$$

where

$$\begin{aligned} \boldsymbol{K}(t^*, \boldsymbol{t}) &= \boldsymbol{R} \otimes \kappa_i{}^2 e^{-\rho^2(t^* - \boldsymbol{t})^2}, \\ \boldsymbol{K}_X^{-1} &= (\boldsymbol{R} \otimes \boldsymbol{S}_i + \boldsymbol{\Sigma}_\epsilon)^{-1}, \\ \boldsymbol{K}(t^*, t^*) &= \boldsymbol{R} \otimes \boldsymbol{S}_i^*, \end{aligned} \tag{5.19}$$

where $\boldsymbol{R}$ and $\boldsymbol{S_i}$ were defined earlier in Section 5.2.4. $\boldsymbol{S}_i^*$ is defined similarly as $\boldsymbol{S}_i$ except that t is replaced with t^*. $\boldsymbol{\beta}^L_{i0}$ is a column vector size 2 by 1 with random intercept of the first biomarker, $\beta^{(1)}_{i0}$, as its first element and random intercept of the second biomarker, $\beta^{(2)}_{i0}$, as it's second element. $\boldsymbol{\beta^{(L)}}_{i0}$ is a column vector of size $2l_i \times 1$, where the first l_i elements are all the random intercept for the first biomarker and the remaining l_i elements are all the random intercept for the second biomarker.

In order to avoid an explicit distributional assumption on the survival times, we specify our survival model as an infinite mixture of Weibull distributions mixed on the $\beta^{(s)}_{i0}$ parameter.

To do so, we use the Dirichlet process mixture of Weibull distributions defined as

$$\begin{aligned} \beta_{i0}^{(s)} | \mu_i, \sigma^2_{\beta_0^{(s)}} &\sim N(\mu_i, \sigma^2_{\beta_0^{(s)}}), && (5.20) \\ \mu_i | G &\sim G, && (5.21) \\ G &\sim DP\big(\alpha^{(S)}, G_0\big), && (5.22) \end{aligned}$$

where $\sigma^2_{\beta_0^{(s)}}$ is a fixed parameter, μ_i is a subject-specific mean parameter from a distribution G with a DP prior, $\alpha^{(S)}$ is the concentration parameter of the DP and G_0 is the base distribution. By using the Dirichlet process prior on the distribution of $\beta_{i0}^{(s)}$, we allow patients with similar baseline hazards to cluster together which subsequently provides a stronger likelihood to estimate the baseline hazards. For other coefficients in the survival model, we assume a multivariate normal prior as

$$\big(\boldsymbol{\zeta^{(s)}}, \zeta^{X^{(1)}}, \zeta^{X^{(2)}}\big) \sim MVN(\mathbf{0}, \Sigma = {\sigma_0}^2 I),$$

where $\boldsymbol{\zeta^{(s)}}$ is a set of coefficients associated with the baseline survival covariates, $\zeta^{X^{(1)}}$ and $\zeta^{X^{(2)}}$ are coefficients associated with value of the first and the second biomarkers respectively, ${\sigma_0}^2$ is a prior variance for each coefficient, and I is the identity matrix.

For the shared shape parameter ν, we consider a log-Normal prior, $\nu \sim LogNormal(a_\nu, b_\nu)$, and specify the prior on the concentration parameter of our DP model to be $\alpha^{(S)} \sim \Gamma(a_\alpha^{(S)}, b_\alpha^{(S)})$.

Finally, since the hazard function includes time-varying covariates, evaluation of the log likelihood that involves integration of the hazard function overtime is done using the rectangular integration that was discussed in detail in Section 4.2.5.

The Posterior Distribution

Consider the joint longitudinal-survival likelihood function, $f_{\boldsymbol{L},\boldsymbol{S}}$, introduced in equation 5.5. Let $\boldsymbol{\omega}$ be a vector of all model parameters with the joint prior distribution $\pi(\boldsymbol{\omega})$. The posterior distribution of the parameter vector $\boldsymbol{\omega}$ can be written as

$$\pi(\boldsymbol{\omega}|\boldsymbol{X},\boldsymbol{Y}) \quad \propto \quad f_{L,S} \times \pi(\boldsymbol{\omega}), \tag{5.23}$$

where $\boldsymbol{X}$ and $\boldsymbol{Y}$ denote longitudinal and time-to-event data respectively, and $f_{L,S}$ is the joint model likelihood function (equation 5.5).

The posterior distribution of the parameters in our proposed joint model is not available in closed form. Hence, samples from the posterior distribution of the model parameters are obtained via Markov Chain Monte Carlo (MCMC) methods. In particular, we use the Hamiltonian Monte Carlo (Neal (2011)) to draw samples from the posterior distribution. Prior distributions on parameters of the joint model were explained in details under the longitudinal and survival component specification, and we assume independence among model parameters in the prior (ie. $\pi(\boldsymbol{\omega})$ is the product of the prior components specified previously). We provide further detail on less standard techniques for sampling from the posterior distribution when using a multivariate GP prior and we explain how to evaluate the survival portion of the likelihood function when time-varying covariates are incorporated into the model.

Evaluation of the Longitudinal Likelihood

Consider equation (5.11) and equation (5.12) where we introduced a flexible longitudinal model to simultaneously model multiple longitudinal biomarkers by using the Gaussian pro-

cess prior. By marginalizing over $\boldsymbol{W}_i$ in equation (5.11), one can show

$$\boldsymbol{X}_i|\boldsymbol{\beta}_{i0}^{(L)}, \sigma_1^2, \sigma_2^2 \sim N(\boldsymbol{\beta}_{i0}^{(L)}, \boldsymbol{R} \otimes \boldsymbol{S}_i + \boldsymbol{\Sigma}_\epsilon). \tag{5.24}$$

In order to sample from the posterior distribution of the parameters of the joint longitudinal-survival model introduced in Section 5.2.3, at each iteration of the MCMC, we need to compute the log posterior probability. Computing the log posterior probability involves evaluation of $log|\boldsymbol{R} \otimes \boldsymbol{S}_i + \boldsymbol{\Sigma}_\epsilon|$ and $(\boldsymbol{R} \otimes \boldsymbol{S}_i + \boldsymbol{\Sigma}_\epsilon)^{-1}$. This requires a memory space of $O({l_i}^2)$ and a computation time of $O({l_i}^3)$ per subject i. Consider matrix $\boldsymbol{S}_i = \kappa_i^2 \boldsymbol{K}_i$, where the $(i,j)^{th}$ element of $\boldsymbol{K}$ is $\boldsymbol{K}_i(i,j) = exp\{-\rho^2(t_{ij} - t_{ij'})^2\}$. $\boldsymbol{K}_i$ can be pre-computed prior to starting the MCMC process. Further, by using the eigen-value decomposition technique, one may make the calculation of the matrix determinant and inverse of the covariance matrix more computationally efficient. Using a similar idea proposed by Flaxman et al. (2015), we propose the following fast multivariate Gaussian process computation approach. We start pre-computing the $\boldsymbol{K}_i$ matrix. Also, we can pre-compute the eigen-vale decomposition of this matrix prior to starting the MCMC process. Consider an eigen-value decomposition of the following form

$$\boldsymbol{K}_i = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^T,$$

where $\boldsymbol{U}$ is a matrix of eigen-vectors and $\boldsymbol{\Lambda}$ is a diagonal matrix of eigen-values. For a scalar κ_i^2, the eigen-value decomposition of $\kappa_i^2 \boldsymbol{K}_i$ is of the form

$$\begin{aligned}\boldsymbol{S}_i &= \kappa_i^2 \boldsymbol{K}_i \\ &= \boldsymbol{U}(\kappa_i^2 \boldsymbol{\Lambda})\boldsymbol{U}^T.\end{aligned}$$

At each iteration of the MCMC, one can obtain the eigen-value decomposition of the matrix $\boldsymbol{R}$,that is of the form

$$\boldsymbol{R} = \boldsymbol{V}\boldsymbol{D}\boldsymbol{V}^T,$$

where $\boldsymbol{V}$ is a matrix of eigen-vectors and $\boldsymbol{D}$ is a diagonal matrix of eigen-values. One can then compute efficiently the log-determinant of the cross-covariance matrix, $log|\boldsymbol{R}\otimes\boldsymbol{S}_i+\boldsymbol{\Sigma}_\epsilon|$, as

$$\begin{aligned}
log|R \otimes S_i + \Sigma_\epsilon| &= log|(VDV^T) \otimes (U(\kappa_i^2\Lambda)U^T) + \Sigma_\epsilon| \\
&= log|(V \otimes U)(D \otimes \kappa_i^2\Lambda)(V \otimes U)^T + \Sigma_\epsilon| \\
&= log|(V \otimes U)(D \otimes \kappa_i^2\Lambda + \Sigma_\epsilon)(V \otimes U)^T| \\
&= log|(D \otimes \kappa_i^2\Lambda) + \Sigma_\epsilon| \\
&= 2l_i \sum_{k=1}^{2}\sum_{j=1}^{l_i} log(d_{kk}\lambda_{jj} + \sigma_1^2 I[k == 1] + \sigma_2^2 I[k == 2]),
\end{aligned} \tag{5.25}$$

and the inverse of the cross-covariance matrix, $(\boldsymbol{R}\otimes\boldsymbol{S}_i+\boldsymbol{\Sigma}_\epsilon)^{-1}$, can be efficiently computed as

$$\begin{aligned}
(R \otimes S_i + \Sigma_\epsilon)^{-1} &= \big((VDV^T) \otimes (U(\kappa_i^2\Lambda)U^T) + \Sigma_\epsilon\big)^{-1} \\
&= \big((V \otimes U)(D \otimes \kappa_i^2\Lambda)(V \otimes U)^T + \Sigma_\epsilon\big)^{-1} \\
&= \big((V \otimes U)(D \otimes \kappa_i^2\Lambda + \Sigma_\epsilon)(V \otimes U)^T\big)^{-1} \\
&= \big((V \otimes U)(D \otimes \kappa_i^2\Lambda + \Sigma_\epsilon)^{-1}(V \otimes U)^T\big),
\end{aligned} \tag{5.26}$$

In equation (5.26), computation of the inverse of the term $(D \otimes \kappa_i^2\Lambda + \Sigma_\epsilon)$ in the middle is very easy as it's a diagonal matrix. Using our proposed efficient computation technique introduced here, we noticed a 30 times faster computation speed in our simulations.

Evaluation of the Survival Likelihood

Similar to Chapter 4, we evaluate the survival likelihood using piece-wise integration. Consider the survival time for subject i that is denoted by t_i and is distributed according to a Weibull distribution with shape parameter τ and scale parameter $exp(\lambda_i)$, where $\lambda_i = \boldsymbol{\zeta^{(S)}} \boldsymbol{Z}_i^{(S)} + \boldsymbol{\zeta^{(L)}} \boldsymbol{Z}_i^{(L)}(t)$, where $\boldsymbol{Z}_i^{(S)}$ and $\boldsymbol{Z}_i^{(L)}(t)$ are vectors of covariates for subject i, with potentially time-varying covariates, corresponding to the survival and the longitudinal covariates respectively, and $\boldsymbol{\zeta^{(S)}}$ and $\boldsymbol{\zeta^{(L)}}$ are vectors of survival and longitudinal coefficients respectively. One can write the hazard function $h_i(t)$ as

$$h_i(t) = \tau t^{\tau-1} exp(\lambda_i - exp(\lambda_i) t^{\tau}). \tag{5.27}$$

The survival function $S_i(t)$ can be written as

$$S_i(t) = exp\{-\int_0^t h_i(w)dw\}.$$

Consider survival data on n subjects, some of whom may have been censored. Let event indicator δ_i that is 1 if the event is observed, and 0 otherwise. The survival likelihood contribution of subject i can be written in terms of the the hazard function $h_i(t)$ and the survival function $S_i(t)$ as

$$\begin{aligned} f_{S|L}^{(i)} &= h_i(t_i)^{\delta_i} S_i(t_i) \\ &= h_i(t_i)^{\delta_i} e^{-\int_0^{t_i} h_i(w)dw}. \end{aligned}$$

The overall survival log-likelihood can be written as

$$
\begin{aligned}
log(L) &= \sum_{i=1}^{n} log(f_{S|L}^{(i)}) \\
&= \sum_{i=1}^{n} \big(\delta_i log(h_i(t_i)) - \int_0^{t_i} h_i(w)dw\big).
\end{aligned}
$$

The hazard function in the equation (5.27) includes some time-varying covariates which often makes the integral of the hazard function non-tractable. In this case, one can estimate the integral using the rectangular integration as follows:

Integration of Survival Hazard with Time-Varying Covariates

1. Set a fixed number of rectangles m and set $A = 0$
2. Divide $(0, t_i)$ interval into m equal pieces each of length $L = t_i/m$

for $i \in \{1, \ldots, m\}$ **do**
  $t_{mid} \leftarrow L/2 + (i-1) * L$
  $A_{temp} \leftarrow L * h_i(t_{mid})$
  $A \leftarrow A + A_{temp}$
end for

5.2.5 A Mutivariate Gaussian Process Model for Modeling Non-Overlapping Biomarker Measures

Consider two longitudinal biomarkers $X^{(1)}$ and $X^{(2)}$, each with $l_i^{(1)}$ and $l_i^{(2)}$ longitudinal measures respectively. The obtained longitudinal measures need not to be taken at the same time for both biomarkers. The two biomarkers may or may not have any measurement time overlap. Consider biomarker observed time points of the form $t_{ij}^{(1)}, j = 1, 2, \ldots, l_i^{(1)}$ and $t_{ik}^{(2)}, k = 1, 2, \ldots, l_i^{(2)}$ for biomarker1 and biomarker2, respectively. Define $\tilde{t}_{ij}$ as a set of unique time points out of a pool of all biomarker observed times from both biomarkers. Define l_i as the number of unique time-points $\tilde{t}_{ij}$. It's obvious that $max\{l_i^{(1)}, l_i^{(2)}\} \leq l_i \leq l_i^{(1)} + l_i^{(2)}$. Also, out of all the l_i unique time points, $l_i^{(1)}$ one of them are the measurement time points where the biomarker 1 measure were obtained and $l_i^{(2)}$ of them are when the biomarker 2

measurements were obtained. For biomarker 1, biomarker measurements at the remaining $(l_i - l_i^{(1)})$ time points can be treated as missing values. Similarly, for biomarker 2, there are $(l_i - l_i^{(2)})$ missing biomarker 2 measured values. One can consider a similar model as in equation (5.13), where observed biomarkers column vector $\boldsymbol{X_i}$ includes $(l_i - l_i^{(1)})$ missing values for biomarker 1 and $(l_i - l_i^{(2)})$ missing values for biomarker2. Despite some missing biomarker measures at some time-points, cross-covariance function can be fully specified as it only depends on observed time points $\tilde{t}_i$ which are all observed.

Under the Bayesian inference, any missing data point can be represented as a parameter that can be estimated using posterior samples in the same way as any other parameter in the model Gelman et al. (2014). Hence, our modeling approach introduced in Section 5.2.3 is not limited at all to overlapping biomarker measures, and can cover a general case where obtained biomarker measures may or may not overlap in time. In non-overlapping case, additional missing biomarker values are introduced in the problem that are treated as parameters and can be easily estimated using posterior samples of those parameters.

5.3 Simulation Studies

In this section, we evaluate our proposed model using simulation studies. We first start by focusing on our proposed longitudinal modeling technique in order to model multiple longitudinal biomarkers simultaneously. One natural question is to ask whether there is any benefit in modeling longitudinal processes simultaneously by taking the correlation between processes into account as opposed to multiple independent longitudinal models each dealing with one longitudinal process. Next, we provide simulation studies on our proposed joint longitudinal-survival model.

5.3.1 Multivariate Gaussian Process Model vs. Multiple Univariate Gaussian Processes

We consider two longitudinal biomarkers of $\boldsymbol{X}^{(1)}$ and $\boldsymbol{X}^{(2)}$. By using summary statistics from real data from the United States Renal Data System, we generate synthetic data for 100 subjects where the first longitudinal biomarker resembles measured albumin biomarker values and the second longitudinal biomarker resembles BMI measured values. We generate synthetic biomarker 1 and biomarker 2 values each with 60 longitudinal albumin and 60 longitudinal BMI measures. Albumin and BMI values are simulated out of multivariate Gaussian process models of the form explained in the equation (5.24). The $kappa_i^2$ values are simulated from the $Uniform(0,1)$ distribution. $\beta_0^{albumin}$ and β_0^{BMI} are simulated from the $Normal(\mu = 5, \sigma = 1)$ and the $Normal(\mu = 20, \sigma = 2)$ distributions, respectively. Measurement errors $\sigma^2_{albumin}$ and σ^2_{BMI} are both set to be equal to 0.3.

Our goal is to now to compare a bivariate Gaussian process technique, which models the trajectory of both biomarkers simultaneously, with an alternative modeling approach that independently models each biomarker using a univariate Gaussian processes approach that was outlined in Chapter 4. Our general comparison scheme is to randomly remove some obtained values out of the biomarkers values and to compare the two modeling techniques in terms of the mean squared prediction error (MSE) of predicting the missing values. In particular, we consider three simulation scenarios below

1. **Scenario 1**: In this scenario, we assume both biomarkers have the same number of missing values. The interest is in comparing the two modeling techniques in terms of MSE and as a function of the correlation between the two processes and the amount of time overlap in missing values.

2. **Scenario 2**: In the second scenario, we assume that one biomarker is less-frequently observed compared to the other biomarker. The interest is in comparing the two models

in terms of MSE and as a function of the true correlation between the two processes and how less-frequent one process is compared to the other process.

3. **Scenario 3**: Finally, in the simulation scenario, we consider a case where the measure of biomarker volatility, $kappa_i^2$, is different between the two biomarker processes for each individual.

Scenario 1

Under this scenario, we remove 20 values out of each biomarker and treat them as missing. These 20 values are removed in three different ways as follows:

1. One when out of 60 time-points where we observed both biomarkers, we choose 20 time points and we remove both biomarkers at each of those 20 selected time points. In this case, missing values between the two biomarkers have 100% time-overlap.

2. Next, we choose 40 time-points out of the overall 60 observed time-points. In 20 of them, we remove the first biomarker values but not the second biomarker. In the remaining 20, we remove the second biomarker values but not the first one. In this case, missing values between the two biomarkers have 0% time-overlap.

3. Finally, we choose missing values so that 50% of the missing biomarker values have time-overlap. In this case, we choose 30 time points out of 60, in 10 of them we remove both biomarker values and treat them as missing. We choose another 10 time-points where we remove the first biomarker values but not the second biomarker values. In the remaining 10 time-points, we remove the second biomarker values but not the first biomarker values.

For each case, we consider 5 different correlation values between the two biomarker processes. We show results from 100 simulated datasets in each case. Table 5.1 compares the

average MSE between our proposed multivariate model (Multi) with the alternative way of independently modeling biomarkers using the univariate approach proposed in Chapter 4 (Uni). As results in Table 5.1 show, overall, modeling two longitudinal biomarkers using a multivariate GP, which takes the correlation between the biomarkers into account, leads to a smaller MSE. %Dec. indicates the percentage decrease in MSE comparing the multivariate model with the univariate model. As correlation between the two processes increases, the multivariate model leads to a smaller MSE of predicting missing biomarker values compared to the univariate model. As it's expected, the amount of overlap between missing values from the two processes has no effect on the univariate model. However, when we simultaneously model the two biomarkers and we take the correlation between the two biomarkers using the multivariate into account, information from an observed biomarker can help to better predict the missing values on the other biomarker. As the results in Table 5.1 show, when missing overlap decreases, which indicates that when a biomarker value is missing, the other biomarker is more likely to be observed, our proposed multivariate model can use the information on the non-missing biomarker to better predict the other missing biomarker, and hence, leads to smaller MSE values.

Scenario 2

Often times, due to the cost of the procedure or the difficulty of obtaining biomarker measures, there are some biomarkers that are measured less frequently compared to other biomarkers. Under this simulation scenario, we model two longitudinal biomarkers where one is less-frequently measured compare to the other biomarker. We compare our proposed multivariate longitudinal model with an alternative univariate model that models biomarkers independently. In particular, we consider three cases of one where per every 5 obtained measures of one biomarker, only one measure of the other biomarker is obtained (20%), the second case where per every two obtained measures of one biomarker, we obtain one measure

Correlation	100% Overlap			50% Overlap			0% Overlap		
ρ	Multi	Uni	%Dec.	Multi	Uni	%Dec.	Multi	Uni	%Dec.
0.1	0.237	0.243	2.5%	0.202	0.206	1.9%	0.207	0.212	2.4%
0.3	0.201	0.207	2.9%	0.215	0.227	5.3%	0.206	0.217	5.1%
0.5	0.209	0.217	3.7%	0.209	0.228	9.3%	0.196	0.227	13.7%
0.7	0.207	0.217	4.6%	0.189	0.228	17.1%	0.185	0.235	21.3%
0.9	0.194	0.204	4.9%	0.158	0.219	27.9%	0.139	0.237	41.4%

Table 5.1: Comparing our proposed multivariate longitudinal approach (Multi) to model two biomarkers by taking the correlation between those biomarkers into account as opposed to an alternative modeling approach that models each biomarker independently (Uni) using our proposed simulation scenario 1. ρ indicates the correlation between the two synthetic biomarker values. Each biomarker has 60 measured values where 20 of them are randomly selected to be missing in three different fashions of one when missing values between the two biomarkers have 100% time-overlap, another when missing values between the two biomarkers have 50% time-overlap, and lastly, when missing values between the two biomarkers have 0% time-overlap.

of the other biomarker (50%), and finally the third case where per every 5 obtained measures of one biomarker, we obtain four measures of the other biomarker (80%). We generate synthetic data for two biomarkers each with 60 measurements and at five different correlation levels between the two biomarker processes. We then randomly remove measurements from the first biomarker in order to simulate the 20%, 50%, and 80% cases explained above. Removed measurements are considered as missing values. Our goal is now to model the two biomarkers once jointly using our proposed multivariate longitudinal model and another time using two independent univariate models. Under each model, we predict the missing biomarker values and by comparing the true values and the predicted values, we compute mean squared error of prediction (MSE). MSE is then considered as a means to compare our proposed multivariate longitudinal model with an alternative univariate model that models each biomarker independently.

Table 5.2 compares the average MSE between our proposed multivariate longitudinal model (Multi) and an alternative univariate model (Uni). As the results in Table 5.2 show, the proposed multivariate model leads to lower MSEs compared to the univariate model. %Dec

indicates the percentage decreases in the MSE between the multivariate model and the univariate model. As the correlation between the two processes increases, we get higher prediction precision by using our proposed multivariate longitudinal model that is capable of taking the correlation between the processes into account and is capable of using the information on one longitudinal process to better impute the other process.

Correlation	20% Freq.			50% Freq.			80% Freq		
ρ	Multi	Uni	%Dec.	Multi	Uni	%Dec.	Multi	Uni	%Dec.
0.1	0.410	0.420	2.4%	0.281	0.286	1.8%	0.181	0.185	2.2%
0.3	0.391	0.423	7.6%	0.248	0.261	5.0%	0.188	0.197	4.6%
0.5	0.343	0.413	17.0%	0.226	0.264	14.4%	0.165	0.187	11.8%
0.7	0.319	0.407	21.6%	0.222	0.266	16.6%	0.152	0.188	19.2%
0.9	0.294	0.397	26.0%	0.190	0.245	22.5%	0.126	0.189	33.3%

Table 5.2: Comparing our proposed multivariate longitudinal approach (Multi) to model two biomarkers by taking the correlation between those biomarkers into account as opposed to an alternative modeling approach that models each biomarker independently (Uni) under the simulation scenario 2. ρ indicates the correlation between the two synthetic biomarker values. Measurements under one biomarker is obtained at a lower frequency compared to the second biomarker. The higher-frequency biomarker has 60 measurements whereas the second biomarker measurements are either at the 20% frequency, or at the 50% frequency, or at the 80% frequency. Also, we consider five different correlation levels as shown by ρ.

Scenario 3

In building our proposed multivariate longitudinal model, we assumed a separable cross-covariance structure. Our proposed model assumes that different longitudinal processes share the same temporal covariance. In particular, our model assumes a shared volatility measure κ_i^2 across different biomarkers. We, however, claim that despite a shared κ_i^2 across longitudinal biomarkers, differences in additional within-biomarker variances can be easily captured in our model through the $\boldsymbol{R}$ covariance matrix that is introduced earlier in equation (5.9). To show this, we simulate two longitudinal biomarkers of with different volatility measure κ_i^2 values. Obviously, an independent univariate Gaussian processes can estimate

different κ_i^2 values across different biomarker processes. Table 5.2 supports our claim that despite a shared κ_i^2 parameters in our proposed multivariate Gaussian process model, the different variability measures across the processes can be easily handled using the $\boldsymbol{R}$ covariance matrix. Similar to Table 5.1, in Table 5.3 we compare our proposed multivariate longitudinal Gaussian process model with a univariate longitudinal Gaussian process in terms of the mean squared error of the predicted values and as a function of the correlation and the percentage of time-overlap in missing values. Table 5.3 shows that our proposed multivariate longitudinal model is capable of handling processes with different volatility measures and leads to a smaller MSE compared to the independent univariate Gaussian models that model each biomarker independently of other present biomarkers.

Correlation	100% Overlap			50% Overlap			0% Overlap		
ρ	Multi	Uni	%Dec.	Multi	Uni	%Dec.	Multi	Uni	%Dec.
0.1	0.263	0.270	2.6%	0.250	0.260	3.9%	0.258	0.267	3.1%
0.3	0.247	0.256	3.6%	0.252	0.263	4.1%	0.249	0.265	5.9%
0.5	0.264	0.275	3.8%	0.232	0.250	6.9%	0.252	0.277	9.1%
0.7	0.236	0.247	4.4%	0.254	0.282	10.2%	0.207	0.243	14.7%
0.9	0.231	0.245	5.9%	0.211	0.254	16.8%	0.214	0.275	22.4%

Table 5.3: Comparing our proposed multivariate longitudinal approach (Multi) to model two biomarkers by taking the correlation between those biomarkers into account as opposed to an alternative modeling approach that models each biomarker independently (Uni) under the third proposed simulation scenario. ρ indicates the correlation between the two synthetic biomarker values. Each biomarker has 60 measured values where 20 of them are randomly selected to be missing in three different fashions of one when missing values between the two biomarkers have 100% time-overlap, another when missing values between the two biomarkers have 50% time-overlap, and lastly, when missing values between the two biomarkers have 0% time-overlap. Biomarkers were simulated with different volatility measures.

5.3.2 Simulation Studies Using the Proposed Joint Multivariate Longitudinal-Survival Model

In this section, we evaluate our proposed joint multivariate longitudinal-survival model using simulation study. We simulated 200 datasets that resembled the real data on end-stage renal disease patients that was obtained from the United States Renal Data System (USRDS). Each dataset included 300 subjects. We first simulated longitudinal trajectories for two biomarkers of $\boldsymbol{X}^{(1)}$ and $\boldsymbol{X}^{(2)}$, with 16 within subject biomarker 1 and 8 within subject biomarker 2 measures. Both biomarkers are generated from a joint multivariate Gaussian process with a high-correlation of 0.9 between the processes. In particular, biomarkers measures for each subject i are simulated from

$$\boldsymbol{X}_i|\boldsymbol{\beta}_{i0}^{(L)}, \sigma_1^2, \sigma_2^2 \sim N(\boldsymbol{\beta}_{i0}^{(L)}, \boldsymbol{R} \otimes \boldsymbol{S}_i + \boldsymbol{\Sigma}_\epsilon),$$

where $\boldsymbol{\beta}_{i0}^{(L)}$ is a stack of two subject specific random intercepts for the two processes. Subject-specific random intercepts for the first biomarker process are simulated from the Normal distribution $N(\mu = 5, \sigma = 1)$ and the random intercepts for the second biomarker are simulated from the Normal distribution $N(\mu = 20, \sigma = 2)$. σ_1^2 and σ_2^2 are both set to 0.3. $\boldsymbol{S}_i$ is considered to be equal to $\boldsymbol{S}_i = \kappa_i^2 \boldsymbol{K}_i$, where κ_i^2's are simulated from the uniform $U(min = 0, max = 1)$ distribution. $\boldsymbol{K}_i$ is the distance matrix of the form $exp\{-\rho^2(t_{ij} - t_{ij'})\}$, with a fixed ρ^2 of 0.1 and time points t_{ij}'s that are sampled uniformly from a followup time with a maximum of 15 months. The $\boldsymbol{R}$ matrix is the covariance matrix of the two biomarkers with a high correlation of 0.9 between the processes and with scaled variances for each biomarker process. Survival times are generated from a Weibull distribution of the form

$$T_i \sim Weibull(\nu, \lambda_i),$$

with the shape parameter of the distribution, ν, set to 1.5 and the log-scale parameter λ_i that is of the form

$$\lambda_i = \beta_{i0}^{(s)} + \zeta^{(X^{(1)})} X_i^{(1)}(T_i) + \zeta^{(X^{(2)})} X_i^{(2)}(T_i),$$

where $\beta_{i0}^{(s)}$ were sampled from an equally weighted mixture of two Normal distributions of $N(\mu = -1.5, \sigma = 1)$ and $N(\mu = 1.5, \sigma = 1)$. $\zeta^{(X^{(1)})}$ is fixed to 0.5 and $\zeta^{(X^{(2)})}$ is fixed to -0.3. Censoring times are generated from a uniform distribution and independent of the survival times T_i with parameters that ensure a 20% censoring rate.

We fit three models to our simulated data. We first fit a last-observation carried forward proportional hazard Cox model. Next, we fit our proposed joint longitudinal-survival that was introduced in Chapter 4, where the model uses a univariate longitudinal component that models each biomarker independently from the other biomarker. Finally, we fit our proposed joint longitudinal-survival model introduced in this chapter with a longitudinal component capable of modeling multiple biomarkers simultaneously by taking the correlation between the biomarkers into account.

We put the Normal prior $N(\mu = 5, \sigma = 2)$ and the Normal prior $N(\mu = 20, \sigma = 5)$ on the random intercepts for biomarker 1 and biomarker 2, respectively. We consider the $log - Normal(-1, 2)$ prior on κ_i^2. Also, we put the $log - Normal(-1, 1)$ prior on σ_1^2 and σ_2^2 parameters. We consider the $\boldsymbol{R}$ matrix to be of the form $\tau\Omega\tau$, where τ is a 2×2 matrix with diagonal elements of 1 and τ_2 and Ω is the correlation matrix between the two biomarkers that is of size 2×2. We consider a $Cauchy(0, 2.5)$ prior on τ_2 and an $LKJ(1)$ prior on Ω. We put the $log - Normal(0, 1)$ prior on the Weibull shape parameter ν. We also consider independent $N(\mu = 0, \sigma = 5)$ priors on $\zeta^{(X^{(1)})}$ and $\zeta^{(X^{(2)})}$. $\beta_{i0}^{(s)}$ are assumed to be distributed according to an unknown distribution G with the Dirichlet process $DP(\alpha, G_0)$ prior. We consider the $\Gamma(3, 3)$ prior on α and we consider G_0 to be the standard Normal distribution

$N(\mu = 0, \sigma = 1)$.

Table 5.4 shows the simulation results. As it was expected, the last observation carried forward Cox model leads to estimates that are shrunk towards 0 as this model is blind to the differential subject-specific baseline hazards that are induced by subject-specific $\beta_{i0}^{(s)}$ values. The estimates under the Cox model are marginalized over all subjects and due to non-collapsibility of these models, the estimates are shrunk toward the null. Further, this model carries the most recent longitudinal measures forward to the event time where as the longitudinal measures at the event time might be quite different from the most recent measures. This is also caused the estimates to shrink to ward 0. Next, our proposed joint longitudinal-survival model in Chapter 4 provides estimates that are closer to the true values compared to the Cox model as the model is capable of detecting baseline subject-specific β_i^0 values as well as providing a good prediction of the two biomarkers at time t by flexibly modeling the trajectory of the biomarkers. Third, our proposed joint multivariate longitudinal-survival model in this chapter, as an extension of our proposed model in Chapter 4, is capable of modeling multiple biomarkers simultaneously by taking the correlation between the biomarkers and hence, leads to even closer coefficient estimates compared to our joint univariate longitudinal-survival model.

Covariate of	True Conditional	LOCF Cox			Uni. Joint Model			Multi Joint Model		
Interest	Estimand	Mean	SD	MSE	Mean	SD	MSE	Mean	SD	MSE
Albumin(t)	0.5	0.242	0.089	0.077	0.443	0.124	0.012	0.481	0.101	0.009
BMI(t)	-0.3	-0.135	0.055	0.027	-0.278	0.127	0.003	-0.287	0.109	0.003

Table 5.4: A simulation study with 200 simulated longitudinal-survival datasets each with 300 subjects and two longitudinal biomarker processes one with 16 within subject measurements and another with 8 within subject measurements. We consider three models of the last-observation carried forward Cox model, our proposed joint univariate longitudinal-survival model (Ch. 4), and our proposed joint multivariate longitudinal-survival models. Coefficient estimates under these models are reported in the table along with the corresponding standard deviation and mean-squared error values per estimated coefficient.

5.4 Application of the Proposed Joint Multivariate Longitudinal Survival Model to DaVita Data on Hemodialysis Patients

In this section, we apply our proposed joint multivariate longitudinal-survival model to data on $n = 929$ hemodialysis patients. The data on these patients were obtained from a 5-year (January 2007-December 2011) cohort of patients who were treated for dialysis in dialysis clinics in the United States. For every participating patient in the study, up to 16 longitudinal albumin and 16 longitudinal calcium measures were taken uniformly over the five years of followup. The censoring rate in the data is 25.6%.

The analysis cohort used in the study were selected from a bigger cohort with 109,718 hemodialysis patients studied by Ravel et al. (2015) who had at least 8 longitudinal measures of albumin and at least 8 longitudinal measures of calcium. Also, In order to adjust for potential confounder factors, other than longitudinal albumin and calcium measures, our proposed model also includes age, sex, race, a baseline measure of phosphorus, and a baseline measure of iron.

In terms of the number of longitudinal albumin measures and longitudinal calcium measures, study subjects have at least 8 measures of each and at most 16 measures of each. Longitudinal measures of phosphorus and iron where also available, however, due to the small correlation between these biomarkers and albumin and calcium, we chose to only consider the baseline measure of phosphorus and iron. Table 5.5 shows the correlations between albumin, calcium, phosphorus, and iron.

Although the longitudinal albumin and calcium biomarkers are supposed to be measured during every lab visit, however, we noticed that in our study cohort, on average at 12.4% of lab visits neither of these two biomarkers were measured. We also noticed that on average

	Albumin	Calcium	Phosphorus	Iron
Albumin	1.000	0.462	0.175	0.257
Calcium	0.462	1.000	0.003	0.158
Phosphorus	0.175	0.003	1.000	0.071
Iron	0.257	0.158	0.071	1.000

Table 5.5: Correlations between the four biomarkers of albumin, calcium, phosphorus, and iron among the study cohort.

in 26.8% of lab visits there were no measured albumin biomarker and in 15.1% lab visits, there were no measured calcium biomarker.

With the aim of testing the association between mortality and the value of the biomarkers of interest, albumin and calcium, among hemodialysis patients, we analyze the data once using last-observation carried forward Cox model, another time using our proposed joint longitudinal-survival model that was introduced in Chapter 4 (Uni. Joint Model), and finally using our recent multivariate joint longitudinal-survival model (Multi. Joint Model).

Unlike the last observation carried forward Cox model that uses the most recent albumin and calcium biomarker values as the values of these biomarkers at each event time, our proposed joint longitudinal-survival models flexibly model the trajectory of these biomarkers over time and at each event time, the models impute the most relevant biomarker values according to the trajectories of those biomarkers. Further, unlike the Cox model that marginalize covariate effects across all subjects, our proposed joint models are capable of detecting differential subject-specific baseline hazards. Our joint longitudinal-survival model proposed in Chapter 4 models the longitudinal albumin and calcium processes independently of each other. Our proposed multivariate joint longitudinal-survival model introduced in this chapter, however, models the two processes simultaneously. Simultaneously modeling the two processes will allow a better longitudinal trajectory specification as the model can borrow information from measurements of one process when measurements of the other process are missing.

Table 5.6 shows the results of analyzing the data using the three models of last-observation

Covariates	No. of Cases	No. of Deaths	LOCF Cox Model Relative Risk (95% CI)		Uni. Joint Model Relative Risk (95% CR)	Multi. Joint Model Relative Risk (95% CR)
Age (10y)	929	691	1.33 (1.20,1.48)	<.001	1.50 (1.33,1.68)	1.56 (1.35,1.76)
Sex						
Men	546	412	1.0		1.0	1.0
Women	383	279	1.02 (0.78,1.32)	0.90	1.01 (0.72,1.41)	1.01 (0.73,1.41)
Race						
White	489	337	1.0		1.0	
Black	264	204	1.79 (0.90,3.58)	0.11	1.76 (0.82,3.69)	1.78 (0.84,3.76)
Hispanic	118	102	0.71 (0.40,1.25)	0.23	0.71 (0.39,1.26)	0.73 (0.37,1.28)
Other	58	48	0.55 (0.28,1.11)	0.10	0.57 (0.20,1.18)	0.58 (0.20,1.18)
Phosphorus (mg/dL)	929	691	1.07 (0.99,1.17)	0.08	1.10 (0.99,1.23)	1.10 (0.99,1.23)
Iron (g/dL)	929	691	0.99 (0.98,0.99)	0.002	0.98 (0.94,1.02)	0.98 (0.94,1.02)
Serum albumin(t) (1-g/dL decrement)	929	691	3.36 (2.64,4.29)	<0.0001	4.17 (2.78,5.72)	5.11 (3.86,6.29)
Calcium (mg/dL)	929	691	1.09 (0.87,1.37)	0.45	1.19 (0.72,1.86)	1.27 (0.92,1.69)

Table 5.6: Results of analyzing the association between the longitudinal albumin and calcium biomarkers and mortality among hemodialysis patients. A cohort of 929 hemodialysis subjects were followed over a maximal follow-up time of 5 years. Three separate models of last-observation carried forward Cox, univariate joint longitudinal-survival model, and multivariate joint longitudinal-survival model were fit to the data.

carried forward Cox, our proposed univariate joint longitudinal-survival model, our proposed multivariate joint longitudinal-survival model. The results from all three models consistently show that age and albumin value at the time of death are significant risk factors of mortality among hemodialysis patients. The results from the Cox model show that each 1 g/dL decrement in albumin level corresponds to 3.4 times higher risk of death. The risk of death per each 1 g/dL decrement in albumin is estimated to be 4.2 and 5.1 times higher under our proposed univariate joint model and multivariate joint model, respectively. Further, compared to the univariate joint model, the multivariate joint model leads to 32% and 36% reduction in the 95% credible region of the estimated effect of every one unit decrement in albumin and calcium, respectively. This observation was expected as the multivariate joint model by simultaneously modeling albumin and calcium trajectories, estimate those trajectories with higher precision compared with the univariate joint model.

5.5 Discussion

When monitoring the health of subjects, often times multiple risk factors are measured over time. Collected longitudinal risk factors are often correlated with each other as they are measures taken on the same subject. Modeling these longitudinal risk factors simultaneously where the correlation between the risk factors are taken into account can be beneficial, specially when there exists differential measuring density in the collected risk factors. Further, the association between the collected risk factors and the survival outcomes is often the practitioners' primary interest. In this chapter, we proposed a joint longitudinal-survival modeling framework with a longitudinal component capable of modeling multiple longitudinal processes simultaneously with the correlation between those processes taken into account. Our modeling framework is robust to common distributional assumptions as by using the Bayesian non-parameteric Gaussian process and Dirichlet process techniques, we avoid common functional and distributional assumptions in the model.

We used synthetic data in order to show the benefit of simultaneously modeling the trajectories of multiple longitudinal processes using our proposed multivariate longitudinal model as opposed to separate independent longitudinal models each modeling the trajectory of one longitudinal process independently from other longitudinal processes. Our findings show that a multivariate model has more precision in estimating the underlying trajectories of the longitudinal risk factors. Next, using synthetic data we showed that our proposed joint multivariate longitudinal-survival model in this chapter performs better in terms of mean-squared error of the estimated survival coefficients compared to the modeling framework introduced in Chapter 4 where the longitudinal biomarkers modeled independently.

In order to test the association between the longitudinal albumin and calcium biomarkers and mortality among hemodialysis patients, we used data on 929 hemodialysis patients. We analyzed the data using three models of last-observation carried forward Cox model, the

univariate joint longitudinal-survival model proposed in Chapter 4, and the multivariate longitudinal-survival model that was proposed in this chapter. While the results are consistent across all models, our proposed multivariate joint model that is capable of modeling the trajectory of longitudinal biomarkers with higher precision, leads to stronger estimated biomarker with higher precision for the estimated effect.

Our proposed modeling framework has some limitations. Our modeling framework is limited to the proportional hazards models only. Further, our method is computationally demanding and may not be scalable as number of subjects and within-subject measurements increase. In future, our modeling framework can be extended by relaxing the proportional hazard assumption on the survival component. Also, one by using alternatives to the conventional MCMC techniques, including variational methods can make our modeling framework more computationally efficient. In the next chapter, we will conclude the dissertation by summarizing our work and providing avenues for future work.

Chapter 6

Future Work

Our work in this thesis can be summarized as follows. In Chapter 3, we proposed Dirichlet process mixture models to model longitudinal data with latent sub-group random intercepts. Using extensive simulation studies, we showed that our proposed Dirichlet process mixture models compared to the common longitudinal models, perform the best in terms of the mean squared error of estimating conditional covariate effects. This is particularly useful in non-collapsible models including the logistic regression models and the proportional hazards models. Using sensitivity analyses, we showed that our proposed Dirichlet process mixture models are robust to the number of within-cluster measurements and to the underlying distribution of the latent cluster-specific random intercepts.

In Chapter 4, we extended our proposed Dirichlet process mixture proportional hazards model in Chapter 3 into a joint longitudinal-survival modeling framework. We proposed a modeling framework with a flexible longitudinal and survival component that avoid commonly assumed distributional assumptions in these models. Our proposed modeling framework proposes a stronger link between the longitudinal measures and the survival outcomes as our model is capable of adjusting the biomarker value at time t in the model, adjusting

for the average derivative of the biomarker trajectory over time, and adjusting for the summary measures of longitudinal trajectories including the higher order summary measure of volatility.

In Chapter 5, we extend our proposed joint longitudinal-survival model introduced in Chapter 4 into a joint longitudinal-survival model with a longitudinal component capable of simultaneously modeling the trajectories of multiple biomarkers by taking the correlation between biomarker processes into account. Using synthetic data we showed when multiple longitudinal biomarkers are collected, by simultaneously modeling these biomarkers, where the correlation between these biomarkers are taken into account, one can gain better precision in estimating the trajectory of each biomarker compared to an approach that models biomarkers independently. Our introduced modeling framework in Chapter 5 provides a strong tool for practitioners to test the association between multiple risk factors and survival outcomes.

Bayesian non-parameteric techniques, despite their flexibility, are often computationally demanding. Although MCMC methods are considered as gold standards as they provide asymptotically accurate posterior samples, however, they are not scalable with big data. Alternatively, many authors have proposed variational techniques, including the variational Gaussian process (VGP) technique (Tran et al. (2016)), the variational Bayesian inference to Dirichlet process mixture models (Blei and Jordan (2004)), the variational inference for Dirichlet process mixtures (Blei et al. (2006)), and the collapsed variational Dirichlet process mixture models (Kurihara et al. (2007)). Variational techniques are computationally more efficient than MCMC techniques, however, the accuracy of their posterior samples is debatable. Alternatively, Williamson et al. (2013) and Lovell et al. (2013) proposed a new parameterization of Dirichlet process mixture models, where they introduce conditional independence structure in the model with which one can run parallel MCMC methods. In future, by using alternatives to MCMC, our proposed methods can become more computationally

efficient and more scalable with large data.

Throughout the dissertation, our proposed models were based on the proportional hazards assumption. In future, our proposed models can be extended to also include non-proportional hazards models. De Iorio et al. (2009), proposed a non-proportional hazards survival model by using dependent Dirichlet process technique. Further, our proposed survival models can be extended to allow for covariates effect to vary over-times. Sebastien et al. (2008) proposed a random-split time approach to model the time-varying effect of kidney transplant on the risk of getting lymphoma. Similar ideas can be adapted in our modeling framework to allow for a more general survival model that does not rely on proportional hazards assumption and can include covariates with time-dependent effects.

Our proposed joint longitudinal-survival modeling framework provides a stronger link between longitudinal and survival data by introducing new ways of adjusting longitudinal covariates. We introduced new ideas on a joint longitudinal-survival model where we adjust for the biomarker values at time t, the average derivative of biomarkers, and the volatility of biomarkers. In future, one can extend our models by adjusting for other summary measures of biomarker trajectories. Additionally, when adjusting for biomarker values at time t, we assume the effect of biomarker values on the risk of death are immediate. In reality, however, there might be a time lag between the effect of a biomarker value and a risk of death. In future, one can incorporated this time-lag effect into our modeling framework, where the lag itself is treated as a parameter in the model and can be estimated using its posterior samples.

Bibliography

Abramowitz, M., Stegun, I. A., et al. (1966), "Handbook of mathematical functions," *Applied mathematics series*, 55, 62.

Andersen, P. K. and Gill, R. D. (1982), "Cox's regression model for counting processes: a large sample study," *The annals of statistics*, 1100–1120.

Antoniak, C. E. (1974), "Mixture of Dirichlet Process with Applications to Bayesian Nonparametric Problems," *Annals of Statistics*, 273(5281), 1152–1174.

Aslanidou, H., Dey, D. K., and Sinha, D. (1998), "Bayesian analysis of multivariate survival data using Monte Carlo methods," *Canadian Journal of Statistics*, 26, 33–48.

Banerjee, S., Carlin, B. P., and Gelfand, A. E. (2014), *Hierarchical modeling and analysis for spatial data*, Crc Press.

Berger, J. B. M. B. J., Dawid, A., Smith, D. H. A., and West, M. (2003), "Markov chain Monte Carlo-based approaches for inference in computationally intensive inverse problems," in *Bayesian Statistics 7: Proceedings of the Seventh Valencia International Meeting*, Oxford University Press, p. 181.

Bernardo, J., Berger, J., Dawid, A., Smith, A., et al. (1998), "Non-Stationary Spatial Modeling," .

Blackwell, D. and MacQueen, J. B. (1973), "Ferguson Distributions via Polya Urn Scheme," *Annals of Statistics*, 1, 353–355.

Blei, D. M. and Jordan, M. I. (2004), "Variational methods for the Dirichlet process," in *Proceedings of the twenty-first international conference on Machine learning*, ACM, p. 12.

Blei, D. M., Jordan, M. I., et al. (2006), "Variational inference for Dirichlet process mixtures," *Bayesian analysis*, 1, 121–144.

Breslow, N. E. and Clayton, D. G. (1993), "Approximate inference in generalized linear mixed models," *Journal of the American statistical Association*, 88, 9–25.

Brown, E. and Ibrahim, J. (2003), "A Bayesian semiparametric joint hierarchical model for longitudinal and survival data," *Biometrics*, 59, 221–228.

Bycott, P. and Taylor, J. (1998), "A comparison of smoothing techniques for CD4 data measured with error in a time-dependent Cox proportional hazards model," *Statistics in medicine*, 17, 2061–2077.

CARLSON, D. M., DUNCAN, D. A., Naessens, J. M., and JOHNSON, W. J. (1984), "Hospitalization in dialysis patients," in *Mayo Clinic Proceedings*, Elsevier, vol. 59, pp. 769–775.

Chung, S. H., Lindholm, B., and Lee, H. B. (2003), "Is malnutrition an independent predictor of mortality in peritoneal dialysis patients?" *Nephrology Dialysis Transplantation*, 18, 2134–2140.

Circle, T. H. B. (2016), *Digital image*, http://newhopeforkidneypatients.com/hemodialysis-and-peritoneal-dialysis-3/. N.

Clayton, D. G. (1991), "A Monte Carlo method for Bayesian inference in frailty models," *Biometrics*, 467–485.

Clinic, M. (2016), *Peritoneal Dialysis - Mayo Clinic*, http://www.mayoclinic.org/tests-procedures/peritoneal-dialysis/home/ovc-20202856.

Cox, D. R. (1972), "Regression Models and Life-Tables," *Journal of the Royal Statistical Society. Series B (Methodological)*, 34, 187–220.

— (1975), "Partial likelihood," *Biometrika*, 62, 269–276.

Dafni, U. G. and Tsiatis, A. A. (1998), "Evaluating surrogate markers of clinical outcome when measured with error," *Biometrics*, 1445–1462.

De Gruttola, V. and Tu, X. M. (1994), "Modelling progression of CD4-lymphocyte count and its relationship to survival time," *Biometrics*, 1003–1014.

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

Diggle, P., Ribeiro, P., and Geostatistics, M.-b. (2007), *Springer Series in Statistics*, Springer.

Doss, H. (1994), "Bayesian nonparametric estimation for incomplete data via successive substitution sampling," *The Annals of Statistics*, 1763–1786.

Doss, H. and Huffer, F. W. (2000), "Monte Carlo methods for Bayesian analysis of survival data using mixtures of Dirichlet priors," .

Doss, H. and Narasimhan, B. (1998), "Dynamic display of changing posterior in Bayesian survival analysis," in *Practical nonparametric and semiparametric Bayesian statistics*, Springer, pp. 63–87.

Eknoyan, G., Beck, G. J., Cheung, A. K., Daugirdas, J. T., Greene, T., Kusek, J. W., Allon, M., Bailey, J., Delmez, J. A., Depner, T. A., et al. (2002), "Effect of dialysis dose and membrane flux in maintenance hemodialysis," *New England Journal of Medicine*, 347, 2010–2019.

Escobar, M. D. (1994), "Estimating normal means with a Dirichlet process prior," *Journal of the American Statistical Association*, 89, 268–277.

Escobar, M. D. and West, M. (1995a), "Bayesian density estimation and inference using mixtures," *Journal of the american statistical association*, 90, 577–588.

— (1995b), "Bayesian density estimation and inference using mixtures," *Journal of American Statistical Society*, 90, 577–588.

Faucett, C. L. and Thomas, D. C. (1996), "Simultaneously modelling censored survival data and repeatedly measured covariates: a Gibbs sampling approach," *Statistics in medicine*, 15, 1663–1685.

Ferguson, T. S. (1973), "A Bayesian Analysis of Some Nonparametric Problems," *Annals of Statistics*, 1, 209–230.

Ferguson, T. S. and Phadia, E. G. (1979), "Bayesian nonparametric estimation based on censored data," *The Annals of Statistics*, 163–186.

Fitzmaurice, G. M., Laird, N. M., and Ware, J. H. (2012), *Applied longitudinal analysis*, vol. 998, John Wiley & Sons.

Flaxman, S., Gelman, A., Neill, D., Smola, A., Vehtari, A., and Wilson, A. G. (2015), "Fast hierarchical Gaussian processes," *Manuscript in preparation*.

Fung, F., Sherrard, D. J., Gillen, D. L., Wong, C., Kestenbaum, B., Seliger, S., Ball, A., and Stehman-Breen, C. (2002), "Increased Risk for Cardiovascular Mortality Among Malnourished End-stage Renal Disease Patients," *American Journal of Kidney Diseases*, 40, 307–314.

Gail, M. H. (1986), "Adjusting for covariates that have the same distribution in exposed and unexposed cohorts," *Modern statistical methods in chronic disease epidemiology*, 3–18.

Gail, M. H., Wieand, S., and Piantadosi, S. (1984), "Biased estimates of treatment effect in randomized experiments with nonlinear regressions and omitted covariates," *Biometrika*, 71, 431–444.

Gelfand, A. E. and Banerjee, S. (2010), "Multivariate spatial process models," *Handbook of Spatial Statistics*, 495–515.

Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. (2014), *Bayesian data analysis*, vol. 2, Chapman & Hall/CRC Boca Raton, FL, USA.

Greenland, S., Robins, J. M., and Pearl, J. (1999), "Confounding and collapsibility in causal inference," *Statistical Science*, 29–46.

Grzebyk, M. and Wackernagel, H. (1994), "Multivariate analysis and spatial/temporal scales: real and complex models," in *Proceedings of the XVIIth International Biometrics Conference*, Citeseer, vol. 1, pp. 19–33.

Gustafson, P. (1997), "Large hierarchical Bayesian analysis of multivariate survival data," *Biometrics*, 230–242.

Habach, G., Bloembergen, W. E., Mauger, E. A., Wolfe, R. A., and Port, F. K. (1995), "Hospitalization among United States dialysis patients: hemodialysis versus peritoneal dialysis." *Journal of the American Society of Nephrology*, 5, 1940–1948.

Hanson, T. E., Branscum, A. J., and Johnson, W. O. (2011a), "Predictive comparison of joint longitudinal-survival modeling: a case study illustrating competing approaches," *Lifetime data analysis*, 17, 3–28.

— (2011b), "Predictive comparison of joint longitudinal-survival modeling: a case study illustrating competing approaches," *Lifetime data analysis*, 17, 3–28.

Hawkins, J. W. and Dugaiczyk, A. (1982), "The human serum albumin gene: structure of a unique locus," *Gene*, 19, 55–58.

Hogan, J. W. and Laird, N. M. (1997), "MODEL-BASED APPROACHES TO ANALYSING INCOMPLETE LONGITUDINAL AND FAILURE TIME DATA," *Statistics in medicine*, 16, 259–272.

Holsclaw, T., Shahbaba, B., and Gillen, D. (2014), "Quantifying the Association Between Within-Subject Volatility in Serum Albumin and Mortality via a Gaussian Process Model," *Submitted: Journal of the American Statistical Association*.

Ibrahim, J. G., Chen, M.-H., and Sinha, D. (2005), *Bayesian survival analysis*, Wiley Online Library.

Kalantar-Zadeh, K., Ikizler, T. A., Block, G., Avram, M. M., and Kopple, J. D. (2003), "Malnutrition-inflammation complex syndrome in dialysis patients: causes and consequences," *American Journal of Kidney Diseases*, 42, 864–881.

Kalantar-Zadeh, K. and Kopple, J. D. (2001), "Relative contributions of nutrition and inflammation to clinical outcome in dialysis patients," *American journal of kidney diseases*, 38, 1343–1350.

Kalantar-Zadeh, K., Kopple, J. D., Block, G., and Humphreys, M. H. (2001), "Association among SF36 quality of life measures and nutrition, hospitalization, and mortality in hemodialysis," *Journal of the American Society of Nephrology*, 12, 2797–2806.

Kopple, J. D. (1997), "Nutritional Status as a Predictor of Morbidity in Maintenance Dialysis Patients." *Asaio Journal*, 43, 246–250.

Kopple, J. D., Zhu, X., Lew, N. L., and Lowrie, E. G. (1999), "Body weight-for-height relationships predict mortality in maintenance hemodialysis patients," *Kidney international*, 56, 1136–1148.

Kurihara, K., Welling, M., and Teh, Y. W. (2007), "Collapsed Variational Dirichlet Process Mixture Models." in *IJCAI*, vol. 7, pp. 2796–2801.

Laird, N. M. and Ware, J. H. (1982), "Random-effects models for longitudinal data," *Biometrics*, 963–974.

Law, N. J., Taylor, J. M., and Sandler, H. (2002a), "The joint modeling of a longitudinal disease progression marker and the failure time process in the presence of cure," *Biostatistics*, 3, 547–563.

— (2002b), "The joint modeling of a longitudinal disease progression marker and the failure time process in the presence of cure," *Biostatistics*, 3, 547–563.

Lee, E. W., Wei, L., Amato, D. A., and Leurgans, S. (1992), "Cox-type regression analysis for large numbers of small groups of correlated failure time observations," in *Survival analysis: state of the art*, Springer, pp. 237–247.

Liang, K.-Y. and Zeger, S. L. (1986), "Longitudinal data analysis using generalized linear models," *Biometrika*, 73, 13–22.

Liu, Q. and Pierce, D. A. (1994), "A note on Gauss?Hermite quadrature," *Biometrika*, 81, 624–629.

Lo, A. Y. et al. (1984), "On a class of Bayesian nonparametric estimates: I. Density estimates," *The annals of statistics*, 12, 351–357.

Lovell, D., Malmaud, J., Adams, R. P., and Mansinghka, V. K. (2013), "Clustercluster: parallel Markov chain Monte Carlo for Dirichlet process mixtures," *arXiv preprint arXiv:1304.2302*.

MacEachern, S. N. and Müller, P. (1998), "Estimating mixture of Dirichlet process models," *Journal of Computational and Graphical Statistics*, 7, 223–238.

Majumdar, A. and Gelfand, A. E. (2007), "Multivariate spatial modeling for geostatistical data using convolved covariance functions," *Mathematical Geology*, 39, 225–245.

Martinussen, T. and Vansteelandt, S. (2013), "On collapsibility and confounding bias in Cox and Aalen regression models," *Lifetime Data Analysis*, 19, 279–96.

Matheron, G. (1973), "The intrinsic random functions and their applications," *Advances in applied probability*, 439–468.

McDonald, S. P., Marshall, M. R., Johnson, D. W., and Polkinghorne, K. R. (2009), "Relationship between dialysis modality and mortality," *Journal of the American Society of Nephrology*, 20, 155–163.

Neal, R. M. (1992), "Bayesian mixture modeling," in *Maximum Entropy and Bayesian Methods*, Springer, pp. 197–211.

— (2000), "Markov Chain Sampling Methods for Dirichlet Process Mixture Models," *Journal of Computational and Graphical Statistics*, 9, 249–265.

— (2011), *Handbook of Markov Chain Monte Carlo*, Chapman & Hall CRC.

— (2012), *Bayesian learning for neural networks*, vol. 118, Springer Science & Business Media.

O'Hagan, A. and Kingman, J. (1978), "Curve fitting and optimal design for prediction," *Journal of the Royal Statistical Society. Series B (Methodological)*, 1–42.

Prentice, R. (1982), "Covariate measurement errors and parameter estimation in a failure time regression model," *Biometrika*, 69, 331–342.

R Brown, E. and G Ibrahim, J. (2003), "A Bayesian semiparametric joint hierarchical model for longitudinal and survival data," *Biometrics*, 59, 221–228.

Rasmussen, C. E. (2006), "Gaussian processes for machine learning," .

Rasmussen, C. E. and Williams, C. K. I. (2006), *Gaussian Processes for Machine Learning*, MIT Press, 2nd ed.

Ravel, V., Streja, E., Molnar, M. Z., Rezakhani, S., Soohoo, M., Kovesdy, C. P., Kalantar-Zadeh, K., and Moradi, H. (2015), "Association of aspartate aminotransferase with mortality in hemodialysis patients," *Nephrology Dialysis Transplantation*, gfv310.

Sahu, S. K., Dey, D. K., Aslanidou, H., and Sinha, D. (1997), "A Weibull regression model with gamma frailties for multivariate survival data," *Lifetime data analysis*, 3, 123–137.

Sargent, D. J. (1998), "A general framework for random effects survival analysis in the Cox proportional hazards setting," *Biometrics*, 1486–1497.

Schluchter, M. D. (1992), "Methods for the analysis of informatively censored longitudinal data," *Statistics in medicine*, 11, 1861–1870.

Schluchter, M. D., Greene, T., and Beck, G. J. (2001), "Analysis of change in the presence of informative censoring: application to a longitudinal clinical trial of progressive renal disease," *Statistics in medicine*, 20, 989–1007.

Schmidt, A. M. and Gelfand, A. E. (2003), "A Bayesian coregionalization approach for multivariate pollutant data," *Journal of Geophysical Research: Atmospheres*, 108.

Sebastien, J.-P. H., Rudser, K. D., and Gillen, D. L. (2008), "The separation of timescales in Bayesian survival modeling of the time-varying effect of a time-dependent exposure," *Biostatistics*, 9, 400–410.

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

Sinha, D. and Dey, D. K. (1997), "Semiparametric Bayesian analysis of survival data," *Journal of the American Statistical Association*, 92, 1195–1212.

Song, X., Davidian, M., and Tsiatis, A. A. (2002), "An estimator for the proportional hazards model with multiple longitudinal covariates measured with error," *Biostatistics*, 3, 511–528.

Struthers, C. and Kalbfliesch, J. (1986), "An estimator for the proportional hazards model with multiple longitudinal covariates measured with error," *Biometrika*, 73, 363–9.

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

Teh, Y. W. (2011), "Dirichlet process," in *Encyclopedia of machine learning*, Springer, pp. 280–287.

Tran, D., Ranganath, R., and Blei, D. M. (2016), "THE VARIATIONAL GAUSSIAN PROCESS," *stat*, 1050, 23.

Tsiatis, A., Degruttola, V., and Wulfsohn, M. (1995), "Modeling the relationship of survival to longitudinal data measured with error. Applications to survival and CD4 counts in patients with AIDS," *Journal of the American Statistical Association*, 90, 27–37.

Tsiatis, A. A. and Davidian, M. (2001), "A semiparametric estimator for the proportional hazards model with longitudinal covariates measured with error," *Biometrika*, 88, 447–458.

— (2004), "Joint modeling of longitudinal and time-to-event data: an overview," *Statistica Sinica*, 809–834.

Vaupel, J. W., Manton, K. G., and Stallard, E. (1979), "The impact of heterogeneity in individual frailty on the dynamics of mortality," *Demography*, 16, 439–454.

Vendrely, B., Chauveau, P., Barthe, N., El Haggan, W., Castaing, F., De Précigout, V., Combe, C., and Aparicio, M. (2003), "Nutrition in hemodialysis patients previously on a supplemented very low protein diet," *Kidney international*, 63, 1491–1498.

Ver Hoef, J. M., Cressie, N., and Barry, R. P. (2004), "Flexible spatial models for kriging and cokriging using moving averages and the Fast Fourier Transform (FFT)," *Journal of Computational and Graphical Statistics*, 13, 265–282.

Wahba, G. (1990), *Spline models for observational data*, vol. 59, Siam.

Wang, Y. and Taylor, J. M. G. (2001a), "Jointly modeling longitudinal and event time data with application to acquired immunodeficiency syndrome," *Journal of the American Statistical Association*, 96, 895–905.

— (2001b), "Jointly modeling longitudinal and event time data with application to acquired immunodeficiency syndrome," *Journal of the American Statistical Association*, 96, 895–905.

Williamson, S., Dubey, A., and Xing, E. P. (2013), "Parallel Markov Chain Monte Carlo for Nonparametric Mixture Models." in *ICML (1)*, pp. 98–106.

Wulfsohn, M. S. and Tsiatis, A. A. (1997a), "A joint model for survival and longitudinal data measured with error," *Biometrics*, 330–339.

— (1997b), "A joint model for survival and longitudinal data measured with error," *Biometrics*, 330–339.

Yadrenko, M. (1987), "Correlation theory of stationary and related random functions, Vol I, Basic results," .

Zeger, S. L., Liang, K.-Y., and Albert, P. S. (1988), "Models for longitudinal data: a generalized estimating equation approach," *Biometrics*, 1049–1060.

Appendix A

Programming Codes

Here I provide the codes I wrote to fit Bayesian models used in this dissertation.

A.1 Chapter 3 Models

A.1.1 Bayesian Logistic Models

Bayesian Logistic Regression - STAN

```
data {
  int<lower = 1> N;
  int<lower = 1> Mi;
  int<lower = 0, upper = 1> Y[N*Mi];
  matrix[N, Mi] t;
}
parameters {
```

```
  real beta0;
  real beta1;
}
model {
  //int<lower = 0,upper = 1> YTmp[Mi];
  beta0 ~ normal(0, 1);
  beta1 ~ normal(0, 1);

  for (i in 1:N){
    for (j in 1:Mi){
      Y[(i - 1)*Mi + j] ~ bernoulli_logit(beta0 + beta1*t[i,j]);
    }
  }
}
```

Hierarchical Bayesian Logistic Regression - STAN

```
data {
  int<lower = 1> N; // number of subjects
  int<lower = 1> Mi; // number of longitudinal measures per subject
  int<lower = 0,upper = 1> Y[N*Mi]; // Long. Time
  matrix[N, Mi] t; // albumin val.
}
parameters {
  vector[N] beta0_i;
  real beta0;
  real beta1;
}
model {
  beta0 ~ normal(0, 1);
  beta1 ~ normal(0, 1);
  beta0_i ~ normal(0, 1);

  for (i in 1:N){
    for (j in 1:Mi){
      Y[(i - 1)*Mi + j] ~ bernoulli_logit(beta0_i[i] + beta0 + beta1*t[i,j]);
    }
  }
}
```

Dirichlet Process Mixture Logistic Regression (Mean Mixture) - JAGS

```
model {
  B[1] <- b[1]
  for (k in 2:K){
    B[k] <- b[k]*((1- b[k-1])*B[k-1]/b[k-1])
  }
  B.sum <- sum(B)

  for (k in 1:K){
    b[k] ~ dbeta(1, 1)
    BetaProb[k] <- B[k]/B.sum
  }

  for (k in 1:K){
    muDP[k] ~ dnorm(0, 0.01)
  }

  for (n in 1:N){
    z[n] ~ dcat(BetaProb)
  }

  for (i in 1:N){
    for (j in 1:Mi){
      Y[i,j] ~ dbern(mu[i,j])
      mu[i,j] <- 1/(1 + exp(-(beta0_i[i] + beta0 + beta1*t[i,j])))
    }
```

```
  }

  for (i in 1:N){
    beta0_i[i] ~ dnorm(Mu_i[i], 1)
    Mu_i[i] <- muDP[z[i]]
  }

  beta0 ~ dnorm(0, 0.01)
  beta1 ~ dnorm(0, 0.01)
}
```

Dirichlet Process Mixture Logistic Regression (Sigma Mixture) - JAGS

```
model {
  B[1] <- b[1]
  for (k in 2:K){
    B[k] <- b[k]*((1- b[k-1])*B[k-1]/b[k-1])
  }
  B.sum <- sum(B)

  for (k in 1:K){
    b[k] ~ dbeta(1, 1)
    BetaProb[k] <- B[k]/B.sum
  }

  for (k in 1:K){
    sigma2DP[k] ~ dlnorm(0, 5)
  }

  for (n in 1:N){
    z[n] ~ dcat(BetaProb)
  }

  for (i in 1:N){
    for (j in 1:Mi){
      Y[i,j] ~ dbern(mu[i,j])
      mu[i,j] <- 1/(1 + exp(-(beta0_i[i] + beta0 + beta1*t[i,j])))
    }
```

```
  }

  for (i in 1:N){
    beta0_i[i] ~ dnorm(0, tau0_i[i])
    tau0_i[i] <- 1/sigma2DP[z[i]]
  }

  beta0 ~ dnorm(0, 0.01)
  beta1 ~ dnorm(0, 0.01)
}
```

A.1.2 Bayesian Survival Models

Bayesian Survival Marginal Model - STAN

```
data {
  int<lower = 1> N_times_Mi; // number of subjects
  int<lower = 1> Num_obs; // Num of rows with nu = 1
  int<lower = 1> Num_cens; // Num of rows with nu = 0
  vector[Num_obs] X_obs;
  vector[Num_cens] X_cens;
  vector[Num_obs] t_obs;
  vector[Num_cens] t_cens; // these are the current censoring times
  vector[Num_obs] subjIDs_obs; // censoring indicator!
  vector[Num_cens] subjIDs_cens; // censoring indicator!
}
parameters {
  vector<lower=1>[Num_cens] myPrimitiveVec;
  real<lower = 0> tau;
  real beta0;
  real beta1;
}
transformed parameters{
  vector[Num_cens] t_actual_cens;
  t_actual_cens <- myPrimitiveVec .* t_cens;
}
model {
  real lambda_Tmp;
```

```
  tau ~ lognormal(-2, 3);
  beta0 ~ normal(0, 1);
  beta1 ~ normal(0, 1);

  // likelihood - observed times:
  for (i in 1:Num_obs){
    lambda_Tmp <- beta0 +  beta1*X_obs[i];
    increment_log_prob(log(tau) + (tau - 1)*log(t_obs[i]) + lambda_Tmp
                       - exp(lambda_Tmp)*(pow(t_obs[i], tau)));
  }

  // likelihood - censored times:
  for (i in 1:Num_cens){
    lambda_Tmp <- beta0 + beta1*X_cens[i];
    increment_log_prob(log(tau) + (tau - 1)*log(t_actual_cens[i])
                       + lambda_Tmp - exp(lambda_Tmp)*(pow(t_actual_cens[i], tau)));
  }
}
```

Hierarchical Bayesian Survival Model - STAN

```
data {
  int<lower = 1> N; // number of subjects
  int<lower = 1> N_times_Mi; // number of rows in the survival data
  int<lower = 1> Num_obs; // Num of rows with nu = 1
  int<lower = 1> Num_cens; // Num of rows with nu = 0
  vector[Num_obs] X_obs;
  vector[Num_cens] X_cens;
  vector[Num_obs] t_obs;
  vector[Num_cens] t_cens; // these are the current censoring times
  int<lower = 1> subjIDs_obs[Num_obs]; // censoring indicator!
  int<lower = 1> subjIDs_cens[Num_cens]; // censoring indicator!
}
parameters {
  vector<lower=1>[Num_cens] myPrimitiveVec;
  real<lower = 0> tau;
  vector[N] beta0_i;
  real beta0;
  real beta1;
}
transformed parameters{
  vector[Num_cens] t_actual_cens;
  t_actual_cens <- myPrimitiveVec .* t_cens;
}
model {
  real lambda_Tmp;
```

```
  tau ~ lognormal(-2.0, 3.0);
  beta0_i ~ normal(0, 1.0);
  beta0 ~ normal(0.0, 1.0);
  beta1 ~ normal(0.0, 1.0);

  // likelihood - observed times:
  for (i in 1:Num_obs){
    lambda_Tmp <- beta0_i[subjIDs_obs[i]] + beta0 +  beta1*X_obs[i];
    increment_log_prob(log(tau) + (tau - 1)*log(t_obs[i]) +
                lambda_Tmp - exp(lambda_Tmp)*(pow(t_obs[i], tau)));
  }

  // likelihood - censored times:
  for (i in 1:Num_cens){
    lambda_Tmp <- beta0_i[subjIDs_cens[i]] + beta0 + beta1*X_cens[i];
    increment_log_prob(log(tau) + (tau - 1)*log(t_actual_cens[i]) +
          lambda_Tmp - exp(lambda_Tmp)*(pow(t_actual_cens[i], tau)));
  }
}
```

Dirichlet Process Mixture Survival (Mean Mixture) - JAGS

```
model {
  B[1] <- b[1]
  for (k in 2:K){
    B[k] <- b[k]*((1- b[k-1])*B[k-1]/b[k-1])
  }
  B.sum <- sum(B)

  for (k in 1:K){
    b[k] ~ dbeta(1, 1)
    BetaProb[k] <- B[k]/B.sum
  }

  for (k in 1:K){
    muDP[k] ~ dnorm(0, 0.01)
  }

  for (n in 1:N){
    z[n] ~ dcat(BetaProb)
  }

  for (i in 1:N){
    for (j in 1:Mi){
      is.censored[i,j] ~ dinterval(t.to.death[i,j], t.cen[i,j])
      t.to.death[i,j] ~ dweib(tau, lambda[i,j])
      log(lambda[i,j]) <- beta0_i[i] + beta0 + beta1*X[i]
```

```
    }
  }

  for (i in 1:N){
    beta0_i[i] ~ dnorm(Mu_i[i], 1)
    Mu_i[i] <- muDP[z[i]]
  }

  beta0 ~ dnorm(0, 0.01)
  beta1 ~ dnorm(0, 0.01)
  tau ~ dlnorm(1, 2)
}
```

Dirichlet Process Mixture Survival (Sigma Mixture) - JAGS

```
model {
  B[1] <- b[1]
  for (k in 2:K){
    B[k] <- b[k]*((1- b[k-1])*B[k-1]/b[k-1])
  }
  B.sum <- sum(B)

  for (k in 1:K){
    b[k] ~ dbeta(1, 1)
    BetaProb[k] <- B[k]/B.sum
  }

  for (k in 1:K){
    sigma2DP[k] ~ dlnorm(0, 5)
  }

  for (n in 1:N){
    z[n] ~ dcat(BetaProb)
  }

  for (i in 1:N){
    for (j in 1:Mi){
      is.censored[i,j] ~ dinterval(t.to.death[i,j], t.cen[i,j])
      t.to.death[i,j] ~ dweib(tau, lambda[i,j])
      log(lambda[i,j]) <- beta0_i[i] + beta0 + beta1*X[i]
```

```
    }
  }

  for (i in 1:N){
    beta0_i[i] ~ dnorm(0, tau0_i[i])
    tau0_i[i] <- 1/sigma2DP[z[i]]
  }

  beta0 ~ dnorm(0, 0.01)
  beta1 ~ dnorm(0, 0.01)
  tau ~ dlnorm(1, 2)
}
```

A.2 Chapter 4 Models

Our programming code for our proposed models in Chapter 4 include three main parts:

1. DP Sampler: Our code to sample from the posterior of Dirichlet process model.
2. HMC code: all sampling is done using our own HMC sampler.
3. Stan code: We used Stan software to calculation the log posterior posterior probability and the gradients required by our HMC sampler.

A.2.1 DP Sampler

Our code to sample from posterior of Dirichlet process using Neal's alg. 8 (Neal, R. M. (2000)).

```
get.DP.alpha.L <- function(nSub, DP.alpha.curr, DPM_MU.star,
                           DP.alpha.a, DP.alpha.b){
  k <- length(DPM_MU.star)
  M <- nSub
  eta <- rbeta(1, DP.alpha.curr + 1, M)
  Pi.eta <- (DP.alpha.a + k - 1)/( (DP.alpha.a + k - 1)
                                   + (M*(DP.alpha.b - log(eta))) )
  sampProb <- c(Pi.eta, (1 - Pi.eta))
  DP.alpha.pool <- c(rgamma(1, shape = (DP.alpha.a + k),
                            scale = 1/(DP.alpha.b - log(eta)) ),
                     rgamma(1, shape = (DP.alpha.a + k - 1),
                            scale = 1/(DP.alpha.b - log(eta)) ) )
  DP.alpha.new <- sample(DP.alpha.pool, 1, prob = sampProb)
```

```
  return(DP.alpha.new)
}

# We borrowed ideas from Dr. Shahbaba in writing this code.
get.Mui <- function(nSub, Beta0_current, Mu.star, indSet,
                    freqSet, DP.alpha, DP.G0.mean,
                    DP.G0.sd, sd_for_beta0_i_S){
  M = 5
  for (i in 1:nSub){
    phi <- NULL
    curInd <- indSet[i]
    freqSet[curInd] <- freqSet[curInd] - 1
    if (freqSet[curInd] == 0){
      phi <- Mu.star[curInd]
      freqSet <- freqSet[-(curInd)]
      Mu.star <- Mu.star[-(curInd)]
      indSet[indSet > curInd] <- indSet[indSet > curInd] - 1
      kMinus <- length(freqSet)
      Mu.star[kMinus + 1] <- phi
      for (m in 1:(M - 1)){
        Mu.star[kMinus + 1 + m] <- rnorm(1, mean = DP.G0.mean,
                                         sd = DP.G0.sd)
      }

    }else{
      kMinus <- length(freqSet)
      for (m in 1:M){
```

```
      Mu.star[kMinus + m] <- rnorm(1, mean = DP.G0.mean,
                                   sd = DP.G0.sd)
    }
  }
  # q1: vector of sampling prob for exisiting parameters
  # q2: vector of sampling prob for auxiliary parameters

  # Prob. of Sampling Existing Parameters:
  # Here, we are looking at beta0_i for subj i to write the likelihood
  q1 <- rep(0, kMinus)
  for (k in 1:kMinus){
    q1[k] <- dnorm(Beta0_current[i], mean = Mu.star[k],
                   sd = sd_for_beta0_i_S, log = T)
  }
  q1 <- q1 + log(freqSet) - log(nSub - 1 + DP.alpha)
  q2 <- rep(0, M)
  for (k in 1:M){
    q2[k] <- dnorm(Beta0_current[i], mean = Mu.star[kMinus + k]
                   , sd = sd_for_beta0_i_S, log = T)
  }
  q2 <- q2 + log(DP.alpha/M) - log(nSub - 1 + DP.alpha)
  myProb <- c(q1, q2)
  myProb.Max <- max(myProb)
  myProb.Rel <- myProb - myProb.Max
  myProb <- exp(myProb.Rel)
  myProb <- myProb/sum(myProb)
  picked <- which(rmultinom(1, 1, myProb) == 1)
```

```
    if (picked <= kMinus){
      indSet[i] <- picked
      freqSet[picked] <- freqSet[picked] + 1
      Mu.star <- Mu.star[-( (kMinus + 1):(kMinus + M) )]
      myProb <- myProb[-( (kMinus + 1):(kMinus + M) )]
    }else{
      indSet[i] <- kMinus + 1
      freqSet <- c(freqSet, 1)
      phi <- Mu.star[picked]
      Mu.star <- Mu.star[-( (kMinus + 1):(kMinus + M) )]
      myProb <- myProb[-( (kMinus + 1):(kMinus + M) )]
      Mu.star[kMinus + 1] <- phi
    }
  }
  return(list(Mu.star = Mu.star, indSet = indSet, freqSet = freqSet))
}

Remixing.Mu <- function(nSub, Mu.star, indSet, Beta0_current,
                        DP.G0.mean, DP.G0.sd, sd_for_beta0_i_S){

  Mu_Uniq_n <- length(Mu.star)
  for (i in 1:Mu_Uniq_n){
    idsToConsider <- (1:nSub)[indSet == i]
    Beta0_for_Mu_i <- Beta0_current[idsToConsider]
    n_tmp <- length(idsToConsider)
    mean_tmp <- ((sd_for_beta0_i_S^2/n_tmp)/
                   ((sd_for_beta0_i_S^2/n_tmp) + DP.G0.sd^2))*DP.G0.mean
```

```
                    + ((DP.G0.sd^2)/((sd_for_beta0_i_S^2/n_tmp) + DP.G0.sd^2))
                      *(mean(Beta0_for_Mu_i))
    var_tmp <- (sd_for_beta0_i_S^2/n_tmp)*(DP.G0.sd^2)
                /((sd_for_beta0_i_S^2/n_tmp) + DP.G0.sd^2)
    Mu.star[i] <- rnorm(1, mean = mean_tmp, sd = sqrt(var_tmp))
  }
  return(Mu.star)
}
```

A.2.2 Joint Longitudinal-Survival Model HMC Code

Our generic HMC sampler that is shared for all proposed models in Chapter 4.

```
# calculating potential engergy:
U <- function(fit, un){
  return(as.numeric(-log_prob(fit, upars = un, adjust_transform = T)))
}

# calculating gradient:
grad_U <- function(fit, un){
  return(as.numeric(-grad_log_prob(fit, upars = un, adjust_transform = T)))
}

# HMC sampler (idea from Neal 2012 - MCMC using Hamiltonian Dynamics)
HMC <- function (fit, epsilon, L_low, L_hi, current_q_un){

  L <- sample(L_low:L_hi, 1)

  q = current_q_un
  p = rnorm(length(q),0,1)  # independent standard normal variates
  current_p = p

  # Make a half step for momentum at the beginning
  p = p - epsilon * grad_U(fit, q) / 2

  # Alternate full steps for position and momentum
  for (i in 1:L){
```

```
  # Make a full step for the position
  q = q + epsilon * p

  # Make a full step for the momentum, except at end of trajectory
  if (i!=L) p = p - epsilon * grad_U(fit, q)
}

# Make a half step for momentum at the end.
p = p - epsilon * grad_U(fit, q) / 2

# Negate momentum at end of trajectory to make the proposal symmetric
p = -p

# Evaluate potential and kinetic energies at start and end of trajectory
current_U = U(fit, current_q_un)
current_K = sum(current_p^2) / 2
proposed_U = U(fit, q)
proposed_K = sum(p^2) / 2

# Accept or reject the state at end of trajectory, returning either
# the position at the end of the trajectory or the initial position

if (runif(1) < exp(current_U-proposed_U+current_K-proposed_K))
{
  return (list(q = q, acc = 1))  # accept
}
```

```
  else
  {
    return (list(q = current_q_un, acc = 0))  # reject
  }
}
```

A.2.3 Joint Longitudinal-Survival Model Stan Code

Model I

```
data {
  int<lower = 1> N;
  int<lower = 1> Mi;
  matrix[N, Mi] time;
  matrix[N, Mi] albumin;
  vector[N] all_surv_times;
  vector[N] nu;
  vector[N] MU_for_beta0_i_S;
  real sd_for_beta0_i_S;
}
transformed data {
  matrix[Mi, Mi] Dist_GP[N];
  matrix[Mi, Mi] Q1[N];
  vector[Mi] R1[N];
  for (i in 1:N){
    for (j in 1:Mi) {
      for (k in 1:Mi){
        Dist_GP[i][j,k] <- exp(- pow(time[i,j] - time[i,k], 2.0));
      }
    }
    Dist_GP[i] <- Dist_GP[i] + diag_matrix(rep_vector(.00001, Mi));
    Q1[i] <- eigenvectors_sym(Dist_GP[i]);
    R1[i] <- eigenvalues_sym(Dist_GP[i]);
```

```
  }
}
parameters {
  vector[N] beta0_alb;
  vector<lower = 0>[N] kappa_sq_alb;
  real<lower = 0> sigma_sq_alb;
  real<lower = 0> survShape;
  vector[N] beta0S;
  real beta0_shared;
  real beta_alb;
}
model {
  row_vector[Mi] mu_alb;
  vector[Mi] YtQ_alb;
  real alb_pred_at_t;
  matrix[Mi,1] K_star_alb;
  matrix[Mi,Mi] SigmaPred_alb;
  matrix[1,Mi] K_transpose_div_SigmaPred_alb;
  int censInd;
  int obsInd;
  real lambda_tmp;
  real hazard_AUC;
  real myStepSize;
  real midPoint;
  real intLength;
  real Hazard_PieceWise_Const;
  for (i in 1:N){
```

```
  mu_alb <- rep_row_vector(beta0_alb[i], Mi);
  YtQ_alb <- (albumin[i,]*Q1[i])';
  increment_log_prob(-0.5 * (sum(log(kappa_sq_alb[i] * R1[i]
                     + sigma_sq_alb)))

    - 0.5*sum((YtQ_alb - (mu_alb*Q1[i])').*(YtQ_alb - (mu_alb*Q1[i])')
    ./(R1[i] * kappa_sq_alb[i] + sigma_sq_alb)));
}
for (i in 1:N){
  myStepSize <- all_surv_times[i]/10.0;
  intLength <- myStepSize;
  hazard_AUC <- 0.0;
  for (aucInd in 1:10){
    midPoint <- intLength/2.0 + (aucInd - 1.0)*intLength;
    SigmaPred_alb <- kappa_sq_alb[i]*Dist_GP[i]
                     + diag_matrix(rep_vector(sigma_sq_alb, Mi));

    for (j in 1:Mi){
      K_star_alb[j, 1] <- kappa_sq_alb[i] * exp(- pow(time[i,j]
                                                - midPoint, 2.0));
    }
    K_transpose_div_SigmaPred_alb <- (K_star_alb)' / SigmaPred_alb;
    alb_pred_at_t <- beta0_alb[i]
                   + to_array_1d(K_transpose_div_SigmaPred_alb
                                 * (row(albumin, i) - beta0_alb[i])')[1];
  lambda_tmp <- beta0S[i] + beta0_shared + beta_alb*alb_pred_at_t;
  Hazard_PieceWise_Const <- exp(lambda_tmp)*survShape
```

```
                              *pow(midPoint, survShape - 1.0);
    hazard_AUC <- hazard_AUC + intLength*Hazard_PieceWise_Const;
    }
    increment_log_prob(nu[i]*log(Hazard_PieceWise_Const) - hazard_AUC);
  }
  beta0_alb ~ normal(5, 2);
  kappa_sq_alb ~ lognormal(-1, 2);
  sigma_sq_alb ~ lognormal(-1, 1);
  survShape ~ lognormal(0, 1);
  beta0_shared ~ normal(0, 5);
  beta_alb ~ normal(0, 5);
  for (i in 1:N){
    beta0S[i] ~ normal(MU_for_beta0_i_S[i], sd_for_beta0_i_S);
  }
}
```

Model II

```
# This code is for the Area Under the Curve Model:
data {
  int<lower = 1> N;
  int<lower = 1> Mi;
  matrix[N, Mi] time;
  matrix[N, Mi] albumin;
  vector[N] all_surv_times;
  vector[N] nu;
  vector[N] MU_for_beta0_i_S;
  real sd_for_beta0_i_S;
}
transformed data {
  matrix[Mi, Mi] Dist_GP[N];
  matrix[Mi, Mi] Q1[N];
  vector[Mi] R1[N];
  for (i in 1:N){
    for (j in 1:Mi) {
      for (k in 1:Mi){
        Dist_GP[i][j,k] <- exp(- pow(time[i,j] - time[i,k], 2.0));
      }
    }
    Dist_GP[i] <- Dist_GP[i] + diag_matrix(rep_vector(0.00001, Mi));
    Q1[i] <- eigenvectors_sym(Dist_GP[i]);
    R1[i] <- eigenvalues_sym(Dist_GP[i]);
  }
```

```
}
parameters {
  vector[N] beta0_alb;
  vector<lower = 0>[N] kappa_sq_alb;
  real<lower = 0> sigma_sq_alb;
  real<lower = 0> survShape;
  vector[N] beta0S;
  real beta0_shared;
  real beta_alb;
  real beta_alb_drev;
}
model {
  row_vector[Mi] mu_alb;
  vector[Mi] YtQ_alb;
  real alb_pred_at_t;
  real alb_drev_pred_at_t;
  matrix[Mi,1] K_star_alb;
  matrix[Mi,Mi] SigmaPred_alb;
  matrix[1,Mi] K_transpose_div_SigmaPred_alb;
  int censInd;
  int obsInd;
  real lambda_tmp;
  real hazard_AUC;
  real myStepSize;
  real midPoint;
  real intLength;
  real Hazard_PieceWise_Const;
```

```
for (i in 1:N){
  mu_alb <- rep_row_vector(beta0_alb[i], Mi);
  YtQ_alb <- (albumin[i,]*Q1[i])';
  increment_log_prob(-0.5 * (sum(log(kappa_sq_alb[i] * R1[i]
                                     + sigma_sq_alb)))
    - 0.5*sum((YtQ_alb - (mu_alb*Q1[i])').*(YtQ_alb - (mu_alb*Q1[i])')
      ./(R1[i] * kappa_sq_alb[i] + sigma_sq_alb)));
}
for (i in 1:N){
  myStepSize <- all_surv_times[i]/10.0;
  intLength <- myStepSize;
  hazard_AUC <- 0.0;
  for (aucInd in 1:10){
    midPoint <- intLength/2.0 + (aucInd - 1.0)*intLength;
    SigmaPred_alb <- kappa_sq_alb[i]*Dist_GP[i]
                     + diag_matrix(rep_vector(sigma_sq_alb, Mi));
    for (j in 1:Mi){
      K_star_alb[j, 1] <- kappa_sq_alb[i] * exp(- pow(time[i,j]
                                                - midPoint, 2.0));
    }
    K_transpose_div_SigmaPred_alb <- (K_star_alb)' / SigmaPred_alb;
    alb_pred_at_t <- beta0_alb[i]
    + to_array_1d(K_transpose_div_SigmaPred_alb *
                  (row(albumin, i) - beta0_alb[i])')[1];

    alb_drev_pred_at_t <- alb_pred_at_t - albumin[i,1];
    lambda_tmp <- beta0S[i] + beta0_shared + beta_alb*alb_pred_at_t
```

```
                    + beta_alb_drev*alb_drev_pred_at_t;


      Hazard_PieceWise_Const <- exp(lambda_tmp)*survShape
                                *pow(midPoint, survShape - 1.0);
      hazard_AUC <- hazard_AUC + intLength*Hazard_PieceWise_Const;
    }
    increment_log_prob(nu[i]*log(Hazard_PieceWise_Const) - hazard_AUC);
  }
  beta0_alb ~ normal(5, 2);
  kappa_sq_alb ~ lognormal(-1, 2);
  sigma_sq_alb ~ lognormal(-2.5, 1);
  survShape ~ lognormal(0, 1);
  beta0_shared ~ normal(0, 2);
  beta_alb ~ normal(0, 1);
  beta_alb_drev ~ normal(0, 1);
  for (i in 1:N){
    beta0S[i] ~ normal(MU_for_beta0_i_S[i], sd_for_beta0_i_S);
  }
}
```

Model II

```
# Model II: Code for Pointwise derivative model
data {
  // longitudinal data!
  int<lower = 1> N;
  int<lower = 1> Mi;
  matrix[N, Mi] time;
  matrix[N, Mi] albumin;
  vector[N] all_surv_times;
  vector[N] nu;
  vector[N] MU_for_beta0_i_S;
  real sd_for_beta0_i_S;
}
transformed data {
  matrix[Mi, Mi] Dist_GP[N];
  matrix[Mi, Mi] Q1[N];
  vector[Mi] R1[N];
  for (i in 1:N){
    for (j in 1:Mi) {
      for (k in 1:Mi){
        Dist_GP[i][j,k] <- exp(- pow(time[i,j] - time[i,k], 2.0));
      }
    }
    Dist_GP[i] <- Dist_GP[i] + diag_matrix(rep_vector(0.00001, Mi));
    Q1[i] <- eigenvectors_sym(Dist_GP[i]);
    R1[i] <- eigenvalues_sym(Dist_GP[i]);
```

```
  }
}
parameters {
  vector[N] beta0_alb;
  vector<lower = 0>[N] kappa_sq_alb;
  real<lower = 0> sigma_sq_alb;
  real<lower = 0> survShape;
  vector[N] beta0S;
  real beta0_shared;
  real beta_alb;
  real beta_alb_drev;
}
model {
  row_vector[Mi] mu_alb;
  vector[Mi] YtQ_alb;
  real alb_pred_at_t;
  real alb_drev_pred_at_t;
  matrix[Mi,1] K_star_alb;
  matrix[Mi,1] K_der;
  matrix[Mi,Mi] SigmaPred_alb;
  matrix[1,Mi] K_transpose_div_SigmaPred_alb;
  matrix[1,Mi] K_der_transpose_div_SigmaPred;
  int censInd;
  int obsInd;
  real lambda_tmp;
  real hazard_AUC;
  real myStepSize;
```

```
real midPoint;
real intLength;
real Hazard_PieceWise_Const;
for (i in 1:N){
  mu_alb <- rep_row_vector(beta0_alb[i], Mi);
  YtQ_alb <- (albumin[i,]*Q1[i])';
  increment_log_prob(-0.5 * (sum(log(kappa_sq_alb[i] * R1[i] + sigma_sq_alb)))
                     - 0.5*sum((YtQ_alb - (mu_alb*Q1[i])').
                       *(YtQ_alb - (mu_alb*Q1[i])')
                       ./(R1[i] * kappa_sq_alb[i] + sigma_sq_alb)));
}
for (i in 1:N){
  myStepSize <- all_surv_times[i]/10.0;
  intLength <- myStepSize;
  hazard_AUC <- 0.0;
   for (aucInd in 1:10){
    midPoint <- intLength/2.0 + (aucInd - 1.0)*intLength;
    SigmaPred_alb <- kappa_sq_alb[i]*Dist_GP[i]
                     + diag_matrix(rep_vector(sigma_sq_alb, Mi));
    for (j in 1:Mi){
      K_star_alb[j, 1] <- kappa_sq_alb[i] * exp(- pow(time[i,j] - midPoint, 2.0));
      K_der[j,1] <- -1.99999*1.0*(midPoint - time[i,j])*K_star_alb[j,1];
    }
    K_transpose_div_SigmaPred_alb <- (K_star_alb)' / SigmaPred_alb;
    K_der_transpose_div_SigmaPred <- (K_der)' / SigmaPred_alb;
    alb_pred_at_t <- beta0_alb[i]
                     + to_array_1d(K_transpose_div_SigmaPred_alb
```

```
                          * (row(albumin, i) - beta0_alb[i])')[1];

      alb_drev_pred_at_t <- to_array_1d(K_der_transpose_div_SigmaPred
                                        * (row(albumin, i) - beta0_alb[i])')[1];

      lambda_tmp <- beta0S[i] + beta0_shared
                    + beta_alb*alb_pred_at_t
                    + beta_alb_drev*alb_drev_pred_at_t;

      Hazard_PieceWise_Const <- exp(lambda_tmp)*survShape
                                  *pow(midPoint, survShape - 1.0);

      hazard_AUC <- hazard_AUC + intLength*Hazard_PieceWise_Const;
    }
    increment_log_prob(nu[i]*log(Hazard_PieceWise_Const) - hazard_AUC);
  }
  beta0_alb ~ normal(5, 2);
  kappa_sq_alb ~ lognormal(-1, 2);
  sigma_sq_alb ~ lognormal(-2.5, 1);
  survShape ~ lognormal(0, 1);
  beta0_shared ~ normal(0, 2);
  beta_alb ~ normal(0, 1);
  beta_alb_drev ~ normal(0, 1);
  for (i in 1:N){
    beta0S[i] ~ normal(MU_for_beta0_i_S[i], sd_for_beta0_i_S);
  }
}
```

Model III

```
data {
  int<lower = 1> N;
  int<lower = 1> Mi;
  matrix[N, Mi] time;
  matrix[N, Mi] albumin;
  vector[N] all_surv_times;
  vector[N] nu;
  vector[N] Continuous_Cov;
  vector[N] MU_for_beta0_i_S;
  real sd_for_beta0_i_S;
}
transformed data {
  matrix[Mi, Mi] Dist_GP[N];
  matrix[Mi, Mi] Q1[N];
  vector[Mi] R1[N];
  for (i in 1:N){
    // dealing with distance matrix eigenvalue decomposition!
    for (j in 1:Mi) {
      for (k in 1:Mi){
        Dist_GP[i][j,k] <- exp(- pow(time[i,j] - time[i,k], 2.0));
      }
    }
    Dist_GP[i] <- Dist_GP[i] + diag_matrix(rep_vector(.00001, Mi));

    // eigenvalue decomposition:
```

```
    Q1[i] <- eigenvectors_sym(Dist_GP[i]);
    R1[i] <- eigenvalues_sym(Dist_GP[i]);
  }
}
parameters {
  vector[N] beta0_alb;
  vector<lower = 0>[N] kappa_sq_alb;
  real<lower = 0> sigma_sq_alb;
  real<lower = 0> survShape;
  vector[N] beta0S;
  real beta0_shared;
  real beta_cont_cov;
  real beta_randIntercept;
  real beta_volatility;
}
model {
  row_vector[Mi] mu_alb;
  vector[Mi] YtQ_alb;
  int censInd;
  int obsInd;
  real lambda_tmp;
  real hazard_AUC;
  real myStepSize;
  real midPoint;
  real intLength;
  real Hazard_PieceWise_Const;
  for (i in 1:N){
```

```
  mu_alb <- rep_row_vector(beta0_alb[i], Mi);
  YtQ_alb <- (albumin[i,]*Q1[i])';
  increment_log_prob(-0.5 * (sum(log(kappa_sq_alb[i] * R1[i] + sigma_sq_alb)))
                     - 0.5*sum((YtQ_alb - (mu_alb*Q1[i])')
                     .*(YtQ_alb - (mu_alb*Q1[i])')
                     ./(R1[i] * kappa_sq_alb[i] + sigma_sq_alb)));
}
for (i in 1:N){
  myStepSize <- all_surv_times[i]/10.0;
  intLength <- myStepSize;
  hazard_AUC <- 0.0;
  for (aucInd in 1:10){
    midPoint <- intLength/2.0 + (aucInd - 1.0)*intLength;
    lambda_tmp <- beta0S[i] + beta0_shared
    + beta_cont_cov*Continuous_Cov[i]
    + beta_randIntercept*beta0_alb[i]
    + beta_volatility*kappa_sq_alb[i];

    Hazard_PieceWise_Const <- exp(lambda_tmp)*survShape
                              *pow(midPoint, survShape - 1.0);

    hazard_AUC <- hazard_AUC
                  + intLength*Hazard_PieceWise_Const;
  }
  increment_log_prob(nu[i]*log(Hazard_PieceWise_Const) - hazard_AUC);
}
beta0_alb ~ normal(5, 2);
```

```
  kappa_sq_alb ~ lognormal(-1, 2);
  sigma_sq_alb ~ lognormal(-1, 1);
  survShape ~ lognormal(0, 1);
  beta0_shared ~ normal(0, 1);
  beta_cont_cov ~ normal(0, 1);
  beta_randIntercept ~ normal(0, 1);
  beta_volatility ~ normal(0, 1);
  for (i in 1:N){
    beta0S[i] ~ normal(MU_for_beta0_i_S[i], sd_for_beta0_i_S);
  }
}
```

Model II

```
data {
  int<lower = 1> N;
  int<lower = 1> Mi;
  matrix[N, Mi] time;
  matrix[N, Mi] albumin;
  vector[N] all_surv_times;
  vector[N] nu;
  vector[N] Continuous_Cov;
  vector[N] MU_for_beta0_i_S;
  real sd_for_beta0_i_S;
}
transformed data {
  matrix[Mi, Mi] Dist_GP[N];
  matrix[Mi, Mi] Q1[N];
  vector[Mi] R1[N];
  for (i in 1:N){
    for (j in 1:Mi) {
      for (k in 1:Mi){
        Dist_GP[i][j,k] <- exp(- pow(time[i,j] - time[i,k], 2.0));
      }
    }
    Dist_GP[i] <- Dist_GP[i] + diag_matrix(rep_vector(.00001, Mi));
    Q1[i] <- eigenvectors_sym(Dist_GP[i]); // would be the same for Kernel GP
    R1[i] <- eigenvalues_sym(Dist_GP[i]);
  }
```

```
}
parameters {
  vector[N] beta0_alb;
  vector<lower = 0>[N] kappa_sq_alb;
  real<lower = 0> sigma_sq_alb;
  real<lower = 0> survShape;
  vector[N] beta0S;
  real beta0_shared;
  real beta_cont_cov;
  real beta_randIntercept;
  real beta_volatility;
}
model {
  row_vector[Mi] mu_alb;
  vector[Mi] YtQ_alb;
  int censInd;
  int obsInd;
  real lambda_tmp;
  real hazard_AUC;
  real myStepSize;
  real midPoint;
  real intLength;
  real Hazard_PieceWise_Const;
  for (i in 1:N){
    mu_alb <- rep_row_vector(beta0_alb[i], Mi);
    YtQ_alb <- (albumin[i,]*Q1[i])';
    increment_log_prob(-0.5 * (sum(log(kappa_sq_alb[i]
```

```
                        * R1[i] + sigma_sq_alb)))
                        - 0.5*sum((YtQ_alb - (mu_alb*Q1[i])')
                        .*(YtQ_alb - (mu_alb*Q1[i])')
                        ./(R1[i] * kappa_sq_alb[i] + sigma_sq_alb)));
}
for (i in 1:N){
  myStepSize <- all_surv_times[i]/10.0;
  intLength <- myStepSize;
  hazard_AUC <- 0.0;
  for (aucInd in 1:10){
    midPoint <- intLength/2.0 + (aucInd - 1.0)*intLength;
    lambda_tmp <- beta0S[i] + beta0_shared
                  + beta_cont_cov*Continuous_Cov[i]
                  + beta_randIntercept*beta0_alb[i]
                  + beta_volatility*kappa_sq_alb[i];

    Hazard_PieceWise_Const <- exp(lambda_tmp)*survShape
                              *pow(midPoint, survShape - 1.0);

    hazard_AUC <- hazard_AUC + intLength*Hazard_PieceWise_Const;
  }
  increment_log_prob(nu[i]*log(Hazard_PieceWise_Const) - hazard_AUC);
}
beta0_alb ~ normal(5, 2);
kappa_sq_alb ~ lognormal(-1, 2);
sigma_sq_alb ~ lognormal(-1, 1);
survShape ~ lognormal(0, 1);
```

```
  beta0_shared ~ normal(0, 1);
  beta_cont_cov ~ normal(0, 1);
  beta_randIntercept ~ normal(0, 1);
  beta_volatility ~ normal(0, 1);
  for (i in 1:N){
    beta0S[i] ~ normal(MU_for_beta0_i_S[i], sd_for_beta0_i_S);
  }
}
```

A.3 Chapter 5 Models

Similiar to the code in Chapter 4, our programming codes for our proposed model in Chapter 5 include three main parts: 1) DP Sampler, 2) HMC code, 3) Stan code.

A.3.1 DP Sampler

```
get.DP.alpha.L <- function(nSub, DP.alpha.curr, DPM_MU.star,
                           DP.alpha.a, DP.alpha.b){
  k <- length(DPM_MU.star)
  M <- nSub
  eta <- rbeta(1, DP.alpha.curr + 1, M)
  Pi.eta <- (DP.alpha.a + k - 1)/( (DP.alpha.a + k - 1)
                                   + (M*(DP.alpha.b - log(eta))) )
  sampProb <- c(Pi.eta, (1 - Pi.eta))
  DP.alpha.pool <- c(rgamma(1, shape = (DP.alpha.a + k),
                            scale = 1/(DP.alpha.b - log(eta)) ),
                     rgamma(1, shape = (DP.alpha.a + k - 1),
                            scale = 1/(DP.alpha.b - log(eta)) ) )
  DP.alpha.new <- sample(DP.alpha.pool, 1, prob = sampProb)
  return(DP.alpha.new)
}

# We borrowed ideas from Dr. Shahbaba in writing this code.
get.Mui <- function(nSub, Beta0_current, Mu.star, indSet,
                    freqSet, DP.alpha, DP.G0.mean, DP.G0.sd, sd_for_beta0_i_S){
  M = 5
```

```
for (i in 1:nSub){
  phi <- NULL
  curInd <- indSet[i]
  freqSet[curInd] <- freqSet[curInd] - 1
  if (freqSet[curInd] == 0){
    phi <- Mu.star[curInd]
    freqSet <- freqSet[-(curInd)]
    Mu.star <- Mu.star[-(curInd)]
    indSet[indSet > curInd] <- indSet[indSet > curInd] - 1
    kMinus <- length(freqSet)
    Mu.star[kMinus + 1] <- phi
    for (m in 1:(M - 1)){
      Mu.star[kMinus + 1 + m] <- rnorm(1, mean = DP.G0.mean,
                                       sd = DP.G0.sd)
    }

  }else{
    kMinus <- length(freqSet)
    for (m in 1:M){
      Mu.star[kMinus + m] <- rnorm(1, mean = DP.G0.mean,
                                   sd = DP.G0.sd)
    }
  }
  # q1: vector of sampling prob for exisiting parameters
  # q2: vector of sampling prob for auxiliary parameters

  # Prob. of Sampling Existing Parameters:
```

```
# Here, we are looking at beta0_i for subj i to write the likelihood
q1 <- rep(0, kMinus)
for (k in 1:kMinus){
  q1[k] <- dnorm(Beta0_current[i], mean = Mu.star[k],
                 sd = sd_for_beta0_i_S, log = T)
}
q1 <- q1 + log(freqSet) - log(nSub - 1 + DP.alpha)
q2 <- rep(0, M)
for (k in 1:M){
  q2[k] <- dnorm(Beta0_current[i], mean = Mu.star[kMinus + k]
                 , sd = sd_for_beta0_i_S, log = T)
}
q2 <- q2 + log(DP.alpha/M) - log(nSub - 1 + DP.alpha)
myProb <- c(q1, q2)
myProb.Max <- max(myProb)
myProb.Rel <- myProb - myProb.Max
myProb <- exp(myProb.Rel)
myProb <- myProb/sum(myProb)
picked <- which(rmultinom(1, 1, myProb) == 1)
if (picked <= kMinus){
  indSet[i] <- picked
  freqSet[picked] <- freqSet[picked] + 1
  Mu.star <- Mu.star[-( (kMinus + 1):(kMinus + M) )]
  myProb <- myProb[-( (kMinus + 1):(kMinus + M) )]
}else{
  indSet[i] <- kMinus + 1
  freqSet <- c(freqSet, 1)
```

```
      phi <- Mu.star[picked]
      Mu.star <- Mu.star[-( (kMinus + 1):(kMinus + M) )]
      myProb <- myProb[-( (kMinus + 1):(kMinus + M) )]
      Mu.star[kMinus + 1] <- phi
    }
  }
  return(list(Mu.star = Mu.star, indSet = indSet, freqSet = freqSet))
}

Remixing.Mu <- function(nSub, Mu.star, indSet, Beta0_current,
                        DP.G0.mean, DP.G0.sd, sd_for_beta0_i_S){

  Mu_Uniq_n <- length(Mu.star)
  for (i in 1:Mu_Uniq_n){
    idsToConsider <- (1:nSub)[indSet == i]
    Beta0_for_Mu_i <- Beta0_current[idsToConsider]
    n_tmp <- length(idsToConsider)
    mean_tmp <- ((sd_for_beta0_i_S^2/n_tmp)/
                   ((sd_for_beta0_i_S^2/n_tmp) + DP.G0.sd^2))*DP.G0.mean
                    + ((DP.G0.sd^2)/((sd_for_beta0_i_S^2/n_tmp) + DP.G0.sd^2))
                      *(mean(Beta0_for_Mu_i))
    var_tmp <- (sd_for_beta0_i_S^2/n_tmp)*(DP.G0.sd^2)
                /((sd_for_beta0_i_S^2/n_tmp) + DP.G0.sd^2)
    Mu.star[i] <- rnorm(1, mean = mean_tmp, sd = sqrt(var_tmp))
  }
  return(Mu.star)
}
```

A.3.2 Joint Multivariate Longitudinal-Survival Model HMC Code

```
# calculating potential engergy:
U <- function(fit, un){
  return(as.numeric(-log_prob(fit, upars = un, adjust_transform = T)))
}

# calculating gradient:
grad_U <- function(fit, un){
  return(as.numeric(-grad_log_prob(fit, upars = un, adjust_transform = T)))
}

# HMC sampler (idea from Neal 2012 - MCMC using Hamiltonian Dynamics)
HMC <- function (fit, epsilon, L_low, L_hi, current_q_un){

  L <- sample(L_low:L_hi, 1)

  q = current_q_un
  p = rnorm(length(q),0,1)  # independent standard normal variates
  current_p = p

  # Make a half step for momentum at the beginning
  p = p - epsilon * grad_U(fit, q) / 2

  # Alternate full steps for position and momentum
  for (i in 1:L){
```

```
  # Make a full step for the position
  q = q + epsilon * p

  # Make a full step for the momentum, except at end of trajectory
  if (i!=L) p = p - epsilon * grad_U(fit, q)
}

# Make a half step for momentum at the end.
p = p - epsilon * grad_U(fit, q) / 2

# Negate momentum at end of trajectory to make the proposal symmetric
p = -p

# Evaluate potential and kinetic energies at start and end of trajectory
current_U = U(fit, current_q_un)
current_K = sum(current_p^2) / 2
proposed_U = U(fit, q)
proposed_K = sum(p^2) / 2

# Accept or reject the state at end of trajectory, returning either
# the position at the end of the trajectory or the initial position

if (runif(1) < exp(current_U-proposed_U+current_K-proposed_K))
{
  return (list(q = q, acc = 1))  # accept
}
else
```

```
  {
    return (list(q = current_q_un, acc = 0))  # reject
  }
}
```

A.3.3 Joint Longitudinal-Survival Model Stan Code

```
functions {
 matrix kron_mvprod(matrix A, matrix B, matrix V) {
   return transpose(A * transpose(B * V));
 }
 matrix calculate_eigenvalues(vector A, vector B, int n1,
                              real sigma_pr1, real sigma_pr2) {
   matrix[2,n1] e;
   for(j in 1:n1) {
      e[1,j] <- (A[j] * B[1] + sigma_pr1);
      e[2,j] <- (A[j] * B[2] + sigma_pr2);
    }
    return(e);
  }
}
data {
  int<lower = 1> N;
  int<lower = 1> Mi;
  int<lower = 1> N_Miss;
  matrix[N, Mi] time;
  vector[(2*N*Mi - N_Miss)] alb_bmi_obs;
  int miss_ind[N_Miss, 2];
  int obs_ind[(2*N*Mi - N_Miss), 2];
  vector[N] all_surv_times;
  vector[N] nu;
  vector[N] MU_for_beta0_i_S;
```

```
  real sd_for_beta0_i_S;
}
transformed data {
  matrix[Mi, Mi] Dist_GP;
  matrix[Mi, Mi] Q1[N];
  vector[Mi] R1[N];
  for (i in 1:N){
    for (j in 1:Mi) {
      for (k in 1:Mi){
        Dist_GP[j,k] <- exp(- pow(time[i,j] - time[i,k], 2.0));
      }
    }
    Q1[i] <- eigenvectors_sym(Dist_GP);
    R1[i] <- eigenvalues_sym(Dist_GP);
  }
}
parameters {
  vector[N] beta0_proc1;
  vector[N] beta0_proc2;
  vector<lower = 0>[N] kappa_sq;
  real<lower = 0> sigma_sq_proc1;
  real<lower = 0> sigma_sq_proc2;
  real<lower = 0> tau2;
  corr_matrix[2] Omega;
  vector[N_Miss] alb_bmi_Miss;
  real<lower = 0> survShape;
  vector[N] beta0S;
```

```
  real beta0_shared;
  real beta_alb;
  real beta_bmi;
}
transformed parameters{
  matrix[N, 2*Mi] alb_bmi;
  for (i in 1:N_Miss){
    alb_bmi[miss_ind[i,1], miss_ind[i,2]] <- alb_bmi_Miss[i];
  }
  for (i in 1:(2*N*Mi - N_Miss)){
    alb_bmi[obs_ind[i,1], obs_ind[i,2]] <- alb_bmi_obs[i];
  }
}
model {
  matrix[2,2] C;
  matrix[2, Mi] mu;
  vector[Mi] R1_K_GP;
  matrix[2, Mi] y[N];
  matrix[2, Mi] myTempTransformer1;
  matrix[2*Mi, 1] myTempTransformer2;
  matrix[2, 2] Q2;
  vector[2] R2;
  matrix[2,Mi] eigenvalues;
  vector[2] tau;
  vector[2] alb_bmi_pred_at_t;
  row_vector[Mi] K_star;
  matrix[2,2*Mi] C_kron_K_star;
```

```
int censInd;
int obsInd;
real lambda_tmp;
real hazard_AUC;
real myStepSize;
real midPoint;
real intLength;
real Hazard_PieceWise_Const;
tau[1] <- 1;
tau[2] <- tau2;
C <- quad_form_diag(Omega, tau);
Q2 <- eigenvectors_sym(C);
R2 <- eigenvalues_sym(C);
for (i in 1:N){
  for (j in 1:Mi){
    y[i][1,j] <- alb_bmi[i,j];
    y[i][2,j] <- alb_bmi[i,Mi + j];
  }
  mu[1,] <- rep_row_vector(beta0_proc1[i], Mi);
  mu[2,] <- rep_row_vector(beta0_proc2[i], Mi);
  R1_K_GP <- kappa_sq[i] * R1[i];
  eigenvalues <- calculate_eigenvalues(R1_K_GP, R2, Mi,
                                       sigma_sq_proc1, sigma_sq_proc2);
  increment_log_prob(-0.5 * sum((y[i] - mu)
                      .* kron_mvprod(Q1[i],Q2, kron_mvprod(transpose(Q1[i]),
                        transpose(Q2), (y[i] - mu)) ./ eigenvalues)) - 0.5
                        * sum(log(eigenvalues)));
```

```
  }
  for (i in 1:N){
    myStepSize <- all_surv_times[i]/10.0;
    intLength <- myStepSize;
    hazard_AUC <- 0.0;
    for (aucInd in 1:10){
      midPoint <- intLength/2.0 + (aucInd - 1.0)*intLength;
      for (j in 1:Mi){
        K_star[j] <- kappa_sq[i] * exp(- pow(time[i,j] - midPoint, 2.0));
      }
      C_kron_K_star[1, 1:Mi] <- C[1,1] * K_star;
      C_kron_K_star[2, 1:Mi] <- C[2,1] * K_star;
      C_kron_K_star[1, (Mi + 1):(2*Mi)] <- C[1,2] * K_star;
      C_kron_K_star[2, (Mi + 1):(2*Mi)] <- C[2,2] * K_star;
      mu[1,] <- rep_row_vector(beta0_proc1[i], Mi);
      mu[2,] <- rep_row_vector(beta0_proc2[i], Mi);
      R1_K_GP <- kappa_sq[i] * R1[i];
      eigenvalues <- calculate_eigenvalues(R1_K_GP, R2, Mi, sigma_sq_proc1,
                                                            sigma_sq_proc2);

myTempTransformer1 <- kron_mvprod(Q1[i],Q2, kron_mvprod(transpose(Q1[i]),
                                        transpose(Q2), (y[i] - mu))
                                                    ./ eigenvalues);

      myTempTransformer2[1:Mi,1] <- transpose(myTempTransformer1[1,]);
      myTempTransformer2[(Mi + 1):(2*Mi),1] <- transpose(myTempTransformer1[2,]);
      alb_bmi_pred_at_t <- mu[,1] + to_vector(C_kron_K_star * myTempTransformer2);
```

```
      lambda_tmp <- beta0S[i] + beta0_shared
                    + beta_alb*alb_bmi_pred_at_t[1]
                      + beta_bmi*alb_bmi_pred_at_t[2];

      Hazard_PieceWise_Const <- exp(lambda_tmp)*survShape
                                  *pow(midPoint, survShape - 1.0);

      hazard_AUC <- hazard_AUC + intLength*Hazard_PieceWise_Const;
    }
    increment_log_prob(nu[i]*log(Hazard_PieceWise_Const) - hazard_AUC);
  }
  beta0_proc1 ~ normal(5, 2);
  beta0_proc2 ~ normal(20, 5);
  kappa_sq ~ lognormal(-1, 2);
  sigma_sq_proc1 ~ lognormal(-1, 1);
  sigma_sq_proc2 ~ lognormal(-1, 1);
  tau2 ~ cauchy(0, 2.5);
  Omega ~ lkj_corr(1);
  survShape ~ lognormal(0, 1);
  beta0_shared ~ normal(0, 5);
  beta_alb ~ normal(0, 5);
  beta_bmi ~ normal(0, 5);
  for (i in 1:N){
    beta0S[i] ~ normal(MU_for_beta0_i_S[i], sd_for_beta0_i_S);
  }
}
```

Appendix B

Posterior Samples Traceplots

Here in this chapter, we provide trace-plots for posterior samples of the Bayesian models used in this dissertation.

B.1 Chapter 3 Traceplots

B.1.1 Bayesian Logistic Models

Normally Distributed Random Intercepts

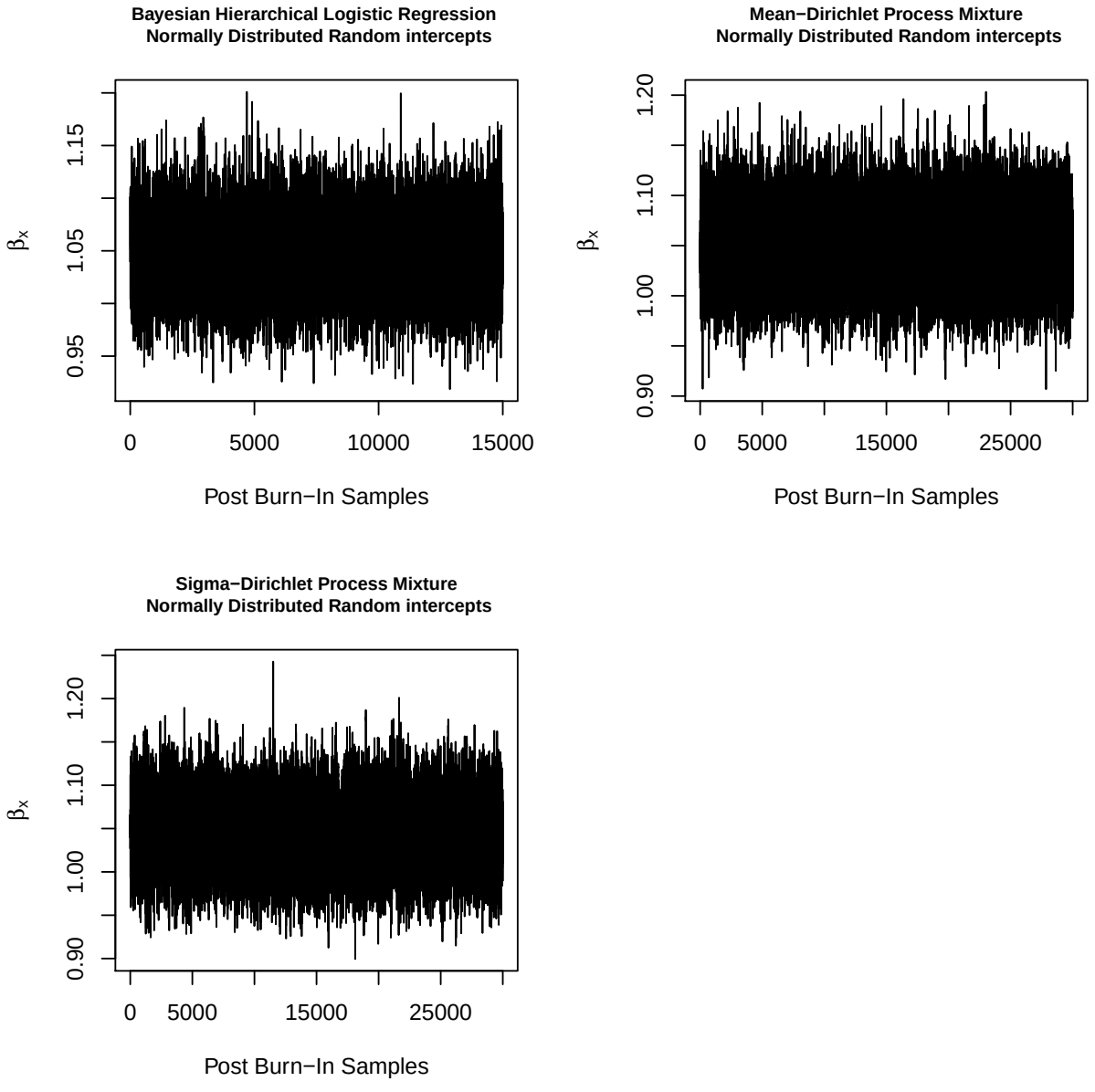


Figure B.1: Traceplots for Posterior samples under a simulation scenario when random intercepts are Normally distributed

Random Intercepts from Mixture of Two Normals with Different Means

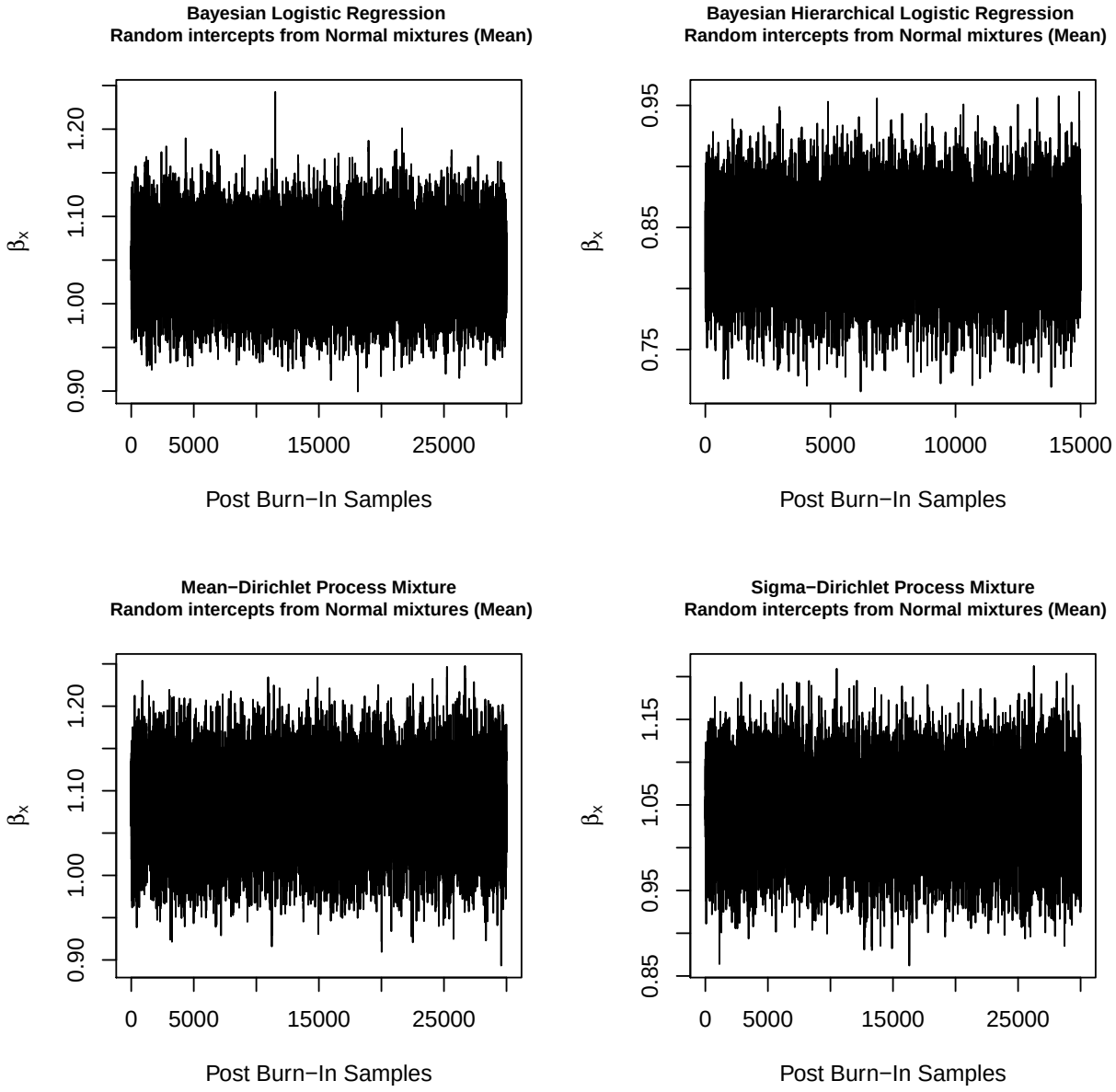


Figure B.2: Traceplots for Posterior samples under a simulation scenario when random intercepts are from a mixture N(-1.5, 1) and N(1.5, 1)

Random Intercepts from Mixture of Two Normals with Different Standard Deviations

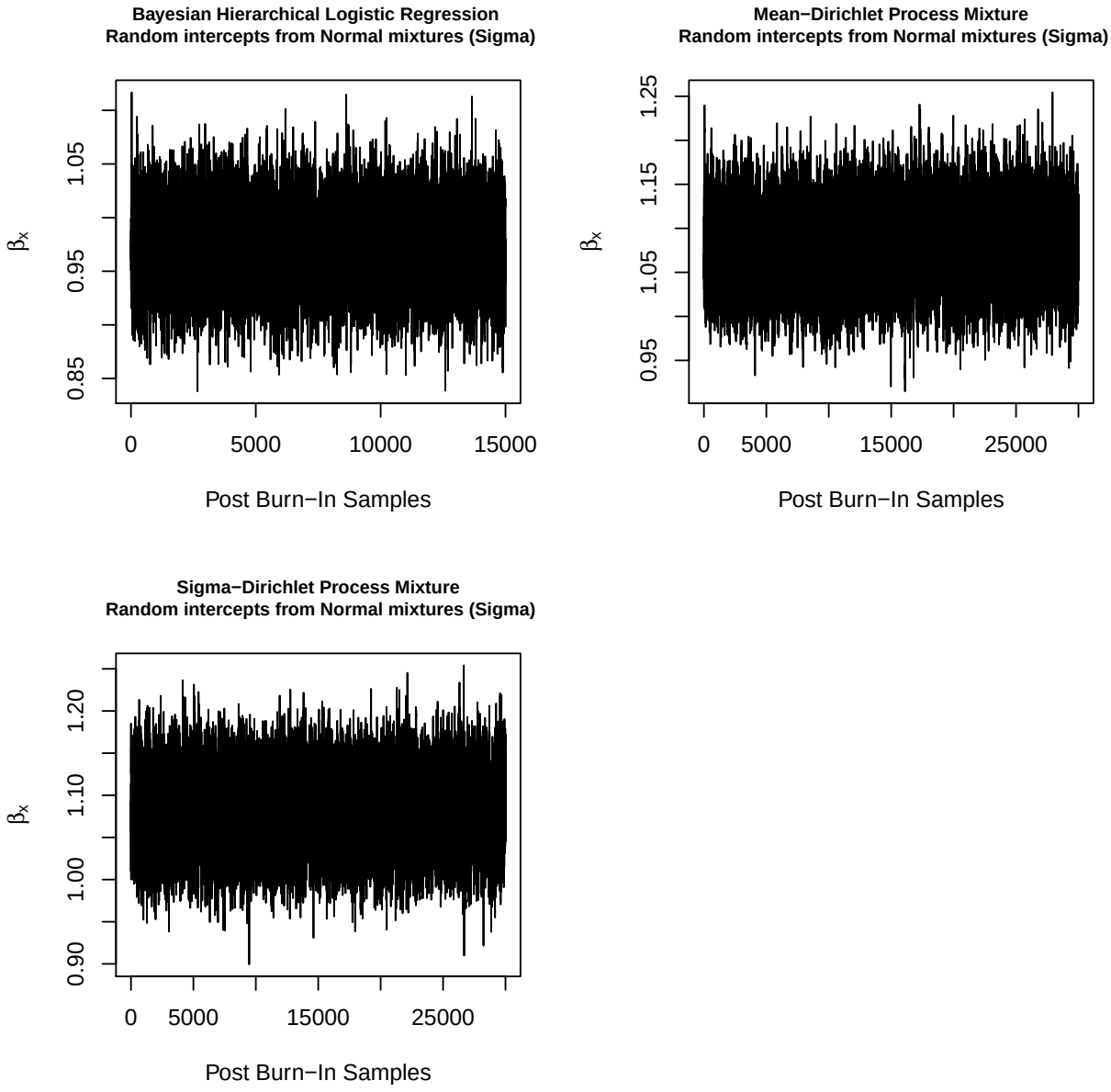


Figure B.3: Traceplots for Posterior samples under a simulation scenario when random intercepts are from a mixture $N(0, \sigma = 1)$ and $N(0, \sigma = \sqrt{5})$

B.1.2 Bayesian Survival Models

Normally Distributed Random Intercepts

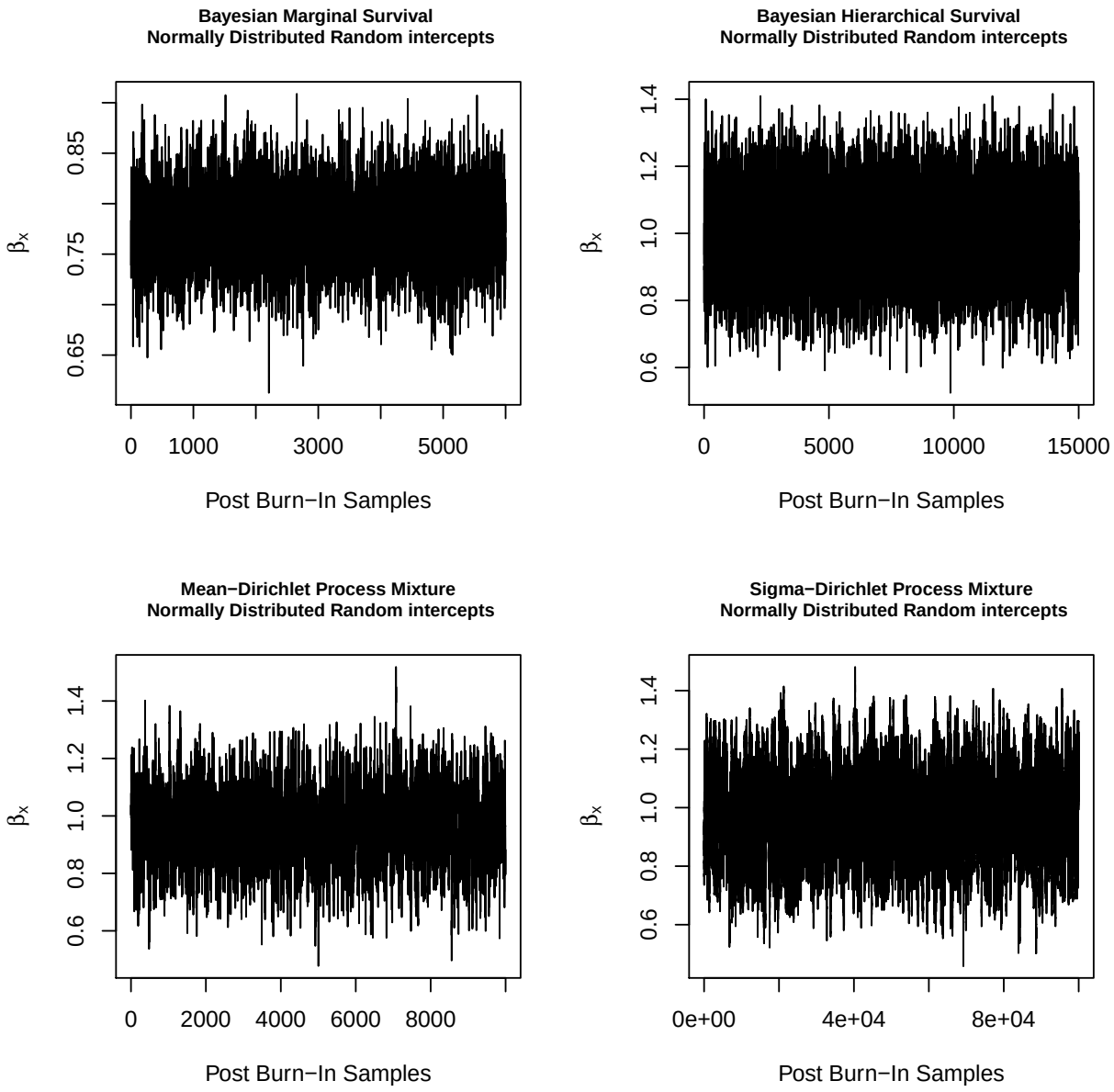


Figure B.4: Traceplots for Posterior samples under a simulation scenario when random intercepts are Nornally distributed

Random Intercepts from Mixture of Two Normals with Different Means

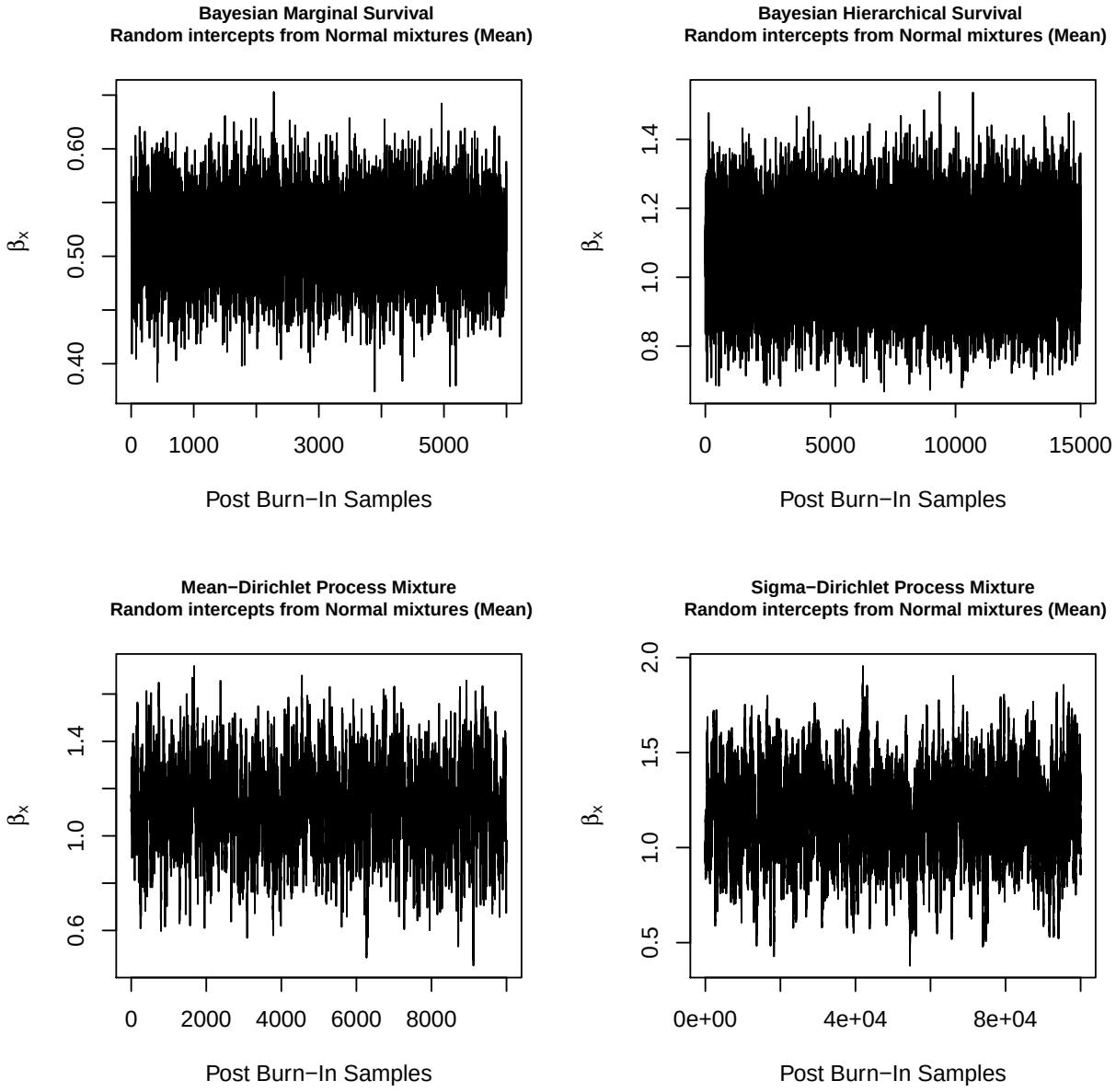


Figure B.5: Traceplots for Posterior samples under a simulation scenario when random intercepts are from a mixture N(-1.5, 1) and N(1.5, 1)

Random Intercepts from Mixture of Two Normals with Different Standard Deviations

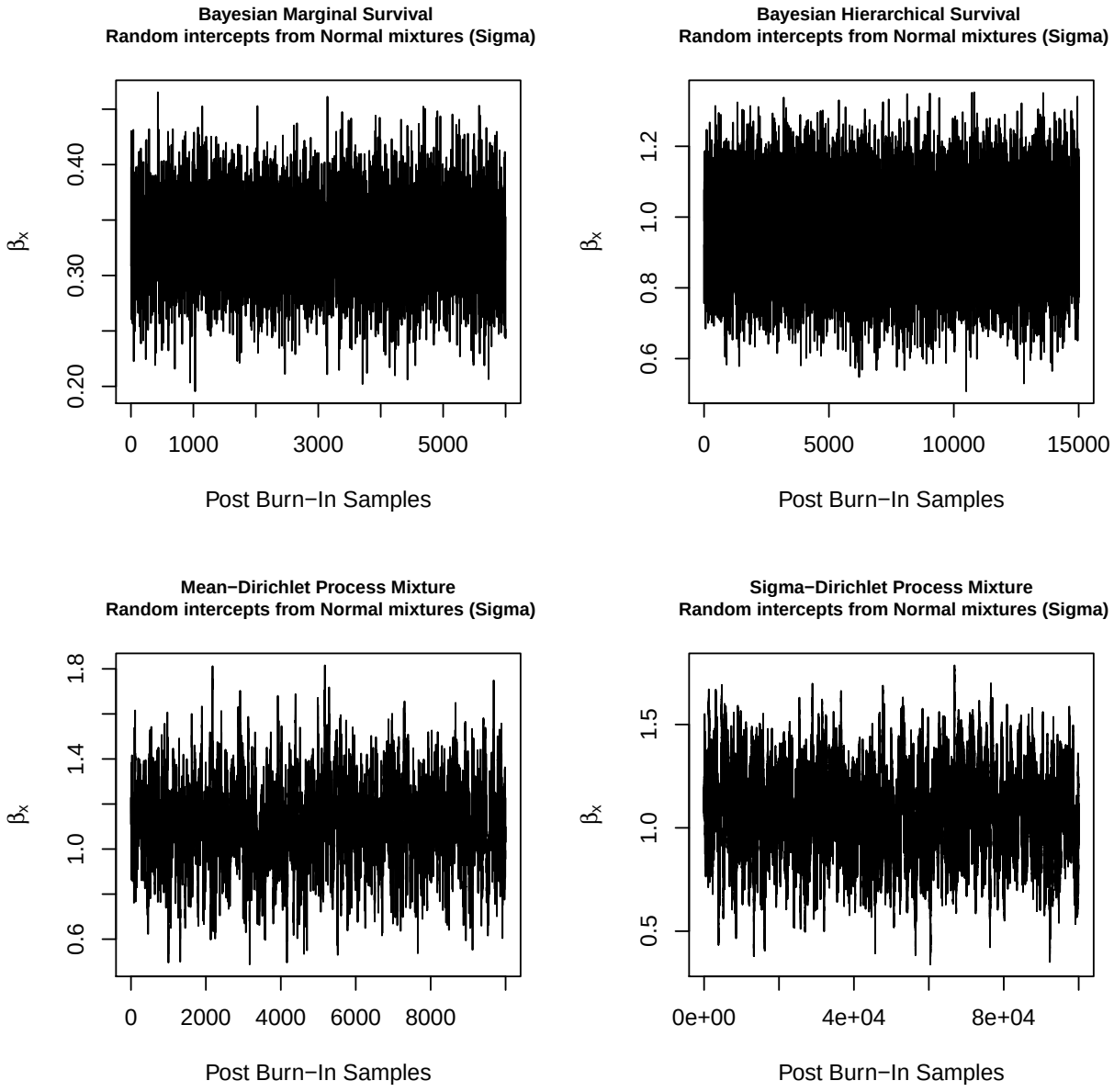


Figure B.6: Traceplots for Posterior samples under a simulation scenario when random intercepts are from a mixture $N(0, \sigma = 1)$ and $N(0, \sigma = \sqrt{5})$

B.2 Chapter 4 Traceplots

In this section, we provide traceplots for survival coefficients in the simulation studies presented in Section 4.3.

B.2.1 Model I

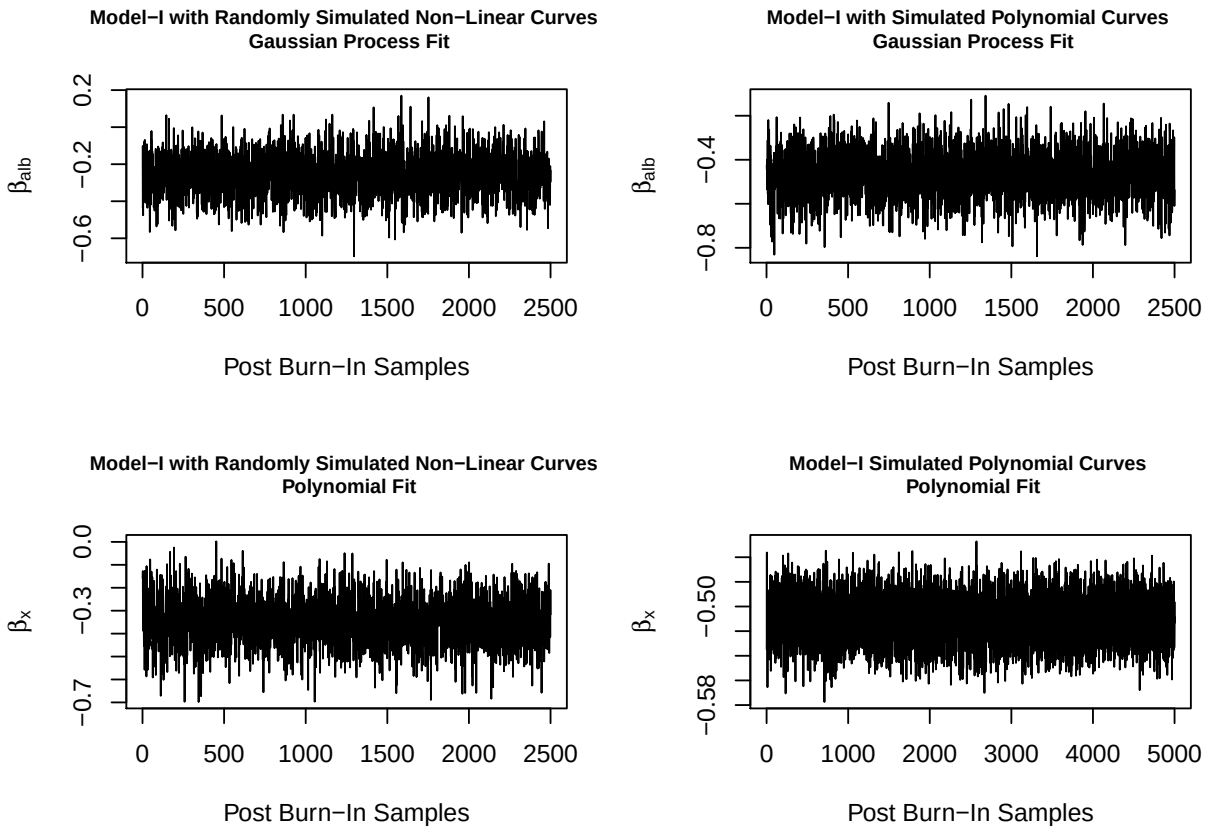


Figure B.7: Traceplot of the β_{alb} coefficient under the Model I simulation in Section 4.3

B.2.2 Model II

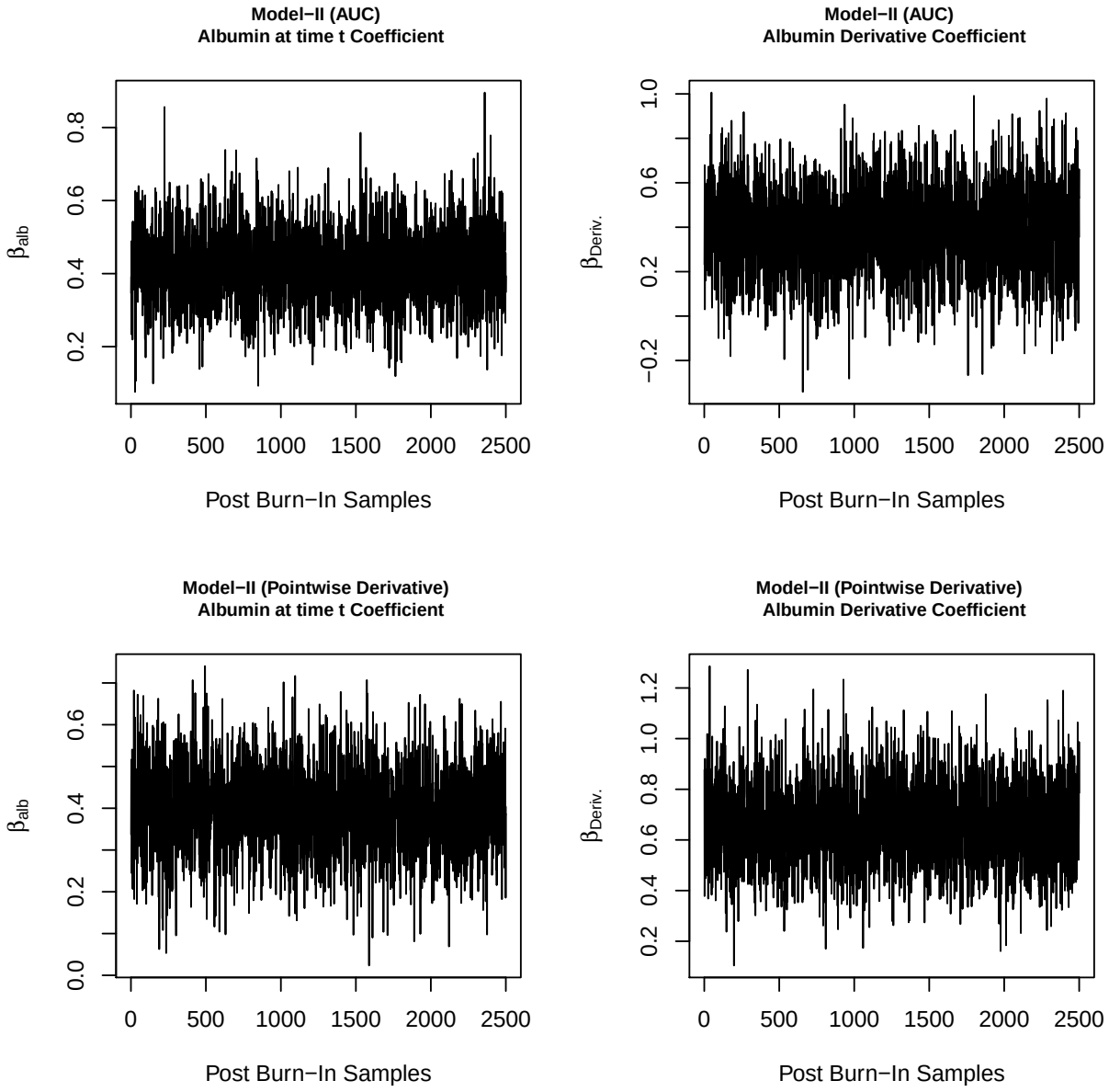


Figure B.8: Traceplot of the β_{alb} coefficient under the Model I simulation in Section 4.3

B.2.3 Model III

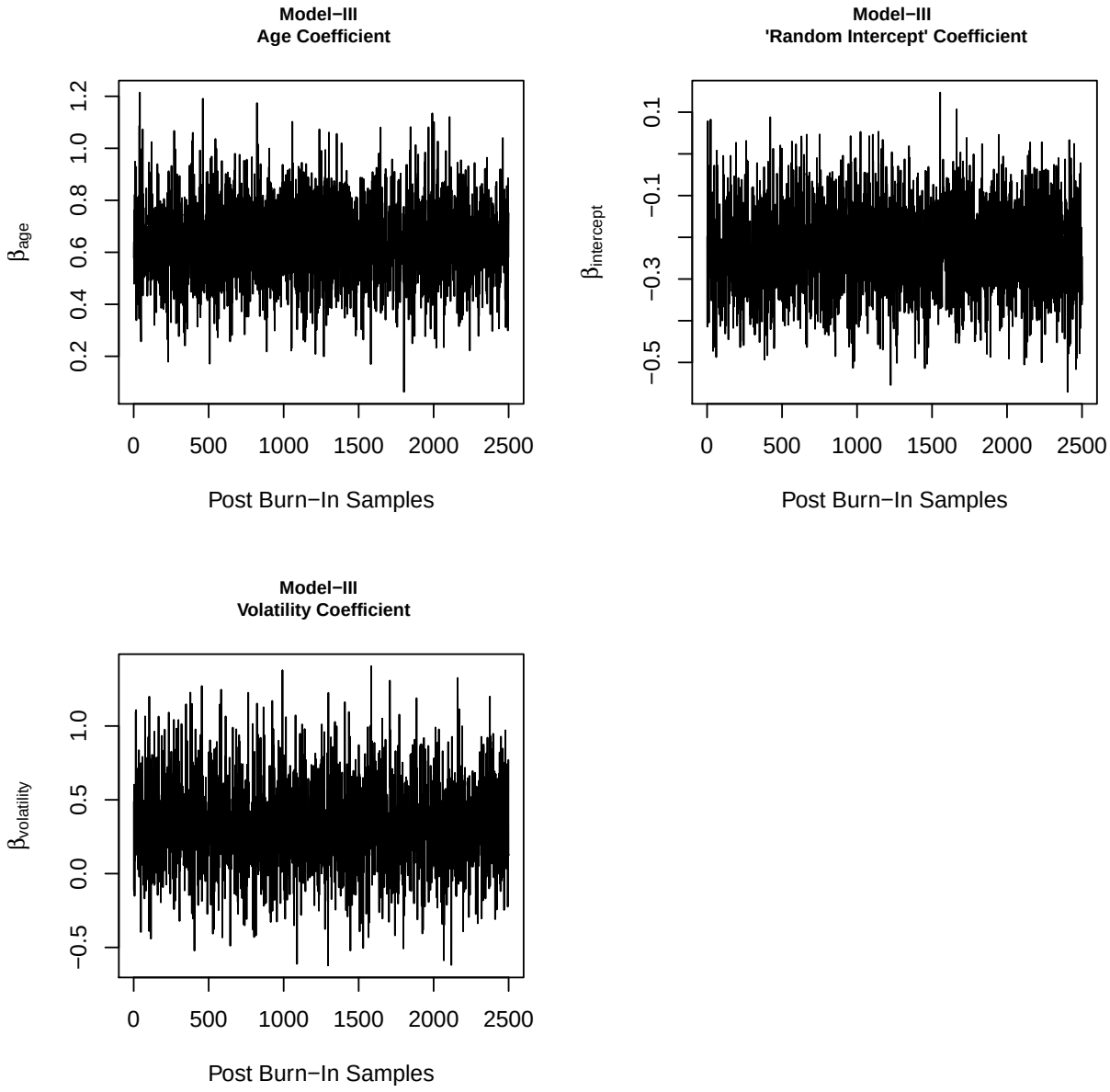
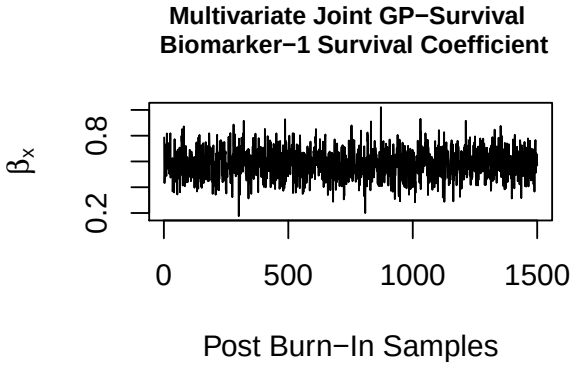


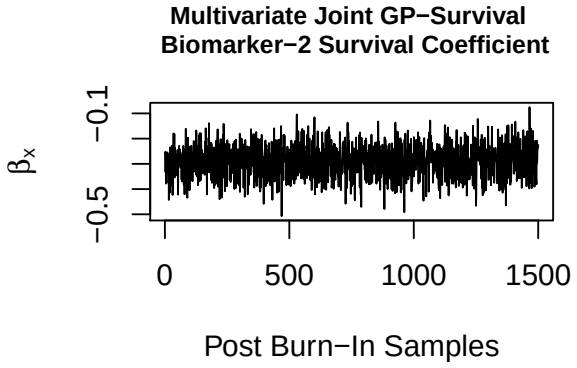
Figure B.9: Traceplot of the β_{alb} coefficient under the Model I simulation in Section 4.3

B.3 Chapter 5 Traceplots

In this section, we provide traceplots for survival coefficients β_{x_1} and β_{x_2} in the simulation studies presented in Section 5.3.

B.3.1 MGP Model





B.3.2 UGP Model