The Discrete Threshold Regression Model

Dissertation

Presented in Partial Fulfillment of the Requirements for the Degree
Doctor of Philosophy in the Graduate School of The Ohio State
University

By

John Stettler, M.S.

Graduate Program in Statistics

The Ohio State University

2015

Dissertation Committee:

Mario Peruggia, Advisor
Steven N. MacEachern
Christopher M. Hans
David Sivakoff
Threshold regression (TR) models have recently become a popular tool for survival analysis. A number of different stochastic processes have been considered as the underlying latent process used to construct threshold regression models, the most popular being Brownian motion with drift and the Ornstein-Uhlenbeck process. TR models have a number of advantages over competing models such as proportional hazards regression. For example, TR models can have proportional or nonproportional hazards. The effects of covariates on the hazard are not required to be multiplicative. Also, because threshold regression models use a latent stochastic process to model an unobserved mechanism that leads individuals to events of interest, the interpretation of the parameters can be very intuitive and insightful.

The TR models that have received the most attention are based on continuous latent models. Continuous models offer much flexibility in terms of drift and the ability to adjust the time scale. However, it is very difficult to incorporate time-varying covariates into a continuous model in a way that is natural and intuitive. This is not the case for TR models based on a latent discrete-time Markov chain. Discrete latent Markovian models have a different level of flexibility compared to continuous models. Because quantities like hazard rates, first hitting times, and survival probabilities are calculated by matrix multiplication, the process can be allowed to change over
time by changing the transition probability matrix in accordance with changes in the
covariates over time. This adds little to the computational burden.

In this dissertation, a discrete threshold regression model is presented that is
based on a discrete-state, discrete-time Markov chain. This model simplifies the
incorporation of time-varying coefficients while retaining a good deal of flexibility
in fitting hazard functions with various shapes, estimating event probabilities when
there are two events that depend on the same underlying process, and estimating
hitting time distributions for two events of interest.
This work is dedicated to my wife Kimmy and daughter Abigail.
Acknowledgments

I would like to express my gratitude to Mario Peruggia for providing advice, guidance, and direction in my research, and also for serving as my advisor. Thanks also to David Sivakoff for introducing me to the RWRE literature, for all of his insights into my research, and for serving on my dissertation committee.

In addition, I would like to thank Steven MacEachern and Christopher Hans for serving on my dissertation committee and for their helpful comments. Thank you also to Christopher Hans, Xinyi Xu, and Michael Pennell for serving on my candidacy exam committee and for providing feedback and guidance for my research.

Thank you to all of the faculty of the Department of Statistics at The Ohio State University for giving me a solid academic foundation in the field of statistics. Finally, I would like to thank my parents and wife for all of the support I received while pursuing my PhD in statistics.

This work was supported in part by the National Science Foundation under award number SES-1024709.
Vita

2009 ................................. B.S. Industrial Management,  
          University of Cincinnati

2012 ................................. M.S. Statistics,  
          The Ohio State University

2009-present  ....................... Graduate Teaching Associate,  
          The Ohio State University.

Fields of Study

Major Field: Statistics
# Table of Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abstract</td>
<td>ii</td>
</tr>
<tr>
<td>Dedication</td>
<td>iv</td>
</tr>
<tr>
<td>Acknowledgments</td>
<td>v</td>
</tr>
<tr>
<td>Vita</td>
<td>vi</td>
</tr>
<tr>
<td>List of Tables</td>
<td>x</td>
</tr>
<tr>
<td>List of Figures</td>
<td>xii</td>
</tr>
<tr>
<td>1. Introduction</td>
<td>1</td>
</tr>
<tr>
<td>1.1 Motivation</td>
<td>1</td>
</tr>
<tr>
<td>1.2 Organization of This Dissertation</td>
<td>5</td>
</tr>
<tr>
<td>2. Survival Analysis</td>
<td>7</td>
</tr>
<tr>
<td>2.1 Basic Quantities</td>
<td>7</td>
</tr>
<tr>
<td>2.2 Parametric Families</td>
<td>11</td>
</tr>
<tr>
<td>2.3 Semiparametric Estimation</td>
<td>14</td>
</tr>
<tr>
<td>3. Threshold Regression</td>
<td>18</td>
</tr>
<tr>
<td>3.1 Process Point of View</td>
<td>20</td>
</tr>
<tr>
<td>3.2 Models with One Absorbing State</td>
<td>26</td>
</tr>
<tr>
<td>3.3 Models with Two Absorbing States</td>
<td>32</td>
</tr>
<tr>
<td>3.4 Time Varying Covariates</td>
<td>38</td>
</tr>
</tbody>
</table>
4. Random Walks in Random Environments ........................................ 46
   4.1 Hitting Times ................................................................. 48
      4.1.1 One boundary ....................................................... 48
      4.1.2 Two boundaries .................................................... 52
      4.1.3 Expected hitting time ............................................. 60
      4.1.4 Hitting time central limit theorem ............................... 62
   4.2 Random Walk in a Dynamic Markovian Random Environment ............ 67
      4.2.1 Drift ................................................................. 68
      4.2.2 Convergence to a Continuous Process ............................ 69
      4.2.3 Boundary hitting times ........................................... 71

5. Discrete Threshold Regression Model ........................................... 74
   5.1 Discrete State, Discrete Time Process .................................. 74
      5.1.1 Initial state ....................................................... 76
      5.1.2 Transition probabilities ......................................... 79
      5.1.3 Hazard functions .................................................. 81
   5.2 Bayesian Formulation ..................................................... 84
      5.2.1 Prior selection ..................................................... 85
      5.2.2 Likelihood ......................................................... 88
   5.3 Extensions ................................................................. 89
      5.3.1 Censoring ............................................................ 89
      5.3.2 Sudden jump to a boundary ....................................... 90
      5.3.3 Time-varying covariates ......................................... 91

6. Simulations and Real Data Applications ....................................... 92
   6.1 Simulated Data ............................................................ 92
   6.2 Heart Attack Data ........................................................ 112
      6.2.1 Comparison with competing models ............................... 119
   6.3 Bond Data ................................................................. 128
      6.3.1 Comparison with other models .................................... 137
      6.3.2 Time-varying covariates ........................................ 143

7. Conclusions and Future Work .................................................. 153
   7.1 Conclusions ............................................................... 153
   7.2 Future Work ............................................................... 154
      7.2.1 Survival distributions with time-varying covariates .......... 154
      7.2.2 Time-varying boundaries ........................................ 155
      7.2.3 Time-delayed covariates ........................................ 157
# List of Tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.1</td>
<td>Parameter values used to create the hazard functions in Figure 5.4</td>
</tr>
<tr>
<td>6.1</td>
<td>Hitting probabilities at each boundary in simple simulation</td>
</tr>
<tr>
<td>6.2</td>
<td>Boundary hitting probability estimates for $\sigma^2 = 0.25$</td>
</tr>
<tr>
<td>6.3</td>
<td>Quantile estimates of the hitting time distributions for $\sigma^2 = 0.25$</td>
</tr>
<tr>
<td>6.4</td>
<td>Deviance information criterion (DIC) and effective number of parameters ($p_D$) for $\sigma^2 = 0.25$</td>
</tr>
<tr>
<td>6.5</td>
<td>Boundary hitting probability estimates for $\sigma^2 = 0.5$</td>
</tr>
<tr>
<td>6.6</td>
<td>Quantile estimates of the hitting time distributions for $\sigma^2 = 0.5$</td>
</tr>
<tr>
<td>6.7</td>
<td>Deviance information criterion (DIC) and effective number of parameters ($p_D$) for $\sigma^2 = 0.5$</td>
</tr>
<tr>
<td>6.8</td>
<td>Boundary hitting probability estimates for $\sigma^2 = 0.9$</td>
</tr>
<tr>
<td>6.9</td>
<td>Quantile estimates of the hitting time distributions for $\sigma^2 = 0.9$</td>
</tr>
<tr>
<td>6.10</td>
<td>Deviance information criterion (DIC) and effective number of parameters ($p_D$) for $\sigma^2 = 0.9$</td>
</tr>
<tr>
<td>6.11</td>
<td>DIC and effective number of parameters ($p_D$) for heart attack data</td>
</tr>
<tr>
<td>6.12</td>
<td>Logistic regression estimates for heart attack data</td>
</tr>
<tr>
<td>6.13</td>
<td>Proportional hazards estimates for heart attack data</td>
</tr>
</tbody>
</table>
6.14 Continuous TR estimates for heart attack data. .................. 122
6.15 Cross validation mean and median absolute deviations. ........ 127
6.16 Cross validation boundary predictions from the DTR model .... 128
6.17 Cross validation boundary predictions from the logistic regression model 128
6.18 DIC and effective number of parameters, $p_D$ for bond data. ...... 133
6.19 Proportional hazards estimates for bond data. ..................... 138
6.20 Continuous TR estimates for bond data. ........................... 140
6.21 Cross validation mean and median absolute deviations. .......... 143
6.22 Cross validation boundary predictions from the DTR model. .... 144
6.23 Cross validation boundary predictions from the logistic regression model 144
6.24 DIC and effective number of parameters, $p_D$, for bond data with time-varying sign indicators. .............................. 145
6.25 Cross validation boundary predictions from the time-varying DTR model 150
6.26 Cross validation mean and median absolute deviations for all models. 151
## List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>4</td>
</tr>
<tr>
<td>2.1</td>
<td>13</td>
</tr>
<tr>
<td>3.1</td>
<td>20</td>
</tr>
<tr>
<td>3.2</td>
<td>23</td>
</tr>
<tr>
<td>3.3</td>
<td>25</td>
</tr>
<tr>
<td>3.4</td>
<td>33</td>
</tr>
<tr>
<td>3.5</td>
<td>40</td>
</tr>
<tr>
<td>3.6</td>
<td>43</td>
</tr>
<tr>
<td>4.1</td>
<td>51</td>
</tr>
<tr>
<td>4.2</td>
<td>53</td>
</tr>
<tr>
<td>4.3</td>
<td>65</td>
</tr>
<tr>
<td>5.1</td>
<td>75</td>
</tr>
<tr>
<td>5.2</td>
<td>76</td>
</tr>
<tr>
<td>5.3</td>
<td>78</td>
</tr>
<tr>
<td>Section</td>
<td>Title</td>
</tr>
<tr>
<td>---------</td>
<td>----------------------------------------------------------------------</td>
</tr>
<tr>
<td>5.4</td>
<td>Examples of hazard functions from the discrete model.</td>
</tr>
<tr>
<td>5.5</td>
<td>The effect of the prior variance of $z\gamma_0$ on the prior for $\xi$.</td>
</tr>
<tr>
<td>5.6</td>
<td>Beta distributions for several values of $\xi$ and $\beta$.</td>
</tr>
<tr>
<td>6.1</td>
<td>Initial state distribution and hazard functions for the simple simulation example with $\sigma^2 = 0.5$; upper boundary hazard in black and lower boundary hazard in red.</td>
</tr>
<tr>
<td>6.2</td>
<td>Example trace plots for 11 states and $\sigma^2 = 0.25$.</td>
</tr>
<tr>
<td>6.3</td>
<td>Posterior distributions of the model parameters for 11 states and $\sigma^2 = 0.25$.</td>
</tr>
<tr>
<td>6.4</td>
<td>Posterior distribution of $\sigma^2$ for models with 3, 11, and 15 states when the true value of $\sigma^2$ is 0.25.</td>
</tr>
<tr>
<td>6.5</td>
<td>Posterior distribution of $\sigma^2$ for models with 3, 11, and 15 states when the true value of $\sigma^2$ is 0.5.</td>
</tr>
<tr>
<td>6.6</td>
<td>Posterior distribution of $\sigma^2$ for models with 3, 11, and 15 states when the true value of $\sigma^2$ is 0.9.</td>
</tr>
<tr>
<td>6.7</td>
<td>Posterior distributions for heart attack data.</td>
</tr>
<tr>
<td>6.8</td>
<td>Posterior predictive distributions for the initial state.</td>
</tr>
<tr>
<td>6.9</td>
<td>Heart attack data survival curves.</td>
</tr>
<tr>
<td>6.10</td>
<td>Bond data sectors.</td>
</tr>
<tr>
<td>6.11</td>
<td>Bond data posteriors for initial state distribution.</td>
</tr>
<tr>
<td>6.12</td>
<td>Posterior initial state distributions for unsigned consumer staples, technology, and financial sectors.</td>
</tr>
<tr>
<td>6.13</td>
<td>Bond data posteriors for the drift.</td>
</tr>
<tr>
<td>6.14</td>
<td>Bond data posteriors for $\sigma^2$ and $\beta$.</td>
</tr>
</tbody>
</table>
6.15 Histogram of credit downgrade times for unsigned and plus sign bonds. 139

6.16 Bond data survival curves. .............................................................. 142

6.17 Bond data posteriors for the initial state distribution with time varying sign indicators. .............................................................. 146

6.18 Histograms of hitting times at the upper and lower boundaries for bonds that have a plus sign at time 0. .................................................. 147

6.19 Bond data posteriors for the drift with time varying sign indicators. . 148

6.20 Bond data posteriors for \( \sigma^2 \) and \( \beta \) with time varying sign indicators. . 149

6.21 Posterior hitting time pmf at lower boundary (downgrade) for one bond. 152

7.1 Histogram of divorce times ............................................................... 156
Chapter 1: Introduction

1.1 Motivation

First hitting time models have recently been considered as an alternative to proportional hazards models in the analysis of survival data (Lee and Whitmore, 2006). Let $T$ be a random variable representing the time at which an event of interest occurs. Let $f(t)$ be the density of $T$ and $S(t) = 1 - F(t)$ be the survival function. Finally, let $h(t) = f(t)/S(t)$ be the hazard function of $T$. (These quantities will be presented more formally in Chapter 2).

A proportional hazards model is a regression model for $T$ in which the hazard is modeled as a function of a vector of covariates $z$ such that

$$h(t|z) = h_0(t) \exp (\beta^\top z),$$

where $\beta$ is a vector of regression coefficients and $h_0(t)$ is a baseline hazard function that does not depend on $z$ (Cox, 1972). The effects of the covariates in this model are multiplicative on the hazard, since

$$\exp (\beta^\top z) = \exp (\beta_1 z_1 + \beta_2 z_2 + \ldots) = \exp (\beta_1 z_1) \cdot \exp (\beta_2 z_2) \cdot \ldots.$$

This model yields a straightforward interpretation of the covariate effects, especially if $z_i$ is an indicator variable. To see this, let $z$ be a single variable that takes values
one and zero. Then,

\[ h(t|z = 1) = h_0(t) \exp(\beta_1), \]

\[ h(t|z = 0) = h_0(t) \exp(0) = h_0(t), \]

and therefore

\[ h(t|z = 1) = \exp(\beta_1) \cdot h(t|z = 0). \]

Suppose that \( \exp(\beta_1) = 1.1 \). Then an individual with \( z = 1 \) has a hazard that is 10% greater than an individual who has \( z = 0 \).

While a straightforward interpretation is desirable, there are cases when covariates do not have a multiplicative effect on the hazard function (Pennell, Whitmore, and Lee, 2009, van den Berg, Arentze, and Timmermans, 2012). In addition, modeling the hazard function directly may put too much emphasis on the event itself at the expense of gaining insight into the process that causes the event (Aalen and Gjessing, 2001). First hitting time models are flexible enough to model proportional and non proportional hazards (Lee and Whitmore, 2010) and they focus on the underlying determinants of survival (Aalen and Gjessing, 2001).

First hitting time models use a latent (unobserved) stochastic process as a model for some unobservable physical process. While the physical process itself may not be observable, it eventually reaches an endpoint that can be observed directly. For example, an engine has a lifetime during which it will run, and over time it will wear out from use. Eventually, it will be worn out to the point that it fails and must be replaced. The driver or operator may not know at any given time how close the engine is to the end of its life, but when the engine fails it will be very clear. In this case, the deterioration of the engine is a physical process that eventually leads
to engine failure. While the driver cannot observe this process directly, the failure of the engine can be observed.

To build a first hitting time model for an engine, first a stochastic process \( \{X_t\} \) must be specified, where, for the purposes of this dissertation, the index \( t \) represents time and \( X_t \) is a real-valued random variable for \( t \geq 0 \). This process could take either discrete or continuous values, and time itself may be measured on either a discrete or a continuous scale. Figure 1.1 shows what a realization of such a stochastic process might look like. Over time the engine deteriorates, which is reflected by the decreasing values of \( X_t \) over time. At time 135 this engine fails, which is indicated by the path of the stochastic process reaching the lower boundary.

The goal in using first hitting time models is to model the unobservable physical process using a stochastic process whose distribution of hitting times at the boundary is similar to the distribution of actual event times. First hitting time models that include regression structures for covariates are called threshold regression (TR) models (Lee and Whitmore, 2006).

TR models are appealing for a number of reasons. They focus attention not only on the event of interest, but also on the underlying process that leads an individual to experience that event (Aalen and Gjessing, 2001). In addition, the parameter estimates from a TR model give insight into the underlying process.

TR models exist that use as the underlying process a discrete-state, continuous-time Markov chain (Aalen, 1995), Brownian motion with drift (Whitmore, 1983), Ornstein-Uhlenbeck processes (Aalen and Gjessing, 2004 and Erich and Pennel, 2015), and others (Lee and Whitmore, 2006). TR models frequently focus on one event of
interest, but there are models for two observable events which are tied to the same underlying process (Horrocks and Thompson, 2004).

Continuous stochastic processes are prevalent in the TR literature. They offer flexibility in a number of ways, as the drift of the process can be any real number and the time scale can be easily accelerated (Lee et al. 2004). However, incorporating time-varying covariates is difficult for a continuous stochastic process. Lee, Whitmore, and Rosner (2008) offer an approach that splits up an individual’s observation period into sub-intervals, and treats each of these sub-intervals as an independent observation. This approach disregards the history of the process at the beginning.

Figure 1.1: Example of a first hitting time model for an engine.
of each sub-interval, fixing a new starting point based on a regression equation. Incorporation of time-varying covariates into the drift term results in first hitting time distributions that are difficult to calculate. Molini et al. (2011) and Urdapilleta (2012) derive these first hitting time distributions for a few special cases, all of which restrict the manner in which the drift can change over time.

An underlying discrete-time Markov chain offers more flexibility in the ability to easily change the drift of the process at any time point. The first hitting time probabilities, and therefore the hazard functions, of a discrete-time Markov chain are calculated by multiplying a transition probability matrix by itself once for each time period, and this matrix can be modified at any point to reflect changes in an individual under study or in their environment. In addition, discrete-time Markov chains enjoy a good deal of flexibility in the hazard functions they can produce and the survival distributions they are able to fit.

In this dissertation, a discrete threshold regression (DTR) model is presented that models event times of two observable events that are tied to a single underlying process and easily incorporates time-varying covariates.

1.2 Organization of This Dissertation

In Chapter 2, fundamental concepts and quantities from survival analysis are reviewed. Chapter 3 contains a review of the TR literature, including details of several models. Chapter 4 introduces some fundamental concepts of Random Walks in Random Environments. This chapter includes derivations of some of the quantities that are needed to implement the DTR model. The DTR model is introduced in Chapter 5. This chapter presents the model itself as well as some details necessary
for a Bayesian implementation. It also includes some extensions of the model. In chapter 6 the DTR model is fit to several datasets. These include a simulation study, a medical data set, and a financial dataset. Several other models are also fit to these data sets and comparisons are made. Chapter 7 provides some discussion and conclusions as well as suggestions for future work.
Chapter 2: Survival Analysis

Survival analysis is concerned with inferences about a random variable $T$, called the survival time, which represents the time at which an event of interest occurs. Some examples of events of interest include development of a disease, death after diagnosis with a disease, breakdown of a mechanical part, etc. There is usually a population of interest, and this population may be homogeneous or heterogeneous. Here homogeneous means that all individuals in the population have survival times that come from a common distribution. If a population is heterogeneous then an individual’s survival time comes from a distribution that is defined in terms of some of the individual’s characteristics (e.g., presence of risk factors).

2.1 Basic Quantities

An individual’s survival time $T$ may be either discrete or continuous. If $T$ is discrete, call its probability mass function (pmf) $p(t_j), j = 1, 2, \ldots$ and if $T$ is continuous call its probability density function (pdf) $f(t)$. The hazard function $h(t)$ is defined as:

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

The hazard function can be interpreted as the instantaneous rate of occurrence of the event of interest for individuals who have not yet experienced the event. Its
analytic expression will differ depending on whether \( T \) has a continuous or a discrete distribution. Further details of these quantities can be found in Klein and Moeschberger (2003).

For a continuous distribution, the hazard function can be written as

\[
 h(t) = \lim_{\Delta t \to 0} \frac{P([t \leq T \leq t + \Delta t] \cap [T > t])}{P(T \geq t)} \frac{1}{\Delta t} = \lim_{\Delta t \to 0} \frac{P(t \leq T \leq t + \Delta t)}{P(T \geq t)} \frac{1}{\Delta t} = \frac{f(t)}{1 - F(t)},
\]

where \( F(t) \) is the cdf of \( T \). For a discrete distribution \( p(t_j), j = 1, 2, \ldots \), the hazard function is zero except at points \( t_j \) where an event occurs with positive probability. At those points,

\[
 h(t_j) = P(T = t_j | T \geq t_j) = \frac{p(t_j)}{1 - F(t_{j-1})}, \quad j = 1, 2, \ldots,
\]

where \( F(t_0) = 0 \).

Other important functions include the cumulative hazard function and the survival function. The cumulative hazard function \( H(t) \) and the survival function \( S(t) = P(T > t) = 1 - F(t) \) are given by:

\[
 H(t) = \int_0^t h(u)du, \quad \text{(Continuous case)},
\]

\[
 H(t) = \sum_{t_j \leq t} h(t_j), \quad \text{(Discrete case)},
\]

and

\[
 S(t) = \int_t^\infty f(u)du, \quad \text{(Continuous case)},
\]

\[
 S(t) = \sum_{t_j > t} p(t_j), \quad \text{(Discrete case)}.
\]
It is often convenient to express the hazard function and cumulative hazard function in terms of the survival function. The following relationships are useful. For a continuous distribution, \( f(t) \),

\[
f(t) = -\frac{d}{dt}S(t) \quad \text{and} \quad h(t) = \frac{f(t)}{S(t)} = -\frac{d}{dt}\log[S(t)],
\]

so that

\[
H(t) = -\log[S(t)],
\]

which implies

\[
S(t) = \exp\left(-H(t)\right) = \exp\left(- \int_0^t h(u)du \right).
\]

For a discrete distribution, \( p(t_j) \),

\[
h(t_j) = \frac{p(t_j)}{S(t_{j-1})} \quad \text{and} \quad p(t_j) = S(t_{j-1}) - S(t_j) \quad \text{which implies} \quad h(t_j) = 1 - \frac{S(t_j)}{S(t_{j-1})},
\]

so that, for \( t_i = \max\{t_j : t_j \leq t\} \),

\[
S(t) = S(t_i) = \frac{S(t_i)}{S(t_{i-1})} \frac{S(t_{i-1})}{S(t_{i-2})} \cdots \frac{S(t_1)}{S(t_0)} = \prod_{t_j \leq t} S(t_j) = \prod_{t_j \leq t} [1 - h(t_j)].
\]

Hazard functions can take several shapes. In particular, the hazard function might be constant, monotone increasing, monotone decreasing, hump shaped (initially increasing and then decreasing), or tub shaped (initially decreasing and then increasing). Outlined below are situations in which each of these shapes might occur.

**Constant:** An important application of the constant hazard function is in connection to Poisson processes. The Poisson process is a continuous-time counting process with stationary, independent increments in which the number of occurrences in a
given time interval follows a Poisson distribution. The waiting time between occurrences of a Poisson process follows an exponential distribution, which has a constant hazard rate. Poisson processes are used to model radioactive decay. For example, Buczyk (2009) compared the number of gamma rays emitted from a radiation source over time to the distribution that would be expected from a Poisson process and found the two to be very similar.

**Monotone Increasing**: An increasing hazard rate is appropriate for applications in which the event of interest is failure due to “wear out.” If the event of interest is time that a mechanical component must be replaced, then the longer the component is used, the more likely it is to wear out.

**Monotone Decreasing**: A decreasing hazard rate is appropriate when there is a high risk of the event occurring initially and the risk drops over time. Lee et al (2010) discuss a study where the lifetime of orthodontic implants was of interest. Some of the implants had issues and needed to be either fixed or redone. In the study the hazard rate was decreasing, meaning that the risk of implant failure was highest in the first weeks after surgery, and decreased afterward. It might be inferred that the high initial risk is an indicator of the level of success of the surgery. As long as the surgery is successful, the risk of failure will be very low; if it is not the implant will fail quickly.

**Hump Shaped**: Motivated by the effect of social activities on traffic and transportation, Van den Berg, Arentze, and Timmerman (2012) modeled the duration of social activities and its dependence on the nature of the activity. They proposed a log-Gaussian model for the event time $T$ such that $\log(t) = X\beta + Z$ where $Z$ has
a normal distribution. The log Gaussian model has a hump shaped hazard rate. In their model, $X$ is a vector of covariates that indicate the nature of the social activity, the relationship between individuals involved, and some socio-demographic data. The hump shaped hazard function indicates that the duration of most social activities is moderate - not very short or very long. Another important illustration of a hump shaped hazard rate is in connection to the first hitting time models that will be discussed in the next chapter. In particular, if the event times are modelled as the first hitting time at a constant boundary of a Weiner process then the first hitting times will have an inverse Gaussian distribution, which has a hump-shaped hazard.

**Tub Shaped:** A tub shaped hazard rate is appropriate when the risk is initially decreasing and then increasing. This type of hazard rate is useful for mortality data since there is a risk of mortality in infants but those who survive infancy tend to live for quite a while longer before risk of death increases again. A tub shaped hazard rate is also appropriate for modelling the lifetime of a randomly sampled mechanical part when some parts in the population are defective. A defective part will fail very quickly, leading to a hazard rate that is high initially, but a part that is not defective will survive much longer before the risk again increases due to the part wearing out.

### 2.2 Parametric Families

Often the distribution of $T$ is assumed to belong to a parametric family. These families are typically right-skewed and have hazard rates that depend on the parameters of the distribution. Common parametric families for survival data include the exponential, Weibull, gamma, log-normal, log-logistic, exponential power, and inverse Gaussian (Klein and Moeschberger, 2003). Below is a brief summary of the hazard
function shapes for each of these families. If the hazard function has a simple analytic form it is included in the summary; in all cases \( h(t) = \frac{f(t)}{S(t)} \) as given above. Figure 1 shows the hazard functions for several choices of parameters.

**Exponential:** Written in terms of its rate \( \lambda > 0 \), the exponential distribution has density \( f(t) = \lambda e^{-\lambda t} \) and constant hazard function \( h(t) = \lambda \) for \( t > 0 \).

**Weibull:** The Weibull distribution with parameters \( \alpha > 0 \) and \( \lambda > 0 \) has density \( f(t) = \alpha \lambda t^{\alpha - 1} \exp(-\lambda t^\alpha) \) and hazard function \( h(t) = \alpha \lambda t^{\alpha - 1} \) for \( t > 0 \). The hazard function can be increasing, decreasing, or constant depending on the choice of \( \alpha \).

**Gamma:** The gamma distribution with parameters \( \alpha > 0 \) and \( \beta > 0 \) has density \( f(t) = \beta^\alpha / \Gamma(\alpha) t^{\alpha - 1} \exp(-\beta t^\alpha) \) for \( t > 0 \). Its hazard function does not have a simple analytic form. It can be increasing, decreasing, or constant depending on the choice of parameters.

**Log Normal:** The log normal distribution with parameters \( \mu \in \mathbb{R} \) and \( \sigma > 0 \) has density \( f(t) = \frac{1}{\sqrt{2\pi} \sigma t} \exp \left( - \frac{(\log t - \mu)^2}{2\sigma^2} \right) \) for \( t > 0 \) and a hazard function that does not have a simple form. The hazard function is hump-shaped.

**Log Logistic:** The log logistic distribution with parameters \( \alpha > 0 \) and \( \lambda > 0 \) has density \( f(t) = \frac{\alpha \lambda^\alpha / [1 + \lambda t^\alpha]^2}{1 + \lambda t^\alpha} \) and hazard function \( h(t) = \frac{\alpha \lambda^\alpha / [1 + \lambda t^\alpha]^2}{1 + \lambda t^\alpha} \) for \( t > 0 \). The hazard function is hump-shaped.

**Exponential Power:** The exponential power distribution with parameters \( \alpha > 0 \) and \( \lambda > 0 \) has hazard function \( h(t) = \alpha \lambda t^{\alpha - 1} e^{(\lambda t)^\alpha} \) for \( t > 0 \). The hazard function is tub-shaped.
**Inverse Gaussian**: The inverse Gaussian distribution with parameters $\mu > 0$ and $\lambda > 0$ has density $f(t) = \sqrt{\frac{\lambda}{2\pi t^3}} \exp\left(\frac{\lambda(x-\mu)^2}{2\mu^2 t}\right)$ for $t > 0$ and a hazard function that does not have a simple form. The hazard function is hump-shaped.

![Hazard function shapes for several parametric families](image)

Figure 2.1: Examples of hazard function shapes for several parametric families.
Regression models exist for some of these parametric models. Klein and Moeschberger (chapter 12) give details for the Weibull, log logistic, log normal, and generalized gamma distributions.

2.3 Semiparametric Estimation

The Cox proportional hazards model (Cox, 1972) is a regression model for the hazard rate. Let z be a vector of covariates, \( \beta \) be a vector of regression coefficients, and \( h_0(t) \) be a baseline hazard function. Then the Cox model for the hazard function is

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

This model has proportional hazards, i.e., the effect of covariates on the hazard rate is multiplicative. For example, let \( z = (1, z) \) and \( \beta = (\beta_0, \beta_1) \). That is, let \( \beta_0 \) be an intercept term and let \( z \) be a covariate with regression coefficient \( \beta_1 \). Then consider the effect on the hazard due to increasing \( z \) by one unit.

\[
h(t|z + 1) = h_0(t)e^{\beta_0 + \beta_1(z + 1)} = h_0(t)e^{\beta_0 + \beta_1 z}e^{\beta_1}
\]

\[
= e^{\beta_1} h(t|z).
\]

As a result, increasing a covariate with \( \beta_1 < 0 \) will decrease the hazard rate and increasing a covariate with \( \beta_1 > 0 \) will increase the hazard rate.

The baseline hazard rate \( h_0(t) \) may or may not be assumed to have a fixed form. In most cases, no specific form is imposed on it and it may have any of the shapes
described above. The regression function \( \exp(\beta^\top z) \) expresses the effect of the parameter values \( z \) on the hazard rate and does not depend on \( T \). As a result the shape of the hazard function for an individual is dependent only on \( h_0(t) \).

The typical method of estimation for \( \beta \) in this model uses the partial likelihood, which doesn’t depend on the baseline hazard \( h_0(t) \). This allows for inferences about the covariate effects without explicit estimation of the shape of the hazard function. If the motivation of the analysis is to determine the effects of the covariates on the hazard rate (and the proportional hazards assumption is valid) then there is no need to estimate the baseline hazard function. If one of the goals of the analysis is estimation of the survival function for an individual with a specific set of covariates, then the baseline hazard must be estimated.

The partial likelihood \( L(\beta) \) is given by the following formula, where \( d \) is the number of observed event times and \( R(t_i) \) is the set of individuals who are at risk of experiencing an event just before time \( t_i \):

\[
L(\beta) = \prod_{i=1}^{d} \frac{\exp(\beta^\top z_i)}{\sum_{j \in R(t_i)} \exp(\beta^\top z_j)}.
\]

The partial likelihood, or its log, can be maximized numerically to find the MLE of \( \beta \).

To estimate the survival function for an individual with a given set of covariates \( z \), Breslow’s method can be used (note at the end of Cox, 1972, as well as Breslow, 1974). This method estimates the baseline cumulative hazard function \( H_0(t) \) with a step function \( \hat{H}_0(t) \) and this estimate can be used to estimate the baseline survival
function \( S_0(t) = \exp(-H_0(t)) \). Breslow’s estimate when there are no ties present is

\[
\hat{H}_0(t) = \sum_{t_i \leq t} \frac{1}{\sum_{j \in R(t_i)} \exp(\hat{\beta}^\top z_j)},
\]

where \( \hat{\beta} \) is the vector that maximizes the partial likelihood \( L(\beta) \).

If some of the covariates in the proportional hazards model change over time, then an adjustment can be made to the partial likelihood (Anderson and Gill, 1982). Imagine that covariates change over time, and that the current measurement is always available. That is, let \( z(t) \) be a continuous function of time. Then the hazard function is

\[
h(t|z(t)) = h_0(t) \exp(\beta^\top z(t)).
\]

To estimate \( \beta \), the partial likelihood \( L(\beta) \) is given by

\[
L(\beta) = \prod_{i=1}^{d} \frac{\exp(\beta^\top z_i(t_i))}{\sum_{j \in R(t_i)} \exp(\beta^\top z_j(t_i))}
\]

and the value \( \hat{\beta} \) that maximizes it can then be used to estimate the baseline cumulative hazard \( H_0(t) \), given by

\[
\hat{H}_0(t) = \sum_{t_i \leq t} \frac{1}{\sum_{j \in R(t_i)} \exp(\hat{\beta}^\top z_j(t_i))}.
\]

The role of the hazard rate in proportional hazards regression is to model the effect of covariates. There is a tradeoff here between the restrictive form of the model (covariate effects on the hazard must be multiplicative) and the ease of interpretation. As an example of interpretation, consider Lee et al. (2010). They used Cox proportional hazards regression to investigate the failure time of orthodontic implants and its dependence on a patient’s age, implantation side (right or left), the clinician
who placed the implant, sex, and an indicator for good oral hygiene. In their study, the only variable that was found to be significant was age. A 95% confidence interval for the multiplicative effect of age was given as (0.868, 0.985) and the point estimate was 0.925. This means that older patients were at a lower risk of implant failure, and the hazard rate decreases by about 7.5% for each additional year of age at the time of surgery.
Chapter 3: Threshold Regression

Threshold regression (TR) uses the first hitting time of a stochastic process at a boundary to model survival data (Lee and Whitmore, 2006). Let $X_t$ be a stochastic process, with the subscript $t$ representing time. Time may be either discrete or continuous. Let $\mathcal{S}$ be the state space of $X_t$, which may be discrete or continuous, and let $\mathcal{B} \subset \mathcal{S}$ be a subset of states. In a TR model the states that belong to $\mathcal{B}$ are absorbing and are called the threshold or boundary states. The states that belong to $\mathcal{S} \setminus \mathcal{B}$ are transient. The initial state $X_0 \notin \mathcal{B}$ may be either fixed or random. Let $\tau = \min\{t : X_t \in \mathcal{B}\}$. Then $\tau$ is called the first hitting time or first passage time of the process $X_t$ in the boundary set $\mathcal{B}$. Frequently, in practice, $\mathcal{B} \equiv b$, i.e. the boundary set is a single state.

In a TR model, the stochastic process $X_t$ is chosen to represent some underlying physical process. The stochastic process is typically latent, so it can not be directly observed. Eventually this latent process will reach the boundary $b$, which triggers the event of interest. While the process itself is unobservable, the event time (and therefore the first hitting time at a boundary) is observable. TR models use regression to model the effect of covariates, and they can accommodate both proportional and non-proportional hazards (Lee and Whitmore, 2010).
Figure 3.1 shows two realizations of a TR process. The horizontal axis shows time, starting from time 0. The vertical axis shows values of the process. There is a single threshold state, represented by the horizontal lower boundary line, and all values of $y$ that are greater than this threshold are transient states. For illustration, one might imagine that time 0 is the time a person is diagnosed with a fatal disease, and the event of interest is death from the disease. During the time between diagnosis and death, the person’s health may improve or decline by small amounts from day to day. Eventually their health will decline to the point of death. The latent process in this case represents a gauge of the person’s health. After diagnosis at time 0 the process moves up and down from day to day tracking the person’s health, and eventually reaches the lower boundary at the time that the person dies.

In figure 3.1 two sample paths are shown. The dark sample path belongs to one Person A, and the lighter sample belongs to another Person B. Person A’s sample path starts close to the lower boundary, indicating very poor health at the time of diagnosis, possibly because of some other co-morbidity at the time of diagnosis. Person B was much better off initially, which is shown by the higher starting point of his process, possibly due to the absence of other medical conditions. However, Person B’s process moves more quickly toward the lower boundary than Person A’s, while Person A’s fluctuates around a constant level for a while before reaching the boundary. This might happen if Person A received a more effective treatment, which kept his health from declining due to the disease of interest without addressing his other health issues. Person B’s quick decline in health could be due to his disease having been diagnosed at a later stage so that it was already very aggressive.
3.1 Process Point of View

Aalen and Gjessing (2001) propose what they call a “process point of view” of models for survival data. They argue that proportional hazards models completely focus on the endpoint of interest, ignoring the process that takes an individual to that endpoint. While the exact nature of this process may be unknown, a flexible enough model may fit survival data well in practice. The use of a latent stochastic process also allows an investigator to separate the effects of different types of covariates, those
that influence or measure an individual’s starting point (in relation to the absorbing state) and those that influence the path of the process itself.

In a TR model the shape of the hazard rate is determined, at least in part, by the distance between an individual’s starting point and the absorbing state. If an individual starts very close to the boundary, then their hazard rate will be decreasing (perhaps after a very brief increase). If they start far from the boundary, then their hazard rate will be increasing. An intermediate starting point will lead to a hazard rate that is increasing for a while and then decreasing. It is important to note that the notion of “close” versus “far” is dependent on the stochastic process that is chosen to model the latent process. Aalen and Gjessing (2001) make the concept of “closeness” more precise using the quasi-stationary distribution. The quasi-stationary distribution is a distribution over the transient state space such that, if $X_0$ has this distribution, then the hazard function will be constant. The hazard rate of a given process will then take its shape based on how the distribution of $X_0$ compares to the quasi-stationary distribution. In particular, if the distribution puts its mass “closer” to the boundary than the quasi-stationary distribution then the process will have a decreasing hazard and if it puts its mass “farther” from the boundary then the process will have an increasing hazard.

In addition to the initial state, the drift plays a role in determining the hazard rate. The drift is the overall tendency of a stochastic process to move upward or downward. A process with a positive drift tends to move upward over time and a process with a negative drift tends to move downward.
To illustrate how a TR model separates different types of effects, four different scenarios will be considered in which the quantity being modeled is the time until death after enrollment in a study for individuals with a fatal disease. At the beginning of the study, some participants receive an experimental treatment and others do not. The group that does not receive a treatment act as a control group. The first two of these scenarios are the same as those considered in Lee and Whitmore (2010), where the authors perform a simulation study and compare of TR models with a proportional hazard model. These authors used Brownian motion as the latent stochastic process, which is a popular choice (Lee and Whitmore (2006), Lee and Whitmore (2010), and Aalen and Gjessing (2001)). Brownian motion will also be assumed here.

**Scenario 1** The treatment of interest is a surgery. This surgery yields an immediate improvement in a patient’s health but does not slow the progression of the disease. As a result, the group that receives this treatment has better health than the control group initially but their health deteriorates at the same rate as the control group.

**Scenario 2** The treatment is non-surgical and slows the progression of the disease. This treatment does not improve an individual’s health, but only slows its decline.

**Scenario 3** The treatment combines the benefits of the first two: it causes an initial improvement in an individual’s health and also slows down the progression of the disease afterward.
**Scenario 4** The treatment is a surgery that is dangerous: it initially causes a decline in an individual’s health from which the individual may or may not recover. However, those individuals who recover will experience an improvement in health and possibly be cured of the disease entirely.

Figure 3.2: Starting points and drifts for 4 scenarios
Figure 3.2 shows how a TR model would fit each of these scenarios. All four treatments are compared to the same control and the threshold is zero for all scenarios. In Scenario 1, the stochastic process for individuals in the treatment group has a higher starting point than the control group, indicating better initial health as a result of their surgery. Since the surgery does not slow down the progression of the disease, the drifts are equal in the treatment and control groups. The mean hitting time for the treatment group is twice that of the control group. For Scenario 2, the starting point of the process is the same for individuals in the treatment and control groups because the treatment does not improve a participant’s health initially. The treatment does, however, slow the progression of the disease and so the health of subjects in the treatment group deteriorates more slowly as implied by the shallower drift in the treatment group. As in Scenario 1, the mean hitting time in the treatment group is twice that of the control group. For Scenario 3, the starting point is higher for individuals in the treatment group due to the initial improvement in health and the drift is shallower due to the slower disease progression. The mean hitting time in the treatment group is four times that of the control group. For Scenario 4, the stochastic process for individuals in the treatment group starts closer to the boundary because of the initial effect of the dangerous surgery. The drift is in a direction away from the boundary because those who recover after the surgery experience an improvement in health afterward. If Brownian motion is used as the stochastic process, then for treatment group 4 there is a positive probability that individuals in the treatment group will never hit the boundary, meaning that they are cured. Because there is a positive probability that the hitting time is not finite, the mean hitting time for this group is undefined.
Figure 3.3 shows the hazard functions for all four scenarios, assuming that $X_t$ is Brownian motion with drift. None of the four scenarios yield proportional hazards. The hazard functions are easy to interpret in terms of the underlying process (health). The surgery in Scenario 1 lowers the hazard immediately, but that benefit disappears after a few years. The treatment in Scenario 2 doesn’t lower the hazard much for the first three months, but by six months the difference is very large and stays large. This
treatment has a longer lasting effect that the surgery in Scenario 1. The treatment in Scenario 3 lowers the hazard immediately and the difference stays large. This treatment combines the effects of the first two treatments. The surgery in Scenario 4 leads to an initially higher hazard because of the immediate decline in health after surgery, but the hazard quickly drops to almost zero because those who recover after surgery experience an improvement in health and a possible cure.

3.2 Models with One Absorbing State

In an early paper on threshold regression, Whitmore (1983) used Brownian motion with drift to model survival data with right censored observations using the following formulation. Let $X_t$ be a Brownian motion process with drift $\delta$ and diffusion coefficient $\nu$, $\nu > 0$. Let the absorbing boundary be fixed at one and let $X_0 = 0$. Letting $Y = \min\{t : X_t \geq 1\}$, the density function for $Y$ is given by

$$f(y) = (2\pi y^3 \nu)^{-1/2} \exp\left(-\frac{(1-\delta y)^2}{2\nu y}\right), \quad y > 0,$$

which is an inverse Gaussian distribution. Parametrizing the drift as $\delta = Z\beta$, where $Z$ is a row vector of covariate values and $\beta$ is a column vector of regression coefficients, Whitmore shows that, assuming no censoring, the maximum likelihood estimates of $\beta$ and $\nu$ are

$$\hat{\beta} = (Z^T Y Z)^{-1} Z^T 1$$

and

$$\hat{\nu} = \frac{1^T Y^{-1} 1 - 1^T Z \hat{\beta}}{n},$$

where $1$ is a column vector of ones. If some observations are censored, then the EM algorithm can be used to find the maximum likelihood estimates.
When survival times are modeled as the first hitting time at an absorbing state of a Brownian motion with drift, the hitting times are a function of three parameters: $X_0$ (or the distance of the starting point from the absorbing boundary), the drift $\delta$, and the diffusion coefficient $\nu$. However, there are only two free parameters. Let the starting state of the process $X_0$ be given. The density function of $Y$ can be written in terms of $\mu = \frac{X_0}{\delta}$ and $\lambda = \frac{X_0^2}{\nu^2}$ in the following way (Doksum and Hoyland, 1992):

$$f(y) = \sqrt{\frac{\lambda}{2\pi y^3}} \exp\left(-\frac{\lambda(y - \mu)^2}{2\mu^2y}\right), \quad y > 0.$$  

Thus, one of the three parameters can be fixed without putting any constraint on the model. Whitmore (1983) chose to constrain $X_0$ to be zero. Lee and Whitmore (2006) chose instead to constrain the diffusion parameter $\nu$ to be one, and used a separate regression function to model $X_0$. This allows different individuals to have different starting points, leading to some of the benefits described in the previous section. Lee and Whitmore (2006) used a log link function for $X_0$; Aalen and Gjessing (2001) discussed using a gamma distribution as it is the quasi-stationary distribution.  

Whitmore, Crowder and Lawless (1998) took a different approach, using an observable process in addition to the latent process of interest. Letting the latent process be denoted by $X_t$ as before, now consider in addition a second process $Y_t$ that can be observed and is believed to be correlated with $X_t$. $Y_t$ is called a marker process. The vector $(X_t, Y_t)$ is assumed to be a bivariate Wiener process with drift $(\mu_x, \mu_y)$ and covariance matrix

$$\Sigma = \begin{pmatrix} \sigma_{xx} & \sigma_{xy} \\ \sigma_{yx} & \sigma_{yy} \end{pmatrix},$$
where $\Sigma$ is assumed to be positive definite. It is further assumed that $(X_0,Y_0)=(0,0)$ and the event of interest happens when $X_t$ first crosses the boundary $a > 0$. If a sample of size $n$ is observed and the event times are denoted by $S_i$, $i = 1, \ldots, n$, and the marker process values at the event times are denoted by $y(s_i)$, $i = 1, \ldots, n$, then the MLEs are

$$\hat{\mu}_y = \frac{\sum_{i=1}^{n} y_i}{\sum_{i=1}^{n} s_i},$$

$$\hat{\mu}_x = \frac{n a}{\sum_{i=1}^{n} s_i},$$

$$\hat{\sigma}_{yy} = \frac{\sum_{i=1}^{n} (y_i - \hat{\mu}_y s_i)^2}{n s_i},$$

$$\hat{\sigma}_{yx} = \hat{\sigma}_{xy} = \frac{\sum_{i=1}^{n} (y_i - \hat{\mu}_y s_i)(a - \hat{\mu}_x s_i)}{n s_i},$$

$$\hat{\sigma}_{xx} = \frac{\sum_{i=1}^{n} (a - \hat{\mu}_x s_i)^2}{n s_i}.$$

Note that these formulas are only valid if an event time is observed for every individual in the sample. If some observations are censored then numerical methods can be used to derive the MLEs. Details are given in Whitmore, Crowder and Lawless (1998).

In the marker process model detailed above, the boundary $a$ must be specified, and can be set at 1 for simplicity. Lee, DeGruttola, and Schoenfeld (1998) extended the above model to incorporate covariates. Their extension used 0 as the absorbing boundary of $X_t$ and required that the starting value $X_0 > 0$ and the other parameters be functions of a vector of covariates, $Z$. Specifically, the initial state of the latent process $(X_0)$, the drifts $\mu_x$ and $\mu_y$, the diffusion parameter of the observable process $\sigma_{yy}$, and the correlation parameter $\rho = \frac{\sigma_{xy}}{\sqrt{\sigma_{xx} \sigma_{yy}}}$ are determined by $Z$ through
appropriate link functions. Lee et al. (1998) recommend the following link functions, where $\beta_x$ and $\beta_y$ denote vectors of regression parameters:

\[
X_0 = \exp(Z\beta_x),
\]
\[
\mu_x = Z\beta_x,
\]
\[
\mu_y = Z\beta_y,
\]
\[
\sigma_{yy} = \exp(Z\beta_{yy}),
\]
and

\[
\rho = \frac{\exp(Z\beta_\rho) - 1}{\exp(Z\beta_\rho) + 1}.
\]

The link functions for $X_0$, $\sigma_{yy}$, and $\rho$ are chosen to enforce appropriate constraints.

In addition to Brownian motion, some authors have considered using the Ornstein-Uhlenbeck (OU) process for threshold regression. Let $W_t$ be a Brownian motion process. Then $X_t$ satisfying the stochastic differential equation

\[
dX_t = (a - bX_t)dt + \sigma dW_t
\]

is an Ornstein-Uhlenbeck process. A key feature of OU processes is mean reversion: this process will tend to stabilize around an equilibrium point $\frac{a}{b}$ (Aalen and Gjessing, 2004). Thus, if a person’s health exhibits some random day to day variation but is relatively stable over time, then the OU process may provide a good latent process for a TR model. Unfortunately, the first hitting time distribution for a general OU process does not have a closed form solution. For the specific case of $a = 0$, $b = 1$, and $\sigma^2 = 2$ explicit formulas do exist (Ricciardi and Sato, 1988). Letting $Y$ denote the
first hitting time at 0 of an OU process after starting at \( x_0 \), the hitting time density and survival function are given by

\[
f(y) = \sqrt{\frac{2}{\pi}} x_0 \frac{e^{2y}}{(e^{2y} - 1)^{3/2}} \exp \left( -\frac{x_0^2}{2(e^{2y} - 1)} \right), \quad y > 0
\]

and

\[
S(y) = 2\Phi \left( \frac{x_0}{\sqrt{e^{2y} - 1}} \right) - 1, \quad y > 0.
\]

See Ricciardi and Sato (1988) for details in the derivation of these quantities.

This distribution can be used to construct a TR model, but the fact that \( x_0 \) is the only free parameter limits the model’s flexibility. Various choices of \( x_0 \) will lead to increasing, decreasing, and hump-shaped hazards (Aalen and Gjessing, 2004, and Erich and Pennell, 2015), but the time scale is fixed because \( \sigma = 2 \) is fixed. For any choice of \( x_0 \) the hazard rate will eventually converge to 1, and will be close to 1 starting around time 2.

Erich and Pennell (2014) addressed this limitation by using a change of time scale. Let \( Y = Y^*/\alpha, \alpha > 0 \), where \( Y^* \) is the first hitting time of the OU process with \( a = 0 \), \( b = 1 \), and \( \sigma = 2 \). Then the pdf of \( Y \) is

\[
f(y) = \sqrt{\frac{2}{\pi}} \alpha x_0 \frac{e^{2\alpha y}}{(e^{2\alpha y} - 1)^{3/2}} \exp \left( -\frac{x_0^2}{2(e^{2\alpha y} - 1)} \right).
\]

Here, a regression model can be specified for \( X_0 \) and \( \alpha \) can also be estimated to improve the fit.

Another stochastic process that has been used as a first hitting time model is a homogeneous Markov chain with discrete state space and continuous time. The hitting time distributions for this process are called phase-type distributions (Aalen, 1995). The states of these models are chosen to represent different intermediate steps
in the process that lead an individual from their starting state to the event of interest. These stages are known, but transitions between them are not observed. These models may move in only one direction from the starting state to the absorbing state, or they may allow the possibility of moving in both directions. They may require movement through the transient states in a specified order, or they may allow for different paths. For example, if an individual is given a medical treatment then their health status may take a different path than that of an individual who did not receive treatment.

A phase-type distribution is defined in the following way. Let \( X_t \) be a discrete space, continuous time Markov chain with \( n \) states. The intensity matrix of a discrete-state continuous-time is denoted by \( A = [\lambda_{ij}]_{i,j=1}^n \), where \( \lambda_{ij} \) is the instantaneous rate of transition from state \( i \) to state \( j \). Let \( A \) be the intensity matrix of \( X_t \) and let \( A^* \) be the part of intensity matrix corresponding to only the transient states. Let \( T \) be the time at which \( X_t \) enters the absorbing state. Then the survival function for \( T \) is

\[
S(t) = \sum_{k=1}^n \sum_{r=0}^{m_k-1} c_{kr} t^r \exp(\rho_k t),
\]

where \( \rho_k \) are the eigenvalues of \( A \) with multiplicities \( m_k \) and the \( c_{kr} \) are constants (Aalen, 1995). Let \( \rho \) be the nonzero eigenvalue with the least absolute value (all nonzero eigenvalues are negative). The hazard rate will converge to \( \rho \) as \( t \to \infty \). Then the (normalized) left eigenvector of \( A^* \) corresponding to \( \rho \) will give the quasistationary distribution.

Phase-type distributions have a number of attractive features. They can produce increasing, decreasing, hump shaped, tub shaped, and constant hazard functions. The quasistationary distribution will exist as long as \( A^* \) is irreducible (Aalen, 1995), and is easy to find. It is possible to include a second absorbing state in the state space.
to which some or all of the transient states can transition. This extra state can be
though of as a “cure” to the event tied to the first absorbing state—transitions into
this state may not be observable, but its inclusion facilitates the estimation of a cure
rate.

Lee and Whitmore (2006) discuss several other survival models based on the first
hitting time of a stochastic process at a boundary. These include Bernoulli processes,
Poisson processes, Gamma processes, Markov chains, and semi-Markov processes.

3.3 Models with Two Absorbing States

In some cases, there are two events of interest that can be tied to the same
underlying physical process. In this case, a first hitting time model with two absorbing
states can be used. Figure 3.4 shows two sample paths from such a model. One
boundary is at state $a$ and the other is at state $-b$, $a, b > 0$. In this model, the latent
process will diffuse up and down until reaching one of the absorbing boundaries. Once
a boundary is reached, the process is absorbed. As a result, it will always be the case
that the hitting time at one boundary is finite and the hitting time at the other
boundary is infinite.

The model illustrated in Figure 3.4 has been used extensively in the context of
two choice tasks (Ratcliff 1978, Blurton et al 2014, Ratcliff et al. 1999, Ratcliff and
McKoon 2008). For example, participants in an experiment may be given a word
and be asked if it belongs to a study set of words that were presented earlier. The
participant will then go through some mental process to determine whether or not
the word belongs to the study set, eventually deciding yes or no. Figure 3.4 displays
two prototypical sample paths of the latent process. One path crosses the upper
boundary first, corresponding to a “yes” decision. The other path crosses the lower boundary first, corresponding to a “no” decision. Only one decision is possible, so only one boundary will be reached by each participant.

If the process is Brownian motion with drift, then the probability $\gamma_-$ of hitting the lower boundary and the probability $\gamma_+$ of hitting the upper boundary are known. Let $\mu$ be the drift, $\sigma$ be the diffusion parameter, $a > 0$ be the upper boundary, and
−b be the lower boundary where \( b > 0 \). Also, let \( X_0 = 0 \). Then \( \gamma_- \) and \( \gamma_+ \) are

\[
\gamma_- = \frac{e^{-\frac{2\mu(a+b)}{\sigma^2}} - e^{-\frac{2\mu b}{\sigma^2}}}{e^{-\frac{2\mu(a+b)}{\sigma^2}} - 1} \quad \text{and} \quad \gamma_+ = \frac{e^{-\frac{2\mu(a+b)}{\sigma^2}} - e^{\frac{2\mu a}{\sigma^2}}}{e^{-\frac{2\mu(a+b)}{\sigma^2}} - 1}.
\]

The hitting time sub-densities are

\[
f_- (t) = \frac{\pi \sigma^2}{(a+b)^2} e^{-\frac{\mu}{\sigma^2}} \sum_{k=1}^{\infty} k \sin \left( \frac{\pi bk}{a+b} \right) \exp \left( -\frac{t}{2} \left( \frac{\mu^2}{\sigma^2} + \frac{\pi^2 k^2 \sigma^2}{(a+b)^2} \right) \right), \quad t > 0,
\]

and

\[
f_+ (t) = \frac{\pi \sigma^2}{(a+b)^2} e^{\frac{\mu}{\sigma^2}} \sum_{k=1}^{\infty} k \sin \left( \frac{\pi ak}{a+b} \right) \exp \left( -\frac{t}{2} \left( \frac{\mu^2}{\sigma^2} + \frac{\pi^2 k^2 \sigma^2}{(a+b)^2} \right) \right), \quad t > 0,
\]

where \( \int_0^{\infty} f_- (t) dt = \gamma_- \) and \( \int_0^{\infty} f_+ (t) dt = \gamma_+ \), i.e. these densities are unnormalized. The expressions for \( f_- (t) \) and \( f_+ (t) \) include an infinite sum which can not be computed exactly. In practice, the sum is truncated (the summands are on the order of \( k \times e^{-k^2} \), so the sum converges quickly). These formulas can be found in Ratcliff (1978).

Horrocks and Thompson (2004) used Brownian motion with two absorbing barriers to model the length of stay at a hospital. The upper boundary represented discharge from the hospital (healthy) and the lower boundary represented death in the hospital. The latent process represented a patient’s level of health. After admission to the hospital, a patient’s health will improve or decline over time until they are well enough to leave the hospital or until they die in the hospital. To include the effect of covariates, they first fixed the diffusion parameter at \( \sigma = 1 \) and required the same drift \( \mu \) for all individuals. The boundaries for individual \( i \) were given as

\[
a_i = \beta_a^T Z_i,
\]

and

\[
b_i = \beta_b^T Z_i.
\]

34
where $Z$ is a column vector of covariates and $\beta_a$ and $\beta_b$ are column vectors of regression coefficients. For estimation, they constructed the likelihood as

$$L(\mu, \sigma, \beta_a, \beta_b) = \prod_{i \in U} f_+(t_i) \times \prod_{i \in L} f_-(t_i) \times \prod_{i \notin U, j \notin L} (1 - F_-(t_i) - F_+(t_i)),$$

where $U$ is the set of individuals who were discharged from the hospital and $L$ is the set of individuals who died in the hospital.

Balka, Desmond, and McNicholas (2009) used Brownian motion with two absorbing barriers to model the time to infection for burn patients. The upper boundary represented infection and the lower boundary represented an immunity threshold. Immunity is not an observable event, so the lower boundary was meant to “catch” the individuals who did not experience infection. The drift was determined by a regression function with the identity link and the lower boundary was determined by a regression function with an exponential link.

Diederich (1995, 1997, 2003) proposed a model using the Ornstein-Uhlenbeck process with two absorbing boundaries. The hitting time distributions aren’t available in closed form for the OU process with two boundaries, so a discrete approximation was used. Let $X_t$ be an OU process defined by the stochastic differential equation

$$dX_t = (\delta - \gamma X_t)dt + \sigma W_t.$$

To approximate this process, Diederich used a discrete-state, discrete-time Markov chain with transition probabilities $p_{i,j} = P(X_{t+1} = j|x_t = i)$, where

$$p_{i,j} = \begin{cases} 
\frac{1}{2} \left(1 - \frac{(\delta - \gamma i)\Delta}{\sigma} \sqrt{\tau}\right) & \text{for } j = i - 1, \\
\frac{1}{2} \left(1 + \frac{(\delta - \gamma i)\Delta}{\sigma} \sqrt{\tau}\right) & \text{for } j = i + 1, \\
1 - p_{i,i-1} - p_{i,i+1} & \text{for } j = i, \\
0 & \text{otherwise,}
\end{cases}$$
where $\Delta$ is the state space increment, $\tau$ is the time increment, and $\tau$ is chosen sufficiently small to make
\[
\left| \frac{\delta - \gamma i \Delta}{\sigma} \sqrt{\tau} \right| < 1.
\]

The specification of $\Delta$ and $\tau$ here allow the discrete approximation to converge to the OU process when $\Delta$ converges to zero and $\tau = \frac{\Delta^2}{\sigma^2}$. Details can be found in Diederich (1997). The boundaries are chosen to be $\pm l\Delta$, for some positive integer $l$. Suppose the number of states is $m$ ($= 2l + 1$). Let $P$ be a transition probability matrix with the above transition probabilities, but with the last row and column moved to be the second row and column as shown below. Then let $R$ and $Q$ be defined by

\[
P = \begin{pmatrix}
 I & 0 \\
 R & Q
\end{pmatrix}
\]

\[
= \begin{pmatrix}
 1 & 0 & 0 & \ldots & 0 & 0 \\
 0 & 1 & 0 & \ldots & 0 & 0 \\
 p_{21} & 0 & p_{22} & p_{23} & \ldots & 0 & 0 \\
 0 & 0 & p_{32} & p_{33} & \ldots & 0 & 0 \\
 0 & 0 & 0 & p_{43} & \ldots & 0 & 0 \\
 \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
 0 & 0 & 0 & 0 & \ldots & p_{m-3,m-2} & 0 \\
 0 & 0 & 0 & 0 & \ldots & p_{m-2,m-2} & p_{m-2,m-1} \\
 0 & p_{m-1,m} & 0 & 0 & \ldots & p_{m-1,m-2} & p_{m-1,m-1}
\end{pmatrix}.
\]

Using this notation, $R$ contains the transition probabilities from transient to absorbing states and $Q$ is the transition probability matrix for the transient states. Let $R_b$ be the first column of $R$ and $R_a$ be the second column of $R$. Also, let the probability distribution for $X_0$ be given by a vector $z$. The probability of hitting the upper and lower boundaries and the distribution of the hitting times at each boundary can be derived in terms of $z$, $Q$, and $R$. Let $\tau_a$ be the first hitting time at the upper boundary and $\tau_b$ be the first hitting time at the lower boundary. Then the probability
of hitting boundaries $a (= l\Delta)$ and $b (= -l\Delta)$ are

$$P(\tau_a < \tau_b) = z^\top (I - Q)^{-1} R_a,$$

and

$$P(\tau_b < \tau_a) = z^\top (I - Q)^{-1} R_b.$$

The hitting time probabilities for the two boundaries are

$$P(\tau_a = n) = z^\top Q^{n-1} R_a,$$

and

$$P(\tau_b = n) = z^\top Q^{n-1} R_b.$$

An extension of this discrete approximation of the OU process allows the process to move through several “attributes” (Diederich, 1997). If the process represents a person’s thought process during a decision task, then the person may consider several aspects (attributes) of the task before reaching a decision. While considering the first attribute, the process has parameters $\delta_1$, $\gamma_1$, and $\sigma_1$. If at some time $n_1$ the person begins to consider a second attribute and they have not yet made a decision (reached a boundary), then starting at time $n_1$ the process has parameters $\delta_2$, $\gamma_2$, and $\sigma_2$. After thinking about the second attribute for a while, the person may start thinking about a third attribute (or about the first attribute again). So there will be several versions of the matrices $R$ and $Q$ above, called $R_1$ and $Q_1$, $R_2$ and $Q_2$, etc.

Suppose there are 2 attributes and a person thinks about the first attribute for $n_1$ time increments and then thinks about attribute 2 until reaching a decision. Then the probability of choosing $a$ at time $n$, where $n > n_1$, is

$$z^\top Q_1^{n_1} Q_2^{n-n_1} R_{a,1}$$
where $R_{a,1}$ is the second column of $R_1$.

### 3.4 Time Varying Covariates

The models discussed in the previous sections used covariates to influence the initial state and path of a stochastic process. The covariates that are used to determine the starting state $X_0$ of the process will generally be fixed, but the covariates that influence the path (drift) of the process may change over time. These time-varying covariates can be incorporated into threshold regression models.

An important simplification of the likelihood function is possible when the latent process in a TR model is a Markov process. In that case, let $t_0 < t_1 < \cdots < t_k$ be distinct times and let $P(x|\theta)$ be a pdf (or pmf), where $\theta$ is a vector of parameters. Then
\[
P(x_{t_i}|x_{t_i-1}, x_{t_i-2}, \ldots, x_{t_0}, \theta) = P(x_{t_i}|x_{t_i-1}, \theta)
\]
and
\[
P(x_{t_0}, x_{t_1}, \ldots, x_{t_k}|\theta) = P(x_{t_0}|\theta) \prod_{i=1}^{k} P(x_{t_i}|x_{t_{i-1}}, \theta).
\]
An important consequence of this is that the process can be treated as $k$ separate processes, each with a random starting point that is determined by the endpoint of the previous process. See Lee, Whitmore, and Rosner (2010) for more details on this decomposition.

Several authors have used the above decomposition in modeling the effects of time varying covariates. Lee, Whitmore, and Rosner (2010) presented the “Markov TR” model. The Markov TR model uses Brownian motion to model the effect of age and smoking on time to lung cancer for a cohort of 121,700 nurses who completed
a questionnaire roughly every two years. Smoking history was measured in pack years, the total number of years a person has smoked multiplied by the number of packs of cigarettes that were smoked per day in those years. The data were interval censored, since a cancer diagnosis could be observed to be present on one questionnaire but not on any previous questionnaires. The starting point of each interval was estimated using a log link function each time a participant in the study completed a questionnaire. To construct the sample likelihood function, suppose an individual was diagnosed with lung cancer between questionnaire \( j - 1 \) and questionnaire \( j \). Let this person’s covariate values (age and pack years) at questionnaire \( i \) be denoted by \( z_i \). Let \( \beta \) be a vector of regression parameters for the log link that determines the initial state in each time period, let \( \mu \) be the drift of the Brownian motion process, and let \( \sigma \), the diffusion parameter, be set equal to one. Then, letting \( F(t|\theta) \) be the cdf for time to lung cancer and \( S(t|\theta) \) be the survival function, \( \theta = (\mu, \beta) \), their contribution to the likelihood would be

\[
F(t_j - t_{j-1}|z_{j-1}, \theta) \times \prod_{i=1}^{j-1} S(t_i - t_{i-1}|z_{i-1}, \theta)
\]

where \( z_0 \) comes from a baseline questionnaire that was completed at the beginning of the study.

Figure 3.5 shows an example of a sample path of the latent Brownian motion process in the Markov TR model. The sample path starts at \( X_{t_0} \), where \( t_0 \) is the time of the baseline questionnaire and \( \log(X_{t_0}) = z_{t_0}\beta \). Here, \( z_{t_0} \) is a row vector of covariates and \( \beta \) is a column vector of regression coefficients. The sample path does not reach the absorbing lower boundary before time \( t_1 \), the time of the first follow-up questionnaire, so at that time the value of the latent process is set equal to \( X_{t_1} \), where
log(\(X_{t_1}\)) = \(z_{t_1}\beta\). The sample path also does not reach the absorbing boundary before times \(t_2\) and \(t_3\), the times of the second and third follow-up questionnaires, so at those times the value of the latent process is set equal to \(X_{t_2}\) and \(X_{t_3}\), respectively, where log(\(X_{t_2}\)) = \(z_{t_2}\beta\) and log(\(X_{t_3}\)) = \(z_{t_3}\beta\). Finally, at a time between \(t_3\) and \(t_4\) the latent process reaches the lower boundary and is absorbed. The likelihood contribution for this sample path is

\[
S(t_1 - t_0|\mu, z_{t_0}, \beta) \cdot S(t_2 - t_1|\mu, z_{t_1}, \beta) \cdot S(t_3 - t_2|\mu, z_{t_2}, \beta) \cdot F(t_4 - t_3|\mu, z_{t_3}, \beta),
\]

where \(S(\cdot)\) and \(F(\cdot)\) are the survival function and cdf, respectively, and the cdf is used rather than the pdf \(f(\cdot)\) because the data in the lung cancer study were interval censored.

Figure 3.5: Example a sample path from the Markov TR model.
In the Markov TR model’s treatment of time varying-covariates, the sample path jumps at the time of each follow-up questionnaire to a new starting point which is determined by new covariate values. In effect, this model treats the latent process values in the time intervals \((t_0, t_1]\), \((t_1, t_2]\), \((t_2, t_3]\), and \((t_3, t_4]\) as independent (conditional on covariate values). While the Markov assumption justifies treating the increments of the latent process as independent, the Markov TR model goes a step further in treating the values of the process across time intervals as independent (again, conditional on covariate values). This is in conflict with the “process point of view” presented by Aalen and Gjessing (2001), which regards the latent process as a representation of an underlying physical process. The underlying process in the lung cancer study presented in Lee, Whitmore, and Rosner (2010) is development of lung cancer. It is conceivable that this development may move in jumps, but those jumps should correspond to events that increase the risk of cancer, not to times when an individual has filled out a questionnaire.

Figure 3.6 shows an illustration of survival curves for the Markov TR model. This illustration is not related to the lung cancer study that is discussed in Lee, Whitmore, and Rosner (2010), but is meant to underscore the effect of follow-up frequency on the Markov TR model. It is assumed in this illustration that the covariates are measured at several follow-up times, and that it is possible for an individual to complete each follow-up questionnaire with the same answers. The survival curves presented here are for an individual who completes several follow-up questionnaires but whose covariate values do not change. In the upper left panel of Figure 3.6 is shown a survival curve based on a starting point \(X_{t_0} = 2\) and drift \(\mu = 0\). This survival curve corresponds to an individual who completes the baseline questionnaire at time \(t_0\) but
does not complete any follow-up questionnaires. The upper right panel again shows this survival curve, along with the survival curve for an individual who completes a follow-up questionnaire weekly (shown in red). The difference between the two comes from the fact that the Markov TR model fixes the value of the latent process to $X_{t_i} = 2$, $i = 1, 2, \ldots$, at each weekly follow-up. Because the value of the latent process is reset to equal 2 each week, it is very unlikely that a sample path will ever drift very far away from the lower absorbing boundary. The lower left panel shows the same two survival curves as the upper right panel as well as the survival curve for an individual who completes a follow-up questionnaire every two weeks (shown in green). The lower right panel shows the survival curves for individuals who complete a follow-up questionnaire at frequencies of one to 50 weeks. Each of these curves is equal to the black curve until the time of the first follow-up, where the black curve corresponds to an individual who completes only a baseline questionnaire and does not complete any follow-up questionnaires.

Two issues with the Markov TR model are apparent from Figure 3.6. The first is that the survival curve depends on the frequency of follow-up, even for an individual whose covariates are not changing over time. For example, a non-smoker will indicate a smoking history of zero pack years on each questionnaire. It does not seem reasonable to predict a higher risk of lung cancer for a non-smoker who completes a follow-up questionnaire every two years than a non-smoker who completes a follow-up questionnaire every five years. The second issue is that completing any number of follow-up questionnaires has the effect of decreasing the survival function $S(t)$ substantially, especially for large $t$. Both of these issues can be remedied by allowing the drift, $\mu$, of the latent process to change in response to changing covariates rather than
Figure 3.6: Survival curves for Markov TR model with differing frequencies of follow-up.

the starting point of the latent process in each time interval. The discrete threshold regression (DTR) model presented in Chapter 5 addresses these issues.

Lee, Chang, and Whitmore (2008) used Brownian motion to model time to disease progression and time to death, where disease progression was an observable event. Patients were given one of two treatments at the beginning of the study and stayed on that treatment until they experienced disease progression. If a patient did not experience disease progression during the study, then they stayed on the primary treatment. If they did experience disease progression then they were switched to the
other treatment. In this study, it was believed that the deterioration of a patient’s health would happen at a different rate after disease progression than before. This effect was accounted for by using a composite time scale $r$, defined by

$$r = \alpha t_1 + t_2,$$

where $t_1$ is the time on the primary treatment and $t_2$ is the time on the secondary treatment. Thus, the event times used in fitting the TR model were equal to the time on primary treatment plus $\alpha$ times the time on secondary treatment. If $\alpha > 1$, then using this composite time scale $r$ has the effect of accelerating the deterioration of health after disease progression compared to deterioration before disease progression.

Blurton, Greenlee, and Gondan (2014) used Brownian motion with two absorbing barriers to model response times in a two choice decision task. Participants in their study were exposed to a stimulus and asked to decide between two alternatives. If a response was not given by a certain time, then another stimulus was presented in addition to the first. This second stimulus changed the drift of the process. Numerical integration was used to calculate quantities of interest. This integration becomes very difficult if drift changes at several time points are allowed. The DTR model presented in Chapter 5 allows the drift to change several times without the burden of numerical integration over several variables.

The examples presented in this chapter were chosen to give an overview of the first hitting time models that appear in the literature. The threshold regression literature focuses mostly on models with one boundary, and typically include some sort of regression structure to model the effect of covariates. The psychology literature
focuses on models with two boundaries which do not include any covariates. Latent processes with continuous state space are generally preferred, with the exception of Aalen’s state space models. In the models with two boundaries in the threshold regression literature, the second boundary is often an unobserved “cure” event that absorbs some individuals. In these models it is not known whether an individual is cured or simply censored. Horrocks and Thompson (2004) use two boundaries that are both for observable events. However, they use Brownian motion which makes it difficult to incorporate the effects of time-varying covariates. Blurton, Greenlee, and Gondan (2014) did allow one change of drift in a two boundary model with Brownian motion, but did not incorporate any covariates and used numerical methods to fit their model. Numerical methods become difficult when the number of parameters increases and multiple time changes are permitted because the density and survival function calculations to derive the likelihood function require the evaluation of numerical integrals. A two boundary TR model that easily incorporates time-varying covariates is needed.
In the first hitting time models described in Chapter 3, event times are modeled by the first hitting time of a latent stochastic process at a boundary. Properties of this latent process were permitted to vary between individuals. For example, the latent process could have different drifts for different individuals (perhaps depending on some covariates), and in some cases the drift might change over time. This, as explained below, leads to a connection between threshold regression and random walks in random environments (RWRE).

Zeitouni (2004) provides an overview of the main results related to random walks in random environments. In the RWRE literature, an “environment” is a set of transition probabilities for a random walk. For example, suppose $X_t$ is a random walk on the integers. Then, for $t \geq 0$,

$$X_{t+1} = \begin{cases} X_t + 1, & \text{with probability } p_+, \\ X_t, & \text{with probability } p_0, \\ X_t - 1, & \text{with probability } p_-, \end{cases}$$

for some probabilities $p_+, p_0,$ and $p_-$ such that $p_+ + p_0 + p_- = 1$. If these probabilities are chosen at random according to some distribution then this is a random environment. A random walk that takes place according to this randomly chosen set of transition probabilities is a random walk in a random environment.
To make the connection with threshold regression clear, imagine that a researcher is modeling the time of a second heart attack after a first heart attack, so that the time of the first heart attack is the initial time \( t=0 \) and the amount of time that passes before the second heart attack is the quantity of interest. Individuals who do not have a second heart attack will have a censored event time. Diabetes is known to increase the risk of heart attack, and an indicator for diabetes can be incorporated into the model as a covariate. To model the time to a second heart attack, let \( X_t \) be a discrete state discrete time Markov chain with \( X_0 = a \) with probability 1 for some positive integer \( a \), and let 0 be an absorbing state. Let \( \tau \) be the time of the heart attack, where \( \tau = \min_{t=1,2,...}\{t : X_t = 0\} \). Since diabetes increases the risk of heart attack, individuals with diabetes should tend to reach state 0 sooner. In other words, the quantity \( p_- \) should be greater for subjects with diabetes than for subjects without diabetes. Let \( (p_0^+, p_0^0, p_0^-) \) be the transition probabilities for subjects without diabetes and \( (p_D^+, p_D^0, p_D^-) \) be the transition probabilities for subjects with diabetes. Then there are two environments: “diabetes” and “no diabetes.” If a person who has had a first heart attack is chosen randomly and the time until their second heart attack is observed, then this is like choosing an environment at random and observing the first passage time at state 0 of a random walk in that random environment.

In some cases, a covariate will change over time. In the survival analysis literature, this is referred to as a time-varying covariate. In the RWRE literature, this is referred to as a random walk in a *dynamic* random environment (Bandyopadhyay and Zeitouni, 2006). To construct this type of RWRE, a new environment is randomly chosen at each time point and a step is taken in the chosen environment.
The probabilities of various environments at each time might depend on the current environment, making the sequence of environments a Markov chain.

In the remainder of this chapter, interesting quantities and properties of RWREs are explored. In the next chapter, this exploration continues with the aim of using a random walk in a random environment as a first hitting time model.

4.1 Hitting Times

For this section, consider a discrete-state, discrete-time random walk \( \{X_t\} \) that starts at state \( X_0 = 0 \) with probability 1 and walks until reaching one of two boundaries \( a \) or \( -b \) for the first time, where \( a, b > 0 \). Let \( \xi_t \) be the increment process, so that \( \xi_t = X_t - X_{t-1} \). Each increment in the walk is of size 1, 0, or \(-1\) so that the process may move one state in either direction or stay in the same state at each time increment. The quantities \( p_+, p_0, \) and \( p_- \) will be as described above. Both the one-boundary and two-boundary cases will be considered.

4.1.1 One boundary

Consider the case where \( b = \infty \), so that the random walk effectively has only one boundary. The number of paths that start at state 0 at time 0 and are in state \( a \) at time \( t \) can be counted in the following way. Since the process starts at 0 there can be no hits before time \( t = a \). For a given time \( t \geq a \), let \( n_1 \) be the number of upward transitions that have occurred up to time \( t \), \( n_2 \) be the number of times the walker has stayed in the same state for two consecutive time periods up to time \( t \), and \( n_3 \) be the number of downward transitions up to time \( t \). It must be true that \( n_1 + n_2 + n_3 = t \),
and that $n_1 - n_3 = a$. As a result, for simplicity, the counts of all transition types can be written in terms of $n_1$ as follows.

- $n_1 = n_1$,
- $n_3 = n_1 - a$,
- $n_2 = t - n_1 - n_3 = t - 2n_1 + a$.

Let the pair $(m, x_m)$ indicate the event that $X_t$ is in state $x_m$ at time $m$. The total number of paths that lead from $(m, x_m) = (0, 0)$ to $(t, a)$ with $n_1$ upward transitions can be found using the multinomial coefficient

$$
\frac{t!}{n_1!n_2!n_3!} = \frac{t!}{n_1!(n_1-a)!(t-2n_1+a)!},
$$

for $a \leq n_1 \leq \lfloor \frac{t+a}{2} \rfloor$, where $\lfloor \cdot \rfloor$ is the floor function. The second inequality rules out values of $n_1$ for which $n_1 - n_3$ cannot be equal to $a$. Therefore, the total number of paths that lead from $(0, 0)$ to $(t, a)$ is

$$
\sum_{n=a}^{\lfloor \frac{t+a}{2} \rfloor} \frac{t!}{n!(n-a)!(t-2n+a)!},
$$

and the probability of moving from $(0, 0)$ to $(t, a)$ is

$$
P(t, a) = \sum_{n=a}^{\lfloor \frac{t+a}{2} \rfloor} \frac{t!}{n!(n-a)!(t-2n+a)!} p^n p_{-a}^{n-a} p_0^{t-2n+a}.
$$

Some of the paths that are counted by the above expressions are not hitting the boundary $a$ for the first time at $t$, so the formula above is not the first hitting time probability for time $t$. For $X_t = a, t \geq a$, there are three possible cases:

1. $X_{t-1} = a + 1$,
2. $X_{t-1} = a$, 

3. $X_{t-1} = a - 1$.

Clearly paths from Cases 1 and 2 do not reach state $a$ for the first time at $t$. The probability of all sample paths that fall into case 1 is equal to the probability of moving from $(m, x_m) = (0, 0)$ to $(t - 1, a + 1)$ and then transitioning down to $a$ at time $t$, or $p_\downarrow \cdot P(t - 1, a + 1)$. Similarly, the probability of all sample paths that fall into Case 2 is $p_0 \cdot P(t - 1, a)$.

For Case 3 note that some of the paths will be hitting the boundary for the first time and others will not. The reflection principle (Karlin and Taylor, 1975) can be used to find out how many of these paths have hit boundary $a$ before time $t$. The reflection principle states that for any path in Case 3 that reached state $a$ before $t$, there is a path that was identical up to the first hitting time at $a$ and then followed a path that was a reflection about $a$ of one of the paths in Case 3. Figure 4.1 illustrates a process that hits boundary $a$ at time 19 and is in state $a - 1$ at time 18, but has already hit boundary $a$ at time 6. After the first hit at time 6, the reflection of the process is shown in gray. Note that both processes have the same number of upward transitions (7), downward transitions (4), and transitions into the same state (8). The reflected path shown in gray will therefore have the same probability as the path shown in black.

By the reflection principle, the number of paths in Case 3 that have hit $a$ before is equal to the number of paths that move from $(m, x_m) = (0, 0)$ to $(t - 1, a + 1)$, which is given by

$$\sum_{n=a+1}^{[a+c]} \frac{(t - 1)!}{n!(n - a - 1)!(t - 2n + a)!}.$$
Figure 4.1: Path that hits $a$ at times 6 and 19 and its reflection about $a$.

Let $\tau_a = \min_{t \geq 0} \{t : X_t = a\}$ be the first hitting time at boundary state $a$. The probability of hitting $a$ for the first time at $t$ is

$$P(\tau_a = t | b = \infty) = p_+ \cdot P(t-1, a-1) - p_- \cdot P(t-1, a+1)$$

$$= p_+ \cdot \sum_{n=a-1}^{\lfloor \frac{t+a-2}{2} \rfloor} \frac{(t-1)!}{n!(n-a+1)!(t-2n+a-2)!} p_+^{n-1} p_-^{n-a-1} p_0^{t-2n+a-2}$$

$$- p_- \cdot \sum_{n=a+1}^{\lfloor \frac{t+a}{2} \rfloor} \frac{(t-1)!}{n!(n-a-1)!(t-2n+a)!} p_+^{n} p_-^{n-a-1} p_0^{t-2n+a}.$$  

After some rearrangement, this expression can be written more succinctly as

$$P(\tau_a = t | b = \infty) = \frac{a}{t} \sum_{n=0}^{m_a(t)} \frac{t!}{(a+n)!n!(t-a-2n)!} p_+^{a+n} p_-^{n} p_0^{t-a-2n},$$

where $m_a(t) = \lfloor \frac{t-a}{2} \rfloor$.  

51
The first hitting time probability at time $t$ for boundary $-b$ can be found similarly in the case that $a = \infty$ to be

$$P(\tau_{-b} = t | a = \infty) = p_{-} \cdot P(t - 1, -(b - 1)) - p_{+} \cdot P(t - 1, -(b + 1))$$

$$= \frac{b}{t} \sum_{n=0}^{\infty} \frac{t!}{n!(b + n)!(t - b - 2n)!} p_{+}^{n} p_{-}^{n} p_{0}^{l-b-2n},$$

where $m_{b}(t) = \lfloor \frac{t-b}{2} \rfloor$.

### 4.1.2 Two boundaries

The first hitting time probabilities found in the previous section will not be correct when both boundaries are finite because, for example, $P(\tau_{a} = t | b = \infty)$ includes the probability of paths that reach any given finite state $-b_{0}$, $0 < b_{0} < \infty$, before hitting $a$. In order to find the first hitting time probabilities that have been corrected for the fact that the process can be absorbed at either boundary ($a$ or $b_{0}$), there are two cases to consider.

1. The walker reaches $a$ for the first time at $t$ without reaching $-b_{0}$ first. This will always be the case for paths that hit $a$ before $t = 2b_{0} + a$.

2. The walker reaches $a$ for the first time at $t$ after already reaching $-b_{0}$.

These first hitting time probabilities are presented for the upper boundary $a$; the first hitting time probabilities for the lower boundary can be found similarly. Since both boundaries are absorbing, the first hitting time probabilities at $a$ should not include the probability of paths in Case 2. One way to find the correct first hitting time probabilities is to first calculate $P(\tau_{a} = t | b = \infty)$ and then subtract out the probability of the paths in Case 2.
Figure 4.2: Depictions of paths that require extra terms in $P(\tau_a = t|b = b_0)$.

Figure 4.2 shows four sample paths of $X_t$ that fall into Case 2 and must be addressed in the calculation of $P(\tau_a = t|b = b_0)$. In the upper left pane, a sample path shown in black starts at $X_0 = 0$ and then drops below $-b_0$ before reaching boundary $a$ for the first time. The quantity $P(\tau_a = t|b = \infty)$ includes the probability of this path, and it must be subtracted out. The reflection principle is used here. The gray path shown in the upper left pane of Figure 4.2 is the same as the black path up to the first hitting time at $-b_0$ and thereafter is the reflection about $-b_0$ of the black path. The gray path first reaches state $-a - 2b_0$ at the same time the
black path reaches $a$. Every path that reaches $-b_0$ before hitting $a$ will have such a reflected path, and, as a result, the number of paths that reach state $-a-2b_0$ for the first time at $t$ is the same as the number of paths that reach boundary $a$ for the first time at $t$ after first reaching state $-b_0$. Due to symmetry, the number of paths that first reach $a+2b_0$ at time $t$ is also the same as the number of paths that first reach boundary $a$ for the first time after first reaching state $-b_0$.

While the number of paths that reach $a+2b_0$ is the same as the number of paths that hit $-b_0$ and then reach $a$ for the first time at $t$, they do not have the same probability. The paths that reach $a$ at time $t$ have had $a$ more upward transitions than downward transitions. The paths that reach $a+2b_0$ have had $a+2b_0$ more upward transitions than downward transitions. The probability of a path that first reaches $a+2b_0$ at time $t$ is therefore equal to $(\frac{p_+}{p_-})^{b_0}$ times the probability of the corresponding path that reaches $a$. As a result,

$$P(\tau_a = t \mid b = \infty) - P(\tau_{a+2b_0} = t \mid b = \infty) \cdot \left(\frac{p_-}{p_+}\right)^{b_0},$$

adjusts for paths like that shown in the top left pane of Figure 4.2.

The upper right pane of Figure 4.2 shows another sample path in black. This path reaches state $a$ before reaching state $-b_0$, and then enters state $a$ again at time $t$. This path is not included in $P(\tau_a = t \mid b = \infty)$ because it does not hit $a$ for the first time at $t$, but its reflection is included in $P(\tau_{a+2b_0} = t \mid b = \infty)$ because it hits $-a-2b_0$ for the first time at $t$. As a result, the expression

$$P(\tau_a = t \mid b = \infty) - P(\tau_{a+2b_0} = t \mid b = \infty) \cdot \left(\frac{p_-}{p_+}\right)^{b_0},$$

should be corrected by adding this path back in. To do this, the reflection principle can again be used. The lower left pane in Figure 4.2 shows the same black sample
path as the upper right pane as well as a gray path that is the same as the black path up to the time when it first reaches $-b_0$ and thereafter is the reflection of the black path about $-b_0$. There is also a blue path, which is the same as the black path up to the time when it first reaches $a$, then is the reflection about $a$ of the black path until it reaches $-b_0$, and thereafter is the reflection about $a$ of the gray path. Every path that reaches $a$, then $-b_0$, then $a$ again at time $t$ has such a reflected path and this path reaches $3a + 2b_0$ for the first time at $t$. This allows calculation of the number of paths that need to be added back in, but paths that hit $3a + 2b_0$ at time $t$ do not have the same probability as paths that hit $a$ at time $t$. The probabilities can be adjusted by multiplying $P(\tau_{3a+2b_0} = t|b=\infty)$ by \((\frac{p_-}{p_+})^{a+b_0}\).

The expression

$$P(\tau_a = t|b=\infty) - P(\tau_{a+2b_0} = t|b=\infty) \cdot \left(\frac{p_-}{p_+}\right)^{b_0} - P(\tau_{3a+2b_0} = t|b=\infty) \cdot \left(\frac{p_-}{p_+}\right)^{a+b_0}$$

adjusts $P(\tau_a = t|b=\infty)$ for paths that hit $-b_0$ before hitting $a$, including those that hit $a$ first, then hit $-b_0$, then $a$ again at $t$.

The lower right pane of Figure 4.2 shows a path in black that hits $-b_0$, then $a$, then $-b_0$ again, and finally $a$ again at time $t$. This path is not included in $P(\tau_a = t|b=\infty)$ because it hits $a$ before time $t$. The gray path is the same as the black path up to the time when if first reaches $-b_0$ and thereafter is the reflection about $-b_0$ of the black path. Since this gray path reaches $-a - 2b_0$ before time $t$, its reflection reaches $a + 2b_0$ before time $t$ and as a result this path’s probability is also not included in $P(\tau_{a+2b_0} = t|b=\infty)$. Finally, the blue path is the same as the black path up to the time it first reaches $a$, then is the reflection about $a$ of the black path up to the time
it first reaches \(-b_0\), and thereafter is the reflection about \(a\) of the gray path. This blue path is included in \(P(\tau_{3a+2b_0} = t|b = \infty)\) because it reaches \(3a + 2b_0\) for the first time at \(t\). As a result, the probability of the black path shown in the lower right pane of Figure 4.2 needs to be subtracted out of the expression given above.

To correctly calculate \(P(\tau_a = t|b = b_0)\) using only one sided hitting probabilities like \(P(\tau_a = t|b = \infty)\), \(P(\tau_{a+2b_0} = t|b = \infty)\), and \(P(\tau_{3a+2b_0}|b = \infty)\), it is necessary to subtract some probabilities out and add others back in as demonstrated in Figure 4.2. As \(t\) increases, more adjustment terms will need to be subtracted out or added in. To do this for a given \(t\), first let \(k_1 = \left\lfloor \frac{t+a}{2(b_0+a)} \right\rfloor\) and let \(k_2 = \left\lfloor \frac{t-a}{2(b_0+a)} \right\rfloor\). Then \(k_1\) is the number of terms that need to be subtracted out and \(k_2\) is the number of terms that need to be added back in. Piecing all of this together,

\[
P(\tau_a = t|b = b_0) = P(\tau_a = t|b = \infty) \\
- \mathbf{1}_{\{t \geq a+2b_0\}} \sum_{j=1}^{k_1} P(\tau_{2j(a+b_0)-a} = t|b = \infty) \left( \frac{p_-}{p_+} \right)^{j(a+b_0)-a} \\
+ \mathbf{1}_{t \geq 3a+2b_0} \sum_{j=1}^{k_2} P(\tau_{2j(a+b_0)+a} = t|b = \infty) \left( \frac{p_-}{p_+} \right)^{j(a+b_0)},
\]

where \(\mathbf{1}_{\{\cdot\}}\) is an indicator variable.

When both \(a\) and \(b\) are finite, it is also possible to calculate \(P(\tau_a = t|b)\) using matrix algebra. Let \(P\) be the transition probability matrix for \(X_t\). That is, let

\[
P = \begin{pmatrix}
-b & -b+1 & -b+2 & \cdots & a \\
-b & 1 & 0 & 0 & \cdots & 0 \\
-b+1 & p_- & p_0 & p_+ & \cdots & 0 \\
-b+2 & 0 & p_- & p_0 & \cdots & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
a & 0 & 0 & 0 & \cdots & 1
\end{pmatrix}
\]

56
where, for convenience, the rows and columns of the matrix are indexed by \((-b, -b+1, \cdots -1, 0, 1, \ldots, a-1, a)\). Using these indices, \(P_{-1,0}\) is the probability of transitioning from state \(-1\) to state \(0\) and is equal to the entry of \(P\) in the \(-1\)st row and 0th column. The probability distribution of \(X_t\), for a fixed \(t \geq 1\) is given by the 0th row of \(P^t\), where \(P^t\) is the \(t\)th power of the transition probability matrix (Karlin and Taylor, 1975, pages 58-59). The probability that \(X_t\) is absorbed into state \(-b\) by time \(t\) is the first entry of this row and the probability that \(X_t\) is absorbed into state \(a\) by time \(t\) is the last entry of this row. The corresponding first hitting time probabilities are

\[
P(\tau_a = t|b) = P_{0,a}^t - P_{0,a}^{t-1}
\]

\[
P(\tau_{-b} = t|a) = P_{0,-b}^t - P_{0,-b}^{t-1}
\]

These matrix calculations were used for all implementations in Chapter 6.

Another quantity of interest is the probability of ever hitting a given boundary. Let \(P(\tau_a < \infty|X_0 = i, b)\) denote the probability of hitting boundary \(a\) at any time after starting at state \(i\), where \(i\) is an integer and \(-b < i < a\). Note that up to this point it was assumed that \(X_0 = 0\). The boundary hitting probabilities are

\[
P(\tau_a < \infty|X_0 = i, b) = \begin{cases} 
\frac{i+b}{a+b}, & \text{if } p_+ = p_- \\
\frac{1 - \left(\frac{p_+}{p_-}\right)^{i+b}}{1 - \left(\frac{p_+}{p_-}\right)^{a+b}}, & \text{if } p_+ \neq p_- 
\end{cases}
\]

\[
P(\tau_{-b} < \infty|X_0 = i, a) = \begin{cases} 
\frac{a-i}{a+b}, & \text{if } p_+ = p_- \\
\frac{1 - \left(\frac{p_+}{p_-}\right)^{a-i}}{1 - \left(\frac{p_+}{p_-}\right)^{a+b}}, & \text{if } p_+ \neq p_- 
\end{cases}
\]

These can be derived in the following way. For simplicity, in this derivation, transform the state space to \(\{0, 1, \ldots, a+b\}\) with states 0 and \(a+b\) being absorbing, and consider \(i \in \{1, 2, \ldots, a+b-1\}\). The location of the lower boundary, which
is at state 0, will be suppressed in the notation so that $P(\tau_{a+b} = t|X_0 = i)$ is the probability of reaching the upper boundary at time $t$. Then

$$P(\tau_{a+b} < \infty|X_0 = i) = p_+ \cdot P(\tau_{a+b} < \infty|X_0 = i + 1)$$

$$+ p_0 \cdot P(\tau_{a+b} < \infty|X_0 = i)$$

$$+ p_- \cdot P(\tau_{a+b} < \infty|X_0 = i - 1),$$

which can be rewritten as

$$P(\tau_{a+b} < \infty|X_0 = i) = \frac{p_+}{1 - p_0} \cdot P(\tau_{a+b} < \infty|X_0 = i + 1)$$

$$+ \frac{p_-}{1 - p_0} \cdot P(\tau_{a+b} < \infty|X_0 = i - 1).$$

Rearranging this further using the fact that $\frac{p_+}{1 - p_0} + \frac{p_-}{1 - p_0} = 1$ gives

$$P(\tau_{a+b} < \infty|X_0 = i + 1) - P(\tau_{a+b} < \infty|X_0 = i) =$$

$$\frac{p_-}{p_+} (P(\tau_{a+b} < \infty|X_0 = i) - P(\tau_{a+b} < \infty|X_0 = i - 1)).$$

Since $P(\tau_{a+b} < \infty|X_0 = 0) = 0$ (state 0 is absorbing),

$$P(\tau_{a+b} < \infty|X_0 = 1) - P(\tau_{a+b} < \infty|X_0 = 0) = P(\tau_{a+b} < \infty|X_0 = 1),$$

$$P(\tau_{a+b} < \infty|X_0 = 2) - P(\tau_{a+b} < \infty|X_0 = 1) = \left(\frac{p_-}{p_+}\right) P(\tau_{a+b} < \infty|X_0 = 1),$$

$$P(\tau_{a+b} < \infty|X_0 = 3) - P(\tau_{a+b} < \infty|X_0 = 2) = \left(\frac{p_-}{p_+}\right)^2 P(\tau_{a+b} < \infty|X_0 = 1),$$

$$\vdots$$

$$P(\tau_{a+b} < \infty|X_0 = i + 1) - P(\tau_{a+b} < \infty|X_0 = i) = \left(\frac{p_-}{p_+}\right)^i P(\tau_{a+b} < \infty|X_0 = 1).$$

58
To find an expression for $P(\tau_{a+b} < \infty|X_0 = i)$, note that

$$P(\tau_{a+b} < \infty|X_0 = i + 1) - P(\tau_{a+b} < \infty|X_0 = 1) =$$

$$\left( P(\tau_{a+b} < \infty|X_0 = 2) - P(\tau_{a+b} < \infty|X_0 = 1) \right)$$

$$+ \left( P(\tau_{a+b} < \infty|X_0 = 3) - P(\tau_{a+b} < \infty|X_0 = 2) \right)$$

$$\vdots$$

$$+ \left( P(\tau_{a+b} < \infty|X_0 = i + 1) - P(\tau_{a+b} < \infty|X_0 = i) \right)$$

$$= \sum_{k=1}^{i} \left( P(\tau_{a+b} < \infty|X_0 = k + 1) - P(\tau_{a+b} < \infty|X_0 = k) \right)$$

$$= \sum_{k=1}^{i} \left( \frac{p_-}{p_+} \right)^k P(\tau_{a+b} < \infty|X_0 = 1),$$

which implies that

$$P(\tau_{a+b} < \infty|X_0 = i + 1) = P(\tau_{a+b} < \infty|X_0 = 1) + \sum_{k=1}^{i} \left( \frac{p_-}{p_+} \right)^k P(\tau_{a+b} < \infty|X_0 = 1),$$

$$= \sum_{k=0}^{i} \left( \frac{p_-}{p_+} \right)^k P(\tau_{a+b} < \infty|X_0 = 1),$$

$$= P(\tau_{a+b} < \infty|X_0 = 1) \sum_{k=0}^{i} \left( \frac{p_-}{p_+} \right)^k,$$

$$= \begin{cases} 
(x + 1)P(\tau_{a+b} < \infty|X_0 = 1), & \text{if } p_- = p_+, \\
\frac{1 - (\frac{p_-}{p_+})^{x+1}}{1 - \frac{p_-}{p_+}} P(\tau_{a+b} < \infty|X_0 = 1), & \text{if } p_- \neq p_. 
\end{cases}$$

To solve for $P(\tau_{a+b} < \infty|X_0 = 1)$, and thus obtain an expression for all $i$, note that

$$1 = P(\tau_{a+b} < \infty|X_0 = a + b) = \begin{cases} 
(a + b)P(\tau_{a+b} < \infty|X_0 = 1), & \text{if } p_- = p_+, \\
\frac{1 - (\frac{p_-}{p_+})^{a+b}}{1 - \frac{p_-}{p_+}} P(\tau_{a+b} < \infty|X_0 = 1), & \text{if } p_- \neq p_. 
\end{cases}$$
This gives $P(\tau_{a+b} < \infty | X_0 = 1) = \frac{1 - p_-}{1 - (\frac{p_+}{p_-} + \sigma)}$ for the case when $p_- \neq p_+$ and $P(\tau_{a+b} < \infty | X_0 = 1) = \frac{1}{a+b}$ for the case when $p_+ = p_-$. Substituting this into the above formula and transforming the state space back to $\{-b, \ldots, a\}$ yields the expressions

$$P(\tau_a < \infty | X_0 = i, b) = \begin{cases} \frac{i+b}{a+b} & \text{if } p_+ = p_-, \\ \frac{1 - (\frac{p_+}{p_-})^{i+b}}{1 - (\frac{p_+}{p_-})^{a+b}} & \text{if } p_+ \neq p_- \end{cases}$$

and

$$P(\tau_{-b} < \infty | X_0 = i, a) = \begin{cases} \frac{a-i}{a+b} & \text{if } p_+ = p_-, \\ \frac{1 - (\frac{p_+}{p_-})^{a-i}}{1 - (\frac{p_+}{p_-})^{a+b}} & \text{if } p_+ \neq p_- \end{cases}$$

### 4.1.3 Expected hitting time

Let $\tau$ be the first time that the process $X_t$ hits a boundary at either $a$ or $-b$. The mean hitting time, $E(\tau)$, can be derived using an argument similar to that used to find $P(\tau_a < \infty | X_0 = i)$ and $P(\tau_{-b} < \infty | X_0 = i)$ above. As before, to simplify the derivation, let the state space be $\{0, 1, \ldots, a+b\}$. Also, let $k_i = E(\tau | X_0 = i)$ and consider first the case where $p_+ = p_- > 0$. Each term $k_i$ can be written in terms of $k_{i+1}$ and $k_{i-1}$ as

$$k_i = 1 + p_+ k_{i+1} + p_0 k_i + p_- k_{i-1},$$

which implies that

$$k_i = \frac{1}{1 - p_0} + \frac{1}{2} k_{i+1} + \frac{1}{2} k_{i-1},$$

since $p_+ = p_-$ by assumption. Rearranging this equality yields

$$k_{i+1} - k_i = k_i - k_{i-1} - \frac{2}{1 - p_0}.$$
Writing \( k_0 \equiv 0 \) since 0 is an absorbing state, the first few terms of this relationship are

\[
\begin{align*}
k_1 - k_0 &= k_1, \\
k_2 - k_1 &= k_1 - k_0 - \frac{2}{1 - p_0} = k_1 - \frac{2}{1 - p_0}, \\
k_3 - k_2 &= k_2 - k_1 - \frac{2}{1 - p_0} = k_1 - \frac{4}{1 - p_0}, \\
&\quad \vdots \\
k_{i+1} - k_i &= k_1 - \frac{2i}{1 - p_0},
\end{align*}
\]
for \( i \in \{0, 1, \ldots, a + b\} \). Now,

\[
\begin{align*}
k_{i+1} - k_1 &= (k_{i+1} - k_i) + (k_i - k_{i-1}) + \cdots + (k_2 - k_1), \\
&= \sum_{j=1}^{i} (k_{j+1} - k_j), \\
&= \sum_{j=1}^{i} (k_1 - \frac{2j}{1 - p_0}), \\
&= i k_1 - \frac{2i}{1 - p_0} \frac{i(i+1)}{2}, \\
k_{i+1} &= (i + 1)(k_1 - \frac{i}{1 - p_0}).
\end{align*}
\]

To solve for \( k_1 \), note that

\[
k_{a+b} = 0 = (a + b) \left( k_1 + \frac{a + b - 1}{1 - p_0} \right),
\]

which implies that

\[
k_1 = \frac{a + b - 1}{1 - p_0}.
\]

Thus,

\[
k_i = E(\tau | X_0 = i) = i \left( \frac{a + b - 1}{1 - p_0} - \frac{i - 1}{1 - p_0} \right) = \frac{i(a + b - i)}{1 - p_0}.
\]
In terms of the original state space \(-b, -b + 1, \ldots, a - 1, a\), this can be written as

\[
E(\tau | X_0 = i) = \frac{(i + b)(a - i)}{1 - p_0} \quad i = -b, -b + 1, \ldots, a - 1, a.
\]

This expression will be finite as long as \(p_0 \neq 1\), which is true by assumption (it was assumed that \(p_+ = p_- > 0\)).

When \(p_+ \neq p_-\), the argument is similar. Assume without loss of generality that \(p_+ > p_-\). With \(k_i\) defined as before,

\[
k_i = \frac{1}{p_0} + \frac{p_+}{1 - p_0} k_{i+1} + \frac{p_-}{1 - p_0} k_{i-1}.
\]

The same steps used above for the \(p_+ = p_-\) case can be used to show that, when \(p_+ > p_-\),

\[
k_i = \frac{a + b}{p_+ - p_-} \frac{1 - \left(\frac{p_-}{p_+}\right)^i}{1 - \left(\frac{p_-}{p_+}\right)^{a+b}} - \frac{i}{p_+ - p_-} \quad i = 0, 1, \ldots, a + b - 1, a + b.
\]

In terms of the original state space, this can be written as

\[
k_i = \frac{a + b}{p_+ - p_-} \frac{1 - \left(\frac{p_-}{p_+}\right)^{i+b}}{1 - \left(\frac{p_-}{p_+}\right)^{a+b}} - \frac{i + b}{p_+ - p_-} \quad i = -b, -b + 1, \ldots, a - 1, a.
\]

### 4.1.4 Hitting time central limit theorem

Again, consider a random walk \(\{X_t\}\) that begins at state 0 and continues until reaching one of the boundaries \(a\) or \(-b\). In this section, assume that \(p_+ - p_- > \delta\) for some \(\delta > 0\) and that \(p_+, p_- > 0\). This process has a tendency to drift toward boundary \(a\).

For the special case \(b = \infty\), the process would eventually reach boundary \(a\) with probability one. In this case a central limit theorem for the hitting time is fairly
straightforward. Let $\tau_1$ be the time when $X_t$ first reaches state 1, $\tau_2$ be the time when $X_t$ first reaches state 2, and so on. Also, let $\tau_0 = 0$. Then $\tau_a = \sum_{i=1}^a \tau_i - \tau_{i-1}$. The differences on the right side of this equation are independent and identically distributed random variables because the process is time-homogeneous. The first two moments of $\tau_1$ are

$$E(\tau_1) = \frac{1}{p_+ - p_-},$$
$$E(\tau_1^2) = \frac{1}{p_+ - p_-} + \frac{2(p_0 + 2p_-)}{(p_+ - p_-)^2} + \frac{2p_-}{(p_+ - p_-)^3}.$$  

By the classical CLT, as $a \to \infty$

$$\frac{1}{\sqrt{a}} \left( \tau_a - \frac{a}{p_+ - p_-} \right) \overset{d}{\to} N \left( 0, \frac{1}{p_+ - p_-} + \frac{2(p_0 + 2p_-) - 1}{(p_+ - p_-)^2} + \frac{2p_-}{(p_+ - p_-)^3} \right).$$

When there is a finite lower boundary at $b < \infty$, then the above result depends on $b$ going to infinity in addition to $a$. To see that this is sufficient for the CLT to hold, let $\tau_a$ be the first hitting time at $a$ for a random walk with two absorbing boundaries (at $a$ and at $-b$) and let $T_a$ be the first passage time at $a$ for a random walk with no absorbing boundaries (that is, the process is not absorbed after reaching state $a$ or state $-b$). Similarly, $T_{-b}$ will be the first hitting time at $-b$ with no absorbing boundaries. Note that

$$P(\tau_a \leq t) = P(T_a \leq t \cap T_{-b} > T_a) \geq P(T_a \leq t \cap T_{-b} > t),$$

since $[T_a \leq t \cap T_{-b} > T_a] \supseteq [T_a \leq t \cap T_{-b} > t]$. Then

$$1 - P(\tau_a \leq t) \leq 1 - P(T_a \leq t \cap T_{-b} > t) = P(T_a > t \cup T_{-b} \leq t) \leq P(T_a > t) + P(T_{-b} \leq t)$$

63
Therefore,

\[ P(\tau_a \leq t) \geq 1 - P(T_a > t) - P(T_{-b} \leq t) = P(T_a \leq t) - P(T_{-b} \leq t). \]

The following relationship is therefore true, where the second inequality results from the fact that \([\tau_a \leq t] \subseteq [T_a \leq t]:\)

\[ P(T_a \leq t) - P(T_{-b} \leq t) \leq P(\tau_a \leq t) \leq P(T_a \leq t). \]

Since \(P(T_{-b} \leq t) \to 0\) as \(b \to \infty\), \(P(\tau_a \leq t) \to P(T_a \leq t)\) as \(b \to \infty\).

There will be some error in using the normal approximation for the hitting time at a finite \(a\). To get a bound on this error for a specific \(a\), let \(Y_i = (\tau_i - \tau_{i-1}) - E(\tau_i - \tau_{i-1}) - \frac{\sqrt{\text{var}(\tau_i - \tau_{i-1})}}{\sqrt{a}}\), \(i = 1, 2, \ldots, a\), and let \(\bar{Y} = \frac{1}{a} \sum_{i=1}^{a} Y_i\). Also let \(F_a(\bar{y})\) be the true cdf of \(\bar{Y}\). If \(\lim_{a \to \infty} F_a(\bar{y}) = \Phi(\bar{y})\) then the CLT will hold. The Berry-Esseen inequality puts a bound on the maximum difference between \(F_a(\bar{y})\) and the standard normal cdf for a finite \(a\). There are several versions of this inequality. Shevtsova (2011) gives the following two where \(\Delta_a = \sup_x |F_a(x) - \Phi(x)|\) and \(\beta_3\) is the third absolute moment of \(Y_1\).

\[
\Delta_a \leq 0.3328 \frac{\beta_3 + 0.429}{\sqrt{a}}, \\
\Delta_a \leq 0.33554 \frac{\beta_3 + 0.415}{\sqrt{a}},
\]

The form of \(\beta_3\) is given below, and is finite due to the assumption that \(p_+ - p_- > \delta\) for some \(\delta > 0\).

\[
\beta_3 = \frac{1}{p_+ - p_-} + \frac{6(p_0 + 2p_-)}{(p_+ - p_-)^2} + \frac{6(p_0 + 2p_-)^2 + 12p_- - 1}{(p_+ - p_-)^3} + \frac{12p_-}{(p_+ - p_-)^4} + \frac{18p_- (p_0 + 2p_-)}{(p_+ - p_-)^4} + \frac{2p_- (p_0 + 2p_- - 1)}{(p_+ - p_-)^3} + \frac{2p_-}{(p_+ - p_-)^2} \frac{3}{2} \sum_{i=1}^{[\frac{1}{p_+ - p_-}]} (t - \frac{1}{p_+ - p_-})^3 P(\tau_1 = t | b) \\
- \left( \frac{1}{p_+ - p_-} + \frac{2(p_0 + 2p_- - 1)}{(p_+ - p_-)^2} + \frac{2p_-}{(p_+ - p_-)^2} \right)^{3/2}.
\]
The presence of a finite lower boundary will contribute some additional error which is bounded by $P(\tau_{-b} < \infty|a) = 1 - P(\tau_{-b} < \infty|a)$. Using the first form of the Berry-Esseen inequality above,

$$\Delta_a \leq 0.3328 \frac{\beta_3 + 0.429}{\sqrt{a}} + (1 - P(\tau_a < \infty|b)).$$

For finite $a$, this normal approximation is best when $a$ is large and $p_+ - p_-$ is large. If this approximation is used when a lower boundary $-b$ is present, the lower boundary doesn’t have much of an impact as long as it is not close to zero. Figure 4.3 illustrates the hitting time CLT with $p_+ = 0.7$, $p_0 = 0.25$, $p_- = 0.05$ and $b = 10$ for several choices of $a$.

![Figure 4.3: Normal approximation for hitting time, $a=10, 50, 100$ and $b=10$.](image)

The above derivation assumes that the starting point of the process is fixed at 0. It is not necessary that the starting point be 0 for the asymptotic result to work, but
to use the normal approximation it is necessary for the boundaries to be sufficiently far from the starting point.

**Hitting time CLT with random starting point**

If the starting point $X_0$ is random, then the CLT above may not hold if its support changes with $a$ and $b$. In particular, if as $a \to \infty$ the support of the random starting point $X_0$ increases to allow the starting point to be close to $a$ then the CLT will not hold. Suppose instead that the support of $X_0$ is confined to some range of states $\{c, \ldots, d\}$. Note that the first two moments and third absolute moment ($\beta_3$) of the increments $\xi_t = X_t - X_{t-1}$ do not change.

For this random walk with a random starting point in $\{c, \ldots, d\}$, denote the first hitting time at $a > d$ as $\tau_a^r$ where $r$ denotes a random starting point. Let $\tau_i$ denote the first hitting time at state $i$ for a process that starts at state $0$ as before. Then $\tau_a^r$ has the same distribution as $\sum_{j=c}^d I_{\{X_0=j\}} \tau_{a-j}$. For $j \in \{c, \ldots, d\}$, the first two moments of $\tau_a^r$ conditional on starting in state $j$ are given by the following expressions

$$E\left(\frac{\tau_a^r}{a} \bigg| X_0 = j\right) = E\left(\frac{1}{a} \sum_{l=1}^{a-j} (\tau_l - \tau_{l-1})\right) = \frac{1}{a} \sum_{l=1}^{a-j} E(\tau_l - \tau_{l-1}) = \frac{a-j}{a} E(\tau_1),$$

$$E\left(\left(\frac{\tau_a^r}{a}\right)^2 \bigg| X_0 = j\right) = \frac{1}{a^2} E\left(\sum_{l=1}^{a-j} (\tau_l - \tau_{l-1})^2\right)$$

$$\quad = \frac{1}{a^2} \sum_{l=1}^{a-j} E(\tau_l - \tau_{l-1})^2 \quad \text{(independent increments)}$$

$$\quad = \frac{1}{a^2} \sum_{l=1}^{a-j} E(\tau_1)^2$$

$$\quad = \left(\frac{a-j}{a}\right) \frac{E(\tau_1^2)}{a}.$$
Since \( \frac{a-j}{a} \to 1 \) as \( a \to \infty \), the first two moments of \( \tau_{a}^{r} \) will converge to the first two moments of \( \tau_{a} \) for all \( j \) in the support of \( X_{0} \). Since \( \tau_{a}^{r} | X_{0} = j \) is the sum of iid random variables, by the classical CLT,

\[
\frac{1}{\sqrt{a}} \left( \tau_{a} - \frac{a}{p_{+} - p_{-}} \right) \bigg| X_{0} = j \xrightarrow{d} N \left( 0, \left( \frac{1}{p_{+} - p_{-}} + \frac{2(p_{0} + 2p_{-}) - 1}{(p_{+} - p_{-})^{2}} + \frac{2p_{-}}{(p_{+} - p_{-})^{3}} \right) \right),
\]

for all \( j \in \{c, \ldots, d, \} \).

### 4.2 Random Walk in a Dynamic Markovian Random Environment

Now let \( X_{t} \) be a random walk in a dynamic random environment. The environment at time \( t \) will be called \( \omega_{t} \). For simplicity let there be two environments, each specifying a probability distribution for the next step of the process \( X_{t} \). That is, \( \omega_{t} \in \{1, 2\} \) for each \( t \). The transition probabilities in environment \( i, i = 1, 2 \), will be called \( (p_{+}^{i}, p_{0}^{i}, p_{-}^{i}) \). At each time point, one of these environments is chosen at random and the walker takes his next step according to the probability distribution in that environment. The environment process \( \omega_{t} \) is assumed to be an irreducible recurrent Markov chain, and the probability of moving from environment \( i \) to environment \( j \) in one time increment will be denoted by by \( \alpha_{ij} \), \( i, j = 1, 2 \). The environment process is an example of an alternating renewal process, and by relation 7.7 of chapter 5 in Karlin and Taylor (1975) the limiting distribution is given by

\[
\lim_{t \to \infty} P(\omega_{t} = 1) = \frac{\mu_{1}}{\mu_{1} + \mu_{2}} \quad \text{and} \quad \lim_{t \to \infty} P(\omega_{t} = 2) = \frac{\mu_{2}}{\mu_{1} + \mu_{2}}
\]

where \( \mu_{i} \) is the mean hold time in environment \( i \). The length of these hold times follows a geometric distribution and it can easily be shown that \( \mu_{i} = \frac{1}{\alpha_{ij}} \). The
stationary distribution for the environment process is therefore

\[ \pi(\omega_t = 1) = \frac{\alpha_{21}}{\alpha_{12} + \alpha_{21}} = v_1, \quad \pi(\omega_t = 2) = \frac{\alpha_{12}}{\alpha_{12} + \alpha_{21}} = v_2. \]

In the case that there are \( K \) environments, the stationary distribution can be found in the following way. Let \( A = [\alpha_{ij}]_{i,j=1}^{K} \) be the transition probability matrix for \( \omega_t \) and let \( \mathbf{v} = [v_1, v_2, \ldots, v_K] \) be a row vector of probabilities. The (unique) normalized vector \( \mathbf{v} \) that satisfies \( \mathbf{v}A = \mathbf{v} \) is the stationary distribution for \( \omega_t \). Another way of stating this is, for \( i = 1, 2, \ldots, K \),

\[ v_i = \sum_{j=1}^{K} v_j \alpha_{ji}. \]

### 4.2.1 Drift

When observed for a long time, the process \( X_t \) will have an average "drift" that is given by \( \lim_{t \to \infty} \frac{X_t}{t} \). The strong ergodic theorem can be used to find this limit (Theorem 5.4 of chapter 9 of Karlin and Taylor, 1975). Let \( \xi_t = X_t - X_{t-1} \) be the increments of the random walk. Then assuming the environment process starts from its stationary distribution, the increments are a stationary process with finite mean

\[
E(\xi_t) = E(1 \cdot 1_{\xi_t=1} + (-1) \cdot 1_{\xi_t=-1} + 0 \cdot 1_{\xi_t=0}) \\
= (v_1 \cdot p_+^1 + v_2 \cdot p_-^2) - (v_1 \cdot p_-^1 + v_2 \cdot p_+^2) \\
= v_1(p_+^1 - p_-^1) + v_2(p_-^2 - p_+^2).
\]

By the strong ergodic theorem,

\[
\lim_{t \to \infty} \frac{X_t}{t} = \lim_{t \to \infty} \frac{1}{t} \sum_{i=1}^{t} \xi_i = E(\xi_t) = v_1(p_+^1 - p_-^1) + v_2(p_-^2 - p_+^2).
\]
4.2.2 Convergence to a Continuous Process

A standard argument for convergence of a random walk to a continuous process uses the functional CLT (Bhattacharya and Waymire, 1990). This method requires that the random walk have independent increments, which the random walk in a dynamic random environment does not have (due to the Markov nature of the environment process). Instead, a martingale CLT (Hall, 1980) will be used that does not require independent increments.

By the strong ergodic theorem, $X_t$ has a “drift” that is given by $v_1(p_1^1 - p_1^-) + v_2(p_2^2 - p_2^-)$ and the environment process $\omega_t$ has mean hold times in environment $i$ equal to $\frac{1}{\alpha_{ij}}$, $i, j = 1, 2$. Let $\omega_t^{(n)}$ be the environment process for a random walk that makes transitions at times $\frac{1}{n}, \frac{2}{n}, \ldots$, for $n = 1, 2, \ldots$. As time “scales down” in this way, it will be desirable to retain the same mean hold times in the two environments. To do this, let $\alpha_{ij}^{(n)} = \alpha_{ij} n$ and let the environment process move from environment 1 to environment 2 with probability $\alpha_{12}^{(n)}$ (and stay in environment 1 with probability $1 - \alpha_{12}^{(n)} = \alpha_{11}^{(n)}$).

Let $Y_n$ be a geometric random variable with success probability $\alpha_{12}^{(n)}$ that is the hold time in environment 1 measured in number of transitions. Note that as $n$ increases, there will be more transitions per unit of time. The hold time in environment 1 measured in the original time units is then $\frac{Y_n}{n}$, and for all $n$, $E\left(\frac{Y_n}{n}\right) = \frac{1}{\alpha_{12}}$. 

69
Since the cdf of a geometric random variable is \( F(k) = 1 - (1 - p)^k \), the following is true for the random variable \( Y_n \):

\[
1 - F_n(t) = P\left(\frac{Y_n}{n} > t\right) = P(Y_n > tn) = P(Y_n > \lfloor tn \rfloor)
\]

\[
= (1 - \frac{\alpha_{12}}{n})^{\lfloor tn \rfloor}
\]

\[
= (1 + \frac{-\alpha_{12}}{n})^{n\lfloor tn \rfloor}.
\]

Then,

\[
\log(1 - F_n(t)) = \frac{\lfloor tn \rfloor}{n} \log((1 + \frac{-\alpha_{12}}{n})^{n}),
\]

which implies that

\[
\log(1 - F_n(t)) \xrightarrow{n \to \infty} -t\alpha_{12},
\]

so that

\[
F_n(t) \xrightarrow{n \to \infty} 1 - \exp(-\alpha_{12}t).
\]

In the limit, the hold times of \( \omega_t^{(n)} \) in environment 1 are exponential with mean \( \frac{1}{\alpha_{12}} \). The same argument shows that the hold times in environment 2 are exponential with mean \( \frac{1}{\alpha_{21}} \).

Let \( z_t^{(n)} \) be given by

\[
z_t^{(n)} = \xi_t - 1_{\{\omega_t^{(n)} = 1\}} E(\xi_t | \omega_{t-1}^{(n)} = 1) - 1_{\{\omega_t^{(n)} = 2\}} E(\xi_t | \omega_{t-1}^{(n)} = 2)
\]

Then \( E(z_t^{(n)} | \mathcal{F}_{t-1}) = E(z_t | \omega_{t-1}^{(n)}) = 0 \) and \( z_t^{(n)} \) is a mean zero martingale, where \( \mathcal{F}_{t-1} \) is the \( \sigma \)-algebra generated by the history of \( X_t \) up to time \( t - 1 \). Hall (1980) gives the following theorem (Theorem 3.2).
Theorem. Let \( \{S_{ni}, F_{ni}, i \leq i \leq k_n, n \geq 1\} \) be a mean zero square integrable martingale array with increments \( X_{ni} \) and let \( \eta^2 \) be a.s. finite. Suppose that

\[
\max_i |X_{ni}| \xrightarrow{p} 0, \quad n \to \infty,
\]
\[
\sum_i X_{ni}^2 \xrightarrow{p} \eta^2, \quad n \to \infty,
\]

and also let the \( \sigma \)-fields be nested \( F_{ni} \subset F_{n+1,i} \). Then \( S_{n,k_n} = \sum_i X_{ni} \xrightarrow{d} Z \), where \( Z \sim N(0, \eta^2) \).

To apply this theorem to the situation under study, take \( X_{ni} = \frac{z_i^{(n)}}{\sqrt{n}} \) to be the increments (so that the state space scales appropriately with the time increments). Then,

\[
\max_i |X_{ni}| = \max_i \left| \frac{z_i^{(n)}}{\sqrt{n}} \right| \xrightarrow{p} 0, \quad n \to \infty,
\]

because the \( z_i^{(n)} \) are bounded and

\[
\sum_i \left( \frac{z_i^{(n)}}{n} \right)^2 = \frac{1}{n} \sum_i (z_i^{(n)})^2 \xrightarrow{p} \eta^2, \quad n \to \infty,
\]

where

\[
\eta^2 = E \left( (z_i^{(n)})^2 \right) = v_1(p_1^+ + p_1^-) + v_2(p_2^+ + p_2^-) - (v_1(p_1^+ - p_1^-))^2 + v_2(p_2^+ - p_2^-)^2
\]

Thus, the process \( X_{ni} \) converges to a diffusion process with drift 0 and diffusion coefficient \( \eta^2 \).

4.2.3 Boundary hitting times

For a random walk in a dynamic Markovian random environment, expressions for the hitting time probabilities at one or two boundaries will be different from those in the single environment case. Let \( \pi_i = P(\omega_0 = i), i = 1, \ldots, K \), and let
\( \pi = (\pi_1, \ldots, \pi_K) \) be the starting distribution for the environment process. That is, the environment at time 0 is determined by the distribution \( \pi \). The limiting distribution for the environment process was given earlier by \( \nu \). Let \( v_i(t) \) be the probability of being in environment \( i \) at time \( t \) after starting from distribution \( \pi \). Let \( A = [\alpha_{ij}]_{i,j=1}^K \) as before. Then the following is true, where \( v(t) = (v_1(t), \ldots, v_K(t)) \) and \( A^0 = I \):

\[
v(t) = \pi A^t.
\]

If the starting distribution \( \pi \) is the stationary distribution for the \( \omega_t \) then note that \( v(t) = \nu \) for all \( t \) since \( \nu A = \nu \). In that case there is no need to express \( v(t) \) as a function of time.

Let \( (p_i^+, p_i^0, p_i^-) \) be the transition probabilities for \( X_t \) in environment \( i \). Define the following functions of \( t \):

\[
    p_+(t) = \sum_{i=1}^{K} v_i(t)p_i^+, \quad p_0(t) = \sum_{i=1}^{K} v_i(t)p_i^0, \quad p_-(t) = \sum_{i=1}^{K} v_i(t)p_i^-.
\]

As in the case of a single environment, the number of paths that hit a boundary at \( a \) or \( -b \) at time \( t \) can be counted with the difference of two multinomial coefficients. However, it will not be the case that all sample paths with the same number of upward transitions will have the same probability. As a result, the hitting time probabilities cannot be written in the same way as \( P(\tau_a = t | b) \) and \( P(\tau_{-b} = t | a) \) were written in Section 4.1.2. However, they can still be written using matrix notation.

Let \( P(t) \) be the transition probability matrix for \( X_t \) at time \( t \). Let \( z = (0, \ldots, 0, 1, 0, \ldots, 0) \) be a row vector of length \( a + b + 1 \), with \( b \) zeros followed by a one followed by \( a \) zeros. The vector \( z \) is the (degenerate) starting distribution for the process \( X_t \), which
starts in state zero with probability one. Then $z \cdot \prod_{j=1}^t P(j)$ is a vector that gives the probability of being in each state at time $t$. Since states $-b$ and $a$ are absorbing, the first and last entries are the cumulative probabilities of hitting states $-b$ and $a$ respectively.
Chapter 5: Discrete Threshold Regression Model

5.1 Discrete State, Discrete Time Process

In this chapter, a discrete threshold regression (DTR) model is presented that is based on a Markov chain with a discrete state space and discrete time. Let \( \{X_t\}_{t=0,1,2,...} \) be a Markov chain with integer-valued state space \( S = \{-b, -b + 1, \ldots, a - 1, a\}, a, b > 0 \). The boundary states, \(-b\) and \(a\), are absorbing states. Then \( X_t \) will start off in some state \( X_0 \in S \setminus \{a, -b\} \) and will travel between states until either \( a \) or \(-b\) is reached. Figure 5.1 shows one sample path of \( X_t \) with \( S = \{-3, -2, -1, 0, 1, 2, 3\} \). In the figure, the boundary states are \(-3\) and \(3\).

For motivation, consider the Worcester Heart Attack Study data set (Hosmer et al., 2008). Five hundred patients were admitted to a hospital in the Worcester, Massachusetts area after a heart attack. For each patient, the dataset includes the length of stay at the hospital and whether or not the patient was alive at the end of their hospital stay. There are also some other variables in the data set, including physical measurements taken at the time of hospital admission (such as pulse and blood pressure) and indicators for several risk factors (such as history of cardiovascular disease and history of heart attacks).
Figure 5.1: Sample path of a seven state Markov chain that starts in state 0 and enters absorbing state 3 at time 8.

The model shown in Figure 5.1 can be used to model the length of stay at the hospital. Figure 5.2 shows a sample path of $X_t$ for this example, where the upper absorbing boundary represents healthy discharge from the hospital and the lower boundary represents death in the hospital. The sample path shown is for a patient who was discharged from the hospital after 5 days. Like the continuous models described by Aalen and Gjessing (2001), Lee and Whitmore (2006, 2010), and others, this latent stochastic process represents a real physical process. For this data set, the physical process is that of recovery after a heart attack. From day to day an individual’s health may improve or worsen until they are either healthy enough to leave the hospital or until they have died in the hospital. An important difference between this model and the continuous models is that it moves in discrete steps. These can be thought of as stages of recovery (or deterioration) of the patient’s health. The value of $X_0$ and transitions between these stages are not directly observed, similar to
Figure 5.2: Example of a sample path in the heart attack example.

the phase-type models in Aalen (1995), but unlike those models the exact nature of each stage in the heart attack model does not need to be specified.

5.1.1 Initial state

Since the process is latent, the initial state is unknown. But there may be some covariates available that influence the starting point. For example, in the Worcester Heart Attack Study, initial measurements of each patient’s pulse and blood pressure were taken. Low blood pressure and a weak pulse can be signs of cardiogenic shock, a condition in which the heart is so badly damaged that it can not pump enough blood to a person’s organs. This is a dangerous condition, and in the Worcester Heart Attack Study 59% of patients with cardiogenic shock died while only 5% of patients without cardiogenic shock died. As a result, a weak pulse and low blood pressure can indicate that a patient’s heart attack was severe and that their recovery process $X_t$ should start close to the lower boundary.
A distribution can be constructed over the transient states that uses covariates to determine how close to each of the boundaries an individual’s process $X_t$ should start. In this way the model accounts for the fact that some heart attacks are very severe and dangerous while others are milder and have a higher chance of recovery. This distribution can be constructed in any way that seems appropriate; the approach presented here uses quantiles of a continuous distribution. A discretized version of the beta distribution is a natural choice because of its finite support and flexibility.

The beta distribution is defined on the interval $(0, 1)$ and can have several different shapes including U-shaped, J-shaped (forward or backward), and unimodal. Its density is

$$f(y) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} y^{\alpha-1}(1-y)^{\beta-1}, \quad \text{for } 0 < y < 1,$$

Here, $\Gamma(\cdot)$ is the gamma function, and $\alpha, \beta > 0$. To use a beta distribution for the distribution of $X_0$, let $m = a + b - 1$ be the number of transient states. Then divide up the interval $(0, 1)$ into $m$ equal sized subintervals, $(0, \frac{1}{m}], (\frac{1}{m}, \frac{2}{m}], \ldots, (\frac{m-1}{m}, 1)$. For a beta distributed variable $y$ and $j = -b + 1, \ldots, a - 1$, let $p^0(j) = P(X_0 = j) = P(\frac{j+b-1}{m} < y \leq \frac{j+b}{m})$.

Figure 5.3 shows some shapes of the beta distribution and the corresponding discrete allocation of probability over 10 transient states. The unimodal shape in the example puts highest probability on states that are far from either boundary. This shape implies that $X_t$ will typically not hit a boundary for several time periods. On the other hand, the U-shaped distribution puts a lot of mass near both boundaries. This shape implies that $X_t$ is likely to hit a boundary very quickly, but it could hit either boundary. The J-shaped distribution puts most of the mass near one of the
Figure 5.3: Examples of Beta distributions and the corresponding discrete distribution for 10 transient states.

$X_t$ is very likely to hit this boundary and to hit it very quickly unless the transition probabilities pull strongly toward the other boundary.

To include the effect of covariates on the initial state $X_0$, a regression function is needed to link the covariates to the parameters of the beta distribution. To accomplish this, the beta distribution will be reparameterized in the following way. Let $\xi$ be the mean of a beta distribution with shape parameters $\alpha$ and $\beta$. Then $\xi = \frac{\alpha}{\alpha + \beta}$. Given $\beta$ and $\xi$, $\alpha = \frac{\xi}{1-\xi} \beta$. So the beta distribution can be parameterized by $\xi$ and one of $\alpha$ and $\beta$. Let $\text{logit}(\xi) = z\gamma_0$, where $\gamma_0$ is a column vector of regression coefficients and
\( \mathbf{z} \) is a row vector of covariate values, so that
\[
\xi = \frac{\exp(\mathbf{z} \gamma_0)}{1 + \exp(\mathbf{z} \gamma_0)}.
\]
The logit link function ensures that \( 0 < \xi < 1 \). Now the parameters of the distribution of \( X_0 \) are \( \beta \) and \( \gamma_0 \).

Using the above parametrization, \( \xi \) determines the mean of \( X_0 \) and \( \beta \) determines the spread and shape. For very small values of \( \beta \) the distribution of \( X_0 \) will be U-shaped, and for large values it will be unimodal. For very large values of \( \beta \) the distribution will be unimodal and very concentrated (it will have small variance).

Since the mean \( \xi \) is a function of \( \mathbf{z} \gamma_0 \), it is instructive to consider the values of this mean that will result from various parameter values. If \( \mathbf{z} \gamma_0 = 0 \) then \( \xi = \frac{1}{2} \), leading to a symmetric distribution for \( X_0 \). If \( \mathbf{z} \gamma_0 < 0 \) then the mean will be closer to the lower boundary, and if \( \mathbf{z} \gamma_0 > 0 \) the mean will be closer to the upper boundary. If \( \mathbf{z} \gamma_0 = -3 \) then \( \xi \approx 0.05 \) and if \( \mathbf{z} \gamma_0 = 3 \) then \( \xi \approx 0.95 \), so for most applications \( \mathbf{z} \gamma_0 \in (-3, 3) \) gives an appropriate range.

### 5.1.2 Transition probabilities

Motivated by the approach that Diederich (1997) used to approximate a continuous stochastic process with a discrete-state discrete-time Markov chain, the transition probabilities for \( X_t \) are parametrized in the following way. Let \( \mu \) and \( \sigma^2 \) be constants such that \( 0 < \sigma^2 < 1 \) and \( |\mu| < \sigma^2 \), and let \( p_{ij} = P(X_t = j | X_{t-1} = i) \). Then for \( i \neq a, -b \)
\[
p_{ij} = \begin{cases} 
\frac{1}{2}(\sigma^2 + \mu), & \text{if } i - j = 1, \\
1 - \sigma^2, & \text{if } i = j, \\
\frac{1}{2}(\sigma^2 - \mu), & \text{if } i - j = -1, \\
0, & \text{otherwise.}
\end{cases}
\]
The resulting transition probability matrix is:

\[
P = \begin{pmatrix}
-b & -b + 1 & -b + 2 & \cdots & a \\
1 & 0 & 0 & \cdots & 0 \\
\frac{1}{2}(\sigma^2 - \mu) & 1 - \sigma^2 & \frac{1}{2}(\sigma^2 + \mu) & \cdots & 0 \\
0 & \frac{1}{2}(\sigma^2 - \mu) & 1 - \sigma^2 & \cdots & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
a & 0 & 0 & 0 & \cdots & 1
\end{pmatrix}
\]

Using this parametrization, \(\mu\) plays the role of a drift coefficient. If \(\mu > 0\), then the process will tend to drift toward the upper boundary and if \(\mu < 0\) it will tend to drift toward the lower boundary. If \(\mu = 0\) then the process does not tend to drift in one direction more than the other. If \(\sigma^2\) is large (close to 1), then \(X_t\) will be very mobile because it has a low chance of staying in the same state for more than one time period. Note that \(\sigma^2\) is not a variance. Rather, it is the expected value of the squared change in \(X_t\) in a single transition. These interpretations of \(\mu\) and \(\sigma^2\) can be summarized as

\[
E(X_{t+1} - X_t) = \mu, \quad E(X_{t+1} - X_t)^2 = \sigma^2.
\]

In many applications there is an interest in tying some covariates to the transition probabilities. This can be done by making the drift parameter \(\mu\) a function of some covariates. As before, let \(z\) be a row vector of covariate values. A link function tying \(z\) to \(\mu\) will have to enforce the constraint that \(|\mu| < \sigma^2\). For individual \(i\) with covariates \(z_i\), the function

\[
\mu_i = \sigma^2(2 \logit^{-1}(z_i\gamma) - 1),
\]
accomplishes this, where $\gamma$ is a column vector of regression coefficients and $\text{logit}^{-1}(\cdot)$ is the inverse logit function.

### 5.1.3 Hazard functions

The first hitting time probabilities for this model will only be nonzero on the positive integers since it is a discrete-time process. Also, there are two hitting time distributions since there are two absorbing states. As before, let $P(\tau_a = t | X_0 = x_0, b)$ be the probability that $X_t$ reaches boundary $a$ for the first time at $t$ after starting in state $x_0$. Let $F_a(t) = P(\tau_a \leq t | X_0 = x_0, b)$ be the cumulative probability at boundary $a$ at time $t$. Denote by $F_a(\infty)$ the probability that $X_t$ hits boundary $a$ before hitting boundary $b$ (and is therefore absorbed in state $a$), and let $F_{-b}(\infty)$ be the probability of absorption into state $-b$. Let $m_a(t) = \lfloor \frac{t-a}{2} \rfloor$, $k_1 = \lfloor \frac{t+a}{2(b_0+a)} \rfloor$, and $k_2 = \lfloor \frac{t-a}{2(b_0+a)} \rfloor$ as in Section 4.1.2. The first hitting time probabilities at $a$ using the parameterization of this section are

$$P(\tau_a = t | b = b_0) = P(\tau_a = t | b = \infty)$$

$$- 1_{\{t \geq a+2b_0\}} \sum_{j=1}^{k_1} P(\tau_{2j(a+b_0)-a} = t | b = \infty) \left( \frac{p_-}{p_+} \right)^{j(a+b_0)-a}$$

$$+ 1_{t \geq 3a+2b_0} \sum_{j=1}^{k_2} P(\tau_{2j(a+b_0)+a} = t | b = \infty) \left( \frac{p_-}{p_+} \right)^{j(a+b_0)},$$

where

$$P(\tau_a = t | b = \infty) = \frac{a}{t} \sum_{n=0}^{m_a(t)} \frac{t!}{(a+n)!n!(t-a-2n)!} \left( \frac{\sigma^2 + \mu}{2} \right)^{a+n} \left( \frac{\sigma^2 - \mu}{2} \right)^{n} (1-\sigma^2)^{t-a-2n}.$$
are

\[ F_a(\infty|X_0 = i) = \begin{cases} \frac{1 - \left(\frac{\sigma^2 - \mu}{\sigma^2 + \mu}\right)^{i+b}}{1 - \left(\frac{\sigma^2 - \mu}{\sigma^2 + \mu}\right)^{a+b}}, & \text{if } \mu \neq 0, \\ \frac{i+b}{a+b}, & \text{if } \mu = 0, \end{cases} \]

and

\[ F_{-b}(\infty|X_0 = i) = \begin{cases} \frac{1 - \left(\frac{\sigma^2 + \mu}{\sigma^2 - \mu}\right)^{a-i}}{1 - \left(\frac{\sigma^2 + \mu}{\sigma^2 - \mu}\right)^{a+b}}, & \text{if } \mu \neq 0, \\ \frac{a-i}{a+b}, & \text{if } \mu = 0. \end{cases} \]

See Section 4.1.2 for details. Two quantities of interest for survival analysis are the survival function and the hazard rate. The cumulative distribution \( F(t) \) for the hitting time is equal to the sum \( F(t) = F^{-b}(t) + F^a(t) \). Letting \( p^0(j) = P(X_0 = j) \), hitting time probabilities and cumulative distributions at each boundary can be written as

\[
P(\tau_a = t) = \sum_{j = -b+1}^{a-1} p^0(j) P(\tau_a = t|X_0 = j),
\]

\[
P(\tau_{-b} = t) = \sum_{j = -b+1}^{a-1} p^0(j) P(\tau_{-b} = t|X_0 = j),
\]

\[
F_a(t) = \sum_{j = -b+1}^{a-1} p^0(j) F_a(t|X_0 = j) = \sum_{j = -b+1}^{a-1} p^0(j) \sum_{k \leq t} P(\tau_a = k|X_0 = j, b),
\]

and

\[
F_{-b}(t) = \sum_{j = -b+1}^{a-1} p^0(j) F_{-b}(t|x_0 = j) = \sum_{j = -b+1}^{a-1} p^0(j) \sum_{k \leq t} P(\tau_{-b} = k|x_0 = j, a).
\]

The survival function is then \( S(t) = 1 - F(t) \). Since there are two events of interest, there are two hazard functions which will be denoted by \( h^{-b}(t) \) and \( h^a(t) \).

For \( t = 1, 2, 3, \ldots \),

\[
h_a(t) = \frac{P(\tau_a = t)}{S(t-1)} = \frac{P(\tau_a = t)}{1 - F^a(t-1) - F^{-b}(t-1)},
\]
and similarly for $-b$,

$$h_{-b}(t) = \frac{P(\tau_{-b} = t)}{S(t-1)} = \frac{P(\tau_{-b} = t)}{1 - F^a(t-1) - F^{-b}(t-1)}.$$
Table 5.1: Parameter values used to create the hazard functions in Figure 5.4

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$a = b = 5$</td>
<td>$a = b = 5$</td>
<td>$a = b = 5$</td>
</tr>
<tr>
<td>$z\gamma_0 = 0.41$</td>
<td>$z\gamma_0 = 0.00$</td>
<td>$z\gamma_0 = -1.73$</td>
</tr>
<tr>
<td>$\beta = 15$</td>
<td>$\beta = 15$</td>
<td>$\beta = 15$</td>
</tr>
<tr>
<td>$\sigma^2 = 0.5$</td>
<td>$\sigma^2 = 0.8$</td>
<td>$\sigma^2 = 0.8$</td>
</tr>
<tr>
<td>$z\gamma = 0$</td>
<td>$z\gamma = 0.25$</td>
<td>$z\gamma = 1.95$</td>
</tr>
<tr>
<td>$a = b = 5$</td>
<td>$a = b = 5$</td>
<td>$a = b = 5$</td>
</tr>
<tr>
<td>$z\gamma_0 = -0.85$</td>
<td>$z\gamma_0 = -0.58$</td>
<td>$z\gamma_0 = 0.20$</td>
</tr>
<tr>
<td>$\beta = 0.5$</td>
<td>$\beta = 0.5$</td>
<td>$\beta = 30$</td>
</tr>
<tr>
<td>$\sigma^2 = 0.8$</td>
<td>$\sigma^2 = 0.8$</td>
<td>$\sigma^2 = 0.2$</td>
</tr>
<tr>
<td>$z\gamma = 1.95$</td>
<td>$z\gamma = 0.79$</td>
<td>$z\gamma = 0$</td>
</tr>
<tr>
<td>$a = b = 5$</td>
<td>$a = b = 5$</td>
<td>$a = b = 5$</td>
</tr>
<tr>
<td>$z\gamma_0 = 0.20$</td>
<td>$z\gamma_0 = 2.20$</td>
<td>$z\gamma_0 = 1.39$</td>
</tr>
<tr>
<td>$\beta = 0.5$</td>
<td>$\beta = 1$</td>
<td>$\beta = 1$</td>
</tr>
<tr>
<td>$\sigma^2 = 0.4$</td>
<td>$\sigma^2 = 0.8$</td>
<td>$\sigma^2 = 0.8$</td>
</tr>
<tr>
<td>$z\gamma = 0$</td>
<td>$z\gamma = 0.51$</td>
<td>$z\gamma = -0.51$</td>
</tr>
</tbody>
</table>

at the lower boundary. For simplicity, parameters are chosen so that the hazard at the upper boundary is greater than the hazard at the lower boundary; as a result the black line is the top line in each plot. The hazard functions can be increasing, decreasing, tub-shaped, or hump-shaped. It is possible for the hazard functions for the two boundaries to cross.

### 5.2 Bayesian Formulation

To estimate the model that was proposed in section 5.1, a Bayesian formulation will now be presented. (Parameter estimation can also be carried out via maximum likelihood, although numerical optimization will be necessary as explicit formulas for the MLEs are not available.)
5.2.1 Prior selection

For the distribution of $X_0$, priors must be selected for two of $\gamma_0$, $\alpha$, and $\beta$. Here the focus will be on $\gamma_0$ and $\beta$. Note that the prior distribution for $\gamma_0$, $\pi(\gamma_0)$, implies a prior distribution on $\xi$ (recall that $\text{logit}(\xi) = z\gamma_0$). Since $\xi$ is an easier parameter to interpret, the focus will be on choosing a prior for $\gamma_0$ that results in an appropriate prior on $\xi$. Let $\gamma_0 \sim N(0, \Sigma)$, where $0$ is a vector of zeros and $\Sigma$ is positive definite. Then note that, for a fixed set of covariates $z$, $z\gamma_0 \sim N(0, \delta^2)$ where $\delta^2 = \sum_i \sum_j z_i z_j \text{cov}(\gamma_i, \gamma_j)$. The resulting prior on $\xi$ has density

$$
\pi(\xi) = \frac{1}{\sqrt{2\pi\delta^2}} \left( \frac{1}{\xi} + \frac{1}{1-\xi} \right) \exp \left( -\frac{1}{2\delta^2} \left( \text{log} \left( \frac{\xi}{1-\xi} \right) \right)^2 \right), \quad 0 < \xi < 1.
$$

Figure 5.5 shows the effect of $\delta^2$ on $\pi(\xi)$. If $\delta^2$ is large then there is a high prior probability that $\xi$ will be very close to either zero or one. So if $z\gamma_0$ is given a large variance, this does not indicate an “uninformative” prior on $\xi$ but instead puts very large prior probability on values of $\xi$ near zero and one. It was mentioned in Section 5.1.1 that values of $z\gamma_0$ that fall between $-3$ and 3 will result in values of $\xi$ in the interval $(0.05, 0.95)$, a suitable range for most applications. Given knowledge of the range of values taken on by the covariate vector $z$, $\Sigma$ can be chosen to give high prior probability to $z\gamma_0 \in (-3, 3)$.

The value of $\beta$ will have an effect on the shape and spread of the beta distribution. If $\beta$ is small, then the distribution will be U-shaped or J-shaped. If it is large, the distribution will be unimodal and very concentrated. Figure 5.6 shows the beta distribution for several combinations of $\xi$ and $\beta$. Each row corresponds to a fixed value of $\xi$ and each column corresponds to a fixed value of $\beta$. A large $\beta$ will imply a small variance for $X_0$, which means that a diffuse prior on $\beta$ (putting high probability
Figure 5.5: The effect of the prior variance of $z\gamma_0$ on the prior for $\xi$.

on $\beta$ being “large”) implies a very concentrated distribution for $X_0$. As a result the prior for $\beta$ should be carefully chosen to put most of its mass on an appropriate region. One option is to specify a gamma distribution, with parameters chosen to give the distribution of $X_0$ desired properties. For example, if a unimodal prior for $X_0$ is desired, the gamma distribution with shape parameter between 10 and 16 and rate parameter equal to 1 is an appropriate choice. A shape parameter of 10 puts 95% of the mass between 5 and 17, and a shape parameter of 16 puts 95% of the mass between 9 and 25. This will result in a unimodal distribution (as long as $\xi$ is not extremely close to 0 or 1). A smaller shape parameter (such as 10) will give a more variable prior on $X_0$ while a larger shape parameter (such as 16) will result in a more concentrated prior for $X_0$. 

86
Figure 5.6: Beta distributions for several values of \( \xi \) and \( \beta \).

Priors must also be selected for the parameters of the transition probability matrix, \( \gamma \) and \( \sigma^2 \). Since \( \sigma^2 \in (0, 1) \), a beta distribution is a natural choice. A normal distribution can be used for \( \gamma \). Recall that

\[
\mu = \sigma^2 (2 \text{logit}^{-1}(z\gamma) - 1),
\]

where \( \text{logit}^{-1}(\cdot) \) is the inverse logit function, and that \(|\mu| < \sigma^2 \). The function \( \text{logit}^{-1}(\cdot) \) has a steep slope in the interval \((-2, 2)\) and changes very little for values that are less than \(-3\) or greater than \(3\). So, keeping in mind the scale of the covariates \( z \), the variance of \( z\gamma \) should be chosen to be small enough that \( z\gamma \in (-3, 3) \) with high
probability. A diffuse prior for $\gamma$ puts high prior probability on $\mu \approx \sigma^2$ and $\mu \approx -\sigma^2$.

5.2.2 Likelihood

Once the prior distributions for $\gamma_0$, $\beta$, $\gamma$, and $\sigma^2$ have been selected, a Metropolis-Hastings algorithm can be used to fit the model. The likelihood function is easily expressed as a product of matrices. Let $t = (t_1, t_2, \ldots, t_n)$ be a sample of hitting times for $n$ individuals, let $d = (d_1, d_2, \ldots, d_n)$ be defined as

$$d_i = \begin{cases} 
2, & \text{if individual } i \text{ hit boundary } a, \\
1, & \text{if individual } i \text{ hit boundary } -b,
\end{cases}$$

and let $z_i$ be the set of covariates for individual $i$. Let $q_i$ be the row vector of initial probabilities for $X_{0i}$ and let

$$Q_i = \begin{pmatrix} 
1 - \sigma^2 & \frac{1}{2}(\sigma^2 + \mu_i) & 0 & \ldots & 0 \\
\frac{1}{2}(\sigma^2 - \mu_i) & 1 - \sigma^2 & \frac{1}{2}(\sigma^2 + \mu_i) & \ldots & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
0 & 0 & 0 & \ldots & 1 - \sigma^2
\end{pmatrix}$$

be the part of the transition probability matrix for individual $i$ that corresponds to the transient states $(-b + 1, \ldots, a - 1)$. Recall that $\mu_i = \sigma^2(2 \logit^{-1}(z_i\gamma) - 1)$ and $q_i$ is determined by $z_i$, $\gamma_0$, and $\beta$. Then the likelihood function is

$$L(\gamma_0, \beta, \gamma, \sigma^2; t, d) = \prod_{i=1}^{n} q_i Q_i^{t_i - 1} r_i,$$

where

$$r_i = \begin{cases} 
(0, 0, \ldots, 0, \frac{(\sigma^2 + \mu_i)}{2})', & \text{if } d_i = 2, \\
\left(\frac{(\sigma^2 - \mu_i)}{2}, 0, \ldots, 0\right)', & \text{if } d_i = 1.
\end{cases}$$
5.3 Extensions

The model presented in this chapter can be extended in a number of ways to accommodate censoring, allow for a sudden jump to a boundary state, and model the effects of time-varying covariates.

5.3.1 Censoring

To account for right-censored observations, an adjustment can be made to the likelihood function presented in the previous section. Let \( q_i \) and \( Q_i \) be defined as before, and let \( t \) be a vector of times that may be either event times (hitting times at \( a \) or \( -b \)) or censoring times. Redefine \( d = (d_1, \ldots, d_n) \) as

\[
d_i = \begin{cases} 
2, & \text{if individual } i \text{ hit boundary } a, \\
1, & \text{if individual } i \text{ hit boundary } -b, \\
0, & \text{if individual } i \text{ did not hit either boundary}, 
\end{cases}
\]

and let

\[
r_i = \begin{cases} 
(0, 0, \ldots, 0, \frac{(\sigma^2 + \mu_i)}{2})', & \text{if } d_i = 2, \\
\left(\frac{(\sigma^2 - \mu_i)}{2}, 0, \ldots, 0\right)', & \text{if } d_i = 1, \\
\left(1 - \frac{(\sigma^2 - \mu_i)}{2}, 1, \ldots, 1 - \frac{(\sigma^2 + \mu_i)}{2}\right)', & \text{if } d_i = 0.
\end{cases}
\]

As before, the likelihood function is

\[
L(\gamma_0, \beta, \gamma, \sigma^2; t, d) = \prod_{i=1}^{n} q_i Q_i^{t_i-1} r_i,
\]

with the only difference being the definition of \( r_i \). The likelihood contribution of individual \( i \) with initial state distribution \( q_i \) and transient state transition probability matrix \( Q_i \) will therefore be:

- the probability of hitting the upper boundary for the first time at \( t_i \) if \( d_i = 2 \),
- the probability of hitting the lower boundary for the first time at \( t_i \) if \( d_i = 1 \),
• the probability of reaching time $t_i$ without being absorbed into either boundary if $d_i = 0$ (the survival function).

5.3.2 Sudden jump to a boundary

In some cases, it is sensible to consider the possibility that, regardless of what state the process is currently in, it may jump to a boundary in the next transition. For instance, while a person’s health status may change by a small amount from day to day most of the time, there is a possibility that their health will suddenly deteriorate. For example, in the Worcester Heart Attack Study, a person who is going through the normal stages of recovery after a heart attack may unexpectedly have another heart attack that proves fatal. In order for the model to accommodate this possibility, the transition probability matrix must be adjusted. Consider the following adjustment to the transition matrix $P$

$$
P = \begin{pmatrix}
-b & -b + 1 & -b + 2 & \cdots & a \\
1 & 0 & 0 & \cdots & 0 \\
\frac{1}{2}(\sigma^2 - \mu) & 1 - \sigma^2 & \frac{1}{2}(\sigma^2 + \mu) & \cdots & 0 \\
r & \frac{1}{2}(\sigma^2 - \mu)(1 - r) & (1 - \sigma^2)(1 - r) & \cdots & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
a & 0 & 0 & \cdots & 1
\end{pmatrix},
$$

where $r$ is the probability of jumping to the lower boundary from a non-adjacent state. For estimation, a prior will need to be specified for $r$; since $r \in (0, 1)$ the beta distribution is a natural choice.

Of course, depending on the application, it is also possible for the transition probability matrix to have a more general form, beyond the parameterization that is presented in this chapter.
5.3.3 Time-varying covariates

In some cases, the covariates $z_i$ are measurements that change over time. In this case, each individual will have a covariate matrix $Z_i = (z^0_i, z^1_i \ldots, z^t_i)$, where $z^t_i$ is the covariate vector for individual $i$ at time $t$. Then, let $q_i$ be determined as before using logit($\xi$) = $z^0_i \gamma_0$ and let $\mu_i(t) = \sigma^2(2\text{logit}^{-1}(z^t_i \gamma) - 1)$. Also, let

$$Q_i(t) = \begin{pmatrix}
1 - \sigma^2 & \frac{1}{2}(\sigma^2 + \mu_i(t)) & 0 & \ldots & 0 \\
\frac{1}{2}(\sigma^2 - \mu_i(t)) & 1 - \sigma^2 & \frac{1}{2}(\sigma^2 + \mu_i(t)) & \ldots & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
0 & 0 & 0 & \ldots & 1 - \sigma^2
\end{pmatrix}.$$

With time-varying covariates, the likelihood functions is

$$L(\gamma_0, \beta, \gamma, \sigma^2; t, d) = \prod_{i,d_i=1}^{t} q_i \left( \prod_{j=1}^{t_i} Q_i(j) \right) r_i.$$ 

This likelihood is fairly easy to compute, especially when compared to a similar construction with a continuous process such as Brownian motion. For Brownian motion, since the value of $X_t$ is not observed at any point before the event time, it is necessary to integrate over the distribution $f(x_t|\gamma_0, \beta, \gamma, \sigma^2, \tau_a > t, \tau_b > t)$. This can be accomplished numerically for one or very few covariate changes and a small number of parameters, but becomes computationally difficult for larger numbers of parameters and covariate changes.
Chapter 6: Simulations and Real Data Applications

To illustrate how the discrete threshold regression (DTR) model works, a simulated data study will be presented first, followed by the analysis of two real data sets. The purpose of the simulation study is to demonstrate the properties of the model and provide some intuition on how to make certain modeling choices. The real data examples will be presented starting in Section 6.2.

6.1 Simulated Data

The simulation study in this section illustrates the effect of fitting the DTR model to a given data set using different numbers of transient states. The number of states in the discrete threshold model impacts the complexity of the model and is an important modeling choice. In the extreme case when only one transient state is used, all sample paths that are absorbed into the same boundary at the same time have taken exactly the same path (they stayed in the one transient state for the same number of transitions and then moved into the boundary state). This model lacks complexity and does not allow much varied behavior for individuals with different covariate values. A model with several transient states, instead, allows for some differentiation between individuals. For example, suppose that two individuals hit the upper boundary at the same time. One of them may have started close to the upper boundary with a
shallow drift so that it took a while to be absorbed, while the other may have started close to the lower boundary but with a steep upward drift that drew them toward the upper boundary rather than the lower boundary. The number of states used for a given data set should be large enough to provide sufficient complexity. However, a larger number of states is not always preferable—choosing too many states can lead to a poor fit. In practice there will typically be a range of states that will provide an appropriate amount of complexity to fit a given data set. The simulation study in this section explores data sets simulated from the DTR model with three different values of $\sigma^2$ (0.25, 0.5, and 0.9) and three different sample sizes ($n=200$, $500$, and $1000$). There are a total of nine combinations of $\sigma^2$ and $n$; for each combination a total of 20 simulated data sets were created.

For all simulated datasets, the boundary states were defined by $a = b = 5$, yielding nine transient states and 11 total states. The remaining parameters, aside from $\sigma^2$, were set at $\gamma_0 = (0.25, -0.25)'$, $\beta = 10$, and $\gamma = (0.25, -0.5)'$. The covariate vector $z_0$ for the starting state distribution consisted of an intercept term (equal to 1 for all individuals) and an indicator variable taking values 0 and 1. The covariate vector $z_t$ for the transition probabilities also consisted of an intercept term and an indicator variable taking values 0 and 1. None of the covariates varied in time. For each sample size, the sample was constructed to yield the following design, where $i$ is the subject index, $i = 1, \ldots, n$.

- 25% of the sample had $z_{0i} = (1, 0)$ and $z_{ti} = (1, 0)$,
- 25% of the sample had $z_{0i} = (1, 0)$ and $z_{ti} = (1, 1)$,
- 25% of the sample had $z_{0i} = (1, 1)$ and $z_{ti} = (1, 0)$,
• 25% of the sample had $z_{0i} = (1, 1)$ and $z_{di} = (1, 1)$.

For convenience, these four groups will be called “0-0”, “0-1”, “1-0”, and “1-1”, respectively. Once the samples were simulated, the DTR model with $a=b$ was fit using 3, 5, 7, 11, and 15 total states (including both transient and absorbing). Since 11 is the “correct” number of states, the corresponding model can be expected to provide a good fit to the data. The use of other numbers of states will illustrate the effect of state space misspecification.

![Initial state distribution and hazard functions for the simple simulation example with $\sigma^2 = 0.5$; upper boundary hazard in black and lower boundary hazard in red.](image)

Figure 6.1: Initial state distribution and hazard functions for the simple simulation example with $\sigma^2 = 0.5$; upper boundary hazard in black and lower boundary hazard in red.
Figure 6.1 shows the initial state distributions and hazard functions for each combination of covariate values for $\sigma^2 = 0.5$. Table 6.1 shows the hitting probabilities $F_a(\infty)$ and $F_{-b}(\infty)$ for each of the four groups. Note that the value of $\sigma^2$ does not impact these probabilities, so they are the same for all simulations. To see that this is true, let $\tau_a$ be the first hitting time at boundary $a$ and let $\tau_a = \infty$ if $X_t$ is absorbed into state $-b$. The probability of hitting boundary $a$ is equal to $P(\tau_a < \infty)$. Recall from Section 4.1.2 that

$$ P(\tau_a < \infty | X_0 = x_0) = \begin{cases} \frac{x_0 + b}{a + b} & \text{if } p_+ = p_- \\ \frac{1 - (\frac{p_-}{p_+})^{x_0 + b}}{1 - (\frac{p_-}{p_+})^{a + b}} & \text{if } p_+ \neq p_- \end{cases} $$

In the case when $p_+ = p_-$, the probability of hitting boundary $a$ depends on $a$ and $b$, which here are fixed to be $a = b = 5$, and on $X_0$, which does not depend on $\sigma^2$. In the case when $p_+ \neq p_-$, note that

$$ \frac{1 - \left(\frac{p_-}{p_+}\right)^{x_0 + b}}{1 - \left(\frac{p_-}{p_+}\right)^{a + b}} = \frac{1 - \left(\frac{\sigma^2 - \mu}{\sigma^2 + \mu}\right)^{x_0 + b}}{1 - \left(\frac{\sigma^2 - \mu}{\sigma^2 + \mu}\right)^{a + b}}, $$

which only depends on the transition probabilities through the ratio $\frac{\sigma^2 - \mu}{\sigma^2 + \mu}$. Since $\mu = \sigma^2(\logit^{-1}(z\gamma) - 1)$, where $\logit^{-1}(\cdot)$ is the inverse logit function, the ratio reduces to

$$ \frac{\sigma^2 - \sigma^2(\logit^{-1}(z\gamma) - 1)}{\sigma^2 + \sigma^2(\logit^{-1}(z\gamma) - 1)} = \frac{\sigma^2(1 - (\logit^{-1}(z\gamma) - 1))}{\sigma^2(1 + (\logit^{-1}(z\gamma) - 1))} = \frac{1 - (\logit^{-1}(z\gamma) - 1)}{1 + (\logit^{-1}(z\gamma) - 1)}, $$

which does not depend on $\sigma^2$. Since the distribution of $X_0$ also does not depend on $\sigma^2$, $P(\tau_a < \infty) = \sum_{x=-b+1}^{a-1} P(X_0 = x) P(\tau_a < \infty | X_0 = x)$ does not depend on $\sigma^2$. Thus, $\sigma^2$ does not impact the boundary hitting probabilities.

A Metropolis-Hastings algorithm was used for estimation with each data set and each number of states, with 1000 iterations of adaptive burn-in followed by 10,000
iterations that were regarded as a sample from the posterior distribution. Following the discussion in section 5.2.1, the prior distributions were

\[ \gamma_0 \sim N(0, I), \]

\[ \beta \sim Gamma(10, 1), \]

\[ \gamma \sim N(0, I), \]

\[ \sigma^2 \sim Beta(1, 1), \]

where \( \mathbf{0} \) is a vector of zeros and \( I \) is the identity matrix.

It is known that the data in this simulation study were generated with 11 states. However, in practice, the most appropriate number of states may not be obvious. To determine which numbers of states provide an appropriate amount of complexity, a comparison will be made on the basis of two criteria: The deviance information criterion (DIC) (Spiegelhalter, Best, and Carlin 1998) and the ability of the model to estimate the first hitting time distributions.

DIC is a measure of model fit that is based on the deviance. For a vector of parameters \( \theta \) and a random variable \( Y \), let

\[ D(\theta) = -2 \log p(y|\theta) + 2 \log f(y), \]

Table 6.1: Hitting probabilities at each boundary in simple simulation.

<table>
<thead>
<tr>
<th>Group</th>
<th>( F_a(\infty) )</th>
<th>( F_{-b}(\infty) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0-0</td>
<td>0.81</td>
<td>0.19</td>
</tr>
<tr>
<td>0-1</td>
<td>0.28</td>
<td>0.72</td>
</tr>
<tr>
<td>1-0</td>
<td>0.77</td>
<td>0.23</td>
</tr>
<tr>
<td>1-1</td>
<td>0.23</td>
<td>0.77</td>
</tr>
</tbody>
</table>

96
a term Spiegelhalter et al. (1998) refer to as the “Bayesian deviance.” Here, \( p(y|\theta) \) is proportional to the likelihood function of \( Y \) and \( f(y) \) is a standardizing term that is a function of the data alone. Denote the posterior expectation of the deviance by \( \bar{D} = E_{\theta|y}(D) \) and let
\[
p_D = \bar{D} - D(E_{\theta|y}(\theta)),
\]
where \( p_D \) is referred to as the “effective number of parameters” and is taken to be a measure of a model’s complexity. Then
\[
DIC = \bar{D} + p_D.
\]

To quantify the ability of each model to estimate the first hitting time distributions, posterior estimates of the following are presented: probabilities of hitting each boundary, median first hitting times at each boundary, and first and third quartiles of first hitting times at each boundary. Figures 6.2 and 6.3 show example trace plots of the model parameters and posterior densities for the 11 state model.

Discussion of the simulation results will focus on group 0-0. When \( \sigma^2 \) is fairly small, as in the case of \( \sigma^2 = 0.25 \), the models with too few states (3 and 5) perform poorly while all other models with more states work fairly well. Table 6.2 shows the estimated hitting probabilities for this case. The average estimates are shown in plain text and the standard deviations of the estimates (measured across the 20 replicate datasets) are shown in italics. The first row of the table shows the true probabilities, based on the parameter values that were used for simulation, and the bold numbers in the body of the table are estimates that are more than one standard deviation away from the true hitting probabilities. That is, bolding in the body of the table is used to highlight poor estimates. Both the 3 and 5 state models tend to underestimate the
probability of hitting the upper boundary and overestimate the probability of hitting the lower boundary. This trend holds true as the standard deviations decrease due to increasing sample size, making it clear that the bias is systematic.

Table 6.3 shows the first, second (median), and third quartiles for hitting times at boundaries $a$ and $-b$. Again, the first line gives the true values based on the parameters used for simulation and the boldface numbers in the body of the table are more than one standard deviation away from them. As before, the 3 and 5 state models do not fit the data well. The other models with more states perform well. The 3 state model seems to do well at fitting one quartile of each distribution at the expense of the others. For the upper boundary, the median is estimated well but the first quartile is too small and the third quartile is too large. For the lower boundary, the third quartile is estimated well but the first two are too small. This model is too restrictive to accurately match all of the quantiles—in order to estimate the median
Figure 6.3: Posterior distributions of the model parameters for 11 states and $\sigma^2 = 0.25$.

well, the posterior favors parameter values that make the hitting time distribution at the upper boundary too dispersed.
<table>
<thead>
<tr>
<th>n</th>
<th>States</th>
<th>( \hat{F}_a(\infty) )</th>
<th>( \hat{F}_{-b}(\infty) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>200</td>
<td>3</td>
<td>0.7548</td>
<td>0.2452</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.0362</td>
<td>0.0365</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>0.7882</td>
<td>0.2118</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.0359</td>
<td>0.0359</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>0.8023</td>
<td>0.1977</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.0373</td>
<td>0.0373</td>
</tr>
<tr>
<td></td>
<td>11</td>
<td>0.8017</td>
<td>0.1983</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.0385</td>
<td>0.0385</td>
</tr>
<tr>
<td></td>
<td>15</td>
<td>0.8016</td>
<td>0.1984</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.0384</td>
<td>0.0383</td>
</tr>
<tr>
<td>500</td>
<td>3</td>
<td><strong>0.7771</strong></td>
<td><strong>0.2229</strong></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.0271</td>
<td>0.0272</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>0.7987</td>
<td>0.2013</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.0256</td>
<td>0.0255</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>0.8097</td>
<td>0.1903</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.0253</td>
<td>0.0252</td>
</tr>
<tr>
<td></td>
<td>11</td>
<td>0.8091</td>
<td>0.1909</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.0247</td>
<td>0.0246</td>
</tr>
<tr>
<td></td>
<td>15</td>
<td>0.8076</td>
<td>0.1924</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.0236</td>
<td>0.0235</td>
</tr>
<tr>
<td>1000</td>
<td>3</td>
<td><strong>0.7777</strong></td>
<td><strong>0.2223</strong></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.0126</td>
<td>0.0128</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td><strong>0.7942</strong></td>
<td><strong>0.2058</strong></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.0141</td>
<td>0.0141</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>0.8085</td>
<td>0.1915</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.0121</td>
<td>0.0121</td>
</tr>
<tr>
<td></td>
<td>11</td>
<td>0.8071</td>
<td>0.1929</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.0128</td>
<td>0.0129</td>
</tr>
<tr>
<td></td>
<td>15</td>
<td>0.8066</td>
<td>0.1934</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.0128</td>
<td>0.0128</td>
</tr>
</tbody>
</table>

Table 6.2: Boundary hitting probability estimates for \( \sigma^2 = 0.25 \).
<table>
<thead>
<tr>
<th>$n$</th>
<th>States</th>
<th>$Q_{1}^{a}$</th>
<th>$Q_{2}^{a}$</th>
<th>$Q_{3}^{a}$</th>
<th>$Q_{1}^{b}$</th>
<th>$Q_{2}^{b}$</th>
<th>$Q_{3}^{b}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>200</td>
<td>3</td>
<td>25.1</td>
<td>1.1</td>
<td>59.6</td>
<td>2.7</td>
<td>118.5</td>
<td>5.3</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>30.9</td>
<td>2.7</td>
<td>61.5</td>
<td>4.0</td>
<td>112.6</td>
<td>6.0</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>30.4</td>
<td>2.8</td>
<td>58.9</td>
<td>4.6</td>
<td>108.2</td>
<td>6.7</td>
</tr>
<tr>
<td></td>
<td>11</td>
<td>30.9</td>
<td>3.2</td>
<td>58.7</td>
<td>4.7</td>
<td>106.9</td>
<td>6.9</td>
</tr>
<tr>
<td></td>
<td>15</td>
<td>31.0</td>
<td>3.1</td>
<td>58.5</td>
<td>4.7</td>
<td>106.5</td>
<td>6.7</td>
</tr>
<tr>
<td>500</td>
<td>3</td>
<td>24.8</td>
<td>0.7</td>
<td>58.9</td>
<td>1.9</td>
<td>117.1</td>
<td>3.6</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>31.4</td>
<td>2.1</td>
<td>61.8</td>
<td>3.0</td>
<td>111.6</td>
<td>4.2</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>30.4</td>
<td>2.1</td>
<td>58.7</td>
<td>3.2</td>
<td>107.3</td>
<td>4.6</td>
</tr>
<tr>
<td></td>
<td>11</td>
<td>31.0</td>
<td>2.1</td>
<td>58.4</td>
<td>3.1</td>
<td>106.2</td>
<td>4.6</td>
</tr>
<tr>
<td></td>
<td>15</td>
<td>31.2</td>
<td>1.9</td>
<td>58.6</td>
<td>3.1</td>
<td>106.5</td>
<td>4.8</td>
</tr>
<tr>
<td>1000</td>
<td>3</td>
<td>24.8</td>
<td>0.7</td>
<td>59.0</td>
<td>1.5</td>
<td>117.1</td>
<td>3.1</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>31.3</td>
<td>1.6</td>
<td>61.7</td>
<td>1.9</td>
<td>111.7</td>
<td>2.9</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>30.0</td>
<td>1.5</td>
<td>58.3</td>
<td>2.1</td>
<td>107.0</td>
<td>3.0</td>
</tr>
<tr>
<td></td>
<td>11</td>
<td>30.6</td>
<td>1.5</td>
<td>58.1</td>
<td>1.9</td>
<td>106.1</td>
<td>2.7</td>
</tr>
<tr>
<td></td>
<td>15</td>
<td>30.6</td>
<td>1.6</td>
<td>58.2</td>
<td>2.1</td>
<td>106.2</td>
<td>3.1</td>
</tr>
</tbody>
</table>

Table 6.3: Quantile estimates of the hitting time distributions for $\sigma^2 = 0.25$. 
Table 6.4: Deviance information criterion (DIC) and effective number of parameters ($p_D$) for $\sigma^2 = 0.25$.

For the various models, Table 6.4 shows the DIC and effective number of parameters, $p_D$, for $\sigma^2 = 0.25$. Lower DIC values indicate a better fit. The effective number of parameters is a measure of model complexity that is used as a “penalty” term in the calculation of the DIC (Spiegelhalter, Best, and Carlin 1998). For all three sample sizes, the DIC is very similar for the models with 11 and 15 states, and only slightly higher for the model with 7 states. The DIC is higher for the models with 3 and 5 states, indicating a comparatively poor fit. The complexity, as measured by $p_D$, is a little smaller than the actual number of model parameters for all numbers of states. Note that there is no initial state distribution in the 3 state model, so there are only 3 parameters in that model compared to the 6 parameters in the other models.
To explain why the 3 state model underperformed and the 15 state model performed so well in this case, Figure 6.4 shows for one data set the posterior distribution of $\sigma^2$ for 3, 11, and 15 states. The data was simulated with $\sigma^2 = 0.25$ and 11 states, and the 11 state model correctly estimates $\sigma^2$ to be close to 0.25. When there are only 3 states, the estimated value of $\sigma^2$ is close to 0.012. Because there are fewer transient states, the 3 state model must restrict the movement of the latent Markov chain, making it less “mobile” (i.e., keeping it longer in the only transient state) in an effort to match the quantiles of the hitting time distributions. On the other hand, the 15 state model estimated $\sigma^2$ to be around 0.5. Because there are more transient states, the 15 state model must allow the latent Markov chain to be more mobile than the 11 state model in order to match the quantiles of the hitting time distributions. Since $\sigma^2$ was only 0.25, there is plenty of room in the parameter space to inflate $\sigma^2$ in the 15 state model and still fit the data well. It is worth noting that fitting a model with a larger number of states increases the computational burden. Since the models with 11 and 15 states performed very similarly here, the 11 state model should be preferred for this reason.

![Figure 6.4: Posterior distribution of $\sigma^2$ for models with 3, 11, and 15 states when the true value of $\sigma^2$ is 0.25.](image)

103
When $\sigma^2$ is a little larger, as is the case when $\sigma^2 = 0.5$, again the 3 and 5 state models perform poorly while the other models fit the data quite well. Table 6.5 shows the boundary hitting probabilities for $\sigma^2 = 0.5$. The true values in the first row are the same as for $\sigma^2 = 0.25$ because $\sigma^2$ has no impact on the probability of hitting either boundary. As Table 6.5 shows, the 7, 11, and 15 state models estimate $F_a(\infty)$ and $F_{-b}(\infty)$ well, especially when the sample size is large. As before, the 3 state model tends to underestimate $F_a(\infty)$ and overestimate $F_{-b}(\infty)$. The 5 state model appears to show this same tendency, but the bias is not as large and is less than one standard deviation for all three sample sizes.
Table 6.6: Quantile estimates of the hitting time distributions for $\sigma^2 = 0.5$. 

<table>
<thead>
<tr>
<th>$n$</th>
<th>States</th>
<th>$Q_{1}^{a}$</th>
<th>$Q_{2}^{a}$</th>
<th>$Q_{3}^{a}$</th>
<th>$Q_{1}^{b}$</th>
<th>$Q_{2}^{b}$</th>
<th>$Q_{3}^{b}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>200</td>
<td>3</td>
<td>12.4</td>
<td>0.9</td>
<td>29.3</td>
<td>1.9</td>
<td>58.3</td>
<td>4.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>17.5</td>
<td>1.4</td>
<td>33.2</td>
<td>2.4</td>
<td>58.8</td>
<td>4.1</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>15.1</td>
<td>1.8</td>
<td>30.1</td>
<td>2.9</td>
<td>53.5</td>
<td>4.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>19.6</td>
<td>1.9</td>
<td>35.0</td>
<td>3.1</td>
<td>60.2</td>
<td>4.8</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>15.3</td>
<td>1.7</td>
<td>28.9</td>
<td>2.8</td>
<td>52.7</td>
<td>4.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>19.9</td>
<td>1.9</td>
<td>34.8</td>
<td>3.2</td>
<td>59.6</td>
<td>4.7</td>
</tr>
<tr>
<td></td>
<td>11</td>
<td>16.0</td>
<td>1.4</td>
<td>30.3</td>
<td>2.3</td>
<td>55.1</td>
<td>3.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>20.6</td>
<td>2.0</td>
<td>36.4</td>
<td>2.8</td>
<td>62.2</td>
<td>4.0</td>
</tr>
<tr>
<td></td>
<td>15</td>
<td>5</td>
<td>15.4</td>
<td>0.6</td>
<td>29.2</td>
<td>1.0</td>
<td>53.2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>20.5</td>
<td>1.2</td>
<td>35.0</td>
<td>1.5</td>
<td>59.8</td>
<td>2.0</td>
</tr>
</tbody>
</table>

| 500 | 3      | 12.6        | 0.5         | 29.7        | 1.1         | 59.2        | 2.3         |
|     |        | 18.6        | 1.0         | 33.7        | 1.6         | 58.9        | 2.6         |
|     | 5      | 15.7        | 1.1         | 30.9        | 1.5         | 55.9        | 2.3         |
|     |        | 18.0        | 1.0         | 33.7        | 1.6         | 58.9        | 2.6         |
|     | 7      | 15.0        | 1.2         | 29.0        | 1.6         | 53.3        | 2.4         |
|     |        | 20.4        | 1.2         | 36.2        | 1.8         | 61.7        | 2.5         |
|     | 11     | 15.3        | 1.1         | 28.7        | 1.8         | 52.9        | 2.5         |
|     |        | 20.5        | 1.2         | 36.0        | 1.7         | 60.9        | 2.4         |
|     | 15     | 15.6        | 1.1         | 29.4        | 1.4         | 54.0        | 2.1         |
|     |        | 20.5        | 1.3         | 36.2        | 1.7         | 61.7        | 2.2         |

| 1000| 3      | 12.6        | 0.5         | 29.6        | 0.8         | 58.8        | 1.3         |
|     |        | 17.6        | 0.6         | 32.8        | 1.0         | 57.5        | 1.4         |
|     | 5      | 16.3        | 0.6         | 31.1        | 0.8         | 55.6        | 1.3         |
|     |        | 17.6        | 0.6         | 32.8        | 1.0         | 57.5        | 1.4         |
|     | 7      | 15.1        | 0.8         | 28.9        | 1.1         | 52.9        | 1.4         |
|     |        | 19.9        | 1.0         | 35.4        | 1.5         | 60.2        | 2.2         |
|     | 11     | 15.4        | 0.6         | 28.7        | 1.0         | 52.6        | 1.5         |
|     |        | 20.1        | 1.1         | 35.1        | 1.6         | 59.4        | 2.1         |
|     | 15     | 15.5        | 0.6         | 29.2        | 1.0         | 53.2        | 1.4         |
|     |        | 20.0        | 1.2         | 35.0        | 1.5         | 59.8        | 2.0         |
Table 6.7: Deviance information criterion (DIC) and effective number of parameters ($p_D$) for $\sigma^2 = 0.5$.

<table>
<thead>
<tr>
<th>n</th>
<th>States</th>
<th>DIC</th>
<th>$p_D$</th>
</tr>
</thead>
<tbody>
<tr>
<td>200</td>
<td>3</td>
<td>2119.2</td>
<td>2.87</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>2090.5</td>
<td>5.17</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>2083.0</td>
<td>5.27</td>
</tr>
<tr>
<td></td>
<td>11</td>
<td>2079.9</td>
<td>5.23</td>
</tr>
<tr>
<td></td>
<td>15</td>
<td>2079.7</td>
<td>4.78</td>
</tr>
<tr>
<td>500</td>
<td>3</td>
<td>5287.9</td>
<td>40.1</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>5211.1</td>
<td>45.8</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>5194.9</td>
<td>48.7</td>
</tr>
<tr>
<td></td>
<td>11</td>
<td>5191.2</td>
<td>50.3</td>
</tr>
<tr>
<td></td>
<td>15</td>
<td>5191.0</td>
<td>50.3</td>
</tr>
<tr>
<td>1000</td>
<td>3</td>
<td>10560.7</td>
<td>50.4</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>10400.2</td>
<td>53.7</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>10370.2</td>
<td>54.6</td>
</tr>
<tr>
<td></td>
<td>11</td>
<td>10363.2</td>
<td>56.2</td>
</tr>
<tr>
<td></td>
<td>15</td>
<td>10362.7</td>
<td>55.7</td>
</tr>
</tbody>
</table>

Table 6.6 shows the quantile estimates for $\sigma^2 = 0.5$. As before, the 3 and 5 state models perform poorly while the 11 and 15 state models provide accurate quantile estimates. The 3 state model is able to estimate $Q^a_2$ accurately but underestimates $Q^a_1$ and overestimates $Q^a_3$. At the lower boundary, the 3 state model estimates $Q^b_3$ well but underestimates $Q^b_1$ and $Q^b_2$. The quantile estimates from the 5 state model are only slightly better than the 3 state model. The 7 state model generally does well, but tends to underestimate $Q^a_1$. The 11 and 15 state models both estimate the quantiles accurately.

Table 6.7 shows the DIC and effective number of parameters, $p_D$, for $\sigma^2 = 0.5$. The DIC is similar for the 11 and 15 state models, a little higher for the 7 state model, and a good deal higher for the 3 and 5 state models as before. The effective number
of parameters tends to be a little lower than the number of parameters in the model for all models, as was the case for $\sigma^2 = 0.25$.

Figure 6.5: Posterior distribution of $\sigma^2$ for models with 3, 11, and 15 states when the true value of $\sigma^2$ is 0.5.

The posterior distributions for $\sigma^2$ for 3, 11, and 15 states are shown in figure 6.5. When the data were simulated from an 11 state model with $\sigma^2 = 0.25$, there was plenty of room in the parameter space for the 15 state model to inflate $\sigma^2$, making the latent Markov chain more mobile and fitting the data well. With data simulated from an 11 state model with $\sigma^2 = 0.5$ there is less room in the parameter space for the 15 state model to inflate $\sigma^2$. As figure 6.5 shows, this results in a posterior distribution of $\sigma^2$ that is pressed against the boundary of the parameter space. There is still enough room there for a good fit in this case, but as the value of $\sigma^2$ that is used for simulation increases, the 15 state model will become unable to fit the data well.

When $\sigma^2$ is fairly large, as is the case when $\sigma^2 = 0.9$, the 3 and 5 state models perform poorly as before, but the 15 state model also does a poor job of fitting the
<table>
<thead>
<tr>
<th>n</th>
<th>States</th>
<th>$\hat{F}_a(\infty)$</th>
<th>$\hat{F}_{-b}(\infty)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>200</td>
<td>3</td>
<td>0.7541 0.0320 0.2459 0.0320</td>
<td></td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>0.7906 0.0325 0.2094 0.0325</td>
<td></td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>0.8045 0.0330 0.1955 0.0330</td>
<td></td>
</tr>
<tr>
<td></td>
<td>11</td>
<td>0.8076 0.0320 0.1924 0.0320</td>
<td></td>
</tr>
<tr>
<td></td>
<td>15</td>
<td>0.8965 0.0271 0.1035 0.0271</td>
<td></td>
</tr>
<tr>
<td>500</td>
<td>3</td>
<td>0.7772 0.0194 0.2228 0.0194</td>
<td></td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>0.7980 0.0189 0.2020 0.0189</td>
<td></td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>0.8144 0.0193 0.1856 0.0193</td>
<td></td>
</tr>
<tr>
<td></td>
<td>11</td>
<td>0.8148 0.0181 0.1852 0.0181</td>
<td></td>
</tr>
<tr>
<td></td>
<td>15</td>
<td>0.8989 0.0167 0.1011 0.0167</td>
<td></td>
</tr>
<tr>
<td>1000</td>
<td>3</td>
<td>0.7811 0.0181 0.2189 0.0181</td>
<td></td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>0.7931 0.0176 0.2069 0.0176</td>
<td></td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>0.8071 0.0164 0.1929 0.0164</td>
<td></td>
</tr>
<tr>
<td></td>
<td>11</td>
<td>0.8073 0.0158 0.1927 0.0158</td>
<td></td>
</tr>
<tr>
<td></td>
<td>15</td>
<td>0.8939 0.0162 0.1061 0.0162</td>
<td></td>
</tr>
</tbody>
</table>

Table 6.8: Boundary hitting probability estimates for $\sigma^2 = 0.9$.

Table 6.8 shows the boundary hitting probabilities for $\sigma^2 = 0.9$. As before, the 3 state model tends to underestimate $F_a(\infty)$ and overestimate $F_{-b}(\infty)$. The 15 state model shows the opposite bias, overestimating $F_a(\infty)$ and underestimating $F_{-b}(\infty)$. The 7 and 11 state models estimate the hitting probabilities accurately.
Table 6.9: Quantile estimates of the hitting time distributions for $\sigma^2 = 0.9$. 

<table>
<thead>
<tr>
<th>$n$</th>
<th>States</th>
<th>$Q^a_1$</th>
<th>$Q^a_2$</th>
<th>$Q^a_3$</th>
<th>$Q^b_1$</th>
<th>$Q^b_2$</th>
<th>$Q^b_3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>200</td>
<td>3</td>
<td>7.0</td>
<td>0.3</td>
<td>16.5</td>
<td>0.8</td>
<td>32.3</td>
<td>1.5</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>9.0</td>
<td>0.7</td>
<td>17.2</td>
<td>0.8</td>
<td>30.9</td>
<td>1.6</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>8.9</td>
<td>0.7</td>
<td>16.4</td>
<td>1.1</td>
<td>29.6</td>
<td>1.6</td>
</tr>
<tr>
<td></td>
<td>11</td>
<td>9.1</td>
<td>0.7</td>
<td>16.4</td>
<td>1.0</td>
<td>29.6</td>
<td>1.6</td>
</tr>
<tr>
<td></td>
<td>15</td>
<td>12.8</td>
<td>1.2</td>
<td>23.7</td>
<td>2.1</td>
<td>41.9</td>
<td>3.3</td>
</tr>
<tr>
<td>500</td>
<td>3</td>
<td>7.2</td>
<td>0.4</td>
<td>16.5</td>
<td>0.7</td>
<td>32.7</td>
<td>1.2</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>9.0</td>
<td>0.5</td>
<td>17.2</td>
<td>0.8</td>
<td>30.9</td>
<td>1.3</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>8.5</td>
<td>0.6</td>
<td>16.1</td>
<td>0.7</td>
<td>29.4</td>
<td>1.3</td>
</tr>
<tr>
<td></td>
<td>11</td>
<td>8.5</td>
<td>0.6</td>
<td>16.0</td>
<td>0.8</td>
<td>29.0</td>
<td>1.3</td>
</tr>
<tr>
<td></td>
<td>15</td>
<td>11.8</td>
<td>0.7</td>
<td>23.0</td>
<td>1.1</td>
<td>41.5</td>
<td>2.1</td>
</tr>
<tr>
<td>1000</td>
<td>3</td>
<td>7.1</td>
<td>0.2</td>
<td>16.7</td>
<td>0.5</td>
<td>32.8</td>
<td>0.7</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>9.5</td>
<td>0.5</td>
<td>17.7</td>
<td>0.6</td>
<td>31.0</td>
<td>0.8</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>9.0</td>
<td>0.4</td>
<td>16.4</td>
<td>0.5</td>
<td>29.9</td>
<td>0.9</td>
</tr>
<tr>
<td></td>
<td>11</td>
<td>9.0</td>
<td>0.4</td>
<td>16.2</td>
<td>0.6</td>
<td>29.2</td>
<td>0.9</td>
</tr>
<tr>
<td></td>
<td>15</td>
<td>12.1</td>
<td>0.4</td>
<td>23.5</td>
<td>0.9</td>
<td>41.8</td>
<td>1.7</td>
</tr>
</tbody>
</table>
Table 6.10: Deviance information criterion (DIC) and effective number of parameters \((p_D)\) for \(\sigma^2 = 0.9\).

Table 6.9 shows the quantile estimates for \(\sigma^2 = 0.9\). The 3 state model again estimates \(Q_a^2\) and \(Q_b^3\) well but all other quantiles poorly. The 5 state model estimates \(Q_1^a\) and \(Q_b^3\) well but all other quantiles poorly. The 7 and 11 state models estimate all of the quantiles well. The 15 state model does very poorly, overestimating all but one of the quantiles. Figure 6.6 shows the posterior distribution of \(\sigma^2\) for models with 3, 11, and 15 states. The posterior is very concentrated near one, the upper boundary of the parameter space, for the 15 state model. For the previous simulations that used smaller values of \(\sigma^2\) for simulation, the 15 state model was able to inflate \(\sigma^2\) in order to fit the data well. In this case that is not possible and the result is a poor fit. In particular, the quantile estimates tend to be too large.
Table 6.10 shows the DIC and effective number of parameters, $p_D$, for $\sigma^2 = 0.9$. The 15 state model has the worst DIC for all sample sizes. Also, $p_D$ is lower for the 15 state model that for the other models. Since $p_D$ is often interpreted as a measure of model complexity, one can interpret this as meaning that the de facto constraint on $\sigma^2$ in the model decreases the model’s complexity.

![Figure 6.6: Posterior distribution of $\sigma^2$ for models with 3, 11, and 15 states when the true value of $\sigma^2$ is 0.9.](image)

When choosing a value of $m$ for a particular analysis, there is typically a range of values that will work well. Models with either too few or too many states will perform poorly. To decide how many states to use, one can fit several models with different numbers of states and then make a comparison based on the DIC values. If several numbers of states yield models that have similar DIC values, then a model with fewer states is preferred in the interest of computation time.

As this simulation explains, there is a direct relationship between the number of states and the value of $\sigma^2$ for a given dataset. When the number of states increases, the process will need to move more “quickly” to match the empirical quantiles of
the hitting time distributions of the data set and a larger value of $\sigma^2$ is needed to accomplish this.

6.2 Heart Attack Data

In this section, the Worcester Heart Attack Study data set (SOURCE: Worcester Heart Attack Study data from Dr. Robert J. Goldberg of the Department of Cardiology at the University of Massachusetts Medical School) is analyzed. This is the data set that was used to motivate the DTR model in Section 5.1. The data set includes records for $n=500$ individuals who were admitted to a hospital after a heart attack. The two events of interest are release from the hospital (upper boundary) and death in the hospital (lower boundary).

The data set includes a number of covariates that will be used in the model. For the mean of the initial state distribution, $\xi$, three variables are used that would be available at the time of admission to the hospital. They are age, gender, and an indicator that is equal to one for patients with systolic blood pressure below 90 (a possible indicator of cardiogenic shock). The covariates used for the transition probabilities ($\mu$, the “drift”) are an indicator for gender that is equal to one for women, and indicator variables for cardiogenic shock, atrial fibrillation, congestive heart failure, and complete heart block. Each of these indicators is equal to one if the condition is present. As mentioned in Section 5.1.1, cardiogenic shock is a condition in which the heart is so badly damaged that it cannot pump a sufficient amount of blood to a person’s organs. Atrial fibrillation is a type of irregular heart beat. Congestive heart failure is a condition in which a person’s heart is weakened and their kidneys cause fluid to build up in various parts of the body as a response. Complete heart
block happens when the electrical signals that pass between the heart’s chambers are disrupted.

The prior distributions are

\[
\begin{align*}
\gamma_0 & \sim N \left( 0, \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \right), \\
\beta & \sim Gamma(10, 1), \\
\gamma & \sim N(0, I),
\end{align*}
\]

and

\[
\sigma^2 \sim Beta(1, 1),
\]

where \(0\) is a vector of zeros and \(I\) is the identity matrix.

The prior for \(\beta\) is chosen to put high prior probability on values of \(\beta\) that will lead to a unimodal distribution for \(X_0\). The prior on \(\sigma^2\) is uniform on \((0, 1)\) and is chosen to allow the data to inform the posterior distribution of \(\sigma^2\). The prior for \(\gamma\) is multivariate normal with mean \(0\) and covariance matrix that is equal to the identity matrix. In Section 5.2.1 it is recommended that these priors be chosen such that \(z_\gamma \in (-3, 3)\) with high probability. For a patient with index \(i\), \(z_{ti}\) will contain a one (for the intercept term) and indicator variables for cardiogenic shock, gender, atrial fibrillation, congestive heart failure, and complete heart block. The prior variance of the term \(z_{ti} \gamma\) is equal to one plus the number of these indicator variables that are equal to one. For 75% of the sample, either zero or one of those indicators is equal to one. For those patients, \(z_{ti} \gamma \in (-3, 3)\) with prior probability at least 0.866. There are four patients for whom either four or five of the indicator variables are equal to one. For those patients, \(z_{ti} \gamma \in (-3, 3)\) with probability at least 0.489. This
probability can be increased by using a smaller variance in the prior distribution of $\gamma$. However, if the prior variance of $\gamma$ is \( \frac{1}{2} \cdot I \) instead of $I$ then the patients with no indicator variables equal to one will have $z_{ti}\gamma \in (-1, 1)$ with prior probability 0.954. This is very restrictive unless there is prior information indicating that patients with no indicators equal to one should have a latent process that starts roughly equidistant from the two absorbing boundaries.

Because the prior variance of $z_{ti}\gamma$ depends on the data, it is not possible to have $z_{ti}\gamma \in (-3, 3)$ with high prior probability for all patients without introducing this type of restriction. The choice of $I$ as the prior variance of $\gamma$ ensures that the prior is not overly restrictive.

In Section 5.2.1 it is also recommended that the prior for $\gamma_0$ be chosen such that $z_{0i}\gamma_0 \in (-3, 3)$ with high probability. For patient with index $i$, $z_{0i}$ will contain a one for the intercept term, the patient’s age in years minus 70 (the mean age in the sample), and indicator that is equal to one if the patient is female, and an indicator that is equal to one if the patient has systolic blood pressure less than 90 at the time of admission to the hospital. The age variable takes values between $-40$ and 34, with over 80% of patients having values between $-20$ and 20. There were 300 men and 200 women in the sample, so 200 patients would have the indicator variable for gender equal to one. Low systolic blood pressure at admission was present for 17 patients, 13 men and the other 4 were women. For the covariance matrix given above for $\gamma_0$, $z_{0i}\gamma_0 \in (-3, 3)$ with probability at least 0.89 for 97% of patients.

Including intercept terms for both $\xi$ and $\mu$, there are a total of 12 parameters in this model. A Metropolis-Hastings algorithm was used to fit the model with 1,000
iterations of adaptive burn-in followed by 10,000 iterations that were considered as a sample from the posterior distribution. The model was fit with 5, 7, 9, and 11 states. Table 6.11 shows the DIC values for the various numbers of states. The model with 7 states has the lowest DIC, so only results from that model will be discussed. With 7 total states, \( a = b = 3 \) and the boundary states are -3 and 3. The transient state space is \( \{-2, -1, 0, 1, 2\} \). Figure 6.7 shows the posterior distributions of the model parameters.

Before fitting, the patients’ ages were centered by subtracting 70, the mean age in the sample. As a result, the “baseline” patient is a 70 year old male who did not have low systolic blood pressure at the time of admission and did not experience cardiogenic shock, atrial fibrillation, congestive heart failure, or complete heart block. Most of the posterior mass of the intercept coefficient for \( \xi \) is placed on negative values, which implies that the baseline group tends to start closer to the lower boundary than to the upper boundary. The posterior distributions of the coefficients for age, gender, and low systolic blood pressure concentrate most of their posterior mass either on negative values or around zero, indicating that all of the individuals in the sample had a high probability of starting close to the lower boundary. This can be interpreted as meaning that all patients were in poor health immediately

<table>
<thead>
<tr>
<th>States</th>
<th>DIC</th>
<th>( p_D )</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>2869.2</td>
<td>8.6</td>
</tr>
<tr>
<td>7</td>
<td>2830.4</td>
<td>11.0</td>
</tr>
<tr>
<td>9</td>
<td>2832.2</td>
<td>10.1</td>
</tr>
<tr>
<td>11</td>
<td>2919.7</td>
<td>9.6</td>
</tr>
</tbody>
</table>

Table 6.11: DIC and effective number of parameters \((p_D)\) for heart attack data.
Figure 6.7: Posterior distributions for heart attack data.

following a heart attack. The posterior mean of the coefficient for age is negative, implying that older patients are worse off immediately following a heart attack than younger patients. The posterior mean of the coefficient for gender in the initial state distribution is close to zero, so gender does not appear to have a strong impact on the initial health state. The posterior mean of the coefficient for low systolic blood pressure is negative, meaning that a patient with low blood pressure at the time of...
admission is likely to be in poorer health at that time than a patient who does not have low blood pressure. A posteriori, the parameter $\beta$ is centered close to 8, and is between 6 and 10 with high posterior probability. To illustrate the effect of the covariates on the initial state distribution, figure 6.8 shows the posterior predictive distribution of the initial state for an individual in the baseline group and for a 70 year old male with low systolic blood pressure at the time of admission. Both are unimodal and centered below the middle state (0), but low systolic blood pressure shifts the peak of the distribution closer to the lower boundary.

![Figure 6.8: Posterior predictive distributions for the initial state.](image)

The posterior mean of the intercept for $\mu$ is positive, and most of the posterior mass is concentrated on positive values, indicating that individuals in the baseline group tend to drift toward the upper boundary. It makes sense for there to be an
overall tendency to drift upward since 92% of the patients in the sample survived. The bulk of the posterior mass for the coefficient for cardiogenic shock is on negative values, and the posterior mean is much larger in magnitude than that of the other coefficients. It is almost as large in absolute value as the posterior mean of the intercept term. An individual who experiences cardiogenic shock will have a latent health process that moves up or down with almost equal probability. Given that most individuals start closer to the lower boundary, this means that even in the absence of other risk factors a patient who experiences cardiogenic shock is more likely to die in the hospital than to survive. In the sample, 59% of patients who experienced cardiogenic shock died. While gender did not have much of an effect on the initial state, it does appear to have an impact on the drift of the process. Specifically, with large posterior probability, females have a smaller drift than males. In the sample, 10.5% of female patients died versus 6% of male patients. For patients who survived, the median length of stay at the hospital was 6 days for both men and women, but the first and third quartiles and 60th, 70th, 80th, and 90th percentiles were all longer for women. The posterior means for the coefficients for atrial fibrillation and congestive heart failure are both negative and are similar in size, so both conditions tend to slow down a patient’s recovery and increase the probability of death. Complete heart block has a coefficient with posterior mean that is close to zero, so it does not seem to have a strong effect on the recovery process, at least when the other covariates are accounted for.
6.2.1 Comparison with competing models

The performance of the DTR model will be compared to two popular survival analysis techniques that can be readily implemented with off-the-shelf software. Comparisons will be based on out-of-sample prediction. The models that are used for comparison are proportional hazards regression and a continuous TR model that uses Brownian motion with drift as the latent process. The proportional hazards regression was carried out using the “survival” package in R. The continuous TR was carried out using the “threg” package in R.

One difficulty in making a direct comparison between these models is that the DTR model includes two events of interest, while the other two methods include only one. A consequence of this is that comparisons can only be made for one event. In the heart attack data set, only about 8% of the patients in the sample died in the hospital, which is a very small proportion to use in estimating the hitting time distribution. If death is treated as the event of interest and released patients are treated as censored, a very large majority of the sample will be censored and estimation will be difficult. One might consider instead modeling “time in hospital,” so that either death or release from the hospital is treated as the event of interest. This will lead to some confusion in the interpretation of the covariate effects. For example, consider two patients who experience cardiogenic shock. One of them survives their hospital stay and is released while the other dies in the hospital. The effect of cardiogenic shock on “time in hospital” for these two patients is difficult to interpret. For the first patient, the experience of cardiogenic shock will make their recovery more difficult and longer, increasing their “time in hospital.” For the second patient, the experience
of cardiogenic shock will likely cause their health to deteriorate faster than it would otherwise, shortening their “time in hospital.” The effect of cardiogenic shock would therefore depend on the event that a person experiences.

To avoid working with a mostly censored sample or with covariate effects that are difficult to interpret, the event of interest for model comparisons will be “time to release from hospital.”

A second difficulty in making comparisons with this data set is the question of how to treat times of death for patients who die in the hospital. Treating death times as “censored recovery times” implies that the patients who died may eventually be released from the hospital, which is not true. Models for cure rates are available (Peng and Dear, 2000, Desmond et. al., 2009, Yu et. al., 2011) which allow for the possibility of a “cure” that renders an individual immune to the event of interest. While it could be argued that death renders a person immune to recovery, cure rate models generally assume that the cure event is unobservable. If one of these models is used, it will consider the patients who die in the hospital as being cured with some probability and being censored with some probability. This is not an accurate model, as it is known what happened to the patients who died.

For the sake of model comparisons, two-part models will be used. The first part will be a logistic regression model, implemented using the “glm” function in R, that will estimate the probabilities of death and recovery for each patient. Then, conditional on recovery, the proportional hazards and continuous TR models will be fit. That is, the proportional hazards and continuous TR models will be fit using a dataset that only contains records of patients who were released from the hospital.
Variable | Estimate | p-value
---|---|---
(Intercept) | 3.54189 | <2e-16
age | -0.05593 | 0.00171
gender | -0.39995 | 0.30705
low sys | -1.27551 | 0.08537
cardio shock | -2.89925 | 1.42e-07
a fib | 0.15865 | 0.75077
congestive heart fail | -0.47797 | 0.23292
complete heart block | -0.20035 | 0.83812

Table 6.12: Logistic regression estimates for heart attack data.

The following covariates are used in the logistic regression model: age, and indicators for gender, low systolic blood pressure (< 90), cardiogenic shock, atrial fibrillation, congestive heart failure, and complete heart block. That is, all of the covariates that were used in the DTR model. The parameter estimates and p-values are shown in Table 6.12. Age and cardiogenic shock are the only significant effects.

All of the covariates that were used in the DTR model were also used in the proportional hazards model. The parameter estimates and p-values are shown in Table 6.13.

All of the covariates in the proportional hazards model have negative estimated coefficients, meaning that they tend to increase the time to release from the hospital. This is sensible for the covariates used, as they are all factors expected to hinder recovery. The coefficient for cardiogenic shock is the largest in magnitude. The coefficients for atrial fibrillation and congestive heart failure also have small p-values. The presence of a low systolic blood pressure at the time of admission, age, gender, and complete heart block are not significant in this model.
<table>
<thead>
<tr>
<th>Variable</th>
<th>Estimate</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>age</td>
<td>-0.00269</td>
<td>0.47717</td>
</tr>
<tr>
<td>gender</td>
<td>-0.12452</td>
<td>0.20750</td>
</tr>
<tr>
<td>low sys</td>
<td>-0.07479</td>
<td>0.80518</td>
</tr>
<tr>
<td>cardio shock</td>
<td>-0.70082</td>
<td>0.04610</td>
</tr>
<tr>
<td>a fib</td>
<td>-0.48539</td>
<td>0.00035</td>
</tr>
<tr>
<td>congestive heart fail</td>
<td>-0.28256</td>
<td>0.00853</td>
</tr>
<tr>
<td>complete heart block</td>
<td>-0.30728</td>
<td>0.36516</td>
</tr>
</tbody>
</table>

Table 6.13: Proportional hazards estimates for heart attack data.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Estimate</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>lny0: (Intercept)</td>
<td>1.538847</td>
<td>0.0000</td>
</tr>
<tr>
<td>lny0: age</td>
<td>0.0028316</td>
<td>0.0452</td>
</tr>
<tr>
<td>lny0: gender</td>
<td>0.1021720</td>
<td>0.1519</td>
</tr>
<tr>
<td>lny0: low sys</td>
<td>0.1304188</td>
<td>0.2910</td>
</tr>
<tr>
<td>mu: (Intercept)</td>
<td>-0.7490517</td>
<td>0.0000</td>
</tr>
<tr>
<td>mu: cardio shock</td>
<td>0.2166366</td>
<td>0.0275</td>
</tr>
<tr>
<td>mu: gender</td>
<td>-0.0358000</td>
<td>0.5445</td>
</tr>
<tr>
<td>mu: a fib</td>
<td>0.1754530</td>
<td>0.0001</td>
</tr>
<tr>
<td>mu: congestive heart fail</td>
<td>0.1086487</td>
<td>0.0050</td>
</tr>
<tr>
<td>mu: complete heart block</td>
<td>0.0640992</td>
<td>0.5782</td>
</tr>
</tbody>
</table>

Table 6.14: Continuous TR estimates for heart attack data.

The continuous TR model uses the same parameters as the DTR model. That is, age, gender, and low systolic blood pressure are used for the log of the initial state and gender, cardiogenic shock, atrial fibrillation, congestive heart failure, and complete heart block are used for the drift. Table 6.14 shows the parameter estimates and p-values.
In the continuous TR model, all of the covariates have positive estimated coefficients except gender, meaning that they will tend to lead the latent Brownian motion process away from the boundary for discharge from hospital. This makes sense given the nature of the covariates. Age is the only significant effect for the log of the initial state, $y_0$. Of the covariates used for the drift term, cardiogenic shock has the largest coefficient. As for the DTR model, complete heart block is not significant. However, unlike the DTR model gender is not significant for the drift. In the DTR model gender appeared to be more important for the drift than for the initial state; in the continuous TR model the gender coefficient in the log initial state distribution has a smaller p-value. Since the DTR model included two outcomes, the drift serves not only to direct the latent Markov chain toward or away from “release from hospital” but also to direct some individuals toward “death in hospital.” As a result, the effect of some covariates differ between the discrete and continuous TR models.

Figure 6.9 compares the survival function estimates for the proportional hazards model, the continuous TR model, and the DTR model to the Kaplan-Meier curves (Kaplan and Meier, 1958) for several groups of patients. The Kaplan-Meier estimates are shown with black lines, the proportional hazards estimates with red lines, the continuous TR estimates with blue lines, and the DTR estimates with green lines. In each of the groups shown, the survival function estimates are calculated at the baseline age of 70. The Kaplan-Meier curves are computed for ages within ten years of the baseline, or between 60 and 80 years, to allow a sufficient sample size. Ages in the sample ranged from 30 to 104, with about 40% of the patients between 60 and 80 years old.
Figure 6.9: Heart attack data survival curves.

The upper left panel of Figure 6.9 shows the estimated survival curves for men with congestive heart failure and no other risk factors. The Kaplan-Meier curve starts out flat and then drops steeply between five and fifteen days before leveling out. The estimated survival curves from the three models have the same overall shape, but all of them drop more quickly than the Kaplan-Meier curve prior to five days. For times greater than five days, the DTR model survival curve is very similar to the Kaplan-Meier curve while the proportional hazards model and continuous TR model estimate a lower survival curve.
The upper right panel of Figure 6.9 shows the estimated survival curves for men with atrial fibrillation and no other risk factors. The Kaplan-Meier curve decreases steadily for the first twelve or thirteen days and then levels off at around fifteen days. All three models estimated survival curves that are similar to each other but decline much more quickly than the Kaplan-Meier curve.

The lower left panel of Figure 6.9 shows the estimated survival curves for men with atrial fibrillation and congestive heart failure and no other risk factors. The Kaplan-Meier curve shows a quick drop which levels off between five and fifteen days before dropping again. The sample size for this group is 38, smaller than the previous two, so this Kaplan-Meier curve may not estimate the shape accurately. All of the model-based survival curve estimates look similar and none of them capture the leveling-off between times five and ten that is present in the Kaplan-Meier curve. The DTR model estimates a slightly higher survival curve than the proportional hazards and continuous TR models for times greater than about seven.

The bottom right panel of Figure 6.9 shows the estimated survival curves for men with atrial fibrillation and congestive heart failure and no other risk factors. The Kaplan-Meier curve for this group also appears to drop initially and then level off before dropping again. The sample size for this curve is 19, which again may be too small to give a good estimate of the shape of the survival curve. All of the model-based survival curves look similar early on, with the DTR model decreasing more slowly than the others for larger times. One possible reason for this is that the presence of risk factors such as atrial fibrillation and congestive heart failure increase the probability of death (lower boundary hit) in the DTR model. This pulls the drift downward, leading to longer estimated survival times at the upper boundary.
To assess out-of-sample prediction accuracy for the proportional hazards, DTR, and continuous TR models, the heart attack data set was randomly divided into ten subsets of size \( n_\ell = 50 \), \( \ell = 1, 2, \ldots, 10 \). Let \( t \) be the vector of event times for the full dataset. Let \( t_\ell \) be the vector of event times for individuals in the \( \ell^{th} \) subset, and \( t_{-\ell} \) be the subset of event times that excludes the \( \ell^{th} \) subset. For each \( \ell \), each of the three models (DTR, proportional hazards, and continuous TR) was fit to \( t_{-\ell} \). The proportional hazards and continuous TR models were fit only to the released patients’ data in \( t_{-\ell} \), while the DTR model was fit to the whole data set. The survival functions were then estimated for each individual in subset \( t_\ell \) based on the fit. Posterior predictive distributions were used for the DTR model, maximum partial likelihood estimates were used for the proportional hazards model, and maximum likelihood estimates were used for the continuous TR model. A logistic regression model was used to estimate the probabilities of death and release for each hold-out subset.

Using the survival functions estimated from each of the models, the median survival time \( \hat{m}_i \) is estimated for each patient. Since each patient belongs to exactly one of the ten subsets and was therefore “left out” of the sample for one fitting of each of the comparison models, each patient has one estimated survival curve for each model. Let

\[
d_i = \left| t_i - \hat{m}_i \right|
\]

be the absolute deviation of observation \( i \) from its estimated median. The absolute deviation from the median is used here as a measure of predictive accuracy because the hitting time distributions are skewed. The mean and median of \( d_i \) are given in
<table>
<thead>
<tr>
<th>Model</th>
<th>Mean $d_i$</th>
<th>Median $d_i$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proportional hazards</td>
<td>2.8</td>
<td>2</td>
</tr>
<tr>
<td>Continuous TR</td>
<td>2.87</td>
<td>2</td>
</tr>
<tr>
<td>DTR</td>
<td>2.76</td>
<td>2</td>
</tr>
</tbody>
</table>

Table 6.15: Cross validation mean and median absolute deviations.

Table 6.15 for the three models. These means and medians are calculated across all patients.

For the DTR model, the medians $m_i$ were calculated as the median hitting time at the upper boundary (release from hospital) *conditional* on hitting the upper boundary. For all three models, the mean and median given in Table 6.15 are based only on individuals who were released from the hospital. That is, all observations were used to fit the DTR model but the values of $d_i$ for patients who died were not used to calculate the mean and median in Table 6.15. All three models had a median $d_i$ of 2, but the means differ. The DTR model had the lowest mean $d_i$ and the continuous TR model had the highest mean $d_i$.

The DTR model and the logistic regression model were used to estimate the probabilities of death and release from the hospital for each patient in the sample. Tables 6.16 and 6.17 show the classification results, where an individual whose probability of death was estimated to be greater than 0.5 was assigned a prediction of “death” and an individual whose probability of recovery was estimated to be greater than 0.5 was assigned a prediction of “release.” The results are very similar. Both models predicted release for 30 of the 39 patients who died. The DTR model did slightly
Table 6.16: Cross validation boundary predictions from the DTR model

<table>
<thead>
<tr>
<th>Predicted</th>
<th>Truth</th>
</tr>
</thead>
<tbody>
<tr>
<td>Death</td>
<td>9</td>
</tr>
<tr>
<td>Release</td>
<td>30</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Predicted</th>
<th>Release</th>
</tr>
</thead>
<tbody>
<tr>
<td>Death</td>
<td>5</td>
</tr>
<tr>
<td>Release</td>
<td>456</td>
</tr>
</tbody>
</table>

Table 6.17: Cross validation boundary predictions from the logistic regression model

<table>
<thead>
<tr>
<th>Predicted</th>
<th>Truth</th>
</tr>
</thead>
<tbody>
<tr>
<td>Death</td>
<td>9</td>
</tr>
<tr>
<td>Release</td>
<td>30</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Predicted</th>
<th>Release</th>
</tr>
</thead>
<tbody>
<tr>
<td>Death</td>
<td>7</td>
</tr>
<tr>
<td>Release</td>
<td>454</td>
</tr>
</tbody>
</table>

better at predicting release for patients who were released, correctly identifying two more than the logistic regression model.

6.3 Bond Data

Credit rating agencies such as Standard and Poor’s (S & P), Moody’s, and Fitch provide investors with credit ratings, which are designed to inform investors about the creditworthiness of a debtor. S & P uses letter ratings, with AAA being the highest and D (default) being the lowest. In between are, in decreasing order of creditworthiness, AA, A, BBB, BB, B, CCC, CC, and C. Each of these rating classes is further divided into subcategories by the inclusion of a plus sign, a minus sign, or neither. A plus sign indicates greater creditworthiness than no sign, which in turn indicates greater creditworthiness than a minus sign. An interesting category is the BBB category. The BBB rating category, including the plus and minus and
unsigned subcategories, is the lowest rating category that is considered to be “investment grade.” Ratings that are lower than BBB are considered to be “non-investment grade” or “speculative.” Ratings are adjusted over time, so a debtor who has a BBB rating now could have their credit rating upgraded or downgraded as their financial situation changes. A downgrade from BBB moves a debtor from “investment grade” to “non-investment grade” and will likely affect the price at which they (and those who own their debt) are able to sell their debt. In this section, the DTR model will be used to model the time to credit rating change for bonds with a BBB rating.

The data come from Datastream, which provides access to a historical financial database. The data used in this section was obtained on January 28, 2013. Datastream contains records of 1,419 unique non-government bond issues that were in Standard & Poor’s BBB credit rating category at some point between the beginning of 1990 and the beginning of 2010. The two events of interest for this data are a credit upgrade and a credit downgrade from the BBB category. One might imagine that the underlying latent process represents a bond issuer’s creditworthiness and that there is a range, from some lower boundary to some upper boundary, that is consistent with a BBB rating. If the latent process drops down to the lower boundary, the bond issuer’s credit rating is downgraded. If the latent process rises up to the upper boundary, the bond issuer’s credit rating is upgraded. If a bond issue reaches maturity (is completely paid off) with a BBB rating or still has rating BBB at the beginning of 2010 then it is a censored observation.

A credit rating is typically applied to an issuer rather than to a specific bond issue, so that two separate issues from the same company will change ratings at the same time. To avoid replication in the data, only one bond per issuer is used. There are
663 unique issuers in the data set that was obtained through Datastream. Changes in macroeconomic conditions can have a similar effect on firms in the same industry, but affect different industries differently. To account for this, business sector will be used as a covariate. S & P divides firms into ten business sectors: consumer discretionary, consumer staples, industrials, materials, health, financials, utilities, technology, telecommunications, and energy. Some of the bonds in the dataset do not have this information provided. For those bonds whose sectors are missing, the issuing company was researched using Google Finance and if the sector was indicated there it was recorded. There are 167 bond issues that do not have a sector indicated and they are dropped for this analysis. The analysis presented in this section is based on 496 bond issues. Time to credit rating change is measured in months.

Figure 6.10 shows the prevalence of various sectors in the bond data set. Consumer discretionary, utilities, and energy are the most common. Health, financial, technology, and telecommunications are the smallest categories. Across all sectors, 174 bonds had a plus sign at the time of entry into the BBB category, 108 had a minus sign, and 214 were unsigned. Two indicator variables are created for the sign, the first equal to one for bonds with a plus sign and the second equal to one for bonds with a minus sign. The unsigned bonds serve as a baseline group. Indicator variables are also created for each of the sectors. All of these indicators are used as covariates for the initial state distribution and for the drift. No intercept terms are used, so there is no issue with using all ten sector indicators. This leads to a total of 12 indicator variables for the initial state distribution (2 for sign indicators and 10 for sector indicators) and 12 for the drift. Adding $\beta$ and $\sigma^2$ to these leads to a total
of 26 variables in the model. The prior distributions used for this analysis are

\[ \gamma_0 \sim N(0, I), \]

\[ \beta \sim \text{Gamma}(10, 1), \]

\[ \gamma \sim N(0, I), \]

and

\[ \sigma^2 \sim \text{Beta}(1, 1), \]

where 0 is a vector of zeros and I is the identity matrix.
Recall that Section 5.2.1 contains a discussion on the topic of prior selection. The prior for $\beta$ that is used here is chosen to put high prior probability on values of $\beta$ that will lead to a unimodal prior on $X_0$. The prior on $\sigma^2$ is uniform on $(0,1)$ and is chosen to allow the data to inform the posterior distribution of $\sigma^2$. The priors for $\gamma_0$ and $\gamma$ are multivariate normal with mean $0$ and covariance matrix that is equal to the identity matrix. In Section 5.2.1 it is recommended that these priors be chosen so that $z_t \gamma \in (-3, 3)$ and $z_0 \gamma_0 \in (-3, 3)$ with high probability. For a bond with index $i$, $z_{it}$ will be a vector with one or two entries equal to one and the rest of the entries equal to zero. This is because a bond can only belong to one sector and can have up to one of the sign indicators equal to one. Thus, the prior variance of the quantity $z_{it} \gamma$ is equal to one or two. The same is true for $z_{0i} \gamma_0$. This puts a prior probability of either 0.997 or 0.866 on the interval $(-3, 3)$, depending on whether one of the sign indicators is equal to one.

A Metropolis-Hastings algorithm is employed, with 2,000 iterations of adaptive burn-in followed by 100,000 iterations that are regarded as a sample from the posterior distribution. Models with 5, 7, 9, and 11 states were fit, and Table 6.18 shows the DIC and effective number of parameters, $p_D$, for each number of states. The model with 7 states has the lowest DIC. Results from that model will be presented.

Figure 6.11 shows the posterior distributions of the sign and sector coefficients for the initial state distribution. Since the posterior distribution for the plus sign coefficient (BBB+) has its mass concentrated around 0, a bond with a plus sign does not have a significantly different starting state distribution compared to an unsigned bond in the same sector. Of the bonds that had a plus sign, 23% were upgraded and 21% were downgraded (the rest were censored). Comparing this with the unsigned
Table 6.18: DIC and effective number of parameters, $p_D$ for bond data.

<table>
<thead>
<tr>
<th>States</th>
<th>DIC</th>
<th>$p_D$</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>3047.6</td>
<td>19.2</td>
</tr>
<tr>
<td>7</td>
<td>3046.7</td>
<td>23.2</td>
</tr>
<tr>
<td>9</td>
<td>3047.2</td>
<td>23.7</td>
</tr>
<tr>
<td>11</td>
<td>3048.5</td>
<td>24.2</td>
</tr>
</tbody>
</table>

bonds, of which 22% were upgraded and 29% were downgraded, it seems sensible that the difference between them should be small. The posterior distribution of the minus sign (BBB-) coefficient puts most of its mass below zero, so bonds with a minus sign are more likely to start near the lower boundary and to have a credit downgrade than other bonds in the same sector. The sector coefficients can be interpreted as representing the baseline case, which is unsigned (no plus or minus sign). The posterior distributions of the coefficients for consumer discretionary, consumer staples, materials, utilities, and energy all put most of their mass close to zero. Unsigned bonds in those sectors are likely to have a starting state that is equidistant from the boundaries. The left panel of Figure 6.12 shows the posterior initial state distribution for the consumer staples sector. It is unimodal and nearly symmetric. Note that with 7 states, only 5 are transient. Taking $a = b = 3$, the transient states are $\{-2, -1, 0, 1, 2\}$. These states are displayed on the horizontal axis in the figure. The posterior distributions of the coefficients for health and technology put more mass below zero than above it. Unsigned bonds in these sectors are likely to have a starting state that is a little closer to the lower boundary than to the upper boundary. The middle panel of Figure 6.12 shows the posterior initial state distribution for the technology sector. It is unimodal and is centered closer to the lower boundary than
to the upper boundary. The posterior distributions of the coefficients for industrials, financials, and telecommunications put most of their mass below zero. With high
posterior probability, bonds in these sectors will have a starting state that is close to
the lower boundary. The right panel of Figure 6.12 shows the posterior initial state
distribution for the financial sector. The highest probability is for state $-2$, which is
next to the lower boundary.

Figure 6.12: Posterior initial state distributions for unsigned consumer staples, tech-
nology, and financial sectors.

Figure 6.13 shows the posterior distributions of coefficients for the drift of the
process. The posterior distribution of the coefficient for a plus sign puts most of its
mass above zero, meaning that a bond with a plus sign will tend to have a more
positive drift (or a less negative drift) than other bonds in the same sector. The
posterior distribution of the coefficient for a minus sign puts most of its mass below
zero. A minus sign therefore has a negative effect on the drift. A bond with a minus
sign will tend to have a more negative drift (or a less positive one) than other bonds
in the same sector. As with the initial state distribution above, the sector coefficients
can be interpreted as representing the baseline case, which is unsigned. The posterior
distributions of the coefficients for consumer staples, telecommunications, and energy
Figure 6.13: Bond data posteriors for the drift.

sectors put most of their mass near zero, so unsigned bonds in those sectors are about as likely to drift up and they are to drift down. The posterior distributions of the
coefficients for the consumer discretionary, materials, and financials sectors all put a little more mass below zero than above zero, so unsigned bonds in those sectors are more likely to drift down than up. The posterior distributions of the coefficients for industrials, health, and utilities sectors put a little more mass above zero than below it, so unsigned bonds in those sectors are more likely to drift up than down.

Figure 6.14 shows the posterior distributions for $\sigma^2$ and $\beta$. The posterior for $\sigma^2$ is mostly concentrated between 0.055 and 0.075. It is not “pressed up” against either of the boundaries of the parameter space, so there is no evidence that the choice of 7 states is restrictive for $\sigma^2$ based on the discussion in Section 6.1. The posterior distribution for $\beta$ is concentrated between 2 and 8, which indicates that the initial state distributions have moderate variance.

6.3.1 Comparison with other models

As with the heart attack dataset in the previous section, proportional hazards and continuous TR models are fit to the bond data and the results are compared to the
DTR model. The event of interest will be a credit downgrade from the BBB category. A bond issue that matures without transitioning out of the BBB category is treated as censored, as is a bond issuer whose credit rating is upgraded. This is different from the heart attack data, where death was not treated as censored. The rationale for this difference is that while death is permanent, a bond that has a credit upgrade is still at risk of being downgraded later on. It makes sense here to treat the upgraded bonds as censored.

The covariates used in the proportional hazards model are the sign and sector indicators. The health sector was used as a baseline. The parameter estimates and p-values are shown in Table 6.19.

The coefficient for a plus sign is positive, indicating that a plus sign increases the hazard for a credit downgrade compared to an unsigned bond. This is difficult to interpret. A possible explanation can be found in Figure 6.15, which shows histograms of the credit downgrade times for unsigned and plus sign bonds. Of the bonds with

<table>
<thead>
<tr>
<th>Variable</th>
<th>Estimate</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>plus</td>
<td>0.3686</td>
<td>0.00054</td>
</tr>
<tr>
<td>minus</td>
<td>-0.2845</td>
<td>0.18117</td>
</tr>
<tr>
<td>Consumer Discretionary</td>
<td>0.4515</td>
<td>0.03020</td>
</tr>
<tr>
<td>Consumer Staples</td>
<td>0.4964</td>
<td>0.03228</td>
</tr>
<tr>
<td>Industrials</td>
<td>-0.6358</td>
<td>0.18344</td>
</tr>
<tr>
<td>Materials</td>
<td>-0.1985</td>
<td>0.55400</td>
</tr>
<tr>
<td>Financials</td>
<td>0.0908</td>
<td>0.82265</td>
</tr>
<tr>
<td>Utilities</td>
<td>1.7886</td>
<td>1.2e-07</td>
</tr>
<tr>
<td>Technology</td>
<td>-0.8864</td>
<td>0.01712</td>
</tr>
<tr>
<td>Telecommunications</td>
<td>0.2189</td>
<td>0.57802</td>
</tr>
<tr>
<td>Energy</td>
<td>1.0590</td>
<td>0.00379</td>
</tr>
</tbody>
</table>

Table 6.19: Proportional hazards estimates for bond data.
a plus sign, 21% are downgraded while 29% of unsigned bonds are downgraded. However, for the plus sign bonds that are downgraded, the event happens more quickly than for the unsigned bonds.

![Histogram of credit downgrade times for unsigned and plus sign bonds.](image)

Figure 6.15: Histogram of credit downgrade times for unsigned and plus sign bonds.

The coefficient for a minus sign is positive, indicating that a minus sign increases the risk of a downgrade. The consumer discretionary, consumer staples, utilities, and energy sectors have coefficients that are significant and positive. For this model that means that bonds in these sectors have a higher risk of downgrade than bonds in the health sector, which is the baseline sector. The technology sector has a significant negative coefficient, so bonds in that sector have a lower risk of downgrade than bonds in the health sector. The industrials, materials, financials, and telecommunications sectors do not have significant coefficients in this model, so bonds in those sectors have a risk of downgrade that is not significantly different from that of bonds in the health sector.
### Table 6.20: Continuous TR estimates for bond data.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Estimate</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>lny0: (Intercept)</td>
<td>1.9102970</td>
<td>0.0000</td>
</tr>
<tr>
<td>lny0: plus</td>
<td>-0.2957673</td>
<td>0.0282</td>
</tr>
<tr>
<td>lny0: minus</td>
<td>-0.1625148</td>
<td>0.1819</td>
</tr>
<tr>
<td>lny0: Consumer Discretionary</td>
<td>0.0395716</td>
<td>0.8776</td>
</tr>
<tr>
<td>lny0: Consumer Staples</td>
<td>0.1185787</td>
<td>0.6995</td>
</tr>
<tr>
<td>lny0: Industrials</td>
<td>-0.2381621</td>
<td>0.3909</td>
</tr>
<tr>
<td>lny0: Materials</td>
<td>0.2070077</td>
<td>0.4629</td>
</tr>
<tr>
<td>lny0: Financials</td>
<td>-0.1973248</td>
<td>0.5041</td>
</tr>
<tr>
<td>lny0: Utilities</td>
<td>0.8819395</td>
<td>0.0020</td>
</tr>
<tr>
<td>lny0: Technology</td>
<td>0.0662554</td>
<td>0.8282</td>
</tr>
<tr>
<td>lny0: Telecommunications</td>
<td>-0.3343729</td>
<td>0.2709</td>
</tr>
<tr>
<td>lny0: Energy</td>
<td>0.2950699</td>
<td>0.2614</td>
</tr>
<tr>
<td>mu: (Intercept)</td>
<td>0.0236920</td>
<td>0.6759</td>
</tr>
<tr>
<td>mu: plus</td>
<td>0.0734713</td>
<td>0.0112</td>
</tr>
<tr>
<td>mu: minus</td>
<td>-0.0279507</td>
<td>0.3125</td>
</tr>
<tr>
<td>mu: Consumer Discretionary</td>
<td>-0.0579619</td>
<td>0.3096</td>
</tr>
<tr>
<td>mu: Consumer Staples</td>
<td>0.0400443</td>
<td>0.5452</td>
</tr>
<tr>
<td>mu: Industrials</td>
<td>0.0751196</td>
<td>0.2360</td>
</tr>
<tr>
<td>mu: Materials</td>
<td>-0.0185223</td>
<td>0.7649</td>
</tr>
<tr>
<td>mu: Financials</td>
<td>-0.1265145</td>
<td>0.1216</td>
</tr>
<tr>
<td>mu: Utilities</td>
<td>-0.0755673</td>
<td>0.2705</td>
</tr>
<tr>
<td>mu: Technology</td>
<td>-0.0172504</td>
<td>0.8080</td>
</tr>
<tr>
<td>mu: Telecommunications</td>
<td>-0.0487707</td>
<td>0.5313</td>
</tr>
<tr>
<td>mu: Energy</td>
<td>-0.0146216</td>
<td>0.8072</td>
</tr>
</tbody>
</table>

The covariates used in the continuous TR model were the same as those used in the DTR model: sign and sector indicators were used for both the initial state and drift. Health was again used as a baseline sector. The coefficients and p-values are shown in Table 6.20.

As in the proportional hazards model, the plus sign coefficient in the log of the initial state term (lny0: plus) has the opposite sign of what is expected. It is negative and significant, indicating that the latent processes for bonds with a plus sign tend
to start closer to the boundary than the latent processes for unsigned bonds. The only sector that has an initial state coefficient that is significantly different from the health sector is utilities, which has a positive coefficient, indicating that the latent processes for bonds in the utilities sector tend to start further from the boundary than the latent processes for utilities bonds. The plus sign coefficient for the drift is positive and significant, indicating that the latent process for bonds with a plus sign tend to drift away from the boundary more quickly (or towards the boundary more slowly) than unsigned bonds. The minus sign coefficient is negative but is not significant. None of the sectors have a drift coefficient that is significantly different from that of the health sector.

Figure 6.16 shows some of the survival curves. As in the heart attack example, the black line is the Kaplan-Meier curve, the red line is the estimated survival curve from the proportional hazards model, the green line is the estimated survival curve from the DTR model, and the blue line is the estimated survival curve from the continuous TR model. The survival curves for the discrete and continuous TR models tend to have the same overall shape, with one typically decreasing more quickly than the other before leveling off. For the consumer discretionary minus and technology minus sectors, the DTR survival curve decreases more quickly while in the others the continuous TR survival curve decreases more quickly. In the consumer consumer discretionary sector with a minus sign the two are indistinguishable for the first 50 months and then the DTR curve drops lower than the continuous TR curve.

To assess out-of-sample prediction accuracy, the bond dataset was randomly divided into ten subsets of size \( n_\ell = 50, \, \ell = 1, \ldots, 9 \) and \( n_{10} = 46 \). The proportional hazards, continuous TR, and DTR models were fit to each subset \( t_{-\ell} \) and survival
curves were estimated for each of the individuals in the $\ell^{th}$ subset. As with the heart attack dataset, the median survival time $\hat{m}_i$ was estimated for each individual and the absolute deviation $d_i$ was calculated.
Both TR models performed well compared to the proportional hazards model in out-of-sample predictions. The DTR model had a lower mean and median $d_i$ than the continuous TR model.

In addition to the survival models, a logistic regression model was fit to the bond data with each subset left out and probabilities of upgrade and downgrade were estimated for each bond that was left out. Only bonds that had an upgrade or a downgrade were used for fitting (censored bonds were left out). Tables 6.22 and 6.23 show the classification results, where a bond whose probability of downgrade was estimated to be greater than 0.5 was assigned a prediction of “downgrade” and a bond whose probability of upgrade was estimated to be greater than 0.5 was assigned a prediction of “upgrade.” The DTR model performed slightly better at predicting upgrades and the logistic regression model performed slightly better at predicting downgrades.

### 6.3.2 Time varying covariates

In the bond data set, some bonds had one sign initially but changed signs later on before being upgraded or downgraded out of the BBB category or censored. Of the bonds that started out with a plus sign, less than half (44.5%) were absorbed or

<table>
<thead>
<tr>
<th>Model</th>
<th>Mean $d_i$</th>
<th>Median $d_i$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proportional hazards</td>
<td>75.4</td>
<td>52</td>
</tr>
<tr>
<td>Continuous TR</td>
<td>37.7</td>
<td>27</td>
</tr>
<tr>
<td>DTR</td>
<td>28.0</td>
<td>21</td>
</tr>
</tbody>
</table>

Table 6.21: Cross validation mean and median absolute deviations.
Table 6.22: Cross validation boundary predictions from the DTR model.

<table>
<thead>
<tr>
<th>Predicted</th>
<th>Truth</th>
</tr>
</thead>
<tbody>
<tr>
<td>Downgrade</td>
<td>95</td>
</tr>
<tr>
<td>Upgrade</td>
<td>51</td>
</tr>
</tbody>
</table>

Table 6.23: Cross validation boundary predictions from the logistic regression model.

<table>
<thead>
<tr>
<th>Predicted</th>
<th>Truth</th>
</tr>
</thead>
<tbody>
<tr>
<td>Downgrade</td>
<td>99</td>
</tr>
<tr>
<td>Upgrade</td>
<td>47</td>
</tr>
</tbody>
</table>

censored without first becoming unsigned or receiving a negative sign. Nearly half of them (48.9%) became unsigned, and the rest had their sign changed to negative. Of the bonds that started out unsigned, only 28% remained unsigned until the time of absorption or censoring. The bonds that started out with a minus sign were much less likely to have a sign change, and 58.3% of them retained their minus sign until they were either absorbed or censored.

To accommodate these sign changes in the DTR model, let \( z_t^i \) be a vector that contains two sign indicators and ten sector indicators, where the sign indicators indicate the sign that bond \( i \) has at time \( t \), and let \( Z_i = (z_i^0, z_i^1, z_i^2, \ldots, z_i^{t_i}) \), where \( t_i \) is bond \( i \)'s event or censoring time. Let \( q_i \) be a row vector of initial state probabilities, which are determined by \( z_i^0, \gamma_0 \), and \( \beta \). Let \( \mu_i(t) = \sigma^2(2 \logit^{-1}(z_i^t\gamma) - 1) \) for
Table 6.24: DIC and effective number of parameters, $p_D$, for bond data with time-varying sign indicators.

<table>
<thead>
<tr>
<th>States</th>
<th>DIC</th>
<th>$p_D$</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>2908.7</td>
<td>18.3</td>
</tr>
<tr>
<td>7</td>
<td>2922.6</td>
<td>22.3</td>
</tr>
</tbody>
</table>

This time-varying version of the DTR model was fit using the same prior distributions described in Section 6.3 using 5 and 7 states. Table 6.24 shows the DIC and effective number of parameters, $p_D$, for each number of states. These DIC values are quite a bit smaller than those obtained without time-varying sign indicators, which implies that the fit is improved by their inclusion. The DIC is lowest for the 5 state model. Results from that model will be presented. In this section the model that was fit in Section 6.3 without time-varying sign indicators will be referred to as the “time homogeneous” model for convenience.

The posterior distributions for the initial state are shown in Figure 6.17. The posterior distribution for the plus sign coefficient puts essentially all of its mass below zero, which indicates that with high probability bonds with a plus sign start closer to the lower boundary than unsigned bonds. However, the posterior distribution of the plus sign coefficient in the drift term puts essentially all of its mass above zero (Figure 6.19). A negative coefficient in the initial state distribution combined with a very positive coefficient in the drift will lead to a high probability of reaching the
Figure 6.17: Bond data posteriors for the initial state distribution with time varying sign indicators.
lower boundary at early times due to the proximity of the starting state to the lower boundary, followed by a high probability of reaching the upper boundary at later times due to the positive drift. As Figure 6.18 shows, the bonds that have a plus sign at time 0 exhibit this behavior. Several of the posterior distributions, such as those for the minus sign, industrials, financials, and telecommunications coefficients look similar to those obtained using the time-homogeneous model. The posterior distributions of the coefficients for consumer discretionary, consumer staples, materials, health, and utilities sectors are more spread out here than those obtained using the time-homogeneous model. Finally, the posterior distribution of the coefficient for the technology sector is shifted to be a little more positive here than that in the time-homogeneous model, and the posterior distribution of the coefficient for the energy sector is shifted to be a little more negative here than in the time-homogeneous model.

Figure 6.18: Histograms of hitting times at the upper and lower boundaries for bonds that have a plus sign at time 0.
The posterior distributions for the drift coefficients are shown in Figure 6.19. The posterior distribution of the plus sign coefficient puts most of its mass above zero, on
values larger than those for the time homogeneous model. The posterior distribution of the minus sign coefficient puts essentially all of its mass below zero, on values smaller than those for the time homogeneous model. The posterior distributions of the coefficients for the health, technology, consumer staples, and telecommunications sectors have larger variance than those obtained from the time homogeneous model. The posterior distributions of the coefficients for the utilities and energy sectors put more mass below zero than those obtained from the time homogeneous model. Finally, the posterior distributions of the coefficients for the consumer discretionary, industrials, materials, and financials sectors look very similar to those obtained with the time homogeneous model.

Figure 6.20: Bond data posteriors for $\sigma^2$ and $\beta$ with time varying sign indicators.

Figure 6.20 shows the posterior distributions for $\sigma^2$ and $\beta$. For this model, $\sigma^2$ is between 0.022 and 0.035 with high posterior probability and $\beta$ is between 4 and 12 with high posterior probability.
To assess this time-varying DTR model’s out-of-sample prediction accuracy, the same ten subsets that were used for model comparison in Section 6.3.1 are used here to estimate the probabilities of upgrade and downgrade for each bond in a cross-validation exercise. As before, a bond with an estimated probability of upgrade that is greater than 0.5 is assigned a prediction of “upgrade” and a bond with an estimated probability of downgrade that is greater than 0.5 is assigned the prediction “downgrade.”

The probability estimates from the time-varying DTR model lead to better classification. Overall, 87% of bonds that were either upgraded or downgraded received the correct prediction. For the logistic regression model and the time homogeneous DTR model, only about 66% of predictions were correct. Because so many bonds changed signs over time, there is a considerable improvement when the DTR model includes these changes.

This improvement is not evident in the mean and median absolute deviations, as Table 6.24 shows. The time-varying DTR model performed worse than the time homogeneous DTR model and the continuous TR model.

Figure 6.21 shows the posterior hitting time distribution from the time-varying and time homogeneous DTR models for the same bond. This bond was initially unsigned.

<table>
<thead>
<tr>
<th>Predicted</th>
<th>Truth</th>
<th>Downgrade</th>
<th>Upgrade</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>137</td>
<td>22</td>
</tr>
<tr>
<td></td>
<td></td>
<td>9</td>
<td>74</td>
</tr>
</tbody>
</table>

Table 6.25: Cross validation boundary predictions from the time-varying DTR model
<table>
<thead>
<tr>
<th>Model</th>
<th>Mean $d_i$</th>
<th>Median $d_i$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proportional hazards</td>
<td>75.4</td>
<td>52</td>
</tr>
<tr>
<td>Continuous TR</td>
<td>37.7</td>
<td>27</td>
</tr>
<tr>
<td>DTR</td>
<td>28.0</td>
<td>21</td>
</tr>
<tr>
<td>Time-varying DTR</td>
<td>40.8</td>
<td>36.5</td>
</tr>
</tbody>
</table>

Table 6.26: Cross validation mean and median absolute deviations for all models.

for 15 months, then was given a negative sign. The downgrade time is indicated by a vertical black line. The hitting time distribution changes abruptly at fifteen months, and the peak shifts outward a little bit. This increases the median of the estimated distribution, shown as a red vertical line, making it inaccurate compared to the time homogeneous model. Section 7.2.1 contains more discussion on this issue.
Figure 6.21: Posterior hitting time pmf at lower boundary (downgrade) for one bond.
Chapter 7: Conclusions and Future Work

7.1 Conclusions

In this dissertation, the DTR model is presented and implemented on two datasets. The model uses a latent discrete-state, discrete-time Markov chain with two absorbing boundaries as a model for survival data. Each boundary is associated with an observable event. The DTR model allows estimation of event probabilities as well as survival distributions. It is flexible enough to fit survival distributions well, can be used to produce a variety of hazard function shapes, and easily incorporates covariates into the transition probabilities.

Extensions of this model include the addition of time-varying covariates and the possibility of more general forms for the transition probability matrix that do not use the parametrization that is presented here. The ease of changing the transition probability matrix at any given time in response to either a change in covariates or possibly a change in the environment gives the DTR model flexibility to model the survival of individuals whose characteristics evolve over time.

When compared to alternative models that estimate the event probabilities separately from the survival distributions, the DTR model performs well. In addition, time-varying covariates are incorporated into the DTR model in a natural way.
7.2 Future Work

7.2.1 Survival distributions with time-varying covariates

When time varying covariates are incorporated into the DTR model, survival distributions can be estimated up to the event time or censoring time but may not be estimated well beyond that point. This is because the time-varying covariates are not observed beyond the event or censoring time. One option for dealing with this issue is to use the last observed covariate value to estimate the survival curve for all time points beyond the event or censoring time. This could be problematic if covariate values change frequently, and in Section 6.3.2 it was an issue for a model that appeared to fit very well.

Another option is to treat the time-varying covariate as a random variable for time points beyond the observed time interval. Section 4.2 presented some important quantities related to random walks in dynamic Markovian random environments. Each value that a time-varying covariate takes on has associated with it a transition probability matrix for $X_t$. Using the terminology introduced in Section 4.2, these transition probabilities are an “environment.” The sequence of environments for an individual can be modeled as a Markov chain.

For example, the bond data set that was discussed in section 6.3 used a bond rating’s sign as a time-varying covariate. Of the bonds that started out with a plus sign, less than half (44.5%) were absorbed or censored without first having a sign change. Of the bonds that started out unsigned, only 28% remained unsigned until the time of absorption or censoring. The bonds that started out with a minus sign were much less likely to have a sign change, and 58.3% of them retained their minus
sign until they were absorbed or censored. A three state Markov chain can be fit to these rating changes, with one state for “plus,” one state for “unsigned,” and one state for “minus.” The transition probabilities for this Markov chain can be estimated using the observed rating changes in the data. Let \( \omega_t \) be the environment process with \( \omega_t = 1 \) if the sign is negative at time \( t \), \( \omega_t = 2 \) if it is unsigned, and \( \omega_t = 3 \) if it is positive. Then let \( \hat{A} = [\alpha_{ij}]_{i,j=1}^3 \) be the estimated transition probability matrix for bond rating signs and suppose that a bond is censored at time \( t_{ccens} \) while in environment \( i \). To estimate the boundary hitting probabilities for times beyond \( t \), let \( v(t) = (v_1(t), v_2(t), v_3(t)) \) be the \( i \)th row of \( \hat{A}^{t-t_{ccens}} \) and use the transition probabilities

\[
p_+(t) = \sum_{j=1}^3 v_j(t)p_j^+, \quad p_0(t) = \sum_{j=1}^3 v_j(t)p_j^0, \quad p_-(t) = \sum_{j=1}^3 v_j(t)p_j^-,\]

where, for example, \( p_j^+ \) is the probability of an upward transition in environment \( j \).

### 7.2.2 Time-varying boundaries

Some time-varying effects might influence characteristics of the latent process in a DTR model other than the transition probabilities. For example, covariates that are external to the individuals of interest may not affect the transition probabilities of the latent process but may impact the boundaries, making events of interest more or less likely.

Figure 7.1 shows the divorce times in Wyoming for couples who were married in Wyoming in 1972-1974 (Source: National Center for Health Statistics (1972-1977)). The horizontal axis measures months since January, 1972. Each of the vertical green lines marks the beginning of a new year. Within each year, there is a clear trend where more couples divorce in July-October than other times of year. The red line marks the date when a law passed that allowed no-fault divorce in Wyoming. It
Figure 7.1: Histogram of divorce times

appears that the trend changes at that point in time. If a threshold regression model is used for this data, the latent process might be interpreted as the health of a person’s marriage, which changes over time. However, the change in law would be unlikely to directly affect the health of a marriage. Rather, it makes a divorce easier to obtain for a married person who would like to end their marriage but would have difficulty doing so under the old law. The law change could be incorporated into this model by changing the threshold state. At the time of change in law, the lower boundary, representing divorce, could move upward. This boundary change is easily implemented with the DTR model. The change of boundary will cause an increase in the probability of divorce for all times after the law change. In this case, the law change is a covariate that is external to the individuals being studied. Other instances of external covariates might be handled well by the DTR model.
7.2.3 Time-delayed covariates

In some instances, a time-varying covariate will have an impact that is progressive or time-delayed. That is, the new covariate value has an effect that is phased in over time. For example, when a person quits smoking their risk of cancer is immediately lowered, but also continues to decrease over time. In addition, circulation and lung function improve over a period of several months after quitting, and the risk of coronary heart disease and stroke decrease over a period of years after quitting (OncoLink).

Suppose one wishes to use a TR model with “stroke” as the event of interest, and consider an individual who has quit smoking recently. If the researcher knows that this person is currently a nonsmoker, this does not indicate that their risk of stroke is elevated due to past smoking. Also, knowing that this person has a history of smoking does not indicate that the risk of stroke has recently decreased as a result of quitting. Because the effect of quitting on the risk of stroke changes with time, it is desirable to model quitting as an effect that changes in time.

The DTR model can implement changes to the transition probabilities that are time-delayed. In Section 5.3.3, the drift $\mu_i(t)$ was used to construct a time-varying transition probability matrix. Let $z^t$ be an indicator variable whose value changes in time. For example, $z^t = 1$ for an individual who has quit smoking and $z = 0$ for an individual who has not quit smoking. Also, let $t_q$ be the time at which a person quits smoking and let $z_d^t$ be a new covariate that will be constructed from $z^t$. Let $m$ be the number of time units over which the effect of stroke is believed to decrease linearly.

157
after a person quits smoking, and let

\[
\begin{align*}
    z_{d}^{t_{q}} & = \frac{1}{m}, \\
    z_{d}^{t_{q}+1} & = \frac{2}{m}, \\
                      & \vdots \\
    z_{d}^{t_{q}+m} & = 1, \\
    z_{d}^{t_{q}+m+1} & = 1, \\
                      & \vdots
\end{align*}
\]

and so on, so that \( z_{d}^{t} = 1 \) for all times after \( t = t_{q} + m \). Using \( z_{d}^{t} \) as a covariate for \( \mu_{i}(t) \) rather than \( z^{t} \) will create a transition probability matrix that changes progressively over time in response to a person quitting smoking. Alternative functional forms of the delayed effect of covariates over time can just as easily be considered.

### 7.2.4 Refinement of the state space

To increase the flexibility of the DTR model, the state space and transition probability matrix can be adjusted. As was seen in Chapter 6, there will typically be a range of numbers of states that works best in the DTR model that was presented in Chapter 5. However, if the state space is refined, so that there are more transient states, it may be possible to adjust the transition probability matrix to achieve an even better fit.

For example, let \( a \) and \( -b \) be fixed and consider an expanded state space with absorbing boundaries \(-2b\) and \(2a\). Also, consider a transition probability matrix such
that for $i \in \{-2b + 2, \ldots, a - 2\}$,

$$p_{ij} = \begin{cases} 
\frac{r}{2}(\sigma^2 + \mu), & \text{if } j - i = 2, \\
\frac{1-r}{2}(\sigma^2 + \mu), & \text{if } j - i = 1, \\
1 - \sigma^2, & \text{if } j = i, \\
\frac{1-r}{2}(\sigma^2 - \mu), & \text{if } j - i = -1, \\
\frac{r}{2}(\sigma^2 - \mu), & \text{if } j - i = -2, \\
0, & \text{otherwise},
\end{cases}$$

for some $r \in (0, 1)$. For completeness, let

$$p_{-2b+1,j} = \begin{cases} 
\frac{r}{2}(\sigma^2 + \mu), & \text{if } j = -2b + 3, \\
\frac{1-r}{2}(\sigma^2 + \mu), & \text{if } j = -2b + 2, \\
1 - \sigma^2, & \text{if } j = -2b + 1, \\
\frac{1}{2}(\sigma^2 - \mu), & \text{if } j = -2b, \\
0, & \text{otherwise},
\end{cases}$$

and

$$p_{2a-1,j} = \begin{cases} 
\frac{1}{2}(\sigma^2 + \mu), & \text{if } j = 2a, \\
1 - \sigma^2, & \text{if } j = 2a - 1, \\
\frac{1-r}{2}(\sigma^2 - \mu), & \text{if } j = 2a - 2, \\
\frac{r}{2}(\sigma^2 - \mu), & \text{if } j = 2a - 3, \\
0, & \text{otherwise},
\end{cases}$$

Compared to the transition probability matrix presented in Section 5.1.2, this allows for a greater variety of sample paths. These paths can move in steps of size one or two at each transition. Note that the inclusion of $r$ here is only one way of allowing larger steps in the transition probability matrix and is used to provide an example of how larger jumps might be incorporated.

The absorbing boundaries can be placed even further apart, and larger jumps can be permitted in the transition probability matrix to compensate, in order to increase the variety of sample paths allowed by the model. Further details are left for future research.
The DTR model described in Chapter 5 and the refinement described above will both lead to models whose properties may be heavily dependent on the number of states that is chosen. In some cases it may be desirable to have a model that limits this dependence in some sense. One way to accomplish this is to redefine the transition probability matrix in a way that approximates the transition probabilities of Brownian motion.

Let $B_t$ be a Brownian motion process with drift $\mu$ and diffusion parameter $\sigma$. Then the increments $B_1 - B_0$, $B_2 - B_1$, etc. are normally distributed with mean $\mu$ and variance $\sigma^2$. Let $a$ and $-b$ be absorbing boundaries, and assume here that both $a$ and $b$ are integers. Let $X_t$ be a discrete-state, discrete-time Markov chain with absorbing boundaries $a$ and $-b$ and transition probabilities

$$p_{ij} = \int_{j-0.5}^{j+0.5} \frac{1}{\sqrt{2\pi \sigma^2}} \exp \left( -\frac{1}{2\sigma^2} (y - (i + \mu))^2 \right) dy,$$

for $i, j \in \{-b+1, -b+2, \ldots a-1\}$. Let the transition probabilities into the boundary states be

$$p_i -b = \int_{-\infty}^{-b+0.5} \frac{1}{\sqrt{2\pi \sigma^2}} \exp \left( -\frac{1}{2\sigma^2} (y - (i + \mu))^2 \right) dy,$$

and

$$p_i a = \int_{a-0.5}^{\infty} \frac{1}{\sqrt{2\pi \sigma^2}} \exp \left( -\frac{1}{2\sigma^2} (y - (i + \mu))^2 \right) dy.$$

To further refine the state space, let $m = a+b+1$ be the number of states and consider the extension to $2m - 1$ states. This extension can be accomplished by adding an additional state in between two existing states. For example, if $a = b = 1$ then $m = 3$ and the state space is $\{-1, 0, 1\}$. To extend this state space to $2m - 1 = 5$ states, additional states can be added between $-1$ and $0$ and between $0$ and $1$ yielding the new state space $\{-1, -0.5, 0, 0.5, 1\}$. The transition probabilities, $p_{i,j}$, can be defined
in the same way as with the original state space, with the limits of integration changed so that 0.25 is added and subtracted instead of 0.5. The state space can be further increased by continuing to add more states in the same way. This formulation will lead to a discrete-state, discrete-time Markov chain that is always approximating a Brownian motion process with drift $\mu$ and diffusion $\sigma$, regardless of the number of states that is used.
Bibliography


tional/Bonds.

[26] E. L. Kaplan and Paul Meier. “Nonparametric estimation from incomplete ob-
servations”. *Journal of the American Statistical Association*, 53(282):457–481, 
June 1958.


mixture model for assessing treatment efficacy in a multiple myeloma clinical 

markers and latent health status”. *Journal of the Royal Statistical Society*, 

analysis: modeling event times by a stochastic process reaching a boundary”. 

regression: their theoretical and practical connections”. *Lifetime Data Analysis*, 

Garshick. “Assessing lung cancer risk in railroad workers using a first hitting 

[34] Mei-Ling Ting Lee, G. A. Whitmore, and Bernard A. Rosner. “Threshold re-
gression for survival data with time-varying covariates”. *Statistics in Medicine*, 

“Survival analysis of orthodontic mini-implants”. *American Journal of Orthodon-

statistics of Brownian motion with purely time dependent drift and diffusion”. 


