I, Yue Zhang, hereby submit this original work as part of the requirements for the degree of Doctor of Philosophy in Mathematics (Statistics).

It is entitled:
Bayesian Cox Models for Interval-Censored Survival Data

Student's name: Yue Zhang

This work and its defense approved by:

Committee chair: Xia Wang, Ph.D.
Committee member: Bin Zhang, Ph.D.
Committee member: Emily Kang, Ph.D.
Committee member: Bledar Konomi, Ph.D.
Committee member: Siva Sivaganesan, Ph.D.
Committee member: Seongho Song, Ph.D.
Bayesian Cox Models for Interval-censored Survival Data

A dissertation submitted to the
Graduate School
of the University of Cincinnati
in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

in the Department of Mathematical Sciences
of the College of Arts and Sciences

by

Yue Zhang
M.S. University of Cincinnati
July 2016

Committee:
Xia Wang, Ph.D., Chair
Bin Zhang, Ph.D., Co-Advisor
Siva Sivaganesan, Ph.D.
Seongho Song, Ph.D.
Emily Lei Kang, Ph.D.
Bledar Alex Konomi, Ph.D.
Abstract

Interval-censored data arise when failure times cannot be observed exactly but can only be determined to lie within an interval. In this dissertation, our research interest focuses on correlated survival data occur when individuals under study are clustered or experience multiple events of interest. We developed three novel Bayesian Cox models to handle different types of correlated interval-censored survival data. For clustered data, we utilized a shared frailty factor for unobserved correlation between observations within the same cluster. For spatially correlated data, we first used frailty for within cluster correlation, and then assigned a conditional autoregressive distribution prior for frailties to model the spatial dependency between clusters. In the aforementioned two frailty models, we also considered time-varying coefficient for temporal covariate effect. For bivariate data, we applied copula model to account for the dependence between outcomes. Simulation studies and analysis of real data examples illustrate the performance and applications of the proposed methods.
Acknowledgments

The acknowledgment section is a great opportunity to express my gratitude to all those people who have made this mission completed. This dissertation would not have been possible without the precious support of them during my life as a Ph.D. student, which I will cherish forever.

First and foremost, I am extremely grateful to Professor Bin Zhang and Professor Xia Wang for not only the help in finding this interesting topic, but also the advice and supervision through the whole dissertation process. Both of them spent countless hours in discussing project details, revising publication drafts and preparing conference presentations. Their patience, support, and encouragement helped me overcome many challenges during my dissertation research. I appreciate all their contributions of time, ideas, and funding to make my Ph.D. experience productive and stimulating.

For this dissertation, I would like to thank my committee faculty members: Professors Siva Sivaganesan, Seongho Song, Emily Lei Kang and Bledar Alex Konomi, for their reviews and comments on this research topic. I would also like to thank Professors Lin Fei, Nanhua Zhang and Bin Huang from Cincinnati Children’s Hospital Medical Center, for not only the continuous financial support, but also the project experience in statistical methodologies and applications.

Lastly, I would like to thank my family for all their love and encouragement. For my parents who raised me with a love of science and supported me in all my pursuits. I also
dedicate this dissertation to my aunt, who encouraged me to come to U.S. for graduate study
and taught me the values of hard work.
Table of Contents

List of Figures vi
List of Tables viii

1 Introduction 1
  1.1 Survival analysis .................................................. 1
  1.1.1 Censoring ......................................................... 3
  1.1.2 Cox model ........................................................ 5
  1.1.3 Time-varying covariate effect ..................................... 6
  1.2 Bayesian inference .................................................. 8
  1.2.1 Bayes’ theorem .................................................... 8
  1.2.2 Bayesian inference for parameters .................................. 9
  1.2.3 Prior distributions .............................................. 10
  1.2.4 Markov chain Monte Carlo (MCMC) methods ......................... 11
  1.3 Frailty models ..................................................... 17
  1.3.1 Independent frailty model ....................................... 17
  1.3.2 Spatial frailty model ........................................... 18
  1.4 Bivariate survival data ............................................. 18
  1.5 Outline ............................................................... 19

2 Independent frailty model with time-varying coefficient 20
  2.1 Introduction ......................................................... 21
  2.2 Model and the likelihood ............................................ 22
  2.3 Prior and posterior .................................................. 24
    2.3.1 Prior ........................................................... 24
    2.3.2 Posterior computation ......................................... 26
  2.4 Simulation ........................................................... 31
  2.5 Children’s dental health data ........................................ 37
  2.6 Discussion ........................................................... 40

3 Spatial frailty model with time-varying coefficient 42
List of Figures

1.1 Time-varying effect of the performance score on stroke readmission. 7

2.1 Estimate of coefficients for models with one covariate and three clusters. The solid lines in the middle are the true coefficients. The long dashed lines are medians of posterior means. The top and bottom short dashed lines are medians of 95% credible intervals. 34

2.2 Estimate of coefficients for models with two covariates and three clusters. The solid lines in the middle are the true coefficients. The long dashed lines are medians of posterior means. The top and bottom short dashed lines are medians of 95% credible intervals. 36

2.3 Estimate of coefficients in children’s dental health data. Black solid line is posterior mean. Black dashed lines are 95% credible intervals. 38

2.4 Estimate of frailties with 95% credible intervals in the children’s dental health data. 40

3.1 Estimate of coefficients for combinations with one covariate. The solid lines in the middle are the true coefficients. The long dashed lines in the middle are medians of posterior means. The top and bottom dot dash lines are median of 95% credible intervals. 59

3.2 Estimate of coefficients for combinations with two covariates. The solid lines in the middle are the true coefficients. The long dashed lines in the middle are medians of posterior means. The top and bottom dot dash lines are median of 95% credible intervals. 60

3.3 Maps of posterior means for the 45 spatial frailties. 63

3.4 Estimate of coefficients for smoking cessation data. Black solid line is posterior mean. Black dashed lines are 95% credible intervals. The difference of effect in reducing relapse risk between smoking intervention and usual care are much smaller for the time around the two peaks (1.8 years and 3.8 years) than neighbor time period. 65

3.5 Maps of posterior means for the spatial frailties $\omega_i$ over 51 zip codes areas in southeast Minnesota based on Models 1, 2 and 3. The white areas are those zip codes without data. 67
4.1 Trace plots and density plots for $\alpha$, $\beta_1$ and $\beta_2$. True values are 0.7, 1, 1, respectively. 80
List of Tables

1.1 Gibbs sampling algorithm ........................................... 12
1.2 Metropolis-Hastings algorithm ................................... 14
1.3 Reversible jump MCMC algorithm ................................. 16

2.1 Model specifications in the simulation study ...................... 32
2.2 Estimates of frailties and model comparison results .......... 35
2.3 Estimate of frailties with 95% credible intervals in the children’s dental health data 39

3.1 Specifications of six combinations ............................... 58
3.2 LPML comparison result between Model 2 and Model 3 .......... 62
3.3 Comparison of three models based on the number of larger biased frailty estimates out of 100 replicates .................. 63

4.1 Simulation results of the proposed method ..................... 82
4.2 Results of children’s neurodevelopmental data analysis .......... 84
Chapter 1

Introduction

This chapter gives a brief introduction of survival analysis, Bayesian inference and some commonly used sampling algorithms. It also presents the basic concepts of frailty models for clustered survival data and copula model for bivariate survival data.

1.1 Survival analysis

Survival analysis is used to analyze data in which the time until the event is of interest. For example, time to tumor recurrence, time until a machine part fails. Let $T$ denote the continuous nonnegative survival time response of a subject, and all functions with respect to $T$ are defined over the interval $[0, \infty)$. Let $f(t)$ represent the probability density function (pdf) of $T$, then the cumulative distribution function (cdf), $F(t)$, is given by

$$F(t) = P(T \leq t) = \int_0^t f(s)ds.$$
The probability that a subject will survive past time $t$ is defined as the survival function, $S(t)$, which takes the form

$$S(t) = 1 - F(t) = P(T \geq t) = \int_t^\infty f(s)ds.$$  \hspace{1cm} (1.1)

As $t$ ranges from 0 to $\infty$, the survival function in (1.1) has the following properties: (1) It is non-increasing; (2) At time $t = 0$, $S(t) = 1$. In other words, the probability of surviving past time 0 is 1. (3) At time $t = \infty$, $S(t) = S(\infty) = 0$. As time goes to infinity, the survival curve goes to 0. In theory, the survival function is smooth. In practice, we observe events on a discrete time scale (days, weeks, etc.).

The hazard function, $h(t)$, is the instantaneous rate of failure at time $t$ given no previous events

$$h(t) = \lim_{\Delta t \to 0} \frac{P(t < T \leq t + \Delta t | T > t)}{\Delta t}.$$  

The cumulative hazard describes the accumulated risk up to time $t$ and takes the form

$$H(t) = \int_0^t h(s)ds.$$  

Note that we have the following relations between these functions

$$h(t) = \lim_{\Delta t \to 0} \frac{S(t) - S(t + \Delta t)}{\Delta t S(t)}$$
$$= -\frac{S'(t)}{S(t)}$$
$$= -\frac{d}{dt}(\log(S(t))),$$
\[ S(t) = \exp \left( - \int_0^t h(s) ds \right) = \exp(-H(t)), \]
\[ f(t) = h(t) \exp \left( - \int_0^t h(s) ds \right) = h(t)S(t). \]

1.1.1 Censoring

A distinguishing feature of the field of survival analysis is censoring, i.e., the failure time is only partially known.

The most common type of censoring is right censoring, where only the lower bounds on event time are available for some subjects in the study. Right censoring could occur when a subject leaves the study before an event occurs, or the study ends before the event has occurred. The second type of censoring is left censoring, where only the upper bounds on event times are available for some subjects in the study. Left-censored data can be observed when the event of interest has already occurred before enrollment.

The third type of censoring is interval censoring, where the event time \( T \) cannot be observed, but can only be determined to lie in an interval. One special case of interval-censored data is case I or current status data (Sun and Kalbfleisch, 1993). This type of censoring means that each subject is observed only once for the status of the occurrence of the event of interest. Suppose that \( C \) is an “examination” or “observation” time. The only knowledge about the event time \( T \) is whether it has occurred before or after \( C \). One such example is given by the tumorigenicity study and in this situation, the time to tumor onset is usually of interest, but not directly observable (Dinse and Lagakos, 1983). As a matter of
fact, the exact measurement of the observation time is often the death or sacrifice time of the subject. In another important type of interval-censored data, case II interval-censored data (Groeneboom and Wellner, 1992), we only know that \( T \) has occurred either within some time interval, or before the left end point of the time interval, or after the right end point of the time interval. Precisely, let \((L, R]\) be a pair of observation times. The data observed is \((L, R, \delta_1, \delta_2, \delta_3)\), where

\[
\delta_1 = 1(T \leq L),
\]

\[
\delta_2 = 1(L < T \leq R),
\]

\[
\delta_3 = 1(T > R).
\]

Interval censoring often arises in longitudinal studies in which subjects are assessed only periodically at some specific times. For example, HIV infection time is only known to fall between the last visit time with a negative result and the first visit time with a positive result. Examples of interval-censored data in AIDS studies can be found in De Gruttola and Lagakos (1989); Jewell et al. (1994); Kim et al. (1993); Sun (1996); Shiboski and Jewell (1992). Besides AIDS, other studies in demographic, epidemiology and medical science also target interval-censored data. See for example, Diamond et al. (1986); Finkelstein (1986); Finkelstein and Wolfe (1985); Hoel and Walburg (1972); Self and Grossman (1986); Sun and Kalbfleisch (1996). However, most of the studies that have been conducted assumed independent subjects. In this dissertation, we will focus on developing methods to analyze interval-censored data when subjects are correlated.
1.1.2 Cox model

Cox model, also called proportional hazards model, was invented by the British statistician D.R. Cox (Cox, 1972). This is a milestone in statistical history due to its contribution in clinical trials and other medical studies. Cox model specifies that covariates have multiplicative effect on the hazard function of the failure time of interest. Let $T$ denote the time to the event of interest, the Cox model takes the form

$$h(t|x) = h_0(t) \exp(x^T \beta), \quad (1.2)$$

where $x$ is a vector of covariates, $\beta$ is the vector of covariate effects, and $h_0(t)$ is the baseline hazard function. It describes the dependence of the hazard on the time $t$. Cox model does not specify the distribution of the baseline hazard. It is more flexible than those parametric models with a particular probability distribution for event time, such as exponential, Weibull and Gompertz, which is the beauty of the model.

Another reason for its broad application is the simple and easy-to-understand interpretation of Cox model with respect to the ratio of hazards. Consider two individuals with covariates $x_1$ and $x_2$, which are scalars for simplicity. The ratio of their hazards at time $t$ is

$$\frac{h(t|x_1)}{h(t|x_2)} = \frac{h_0(t) \exp(x_1 \cdot \beta)}{h_0(t) \exp(x_2 \cdot \beta)}$$

$$= \frac{\exp(x_1 \cdot \beta)}{\exp(x_2 \cdot \beta)}$$

$$= \exp\{\beta(x_1 - x_2)\}.$$
In other words, \( h(t|x_1) \propto h(t|x_2) \), i.e., the hazards are proportional to each other and do not depend on time. In particular, the hazard for the individual with covariate \( x_1 \) is \( \exp\{\beta(x_1 - x_2)\} \) times that of the individual with covariate \( x_2 \). This term, \( \exp\{\beta(x_1 - x_2)\} \), is called the hazard ratio comparing \( x_1 \) to \( x_2 \).

To estimate regression parameters of Cox model with right-censored data, one only deals with the finite or parametric part of the semiparametric model, e.g., partial likelihood approach (Cox, 1972). Unlike the methods developed for right-censored data, estimating regression parameters under interval-censoring is more challenging since it involves estimation of both nonparametric baseline and coefficients simultaneously. To deal with this, Huang et al. (1996) proposed to use the efficient estimation approach, Pan (1999) extended the iterative convex minorant algorithm to the Cox model, Sinha et al. (1999) generated augmented event time in its observed time interval and parameters piecewisely via Markov chain Monte Carlo.

### 1.1.3 Time-varying covariate effect

The proportionality of the hazards is a fundamental assumption in the Cox model as specified in (1.2). It implies that the factors investigated have a constant impact on the hazard, or risk, over time. In practice, this might be violated when effect is time-dependent. As a result, misleading effect estimates can be derived, and significant effect in the early (or late) follow-up period may be missed. In this dissertation, the time-varying coefficient method is implemented to model time-varying covariate effect. In previous studies, Cai and Betensky
Figure 1.1: Time-varying effect of the performance score on stroke readmission.

(2003) provided a smoothed estimate of the hazard function by maximizing the penalized likelihood. Cai et al. (2007) used a local polynomial method for a marginal survival model. In such case model (1.2) can be extended to

\[ h(t | x) = h_0(t) \exp(x^T \beta(t)). \]  

(1.3)

The model 1.3 can be exemplified by a stroke study analyzed by Yu et al. (2013), where composite stroke performance score had time-varying effect on stroke readmission, see Figure 1.1. In the plot, the solid line is the coefficient of composite score over time; the dotted lines are 95% pointwise confidence interval; the dashed line is the performance effect in constant coefficient model. In addition, there is a “0” solid reference line in the plot.

People usually consider parameter estimation and dimension selection as two steps. In this dissertation, by applying reversible jump MCMC, the model automatically determines
the extent to which the temporal dynamics is needed for each covariate, resulting in smoother
and more stable curve estimates.

1.2 Bayesian inference

Back to 18th century, Bayes’ famous paper (Bayes, 1763) discussed the conditional prob-
ability of failure in a single trial, given some data on the previous number of failures. In
order to learn about the failure probability on the basis of observed data, Bayes proposed
a theorem relating conditional and marginal probabilities of random variables which is used
to calculate the required conditional probability.

1.2.1 Bayes’ theorem

Bayes’ theorem is usually stated in terms of probabilities for observable events. Let $A$ and
$B$ be events, then

$$p(A|B) = \frac{p(B|A)p(A)}{p(B)}.$$ 

The marginal probability of $A$, $p(A)$, is often referred to as the prior probability of $A$. The
word “prior” indicates “before taking account of the information in $B$.” The conditional
probability of $A$ given $B$, $p(A|B)$, is often referred to as the posterior probability of $A$ after
taking account of the information in $B$. The expression $p(B|A)$ is the conditional probability
of $B$ given $A$, and $p(B)$ is the marginal probability of $B$. 

8
1.2.2 Bayesian inference for parameters

In frequentist statistics only the data are assumed to be random variables with associated probability distributions; parameters are assumed to be fixed, and their associated $p$-values and confidence intervals are based on long-run frequency properties under repeated sampling of the data. From a Bayesian perspective, both data and parameters can have probability distributions, and so Bayes’ theorem can be used to learn about probabilities of unobservable parameters as well as observable events. For simplicity, let $\beta$ denote the single parameter and $t$ denote the data. Bayes’ theorem for inference about the parameter can be expressed as

$$ p(\beta|t) = \frac{p(t|\beta)p(\beta)}{p(t)}. $$

The interpretation is analogous to Bayes’ theorem: $p(\beta)$ is the prior distribution for $\beta$ and expresses the uncertainty about the values of $\beta$ before taking account of the observed data; $p(\beta|t)$ is the posterior distribution for $\beta$ and represents the uncertainty about $\beta$ conditional on the data $t$. The conditional distribution $p(t|\beta)$ describes how the data depend on the parameter $\beta$. Since $p(t)$ is a normalizing constant, Bayes’ theorem in this context is often expressed simply as

$$ p(\beta|t) \propto p(t|\beta)p(\beta). $$

Consider the likelihood function of $\beta$, say $L(\beta|t)$, such that $L(\beta|t) \propto p(t|\beta)$. By applying Bayes’ theorem

$$ p(\beta|t) \propto L(\beta|t)p(\beta). $$
Hence Bayes' theorem essentially states that: posterior $\propto$ likelihood $\times$ prior.

### 1.2.3 Prior distributions

A Bayesian analysis requires the assertion of a prior distribution for the unknown parameters. The prior distribution can be viewed as representing the current state of knowledge, or current description of uncertainty about the model parameters prior to data being observed. Approaches to choosing a prior distribution divide into two main categories: informative priors and non-informative priors (Gelman et al., 2014).

The informative priors are constructed based on the knowledge about the substantive knowledge from other data or elicited expert opinion if possible such that the priors properly reflect some information about the unknown parameters. A simple and commonly used class of informative priors are conjugate priors and non-conjugate priors. The chosen form for the conjugate priors result in a posterior which takes the same form as the priors. For example, when likelihood based on data $x$ is from binomial distribution $Binom(n, \theta)$, the beta prior $Beta(\alpha, \beta)$ leads to a beta posterior $Beta(\alpha + x, \beta + n - x)$. Prior information may not always conform to the particular form of a conjugate prior, then a non-conjugate prior is needed. For example, if data $x$ is from normal distribution $N(\theta, \sigma^2)$ and $\theta$ is known to be positive, possible choices of prior distribution are gamma distribution, truncated normal distribution and Cauchy distribution, etc.

The second main approach of choosing a prior distribution is to construct a non-informative
prior that represents ignorance about the model parameters. Besides non-informative, this type of distribution is also called objective prior distribution. Choosing a non-informative prior distribution is an attempt at objectivity by acting as though no prior knowledge about the parameters exists before observing the data. The appeal of this approach is that it directly addresses the criticisms of informative prior distributions as being subjectively chosen. Commonly used objective priors are flat prior, Jeffreys prior, diffuse priors and reference priors. More details were discussed by (Gelman et al., 2014).

1.2.4 Markov chain Monte Carlo (MCMC) methods

In Bayesian inference, Markov chain Monte Carlo methods is a widely used approach for sampling from a probability distribution based on constructing a Markov chain that has the desired distribution. Formally, a sequence of random variables $X^{(0)}, X^{(1)}, X^{(2)}, \ldots$ forms a Markov chain if, for all $t$, the distribution of the $(t + 1)^{th}$ variable in the sequence is given by

$$X^{(t+1)} \sim p_{\text{trans}}(x|X^{(t)} = x^{(t)}).$$

That is, the distribution of $X^{(t+1)}$ is independent of all other preceding values, $X^{(t-1)}, \ldots, X^{(0)}$ conditional on the current value of $X^{(t)}$. The right side of (1.4) is defined as the conditional probability of moving to any particular new value given the current value of the chain, which is also called the transition distribution of the Markov chain. Subject to fairly general regularity conditions (Cox and Miller, 1965), the marginal distribution of $X^{(t+1)}$ will converge to a unique stationary distribution as $t \to \infty$. A brief introduction of several MCMC
Table 1.1: Gibbs sampling algorithm

<table>
<thead>
<tr>
<th>Step 1. Initialize $x^{(0)} \sim p(x)$.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Step 2. for iteration $i = 1, 2, \ldots$ do</td>
</tr>
<tr>
<td>$x_1^{(i)} \sim p(X_1 = x_1</td>
</tr>
<tr>
<td>$x_2^{(i)} \sim p(X_2 = x_2</td>
</tr>
<tr>
<td>$\vdots$</td>
</tr>
<tr>
<td>$x_k^{(i)} \sim p(X_k = x_k</td>
</tr>
<tr>
<td>end for.</td>
</tr>
</tbody>
</table>

algorithms that are applied in this dissertation is given as follows.

**Gibbs sampling**

The Gibbs sampler (Geman and Geman, 1984; Gelfand and Smith, 1990; Casella and George, 1992) is one of the most widely used algorithms for simulating Markov chains. Consider the random variables $X_1, X_2, \ldots, X_k$. We start by setting these variables to their initial values $X_1^{(0)}, X_2^{(0)}, \ldots, X_k^{(0)}$ (often values sampled from a prior distribution $p$). Then $X_1^{(i)}, X_2^{(i)}, \ldots, X_k^{(i)}$, $i = 1, 2, \ldots$, are sampled from the distribution of that variable conditional on all other components sampled. Table 1.1 details the algorithm of Gibbs sampler.

The beauty of Gibbs sampling is that simulation from a complex, high-dimensional joint posterior distribution is reduced to a sequence of algorithm for sampling from one dimension.
or low dimensional distributions. Because we initialize the algorithm with random values, the samples simulated based on this algorithm at early iterations may not necessarily be representative of the actual posterior distribution. However, the theory of MCMC guarantees that the stationary distribution of the samples generated under Gibbs sampling with a sufficiently large number of iterations is the target joint posterior that we are interested in.

**Metropolis-Hastings**

The Metropolis-Hastings algorithm was developed by Metropolis et al. (1953), and subsequently generalized by Hastings (1970). The Metropolis-Hastings algorithm simulates samples from a probability distribution by making use of the full joint density function and proposal distributions for each of the variables of interest. Consider the random variables \(X_1, X_2, \ldots, X_k\). We start by setting these variables to their initial values \(X_1^{(0)}, X_2^{(0)}, \ldots, X_k^{(0)}\), which are often sampled from a prior distribution \(p\). The main loop of Metropolis-Hastings algorithm consists of three components: (1) Generate a proposal (or a candidate) sample \(x_{\text{cand}}\) from the proposal distribution \(q(x^i|x^{(i-1)})\); (2) Compute the acceptance probability via the acceptance function \(\alpha(x_{\text{cand}}|x^{(i-1)})\) based upon the proposal distribution; (3) Accept the candidate sample with probability \(\alpha\), the acceptance probability, or reject it with probability \(1 - \alpha\). Table 1.2 provides the details of a generic Metropolis-Hastings algorithm.
Table 1.2: Metropolis-Hastings algorithm

<table>
<thead>
<tr>
<th>Step 1. Initialize $x^{(0)} \sim p(x)$.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Step 2. for iteration $i = 1, 2, ...$ do</td>
</tr>
<tr>
<td>Propose: $x^{cand} \sim q(x^{(i)}</td>
</tr>
<tr>
<td>Acceptance probability:</td>
</tr>
<tr>
<td>$\alpha(x^{cand}</td>
</tr>
<tr>
<td>$u \sim \text{Uniform } \mathcal{U}(u; 0, 1)$</td>
</tr>
<tr>
<td>if $u &lt; \alpha$ then</td>
</tr>
<tr>
<td>Accept the proposal: $x^{(i)} = x^{cand}$</td>
</tr>
<tr>
<td>else</td>
</tr>
<tr>
<td>Reject the proposal: $x^{(i)} = x^{(i-1)}$</td>
</tr>
<tr>
<td>end if</td>
</tr>
<tr>
<td>end for.</td>
</tr>
</tbody>
</table>

Reversible jump MCMC

The reversible jump MCMC (Green, 1995) provides a general framework for simulation in which the dimension of the parameter space can vary between iterations of the Markov chain. The reversible jump sampler can be viewed as an extension of the Metropolis-Hastings algorithm onto more general state spaces. In the simulation analysis, the parameters in the current state $(k, \theta_k)$ and new state $(k', \theta'_k)$ have different dimensions $k$ and $k'$. We start
by initial values \((k^{(0)}, \theta_k^{(0)})\), where \(\theta_k^{(0)}\) is often sampled from a prior distribution \(p\). The algorithm of reversible jump MCMC is given in Table 1.3.

After setting initial value, the first step of main loop is choosing a move type. The probability to select a birth move is \(b_k\), in which the dimension is increased from \(k\) to \(k + 1\). The probability to select a death move is \(d_k\), in which the dimension is decreased from \(k\) to \(k - 1\). For update move, the selected probability is \(1 - b_k - d_k\), and the dimension remains fixed to \(k\). Green (1995) introduced a way to calculate \(b_k\) and \(d_k\). For \(k = 0\) the death move is impossible, so that \(d_0 = 0\), whereas for \(k = k_{\text{max}}\) the birth move is impossible, so that \(b_{k_{\text{max}}} = 0\). For all other values of \(k\) the birth and death probabilities are taken to be

\[
b_k = c \times \min \left\{ 1, \frac{p(k + 1 | \lambda)}{p(k | \lambda)} \right\},
\]

\[
d_{k+1} = c \times \min \left\{ 1, \frac{p(k | \lambda)}{p(k + 1 | \lambda)} \right\},
\]

where \(p(\cdot | \lambda)\) denotes a Poisson distribution with parameter \(\lambda\), and the constant \(c\) is a tuning parameter to ensure that \(b_k + d_k \leq 0.9\) for all \(k = 0, 1, ..., k_{\text{max}}\).

For birth move, suppose the current state is \((k, \theta_k)\) and we wish to propose to a move to a new state \((k', \theta_{k'}')\), where \(k' = k + 1\). In order to match dimensions between the two states, a random variable \(u\) is generated from know density \(q_{d_k \rightarrow k'}(u)\). The current state \(\theta_k\) and random variable \(u\) are then mapped to the new state \(\theta_{k'} = g_{k \rightarrow k'}(\theta_k, u)\) through a one-to-one mapping function \(g_{k \rightarrow k'} : \mathcal{R}^{n_k} \times \mathcal{R}^1 \rightarrow \mathcal{R}^{n_{k'}}\). The acceptance probability of update move is given by

\[
\alpha[(k, \theta_k), (k', \theta_{k'}')] = \min \left\{ 1, \frac{p(x | k', \theta_{k'}')p(k', \theta_{k'}')q(k' \rightarrow k)}{p(x | k, \theta_k)p(k, \theta_k)q(k \rightarrow k')q_{d_k \rightarrow k'}(u)} \left| \frac{\partial g_{k \rightarrow k'}(\theta_k, u)}{\partial (\theta, u)} \right| \right\}, \quad (1.5)
\]
Table 1.3: Reversible jump MCMC algorithm

Step 1. Initialize $k^{(0)}$ and $\theta^{(0)}_k \sim p(\theta_k)$.

Step 2. for iteration $i = 1, 2, \ldots$ do

Sample $u \sim \text{Uniform} \ U(u; 0, 1)$

if $u \leq b_k$ then birth move

else if $u \leq b_k + d_k$ then death move

else update move

end if

end for.

where $x$ is observed data, $q(k \rightarrow k')$ denotes the probability of proposing a move from $(k, \theta_k)$ to $(k',\theta'_k)$, and the final term is the determinant of the Jacobin matrix. (1.5) can also be expressed as

$$\text{Likelihood ratio} \times \text{Prior ratio} \times \text{Proposal ratio} \times \text{Jacobian term}.$$ 

The acceptance probability of death move, from $(k',\theta'_k)$ to $(k, \theta_k)$, is the reciprocal of the acceptance probability of birth move

$$\alpha[(k',\theta'_k), (k, \theta_k)] = \frac{1}{\alpha[(k, \theta_k), (k', \theta'_k)]}.$$ 

For update move, $k$ is fixed, and the parameters $\theta_k$ can be updated according to any Monte Carlo updating scheme, e.g., adaptive rejection sampling (Gilks and Wild, 1992).
1.3 Frailty models

The concept of frailty provides a suitable way to introduce random effects in the model to account for association and unobserved heterogeneity. In its simplest form, a frailty is an unobserved random factor that modifies multiplicatively the hazard function of an individual or a group or cluster of individuals. The term frailty was first introduced by Vaupel et al. (1979) to describe individual heterogeneity or unknown risk factors of the hazard function. Two categories of frailty models are considered in this dissertation. One is independent frailty model, and the other is spatial frailty model.

1.3.1 Independent frailty model

In this situation, individuals in a group $i$ are supposed to share the same frailty $\omega_i$. The conditional hazard for individual $j$ in group $i$ is

$$h(t_{ij}|\omega_i) = \omega_i h(t_{ij}), \quad (1.6)$$

where $h(t_{ij}) = h_0(t_{ij}) \exp(x_{ij}^T \beta)$ in the Cox regression model. The $\omega_i$ are independent identically distributed following a chosen distribution, e.g., gamma distribution Clayton (1978), log-normal distribution McGilchrist and Aisbett (1991). The model assumes that all time observations are independent given the values of the frailties. In other words, it is a conditional independent model. The value of $\omega_i$ is constant over time and common to the individuals in the group and thus responsible for creating dependence.
1.3.2 Spatial frailty model

The independent assumption of frailties in (1.6) is not always reasonable. In this dissertation, we consider hierarchical survival models for datasets which are spatially correlated. Frailties \( \omega_i \) corresponding to strata in closer proximity to each other might also be similar in magnitude. One way to model such spatial dependence of the strata is based on neighborhood information, where we use only the positions of the strata relative to each other, e.g., which counties neighbor with others. The spatial frailty distribution assumes that \( \omega_i \) is defined only on discretely indexed regions. Here we apply a conditionally autoregressive (CAR) model (Besag et al., 1991; Banerjee et al., 2003) such that \( \omega_i | \lambda \sim \text{CAR}(\lambda) \) with hyperparameter \( \lambda \). See Section 3.2 for details.

1.4 Bivariate survival data

In survival analysis, one area of interest is processes where each individual may experience multiple events, which may be correlated. See for example, the time to development milestones of language and motor skills in children’s autism research in Section 4.5. Consider the two event times, \( T_1 \) and \( T_2 \), the interest will focus on the joint survival

\[
S(t_1, t_2) = p(T_1 > t_1, T_2 > t_2),
\]

where \( S(t, t) \) is the probability that the two events of interest to one individual don’t occur until time \( t \). In this dissertation, we consider a copula model (Clayton, 1978) for joint
survival function of $T_1, T_2$ such that

$$S(t_1, t_2) = C_\alpha(S_1(t_1), S_2(t_2)).$$

Here, $S_1$ and $S_2$ denote the marginal survival functions of $T_1$ and $T_2$, respectively, $C_\alpha$ is a genuine survival function on the unit square (see Section 4.3 for details), and $\alpha \in \mathcal{R}$ is a global association parameter.

### 1.5 Outline

The rest of the thesis is organized as follows. In Chapter 2, independent frailty model is considered with time-varying coefficient for the interval-censored survival data. In Chapter 3, we consider Cox model with time-varying coefficient, and the correlated interval-censored subjects are modeled by spatial frailties. In Chapter 4, we model bivariate data by applying copula model. Chapter 5 summarizes this thesis with discussion of future work.
Chapter 2

Independent frailty model with time-varying coefficient

In this chapter, we consider a Bayesian approach for clustered interval-censored data under a dynamic Cox regression model. Some methods that incorporate right censoring have been developed for clustered data with temporal covariate effects. However, interval-censored data analysis under the same circumstance are much less developed. In this chapter, we estimate piecewise constant coefficients based on a dynamic Cox regression model under Bayesian framework. The dimensions of coefficients are automatically determined by reversible jump Markov chain Monte Carlo algorithm. Meanwhile, we use a shared frailty factor for unobserved heterogeneity or for statistical dependence between observations. Simulation studies are conducted to evaluate the performance of the proposed method. The methodology is exemplified with a pediatric study on children’s dental health data.
2.1 Introduction

The most popular semiparametric regression model in survival literature is proportional hazards model which is also referred as Cox model (Cox, 1972). It specifies that covariates have multiplicative effect on the hazard function of the failure time of interest. Many approaches have been developed for interval-censored data under Cox model. However, there are limitations in existing models. One example is that the relative risk of two subjects may change over time. Hence, it is important to detect the temporal effects on the failure time.

The time-varying coefficients in right-censored data can be estimated by several ways, such as partial likelihood approach (Zucker and Karr, 1990), histogram sieve procedures (Higle and Sen, 1991), and one-step estimation procedure for the cumulative parameter function (Martinussen and Scheike, 2002; Martinussen et al., 2002). However, Cox models with time-varying coefficients for interval-censored data are much less developed. In a recent study, Wang et al. (2013) proposed a Bayesian extension of Cox model by applying an efficient reversible jump Markov chain Monte Carlo (MCMC) algorithm (Green, 1995) and putting dynamics on all coefficients as well as the baseline hazard, which were specified as piecewise constants. However, their methods only considered the case that the subjects are independent, which may not be realistic in some applications. For instance, a lot of clinical trials are multi-center studies especially for rare disease. Thus, the correlation of subjects within each cluster becomes crucial and needs to be addressed in the analysis. The motivating example is a longitudinal oral health study conducted in Flanders (Belgium) — the Signal
Tandmobiel (Vanobbergen et al., 2000). The aim was to access the oral health condition of Flemish school children and to determine the benefit of the intervention. The outcome of interest is the time to emergence of permanent tooth, which is interval-censored due to the annual examination scheme. Children involved in this study were from five provinces. The correlation of the subjects within the same province is nonignorable.

In this study, we propose a frailty Cox regression model when the subjects are correlated. The model also allows the existence of time-varying coefficients. A Bayesian approach is discussed with efficient implementation. The remainder of the chapter is organized as follows. Section 2.2 discusses general data structure, model and associated likelihood function. Section 2.3 describes prior specification and posterior computation details. Section 2.4 shows simulation results of proposed method and comparison with Wang’s model (Wang et al., 2013). A children’s dental health data is analyzed in Section 2.5. Conclusions and discussions are enclosed in Section 2.6.

2.2 Model and the likelihood

Let \( T_{i,j} \) denote the survival time for the \( j^{th} \) subject in the \( i^{th} \) cluster. Assume there are \( n \) clusters in a study and \( m_i \) subjects in each cluster such that \( i = 1, 2, ..., n, j = 1, 2, ..., m_i \), hence there are \( N = \sum_{i=1}^{n} m_i \) subjects in total. Cox model with time-varying regression coefficients is considered. Conditional on a \( p \)-dimensional vector of covariates, \( x_{i,j} \), and the unobserved frailty random variable \( \omega_i \) for the \( i^{th} \) cluster, the hazard function can be written
as

$$\lambda(t|\omega_i, x_{i,j}) = \lambda_0(t)\omega_i \exp(x_{i,j}^T \beta(t)),$$

where $\lambda_0(t)$ is an unknown baseline hazard function common to all subjects, $x_{i,j}$ is the $p \times 1$ covariate vector for the $j^{th}$ subject in the $i^{th}$ cluster and $\beta(t)$ is the $p$-dimensional regression coefficient function of main interest. Shared frailty model, the most common type of frailty model, is used for within-cluster dependence. Note that “shared frailty” implies that individuals in the same group share the common frailty parameter. The frailty $\omega_i$ is assumed to follow a parametric distribution, which can either take on the format of finite mean frailty distributions including but not limited to gamma and log-normal distribution; or take on the format of infinite mean distributions such as the positive stable distribution (Ibrahim et al., 2005).

For interval-censored data, unobserved event time $T_{i,j}$ is located in censoring interval $(L_{i,j}, R_{i,j}]$. The contribution of subject $j$ in $i^{th}$ cluster to the observed data likelihood is

$$\Pr(T_{i,j} \in (L_{i,j}, R_{i,j}]|\omega_i, x_{i,j}) = \Pr(T_{i,j} > L_{i,j}|\omega_i, x_{i,j}) - \Pr(T_{i,j} > R_{i,j}|\omega_i, x_{i,j}).$$

If the unobserved event time $T_{i,j}$’s are given, the interval-censored data reduce to right-censored data. In the proposed model, both $\lambda_0(t)$ and $\beta(t)$ are assumed to be left continuous step functions, where both the number of jumps and the locations of the jumps are random and can be estimated. Let $k = 1, 2, ..., K$ denote all the ordered grids and $0 = \tau_0 < \tau_1 < \tau_2 < ... < \tau_K < \infty$ be corresponding time points. The length of each grid may be taken to be sufficiently small to approximate any hazard and coefficient function. Here we assume
these $K$ grids contain all potential jump points. Let $\lambda_k = \lambda_0(\tau_k)$, and $\beta_k = \beta(\tau_k)$ denote the baseline hazard function and coefficient function evaluated at each grid; $dN_{i,j,k}$ indicates whether or not the event time $T_{i,j}$ falls within the $k^{th}$ interval of grid, i.e., $dN_{i,j,k} = 1(T_{i,j} \in (\tau_{k-1}, \tau_k])$; $Y_{i,j,k}$ denotes the at-risk variable. If $dN_{i,j,k} = 1$ for some value $k$, $Y_{i,j,l} = 1$ for $l < k$ and $Y_{i,j,l} = 0$ for $l > k$, while $Y_{i,j,k} = (T_{i,j} - \tau_{k-1})/\Delta_k$ for $l = k$, where $\Delta_k = \tau_k - \tau_{k-1}$ is the width of the $k^{th}$ interval. The augmented likelihood function for $j^{th}$ subject of $i^{th}$ cluster is

$$
\ell_{i,j}(\Theta|\{dN_{i,j,k}, Y_{i,j,k}\}_{k=1}^{K}, \omega_i, x_{i,j})
= \prod_{k=1}^{K} \{\lambda_k \omega_i \exp(x_{i,j}^T \beta_k)\}^{dN_{i,j,k}} \exp\{-\Delta_k \lambda_k \omega_i \exp(x_{i,j}^T \beta_k)Y_{i,j,k}\},
$$

where $i = 1, 2, ..., n$, $j = 1, 2, ..., m_i$ and $\Theta = \{\lambda_k, \beta_k; k = 1, 2, ..., K\}$.

### 2.3 Prior and posterior

#### 2.3.1 Prior

In the following description, we use $\theta(t)$ to denote either $\log \lambda_0(t)$ or one $\beta(t)$. Assume that the number of jumps $P$ in $\theta(t)$ follows a discrete uniform distribution ranging from 1 to $K$.

For a fixed $P$, the jump times $0 < \tau_1 < \tau_2 < ... < \tau_P = \tau_K$ are randomly selected from all time grids except the last one. Given $P$ and jump times, a hierarchical Markov type process
prior for $\theta(t)$ proposed by Wang et al. (2013) is specified as follows

$$
\theta(\tau_1) \sim \mathcal{N}(0, a_0 \nu), \\
\theta(\tau_p) | \theta(\tau_{p-1}) \sim \mathcal{N}(\theta(\tau_{p-1}), \nu), \quad p = 2, 3, ..., P, \\
\nu \sim \mathcal{IG}(\alpha_0, \eta_0),
$$

where $a_0 > 0$ is a hyper-parameter which can be chosen as a large number to reflect higher uncertainty in the prior input, and $\mathcal{IG}(\alpha_0, \eta_0)$ denotes an inverse-gamma distribution with shape parameter $\alpha_0$ and scale parameter $\eta_0$ such that the mean is $\eta_0 / (\alpha_0 - 1)$. Similar priors have been used in generalized additive models (Fahrmeir and Lang, 2001; Brezger and Lang, 2006). In order to compare simulation results, we set $a_0 = 100, \alpha_0 = 2, \eta_0 = 1$, which are the same as Wang et al. (2013). Gamma distribution, the most commonly used finite mean distribution, is used to model the frailty term $\omega_i$. For finite mean frailty distributions, we need the mean of the frailty distribution to be unity in order for the parameters of the model to be identifiable (Ibrahim et al., 2005). Thus, we assume

$$
\omega_i \overset{\text{i.i.d.}}{\sim} \mathcal{G}(\kappa^{-1}, \kappa^{-1}), \quad i = 1, 2, ..., n,
$$

where $\kappa$ is the variance of the $\omega_i$'s and larger values of $\kappa$ imply greater heterogeneity among clusters. Let $\eta = \kappa^{-1}$ for notational convenience. Vague hyper priors for $\eta$ are commonly used, like uniform distribution $\mathcal{U}(0, a)$ with some large $a$ or gamma distribution $\mathcal{G}(b, b)$ with $b$ close to zero. In this study, vague gamma prior $\mathcal{G}(0.001, 0.001)$ (Ibrahim et al., 2005) is
used, which is denoted as $\pi_{\eta}(\cdot)$. The joint prior density is proportional to

$$
\pi_{\eta}(\eta) \prod_{i=1}^{n}\{\omega_i^{\eta-1}\exp(-\eta \omega_i)\} \frac{\eta_0^{\alpha_0}}{\Gamma(\alpha_0)} \nu^{-\alpha_0-1} \exp\left(-\frac{\eta_0}{\nu}\right) \nu^{-\frac{\nu}{2}} \exp\left\{-\frac{\theta(t)^2}{2\alpha_0\nu}\right\} \times \prod_{p\geq 2} \exp\left[-\frac{\left\{\theta(\tau_p) - \theta(\tau_{p-1})\right\}^2}{2\nu}\right].
$$

### 2.3.2 Posterior computation

The posterior samples are obtained under Gibbs sampling framework based on $j^{th}$ subject of the $i^{th}$ group observed in the $k^{th}$ time interval, where $i = 1, 2, ..., n$, $j = 1, 2, ..., m_i$, $k = 1, 2, ..., K$, and $K$ is the total number of time grids. The parameters of interest include $\theta(t)$, the event time $T_{i,j}$’s, the event indicators $dN_{i,j,k}$’s and at-risk variables $Y_{i,j,k}$’s as well as frailty term $\omega_i$’s. Let $D = \{dN_{i,j,k}, Y_{i,j,k}\}$, $\Theta = \{\theta(t)\}$, $W = \{\omega_i\}$. The Gibbs sampling algorithm draws $D$, $\Theta$, $\nu$, $W$ and $\eta$ iteratively, where $\nu$ and $\eta$ are hyper-parameters.

The first step is to sample event time $T_{i,j}$, event indicators $dN_{i,j,k}$’s and at-risk variables $Y_{i,j,k}$’s for augmented data given $\Theta$ and $W$. This can be decomposed into two steps.

1. Locate grid interval for each event time. For finite interval-censored subject, event indicators $dN_{i,j,k}$’s follow a multinomial distribution with size 1 and probability vector
\((e_{i,j,1}, e_{i,j,2}, \ldots, e_{i,j,k})\), where for \(k = 1, 2, \ldots, K\),

\[
e_{i,j,k} = \frac{p_{i,j,k} \mathbb{1}(\tau_k \in (L_{i,j}, R_{i,j}])}{\sum_{\tau \in (L_{i,j}, R_{i,j})} p_{i,j,\tau}},
\]

\[
p_{i,j,k} = \begin{cases} 
\exp \left\{ - \sum_{l=1}^{k-1} \Delta_l \lambda_l \omega_i \exp(x^T_{i,j} \beta_k) \right\} - \exp \left\{ - \sum_{l=1}^{k} \Delta_l \lambda_l \omega_i \exp(x^T_{i,j} \beta_k) \right\} & \text{if } k > 1, \\
1 - \exp \left\{ - \Delta_1 \lambda_1 \omega_i \exp(x^T_{i,j} \beta_1) \right\} & \text{if } k = 1.
\end{cases}
\]

Thus, if observed interval \((L_{i,j}, R_{i,j}]\) only covers one time grid \(\tau_k\), then \(dN_{i,j,k} = 1\) and all other event indicators equal 0. Otherwise, if \((L_{i,j}, R_{i,j}]\) covers multiple time grids \(\tau_k\)’s, \(dN_{i,j,k}\) is sampled from multinomial distribution with probability vector calculated based on these covered time grids \(\tau_k\)’s. In other words, the time grid \((\tau_{k-1}, \tau_k]\) with \(dN_{i,j,k} = 1\) is sampled in this step.

(2) Within selected time grids, the exact time \(T_{i,j}\) follows a doubly truncated exponential distribution on \((\tau_{k-1}, \tau_k]\) with a distribution function

\[
F(u) = \frac{1 - \exp\{-\lambda_k (u - \tau_{k-1}) \omega_i \exp(x^T_{i,j} \beta_k)\}}{1 - \exp\{-\lambda_k \Delta_k \omega_i \exp(x^T_{i,j} \beta_k)\}}.
\]

Then \(T_{i,j}\) is sampled by inverse distribution method, and at-risk variables \(Y_{i,j,k}\)’s are calculated as defined in Section 2.2.

The next step is to sample each component of baseline hazard \(\log \lambda_0(t)\), regression coefficient \(\beta(t)\) given \(D\) and \(W\). Reversible jump MCMC algorithm (Green, 1995) is applied here.
because the number of jumps \( P \) is random and dimension of posterior distribution could vary from iteration to iteration. The probability of taking a birth, death and update move are set as 0.3, 0.3, and 0.4 (Wang et al., 2013), respectively.

(1) Update move. In this step, both \( P \) and jump times are fixed. The conditional posterior distribution of \( \theta(\tau_p) \) given \( D, W \) and all other components in \( \Theta \) is

\[
\pi(\theta(\tau_p)|\Theta/\{\theta(\tau_p)\}, D, W) \propto \exp \left\{ -\frac{(\theta(\tau_p) - \mu_p)^2}{2\sigma^2_p} \right\} \times \exp \left\{ -\sum_{i=1}^{n} \sum_{j=1}^{m_i} \sum_{k=1}^{K} \mathbb{1}(\tau_k \in (\tau_{p-1}, \tau_p)) \Delta_k \lambda_k \omega_i \exp(x^T_{i,j} \beta_k) Y_{i,j,k} \right\},
\]

where \( \theta(\tau_p) \) is either \( \log \lambda(\tau_p) \) or one \( \beta_{\tau_p} \). The steps of computing \( \mu_p \) and \( \sigma^2_p \) are shown as follows.

\[
\mu_1 = \sigma^2_1 \left\{ \sum_{i=1}^{n} \sum_{j=1}^{m_i} \sum_{k=1}^{K} \mathbb{1}(s_k \in (0, \tau_1]) x_{i,j} dN_{i,j,k} \right\} + \frac{a_0 \theta(\tau_2)}{1 + a_0},
\]

\[
\sigma^2_1 = a_0 \nu/(1 + a_0);
\]

for \( p = 2, \ldots, P - 1, \)

\[
\mu_p = \sigma^2_p \left\{ \sum_{i=1}^{n} \sum_{j=1}^{m_i} \sum_{k=1}^{K} \mathbb{1}(s_k \in (\tau_{p-1}, \tau_p]) x_{i,j} dN_{i,j,k} \right\} + \frac{\theta(\tau_{p-1})}{2} + \frac{\theta(\tau_{p+1})}{2},
\]

\[
\sigma^2_p = \nu/2;
\]
for \( p = P \),

\[
\mu_p = \sigma_p^2 \left\{ \sum_{i=1}^{n} \sum_{j=1}^{m_i} \sum_{k=1}^{K} 1(s_k \in (\tau_{p-1}, \tau_p]) x_{i,j} dN_{i,j,k} \right\} + \theta(\tau_{p-1}),
\]

\[
\sigma_p^2 = \nu.
\]

Since it can be shown that (2.1) is log-concave, adaptive rejection algorithm (Gilks and Wild, 1992) is applied to sample \( \theta(\tau_p) \).

(2) Birth move. A new jump time \( \tau' \) is "born" in this move. This new \( \tau' \) is randomly selected from non-jump time grids. Let \( \{\tau'_p, p = 1, 2, ..., P + 1\} \) and \( \{\tau_p, p = 1, 2, ..., P\} \) denote new and current jump times respectively. Assume \( \tau' \in (\tau_{p-1}, \tau_p) \), then \( \tau' \) and \( \tau'_{p+1} \) need to be sampled by

\[
\theta(\tau'_p) = \pi_1 \theta(\tau_{p-1}) + \pi_2 \{\theta(\tau_p) + u\},
\]

\[
\theta(\tau'_{p+1}) = \pi_1 \{\theta(\tau_p) - u\} + \pi_2 \theta(\tau_{p+1}),
\]

where weights are defined as

\[
\pi_1 = (\tau'_p - \tau'_{p-1})/(\tau'_{p+1} - \tau'_{p-1}),
\]

\[
\pi_2 = (\tau'_{p+1} - \tau'_p)/(\tau'_{p+1} - \tau'_{p-1}),
\]

where \( u \) is generated from a uniform distribution \( U(-\epsilon_0, \epsilon_0) \) and \( \epsilon_0 \) is set to 1 in this study. Variable \( u \) here is an auxiliary variable to the old model, which helps balance out the one dimension increase of the proposed new model. When \( \tau' \) is near the boundaries, set \( \theta(\tau_0 = \tau_1) \) and \( \theta(\tau_{P+1}) = \theta(\tau_P) \). Let \( \theta = \{\theta(\tau_1), \theta(\tau_2), ..., \theta(\tau_P)\} \) and
\[ \theta' = \{ \theta'(\tau_1), \theta'(\tau_2), ..., \theta'(\tau_{P+1}) \}. \]

The acceptance ratio can be computed with posterior distribution \( \pi(\theta'|\Theta/\{\theta\}, \omega, \nu, D) \), uniform density function \( \pi_u \) and Jacobian of the transformation,

\[
R_{\text{birth}} = \frac{\pi(\theta'|\Theta/\{\theta\}, \omega, \nu, D) \partial \theta'}{\pi(\theta|\Theta/\{\theta\}, \omega, \nu, D) \pi(u) \partial(\theta, u)}.
\]

The acceptance probability is defined as \( \min\{1, R_{\text{birth}}\} \).

(3) Death move. One of current jump time \( \tau_p \) is removed, where the index \( p \) is uniformly selected from current jump point set \( \{1, 2, ..., P-1\} \). Then this can be treated as an inverse step of birth move. By using the same transformation function of birth move, the expression of \( \theta'(\tau_p') \) can be computed as follows

\[
\theta'(\tau_p') = \frac{1}{2} \left\{ -\frac{\pi_1}{\pi_2} \theta(\tau_{p-1}) + \frac{1}{\pi_2} \theta(\tau_p) + \frac{1}{\pi_1} \theta(\tau_{p+1}) - \frac{\pi_2}{\pi_1} \theta(\tau_{p+2}) \right\},
\]

where weights are defined as

\[
\pi_1 = (\tau_p - \tau_{p-1})/(\tau_{p+1} - \tau_{p-1}),
\]
\[
\pi_2 = (\tau_{p+1} - \tau_p)/(\tau_{p+1} - \tau_{p-1}),
\]

and acceptance probability is \( \min\{1, R_{\text{birth}}^{-1}\} \).

The hyper-parameter \( \nu \) has a conjugate inverse-gamma prior and the posterior distribution is also an inverse-gamma as

\[
\nu|\Theta, D \sim IG \left[ \alpha_0 + \frac{P}{2}, \eta_0 + \frac{\theta(\tau_1)^2}{2a_0} + \sum_{p \geq 2} \frac{\{\theta(\tau_p) - \theta(\tau_{p-1})\}^2}{2} \right].
\]
The conditional posterior of $\eta$ given $W$ is
\[
\eta|W \propto \left( \prod_{i=1}^{n} \omega_i \right)^{\eta-1} \eta^{n} \exp\left(-\eta \sum_{i=1}^{n} \omega_i \right) \frac{\Gamma(\eta)}{\pi^{\eta}} \pi_{\eta}(\eta),
\]
i.e., the conditional posterior distribution of $\eta$ depends on the data only through $W$. As mentioned before, gamma distribution $G(0.001, 0.001)$ is used as the prior for $\eta$ in the following analysis.

The frailty parameter $\omega_i$ is sampled as follows
\[
\omega_i|\Theta, D, \eta \sim G\left( \eta + \sum_{j=1}^{m_i} \sum_{k=1}^{K} dN_{i,j,k}, \eta + \sum_{j=1}^{m_i} \sum_{k=1}^{K} \Delta_k \lambda_k \exp(x_{i,j,\beta_k}^T Y_{i,j,k}) \right).
\]

### 2.4 Simulation

In this section, we present simulation study results to demonstrate the properties of the methods introduced in previous sections. Both constant and time-varying coefficients are considered in the simulation study, where the constant coefficients are set to 1 and time-varying coefficient function is $\beta(t) = 0.5 + \sin(t\pi/9)$. The time interval of interest is set to be $(0, 9)$. In addition, we consider two types of covariates (binary or continuous). Then we simulate survival time $t$ under six models with various combinations of coefficients and covariates. The details are listed in Table 2.1. Assume that the baseline hazard function is $\lambda_0(t) = 0.1\sqrt{t}$ and $u = S(t)$ follows a uniform distribution $U(0, 1)$. The shared frailty $\omega_i$ is sampled from gamma distribution $G(1, 1)$. When all the coefficients are constant, survival time are simulated from inverse uniform function $(-15 \log(u)/(\omega_i \exp(x_{i,j,\beta_k}^T)))^{1/2}$. If at least
Table 2.1: Model specifications in the simulation study

<table>
<thead>
<tr>
<th></th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
<th>Model 4</th>
<th>Model 5</th>
<th>Model 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta_1$</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0.5 + sin($t\pi/9$)</td>
<td>0.5 + sin($t\pi/9$)</td>
<td>0.5 + sin($t\pi/9$)</td>
</tr>
<tr>
<td>$x_1$</td>
<td>$\mathcal{B}(0.5)^*$</td>
<td>$\mathcal{N}(0, 1)^{**}$</td>
<td>$\mathcal{B}(0.5)$</td>
<td>$\mathcal{B}(0.5)$</td>
<td>$\mathcal{N}(0, 1)$</td>
<td>$\mathcal{B}(0.5)$</td>
</tr>
<tr>
<td>$\beta_2$</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$x_2$</td>
<td>$\mathcal{N}(0, 1)$</td>
<td></td>
<td></td>
<td></td>
<td>$\mathcal{N}(0, 1)$</td>
<td></td>
</tr>
</tbody>
</table>

* $\mathcal{B}(0.5)$ denotes a Bernoulli distribution with the success probability as 0.5.
** $\mathcal{N}(0, 1)$ represents a standard normal distribution.

one coefficient is time-varying, we have

$$u = S(t) = S_0(t)^{\omega_i \exp(x_{ij}^{\top} \beta(t))},$$

where $S_0(t) = \exp(-\int_0^t \lambda_0(z)dz)$ and $t$ is calculated with numerical method after taking natural logarithm on both side of equation. To generate the censoring interval, the log-normal density function $\mathcal{LN}(x; 0, 0.4)$ is used to simulate gap times. If the exact event occur between two consecutive visits, the time interval of the visits is taken as the censoring interval. If the exact event time does not occur before maximum follow up time, the subject would be considered as right-censored with the last visit time as the censoring point.

Each simulated dataset contains three clusters and 100 subjects in each cluster. Six different models are considered with details listed in Table 2.1. Models 1, 2, 4, 5 contain one covariate and Models 3, 6 have two covariates. For each model, 300 datasets were simulated. We ran 12,000 MCMC iterations with the first 2000 iterations discarded as burn-in period.
We checked the convergence based on trace plots, autocorrelation plots and Geweke statistics. LPML (log pseudo marginal likelihood) (Ibrahim et al., 2005) is used for model comparison. LPML is the summation over all CPO\((i)\) (conditional predictive ordinate), where CPO\((i)\) is the posterior predictive probability of the \(i\)th observation given all other observed data under the assumption that the current model is true. Since the dimension of the parameters changes from one iteration to another, LPML is the best choice and it takes the form based on all observations:

\[
\text{LPML} = \sum_{i=1}^{n} \sum_{j=1}^{m_i} \log \text{CPO}_{i,j}. \tag{2.2}
\]

A larger value of LPML implies a better Model. The conditional predictive ordinate (CPO) in (2.2) is defined as the cross-validation posterior density conditional on the rest of the observations (Dey et al., 1997). Sinha et al. (1999) showed that a CPO\(_{i,j}\) can be computed as

\[
\text{CPO}_{i,j} = \left\{ \mathbb{E} \left[ \Pr(T_{i,j} \in (L_{i,j}, R_{i,j})|\Theta, \omega_i, x_{i,j})^{-1} \right] \right\}^{-1}, \tag{2.3}
\]

where \(\Pr(T_{i,j} \in (L_{i,j}, R_{i,j})|\Theta, \omega_i, x_{i,j})\) is the individual likelihood evaluated after taking expectation to the joint posterior of \(\Theta\) and \(\omega_i\). In practice, (2.3) is equivalent to Monte Carlo estimate at MCMC samples:

\[
\text{CPO}_{i,j} = \left[ \frac{1}{R} \sum_{r=1}^{R} \left( \frac{1}{\Pr(T_{i,j} \in (L_{i,j}, R_{i,j})|\Theta, \omega_i, x_{i,j})} \right) \right]^{-1},
\]

where index \(r\) is the \(r\)th iteration and \(R\) is the total number of iterations. Models with higher LPML are preferred to models with lower LPML. When all frailties are equal to one, our model reduces to Wang’s model (Wang et al., 2013).
Figures 2.1 and 2.2 include estimated coefficients from all six models. Figure 2.1 has the results from Models 1, 2, 4, 5 with one covariate, whereas Figure 2.2 displays the results from Models 3 and 6 with two covariates. The upper panel of figure contains models with frailty and the lower panel is for models without frailty.

Panels in the first row consist of plots of results from the proposed model with frailty and those in the second row present Wang’s model without frailty. Each panel includes median of the posterior mean and median of 95% credible intervals of regression coefficients from our model with frailty or Wang’s model. By comparing each pair of plots, it’s obvious that the
Table 2.2: Estimates of frailties and model comparison results

<table>
<thead>
<tr>
<th></th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
<th>Model 4</th>
<th>Model 5</th>
<th>Model 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \omega_1 ) (0.419*)</td>
<td>0.281 (0.019, 0.623)</td>
<td>0.177 (0.001, 0.506)</td>
<td>0.266 (0.012, 0.605)</td>
<td>0.303 (0.028, 0.644)</td>
<td>0.185 (0.001, 0.507)</td>
<td>0.272 (0.016, 0.591)</td>
</tr>
<tr>
<td>( \omega_2 ) (0.936*)</td>
<td>0.648 (0.047, 1.363)</td>
<td>0.387 (0.002, 1.099)</td>
<td>0.588 (0.029, 1.322)</td>
<td>0.685 (0.066, 1.431)</td>
<td>0.414 (0.003, 0.414)</td>
<td>0.595 (0.036, 1.308)</td>
</tr>
<tr>
<td>( \omega_3 ) (2.171*)</td>
<td>1.436 (0.106, 3.046)</td>
<td>0.886 (0.005, 2.532)</td>
<td>1.345 (0.070, 2.897)</td>
<td>1.519 (0.147, 3.200)</td>
<td>0.915 (0.007, 2.392)</td>
<td>1.399 (0.087, 2.984)</td>
</tr>
<tr>
<td>( \eta ) (1*)</td>
<td>1.729 (0.031, 4.951)</td>
<td>1.217 (0.011, 3.905)</td>
<td>1.574 (0.026, 4.694)</td>
<td>1.834 (0.035, 5.098)</td>
<td>1.253 (0.012, 4.163)</td>
<td>1.622 (0.028, 4.735)</td>
</tr>
<tr>
<td>% Diff &gt; 0</td>
<td>95.7%</td>
<td>87%</td>
<td>93.3%</td>
<td>96.3%</td>
<td>87%</td>
<td>91.3%</td>
</tr>
</tbody>
</table>

* True value.
Figure 2.2: Estimate of coefficients for models with two covariates and three clusters. The solid lines in the middle are the true coefficients. The long dashed lines are medians of posterior means. The top and bottom short dashed lines are medians of 95% credible intervals.

The proposed model outperforms Wang’s method by capturing the true value for both constant and time-varying coefficients. Besides, the proposed model produces credible intervals for the regression coefficients that are similar or narrower in width compared to those from Wang’s model.

Table 2.2 shows the estimates of frailties and model comparison results. Each cell contains median of posterior mean and 95% credible interval which covers the true value. In addition, it contains mean and (0.025, 0.975) quantile of “LPML Difference” and percentage of “LPML Difference” greater than zero, where “LPML Difference” is calculated as LPML of
the proposed model minus LPML of Wang’s model. Table 2.2 shows that the mean of difference is at least 30.810 and at least 87% of “LPML Difference” are greater than zero, which further indicates that the proposed model performs consistently better and consideration of within cluster correlation is necessary in some cases.

In order to see the performance of the proposed method under various scenarios, we also tried different cluster numbers (3 & 10), equal or unequal sample size assignment on different clusters as well as different magnitudes of frailties. Although not shown here, the results all indicate that the proposed model performances consistently better than Wang et al. (2013).

2.5 Children’s dental health data

In this section, we apply the methodology described in previous sections to the Signal Tandmobiel project that was conducted in Flanders (Belgium) to estimate the oral health condition of Flemish primary schoolchildren. The children were divided into 15 strata, a combination of 3 educational systems (public, municipal or private) and 5 provinces. Over 6,000 children were recruited which represented approximately 7% of the total target population in Flanders (Vanobbergen et al., 2000). A total of 4468 of them were randomized and examined annually by 16 trained dentists using the standardized and widely accepted criteria recommended by the World Health Organization. We focus on the time to emergence of permanent tooth 24 (central incisor) here.

For the analysis, we considered the covariate \( dmf \) as a dichotomized variable which denotes
the status of the primary predecessor of this tooth (0=sound, 1=decayed, missing or filled).

A random effect of province-by-gender was considered in the regression model described in previous sections. Frailty is assumed to follow a gamma distribution with equal shape and scale which has a gamma hyper-prior $G(0.001, 0.001)$. 12,000 Gibbs samples with a burn-in period of 2,000 were generated. Although not shown, the convergence of MCMC chains was checked by trace plot, autocorrelation plot and Geweke’s statistics.

Figure 2.3 presents the analysis results by applying the proposed method. As tooth 24 does not emerge before age 5, the time scale in Figure 2.3 is the time in years since age 5. The results include the point estimates and 95% credible intervals. The positive estimate indicates that the children with decayed primary predecessor have higher risks than those
Table 2.3: Estimate of frailties with 95% credible intervals in the children’s dental health data

<table>
<thead>
<tr>
<th>Province</th>
<th>Girls</th>
<th>Boys</th>
</tr>
</thead>
<tbody>
<tr>
<td>Antwerpen</td>
<td>1.18 (1.02, 1.35)</td>
<td>0.84 (0.72, 0.96)</td>
</tr>
<tr>
<td>Limburg</td>
<td>1.08 (0.92, 1.25)</td>
<td>0.81 (0.68, 0.93)</td>
</tr>
<tr>
<td>Oost Vlaanderen</td>
<td>1.19 (1.02, 1.36)</td>
<td>0.91 (0.78, 1.05)</td>
</tr>
<tr>
<td>Vlaams Brabant</td>
<td>1.16 (0.98, 1.35)</td>
<td>0.93 (0.79, 1.08)</td>
</tr>
<tr>
<td>West Vlaanderen</td>
<td>1.06 (0.89, 1.23)</td>
<td>0.87 (0.73, 1.00)</td>
</tr>
</tbody>
</table>

without, which is consistent with the results based on ICM algorithm (Gómez et al., 2009). However, there is an obvious trend of the coefficient estimate, and the effect becomes weaker over time. After 6 years, the effect vanished. All those findings were not captured in (Gómez et al., 2009). Table 2.3 shows the frailty estimate. In general, girls have higher risk than boys. Specifically, in provinces Antwerpen and Limburg, the two provinces in north and adjacent to the Netherlands, the difference between girls and boys is more significant than that in provinces Vlaams Brabant and West Vlaanderen, the two provinces in the middle of Belgium. Moreover, boys in Vlaams Brabant, where the capital city Brussels is located, have the highest risk compared to boys in other provinces. But the high risk was not shown in girls (Figure 2.4). There interesting findings were never reported before to the author’s best knowledge and may have significant impacts in health policy and children’s health care. We also fitted Wang’s model (Wang et al., 2013) to children’s dental health data with covariate
Figure 2.4: Estimate of frailties with 95% credible intervals in the children’s dental health data.

$dmf$. The LPML value for the proposed model and Wang’s model are -5480.8 and -5511, respectively. In summary, the proposed model is preferred to the model in Gómez et al. (2009) and Wang et al. (2013).

### 2.6 Discussion

In the present study, we proposed a Bayesian approach for clustered interval-censored Cox regression model with time-varying covariate effects. In order to capture the temporal nature of covariate effects by using time-varying coefficients more precisely, we have shown that it’s crucial to consider dependence structure for clustered outcomes. In our study, we used a gamma frailty model for unobserved heterogeneity. A non-informative hyper prior of gamma
parameter is used for heterogeneity among clusters. Other finite mean frailty models, such as log-normal frailty model, can also be applied for the proposed methodology developed in this study.

A key step in this procedure is to determine the number of jumps, which will affect the estimated smoothness of the estimated curves as well as effectiveness assessment. Previous studies usually specify large number of jumps and then select jumps after model fitting. This may not be appropriate when prior information of jumps is not available. In our approach, we used reversible jump MCMC to automatically select jumps during iterations. Regression coefficients and baseline hazard are piecewise constant and can be estimated given the number of pieces and jump locations.

Although our discussion has been focused on the Cox model with independent frailty, much wider extensions and applications are feasible with the proposed Bayesian approach. In some cases, independent frailty assumption may not hold and the method can be extended to more general frailty distributions, such as “correlated frailty models” (Hens et al., 2009), where the frailty factors are correlated.
Chapter 3

Spatial frailty model with time-varying coefficient

In this chapter, we propose a Bayesian approach to dynamic Cox regression model allowing for spatial correlation with interval-censored time-to-event data. With Bayesian approach, the coefficient curves are piecewise constant and the number of pieces and jump points are estimated from data. A conditional autoregressive distribution is employed to model the spatial dependency. The posterior summaries are obtained via an efficient reversible jump Markov chain Monte Carlo algorithm. The properties of our method are illustrated by conducting simulation studies. In addition, we apply our method to a smoking cessation data in southeastern Minnesota.
3.1 Introduction

Chapter 2 proposed a Bayesian approach to estimate time-varying coefficient for clustered interval-censored data with independent shared frailty model. However, independent frailty assumption may not hold in practice. Many health-science data are spatially correlated due to the development of geographical information systems. A spatial frailty model with conditional autoregressive (CAR) distribution prior is then more appropriate. In previous studies, Hodges et al. (2003) derived the correct power for the precision parameter in the CAR model. Banerjee et al. (2003) modeled spatially correlated frailties under the Weibull model. Banerjee and Carlin (2004) investigated spatial frailties in parametric cure rate models using multivariate conditional autoregressive (MCAR) prior. Pan et al. (2014) studied spatial frailties with CAR prior under Cox model. However, most work for building regression model with spatial data has been conducted under the constant coefficient assumption, which may not be appropriate in real case studies.

The motivating problem is a geographically referenced smoking cessation data set consisting of 223 subjects living in 51 zip code areas in the southeastern corner of Minnesota. The objective of this study is to estimate the effect of the prognostic factors on time to relapse to smoking, after adjusting for the spatial dependency within and among clusters. Each zip code area forms a spatial cluster. The event of interest is relapse to smoking, which was interval-censored due to the fact that subjects were monitored at annual visits for 5 years.
In this chapter, we further propose a Bayesian approach to estimate time-varying coefficient for spatially correlated interval-censored data under Cox model. In spatial analysis, clusters can be formed based on geographical areas, and this feature is imported to this study as spatially correlated frailties. We also apply reversible jump MCMC to generate posterior samples of baseline and coefficients, and the CAR distribution is used as the prior for spatial frailties (Banerjee et al., 2003).

The remainder of this chapter is organized as follows. Section 3.2 introduces data structure, model specification and associated likelihood function. Besides the proposed model, we also provide two other models: one is for constant coefficient and spatially correlated frailties, and the other is for time-varying coefficient and independent frailties. Section 3.3 describes prior specification and Section 3.4 includes posterior inference details. Section 3.5 presents simulation studies of proposed method with model comparison results. A smoking cessation data is analyzed in Section 3.6. Conclusions and discussions are enclosed in Section 3.7.

### 3.2 Model specification

Let \( T_{i,j} \) denote the survival time for the \( j^{th} \) subject in the \( i^{th} \) cluster, where \( i = 1, 2, ..., n, \ j = 1, 2, ..., m_i \). Therefore we have a total of \( N = \sum_{i=1}^{n} m_i \) subjects. Consider a Cox model with constant or time-varying regression coefficients conditional on a \( Q \)-dimensional vector of covariates, \( x_{i,j} \), and the unobserved frailty random variable \( \omega_i \) for the \( i^{th} \) cluster. The
hazard function can be written as

\[ \lambda(t|\omega_i, x_{i,j}) = \lambda_0(t) \exp(x_{i,j}^T \beta(t) + \omega_i), \]

where \( \lambda_0(.) \) is an unknown baseline hazard function common to all subjects, \( x_{i,j} \) is the \( Q \times 1 \) covariate vector for the \( j^{th} \) subject in the \( i^{th} \) cluster and \( \beta(t) \) is the \( Q \)-dimensional regression coefficient function of main interest. The frailty \( \omega_i \) can be either independent or correlated.

In this study, we consider three models specified as follows:

- **Model 1**: \( \beta(t) \) is fixed, \( \omega_i \)'s are spatially correlated.
- **Model 2**: \( \beta(t) \) is time-varying, \( \omega_i \)'s are independent.
- **Model 3**: \( \beta(t) \) is time-varying, \( \omega_i \)'s are spatially correlated.

For interval-censored data, unobserved event time \( T_{i,j} \) is located in observed censoring interval \( (L_{i,j}, R_{i,j}] \). The contribution of subject \( j \) in \( i^{th} \) cluster to the observed data likelihood is

\[
\Pr(T_{i,j} \in (L_{i,j}, R_{i,j}] | \omega_i, x_{i,j}) = \Pr(T_{i,j} > L_{i,j}|\omega_i, x_{i,j}) - \Pr(T_{i,j} > R_{i,j}|\omega_i, x_{i,j}).
\]

When the cumulative hazard function is a totally unspecified nondecreasing function, it's always challenging to estimate regression parameters directly since the partial likelihood does not exist for interval-censored data (Sun, 2007). In this study, we borrowed the idea of Bayesian discretized semiparametric model, which was proposed by Sinha et al. (1999) and also studied by Kim et al. (2007) and Wang et al. (2013). The basic idea is to generate augmented event time in its observed interval and compute the estimates of parameters.
piecewisely. Once the unobserved event time \( T_{i,j} \)'s are given, the interval-censored data reduce to right-censored data. First we assume that \( \lambda_0(t) \) and \( \beta(t) \) are left continuous step functions, where both the number of jumps and the locations of the jumps are random and to be estimated. Let \( k = 1, 2, ..., K \) denote all the ordered grids and \( 0 = \tau_0 < \tau_1 < \tau_2 < ... < \tau_K < \infty \) be the corresponding time points. Here we assume the time points contain all potential jump points. The length of each time interval may be taken to be sufficiently small so that hazard and coefficient function can be appropriately estimated. Denote \( \lambda_k = \lambda_0(\tau_k) \), and \( \beta_k = \beta(\tau_k) \) as the baseline hazards function and coefficient function evaluated at each grid point. Let \( dN_{i,j,k} \) indicate whether or not the event time \( T_{i,j} \) falls within the \( k^{th} \) interval, i.e., \( dN_{i,j,k} = 1(T_{i,j} \in (\tau_{k-1}, \tau_k]) \). Let \( Y_{i,j,k} \) be the at-risk variable. \( Y_{i,j,l} = 1 \) for \( l < k \), \( Y_{i,j,l} = 0 \) for \( l > k \), and \( Y_{i,j,k} = (T_{i,j} - s_{k-1})/\Delta_k \) for \( l = k \), where \( \Delta_k = \tau_k - \tau_{k-1} \) is the width of the \( k^{th} \) interval. Thus, the augmented likelihood function for \( j^{th} \) subject of \( i^{th} \) cluster is

\[
\ell_{i,j}(\Theta | \{dN_{i,j,k}, Y_{i,j,k}\}^{K}_{k=1}, \omega_i, x_{i,j}) = \prod_{k=1}^{K} \{ \lambda_k \exp(x_{i,j}^T \beta_k + \omega_i) \}^{dN_{i,j,k}} \exp\{-\Delta_k \lambda_k \exp(x_{i,j}^T \beta_k + \omega_i) Y_{i,j,k}\}, \tag{3.1}
\]

where \( i = 1, 2, ..., n, \ j = 1, 2, ..., J_i \) and \( \Theta = \{ \lambda_k, \beta_k; k = 1, 2, ..., K \} \). The conditional autoregressive (CAR) distribution, which is originally developed by Besag (1974), is imposed for spatially correlated \( \omega_i \)'s in (3.1). The general form of the CAR model is

\[
\Pr(\omega) \propto \exp\left\{-\frac{1}{2} \omega^T V \omega\right\}, \tag{3.2}
\]

where \( V \) is an \( n \times n \) positive definite symmetric matrix. If we specify that \( V \cdot 1 = 0 \), then we get the intrinsic conditional autoregressive (ICAR) model (Besag and Kooperberg, 1995).
Note that $V$ is positive semi-definite and the variance matrix $V^{-1}$ no longer exists. An algebraic decomposition of the power term in (3.2) is given by

$$\Pr(\omega) \sim \exp\left\{\frac{1}{2} \sum_{i<j} V_{ij} (\omega_i - \omega_j)^2 \right\}. \quad (3.3)$$

Since $\omega_i$'s in (3.3) are actually non-identifiable, in Bayesian implementation of this study, an identifying sum-to-zero constraint is imposed by centering the $\omega_i$'s around zero after each iteration (Carlin and Louis, 1997). In practice, it usually furthermore specifies that $V = \pi_\omega W$, where $\pi_\omega$ is a precision parameter, $W_{ii} = m_i$, $W_{ij} = -1_{(i\sim j)}$, $m_i$ is the number of neighbors for area $i$, $i \sim j$ denotes that areas $i$ and $j$ are neighbors. Then (3.3) becomes

$$\Pr(\omega) \sim \exp\left\{-\frac{\pi_\omega}{2} \sum_{i<j} (\omega_i - \omega_j)^2 1_{(i\sim j)} \right\}, \quad (3.4)$$

and (3.4) is also equivalent to

$$\omega_i | \omega_{-i} \sim \mathcal{N}(\bar{\omega}_{ii}, 1/(m_i \pi_\omega)), \quad (3.5)$$

where $\omega_{-i}$ is the set of all spatial frailties except the one for area $i$, $\bar{\omega}_{ii}$ is area $i$'s neighbor mean of frailties, and $\mathcal{N}(\bar{\omega}_{ii}, 1/(m_i \pi_\omega))$ denotes a normal distribution with mean $\bar{\omega}_{ii}$ and variance $1/(m_i \pi_\omega)$. The conditional distribution in (3.5) is to be used as a prior for $\omega_i$ in MCMC. Given $\pi_\omega$, the prior for $\omega$ is as follows

$$\Pr(\omega) \propto \pi_\omega^{(n-b)/2} \exp\left\{-\frac{\pi_\omega}{2} \sum_{i<j} (\omega_i - \omega_j)^2 1_{(i\sim j)} \right\},$$

where $b$ is the number of disconnected groups of areas, and $b = 0$ if all the areas have at least one neighbor.
3.3 Prior Distributions

In this section, we present specifications of priors for the three aforementioned models in Section 3.2. Since constant and time-varying coefficients, independent and spatially frailties have different prior structures, we list the details of priors for each model in Sections 3.3.1, 3.3.2 and 3.3.3 separately.

3.3.1 Fixed $\beta$ with spatially correlated $\omega_i’s$

We first introduce the prior specifications for Model 1 with constant coefficient and spatially correlated frailties as follows

$$\lambda_k \sim \mathcal{G}(c_k, d_k), \quad k = 1, ..., K,$$

$$\beta_q \sim \mathcal{N}(\mu_0, \sigma^2_0), \quad q = 1, ..., Q,$$

$$\omega_i | \omega_{-i} \sim \mathcal{N}(\bar{\omega}_{ii}, 1/(m_i \pi_{\omega})), \quad i = 1, 2, ..., n,$$

$$\pi_{\omega} \sim \mathcal{G}(0.01, 0.01),$$

where $\mathcal{G}(0.01, 0.01)$ denotes a gamma distribution with mean 1 and variance 100. This kind of prior was also used for spatially correlated data in Pan et al. (2014). However, we provide a more non-informative prior for precision parameter $\pi_{\omega}$. The priors for baseline and coefficients are the same as Sinha et al. (1999).
3.3.2 Time-varying $\beta(t)$ with independent $\omega_i$,s

The correlated interval-censored data with time-varying coefficient was also studied in Zhang et al. (2016). They assigned a conjugate gamma prior to the exponential of $\omega_i$. However, in this study, $\omega_i$ is treated as random intercept with a vague normal prior for model comparison (Model 2 and Model 3) purpose. Then, due to the fact that the logarithm of the baseline function $\log(\lambda_0(t))$ can be considered as a intercept, we specify the same type of priors for both the logarithm of the baseline function and the covariate coefficient functions. In the following, let $\theta(t)$ denote either log baseline $\log(\lambda_0(t))$ or one $\beta(t)$ for prior specifications.

Let $P$ denote the number of jumps in $\theta(t)$, which follows a discrete uniform distribution range from 1 to $K$. With a fixed $P$, the jump times $\{\tau_p\}$, $p = 1, 2, ..., P$, except last one are randomly selected from all time grids. Given $P$ and $\{\tau_p\}$, priors are specified as follows

$$
\theta(\tau_1) \sim \mathcal{N}(0, a_0\nu),
$$

$$
\theta(\tau_p)|\theta(\tau_{p-1}) \sim \mathcal{N}(\theta(\tau_{p-1}), \nu), \quad p = 2, 3, ..., P,
$$

$$
\nu \sim \mathcal{IG}(\alpha_0, \eta_0),
$$

$$
\omega_i \overset{i.i.d.}{\sim} \mathcal{N}(0, \sigma^2), \quad i = 1, ..., n, \quad (3.6)
$$

$$
\sigma^2 \sim \mathcal{IG}(0.01, 0.01),
$$

where $a_0 > 0$ is a hyper parameter which can be chosen as a large number to reflect higher uncertainty in the prior input, and $\mathcal{IG}(\alpha_0, \eta_0)$ denotes an inverse gamma distribution with shape parameter $\alpha_0$ and rate parameter $\eta_0$ with mean $\eta_0/(\alpha_0 - 1)$ for $\alpha_0 > 1$. The proposed
prior for $\theta(t)$ for Model 2 determines the posterior structure of the baseline and coefficient parameters, and controls the smoothness of time-dependent curve. Compared to Sinha’s prior specification for M1 (Sinha et al., 1999), the posterior inferences in this study are more robust to the selections of hyper parameters. Since the number of jumps $P$ and jump times $\tau_p$ are different between models in iterations, reversible jump MCMC (Green, 1995) is implemented to handle this dimension change, and the posterior computation steps are discussed in Section 3.4.

### 3.3.3 Time-varying $\beta(t)$ with spatially correlated $\omega_i$'s

Model 3 proposes an efficient way to simultaneously estimate time-varying coefficient and spatially correlated frailties. The specification of priors incorporates distribution functions for $\omega_i$ in Model 1 and $\theta(t)$ in Model 2:

\[
\begin{align*}
\theta(\tau_1) & \sim N(0, a_0 \nu), \\
\theta(\tau_p) | \theta(\tau_{p-1}) & \sim N(\theta(\tau_{p-1}), \nu), \quad p = 2, 3, ..., P, \\
\nu & \sim IG(\alpha_0, \eta_0), \\
\omega_i | \omega_{-i} & \sim N(\bar{\omega}_{ii}, 1/(m_i \pi_\omega)), \quad i = 1, 2, ..., n, \\
\pi_\omega & \sim G(0.01, 0.01).
\end{align*}
\]

The smoothness of $\theta(t)$ is also controlled by Markov type process prior and dimension of parameters between models is determined by reversible jump MCMC. Prior of frailties is based on ICAR model for model comparison purpose (Model 1 and Model 3). Note that
Model 3 can also easily be extended with other coefficient’s prior with various degrees of smoothness or different frailties’ prior to account for spatial correlation.

3.4 Posterior Computation

The posterior samples can be obtained through MCMC sampling including a mixture of Gibbs sampling, Metropolis-Hastings, and the adaptive rejection algorithm. The computation is based on \(j^{th}\) subject of the \(i^{th}\) group observed in the \(k^{th}\) time interval, where \(i = 1, 2, ..., n, j = 1, 2, ..., m_i, k = 1, 2, ..., K\) and \(K\) is the total number of time grids. The parameters need to be sampled include three parts: (1) the augmented event times \(T_{ij}\)’s, whose information is equivalent to those from our specially defined event indicator \(dN_{i,j,k}\)’s and at-risk variable \(Y_{i,j,k}\)’s; (2) the baseline \(\lambda_k\), coefficients of covariates \(\beta_k\); (3) frailty terms \(\omega_i\)’s. Let \(D = \{dN_{i,j,k}, Y_{i,j,k}\}\), \(\Theta = \{\lambda_k, \beta_k\}\) and \(W = \{\omega_i\}\). The MCMC algorithm draws \(D, \Theta, W\) iteratively.

3.4.1 Sample \(D\)

Since baseline hazards and regression coefficients are assumed to be piecewise constant, the sampling of augmented event time \(T_{ij}\) given \(\Theta\) and \(W\) can be separated into two steps.

(1) Locate grid interval for each event time. For finite interval-censored subject, event indicator \(dN_{i,j,1}, dN_{i,j,2}, ..., dN_{i,j,k}\) follows a multinomial distribution with size 1 and proba-
bility vector \((e_{i,j,1}, e_{i,j,2}, ..., e_{i,j,k})\), where for \(k = 1, 2, ..., K\),
\[
e_{i,j,k} = \frac{p_{i,j,k} \mathbb{I}(s_k \in (L_{i,j}, R_{i,j}])}{\sum_{l \in (L_{i,j}, R_{i,j})} p_{i,j,l}},
\]
\[
p_{i,j,k} = \begin{cases} 
\exp \left\{ -\sum_{l=1}^{k-1} \Delta_l \lambda_l \exp(x_{i,j}^T \beta_k + \omega_i) \right\} - \\
\exp \left\{ -\sum_{l=1}^{k} \Delta_l \lambda_l \omega_i \exp(x_{i,j}^T \beta_k + \omega_i) \right\} & \text{if } k > 1 \\
1 - \exp \left\{ -\Delta_1 \lambda_1 \exp(x_{i,j}^T \beta_1 + \omega_i) \right\} & \text{if } k = 1.
\end{cases}
\]
Thus, if observed interval \((L_{i,j}, R_{i,j}]\) only covers one time grid \(s_k\), then \(dN_{i,j,k} = 1\) and all other event indicators equal 0. Otherwise, if \((L_{i,j}, R_{i,j}]\) covers multiple time grids \(s_k\)'s, \(dN_{i,j,k}\) is sampled from multinomial distribution with probability vector calculated based on these covered time grids \(s_k\)'s. In other words, the time grid \((s_{k-1}, s_k]\) with \(dN_{i,j,k} = 1\) is sampled in this step.

(2) Within selected time grids, the exact time \(T_{i,j}\) follows a doubly truncated exponential distribution on \((s_{k-1}, s_k]\) with a distribution function
\[
F(u) = \frac{1 - \exp\left\{-\lambda_k (u - s_{k-1}) \exp(x_{i,j}^T \beta_k + \omega_i)\right\}}{1 - \exp\left\{-\lambda_k \Delta_k \exp(x_{i,j}^T \beta_k + \omega_i)\right\}}.
\]
Then \(T_{i,j}\) will be sampled by using inverse distribution method and \(Y_{i,j,k}\) can be computed based aforementioned definition.

### 3.4.2 Sample \(\Theta\)

The next step is to sample each component of baseline hazards, regression coefficients given \(D\) and \(W\).
(1) If $\beta(t)$ is fixed as described in Section 3.3.1, we obtain $\lambda_k$ from gamma distribution $G(c_k + \sum_{i=1}^n \sum_{j=1}^{m_i} d_{N_{i,j,k}} + d_k + \sum_{i=1}^n \sum_{j=1}^{m_i} (\Delta_k \exp(x^T \beta + \omega_i))Y_{i,j,k})$ and conditional posterior of $\beta_q$ with the form

$$\Pr(\beta_q|\Theta(\beta_q), D, W, x) \propto \prod_{i=1}^n \prod_{k=1}^K \{\lambda_k \exp(x_{q,i,j} x \beta_q + \omega_i)\} \exp{-\Delta_k \lambda_k \exp(x_{q,i,j} \beta_q + \omega_i)Y_{i,j,k}}$$

$$\times \phi(\beta_q; \beta_0, \sigma_0^2).$$  \hspace{1cm} (3.8)

The complete conditional distributions of $\beta$’s do not correspond to standard statistical distributions. However, it’s easy to show that (3.8) is a log-concave density since all those conditional distributions are log-concave. Thus, we use the adaptive rejection algorithm of Gilks and Wild (1992) to simulate $\beta$’s.

(2) If $\beta(t)$ is time dependent, $\theta(t)$ is used to denote log baseline $\log(\lambda(t))$ and one $\beta(t)$.

Reversible jump MCMC algorithm is applied here because the number of jumps $P$ is random and dimension of posterior distribution could vary from iteration to iteration. The probability of taking a birth, death and update move are set as 0.3, 0.3, and 0.4, respectively. For the update move, both $P$ and jump times are fixed. The conditional posterior distribution of $\theta(\tau_p)$ given $D, W$ and all other components in $\Theta$ is

$$\Pr(\theta(\tau_p)|\Theta/\{\theta(\tau_p)\}, D, W) \propto \exp{\left\{-\frac{(\theta(\tau_p) - \mu_p)^2}{2\sigma_p^2}\right\}}$$

$$\times \exp{\left\{-\sum_{i=1}^n \sum_{j=1}^{m_i} \sum_{k=1}^K \mathbb{1}(s_k \in (\tau_{p-1}, \tau_p]) \Delta_k \lambda_k \exp(x_{i,j}^T \beta_k + \omega_i)Y_{i,j,k}\right\}}.$$

$$\hspace{1cm} (3.9)$$

53
where \( \theta(\tau_p) \) is \( \log \lambda_k \) or some components of \( \beta_k \). The steps of computing \( \mu_p \) and \( \sigma_p^2 \) are shown as follows.

For \( p = 1 \),
\[
\mu_1 = \sigma_1^2 \left\{ \sum_{i=1}^n \sum_{j=1}^{m_i} \sum_{k=1}^K \mathbb{I}(s_k \in (0, \tau_1]) x_{i,j} dN_{i,j,k} \right\} + \frac{a_0 \theta(\tau_2)}{1 + a_0},
\]
\[
\sigma_1^2 = a_0 \nu / (1 + a_0);
\]

for \( p = 2, \ldots, P - 1 \),
\[
\mu_p = \sigma_p^2 \left\{ \sum_{i=1}^n \sum_{j=1}^{m_i} \sum_{k=1}^K \mathbb{I}(s_k \in (\tau_{p-1}, \tau_p) x_{i,j} dN_{i,j,k} \right\} + \frac{\theta(\tau_{p-1})}{2} + \frac{\theta(\tau_{p+1})}{2},
\]
\[
\sigma_p^2 = \nu / 2;
\]

for \( p = P \),
\[
\mu_p = \sigma_p^2 \left\{ \sum_{i=1}^n \sum_{j=1}^{m_i} \sum_{k=1}^K \mathbb{I}(s_k \in (\tau_{P-1}, \tau_P]) x_{i,j} dN_{i,j,k} \right\} + \theta(\tau_{P-1}),
\]
\[
\sigma_p^2 = \nu,
\]

Since it can be shown that (3.9) is log-concave, adaptive rejection algorithm (Gilks and Wild, 1992) can also be applied to sample \( \theta(\tau_p) \). For birth move, a new jump time \( \tau' \) is “born” in this move. This new \( \tau' \) is randomly selected from non-jump time grids. In death move, one of current jump times \( \tau_p \) is removed, where the index \( p \) is uniformly selected from current jump point set \( \{1, 2, \ldots, P - 1\} \). Details of birth move and death move are listed as follows.
Birth move

A new jump time \( \tau' \) is “born” in this move. This new \( \tau' \) is randomly selected from non-jump time grids. Let \( \{ \tau_p', p = 1, 2, ..., P + 1 \} \) and \( \{ \tau_p, p = 1, 2, ..., P \} \) denote new and current jump times respectively. Assume \( \tau' \in (\tau_{p-1}, \tau_p) \), then \( \tau' \) and \( \tau_{p+1}' \) need to be sampled according to

\[
\theta(\tau_p') = \pi_1 \theta(\tau_{p-1}) + \pi_2 \{ \theta(\tau_p) + u \},
\]
\[
\theta(\tau_{p+1}') = \pi_1 \{ \theta(\tau_p) - u \} + \pi_2 \theta(\tau_{p+1}),
\]

where weights are defined as

\[
\pi_1 = (\tau_p' - \tau_{p-1}')/(\tau_{p+1}' - \tau_{p-1}),
\]
\[
\pi_2 = (\tau_{p+1}' - \tau_p')/(\tau_{p+1}' - \tau_{p-1}),
\]

and \( u \) is generated from a uniform distribution \( U(-\epsilon_0, \epsilon_0) \) with \( \epsilon_0 \) set to 1 in this study. Variable \( u \) here is an auxiliary variable to the old model, which helps balance out the one dimension increase of the proposed new model. When \( \tau' \) is near the boundaries, set \( \theta(\tau_0 = \tau_1) \) and \( \theta(\tau_{P+1}) = \theta(\tau_P) \). Let \( \theta = \{ \theta(\tau_1), \theta(\tau_2), ..., \theta(\tau_P) \} \) and \( \theta' = \{ \theta'(\tau_1), \theta'(\tau_2), ..., \theta'(\tau_{P+1}) \} \). The acceptance ratio can be computed with posterior distribution \( \pi(\theta'|\Theta/\{\theta\}, \omega, \nu, D) \), uniform density function \( \pi_u \) and Jacobian of the transformation,

\[
R_{\text{birth}} = \frac{\pi(\theta'|\Theta/\{\theta\}, \omega, \nu, D)}{\pi(\theta|\Theta/\{\theta\}, \omega, \nu, D)\pi(u)} \left| \frac{\partial \theta'}{\partial (\theta, u)} \right|.
\]

The acceptance probability is defined as \( \min\{1, R_{\text{birth}}\} \).
**Death move**

One of current jump time $\tau_p$ is removed, where the index $p$ is uniformly selected from current jump point set $\{1, 2, ..., P - 1\}$. Then this can be treated as an inverse step of birth move. By using the same transformation function of birth move, the expression of $\theta(\tau'_p)$ can be computed as follows

$$\theta(\tau'_p) = \frac{1}{2} \left\{ -\frac{\pi_1}{\pi_2} \theta(\tau_{p-1}) + \frac{1}{\pi_2} \theta(\tau_p) + \frac{1}{\pi_1} \theta(\tau_{p+1}) - \frac{\pi_2}{\pi_1} \theta(\tau_{p+2}) \right\},$$

where weights are defined as

$$\pi_1 = \frac{(\tau_p - \tau_{p-1})}{(\tau_{p+1} - \tau_{p-1})},$$

$$\pi_2 = \frac{(\tau_{p+1} - \tau_p)}{(\tau_{p+1} - \tau_{p-1})},$$

and acceptance probability is $\min\{1, R_{\text{birth}}^{-1}\}$.

### 3.4.3 Sample $W$

As discussed in Section 3.3, $\omega_i$’s could be either independent or spatially correlated with each other. Here, we have different forms for them based on aforementioned prior distributions.

1. If $\omega_i$’s are independent, Metropolis-Hastings algorithm is applied to generate samples based on following full conditional distribution

$$\Pr(\omega_i|\Theta, D, \pi_\omega, \omega_{-i}) \propto \prod_{j=1}^{J_i} \prod_{k=1}^{K} \lambda_k \exp(x_{i,j}^T \beta_k + \omega_i) \exp\{-\Delta_k \lambda_k \exp(x_{i,j}^T \beta_k + \omega_i) \} \Pr(\omega_i),$$
where $\Pr(\omega_i)$ denotes the prior in (3.6). Then sample $\sigma^2$ from $\mathcal{I}G(0.01 + \frac{I}{2}, 0.01 + \sum_{i=1}^{I} \omega_i^2)$.

(2) If $\omega_i$’s are assumed to be spatially correlated, the conditional posterior distribution takes the form

$$
\Pr(\omega_i|\Theta, D, \pi_\omega, \omega_{-i}) \propto \prod_{j=1}^{J} \prod_{k=1}^{K} \{\lambda_k \exp(x_{i,j}^T \beta_k + \omega_i)\} dN_{i,j,k} \exp\{-\Delta_k \lambda_k \exp(x_{i,j}^T \beta_k + \omega_i)Y_{i,j,k}\} \Pr(\omega_i|\omega_{-i}),
$$

where $\Pr(\omega_i|\omega_{-i})$ denotes the prior in (3.7). Metropolis-Hastings algorithm can be applied to sample $\omega_i$’s. Then $\pi_\omega$ is sampled from $\mathcal{G}(0.01 + \frac{I-B}{2}, 0.01 + \frac{1}{2} \sum_{i<j} (\omega_i - \omega_j)^2 1_{(i\sim j)})$.

### 3.5 Simulation studies

The performance of proposed models were evaluated through a simulation study. The spatial structure of simulation study is based on 45 zip code areas in Cincinnati. The Shapefile with boundary information for both Cincinnati and Southeastern Minnesota in the following smoking cessation data was downloaded from United States Census Bureau website. The name of Shapefile is “2014 ZIP Code Tabulation Areas (ZCTAs) Boundary File”. The raw data from website is a national level file. Then zip codes were used as match index to subset data. A total of 100 data sets were generated. Each data set contains 45 zip code areas in Cincinnati and 15 subjects within each area. Assume that the baseline function was $0.1 \sqrt{t}$.

For each data set, failure times were generated from a PH model with spatial frailties

$$
S(t|x_{i,j}, \omega_i) = S_0(t)^{\exp(x_{i,j}^T \beta(t) + \omega_i)},
$$
where $S_0(t) = \exp(\int_0^t \lambda_0(z) dz)$ and the time range of interest was set to be $(0, 9)$. The constant coefficient is set to 1 and the time-varying coefficient function is $\beta(t) = 0.5 + \sin(t\pi/9)$. In addition, two types of covariates (binary or continuous) are considered in this study. Then we simulated survival time $t$ from six combinations of coefficients and covariates, and the details are listed in Table 3.1.

To generate the censoring interval, the log-normal density function $\mathcal{LN}(x; 0, 0.4)$ was used to simulate gap times. If the event occurs between two consecutive visits, the time interval of the visits is taken as the censoring interval. If the event does not occur before maximum follow-up time, the subject would be considered as right censoring with the last visit time as the censoring point. To generate spatial frailties $\omega_i$, we first generated $\omega_i^*$ from multivariate normal $\mathcal{N}_n(0, (\pi_\omega W^*)^{-1})$, where $W^* = W + \text{diag}(0.0001, n)$, and then centered $\omega_i^*$ around...
zero to get the $\phi_i$. $W^*$ was used to make the precision matrix invertible and the sum to zero constraint was imposed to get identifiable $\omega_i$’s. Similar simulation setup of spatially correlated frailties can be found in Pan’s study (Pan et al., 2014).

Table 3.1 listed the details of six combinations of coefficients and covariates. The first three combos have constant coefficients and the other three are with time-varying coefficients. Combo 1, 2, 4, 5 contain one covariate (1, 2 are binary and 4, 5 are continuous), and Combo 3, 6 have two covariates. We ran 12,000 MCMC iterations and the first 2000 iterations were considered as burn-in period. In Models 2 and 3, hyper parameter $a_0$ was fixed at 100 so
Figure 3.2: Estimate of coefficients for combinations with two covariates. The solid lines in the middle are the true coefficients. The long dashed lines in the middle are medians of posterior means. The top and bottom dot dash lines are median of 95% credible intervals.

that the coefficient of first time grid is assigned a noninformative prior. Although not shown, the convergence of the MCMC chains were assessed by trace plot, autocorrelation plot, and Geweke’s diagnostics. Fast convergence and good mixing were observed for all parameters.

Figure 3.1 and Figure 3.2 present estimated coefficients of all six Combos from the 100 replicates. Figure 3.1 has the results from Combos 1, 2, 4, 5 with one covariate, whereas Figure 3.2 displays the results from Combos 3 and 6 with two covariates. The upper panel of figure contains Combos with one covariate and the lower panel is for Combos with two covariates. For each panel, the first row consists of plots from Model 1 with priors fixed
coefficient and spatially correlated frailties; the second row presents results from Model 2 with priors time-varying coefficient and independent frailties; the third row displays estimates from Model 3 with priors for time-varying coefficient and spatially correlated frailties. For constant coefficients, all the three models yield good estimates, and Model 1 gives much narrower credible intervals than Models 2 and 3. This is predictable because Models 2 and 3 provide pointwise estimates. On the other hand, advantage of Models 2 and 3 in recovering the time-varying coefficients is obvious and both of them also generate fairly well credible intervals. However, Model 1 yields biased estimates in Combo 4 and 5 since it misspecifies time-varying coefficients as constant. For every combination, the plot including estimates of either constant or time-varying coefficient from Models 2 and 3 are nearly same. This means the proposed method for time-varying coefficients is robust even spatial structure for frailties is not specified. To the authors’ best knowledge, one possible reason is frailties are treated as random intercept with normality distribution and centered by zero. Thus, in order to compare the performance of Models 2 and 3, we considered a few commonly used statistics, like Bayesian information criterion (BIC), deviance information criterion (DIC) (Spiegelhalter et al., 2002) and log pseudo marginal likelihood (LPML) (Ibrahim et al., 2005).

Table 3.2 shows the LPML comparison results between Model 2 and Model 3. Besides mean and (0.025, 0.975) quantile of LPML difference (Model 3 minus Model 2, so a positive mean supports that Model 3 is better), the percentage of LPML difference greater than zero out of 100 replicates is also reported. The means of LPML difference are all over 100, and the
Table 3.2: LPML comparison result between Model 2 and Model 3

<table>
<thead>
<tr>
<th></th>
<th>Combo 1</th>
<th>Combo 2</th>
<th>Combo 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>LPML</td>
<td>117.6 (-139.0, 391.3)</td>
<td>145.8 (-287.1, 498.3)</td>
<td>125.5 (-92.9, 498.3)</td>
</tr>
<tr>
<td>% Diff &gt; 0</td>
<td>84%</td>
<td>79%</td>
<td>89%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Combo 4</th>
<th>Combo 5</th>
<th>Combo 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>LPML</td>
<td>116.4 (-129.5, 343.7)</td>
<td>115.1 (-159.9, 539.4)</td>
<td>100.6 (-117.9, 357.6)</td>
</tr>
<tr>
<td>% Diff &gt; 0</td>
<td>85%</td>
<td>70%</td>
<td>81%</td>
</tr>
</tbody>
</table>

LPML Diff = LPML of Model 3 − LPML of Model 2. Mean and (0.025, 0.975) quantile from 100 replicates are reported. % Diff > 0 is calculated as percentage of LPML Diff > 0 over 100 replicates.

The largest percentage is 89% from Combo 3 and smallest percentage is 70% from Combo 5. The LPML comparison results indicate that Model 3 performs consistently better than Model 2. Based on the fact that both models provide similar coefficient estimates, the difference might be from estimate of frailties. We summarized results in Table 3.3 and in Figure 3.3 for model comparison.

In Table 3.3, models are compared based on frailty estimates. We first calculated mean bias of 45 frailty estimates from each model, and then counted the number with larger bias. Take Model 1 and Model 2 for example, the number equals 5 for Combo 1, which means there are 5 replicates that Model 1 yield larger mean bias than Model 2. Thus, a smaller number indicates a better model. It’s obvious that Model 1 and Model 3 provide much smaller biased frailty estimates than Model 2, which also explained the LPML comparison result in Table 3.2.
Table 3.3: Comparison of three models based on the number of larger biased frailty estimates out of 100 replicates

<table>
<thead>
<tr>
<th></th>
<th>Combo 1</th>
<th>Combo 2</th>
<th>Combo 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model 1 versus Model 2</td>
<td>5</td>
<td>9</td>
<td>4</td>
</tr>
<tr>
<td>Model 1 versus Model 3</td>
<td>59</td>
<td>58</td>
<td>59</td>
</tr>
<tr>
<td>Model 2 versus Model 3</td>
<td>96</td>
<td>94</td>
<td>94</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Combo 4</th>
<th>Combo 5</th>
<th>Combo 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model 1 versus Model 2</td>
<td>6</td>
<td>6</td>
<td>5</td>
</tr>
<tr>
<td>Model 1 versus Model 3</td>
<td>62</td>
<td>58</td>
<td>52</td>
</tr>
<tr>
<td>Model 2 versus Model 3</td>
<td>96</td>
<td>97</td>
<td>94</td>
</tr>
</tbody>
</table>

Figure 3.3: Maps of posterior means for the 45 spatial frailties.
To investigate the spatial pattern of frailties, we plotted the simulated terms and posterior means of frailty estimates from all three proposed models (Figure 3.3). The attached plot is from Combo 6, which has both constant and time-varying coefficients. Since the plots from six combinations are similar, we only included one here for example. All the four plots in Figure 3.3 share the same grayscale. Note that Model 1 and Model 3 generate a similar spatial frailty pattern to simulated values. However Model 2 fails to capture some large frailties in the Southeast corner.

### 3.6 Smoking cessation data

We apply the proposed methodology to analyze the aforementioned smoking cessation data. More details about the data information were discussed by Murray et al. (1998) in a Lung Health study with 5887 adult smokers followed up for 5 years. The data set used in this study is a subset consisting of 223 subjects who are known to have quit smoking at least once during the study period and have an identifiable Minnesota zip code of residence. The outcome of interest is the time to smoking relapse, which is interval-censored since the subjects were monitored at annual visits. The time scale has the study entry time as origin and the maximum of 5 years. The time to smoking relapse in its particular data set is either interval-censored (64/223 or 29%) or right-censored (159/223 or 71%). Two covariates are considered: gender (0=Male, 1=Female) and treatment (0=usual care (UC) group which received no special antismoking intervention, 1=smoking intervention (SI) group). Subjects
Figure 3.4: Estimate of coefficients for smoking cessation data. Black solid line is posterior mean. Black dashed lines are 95% credible intervals. The difference of effect in reducing relapse risk between smoking intervention and usual care are much smaller for the time around the two peaks (1.8 years and 3.8 years) than neighbor time period.

from the same zip code area are treated as a spatial cluster.

All three models were fitted to the data. We ran 12,000 MCMC iterations and the first 2,000 iterations were discarded as burn-in period. In Models 2 and 3, hyperparameter $a_0$ was fixed at 100 so that the coefficient of first time grid is assigned a noninformative prior. The estimated coefficient with 95% credible intervals for all three models are presented in Figure 3.4. From Model 1, the posterior of $\beta_{\text{gender}}$ has mean 0.355 and 95% credible interval ($-0.125, 0.871$), the posterior of $\beta_{\text{treatment}}$ has mean $-0.409$ and 95% credible interval ($-0.968, 0.132$). Although the credible intervals cover zero, the posterior means show that
women are more likely to relapse than men, and smoking intervention is effective in reducing relapse risk. These findings are consistent with the results in Banerjee and Carlin (2004). For Models 2 and 3, the coefficient of gender is close to straight line, from which we can conclude that the effect is stable over time. However, the coefficient of treatment exhibits obvious trend, suggesting that the inference based on Model 1 could be misleading. The same as simulation study results, Models 2 and 3 have similar trend of estimates: it increases from negative value to about zero around 1.8 years, then gradually approaches to \(-1\) between 2 years and 3 years. After that, the trend reaches the second peak around 3.8 years and ends with a downward curve.

Figure 3.5 maps the posterior means of spatial frailties \((\omega_i)\) for the smoking data by using Models 1, 2 and 3. Each plot contains 89 zip codes: Goodhue county (10); Rice county (6); Steele County (3); Wabasha county (9); Dodge county (6); Mower county (14); Olmsted County (11); Winona County (10); Fillmore County (12); Houston County (7); Waseca County (1). Only 51 of them contains data and they are colored with gray scale from darkest “Grey 0” to lightest “Grey 85”. The other zip code areas without data are plotted with white color. Models 1 and 3 reveal similar spatial patterns for frailties, where a few higher values of \(\omega_i\)’s occur in the northwest region, which happens to be close to the southern suburbs of Minneapolis. This indicates that higher risks of smoking relapse were observed in those regions. These findings also agree with the pattern from Banerjee and Carlin (2004), where a spatial multivariate CAR Weibull cure rate model was fitted. However, Model 2 with independent frailty prior does not show similar spatial trend. Based on the results from
Figure 3.5: Maps of posterior means for the spatial frailties $\omega_i$ over 51 zip codes areas in southeast Minnesota based on Models 1, 2 and 3. The white areas are those zip codes without data.

previous studies and this research, a spatial correlation concern for frailties is preferred for this smoking cession data. The LMPL from Model 3 (-201.2) is larger than LPML from Model 2 (-205.8), which indicates that Model 3 performs better than Model 2.
3.7 Summary and discussion

In this chapter, we have a Bayesian dynamic model for clustered interval-censored data, where the correlation is modeled by spatially correlated frailties. The results of proposed methods show a consistent spatial pattern of frailties with previous studies. However, unlike the constant estimates of smoking intervention effect for smoking relapse in other research, we have found a time-varying trend.

The proposed methods for time-varying coefficient and the spatially correlated frailties can be easily extended to other regression models or frailties’ correlations. However, the work could be complicated for some scenarios, for example, when parametric assumption of frailties is violated. Nonparametric priors may be needed to relax the parametric assumption. In addition, the time-varying coefficient models (Models 2 and 3) provide wider credible intervals than constant coefficient model (Model 1) due to piecewise nature. Thus, the theoretical property of priors for time-varying coefficient remains to be further studied.
Chapter 4

Copula model for bivariate data

In this chapter, we propose a Bayesian method for bivariate data with interval-censored failure time. Copula model is applied to account for the dependence among outcomes. For marginal distribution, Cox model is used to illustrate the properties of failure times. The techniques can be easily extended to other distributions of failure time. The joint estimation of dependence structure and regression parameters is accomplished via Markov chain Monte Carlo (MCMC). Simulation studies are performed to evaluate the proposed method. A pediatric study on children’s language development is exemplified for illustration.

4.1 Introduction

For bivariate data, estimation of association parameter and testing of independence also need to be considered. In survival analysis, copula model (Clayton, 1978; Oakes, 1989) is a
commonly used tool to model bivariate data, which assumes that the joint survival function of the individuals within a cluster is given by a copula function.

Bivariate Cox model for interval-censored data is much less developed compared to right-censored data. For case I interval censoring or current status data, where the failure time of interest is not observed but known only to be smaller or larger than an observation or a monitoring time, Wang et al. (2008) considered efficient estimation or regression and association parameters jointly with copula method. For case II interval-censored data, where the failure time is located within 2 observation times, Sun et al. (2006) studied the problem of estimating the association between two related survival variables when they follow a copula model.

In this study, we propose a Bayesian approach to model bivariate survival data with case II interval censoring and semiparametric marginals. The copula model is considered for the association structure. The idea to estimate model parameters is from Sinha et al. (1999), where augmented event time is generated in its observed time interval and parameters are estimated piecewisely. The study is motivated by a children’s neurodevelopmental data, where the primary interest is the impact of in-utero exposure to selective serotonin reuptake inhibitors (SSRI) on the milestones of language and motor skills.

The remainder of this chapter is organized as follows. Section 4.2 discusses copula model, data structure and marginal likelihood. Section 4.3 describes joint likelihood function, prior specification and posterior inference. Section 4.4 introduces the simulation studies of proposed method. A children language development data is analyzed in Section 4.5. Summary
of conclusions and discussions are enclosed in Section 4.6.

4.2 Model specification

The copula model is one of most commonly used tools to model association structure for bivariate data. An important family of copulas is the Archimedean Copula (Genest and MacKay, 1986), where

\[
C_\alpha(u,v) = \phi_\alpha\{\phi_\alpha^{-1}(u) + \phi_\alpha^{-1}(v)\}, \quad 0 \leq u,v \leq 1,
\]

with \( \phi : [0, +\infty] \rightarrow [0, 1], \phi(0) = 1, \phi' < 0, \phi'' > 0. \) Consider the positive stable copula (Hougaard, 1986a,b), the Laplace transform is \( \phi(u) = \exp(-u^\alpha) \) so that \( \phi^{-1}(u) = (-\log u)^{1/\alpha}. \) The corresponding copula takes the form

\[
C_\alpha(u,v) = \exp\{-[(-\log(u))^{1/\alpha} + (-\log(v))^{1/\alpha}]^\alpha\}, \quad 0 < \alpha \leq 1
\]

Let \((T_1, T_2)\) be continuous random variables, \((S_1, S_2)\) and \((f_1, f_2)\) be corresponding marginal survival and density functions, respectively. Therefore the joint survival function in the positive stable copula is

\[
S(t_1, t_2) = C_\alpha(S_1(t_1), S_2(t_2)) = \exp\{-[(-\log(S(t_1)))^{1/\alpha} + (-\log(S(t_2)))^{1/\alpha}]^\alpha\},
\]

and joint density function is

\[
f(t_1, t_2) = c_\alpha(S_1(t_1), S_2(t_2))f_1(t_1)f_2(t_2),
\]
where $c_{\alpha}(.,.)$ is a density function related to the copula $C_{\alpha}$ and given by

$$c_{\alpha}(S_1(t_1), S_2(t_2)) = \frac{\partial C_{\alpha}(S_1(t_1), S_2(t_2))}{\partial S_1(t_1) \partial S_2(t_2)} = \frac{\partial \exp\{-(\frac{-(\log(S_1(t_1)))}{\alpha} + \frac{-(\log(S_2(t_2)))}{\alpha})^\alpha\}}{\partial S_1(t_1) \partial S_2(t_2)}. \quad \text{(4.1)}$$

Let $S_{\alpha} = (-(\log(S(t_1))))^{1/\alpha} + (-(\log(S(t_2))))^{1/\alpha}$, then (4.1) becomes

$$\frac{\partial \exp\{-S_{\alpha}\}}{\partial S_1(t_1) \partial S_2(t_2)} = \frac{\partial \left(\exp\{-S_{\alpha} S_{\alpha}^{-1}(-(\log(S_1(t_1))))^{1/\alpha-1} \frac{1}{S_1(t_1)}\}\right)}{\partial S_2(t_2)} = \frac{\exp\{-S_{\alpha} S_{\alpha}^{-2}(-(\log(S_1(t_1))))^{1/\alpha-1}(-(\log(S_2(t_2))))^{1/\alpha-1}(S_{\alpha} - 1 + 1/\alpha)}{S_1(t_1) S_2(t_2)}. \quad \text{(4.3)}$$

Let $T_{1,j}$ and $T_{2,j}$ be the unobserved event time, $(L_{1,j}, R_{1,j}]$ and $(L_{2,j}, R_{2,j}]$ be the observed censoring intervals for subject $j$, where $j = 1, 2, ..., n$. Suppose $T_{1,j}$ and $T_{2,j}$ follow the proportional hazards model marginally (Cox, 1972). Conditional on a $P -$ dimensional vector of covariates $x$, the hazard function is

$$\lambda_i(t|x) = \lambda_{0,i}(t) \exp(x^T \beta_i), \quad i = 1, 2,$$

where $\lambda_{0,i}(t)$ denotes the baseline hazard function, and $\beta_i$ is a $P -$ dimensional vector of regression parameters representing covariate effects. Then the contribution of subject $j$ to the marginal observed data likelihood is

$$p(T_{i,j} \in (L_{i,j}, R_{i,j}]|x_i) = p(T_{i,j} > L_{i,j}|x_i) - p(T_{i,j} > R_{i,j}|x_i).$$

When the cumulative hazard function is a totally unspecified nondecreasing function, it’s always challenging to estimate regression parameters directly since the partial likelihood
dose not exist for interval-censored data (Sun, 2007). In this study, we borrowed an idea of Bayesian discretized semiparametric model, which was proposed by Sinha et al. (1999). The basic idea is to generate augmented event time in its observed interval and compute the estimates of parameters piecewisely. Once the unobserved event time $T_{i,j}$’s are given, the interval-censored data reduce to right-censored data. Let $k = 1, 2, \ldots, K$ denote all the ordered grid points and $0 = s_0 < s_1 < s_2 < \ldots < s_K < \infty$ be corresponding time. Assume that $\lambda_{0,i}(t)$ is left continuous step function and denote $\lambda_{i,k} = \lambda_{0,i}(s_k)$. Let $dN_{i,j,k}$ indicate whether or not the event time $T_{i,j}$ falls within the $k^{th}$ interval, i.e., $dN_{i,j,k} = 1(T_{i,j} \in (s_{k-1}, s_k])$. Let $Y_{i,j,k}$ be the at-risk variable. If $dN_{i,j,k} = 1$ for some value $k$, then $Y_{i,j,l} = 1$ for $l < k$, $Y_{i,j,l} = 0$ for $l > k$, and $Y_{i,j,k} = (T_{i,j} - s_{k-1})/\Delta_k$ for $l = k$, where $\Delta_k = s_k - s_{k-1}$ is the width of the $k^{th}$ interval. The augmented likelihood function for $j^{th}$ subject is

$$
\ell_{i,j}(\Theta_i|\{dN_{i,j,k}, Y_{i,j,k}\}_{k=1}^{K}, x_j) = \prod_{k=1}^{K}\{\lambda_{i,k}\exp(x_j^T \beta_i)\}^{dN_{i,j,k}}\exp\{ -\Delta_k \lambda_{i,k}\exp(x_j^T \beta_i)Y_{i,j,k}\},$

(4.4)

where $j = 1, 2, \ldots, n$, and $\Theta_i = \{\lambda_{i,k}, \beta_i; i = 1, 2, k = 1, 2, \ldots, K\}$.

By introducing latent variable $\{dN_{i,j,k}, Y_{i,j,k}\}_{k=1}^{K}$, the hazard function and survival function can be written as

$$
\lambda_i(t|\{dN_{i,j,k}, Y_{i,j,k}\}_{k=1}^{K}, x_j) = \prod_{k=1}^{K}\{\lambda_{i,k}\exp(x_j^T \beta_i)\}^{dN_{i,j,k}},$

and

$$
S_i(t|\{dN_{i,j,k}, Y_{i,j,k}\}_{k=1}^{K}, x_j) = \exp\{ -\sum_{k=1}^{K}\Delta_k \lambda_{i,k}\exp(x_j^T \beta_i)Y_{i,j,k}\},$

(4.5)

respectively.
4.3 Joint likelihood, prior and posterior

4.3.1 Joint likelihood

In the case of the joint estimation for the marginals and association parameter under the Bayesian approach, the joint likelihood distribution is

$$\ell(\alpha, \Theta_1, \Theta_2 | \{dN_{i,j,k}, Y_{i,j,k}\}_{k=1}^{K}, x_j) = \prod_{j=1}^{n} (c_\alpha(S_1(t_1), S_2(t_2)) f_1(t_1) f_2(t_2))^{\delta_{1,j}\delta_{2,j}}$$

$$\times \left( \frac{\partial C_\alpha(S_1(t_1), S_2(t_2))}{\partial S_1(t_1)} (-f_1(t_1)) \right)^{\delta_{1,j}(1-\delta_{2,j})}$$

$$\times \left( \frac{\partial C_\alpha(S_1(t_1), S_2(t_2))}{\partial S_2(t_2)} (-f_2(t_2)) \right)^{(1-\delta_{1,j})\delta_{2,j}}$$

$$\times C_\alpha(S_1(t_1), S_2(t_2))^{(1-\delta_{1,j})(1-\delta_{2,j})}, \quad (4.6)$$

where $c_\alpha(S_1(t_1), S_2(t_2))$ has the form in (4.3), $f_i(t)$ takes the form in (4.4), $S_i(t)$ is defined in (4.5), and $\delta_{i,j}$ indicates that whether event happens in interested time range $(0, s_K]$, i.e., $\delta_{i,j} = 1(\sum_{k=1}^{K} dN_{i,j,k} > 0)$. The partial derivative term with respect to $S_1(t_1)$ is computed in (4.2), and the partial derivative term with respect to $S_2(t_2)$ takes the similar form. Then, (4.6) can be rewritten as

$$\ell(\alpha, \Theta_1, \Theta_2 | \{dN_{i,j,k}, Y_{i,j,k}\}_{k=1}^{K}, x_j)$$

$$= \prod_{j=1}^{n} (S_\alpha^\alpha - 1 + 1/\alpha)^{\delta_{1,j}\delta_{2,j}} \exp(-S_\alpha^\alpha) S_\alpha^{\alpha(\delta_{1,j}+\delta_{2,j}-\delta_{1,j}\delta_{2,j})-\delta_{1,j}-\delta_{2,j}}$$

$$\times \left( \prod_{k=1}^{K} [\lambda_{1,k} \exp(x_j \beta_1)^{dN_{i,j,k}}] \right) \left( \sum_{k=1}^{K} \Delta_k \lambda_{1,k} Y_{i,j,k} \exp(x_j \beta_1) \right)^{\delta_{1,j}(1/\alpha-1)}$$

74
\[ \times \left( \prod_{k=1}^{K} [\lambda_{2,k} \exp(x_j \beta_2)]^{dN_{2,j,k}} \right) \left( \sum_{k=1}^{K} \Delta_k \lambda_{2,k} Y_{2,j,k} \exp(x_j \beta_2) \right)^{d_{2,j}(1/\alpha - 1)}, \] (4.7)

where \( S_\alpha = \left( \sum_{k=1}^{K} \Delta_k \lambda_{1,k} Y_{1,j,k} \exp(x_j \beta_1) \right)^{1/\alpha} + \left( \sum_{k=1}^{K} \Delta_k \lambda_{2,k} Y_{2,j,k} \exp(x_j \beta_2) \right)^{1/\alpha} \).

### 4.3.2 Prior specification

For Bayesian inference, we need to assign priors for \( \{ \alpha, \lambda_{i,k}, \beta_i \} \), where baselines from two marginal distributions of all time grids share the same prior, and so does coefficient. The prior distributions applied in this study are specified as follows

\[
\lambda_{i,k} \sim \mathcal{G}(\eta_k, \gamma_k), \quad i = 1, 2; \quad k = 1, 2, ..., K,
\]

\[
\beta_{i,p} \sim \mathcal{N}(\beta_{i,p,0}, \omega_{i,p,0}^2), \quad i = 1, 2; \quad p = 1, 2, ..., P,
\]

where \( \mathcal{G}(\eta_k, \gamma_k) \) denotes gamma distribution with shape parameter \( \eta_k \) and rate parameter \( \gamma_k \), and \( \mathcal{N}(\beta_{i,p,0}, \omega_{i,p,0}^2) \) is normal distribution with mean \( \beta_{i,p,0} \) and variance \( \omega_{i,p,0}^2 \). We assume that hyperparameters \( \eta_k \) and \( \gamma_k \) are known in advance. Hyperparameters \( \beta_{i,p,0} \) and \( \omega_{i,p,0}^2 \) are assigned conjugate priors with relatively large variance as follows

\[
\beta_{i,p,0} \sim \mathcal{N}(0, 10),
\]

\[
\omega_{i,p,0}^2 \sim \mathcal{IG}(0.1, 0.1),
\]

where \( i = 1, 2, \quad p = 1, 2, ..., P, \) and \( \mathcal{IG}(0.1, 0.1) \) denotes inverse gamma distribution where both shape and scale parameter are assigned 0.1. In addition, the association parameter \( \alpha \) is assigned a uniform prior of \( \mathcal{U}(0, 1) \).
4.3.3 Posterior computation

First we need to sample event time $T_{i,j}$. We give the details for sampling $T_{1,j}$ and sampling of $T_{2,j}$ follows similar steps. We derive the survival function of $t_1$ conditional on $t_2$ and other parameters as follows

$$S_1(t_1|T_2 = t_2) = p(T_1 > t_1|T_2 = t_2)$$

$$= \left( \frac{\partial C_\alpha(S_1(t_1), S_2(t_2))}{\partial S_2(t_2)} (-f_2(t_2)) \right) \times \left( \frac{1}{f_2(t_2)} \right)$$

$$= \exp \left( -S_\alpha^\alpha S_\alpha^{\alpha-1} \frac{(-\log(S_2(t_2)))^{1/\alpha-1}}{S_2(t_2)} \right),$$

Because of the piecewise constant nature of baseline hazards, sampling of $T_{1,j}$ can be decomposed into two steps:

1. Determine which grid interval contains the event time. For subject $j$, vector $(dN_{1,j,1}, dN_{1,j,2}, ..., dN_{1,j,K})$ follows a multinomial distribution with size 1 and probability vector $(e_{1,j,1}, e_{1,j,2}, ... e_{1,j,K})$

$$e_{1,j,k} = \frac{p_{1,j,k} \mathbb{1}(s_k \in (L_{1,j}, R_{1,j}])}{\sum_{s_l \in (L_j, R_j]} p_{1,i,l}},$$

$$p_{1,i,l} = S_1(s_{k-1}|T_2 = t_2) - S_1(s_k|T_2 = t_2).$$

2. Sample the exact time within the selected time grid. If $dN_{1,j,k} = 1$, the event time $T_{1,j}$ follows a doubly truncated distribution on $(s_{k-1}, s_k]$ with distribution form

$$F(u) = \frac{S_1(s_{k-1}|T_2 = t_2) - S_1(u|T_2 = t_2)}{S_1(s_{k-1}|T_2 = t_2) - S_1(s_k|T_2 = t_2)}.$$

Using the inverse distribution function method, sampling of $T_{1,j}$ is then straightforward.
The posterior samples of parameters $\alpha$, $\lambda_{i,k}$, $\beta_{i,p}$ and hyperparameters $\beta_{i,p,0}$, $\omega_{i,p,0}^2$ are generated via Markov chain Monte Carlo (MCMC). The detailed development of the MCMC sampling algorithm is given as follows.

**Metropolis-Hastings update on $\alpha$**

The full condition distribution of $\alpha$ is given by

$$p(\alpha|\cdot) \propto \prod_{j=1}^n (S_{\alpha}^{\alpha} - 1 + 1/\alpha)^{\delta_{1,j} \delta_{2,j}} \exp\left(-S_{\alpha}^{\alpha} S_{\alpha}^{\alpha} (\delta_{1,j} + \delta_{2,j}) - \delta_{1,j} - \delta_{2,j}\right) \times \left(\sum_{k=1}^K \Delta_k \lambda_{1,k} Y_{1,j,k} \exp(x_j \beta_1)\right)^{\delta_{1,j} (1/\alpha - 1)} \times \left(\sum_{k=1}^K \Delta_k \lambda_{2,k} Y_{2,j,k} \exp(x_j \beta_2)\right)^{\delta_{2,j} (1/\alpha - 1)} \times \pi(\alpha),$$

where $\pi(\alpha)$ is a density function of uniform distribution. To facilitate sampling $\alpha$ from the given boundary of the uniform prior, a logit transformation (Wang et al., 2015) is applied as $\eta = \text{logit}\{(\alpha - \alpha_{\text{min}})/(\alpha_{\text{max}} - \alpha_{\text{min}})\}$ and $\eta_{m+1} \sim \mathcal{N}(\eta_m, \alpha_{\eta}^2)$, where $\eta_m$ is the value of $\eta$ at the $m^{th}$ iteration and $\alpha_{\eta}^2$ is the tuning constant to achieve a desired acceptance rate. Then $\alpha_{m+1} = \{\exp(\eta_{m+1}) \alpha_{\text{max}} + \alpha_{\text{min}}\}/\{1 + \exp(\eta_{m+1})\}$. The probability of updating $\alpha$ is decided by the minimum of 0 and $\log\{p(\alpha_{m+1}|\cdot)(\alpha_{m+1} - \alpha_{\text{min}})(\alpha_{\text{max}} - \alpha_{m+1})\} - \log\{p(\alpha_m|\cdot)(\alpha_m - \alpha_{\text{min}})(\alpha_{\text{max}} - \alpha_m)\}$, where $\alpha_{\text{min}} = 0$ and $\alpha_{\text{min}} = 1$. 

77
Metropolis-Hastings update on $\beta_i$

The full condition distribution of $\beta_1$ is given by

$$
p(\beta_1 \mid \cdot) \propto \prod_{j=1}^{n} (S_{\alpha}^a - 1 + 1/\alpha)^{\delta_{1,j} \delta_{2,j}} \exp(-S_{\alpha}^a) S_{\alpha}^a (\delta_{1,j} + \delta_{2,j} - \delta_{1,j} \delta_{2,j}) - \delta_{1,j} - \delta_{2,j}$$

$$\times \left( \prod_{k=1}^{K} [\lambda_{1,k} \exp(x_j \beta_1)]^{dN_{1,j,k}} \right) \left( \sum_{k=1}^{K} \Delta_k \lambda_{1,k} Y_{1,j,k} \exp(x_j \beta_1) \right)^{\delta_{1,j}(1/\alpha - 1)}$$

$$\times \exp \left( -\frac{(\beta_1 - \beta_0)^2}{2\omega_0^2} \right).$$

We draw $\beta_1$ using the Metropolis-Hastings algorithm. For the $m^{th}$-current MCMC iteration, a new value for $\beta_{1,m+1}$ is sampled from $\mathcal{N}(\beta_{1,m}, \omega_0^2 \tau)$, where $\beta_{1,m}$ is the value of $\beta_1$ at the $m^{th}$ iteration, $\omega_0$ is the prior variance of $\beta_1$, and $\tau$ is the tuning parameter which is selected to achieve desired acceptance rate. Each element is updated based on the probability

$$\min \left\{ 1, \frac{p(\beta_{1,m+1} \mid \cdot)}{p(\beta_{1,m} \mid \cdot)} \right\}.$$ The parameter $\beta_2$ is sampled in the same way.

Metropolis-Hastings update on $\lambda_{i,k}$

To accelerate convergence, each element of $\lambda$ is sampled separately. We illustrated $\lambda_{1,1}$ for example, and all other elements of $\lambda$ share the same strategy. The full condition distribution of $\lambda_{1,1}$ takes the form

$$
p(\lambda_{1,1} \mid \cdot) \propto \prod_{j=1}^{n} (S_{\alpha}^a - 1 + 1/\alpha)^{\delta_{1,j} \delta_{2,j}} \exp(-S_{\alpha}^a) S_{\alpha}^a (\delta_{1,j} + \delta_{2,j} - \delta_{1,j} \delta_{2,j}) - \delta_{1,j} - \delta_{2,j}$$

$$\times (\lambda_{1,1} \exp(x_j \beta_1))^{dN_{1,j,1}} \left( \sum_{k=1}^{K} \Delta_k \lambda_{1,k} Y_{1,j,k} \exp(x_j \beta_1) \right)^{\delta_{1,j}(1/\alpha - 1)}$$

$$\times \pi(\lambda_{1,1}).$$
where \( \pi(\lambda_{1,1}) \) is the prior distribution of \( \lambda_{1,1} \). For the \( m^{th} \)-current MCMC iteration, \( \log(\lambda_{1,1}^{m+1}) \) is drawn from \( \mathcal{N}(\log(\lambda_{1,1}^m), \sigma^2) \), where \( \lambda_{1,1}^m \) is the value of \( \lambda_{1,1} \) at the \( m^{th} \) iteration and \( \sigma^2 \) is assigned an appropriate value to adjust acceptance rate. The new value is accepted with the probability given by

\[
\min \left\{ 1, \frac{p(\lambda_{1,1}^{m+1}| \cdot) \times \lambda_{1,1}^{m+1}}{p(\lambda_{1,1}^m| \cdot) \times \lambda_{1,1}^m} \right\}.
\]

**Gibbs sampling update on \( \beta_{i,p,0} \) and \( \omega_{i,p,0}^2 \)**

Both \( \beta_{i,p,0} \) and \( \omega_{i,p,0}^2 \) can be sampled from conditional posterior distribution as follows

\[
\beta_{i,p,0} \sim \mathcal{N}(\beta_{i,p}/11, 11),
\]

\[
\omega_{i,p,0}^2 \sim \mathcal{IG}(0.6, 0.1 + 10\beta_{i,p}^2/22).
\]

### 4.4 Simulation results

In this section, we present simulation study results to demonstrate the properties of the methods introduced in previous sections. The association parameter \( \alpha \) was set to be 0.7, which indicates a moderate association between events. Then \( T_1 \) and \( T_2 \) can be calculated from marginal survival function \( S(t) = \exp\{- \int_0^t \lambda_i(t) \exp(\mathbf{x}_j^T \beta_i) dt\} \). In this study, both \( \lambda_1 \) and \( \lambda_2 \) equal 0.5, and covariate \( \mathbf{x}_j \) is sampled from Bernoulli distribution \( \mathcal{B}(0.5) \). For coefficients, we considered the pairs \((1, 1), (1, 0), (1, -1)\) to illustrate the properties of proposed method. To generate the censoring interval, the log-normal density function \( \mathcal{LN}(t; 0, 0.8) \) was used to simulate gap times. If the event occurs between two consecutive visits, the time
Figure 4.1: Trace plots and density plots for $\alpha$, $\beta_1$ and $\beta_2$. True values are 0.7, 1, 1, respectively.
interval of the visits is taken as the censoring interval. It’s assumed that bivariate data share the same follow up time for each subject, which often occurs in practice because it’s common to have information of multiple events in a single visit. For the simulated data set, to get a better estimate of associate parameter, each event has observed time interval, which means only interval-censored data included. The total time grids $K$ was set to 8, and each time grid contains approximately equal number of observations with covariate equals 0, which improved the estimate of baseline compared to equally spaced time grids. We ran 5,500 MCMC iterations and the first 500 iterations were considered as burn-in period. For the prior hyperparameters of baseline, we set $\eta_k = 0.2$ and $\gamma_k = 0.4$, which was also used in Sinha et al. (1999). Each simulated dataset contains 200 subjects and the results are summarized based on 100 replicates. Figure 4.1 displays trace plot and density plot for $\alpha$, $\beta_1$ and $\beta_2$ for the coefficient pair (1, 1) from one replicate. Although not shown here, the convergence of the MCMC chains were assessed by autocorrelation plot and Geweke’s diagnostics. Fast convergence and good mixing were observed for all parameters.

Table 4.1 summarizes the estimation results of proposed model with three cases of coefficients. For each parameter, true value is reported, the bias is calculated as the average of the 100 posterior means minus the true value (Bias), the sample standard deviation (SSD) is the sample standard deviation of the 100 posterior means, the empirical standard error (ESE) is the average of the 100 estimated standard errors, and the 95% coverage probability (CP) is the percent of the 100 credible intervals for each parameter that contains the true value. The results in Table 4.1 suggest that the propose model works very well for all the three cases.
Table 4.1: Simulation results of the proposed method

<table>
<thead>
<tr>
<th>Parameter</th>
<th>True</th>
<th>Bias</th>
<th>SSD</th>
<th>ESE</th>
<th>CP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Case 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>α</td>
<td>0.7</td>
<td>0.0620</td>
<td>0.0493</td>
<td>0.0632</td>
<td>0.9</td>
</tr>
<tr>
<td>β₁</td>
<td>1</td>
<td>0.0275</td>
<td>0.2032</td>
<td>0.2046</td>
<td>0.95</td>
</tr>
<tr>
<td>β₂</td>
<td>1</td>
<td>0.0014</td>
<td>0.1833</td>
<td>0.2027</td>
<td>0.97</td>
</tr>
<tr>
<td>Case 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>α</td>
<td>0.7</td>
<td>0.0437</td>
<td>0.0503</td>
<td>0.0610</td>
<td>0.93</td>
</tr>
<tr>
<td>β₁</td>
<td>1</td>
<td>0.0418</td>
<td>0.1995</td>
<td>0.2021</td>
<td>0.98</td>
</tr>
<tr>
<td>β₂</td>
<td>-1</td>
<td>0.0043</td>
<td>0.1659</td>
<td>0.1782</td>
<td>0.98</td>
</tr>
<tr>
<td>Case 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>α</td>
<td>0.7</td>
<td>0.0379</td>
<td>0.0487</td>
<td>0.0621</td>
<td>0.95</td>
</tr>
<tr>
<td>β₁</td>
<td>1</td>
<td>0.0230</td>
<td>0.1957</td>
<td>0.2059</td>
<td>0.96</td>
</tr>
<tr>
<td>β₂</td>
<td>0</td>
<td>-0.0356</td>
<td>0.1388</td>
<td>0.2552</td>
<td>1</td>
</tr>
</tbody>
</table>

In terms of point estimates, precision, and coverage probabilities. The bias of parameters is relatively small, and the sample standard deviation is close to empirical standard error. The coverage probabilities are close to the nominal level of 0.95. In Case 3, the coverage probability for $\beta_2 = 0$ is 1 due to larger empirical standard error. This may be caused by the identification issue between coefficient and baseline.
4.5 Children neurodevelopmental data

In this section, we apply the methodology described in previous sections to childrens neurodevelopmental data from National Database for Autism Research (NDAR). Characterized by social and communication deficits, Autism spectrum disorder (ASD) has affected 1 in every 68 children in the U.S. Although the role of genetic factors in ASD has been extensively documented, recent evidence suggests that environmental factors may contribute to 55% of the variance in liability to ASD. The identification of environmental contributors may hence provide a key step towards reducing the increasing rate of ASD. In-utero exposure to selective serotonin reuptake inhibitors (SSRI) has been speculated to affect the risk of ASD. However, the impact of in-utero exposure to SSRI on neurodevelopmental outcomes remains controversial due to mixed findings. The two outcomes of interest are the time to development milestones of language and motor skills. The time to milestone of language development is defined as the age of first word to the age of first phrase. In some literatures, the age of first word was defined as time of language development. However, for the children with ASD, some of them can never speak a complete sentence. Thus, to use the age of first word is not appropriate. The time to milestone of motor skill development is defined as the age of learning how to sit to the age of starting walk. The NDAR data involved 731 children with diagnosis of ASD. The diagnoses of ASD have been made based on Autism Diagnostic Interview-Revised (ADI-R) Lord et al. (1994). Information about ages of developmental milestones related to language and motor skills was based on parental reports. The history
Table 4.2: Results of children’s neurodevelopmental data analysis

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>95% HPD interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha$</td>
<td>0.55</td>
<td>0.02</td>
<td>(0.51, 0.60)</td>
</tr>
<tr>
<td>$\beta_1$ (language)</td>
<td>0.25</td>
<td>0.39</td>
<td>(-0.58, 0.96)</td>
</tr>
<tr>
<td>$\beta_2$ (motor skill)</td>
<td>0.49</td>
<td>0.34</td>
<td>(-0.20, 1.14)</td>
</tr>
</tbody>
</table>

of in-utero exposure to SSRIs was established based on self-reported information.

For the analysis, we considered the covariate SSRI as a dichotomized variable which denotes the status of in-utero exposure to selective serotonin reuptake inhibitors (0=No, 1=Yes). The time grids were set as defined in Section 4.4. We ran 5,500 MCMC iterations and the first 500 iterations were considered as burn-in period. The convergence of the MCMC chains were assessed by trace plot, autocorrelation plot, and Geweke’s diagnostics (the plots and numbers are not shown). Fast convergence and good mixing were observed for all parameters.

Table 4.2 presents the analysis results by applying the proposed method. The results include the point estimates, standard errors, and 95% credible intervals. $\beta_1$ and $\beta_2$ denote the coefficients of SSRI on language and motor skill development, respectively. Note that there is a positive association between the time to development milestones of language and motor skills ($\alpha = 0.55$ with 95% HPD interval (0.51, 0.60)).
4.6 Conclusions and discussions

In this study, we present a Bayesian approach to model bivariate survival data with “case II” interval censoring and semiparametric marginals. The copula model is considered for the association structure. The idea to estimate model parameters is based on augmented event time from observed time interval and parameters are estimated piecewisely. Simulation results in Section 4.5 demonstrates the properties of the proposed method, and children’s neurodevelopmental data analysis in Section 4.6 provides an illustrative example of such a method.

We have assumed Cox model for the censoring time, which is more general than parametric models. This assumption is not essential in the proposed method and can be easily extended to other marginal distributions. In addition, the joint survival function of the two related survival times is assumed to follow a positive stable copula model to illustrate the computation steps of proposed method. Actually, the Bayesian approach with other Archimedean copulas can also be easily implemented, where the only difference is joint survival function.

In the proposed approach, the covariate effect on outcomes is assumed to be constant. However, this assumption may not be true in practice since the relative risk of two subjects may change over time. For example, in asthma study, the effect of airway reactivity on recurrent wheezing is changing over time (Yu et al., 2013). Thus, one possible area of future study is to detect the time-varying covariate effect and model such temporal effect by time
dependent coefficient.
Chapter 5

Summary and future work

In this thesis, we developed novel Bayesian Cox models for interval-censored data. The properties of these models were illustrated by simulation studies. For practical purpose, the proposed models were applied to three real data.

In Chapter 2, to model clustered interval-censored data with time-varying covariate effect, we proposed a Bayesian model with frailty for unobserved heterogeneity or statistical dependence between observations, and time dependent coefficient for time-varying covariate effect. We estimate coefficients based on preassigned time grids and the dimensions of coefficients are automatically determined by reversible jump Markov chain Monte Carlo algorithm. In this chapter, constant shared gamma frailty is utilized. Simulation studies were conducted to evaluate the performance of the proposed method. The methodology was applied to a pediatric study on children’s dental health data, and we found an obvious trend of the coefficient estimate and within cluster correlation among the interaction of gender
and province.

In Chapter 3, we proposed a Bayesian approach to dynamic Cox regression model allowing for spatial correlation with interval-censored survival data. In spatial analysis, clusters can be formed based on geographical areas, and this feature is imported as spatially correlated frailties. We also apply reversible jump MCMC to generate posterior samples of baseline and coefficients, and the conditional autoregressive (CAR) distribution is used as the prior for spatial frailties. We applied the proposed method to smoking cessation data in southeastern Minnesota, and found that ignoring of spatial correlation may lead to bias of estimates for frailties.

In Chapter 4, we proposed a Bayesian copula model for bivariate survival data with “case II” interval censoring and semiparametric marginals. The results of simulation studies illustrate the properties of the proposed method. The proposed method was exemplified by children neurodevelopmental data.

In this thesis, the log-baseline hazard and coefficient are both assigned a dynamic normal prior and the dimension of parameters is controlled by reversible jump MCMC. Although the estimates have much narrower credible intervals and are much smoother compared to previous methods without applying reversible jump MCMC, they still have relatively larger credible intervals than the constant coefficient model to fit simulated data with constant coefficient. Meanwhile, the point estimates close to the right tail are usually biased due to less observations in the right tail of survival curve. We may need to consider either adjust reversible jump MCMC algorithm, or apply a more flexible prior in future work. Further
more, a selection of unbalanced time grids, e.g., time grids are wider in the right tail to include more subjects, can also be considered.

All the priors of parameters are assumed to be parametric distribution in this thesis, e.g., normal prior for coefficient, gamma or log-normal prior for frailty. However, this parametric prior assumption is hard to verify and may not be true. This may lead to poor parameter estimates when the distribution is misspecified. These possibilities can be addressed when the distribution is drawn from a large class of distributions. Such a large distribution could be formed by using nonparametric approaches. See for example, Bayesian Dirichlet process was used for the frailty effects distribution (Ferguson, 1973).

In the third project, constant coefficient was considered due to the reason that proportional hazards assumption is satisfied for the models with SSRI covariate effect on outcomes of language and motor development milestone time. Whereas, this model can also be extended to a more general one with time-varying coefficient if it’s necessary in practice. Since there two events of main interest, we proposed a bivariate model by applying copula method. A multivariate model is also possible to develop if there are more than two outcomes.

Another assumption in this thesis is that censoring is non-informative, meaning that for each individual, the censoring time is statistically independent of the event time. In practice, we may also need to consider informative censoring, or a sensitivity analysis on the estimates of the parameter of interest.

All the three proposed models, constant shared frailty model with time-varying coefficient, spatial frailty model with time-varying coefficient, and copula model for bivariate
data, are based on Cox model due to a simple reason that Cox model is the most popular survival model. In future work, we can also extend these techniques to other models, e.g., proportional odds model, or a even more general model, transformation model for survival data, which includes both the Cox model and proportional odds model as two special cases.
Bibliography


Bayes T (1763) A letter from the late Reverend Mr. Thomas Bayes, FRS to John Canton, MA and FRS. Philosophical Transactions (1683-1775) 53:269–271


92


