Likelihood-based testing and model selection for hazard functions with unknown change-points

Matthew R. Williams

Dissertation submitted to the Faculty of the
Virginia Polytechnic Institute and State University
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in

Statistics

Dong-Yun Kim, Chair
Eric P. Smith
Feng Guo
Marion Reynolds

March 30, 2011
Blacksburg, Virginia

Keywords: Change-point, Donsker class, Gaussian process, hazard function, local asymptotic normality, Likelihood ratio test

Likelihood-based testing and model selection for hazard functions with unknown change-points

Matthew R. Williams

Abstract

The focus of this work is the development of testing procedures for the existence of change-points in parametric hazard models of various types. Hazard functions and the related survival functions are common units of analysis for survival and reliability modeling. We develop a methodology to test for the alternative of a two-piece hazard against a simpler one-piece hazard. The location of the change is unknown and the tests are irregular due to the presence of the change-point only under the alternative hypothesis. Our approach is to consider the profile log-likelihood ratio test statistic as a process with respect to the unknown change-point. We then derive its limiting process and find the supremum distribution of the limiting process to obtain critical values for the test statistic. We first reexamine existing work based on Taylor Series expansions for abrupt changes in exponential data. We generalize these results to include Weibull data with known shape parameter. We then develop new tests for two-piece continuous hazard functions using local asymptotic normality (LAN). Finally we generalize our earlier results for abrupt changes to include covariate information using the LAN techniques. While we focus on the cases of no censoring, simple right censoring, and censoring generated by staggered-entry; our derivations reveal that our framework should apply to much broader censoring scenarios.

Dedication

To David Craig and Hayley Shen, in the pursuit of honor and cookies.

Acknowledgments

I am grateful to my advisor Dong-Yun Kim for her time, talent, and enthusiasm. I could not have found a more caring and dedicated mentor. Thank you to Eric Smith for watching out for me since the beginning and for serving on both my masters and doctoral committees. Thank you also to Feng Guo and Marion Reynolds for their time and insights as committee members. While at Virginia Tech, I was fortunate to be sponsored as a Doctoral Scholar through the Institute for Critical Technology and Applied Science. Thank you to Roop Mahajan and Ann Craig for the administration of this ambitious program. Thank you also to Jeffrey Birch for securing this fellowship and for being so invested in my education. Thank you to the faculty and staff from the Department of Statistics, the Graduate School, and Virginia Tech for this opportunity. Finally, thank you to my fellow students, friends, and family. I lack the skill to adequately express my gratitude to you all. Perhaps I should have used more equations.

Contents

List of Tables

List of Figures

CHAPTER 1

INTRODUCTION

The focus of the research presented here is the development of testing procedures for the existence of change-points in parametric hazard models of various types. Our main approach is to consider the profile log-likelihood ratio test statistic as a process with respect to the unknown change-point. We then derive its limiting process and find the supremum distribution of the limiting process to obtain critical values for the test statistic.

In this chapter, we present a motivating example and provide brief background on hazard functions and change-points. In Chapter 2 we provide an overview of recent results in the literature. With this context, we then demonstrate our main approach with results in Chapters 3, 4, and 5. Finally in Chapter 6, we summarize our main findings and outline potential avenues for further research.

1.1 Motivating Example

Let us consider an epidemiological data set comparing treatment and control groups in their times to the onset of symptoms. Responses are either the time of first record of the symptom or a right censoring time when the study ended or the patient left. Figure 1.1 displays the survival curves for just such a scenario. While the survival of the treatment group seems to fit reasonably to a smooth curve, the control group has a bend or kink requiring a two-piece style curve to obtain a good fit. In practice we could always fit the two-piece curve, but we wish to avoid over-fitting the data. We therefore seek a statistical test to aid in our selection of an appropriate model.

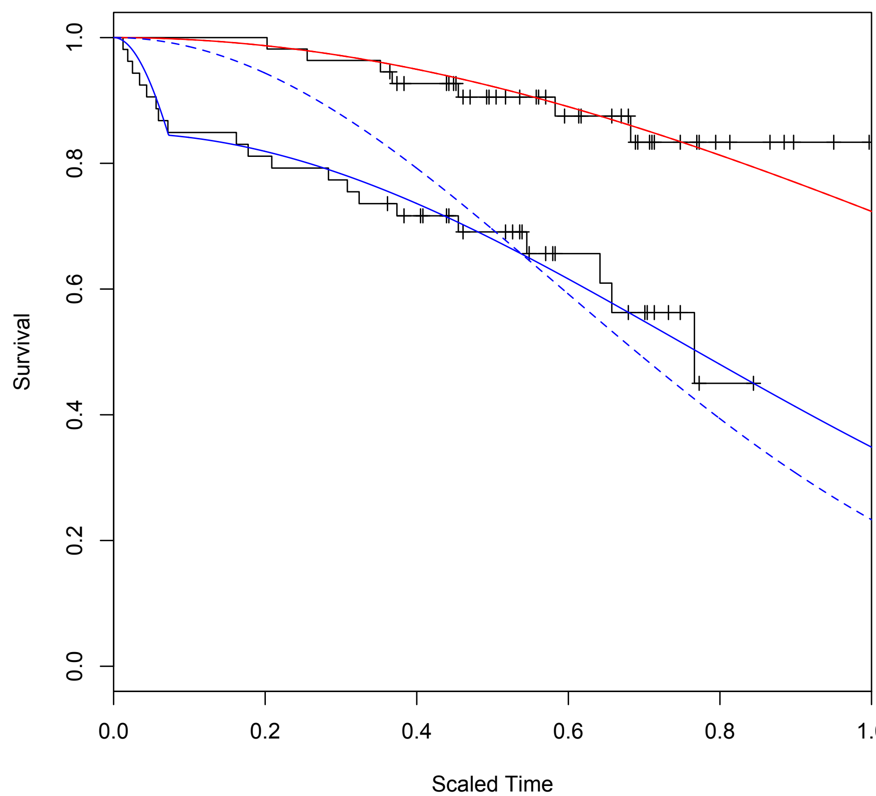

Figure 1.1: Survival curves for control (blue) and treatment (red) groups from a clinical trial. Parametric Change-point model (solid) lines up with non-parametric Kaplan-Meier estimates (black). Parametric No-change model (dashed) is a poor fit.

1.2 Hazard Functions

While the curves in Figure 1.1 are survival functions, we can equivalently model hazard functions. A hazard function is the instantaneous rate of failure (or event occurrence) and is calculated as the ratio of the probability density function over the survival function $h(t) = f(t)/S(t)$. The probability of failure before time t is $F(t) = \int_0^t f(x)\,dx$ and the survival is the complementary probability that a failure does not occur before t: $S(t) = 1 - F(t)$. Any of $h(t)$, $f(t)$, $F(t)$, and $S(t)$ uniquely describes the probabilistic behavior of an event. In fact, from one we can directly derive the other three. For example, $S(t) = \exp(-\int_0^t h(x)\,dx)$, where the integral is known as the cumulative hazard function and is sometimes denoted $H(t)$ or $\Lambda(t)$. We will avoid the latter notation since we later use Λ to refer to likelihood ratios.

In Figure 1.1, the parametric models for survival are specified in terms of linear hazard functions for the treatment $h_T(t)$ and control group $h_C(t)$:

$$\begin{aligned} h_T(t) = &\quad \theta_T t,\ t \geq 0 \\ h_C(t) = &\quad \begin{cases} \theta_{C1} t, & 0 \leq t < \nu \\ \theta_{C2} t, & t \geq \nu \end{cases} \end{aligned} \tag{1.1}$$

where θ_T, θ_{C1}, θ_{C2} are unknown positive slopes and ν is the unknown change-point. The question them becomes how to test the null hypothesis $\theta_{C1} = \theta_{C2}$ versus the alternative $\theta_{C1} \neq \theta_{C2}$. The presence of the change-point ν makes this problem irregular and requires more complex techniques to address.

1.3 The Change-Point Problem

Suppose we observe data $t_1, \ldots, t_n$ which are independent and distributed according to a distribution function $F(t;\theta_i)$ with parameters θ_i. The classic change-point problem is to compare the null hypothesis: $\theta_1 = \ldots = \theta_n$ to the alternative $\theta_1 = \ldots = \theta_{k-1} \neq \theta_k = \ldots = \theta_n$, where the k is considered an unknown change-point. Under the null hypothesis, the data can be grouped together, while under the alternative it can be separated into two groups. What makes this problem interesting is the nuisance parameter k. Standard approaches such as likelihood ratio tests have to be modified to properly account for the uncertainty involved with estimating k. This classic problem has been studied and generalized extensively. See Csörgő and Horváth (1997) for a broad collection of limit theorems based on partial sums and Duggins (2010) for parametric resampling methods and a comprehensive literature review.

We focus our efforts on a different kind of change-point problem. Suppose we have time to event data consistent with hazard model (1.1). This problem is similar to the classic change-point problem in that the nuisance parameter ν is not present under the null model and the uncertainty in its estimation must be accounted for during testing and confidence interval construction. However, unlike the classic change-point model we are not testing whether to model one group vs. two groups. Instead we are testing whether or not to include a structural change in the model for all observations. In this respect our change-point problem is more analogous to modeling a mixture of two distributions.

The change-point model (1.1) represents an abrupt change in the hazard function. Chapter 3 generalizes this model from linear to polynomial hazard functions. Chapter 4 examines two-piece continuous hazard functions. Chapter 5 generalizes model (1.1) and the results from Chapter 3 to include covariate information. In the next Chapter, we introduce available techniques currently in the literature and provide a context in which to introduce our new methodology.

CHAPTER 2

BACKGROUND AND LITERATURE REVIEW

Hazard rates and more general hazard functions are commonly the unit of analysis for survival and reliability applications. In these realms, data come from event times such as death, part failure, or the onset of symptoms. In survival analysis, the emphasis is generally on identifying the efficacy of treatments as well as identifying important covariate information such as gender, age, and heart rate. The underlying distribution is of minimal interest and is therefore often modeled with semi-parametric methods. In contrast, reliability is often performed in the engineering arena. The underlying failure mechanism as well as any covariate information is associated with known functions and parameters which have domain-specific interpretation.

A common model assumption in both survival and reliability analysis is the proportional hazards model:

$$h(t) = \lambda_0(t)\exp(\mathbf{z}^T\boldsymbol{\beta}) \tag{2.1}$$

where $\lambda_0(t)$ is a common baseline function, $\mathbf{z}$ is a vector of covariate information, and $\boldsymbol{\beta}$ are the corresponding regression coefficients. This popular model is attributed to Cox (1972). A proportional relationship for the hazard implies a power relationship for the survival function: $S(t) = S_0(t)^\alpha$ where $\alpha = \exp(\mathbf{z}^T\boldsymbol{\beta})$.

We can now revisit our motivating example: Under the null hypothesis the hazard function (1.1) satisfies (2.1) with $\lambda_0(t) = t$, $\exp(\beta_0) = \theta_T$ for treatment, and $\exp(\beta_0 + \beta_1) = \theta_C$ for control. We notice that a structural change-point in $\lambda_0(t)$ or β_0 still implies that the hazard functions for treatment and control are proportional. In contrast, a change-point in β_1 leads to a non-proportional hazard model. This is an important distinction. Fully parametric methods have historically focused on detecting changes in the baseline hazard $\lambda_0(t)$ and/or common parameter β_0. In contrast, semi-parametric partial likelihood methods by their nature can only model β_1 directly. Therefore any change-point models must also be limited to β_1. Our methods developed in this work can be applied to test for changes in either the baseline (Chapters 3 and 4) or the covariate coefficients (Chapter 5) or both.

2.1 Estimation and Inference for Hazard Models

Modern survival analysis (see O'Quigley 2008) uses the empirical distribution of the data to consistently estimate the baseline survival function $S_0(t)$ or the baseline cumulative hazard function $H_0(t)$. Inference for the regression parameters $\boldsymbol{\beta}$ is based on Cox's (1975) partial likelihood:

$$PL(\boldsymbol{\beta}) = \prod_{i=1}^{n_0} \frac{Y_i(t_i)\exp(\mathbf{z}_i^T\boldsymbol{\beta})}{\sum_{j=1}^{n} Y_j(t_i)\exp(\mathbf{z}_j^T\boldsymbol{\beta})}$$

where $Y_j(t_i)$ is an indicator function which takes the value of 1 when observation $j = 1,\ldots,n$ with associated covariate information z_j is in the risk set at time of failure t_i for $i = 1,\ldots,n_0 \leq n$ and the value of 0 otherwise. The concept of the risk set is very flexible and allows for many kinds of transition states and censoring regimes. The partial likelihood is then used in ways analogous to the traditional likelihood. Estimates $\hat{\boldsymbol{\beta}}$ are derived by maximizing $PL(\boldsymbol{\beta})$ or equivalently solving for the zeros of the partial score function $\partial \ln PL(\boldsymbol{\beta})/\partial \boldsymbol{\beta}$. Properties based on the asymptotic normality of $\hat{\boldsymbol{\beta}}$ are also available.

In modern reliability (see Meeker and Escobar 1998) we generally select an underlying distribution for t such as the Weibull or Lognormal. This implies a fully parametric baseline hazard $\lambda_0(t)$ and survival $S_0(t)$. In reliability, Weibull models are particularly useful due to their flexibility and the interpretation of their parameters. For example, the Weibull hazard function is a polynomial of order $k-1$ where $k > 0$ is the shape parameter. When $k = 1$, we have the exponential distribution which possesses the memoryless property. It's hazard function is a constant. In reliability this is known as random failure. When $k < 1$, the hazard function is decreasing. This means that once units survive past an initial period of high mortality, failures rates decrease and even out. This behavior is known as infant mortality. Finally when $k > 1$, we have increasing hazard functions, which imply failure rates that increase with age. This behavior is known as wear out (or rapid wear out for large k). Once we have decided on a distribution, proper likelihoods are then utilized to perform estimation and inference in the usual manner.

2.2 Deriving Asymptotic Properties

In the survival and reliability literature there are many asymptotic results showing either the consistency of estimators or the distributional convergence of a test statistic. The older works tend to assume that $\lambda_0(t)$ is a constant and there is no covariate information. In this setting, the data are independent and identically exponential under the null hypothesis. In several works, the relationship between exponential data and the Poisson process is utilized and asymptotic results are derived using the machinery of counting processes and renewal theory (For a general background, see for example Siegmund 1985).

Change-point inference results for Cox's partial likelihood tend to rely heavily on martingale central limit theorems. An accessible description of both martingales and their uses in proportional hazards regression can be found in O'Quigley (2008). Formal derivations of asymptotic properties of partial likelihood estimates can be found in Prentice and Self (1983).

Other works rely on transformations to use the classic convergence results of the empirical distribution function: $F_n(t) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}_{\{x_i<t\}}$ for iid X_i with distribution function $F(t)$. By the law of large numbers $F_n(t) \xrightarrow{a.s.} F(t)$ for every t. By the central limit theorem $\sqrt{n}(F_n(t) - F(t)) \xrightarrow{d} N(0, F(t)(1 - F(t)))$ for every t. If we now consider t as an index to a random path or process, we now refer to the empirical process $\mathbb{F}_n(t)$. The Glivenko-Cantelli Theorem extends the law of large numbers to this process: $\sup_t |\mathbb{F}_n(t) - F(t)| \xrightarrow{a.s.} 0$. Similarly Donsker's Theorem extends the central limit theorem $\sqrt{n}(\mathbb{F}_n(t) - F(t)) \Longrightarrow \mathbb{G}(t)$, where $\Longrightarrow$ stands for the weak convergence of a process and $\mathbb{G}(t)$ is the Brownian Bridge process.

In the work presented in Chapters 3, 4, and 5, we rely on empirical processes to justify asymptotic results. While Chapter 3 relies on the classic results mentioned above, in Chapters 4 and 5 we use more modern tools of empirical measure. Consider the properties of the empirical expectation for a function f: $P_n f(t) = \frac{1}{n}\sum_{i=1}^{n} f(x_i, t)$. If we consider each t separately, than the law of large numbers and the central limit theorem are still generally applicable. However, if we now consider t as an index to a random process we then have an empirical process $\mathbb{P}_n f(t)$ with expecation $Pf(t)$. Classes of functions such that $\sup_t |\mathbb{P}_n f(t) - Pf(t)| \xrightarrow{a.s.} 0$ holds are known as Glivenko-Cantelli classes. Similarly, classes of functions such that $\sqrt{n}(\mathbb{P}_n f(t) - Pf(t)) \Longrightarrow \mathbb{G} f(t)$ holds are known as Donsker classes. While abstract results for these classes have been available for decades (Giné and Zinn, 1984; Alexander et al., 1984), it is only more recently that they have been combined authoritatively and illustrated with accessible examples (Dudley, 1999; van der Vaart and Wellner, 1996; van der Vaart, 1998). These works make it much more feasible to directly evaluate membership in these classes for specific functions $f \in \mathscr{F}$.

2.3 Change-point Inference

Most of the results for inference with respect to change-point models in hazard functions have been split between model fitting and testing for the existence of a change-point. We highlight some main results for both areas to demonstrate the variety of techniques that have been applied.

2.3.1 Model Fitting and Estimation

Many approaches to model fitting are variants of constant and step-like hazard functions. Nguyen et al. (1984) established a consistent estimator for the change-point in the basic constant hazard rate problem. Yao (1986) proposed a different estimator based on maximizing the likelihood and derived its non-normal limiting distribution. Pham and Nguyen (1990) modify this slightly and Pham and Nguyen (1993) prove that a parametric bootstrap will converge to the same limiting distribution. More recently Ferger (2007) extend estimation results to smooth hazards with a single jump. Xu and Adak (2002) use a CART-like tree-based approach to fit piecewise constant hazard functions with time-varying regression effects. Dupuy (2006) allows for the unknown change to affect both the hazard $\lambda_0(t)$ and the regression parameters $\boldsymbol{\beta}$ of a parametric survival regression model. The estimators for the change-point and the hazard and regression parameters are shown to be consistent. Zhao et al. (2009) consider a change-point hazard function model with long-term survivors (i.e. non-negligible right censoring). They use pseudo maximum likelihood for estimation and establish large sample properties of their estimators. Kosorok and Song (2007) examine linear transformation models for right censored survival data. The unknown change-point is based on a covariate threshold. Their nonparametric maximum likelihood estimators are shown to be consistent.

Other estimation approaches consider generalizing from constant (exponential) to polynomial (Weibull) hazard functions. Achar and Bolfarine (1989) outline a Bayesian estimation procedure for abrupt change in exponential hazard functions with unknown change-point, but do not implement the full model. Ghosh and Ebrahimi (2008) demonstrate a Bayesian approach for the estimation of changes for the exponential and the Weibull case. Casellas (2007) uses Metropolis-Hastings procedures to fit Weibull survival models with piecewise baseline hazard functions. They use deviance information criteria (Spiegelhalter et al., 2002) to choose the number of change-points to include in the model.

2.3.2 Change-point Testing and Model Selection

When testing whether for the existence of the change-point, the unknown change-point is often a parameter present only under the alternative. Davies (1977) concluded that the aysmptotic results for the standard likelihood ratio test statistics for such situations are not applicable, but inference from a Gaussian process may be possible. Matthews and Farewell (1982; 1985) considered the case of a constant hazard rate with an abrupt change at an unknown time. They developed a likelihood ratio test for such a change-point and simulated critical values. Matthews et al. (1985) demonstrate the weak convergence of a score test statistic to a Brownian Bridge and derive asymptotic critical values for the test. Worsley (1988) noted that the likelihood ratio test depends on the width of the data intervals. Worsley also claims that the maximized likelihood ratio test statistic corresponds to estimates of change-points that are very close to uncensored observations. In practice this makes it unnecessary to scan over the entire range of the change-point. Instead we can focus on a neighborhood around the observed data points. If all observations are uncensored, the likelihood ratio test becomes unbounded (infinite) for values close to the largest observation (Nguyen et al., 1984). Loader (1991) also considered a constant hazard rate with abrupt change using large deviation approximations to the significance level of the likelihood ratio test. Loader demonstrates the method on Stanford heart transplant data.

Other works have expanded the scope of testing for a change-point within the constant hazard rate framework. Kim et al. (2004) developed a likelihood ratio test for a change in hazard rates under staggered-entry and type-I censoring. Their test statistic converges weakly to a stationary Ornstein-Uhlenbeck process, whose tail probabilities can be approximated. Park (1988) tests a constant failure rate versus a change in trend. They assume that the proportion of the population that fails before the change-point is known. Their results are based on differentials of functions of order statistics (Boos, 1979). Na et al. (2005) extend these results to unknown change-points and unknown proportions, and they derive asymptotic properties of the test. Luo et al. (1997) developed a likelihood ratio test that includes binary covariates and tests the null hypothesis of a change-point at $\nu = \nu_0$ vs. the alternative $\nu \neq \nu_0$. Dupuy (2009) allows for the regression coefficients of the covariates to change as well. They were able to obtain non-asymptotic bounds for the significance level and type II error assuming the exponential data with random right censoring. Goodman et al. (2006) test for multiple change-points in piecewise log-linear hazard functions using asymptotic Wald tests. Qin and Sun (1997) examine right-censored data with the hazard function completely specified before the change, but completely unknown after it. They propose a modified maximal censored likelihood ratio test and bootstrap to get critical values. Groenewald and Schoeman (2001) consider fitting exponential, monotonic, and bath tub hazard functions with Bayesian methods. They use fractional Bayes factors (O'Hagan,

1997) to compare models. Liu et al. (2008) propose a maximal score test for detecting a change-point in a Cox proportional hazards model with censored data. They use a Monte Carlo approach for assessing the statistical significance of the tests.

There has been some overlap between estimation and testing. For example, Kosorok and Song (2007) also test for the existence of a change-point by finding the supremum of the score process and use Monte Carlo procedures to evaluate the properties of the test statistic. But perhaps testing and estimation should be separated, at least procedurally. Henderson (1990) demonstrates that reasonable estimates for the change-point that come from testing procedures (such as maximizing a statistic over a range of potential values for the change-point) might not be sufficient under the alternative hypothesis. In other words, there might be more information about the change-point from the data than is contained in the estimates of the change-point that are obtained from maximizing test statistics.

2.4 Our Approach: Quadratic Expansions of the Likelihood

Few of the testing methods in the literature identify the limiting distribution for their tests, requiring simulation or bootstrapping to obtain critical values. Most of these methods avoid incorporating meaningful censoring structures, and few consider hazard functions outside of the constant or proportional hazard framework. Our approach for testing is to approximate the profile likelihood ratio test process (indexed by the unknown change-point) by a quadratic process that is asymptotically equivalent. When applicable, this method allows for the natural incorporation of structural changes, censoring, and covariate information since these are all part of a proper likelihood. Furthermore our quadratic approximation provides a structure which can either be simulated or further approximated to derive closed form solutions.

In Chapter 3 we revisit the conditions in Kim et al. (2004) and generalize the results to include the Weibull distribution with known shape parameter under the null hypothesis of no change-point. We derive the approximation through a direct second order Taylor expansion of the likelihood ratio test. In Chapter 4 we consider two-piece continuous hazard functions. We demonstrate a method that is known as local asymptotic normality (LAN) to find quadratic approximations to the likelihood under the null and the alternative hypothesis. One of the main advantages of LAN is that it does not require closed form estimators of the parameters. Finally in Chapter 5 we use the LAN method to generalize the results of Chapter 3 to include covariate information. While we focus on censoring generated by staggered-entry, several of our derivations reveal that this framework has much broader applicability.

CHAPTER 3

ABRUPT CHANGE UNDER STAGGERED-ENTRY AND RIGHT CENSORING

In this chapter, we derive the limiting distribution for the test of change-point when the data have a Weibull distribution and enter in a staggered fashion at random times, which often occurs in large scale clinical trials or experiments. Also, the data are subject to type I censoring, which arises when some of the subjects are still alive at the end of the fixed study period. Since an exponential distribution is a special case of the Weibull distribution, the current work extends the result of Kim et al. (2004). Although the exponential case has merit, its application is rather limited due to the availability of suitable data. On the other hand, Weibull distributions are flexible and used extensively across many different disciplines.

In Section 3.1 we prove that the profile log-likelihood ratio test statistic converges weakly to a nonstationary Gaussian process, and via a suitable transformation, becomes the Ornstein-Uhlenbeck process. We show that the critical values of the test can be easily computed without extensive simulations from the approximate tail probabilities of the supremum of the stationary process.

In Section 3.2, using computer simulation we evaluate the performance of the asymptotic critical values determined in Corollary 4, Section 2, by studying the empirical error probability and power of the test. We find that the empirical error probability closely matches nominal significance level, and the attained power is reasonably high even for moderate sample sizes. We also report that the test is valid even when a large percentage of data is censored.

In Section 3.3, we illustrate our method using data from a clinical trial in which gamma interferon was used to treat chronic granulomatous disease (CGD), as in Fleming and Garrington (1991). Our test shows that there is a change-point in the hazard function for the control group, and the piecewise Weibull distribution seems a suitable model for the data.

3.1 Main Result

3.1.1 The Weibull Hazard Model

Suppose in a random experiment (e.g., clinical trial) subjects enter at random times $\tau_1 < \ldots < \tau_n$, following a Poisson process $N = N(t)$ with intensity rate γ. We assume the ending time of the experiment, say $T > 0$, is fixed and independent of the number of subjects included in the study. Let τ_i denote the time the i^{th} subject entered the study. Conditioning on $N(T) = n$, it is easy to show that $T - \tau_i$ follows the uniform distribution on $[0, T]$.

Let Y_i denote the survival time of the i^{th} subject. Further we assume that Y_i has the hazard function $h_Y(t)$ of the following form:

$$h_Y(t) = \begin{cases} \theta_1 t^{k-1}, & 0 \le t < \nu \\ \theta_2 t^{k-1}, & t \ge \nu \end{cases} \tag{3.1}$$

where $\nu, \theta_1, \theta_2 > 0$ are unknown, and $k > 0$ is known. The corresponding density $f_Y(t)$ is

$$f_Y(t) = \begin{cases} \theta_1 t^{k-1} \times \exp(-\theta_1 t^k / k), \ t < \nu \\ \theta_2 t^{k-1} \times \exp(-\theta_2 t^k / k + (\theta_2 - \theta_1)\nu^k / k), \ t \ge \nu \end{cases} \tag{3.2}$$

3.1.2 Weak Convergence of the Likelihood Ratio Test

We wish to test $H_0 : \theta_1 = \theta_2$ versus $H_1 : \theta_1 \neq \theta_2$. Under the null hypothesis ν vanishes and (3.2) reduces to a Weibull density. This problem is irregular in the sense that the change-point ν in (3.2) has meaning only under the alternative.

Unless otherwise noted, throughout the rest of this paper we assume that $N(T)=n$ subjects are included in the study. Then the observed data consists of pairs (X_i,δ_i) for $i=1,\dots,n$ where $X_i=\min(Y_i,T-\tau_i)$ denotes the observed lifetime and $\delta_i=\mathbf{1}\{Y_i< T-\tau_i\}$ denotes the indicator, respectively. $\delta_i=1$ indicates the i^{th} observation is not censored. In this case, the log-likelihood can be written as

$$\begin{aligned}\ln L_n(\theta_1,\theta_2,\nu) &= K_1(\nu)\ln(\theta_1)+K_2(\nu)\ln(\theta_2)\\ &\quad - T_1(\nu)\theta_1-T_2(\nu)\theta_2+\textstyle\sum_{i=1}^n(k-1)\delta_i\ln x_i,\end{aligned}$$

where

$$K_1(\nu)=\sum_{i=1}^n\mathbf{1}\{x_i<\nu,\delta_i=1\},\qquad K_2(\nu)=\sum_{i=1}^n\mathbf{1}\{x_i\geq\nu,\delta_i=1\},$$
$$T_1(\nu)=\frac{1}{k}\sum_{i=1}^n(x_i^k\wedge\nu^k),\qquad T_2(\nu)=\frac{1}{k}\sum_{i=1}^n(x_i^k-\nu^k)^+,$$

with $(x\wedge y)=\min\{x,y\}$ and $x^+=\max\{0,x\}$.

Under the null hypothesis, $K_i(\nu)$, $T_i(\nu)$ do not depend on ν so the maximum likelihood estimate (MLE) for θ can be simply written as

$$\hat\theta=\frac{K_1+K_2}{T_1+T_2}=\frac{\sum_{i=1}^n\mathbf{1}\{\delta_i=1\}}{(1/k)\sum_{i=1}^n x_i^k}.$$

Under the alternative hypothesis, for each ν, the MLEs for θ_1, θ_2 are $\hat\theta_1(\nu)=K_1(\nu)/T_1(\nu)$ and $\hat\theta_2(\nu)=K_2(\nu)/T_2(\nu)$, respectively. The profile log-likelihood ratio test statistic $\Lambda_n(\nu)$ is

$$\Lambda_n(\nu)=K_1(\nu)\ln\left(\frac{K_1(\nu)}{T_1(\nu)}\times\frac{T_1+T_2}{K_1+K_2}\right)+K_2(\nu)\ln\left(\frac{K_2(\nu)}{T_2(\nu)}\times\frac{T_1+T_2}{K_1+K_2}\right).$$

Since the likelihood is unbounded at the end points, we restrict the range of ν to $[a,b]$ for some design parameters a,b such that $0<a<b<T$. Under this constraint on ν, the test rejects H_0 if the test statistic $\sup_{a\leq\nu\leq b}2\Lambda_n(\nu)$ exceeds a critical value that is to be determined later in this paper.

We introduce $Z_n(\nu)$ as the following:

$$Z_n(\nu)=\left(\frac{T_1(\nu)}{K_1(\nu)}-\frac{T_2(\nu)}{K_2(\nu)}\right)\left(\frac{K_1(\nu)K_2(\nu)}{K_1+K_2}\right)^{1/2}\left(\frac{K_1+K_2}{T_1+T_2}\right)\tag{3.3}$$

Using a simple second-order Taylor series expansion, we can write

$$2\Lambda_n(\nu)=Z_n^2(\nu)+R_n(\nu).\tag{3.4}$$

and the following result holds. Note that by letting $n \to \infty$ it is understood that the Poisson intensity $\gamma \to \infty$.

3.1.3 Theorems

By $\to^p$ and $\to^d$, we denote the convergence in probability and convergence in distribution, respectively.

Proposition 3.1.1 *For any* $0 < a < b < T$,

$$\sup_{a \le v \le b} |R_n(v)| \to^p 0$$

as $n \to \infty$, *when* H_0 *is true and* $R_n(v)$ *is as in (3.4).*

Proof The proof is analogous to that of Proposition 1 in Kim et al. (2004).

Theorem 3.1.2 *Under* H_0, $Z_n(v)$ *as in (3.3) converges weakly to the zero-mean, unit-variance nonstationary Gaussian process* $Z(v)$, *where for* $v_1 \le v_2$

$$\rho_{12} = Cov(Z(v_1), Z(v_2)) = \left(\frac{P(X < v_1, \delta = 1)}{P(X < v_2, \delta = 1)} \times \frac{P(X \ge v_2, \delta = 1)}{P(X \ge v_1, \delta = 1)} \right)^{1/2}.$$

Proof See Proofs Section.

Although $Z(v)$ is nonstationary, we can transform $Z(v)$ into an Ornstein-Uhlenbeck process $W(t)$, a stationary Gaussian process whose properties have been extensively studied.

Corollary 3.1.3 *Let* $Z(v)$ *be as in Theorem 3.1.2 and let* $g_1(v) = P(X < v, \delta = 1)$ *and* $g_2(v) = P(X \ge v, \delta = 1)$. *Then*

$$W(t) = Z\left(g_2^{-1}\left(\frac{g_2(0)e^t}{1+e^t}\right)\right),\ t \in (-\infty, \infty)$$

is an Ornstein-Uhlenbeck process.

Proof Both g_1 and g_2 are positive, g_1 is increasing, and g_2 is decreasing on $v \in [0, T]$, and $g_1(v) = g_2(0) - g_2(v)$. Furthermore we have

$$g_2(v) = g_2(0)\frac{e^t}{1+e^t},\ t \in (-\infty, \infty).$$

Since $v_1 < v_2$ if and only if $t_1 > t_2$, $\rho_{12} = e^{-\frac{1}{2}|t_1 - t_2|}$.

Asymptotic critical values of the test statistic can be readily computed from the following approximation formula.

Corollary 3.1.4 *Let $Z(v)$ be as in Theorem 3.1.2. Then for large $c > 0$,*

$$P[\sup_{a \le v \le b} Z(v)^2 > c^2] \approx (b^* - a^*) \frac{c}{\sqrt{2\pi}} e^{-c^2/2},$$

where $a^* = \log[g_2(b)/(g_2(0) - g_2(b))]$ and $b^* = \log[g_2(a)/(g_2(0) - g_2(a))]$.

Proof See Theorem 12.2.9 and Remark 12.2.10 in Leadbetter et al. (1983).

3.2 Simulation Study

Critical values of the test can be computed using Theorem 3.1.2 or Corollary 3.1.4. To use the former, we simulate 50,000 realizations of $Z(v)$ for different combinations of parameters, for $v \in [a, b]$ over a grid of 10^{-4} resolution. This provides one set of critical values. Alternatively we can find the approximate critical values using Corollary 3.1.4, which does not require the generation of many random numbers. Table 3.1 shows that two sets of critical values match up reasonably well for the $[0.2, 0.8]$ and $[0.3, 0.7]$ cases but less so for $[0.4, 0.6]$. This does not pose a serious problem though. Based on power analysis later in this section, the use of a narrow interval is not recommended.

3.2.1 Empirical Error Probabilities

Corollay 3.1.4 is applicable regardless of specific values of k in (3.2). For this simulation study however, we generated 10,000 data sets from a Rayleigh distribution, a special case of Weibull distribution with $k = 2$ for each combination of $(a, b) = (0.2, 0.8)$, $(0.3, 0.7)$, $(0.4, 0.6)$ and $\theta = 0.25, 1, 4$, over a 10^{-4} grid for v in $[a, b]$. To simulate the Poisson entry times, we generate $N \sim \text{Poisson}(\gamma)$ with the expected sample size $\gamma \in \{50, 100, 200, 500, 1000\}$. We then take a random sample Y_i for $i = 1, \ldots, n$ from a Rayleigh distribution with parameter θ. Empirical Type I errors were computed and compared to the nominal 5% significance level. The result is summarized in Table 3.2.

Table 3.1: Critical Values for $\sup\{Z(v)^2 : a \leq v \leq b\}$ **for** $k = 2$ **for combinations of** a**,** b**, and** θ_0**. For each, 50,000 realizations were generated by Theorem 3.1.2. Critical values were taken as the 95% and 99% quantiles. Critical values were also computed using Corollary 3.1.4.**

Parameters				Thm. 3.1.2		Cor. 3.1.4		Censoring
θ_0	T	a	b	95%	99%	95%	99%	P($\delta = 0$)
0.25	1	0.2	0.8	9.21	12.69	9.32	12.86	96%
		0.3	0.7	8.20	11.64	8.16	11.74	
		0.4	0.6	7.09	10.49	6.43	10.10	
1	1	0.2	0.8	9.30	12.78	9.36	12.90	85.6%
		0.3	0.7	8.29	11.90	8.20	11.79	
		0.4	0.6	7.04	10.48	6.47	10.14	
4	1	0.2	0.8	9.44	12.97	9.56	13.10	59.8%
		0.3	0.7	8.42	11.83	8.43	12.00	
		0.4	0.6	7.23	10.62	6.71	10.36	

The $\theta_0 = 4$ case corresponds to data sets of which about 60% are censored. Even in such cases, moderate sample size of 50 is sufficient to attain the nominal significance level. On the other hand, for $\theta_0 = 0.25$ case the test is substantially conservative and it requires much larger sample (at least 500, say) to reach the same nominal level. This is not surprising because in this worst case scenario, almost all data points are censored (96%).

3.2.2 Power

Under the alternative hypothesis, we generate 10,000 data sets using $\theta_1 = 0.25, 1, 4$ and the ratio of parameters $\theta_2/\theta_1 \in [1/3, 3]$ with the expected sample size $\gamma = 500$. Recall that θ_1, θ_2 denote parameters in the hazard function before and after the change-point v, respectively. We assume a change-point occurs at $v = 0.5$. Figure 3.1 shows power curves under various combinations of parameters with the fixed design parameter interval $[0.3, 0.7]$. For a given sample size, the test has higher power when (i) the amount of change in parameter is substantial, i.e., for large values of $log(\theta_2/\theta_1)$, and (ii) the censoring proportion is low.

One interesting fact is that the power curves are not symmetric; if we compare the power at $\theta_2/\theta_1 = c$ and the power $\theta_2/\theta_1 = 1/c$ for $c > 1$, the former is higher than the

Table 3.2: Convergence of Empirical Type I errors for 5% Critical Values using Theorem 3.1.2. Rayleigh data was simulated with no change. The proportion of times the test exceeded the critical value (Type I error) approaches 5% as the expected sample size increases. Simulations were repeated 10,000 times for standard errors ≈ 0.0020.

Parameters				Expected Sample Size				
θ_0	T	a	b	50	100	200	500	1000
0.25	1	0.2	0.8	0.0010	0.0055	0.0159	0.0333	0.0426
		0.3	0.7	0.0009	0.0046	0.0188	0.0410	0.0481
		0.4	0.6	0.0001	0.0025	0.0228	0.0486	0.0460
1	1	0.2	0.8	0.0164	0.0246	0.0354	0.0472	0.0503
		0.3	0.7	0.0155	0.0372	0.0484	0.0477	0.0475
		0.4	0.6	0.0173	0.0470	0.0500	0.0486	0.0492
4	1	0.2	0.8	0.0628	0.0445	0.0482	0.0485	0.0468
		0.3	0.7	0.0512	0.0483	0.0490	0.0501	0.0506
		0.4	0.6	0.0449	0.0487	0.0493	0.0463	0.0514

latter, showing a higher power curve on the right side. For an exponential case, Kim et al. (2004) use the Hellinger distance to explain this interesting phenomenon. That is, power is higher on the right side because the Hellinger distance is larger there.

More intuitively, we may explain this using the difference in censoring rates. For a given θ_1 under H_1, the hazard function $h(t)$ satisfies the following: $h(t;(\theta_2/\theta_1) > 1) \geq h(t;(\theta_2/\theta_1) \leq 1)$. So, for $\theta_2/\theta_1 > 1$ the hazard is greater and the data are less likely to be censored at the end of study period T, providing greater power.

Figure 3.2 shows the power curves for three choices of design parameter intervals: $[0.4, 0.6]$, $[0.3, 0.7]$, and $[0.2, 0.8]$. The change-point is at $\nu = 0.5$. In general, the narrower the design parameter interval is, the higher power the test has, if the true change-point is inside the interval considered. Note that the increased power for the narrower interval is very modest. On the other hand, when the interval does not include the change-point, the power deteriorates quickly (not shown here). Given these facts, we suggest using a large enough design interval when performing an actual test.

To compute the critical values from Corollary 3.1.4, one needs to find $b^* - a^*$, which depends on the estimate of θ under H_0. Simulations (not included) indicate that computation of $b^* - a^*$ is insensitive to a wide range of θ_0 parameter values, so use of any reasonable estimator for θ works well in determining the critical values. By using the observed proportion of censored observations as an estimate of $P(\delta = 0)$ and then solving for θ_0, we can obtain an estimator for θ under the null hypothesis. This estimator will be used in the next section.

3.3 Application to CGD Data

We illustrate the proposed test with clinical trial data of gamma interferon in the treatment of chronic granulomatous disease (CGD) (Fleming and Garrington, 1991). CGD is a hereditary disease of the immune system that can lead to serious infections. We only consider the time (in days) elapsed until the first infection since the start of treatment. The patients start treatment at different times which fits the staggered entry scenario. Examination of the event times suggests that a Rayleigh distribution is a reasonable basis for the underlying parametric model with the exception of the possible change-point in the hazard. Weibull plots of the placebo and treatment groups indicate that $k = 2$ is satisfactory (figure not included). In fact, the MLE of the shape parameter k for the treatment group is 1.80, which is well within the 95% confidence interval based on Wald statistic. Likewise, the MLE's of k for the control group are 2.19 and 1.96 before and after the change, respectively, and both are within 95% Wald confidence intervals. Therefore use of $k = 2$ for our model is supported by the data. Based on a Kaplan-Meier probability plot (Figure 3.3), the placebo group and the treatment group appear to have distinct characters. The early failures for the placebo group suggest a potential downward shift in the scale parameter. In contrast, we do not see an apparent change in the treatment group. Therefore we apply the test of change-point separately.

Out of 53 patients in the control group who were given placebo, 20 are complete cases so 62% of the data are censored. Examination of data as well as estimated survival functions shown in Figure 3.3 indicates that a change may have occurred early on. To be inclusive, we choose a sufficiently wide range $\nu \in [0.05, 0.95]$ for the potential change-point. An estimate $\hat{\theta}_0$ as explained at the end of the previous section is 3.60, and 5% and 1% critical values are 11.30 and 14.79 respectively. The observed test statistic is 37.79, so the null hypothesis is rejected at 1%. This strongly suggests the presence of a change-point, which is estimated to be 23.02 days after treatment while $\hat{\theta}_1(\hat{\nu}) = 64.69$ and $\hat{\theta}_2(\hat{\nu}) = 1.78$. There is a dramatic drop in hazard after the change-point.

In the treatment group, there are 55 patients receiving the treatment and only seven observations are complete cases. The observed likelihood ratio test statistic is 1.63, a rather small value so the null hypothesis is not rejected. Figure 3.3 summarizes the survival functions of both control and treatment groups. Our parametric survival functions match up very well with the usual Kaplan-Meier estimates, and a rapid drop in survival probability followed by a more gradual decrease afterward in the control group gives a visual indication of a change-point early on.

3.4 Proofs

Proof of Weak Convergence Let $\mathbb{F}_n(x) = \sqrt{n}[F_n(x) - F(x)]$ and $\tilde{\mathbb{F}}_n(x) = \sqrt{n}[\tilde{F}_n(x) - \tilde{F}(x)]$ denote the empirical and sub-empirical processes, respectively, where $F(x) = P(X \le x)$, $\tilde{F}(x) = P(X \le x, \delta = 1)$, and F_n and $\tilde{F}_n(x)$ are the corresponding empirical versions of $F(x)$ and $\tilde{F}(x)$.

Expressing $T_j(\nu)$ as T_j and $K_j(\nu)$ as K_j for ease of notation, we note that

$$\frac{T_1}{K_1} - \frac{T_2}{K_2} = \frac{1}{K_1 K_2}\left[K_2\left(T_1 - \frac{K_1}{\theta}\right) - K_1\left(T_2 - \frac{K_2}{\theta}\right)\right].$$

Let $D[0, T]$ denote the Skorohod space on $[0, T]$. Define $\phi : D[0, T] \times D[0, T] \to D[0, T] \times D[0, T]$ such that

$$\phi(g, h)(\nu) = \left(-\int_0^\nu x^{k-1} g(x)dx - \tfrac{1}{\theta}h(\nu), -\int_\nu^T x^{k-1} g(x)dx - \tfrac{1}{\theta}[h(T) - h(\nu)]\right)^t$$

Then $Z_n(\nu)$ can be expressed as

$$Z_n(\nu) = C_n\left(\frac{K_2}{n}, \frac{K_1}{n}\right)\phi\left(\mathbb{F}_n, \tilde{\mathbb{F}}_n\right), \tag{3.5}$$

where

$$C_n = \left(\frac{K_1 + K_2}{K_1 K_2}\right)^{1/2} \frac{n^{3/2}}{T_1 + T_2}$$

We establish weak convergence by using Theorem 3 from Breslow and Crowley (1974), in which $\{\mathbb{F}_n, \tilde{\mathbb{F}}_n\}$ converge weakly to a bivariate process $\{G, \tilde{G}\}$. Kim et al. (2004) note that G has the same distribution as $\mathbb{B}(F)$ and $\tilde{G}$ has the same distribution as $\tilde{\mathbb{B}}(F)$, where $\mathbb{B}$ is a Brownian Bridge. This result, along with the continuity of ϕ, implies that $Z_n(v)$ converges weakly to $Z(v)$, where

$$Z(v) = C(v)[\tilde{F}(T) - \tilde{F}(v), -\tilde{F}(v)]\phi[\mathbb{B}(F), \mathbb{B}(\tilde{F})](v), \tag{3.6}$$

with

$$C(v) = \frac{k}{E(X^k)}\left(\frac{\tilde{F}(T)}{\tilde{F}(v)\left(\tilde{F}(T) - \tilde{F}(v)\right)}\right)^{1/2}.$$

Derivation of covariance To find $Cov(Z(v_1), Z(v_2))$, first let $T_1 - K_1\theta^{-1} = \sum_{i=1}^{n} X_i'(v)$ and $T_2 - K_2\theta^{-1} = \sum_{i=1}^{n} X_i''(v)$. By the central limit theorem, for a given v, $\sqrt{n}(\sum X_i'(v), \sum X_i''(v))$ converges in distribution to the bivariate normal random variable $(Z_1(v), Z_2(v))$ with mean vector $E(X', X'')$ and covariance $Cov(X', X'')$. Note that $E(X'(v)) = E(X''(v)) = 0$, and $Cov(X'(v), X''(v)) = 0$. Let $\Sigma = Cov(X'(v_1), X''(v_1), X'(v_2), X''(v_2))$. Now we can write

$$\begin{bmatrix} Z_n(v_1) \\ Z_n(v_2) \end{bmatrix} = A_n \begin{bmatrix} \frac{1}{\sqrt{n}}\sum_{i=1}^{n} X_i'(v_1) \\ \frac{1}{\sqrt{n}}\sum_{i=1}^{n} X_i''(v_1) \\ \frac{1}{\sqrt{n}}\sum_{i=1}^{n} X_i'(v_2) \\ \frac{1}{\sqrt{n}}\sum_{i=1}^{n} X_i''(v_2) \end{bmatrix} \to^d A \begin{bmatrix} Z_1(v_1) \\ Z_2(v_1) \\ Z_1(v_2) \\ Z_2(v_2) \end{bmatrix} = \begin{bmatrix} Z(v_1) \\ Z(v_2) \end{bmatrix},$$

where

$$\begin{aligned} A_n &= \begin{bmatrix} a_{1n}(v_1) & a_{2n}(v_1) & 0 & 0 \\ 0 & 0 & a_{1n}(v_2) & a_{2n}(v_2) \end{bmatrix}, \\ A &= \begin{bmatrix} a_1(v_1) & a_2(v_1) & 0 & 0 \\ 0 & 0 & a_1(v_2) & a_2(v_2) \end{bmatrix}, \end{aligned}$$

$(a_{1n}(v), a_{2n}(v)) = C_n(v)(K_2/n, K_1/n)$, and $(a_1(v), a_2(v)) = C(v)[\tilde{F}(T) - \tilde{F}(v), -\tilde{F}(v)]$ from (3.5) and (3.6). Finally,

$$A\Sigma A^t = \begin{bmatrix} 1 & \rho_{12} \\ \rho_{21} & 1 \end{bmatrix}.$$

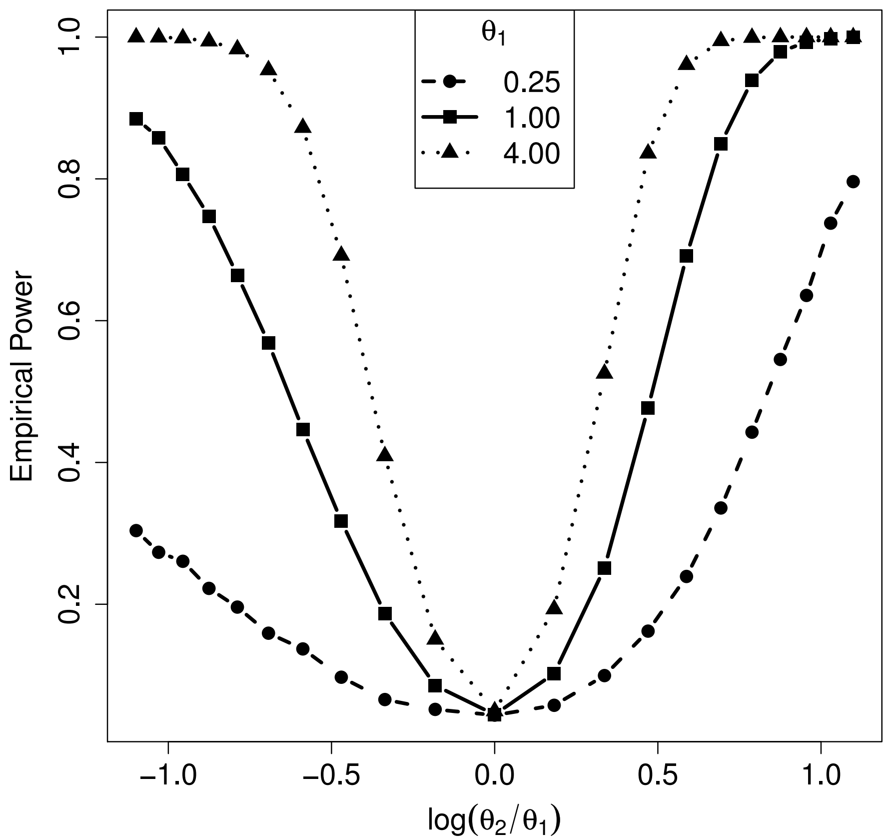

Figure 3.1: Comparing empirical power at the 5% level for different values of θ_1 with a change at $\nu = 0.5$. The expected sample size $\gamma = 500$ and the interval $[a, b] = [0.3, 0.7]$. Each point is based on 10,000 simulations.

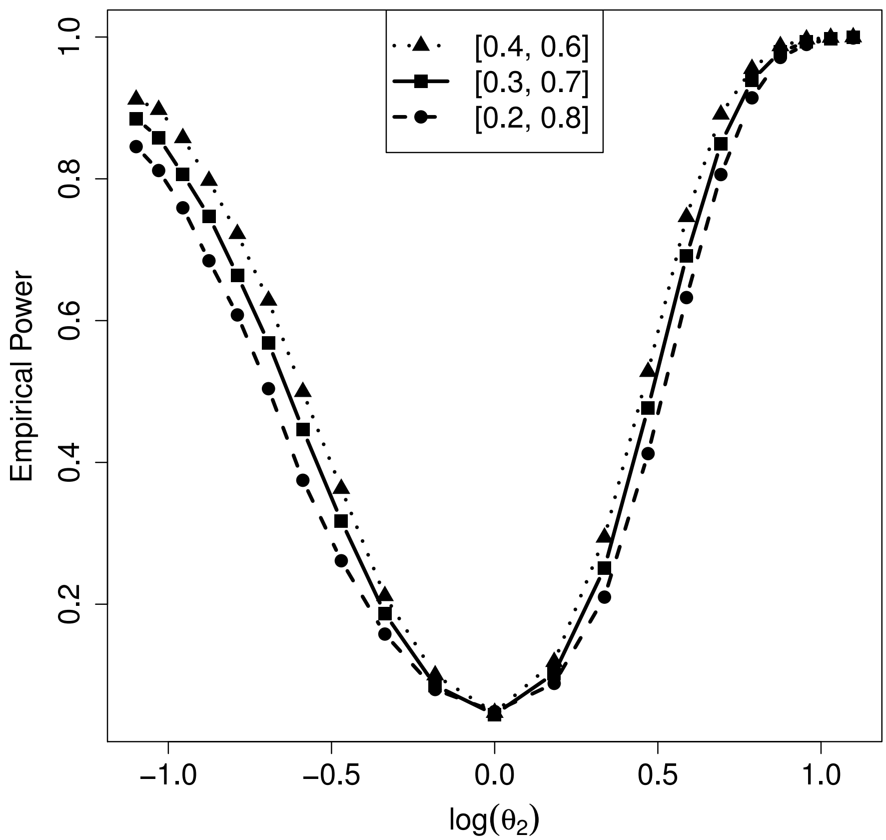

Figure 3.2: Comparing empirical power at the 5% level for different intervals $[a, b]$ with a change at $\nu = 0.5$. The initial slope of the hazard function $\theta_1 = 1$ and the expected sample size $\gamma = 500$. Each point is based on 10,000 simulations.

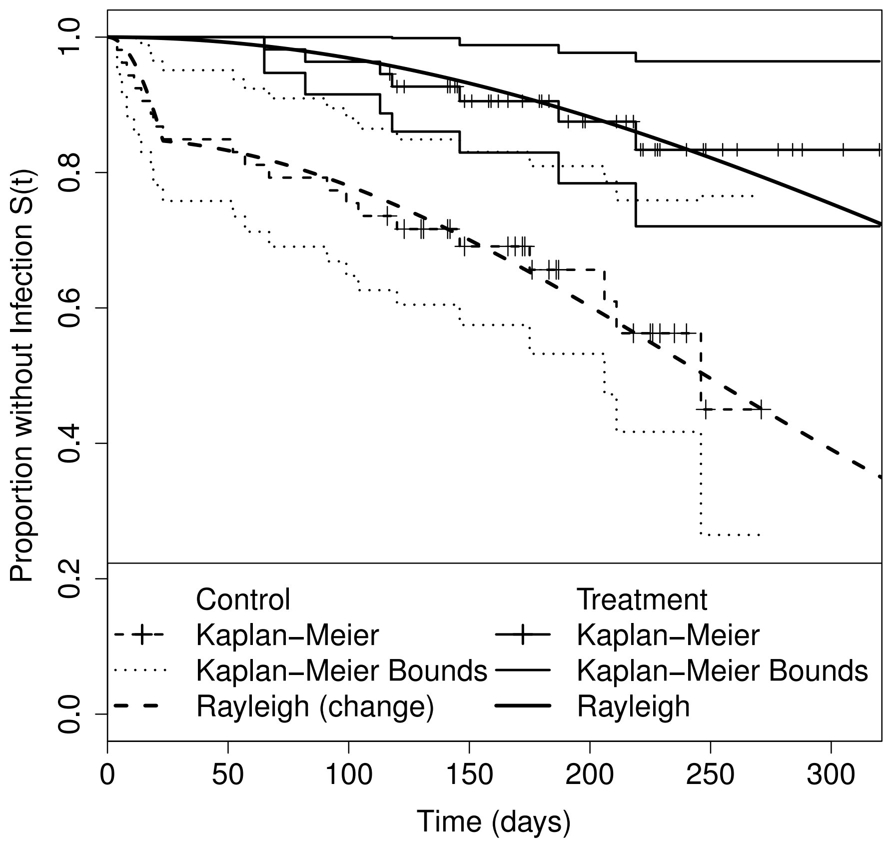

Figure 3.3: Survival curves for control and treatment groups from the chronic granulomatous disease study. Kaplan-Meier estimates line up with Rayleigh models, including a change-point for the control group at 23 days.

CHAPTER 4

CONTINUOUS MONOTONE HAZARDS WITH UNKNOWN CHANGE-POINT

In Chapter 3, we investigated testing for disjoint two-piece hazard functions representing abrupt changes. In this chapter, we consider continuous two-piece hazard functions. One important example is a hazard that is constant up until a change-point then increases in a monotone fashion. In reliability, this is known as random failure followed by a wear-out period. Unlike the step function model, this hazard is a continuous function, but still retains a change-point as nuisance parameter. Park (1988), Na and Lee (2003), and Na et al. (2005) studied a constant hazard against an unspecified trend model in this context.

Here we consider tests for an unknown change-point $\nu \in [a,b] \subset (0,\infty)$ in continuous monotone hazard functions in the presence of uncensored data. Let $\mathbf{t}$ be the vector of n independent identically distributed and uncensored event times t_i for $i = 1,\dots,n$. We examine two models for the hazard function of this time to event data.

$$h_1(t) = \theta_1 \mathbf{1}_{\{t<\nu\}} + \left(\theta_1 + \theta_2(t-\nu)^{(k-1)}\right)\mathbf{1}_{\{t\geq\nu\}}, \tag{4.1}$$

$$h_2(t) = \alpha + \theta_1 t\,\mathbf{1}_{\{t<\nu\}} + (\theta_1\nu + \theta_2(t-\nu))\,\mathbf{1}_{\{t\geq\nu\}}, \tag{4.2}$$

where $\theta_1, \theta_2 > 0$ are unknown parameters, and the shape parameter $k > 1$ may be unknown. $\mathbf{1}_A$ denotes the indicator function for the set A. The hazard $h_1(t)$ is constant up until the change-point ν then monotone increases as a polynomial afterward. Under the null hypothesis $H_0 : \theta_2 = 0$, $h_1(t)$ is a constant hazard. In model (4.2), $h_2(t)$ is two-piece linear function with positive but unknown slopes $\theta_1, \theta_2 > 0$. The intercept $\alpha \geq 0$ may be unknown. Under the null hypothesis $H_0 : \theta_2 = \theta_1$, $h_2(t)$ is a straight line with positive slope. Geometrically speaking, both hazards are of "hockey stick" shapes with varying details under the alternative hypotheses. We note that θ_i , $i = 1, 2$ are generic parameters that have distinct values for each of model (4.1) and model (4.2).

In the next section we present our main result: the weak convergence of the profile log-likelihood ratio test statistics. In section 4.2 we use a local quadratic expansion of the log-likelihood and prove weak convergence by the properties of empirical measure. In section 4.3 we show how to derive the covariance structures for the limiting processes. In the final section, we conduct simulation studies to compare the type I error and power of the new tests to other available tests for some special cases.

4.1 Main Result

Let $\xrightarrow{a.s}$, $\xrightarrow{P}$, $\xrightarrow{d}$, and $\Longrightarrow$ denote almost sure convergence, convergence in probability, convergence in distribution, and weak convergence of a process, respectively. Let $l_\nu(\boldsymbol{\theta};\mathbf{t})$ be the log-likelihood of observations $\mathbf{t} = (t_1, \ldots, t_n)$ for the parameter vector $\boldsymbol{\theta} = (\theta_1, \theta_2, k)^T$ for model (4.1) and $(\theta_1, \theta_2, \alpha)^T$ for model (4.2). Define $\dot{l}_\nu(\boldsymbol{\theta_0};\mathbf{t})$ and $\ddot{l}_\nu(\boldsymbol{\theta_0};\mathbf{t})$ as the first and second derivatives of the log-likelihood with respect to $\boldsymbol{\theta}$ evaluated at $\boldsymbol{\theta} = \boldsymbol{\theta_0}$. The vector $\boldsymbol{\theta_0}$ takes the null hypothesis values $\boldsymbol{\theta_0} = (\theta_0, 0, k)^T$ for model (4.1) and $(\theta_0, \theta_0, \alpha)^T$ for model (4.2). We reparameterize the log-likelihood function of $\boldsymbol{\theta}$ and approximate with a local Taylor expansion:

$$l_\nu(\boldsymbol{\theta};\mathbf{t}) = l_\nu(\boldsymbol{\theta_0} + \mathbf{h}/\sqrt{n};\mathbf{t}) = \sum_{i=1}^{n} l_\nu(\boldsymbol{\theta_0}; t_i) + W_n(\boldsymbol{\theta_0}, \mathbf{h}) \tag{4.3}$$

where

$$W_n(\boldsymbol{\theta_0}, \mathbf{h}) = \sum_{i=1}^{n} \dot{l}_\nu(\boldsymbol{\theta_0}; t_i)^T \mathbf{h}/\sqrt{n} + \mathbf{h}^T \sum_{i=1}^{n} \ddot{l}_\nu(\boldsymbol{\theta_0}; t_i)\mathbf{h}/2n + \sum_{i=1}^{n} R_i(\boldsymbol{\theta_0}, \mathbf{h}),$$

where $\mathbf{h} = (h_1, h_2, h_3)^T$, $h_j \in \mathbb{R}$, and $R_i(\boldsymbol{\theta_0}, \mathbf{h})$ is the remainder term which we revisit in Section 4.2.

For given ν, the profile log-likelihood ratio test statistic $\Lambda_n(\nu)$ is

$$\Lambda_n(\nu) = -(\sup_{\{\mathbf{h}\in\mathcal{H}\}} W_n(\boldsymbol{\theta_0}, \mathbf{h}) - \sup_{\{\mathbf{h}\in\mathcal{H}_0\}} W_n(\boldsymbol{\theta_0}, \mathbf{h})) \tag{4.4}$$

where $\mathscr{H}_0$ is the restricted support of $\mathbf{h}$ under the null hypothesis and $\mathscr{H}$ is the full support.

Define $C_{ij} = Cov(\dot{l}_{\nu_i}(\boldsymbol{\theta_0}; t), \dot{l}_{\nu_j}(\boldsymbol{\theta_0}; t))$ for any $\nu_i, \nu_j \in [a, b]$. For ν_i given, the Fisher information under the null hypothesis is $I_{\boldsymbol{\theta_0}}(\nu_i) = C_{ii}$. Now define $\Sigma_{ij} = I^{-1}_{\boldsymbol{\theta_0}}(\nu_i) C_{ij} I^{-1}_{\boldsymbol{\theta_0}}(\nu_j)$. Let $\{\cdot\}_{lk}$ denote the l, k element of a matrix and let $\{\cdot\}_{-3}$ denote a submatrix with the 3^{rd} row and 3^{rd} column removed. These removed elements correspond to the nuisance parameter k for model (4.1) or α for model (4.2). The 1^{st} and 2^{nd} columns and rows correspond to the θ_1 and θ_2 parameters respectively.

Theorem 4.1.1 *For model (4.1), when H_0: $\theta_2 = 0$ is true, for any $\nu \in [a, b]$, $0 < a < b < \infty$, $2\Lambda_n(\nu)$ converges weakly to $(Z^+(\nu))^2$, where $2\Lambda_n(\nu)$ is in (4.4) with $\boldsymbol{\theta_0} = (\theta_0, 0, k)^T$, $Z^+(\nu) = \max(Z(\nu), 0)$ and $Z(\nu)$ is a zero mean Gaussian process with the covariance function*

$$Cov(Z(\nu_i), Z(\nu_j)) = (\{I_{\boldsymbol{\theta_0}}(\nu_i)\}_{22}\{I_{\boldsymbol{\theta_0}}(\nu_j)\}_{22})^{1/2}\{\Sigma_{ij}\}_{22}$$

Theorem 4.1.2 *For model (4.2), when $H_0 : \theta_1 = \theta_2$ is true, for any $\nu \in [a, b]$, $0 < a < b < \infty$, $2\Lambda_n(\nu)$ converges weakly to $Z(\nu)^2$, where $2\Lambda_n(\nu)$ is in (4.4) with $\boldsymbol{\theta_0} = (\theta_0, \theta_0, \alpha)^T$ and $Z(\nu)$ is a zero mean Gaussian process with the covariance function*

$$Cov(Z(\nu_i), Z(\nu_j)) = \frac{(\{\Sigma_{ij}\}_{11} + \{\Sigma_{ij}\}_{22}) - (\{\Sigma_{ij}\}_{12} + \{\Sigma_{ij}\}_{21})}{\sqrt{\left((I_i^{11} + I_i^{22}) - (I_i^{12} + I_i^{21})\right)\left((I_j^{11} + I_j^{22}) - (I_j^{12} + I_j^{21})\right)}}$$

where $I_i^{lk} = \{\{\{I_{\boldsymbol{\theta_0}}(\nu_i)\}_{-3}\}^{-1}\}_{lk}$.

In Section 4.2 we show that models (4.1) and (4.2) satisfy the conditions for local asymptotic normality (van der Vaart, 1998, p. 104). Then for each ν we have convergence in distribution (van der Vaart, 1998, p. 232): $2\Lambda_n(\nu) \xrightarrow{d} 2\Lambda(\nu)$ with

$$\begin{aligned} 2\Lambda(\nu) &= \inf_{\{\mathbf{h}\in\mathscr{H}_0\}}(\mathbf{X}-\mathbf{h})^T I_{\boldsymbol{\theta_0}}(\mathbf{X}-\mathbf{h}) - \inf_{\{\mathbf{h}\in\mathscr{H}\}}(\mathbf{X}-\mathbf{h})^T I_{\boldsymbol{\theta_0}}(\mathbf{X}-\mathbf{h}) \\ &= (\mathbf{X}-\hat{\mathbf{h}}_0)^T I_{\boldsymbol{\theta_0}}(\mathbf{X}-\hat{\mathbf{h}}_0) - (\mathbf{X}-\hat{\mathbf{h}})^T I_{\boldsymbol{\theta_0}}(\mathbf{X}-\hat{\mathbf{h}}) \end{aligned} \tag{4.5}$$

where $\mathbf{X} \sim N(\mathbf{0}, I^{-1}_{\boldsymbol{\theta_0}})$, and $\sqrt{n}(\sum_{i=1}^n \dot{l}_\nu(\boldsymbol{\theta_0}; t_i)/n) \xrightarrow{d} I_{\boldsymbol{\theta_0}}\mathbf{X}$. Details for the model-specific $\hat{\mathbf{h}} \in \mathscr{H}$ and $\hat{\mathbf{h}}_0 \in \mathscr{H}_0$ are provided in Section 4.3. We then use empirical measure results to demonstrate the weak convergence $2\Lambda_n(\nu) \Longrightarrow 2\Lambda(\nu)$ for all $\nu \in [a, b]$.

Corollary 4.1.3 *For model (4.1),* $\sup_{a\le v\le b} 2\Lambda(v)$ *converges in distribution to* $\sup_{a\le v\le b}[Z^{+}(v)]^2$ *where* $Z^{+}(v)$ *is in Theorem 4.1.1. The significance level* α *for a critical value* c *is then*

$$\alpha = P[\sup_{a\le v\le b}[Z^{+}(v)]^2 > c].$$

Corollary 4.1.4 *For model (4.2),* $\sup_{a\le v\le b} 2\Lambda(v)$ *converges in distribution to* $\sup_{a\le v\le b} Z(v)^2$ *where* $Z(v)$ *is in Theorem 4.1.2. The significance level* α *for a critical value* c *is then*

$$\alpha = P[\sup_{a\le v\le b}[Z(v)]^2 > c].$$

These are direct applications of Slutky's Theorem and the Continuous Mapping Theorem to the results from Theorems 4.1.1 and 4.1.2.

4.2 Proof of Weak Convergence

We consider the likelihood as a density $p_{\boldsymbol{\theta}}(t) = L(\boldsymbol{\theta}; t)$. For our two models $\sqrt{p_{\boldsymbol{\theta}}(t)}$ is continuously differentiable with respect to $\boldsymbol{\theta}$ for every t and v. Furthermore, the Fisher Information matrix $I_{\boldsymbol{\theta}}$ is well defined and continuous with respect to $\boldsymbol{\theta}$ for any $v \in [a,b]$. Then our experiment is differentiable in quadratic mean (van der Vaart, 1998, p. 95). Therefore local asymptotic normality holds for each v (van der Vaart, 1998, p. 104).

Definition (van der Vaart, 1998, p. 269): The *empirical process* evaluated at f is defined as $\mathbb{G}_n f = \sqrt{n}(\mathbb{P}_n f - Pf)$. A class $\mathscr{F}$ of measurable functions $f : \mathscr{X} \mapsto \mathbb{R}$ is called *P-Donsker* (or Donsker) if the sequences of processes $\{\mathbb{G}_n f : f \in \mathscr{F}\}$ converges in distribution to a tight limit process in the space $\ell^\infty(\mathscr{F})$. (Where $\mathbb{P}_n f = n^{-1}\sum_{i=1}^{n} f(x_i)$, $Pf = \int f\,dP = E(f)$, and $\ell^\infty(\mathscr{F})$ is the collection of all bounded functions $z : \mathscr{F} \mapsto \mathbb{R}$.)

For (4.1) and (4.2), we show that all three elements of the vector-valued function $\dot{l}_v(\boldsymbol{\theta_0}; t)$ are in one Donsker class indexed by v. We construct an envelope function $M(t)$ such that $M(t) \ge \|\partial \dot{l}_v(\boldsymbol{\theta_0}; t)/\partial v\|_1$ for $v \in [a,b]$ and $E(|M(t)|) < \infty$. The norm $\|\cdot\|_1$ is the sum of the absolute value of the three elements of $\dot{l}_v(\boldsymbol{\theta_0}; t)$. Note that $M(t)$ is not a function of v. The existence of this envelope function satisfies the conditions for a Donsker class (van der Vaart, 1998, p. 271).

For our two models, if $M = \sup_{\{t\in(0,\infty),\ v\in[a,b]\}} \|\partial\, \dot{l}_v(\boldsymbol{\theta_0}; t)/\partial v\|_1 < \infty$ then set $M(t) = M$. Otherwise construct $M(t)$ in three pieces:

$$M(t) = \begin{cases} \|\partial\, \dot{l}_v(\boldsymbol{\theta_0}; t)/\partial v\|_1,\ t < a \le v \\ \max\{\|\partial\, \dot{l}_v(\boldsymbol{\theta_0}; t)/\partial v\|_1;\ t \in [a, b^*],\ v \in [a,b]\},\ a \le t \le b^* \\ \|\partial\, \dot{l}_v(\boldsymbol{\theta_0}; t)/\partial v\|_1,\ v = v^* \in [a,b],\ t > b^* \end{cases}$$

where $b^* \ge b$ and v^* are selected so that $v^* \in [a,b]$, $\|\partial\, \dot{l}_v(\boldsymbol{\theta_0}; t)/\partial v\|_1$ evaluated at $v = v^*$ dominates all curves with other values of $v \in [a,b]$ when $t > b^*$, and the interval $[a, b^*]$ is finite. Such b^* and v^* can be selected for a given $\boldsymbol{\theta_0}$ vector.

For model (4.1): $\|\partial\, \dot{l}_v(\boldsymbol{\theta_0}; t)/\partial v\|_1 = |(t-v)^{k-2}(1-k+\theta_0(t-v))/\theta_0\ \mathbf{1}_{\{t\ge v\}}|$ and its expectation $E(\|\partial\, \dot{l}_v(\boldsymbol{\theta_0}; t)/\partial v\|_1) = 2\theta_0 \exp(1-k-\theta_0 v)(k-1/\theta_0)^{k-1}$ which is finite for all $v \in [a,b]$. For model (4.2): $\|\partial\, \dot{l}_v(\boldsymbol{\theta_0}; t)/\partial v\|_1 = 2|1/(\alpha_0 + t\theta_0) - (t-v)\ \mathbf{1}_{\{t\ge v\}}|$. Directly evaluating $E(\|\partial\, \dot{l}_v(\boldsymbol{\theta_0}; t)/\partial v\|_1)$ involves many terms. Instead we use $E((\|\partial\, \dot{l}_v(\boldsymbol{\theta_0}; t)/\partial v\|_1)^2) = \exp(\alpha^2/\theta_0)\int_{(\alpha+\theta_0 v)^2/2\theta_0}^{\infty} e^{-t}/t\ dt$ which is finite for $v \in [a,b]$. Then $E(\|\partial\, \dot{l}_v(\boldsymbol{\theta_0}; t)/\partial v\|_1)$ must also be finite for $v \in [a,b]$. By our construction $E(|M(t)|) < \infty$ for both models. Then for $v \in [a,b]$, $\sqrt{n}(\sum_{i=1}^n \dot{l}_v(\boldsymbol{\theta_0}; t_i)/n) \Longrightarrow I_{\boldsymbol{\theta_0}}(v)X(v)$, where $X(v)$ is a zero mean Gaussian process with $Cov(X(v_i), X(v_j)) = \Sigma_{ij}$.

Definition (van der Vaart, 1998, p. 269): A class $\mathscr{F}$ of measurable functions $f : \mathscr{X} \mapsto \mathbb{R}$ is called *P-Glivenko-Cantelli* (or Glivenko-Cantelli) if $\|\mathbb{P}_n f - Pf\|_{\mathscr{F}} = \sup_{f\in\mathscr{F}} |\mathbb{P}_n f - Pf| \xrightarrow{a.s.} 0$. (Where $\mathbb{P}_n f = n^{-1}\sum_{i=1}^n f(x_i)$ and $Pf = \int f\, dP = E(f)$)

For both models, we show that each element in the matrix of functions $\ddot{l}_v(\boldsymbol{\theta_0}; t)$ are in one Glivenko-Cantelli class indexed by v. We can construct an envelope function $F(t)$ such that $F(t) \ge \|\ddot{l}_v(\boldsymbol{\theta_0}; t)\|_1$ for $v \in [a,b]$ and $E(|F(t)|) < \infty$. This process is very similar to that of constructing $M(t)$. The existence of this envelope function satisfies the conditions for a Glivenko-Cantelli class (van der Vaart, 1998, p. 272). Then $\sup_{v\in[a,b]} \|n^{-1}\sum_{i=1}^n \ddot{l}_v(\boldsymbol{\theta_0}; t_i) - I_{\boldsymbol{\theta_0}}\|_1 \xrightarrow{P} 0$.

We next consider the remainder $R_i(\boldsymbol{\theta_0}, \mathbf{h}) = R_i^1(\boldsymbol{\theta_0}, \mathbf{h}) - \mathbf{h}^T \ddot{l}_v(\boldsymbol{\theta_0}; t_i)\mathbf{h}/2n$ where $R_i^1(\boldsymbol{\theta_0}, \mathbf{h}) = l_v(\boldsymbol{\theta_0} + \mathbf{h}/\sqrt{n}; t_i) - l_v(\boldsymbol{\theta_0}; t_i) - \mathbf{h}^T \dot{l}_v(\boldsymbol{\theta_0}; t_i)/\sqrt{n}$. By Taylor's Theorem: $R_i^1(\boldsymbol{\theta_0}, \mathbf{h}) = \mathbf{h}^T \ddot{l}_v(\boldsymbol{\theta_0} + \mathbf{h}_\phi/\sqrt{n}; t_i)\mathbf{h}/2n$, where $\mathbf{h}_\phi^T = (\phi_1 h_1, \ldots, \phi_d h_d)$ for some $\phi_1, \ldots, \phi_d \in [0,1]$ with d the dimension of $\boldsymbol{\theta}$. For fixed $\boldsymbol{\theta_0}$ and $\mathbf{h}$ and for n large enough, we can construct an appropriate envelope function $F(t)$ which now covers functions $\ddot{l}_v(\boldsymbol{\theta_0}+\mathbf{h}_\phi/\sqrt{n};t)$ indexed by the pair $\{v,\boldsymbol{\theta}\}\in[a,b]\times[\boldsymbol{\theta_0},\boldsymbol{\theta_0}+\mathbf{h}/\sqrt{n}]$. Then $\ddot{l}_v(\boldsymbol{\theta_0}+\mathbf{h}_\phi/\sqrt{n};t)$ is Glivenko-Cantelli since $[\boldsymbol{\theta_0},\boldsymbol{\theta_0}+\mathbf{h}_\phi/\sqrt{n}]\subset[\boldsymbol{\theta_0},\boldsymbol{\theta_0}+\mathbf{h}/\sqrt{n}]$. We can construct an appropriate envelope function for the difference $\ddot{l}_v(\boldsymbol{\theta_0}+\mathbf{h}_\phi/\sqrt{n};t)-\ddot{l}_v(\boldsymbol{\theta_0};t)$ using the triangle inequality, so it is also Glivenko-Cantelli. Thus $\sup_{v\in[a,b]}|\sum_{i=1}^n R_i(\boldsymbol{\theta_0},\mathbf{h})|\xrightarrow{P}0$.

4.3 Specifying the Covariance Structures

In order to characterize the $2\Lambda(v)$ process, we will need

$$Cov(\mathbf{X}(v_i),\mathbf{X}(v_j))=I_{\boldsymbol{\theta_0}}^{-1}(v_i)C_{ij}I_{\boldsymbol{\theta_0}}^{-1}(v_j).$$

We can derive $Cov(\dot{l}_{v_i}(\boldsymbol{\theta_0};t),\dot{l}_{v_j}(\boldsymbol{\theta_0};t))=C_{ij}$ directly via integration. Once we characterize the process, we can simulate or approximate critical values for testing.

4.3.1 Model 1

For model (4.1) we use the following reparameterization: $l_v(\theta_1,\theta_2,k;t)=l_v(\theta_0+h_1/\sqrt{n},0+h_2/\sqrt{n},k+h_3/\sqrt{n};t)$, with $h_2=0$ under the null hypothesis and $h_2>0$ under the alternative. Then for (4.5): $\mathbf{X}=(X_1,X_2,X_3)^T$, $\hat{\mathbf{h}}_0=(X_1,0,X_3)^T$, and $\hat{\mathbf{h}}=(X_1,X_2^+,X_3)^T$. Then $2\Lambda(v)=X_2\{I_{\boldsymbol{\theta_0}}\}_{22}X_2-X_2^-\{I_{\boldsymbol{\theta_0}}\}_{22}X_2^-=\{I_{\boldsymbol{\theta_0}}\}_{22}(X_2^+)^2$ where $X_2^-=\min(X_2,0)$, and $X_2^+=\max(X_2,0)$. Then $2\Lambda(v)=(Z^+(v))^2$ as in Theorem 4.1.1.

As a concrete example, we consider $k=2$ case. Then for $v_j\geq v_i$

$$C_{ij}=\begin{bmatrix}1/\theta_0^2 & \exp(-\theta_0 v_j)/\theta_0^3\\ \exp(-\theta_0 v_i)/\theta_0^3 & \exp(-\theta_0 v_j)(2+\theta_0(v_j-v_i))/\theta_0^4\end{bmatrix}$$

and simplifying $(\{I_{\boldsymbol{\theta_0}}(v_i)\}_{22}\{I_{\boldsymbol{\theta_0}}(v_j)\}_{22})^{1/2}\{\Sigma_{ij}\}_{22}$,

$$Cov(Z(v_i),Z(v_j))=\frac{2e^{\theta_0(v_i+v_j)/2}\left(e^{\theta_0 v_i}(\theta_0(v_j-v_i)+2)-1\right)}{\left(2e^{\theta_0 v_i}-1\right)\left(2e^{\theta_0 v_j}-1\right)}\tag{4.6}$$

We note that $Var(Z(v))=2e^{\theta_0 v}/(2e^{\theta_0 v}-1)$, which is between 1 and 2 for $v\in[a,b]\subset(0,\infty)$.

4.3.2 Model 2

For model (4.2) we reparameterize $l_\nu(\theta_1,\theta_2,\alpha;t) = l_\nu(\theta_0 + h_1/\sqrt{n}, \theta_0 + h_2/\sqrt{n}, \alpha + h_3/\sqrt{n}; t)$. Under the null hypothesis $h_1 = h_2$. Then in (4.5): $\mathbf{X} = (X_1, X_2, X_3)^T$, $\hat{\mathbf{h}} = (X_1, X_2, X_3)^T$, and $\hat{\mathbf{h}}_0 = (\hat{h}_0^T, \hat{h}_0^T, X_3)^T = (\hat{h}_0^T \mathbf{1}^T, X_3)^T$ with $\hat{h}_0 = (\mathbf{1}^T I_{-3}\mathbf{1})^{-1}\mathbf{1}^T X_{-3}$, where $\mathbf{X}_{-3} = (X_1, X_2)^T$, $\mathbf{1} = (1,1)^T$ is the vector of ones, and I_{-3} is the submatrix of $I_{\boldsymbol{\theta_0}}$ after removing the 3^{rd} column and row.. Then $2\Lambda(\nu) = \mathbf{X}_{-3}^T(I_{-3} - I_{-3}\mathbf{1}(\mathbf{1}^T I_{-3}\mathbf{1})^{-1}\mathbf{1}^T I_{-3})\mathbf{X}_{-3} = \mathbf{X}_{-3}^T Q \mathbf{X}_{-3}$. It is clear that Q is symmetric and $\mathbf{1}^T Q = 0$. Furthermore $\{Q\}_{11} = |I_{-3}|/\mathbf{1}^T I_{-3}\mathbf{1} = 1/((I^{11} + I^{22}) - (I^{12} + I^{21}))$, where $I^{lk} = \{\{\{I_{\boldsymbol{\theta_0}}(\nu)\}_{-3}\}^{-1}\}_{lk}$ and $|I_{-3}|$ is the determinant of the matrix I_{-3}. Then $2\Lambda(\nu) = \{Q\}_{11}(X_1 - X_2)^2 = Z^2(\nu)$ as in Theorem 4.1.2. If α is known, then $Z^2(\nu)$ is a chi-square process and $Var(Z(\nu_i), Z(\nu_j)) = 1$. Unlike for model (4.1), Σ_{ij} does not take any simpler form for special cases.

4.4 Simulation Study: Type I Error and Power

To obtain critical values, we simulate vectors of normal variates $Z(\nu)$ over a grid of 1000 values of $\nu \in [a,b]$. For model (4.1) we assume $k = 2$, use the simpler covariance (4.6), and find $\sup_{\nu\in[a,b]}(Z^+(\nu))^2$. For model (4.2) we use the covariance from Theorem 4.1.2, assume $\alpha = 0$, and then find $\sup_{\nu\in[a,b]}(Z(\nu))^2$. We repeat 10,000 times and use the 95% and 99% quantiles as the critical values for significance levels of 5% and 1% respectively. In practice we would replace $\boldsymbol{\theta_0}$ by a consistent estimator. However, for this discussion we will treat it as known. We choose $\theta_0 = 1$ with the range of $\nu \in [a,b]$ equal to the 25% and 75% expected percentiles of the data under the null hypothesis. Then the approximate upper 5^{th} and 1^{st} percentiles are 4.44 and 8.98 for model (4.1) and 5.65 and 8.87 for model (4.2).

We compare each test to an alternative test already available in the literature. For model (4.1), we can use the test by Na and Lee (2003) which compares exponential life (constant hazard) against an alternative of increasing then decreasing mean residual life (IDMRL). The non-decreasing hazard function (4.1) satisfies the IDMRL property. We simulate data under the alternative $\theta_1 = 1$ and $\theta_2 = 0, 1/5, 1/2, 1, 2, 5$ using 50 observations with the true change ν equal to the expected 50% quantile. The simulation is repeated $10,000$ times for each setting. Type I error (where $\theta_2 = 0$) for our proposed test is close to the nominal level and the test has much higher power than the IDMRL test (Table 4.1).

Table 4.1: Power simulations for Constant then Increasing Hazard Function ($s.e. < 0.005$)

θ_2	0	0.2	0.5	1	2	5
LAN 5%	0.056	0.165	0.370	0.633	0.884	0.993
LAN 1%	0.007	0.027	0.086	0.239	0.543	0.901
IDMRL 5%	0.054	0.097	0.186	0.339	0.568	0.851
IDMRL 1%	0.010	0.021	0.050	0.120	0.284	0.614

Table 4.2: Power simulations for Two-Piece Linear Hazard Function ($s.e. < 0.005$)

θ_2	0.1	0.2	0.5	1	2	5	10
LAN 5%	0.333	0.228	0.070	0.039	0.149	0.647	0.932
LAN 1%	0.176	0.096	0.017	0.007	0.046	0.376	0.798
Davies 5%	0.385	0.274	0.088	0.041	0.151	0.629	0.914
Davies 1%	0.211	0.124	0.023	0.008	0.044	0.355	0.761

Next we examine power for model (4.2). Since we assume $\alpha = 0$ is known, $2\Lambda(\nu)$ is a chi-square process. We can then use results from Davies (1977, 1987) to approximate the p-values to our log-likelihood ratio test. Again we simulate data under the alternative $\theta_1 = 1$ and $\theta_2 = 1/10, 1/5, 1/2, 1, 2, 5, 10$ using 50 observations with ν equal to the expected 50% quantile. The simulation is repeated $10,000$ times for each setting. From Table 4.2, both tests are conservative in terms of type I error ($\theta_2 = 1$). Our method has slightly higher power for $\theta_2 > \theta_1$ while the Davies approximation performs somewhat better for $\theta_2 < \theta_1$, indicating that they are overall comparable in this scenario. On the other hand, the current method is more general than Davies' approximation because the former can be used regardless of whether the limiting process is a chi-square process or not.

4.5 Conclusions

We have outlined a method to approximate the likelihood ratio test for an unknown change-point in two different hazard functions. We used general tools of empirical measure (Donsker and Glivenko-Cantelli classes) to prove weak convergence of the test statistic to a quadratic process. Critical values for likelihood ratio tests can be computed even when the limiting process is not a chi-square process. Simulation results for some special cases demonstrate the potential of this flexible approach, showing improvement over existing tests. For simplicity we assumed no censoring in this Chapter, but in Chapter 5 we demonstrate that censoring mechanisms can be accommodated with minor modifications.

4.6 Supplemental: Derivatives

We include the first and second derivatives evaluated at $\theta = \theta_0$. From these, we can construct envelope functions $M(t)$ and $F(t)$. We can also calculate $Cov(\dot{l}_{\nu_i}(\boldsymbol{\theta_0};t), \dot{l}_{\nu_j}(\boldsymbol{\theta_0};t))$ through integration. Any derivatives not included (such as $\partial \ln L/\partial \alpha$) are equal to zero.

4.6.1 Model 1

$$
\begin{aligned}
\partial l_\nu(\boldsymbol{\theta};t)/\partial\theta_1 &= \tfrac{1}{\theta_0} - t \\
\partial l_\nu(\boldsymbol{\theta};t)/\partial\theta_2 &= \tfrac{1}{k\theta_0}(t-\nu)^{k-1}(k-\theta_0(t-\nu)),\ t \geq \nu \\
\partial^2 l_\nu(\boldsymbol{\theta};t)/\partial\nu\partial\theta_2 &= \tfrac{1}{\theta_0}(t-\nu)^{k-2}(1-k+\theta_0(t-\nu)),\ t \geq \nu \\
\partial^2 l_\nu(\boldsymbol{\theta};t)/\partial\theta_1^2 &= -\tfrac{1}{\theta_0^2} \\
\partial^2 l_\nu(\boldsymbol{\theta};t)/\partial\theta_2\partial\theta_1 &= -\tfrac{1}{\theta_0^2}(t-\nu)^{k-1},\ t \geq \nu \\
\partial^2 l_\nu(\boldsymbol{\theta};t)/\partial\theta_2^2 &= -\tfrac{1}{\theta_0^2}(t-\nu)^{2(k-1)},\ t \geq \nu \\
\partial^2 l_\nu(\boldsymbol{\theta};t)/\partial k\partial\theta_2 &= -\tfrac{1}{k^2\theta_0}(t-\nu)^{k-1}(\theta_0(t-\nu)+k(k-\theta_0(t-\nu))\ln(t-\nu)),\ t \geq \nu
\end{aligned}
$$

4.6.2 Model 2

$$
\begin{aligned}
\partial l_\nu(\boldsymbol{\theta};t)/\partial\theta_1 &= \tfrac{-t\left(-2+t\alpha_0+t^2\theta_0\right)}{2(\alpha_0+t\theta_0)}\mathbf{1}_{\{t<\nu\}}+\tfrac{\nu\left(2-2t\alpha_0-2t^2\theta_0+\alpha_0\nu+t\theta_0\nu\right)}{2(\alpha_0+t\theta_0)}\mathbf{1}_{\{t\geq\nu\}}\\
\partial l_\nu(\boldsymbol{\theta};t)/\partial\theta_2 &= -\tfrac{(t-\nu)\left(-2+t^2\theta_0-\alpha_0\nu+t(\alpha_0-\theta_0\nu)\right)}{2(\alpha_0+t\theta_0)},\ t\geq\nu\\
\partial l_\nu(\boldsymbol{\theta};t)/\partial\alpha &= -\tfrac{t^2\theta_0+t\alpha_0-1}{\alpha_0+t\theta_0}\\
\partial^2 l_\nu(\boldsymbol{\theta};t)/\partial\nu\partial\theta_1 &= \tfrac{1}{\alpha_0+t\theta_0}-(t-\nu),\ t\geq\nu\\
\partial^2 l_\nu(\boldsymbol{\theta};t)/\partial\nu\partial\theta_2 &= -\tfrac{1}{\alpha_0+t\theta_0}+(t-\nu),\ t\geq\nu\\
\partial^2 l_\nu(\boldsymbol{\theta};t)/\partial\theta_1^2 &= \tfrac{-t^2}{(\alpha_0+t\theta_0)^2}\mathbf{1}_{\{t<\nu\}}+\tfrac{-\nu^2}{(\alpha_0+t\theta_0)^2}\mathbf{1}_{\{t\geq\nu\}}\\
\partial^2 l_\nu(\boldsymbol{\theta};t)/\partial\theta_2\partial\theta_1 &= \tfrac{-\nu(t-\nu)}{(\alpha_0+t\theta_0)^2},\ t\geq\nu\\
\partial^2 l_\nu(\boldsymbol{\theta};t)/\partial\alpha\partial\theta_1 &= \tfrac{-t}{(\alpha_0+t\theta_0)^2}\mathbf{1}_{\{t<\nu\}}+\tfrac{-\nu}{(\alpha_0+t\theta_0)^2}\mathbf{1}_{\{t\geq\nu\}}\\
\partial^2 l_\nu(\boldsymbol{\theta};t)/\partial\theta_2^2 &= \tfrac{-(t-\nu)^2}{(\alpha_0+t\theta_0)^2},\ t\geq\nu\\
\partial^2 l_\nu(\boldsymbol{\theta};t)/\partial\alpha\partial\theta_2 &= \tfrac{-(t-\nu)}{(\alpha_0+t\theta_0)^2},\ t\geq\nu\\
\partial^2 l_\nu(\boldsymbol{\theta};t)/\partial\alpha^2 &= \tfrac{-1}{(\alpha_0+t\theta_0)^2},\ t\geq\nu
\end{aligned}
$$

CHAPTER 5

ABRUPT CHANGE IN THE PRESENCE OF COVARIATE INFORMATION

We now return to the two-piece abrupt hazard function from Chapter 3. In this chapter, we reparameterize to accommodate covariate information through a linear predictor. We are now able to consider an abrupt change after which some of coefficients in the linear predictor change. The resulting hazard function is the following:

$$h(t_i;\mathbf{z}_i,\mathbf{w}_i) = \begin{cases} t_i^{k-1}\exp\left(\mathbf{z}_i^T\boldsymbol{\beta}_1 + \mathbf{w}_i^T\boldsymbol{\alpha}\right), & t_i < \nu \\ t_i^{k-1}\exp\left(\mathbf{z}_i^T\boldsymbol{\beta}_2 + \mathbf{w}_i^T\boldsymbol{\alpha}\right), & t_i \geq \nu \end{cases} \tag{5.1}$$

where $\mathbf{z}_i$ and $\mathbf{w}_i$ are covariate vectors and $\boldsymbol{\beta}_1$, $\boldsymbol{\beta}_2$, and $\boldsymbol{\alpha}$ are their coefficients. The shape parameter $k > 0$ is known, but $\boldsymbol{\alpha}$ may be unknown nuisance parameters. Under H_0 $\boldsymbol{\beta}_1 = \boldsymbol{\beta}_2 = \boldsymbol{\beta}_0$ making the change-point ν vanish.

If we consider the exponential case ($k = 1$), Dupuy (2009) has derived non-asymptotic bounds for significance and type II error rates assuming all coefficients change and data are subject to random right censoring. If we consider semi-parametric approaches using the partial likelihood, then some complexity from the baseline hazard $\lambda_0(t_i) = t_i^{k-1}$ is avoided. In this framework, there are several results for the case when β is a scalar: O'Quigley and Natarajan (2004) demonstrate the use of Davies' approximation (1987) for the score test for a case equivalent to when $\alpha = 0$. Liang et al. (1990) demonstrate the convergence of the score test process to an Ornstein-Uhlenbeck process if a technical assumption about the relationship between β and α is made. Liu et al. (2008) note that this technical assumption is difficult to assess in practice and instead develop a Monte-Carlo method to generate the distribution for the score test statistic. They demonstrate that results can be extended to multiple change-points (although the changing coefficient β is still a scalar) and provide a simulation study example with two change-points.

We develop methodology for the fully parametric case (5.1) that is comparable and complementary to these results and includes the case where $\boldsymbol{\beta}$ is a vector. We demonstrate by assuming a censoring structure, but we will show that our method can be used for other censoring regimes. We now consider the case of staggered entry and type I censoring from Chapter 3. Subjects enter at times $\tau_1 < \ldots < \tau_n$, following a Poisson process. The ending time of the study T_R is independent of the number of subjects or events. Conditioning on the total number of subjects, $N(T_R) = n$, who entered the study before time T_R, the censoring time of each subject is a sample from a uniform distribution, $(T_R - \tau_i) \sim U(0, T_R)$, where τ_i is the time the i^{th} subject entered the study.

The underlying data t follow model (5.1), but we observe the pair $\{x, \delta\}$, where $x = \min(t, T_R - \tau)$ and δ is the indicator $\mathbf{1}_{\{t<T_R-\tau\}}$. The likelihood for an observation$\{x,\delta\}$ is:

$$
\begin{aligned}
L(\theta; x, \delta) = & \left(x^{k-1}\exp(\mathbf{z}^T\boldsymbol{\beta}_1 + \mathbf{w}^T\boldsymbol{\alpha})\right)^{\delta} \exp\left[-x^k \exp(\mathbf{z}^T\boldsymbol{\beta}_1 + \mathbf{w}^T\boldsymbol{\alpha})/k\right] \mathbf{1}_{\{x<\nu\}} \\
& + \left(x^{k-1}\exp(\mathbf{z}^T\boldsymbol{\beta}_2 + \mathbf{w}^T\boldsymbol{\alpha})\right)^{\delta} \exp\left[-(\exp(\mathbf{z}^T\boldsymbol{\beta}_1)\nu^k\right. \\
& \left. + \exp(\mathbf{z}^T\boldsymbol{\beta}_2)(x^k - \nu^k))\exp(\mathbf{w}^T\boldsymbol{\alpha})/k\right] \mathbf{1}_{\{x\geq\nu\}}
\end{aligned}
$$

Under the null hypothesis, $f(x)$ the marginal density for x and $\tilde{f}(x)$ the conditional density for $\{x, \delta = 1\}$ (both conditioned on $\mathbf{z}, \mathbf{w}$) are the following:

$$
\begin{aligned}
f(x) &= \exp(-x^k \exp(\mathbf{z}^T\boldsymbol{\beta} + \mathbf{w}^T\boldsymbol{\alpha})/k)\left(1 + x^{k-1}(T_R - x)\exp(\mathbf{z}^T\boldsymbol{\beta} + \mathbf{w}^T\boldsymbol{\alpha})/k)\right)/T_R \\
\tilde{f}(x) &= \exp(-x^k \exp(\mathbf{z}^T\boldsymbol{\beta} + \mathbf{w}^T\boldsymbol{\alpha})/k)\left(x^{k-1}(T_R - x)\exp(\mathbf{z}^T\boldsymbol{\beta} + \mathbf{w}^T\boldsymbol{\alpha})/k)\right)/T_R
\end{aligned} \tag{5.2}
$$

We will use $f(x)$ and $\tilde{f}(x)$ later to evaluate the expectations $E_{x,\delta|\mathbf{z},w}(\cdot)$.

5.1 Main Result

Let $\xrightarrow{a.s}$, $\xrightarrow{P}$, $\xrightarrow{d}$, and $\Longrightarrow$ represent almost sure convergence, convergence in probability, convergence in distribution, and weak convergence of a process respectively. Let $l_\nu(\boldsymbol{\theta}; x, \delta, \mathbf{z}, \mathbf{w})$ be the log-likelihood of the parameter vector $\boldsymbol{\theta} = (\boldsymbol{\beta}_1^T, \boldsymbol{\beta}_2^T, \boldsymbol{\alpha}^T)^T$ for model (5.1) with an observation $\{x, \delta\}$, covariate vectors $\mathbf{z}$ and $\mathbf{w}$ and a given change-point ν.

Now define $\dot{l}_v(\boldsymbol{\theta_0};x,\delta,\mathbf{z},\mathbf{w})$ and $\ddot{l}_v(\boldsymbol{\theta_0};x,\delta,\mathbf{z},\mathbf{w})$ as the first and second derivatives of the log-likelihood with respect to $\boldsymbol{\theta}$ evaluated at $\boldsymbol{\theta} = \boldsymbol{\theta_0}$. The vector $\boldsymbol{\theta_0}$ takes the null hypothesis values $(\boldsymbol{\beta}_0^T, \boldsymbol{\beta}_0^T, \boldsymbol{\alpha}_0^T)^T$. Now consider $\{\mathbf{x}, \boldsymbol{\delta}\}$ the paired vectors of observations $\{x_1, \delta_1\}, \ldots, \{x_n, \delta_n\}$ with corresponding covariates matrices $\mathbf{Z}$ and $\mathbf{W}$. We reparameterize the log-likelihood and approximate with a local Taylor expansion: $l_v(\boldsymbol{\theta};\mathbf{x},\boldsymbol{\delta},\mathbf{Z},\mathbf{W}) = l_v(\boldsymbol{\theta_0}+\mathbf{h}/\sqrt{n};\mathbf{x},\boldsymbol{\delta},\mathbf{Z},\mathbf{W}) = \sum_{i=1}^n l_v(\boldsymbol{\theta_0};x_i,\delta_i,\mathbf{z}_i,\mathbf{w}_i) + W_n(\boldsymbol{\theta_0},\mathbf{h})$ and

$$W_n(\boldsymbol{\theta_0},\mathbf{h}) = \sum_{i=1}^n \dot{l}_v(\boldsymbol{\theta_0};x_i,\delta_i,\mathbf{z}_i,\mathbf{w}_i)^T\mathbf{h}/\sqrt{n} + \mathbf{h}^T\sum_{i=1}^n \ddot{l}_v(\boldsymbol{\theta_0};x_i,\delta_i,\mathbf{z}_i,\mathbf{w}_i)\mathbf{h}/2n + \sum_{i=1}^n R_i(\boldsymbol{\theta_0},\mathbf{h})$$

where $\mathbf{h}$ is a vector of the same dimension as $\boldsymbol{\theta}$, and $R_i(\boldsymbol{\theta_0},\mathbf{h})$ is the remainder term which we revisit in Section 5.2. The log-likelihood ratio test statistic (given v) is now:

$$2\Lambda_n(v) = -2\left(\sup_{\{\mathbf{h}\in\mathscr{H}\}} W_n(\boldsymbol{\theta_0},\mathbf{h}) - \sup_{\{\mathbf{h}\in\mathscr{H}_0\}} W_n(\boldsymbol{\theta_0},\mathbf{h})\right)$$

where $\mathscr{H}_0$ is the restricted support of $\mathbf{h}$ under the null hypothesis and $\mathscr{H}$ is the full support.

Define $C_{ij} = Cov(\dot{l}_{v_i}(\boldsymbol{\theta_0};x,\delta,\mathbf{z},\mathbf{w}), \dot{l}_{v_j}(\boldsymbol{\theta_0};x,\delta,\mathbf{z},\mathbf{w}))$. For v_i given, the Fisher information under the null hypothesis is $I_{\boldsymbol{\theta_0}}(v_i) = C_{ii}$. Define $I_{\boldsymbol{\beta_0}}(v_i)$ as the subset of the Fisher information applying only to $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$. Now define $\Sigma_{ij} = I_{\boldsymbol{\theta_0}}^{-1}(v_i)C_{ij}I_{\boldsymbol{\theta_0}}^{-1}(v_j)$. Let $\mathbf{X} = (\mathbf{X}_{\beta_1}^T, \mathbf{X}_{\beta_2}^T, \mathbf{X}_\alpha)^T$ be a multivariate Gaussian process with mean zero and covariance Σ_{ij}. Let $\mathbf{X}_\beta = (\mathbf{X}_{\beta_1}^T, \mathbf{X}_{\beta_2}^T)^T$. Finally let $I_{d,2,1} = [I_d, I_d]^T$ where I_d is the identity matrix of dimension $d \times d$, where d is the length of $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$.

Theorem 5.1.1 *For model (5.1) when H_0: $\boldsymbol{\beta}_1 = \boldsymbol{\beta}_2$ is true: for any $v \in [a,b]$ and $0 < a < b < \infty$, $2\Lambda_n(v)$ converges weakly to the quadratic process $\mathbf{X}_\beta^T\mathbf{Q}\mathbf{X}_\beta(v)$, where*

$$\mathbf{Q} = I_{\boldsymbol{\beta_0}} - I_{\boldsymbol{\beta_0}}I_{d,2,1}(I_{d,2,1}^T I_{\boldsymbol{\beta_0}} I_{d,2,1})^{-1} I_{d,2,1}^T I_{\boldsymbol{\beta_0}}$$

Corollary 5.1.2 *Assume $\boldsymbol{\alpha}$ is known. Then $\mathbf{X}_\beta^T\mathbf{Q}\mathbf{X}_\beta(v)$ is a χ_d^2 process with d degrees of freedom.*

Example If we assume $\boldsymbol{\alpha} = \mathbf{0}$, then all covariate coefficients change after v.

Corollary 5.1.3 *When β_1 and β_2 are scalars, then $\mathbf{X}_\beta^T \mathbf{Q} \mathbf{X}_\beta(v) = Z^2(v)$ where $Z(v)$ is a mean zero Gaussian process with covariance*

$$Cov(Z(v_i), Z(v_j)) = \frac{(\{\Sigma_{ij}\}_{11} + \{\Sigma_{ij}\}_{22}) - (\{\Sigma_{ij}\}_{12} + \{\Sigma_{ij}\}_{21})}{\sqrt{\left((I_i^{11} + I_i^{22}) - (I_i^{12} + I_i^{21})\right)\left((I_j^{11} + I_j^{22}) - (I_j^{12} + I_j^{21})\right)}} \tag{5.3}$$

where $I_i^{lk} = \{I_{\boldsymbol{\beta_0}}(v_i)^{-1}\}_{lk}$.

Example If we assume that β_1 and β_2 are scalars, $z_i = 1$, and $\alpha = \mathbf{0}$, then we can use Corollary 5.1.3 to derive results for the case of no covariates and a changing intercept. This result is equivalent to Theorem 3.1.2. For this case, approximations to $\sup\{Z^2(v) : a \le v \le b\}$ are available via Corollary 3.1.4.

In section 5.2 we show that model (5.1) satisfies the conditions for local asymptotic normality (van der Vaart, 1998, p. 104). Then for each v we have convergence in distribution (van der Vaart, 1998, p. 232): $2\Lambda_n(v) \xrightarrow{d} 2\Lambda(v)$ with

$$\begin{aligned} 2\Lambda(v) &= \inf_{\{\mathbf{h}\in\mathscr{H}_0\}}(\mathbf{X}-\mathbf{h})^T I_{\boldsymbol{\theta_0}}(\mathbf{X}-\mathbf{h}) - \inf_{\{\mathbf{h}\in\mathscr{H}\}}(\mathbf{X}-\mathbf{h})^T I_{\boldsymbol{\theta_0}}(\mathbf{X}-\mathbf{h}) \\ &= (\mathbf{X}-\hat{\mathbf{h}}_0)^T I_{\boldsymbol{\theta_0}}(\mathbf{X}-\hat{\mathbf{h}}_0) - (\mathbf{X}-\hat{\mathbf{h}})^T I_{\boldsymbol{\theta_0}}(\mathbf{X}-\hat{\mathbf{h}}) \end{aligned} \tag{5.4}$$

where $\mathbf{X} \sim N(\mathbf{0}, I_{\boldsymbol{\theta_0}}^{-1})$, and $\sqrt{n}(\sum_{i=1}^n \dot{l}_v(\boldsymbol{\theta_0}; x_i, \delta_i, \mathbf{z}_i, \mathbf{w}_i)/n) \xrightarrow{d} I_{\boldsymbol{\theta_0}}\mathbf{X}$. Details for solving for $\hat{\mathbf{h}} \in \mathscr{H}$ and $\hat{\mathbf{h}}_0 \in \mathscr{H}_0$ are provided in Section 5.3. We then use empirical measure results to demonstrate the weak convergence $2\Lambda_n(v) \Longrightarrow 2\Lambda(v)$ for all $v \in [a, b]$.

Corollary 5.1.4 *For model (5.1), $\sup_{a\le v\le b} 2\Lambda(v)$ converges in distribution to $\sup_{a\le v\le b} \mathbf{X}_\beta^T \mathbf{Q} \mathbf{X}_\beta(v)$ where $\mathbf{X}_\beta^T \mathbf{Q} \mathbf{X}_\beta(v)$ is in Theorem 5.1.1. The significance level α for a critical value c is then*

$$\alpha = P[\sup_{a\le v\le b} \mathbf{X}_\beta^T \mathbf{Q} \mathbf{X}_\beta(v) > c].$$

These are direct applications of Slutky's Theorem and the Continuous Mapping Theorem to the results from Theorem 5.1.1.

5.2 Proof of Weak Convergence

For the iid case (for example see Section 4.2), we would consider the likelihood as a density $p_{\boldsymbol{\theta}}(x, \delta) = L(\boldsymbol{\theta}; x, \delta)$. However, covariate information leads to nonidentical distributions. To circumvent this, we treat the covariates as random variables and consider the joint distribution (See generalized linear model example in van der Vaart 1998, p. 233):

$$p_{\boldsymbol{\theta}}(x,\delta,\mathbf{z},\mathbf{w}) = f(x,\delta|\mathbf{z},\mathbf{w})g(\mathbf{z},\mathbf{w}) = L(\boldsymbol{\theta};x,\delta,\mathbf{z},\mathbf{w})L'(\boldsymbol{\gamma};\mathbf{z},\mathbf{w})$$

where $L(\boldsymbol{\theta};x,\delta,\mathbf{z},\mathbf{w})$ is the conditional likelihood given $\mathbf{z},\mathbf{w}$. Then $L'(\boldsymbol{\gamma};\mathbf{z},\mathbf{w})$ is the marginal likelihood of the covariates for unrelated nuisance parameter vector $\boldsymbol{\gamma}$.

For our model $\sqrt{p_{\boldsymbol{\theta}}(x,\delta,\mathbf{z},\mathbf{w})}$ is continuously differentiable with respect to $\boldsymbol{\theta}$ for every (x,δ), $\mathbf{z}$, $\mathbf{w}$, and v. The Fisher Information matrix component for $\boldsymbol{\theta}$ is

$$I_{\boldsymbol{\theta}} = E_{\mathbf{z},\mathbf{w}}(E_{x,\delta|\mathbf{z},\mathbf{w}}(\ddot{l}_v(\boldsymbol{\theta};x,\delta,\mathbf{z},\mathbf{w}))).$$

For model (5.1) under staggered-entry, $E_{x,\delta|\mathbf{z},\mathbf{w}}(\ddot{l}_v(\boldsymbol{\theta};x,\delta,\mathbf{z},\mathbf{w}))$ is well-defined and continuous with respect to $\boldsymbol{\theta}$ for any v. Then for a compatible marginal distribution $g(\mathbf{z},\mathbf{w})$, $I_{\boldsymbol{\theta}}$ is also well-defined and continuous with respect to $\boldsymbol{\theta}$ for any v. We will assume $\mathbf{z},\mathbf{w}$ have a such a compatible distribution and will approximate it using the empirical distribution of the observed covariates. By the above properties of $\sqrt{p_{\boldsymbol{\theta}}(x,\delta,\mathbf{z},\mathbf{w})}$ and $I_{\boldsymbol{\theta}}$, our experiment is differentiable in quadratic mean (van der Vaart, 1998, p. 95). Therefore local asymptotic normality holds for each v (van der Vaart, 1998, p. 104).

Definition (van der Vaart, 1998, p. 269): The *empirical process* evaluated at f is defined as $\mathbb{G}_n f = \sqrt{n}(\mathbb{P}_n f - Pf)$. A class $\mathscr{F}$ of measurable functions $f : \mathscr{X} \mapsto \mathbb{R}$ is called *P-Donsker* (or Donsker) if the sequences of processes $\{\mathbb{G}_n f : f \in \mathscr{F}\}$ convergence in distribution to a tight limit process in the space $\ell^\infty(\mathscr{F})$. (Where $\mathbb{P}_n f = n^{-1}\sum_{i=1}^n f(x_i)$, $Pf = \int f dP = E(f)$, and $\ell^\infty(\mathscr{F})$ is the collection of all bounded functions $z : \mathscr{F} \mapsto \mathbb{R}$.)

For (5.1) we show that all elements of the vector-valued function $\dot{l}_v(\boldsymbol{\theta_0};x,\delta,\mathbf{z},\mathbf{w})$ are in one Donsker class indexed by v. We construct an envelope function $M(x,\delta,\mathbf{z},\mathbf{w})$ such that $M(x,\delta,z,w) \geq \|\partial \dot{l}_v(\boldsymbol{\theta_0};x,\delta,\mathbf{z},\mathbf{w})/\partial v\|_1$ for $v \in [a,b]$ and $E(|M(x,\delta,\mathbf{z},\mathbf{w})|) < \infty$. The norm $\|\cdot\|_1$ is the sum of the absolute value of the elements of $\dot{l}_v(\boldsymbol{\theta_0};x,\delta,\mathbf{z},\mathbf{w})$. Note that $M(x,\delta,\mathbf{z},\mathbf{w})$ is not a function of v. The existence of this envelope function satisfies the conditions for a Donsker class (van der Vaart, 1998, p. 271).

Based on the derivatives (See Section 5.6) we choose $M(x,\delta,\mathbf{z},\mathbf{w}) = |\mathbf{z}\exp(\mathbf{z}^T\beta_0 + \mathbf{w}^T\alpha_0)(v^*)^{k-1}|$ where v^* maximizes v^{k-1} for fixed $k > 0$ and $v \in [a,b]$. If $0 < k < 1$, then

$\nu^* = a$. Otherwise $\nu^* = b$. We note that $E(|M(x,\delta,\mathbf{z},\mathbf{w})|) = E_{\mathbf{z},\mathbf{w}}(E_{x,\delta|\mathbf{z},\mathbf{w}}(|M(x,\delta,\mathbf{z},\mathbf{w})|))$ and $E_{x,\delta|\mathbf{z},\mathbf{w}}(|M(x,\delta,\mathbf{z},\mathbf{w})|) = M(x,\delta,\mathbf{z},\mathbf{w})$. If we can assume that the distribution $g(\mathbf{z},\mathbf{w})$ is such that $E_{\mathbf{z},\mathbf{w}}(|\mathbf{z}\exp(\mathbf{z}^T\beta_0 + \mathbf{w}^T\alpha_0)(\nu^*)^{k-1}|) < \infty$, then $E(|M(x,\delta,\mathbf{z},\mathbf{w})|) < \infty$. By our construction: $\sqrt{n}(\sum_{i=1}^n \dot{l}_\nu(\boldsymbol{\theta_0}; x_i,\delta_i,\mathbf{z}_i,\mathbf{w}_i)/n) \Longrightarrow I_{\boldsymbol{\theta_0}}(\nu)X(\nu)$ for $\nu \in [a,b]$, where $X(\nu)$ is a zero mean Gaussian process with $Cov(X(\nu_i), X(\nu_j)) = \Sigma_{ij}$.

Definition (van der Vaart, 1998, p. 269): A class $\mathscr{F}$ of measurable functions $f : \mathscr{X} \mapsto \mathbb{R}$ is called *P-Glivenko-Cantelli* (or Glivenko-Cantelli) if $||\mathbb{P}_n f - Pf||_{\mathscr{F}} = \sup_{f\in\mathscr{F}} |\mathbb{P}_n f - Pf| \xrightarrow{a.s.} 0$. (Where $\mathbb{P}_n f = n^{-1}\sum_{i=1}^n f(x_i)$ and $Pf = \int f dP = E(f)$)

For our model, we would like to show that each element in the matrix of functions $\ddot{l}_\nu(\boldsymbol{\theta_0}; x,\delta,\mathbf{z},\mathbf{w})$ are Glivenko-Cantelli classes indexed by ν. We construct an envelope function $F(x,\delta,\mathbf{z},\mathbf{w})$, such that $F(x,\delta,\mathbf{z},\mathbf{w}) \geq ||\ddot{l}_\nu(\boldsymbol{\theta_0}; x,\delta,\mathbf{z},\mathbf{w})||_1$ and $E(|F(x,\delta,\mathbf{z},\mathbf{w})|) < \infty$ for $\nu \in [a,b]$. This process is very similar to that of constructing $M(x,\delta,\mathbf{z},\mathbf{w})$. The existence of this envelope function satisfies the conditions for a Glivenko-Cantelli class (van der Vaart, 1998, p. 272). Alternatively, we can show membership of $\ddot{l}_\nu(\boldsymbol{\theta_0}; x,\delta,\mathbf{z},\mathbf{w})$ in a Donkser class directly, which implies membership in a Glivenko-Cantelli class. For this we choose $M_2(x,\delta,\mathbf{z},\mathbf{w}) = (||\mathbf{z}\mathbf{z}^T||_1 + ||2\mathbf{z}\mathbf{w}^T||_1 + ||\mathbf{w}\mathbf{w}^T||_1)\exp(\mathbf{z}^T\beta_0 + \mathbf{w}^T\alpha_0)(\nu^*)^{k-1}$ where ν^* maximizes ν^{k-1} for fixed $k > 0$ and $\nu \in [a,b]$. If $0 < k < 1$, then $\nu^* = a$. Otherwise $\nu^* = b$. Then $\sup_{\nu\in[a,b]} ||n^{-1}\sum_{i=1}^n \ddot{l}_\nu(\boldsymbol{\theta_0}; x_i,\delta_i,\mathbf{z}_i,\mathbf{w}_i) - I_{\boldsymbol{\theta_0}}||_1 \xrightarrow{P} 0$.

We next consider the remainder $R_i(\boldsymbol{\theta_0},\mathbf{h}) = R_i^1(\boldsymbol{\theta_0},\mathbf{h}) - \mathbf{h}^T\ddot{l}_\nu(\boldsymbol{\theta_0}; x_i,\delta_i,\mathbf{z}_i,\mathbf{w}_i)\mathbf{h}/2n$ where $R_i^1(\boldsymbol{\theta_0},\mathbf{h}) = l_\nu(\boldsymbol{\theta_0} + \mathbf{h}/\sqrt{n}; x_i,\delta_i,\mathbf{z}_i,\mathbf{w}_i) - l_\nu(\boldsymbol{\theta_0}; x_i,\delta_i,\mathbf{z}_i,\mathbf{w}_i) - \mathbf{h}^T\dot{l}_\nu(\boldsymbol{\theta_0}; x_i,\delta_i,\mathbf{z}_i,\mathbf{w}_i)/\sqrt{n}$. According to Taylor's Theorem: $R_i^1(\boldsymbol{\theta_0},\mathbf{h}) = \mathbf{h}^T\ddot{l}_\nu(\boldsymbol{\theta_0} + \mathbf{h}_\phi/\sqrt{n}; x_i,\delta_i,\mathbf{z}_i,\mathbf{w}_i)\mathbf{h}/2n$, where $\mathbf{h}_\phi^T = (\phi_1 h_1, \ldots, \phi_d h_d)$ for some $\phi_1,\ldots,\phi_d \in [0,1]$ with d the dimension of $\boldsymbol{\theta}$. For fixed $\boldsymbol{\theta_0}$ and $\mathbf{h}$ and for n large enough, we can construct an appropriate envelope function $F(x,\delta,\mathbf{z},\mathbf{w})$ which now covers functions $\ddot{l}_\nu(\boldsymbol{\theta}; x,\delta,\mathbf{z},\mathbf{w})$ indexed by the pair $\{\nu,\boldsymbol{\theta}\} \in [a,b] \times [\boldsymbol{\theta_0}, \boldsymbol{\theta_0} + \mathbf{h}/\sqrt{n}]$. Then $\ddot{l}_\nu(\boldsymbol{\theta_0} + \mathbf{h}_\phi/\sqrt{n}; x,\delta,\mathbf{z},\mathbf{w})$ is Glivenko-Cantelli since $[\boldsymbol{\theta_0}, \boldsymbol{\theta_0} + \mathbf{h}_\phi/\sqrt{n}] \subset [\boldsymbol{\theta_0}, \boldsymbol{\theta_0} + \mathbf{h}/\sqrt{n}]$. We can construct an appropriate envelope function for the difference $\ddot{l}_\nu(\boldsymbol{\theta_0} + \mathbf{h}_\phi/\sqrt{n}; x,\delta,\mathbf{z},\mathbf{w}) - \ddot{l}_\nu(\boldsymbol{\theta_0}; x,\delta,\mathbf{z},\mathbf{w})$ using the triangle inequality, so it is also Glivenko-Cantelli. Thus $\sup_{\nu\in[a,b]} |\sum_{i=1}^n R_i(\boldsymbol{\theta_0},\mathbf{h})| \xrightarrow{P} 0$.

We note here that our envelopes $M(x,\delta,\mathbf{z},\mathbf{w})$ and $M_2(x,\delta,\mathbf{z},\mathbf{w})$ are not functions of either x or δ. This suggests that any well-defined right-censoring regime for model (5.1) will possess the weak convergence properties demonstrated above. Of course the details of C_{ij} will change, but we expect the results from Theorem 5.1.1 to hold for broader classes of censoring. One example is that of simple right censoring at time T_R. It is a straight-forward exercise to show that the results of the previous sections apply to this simple case as well.

5.3 Specifying the Covariance Structure

To characterize the $2\Lambda(\nu)$ process, we need $Cov(\mathbf{X}(\nu_i),\mathbf{X}(\nu_j)) = I^{-1}_{\boldsymbol{\theta_0}}(\nu_i)C_{ij}I^{-1}_{\boldsymbol{\theta_0}}(\nu_j)$. We can evaluate $Cov(\dot{l}_{\nu_i}(\boldsymbol{\theta_0};x,\delta,\mathbf{z},\mathbf{w}),\dot{l}_{\nu_j}(\boldsymbol{\theta_0};x,\delta,\mathbf{z},\mathbf{w})) = C_{ij}$ directly. We note that $C_{ij} = E_{\mathbf{z},\mathbf{w}}(E_{x,\delta|\mathbf{z},\mathbf{w}}(\dot{l}_{\nu_i}(\boldsymbol{\theta_0};x,\delta,z,w)\dot{l}_{\nu_j}(\boldsymbol{\theta_0};x,\delta,\mathbf{z},\mathbf{w})))$. The inner expectation (which we name C_{ijl}) can be evaluated directly through integration using (5.2), the distributions of x,δ conditional on $\mathbf{z}_l,\mathbf{w}_l$ for $l = 1,\ldots,n$. The integrands change depending on the domain of $x \in [0,\nu_i),[\nu_i,\nu_j),[\nu_j,T]$. The outer expectation can be approximated by $\frac{1}{n}\sum_{l=1}^{n} C_{ijl}$, the empirical expectation with respect to observed $\mathbf{z}_l,\mathbf{w}_l$.

The following are the covariance equations for C_{ijl} conditioned on the covariate values z_l, w_l.

$$C_{ijl} = \begin{bmatrix} C_{\beta_1,\beta_1} & C_{\beta_1,\beta_2} & C_{\beta_1,\alpha} \\ C_{\beta_2,\beta_1} & C_{\beta_2,\beta_2} & C_{\beta_2,\alpha} \\ C_{\alpha,\beta_1} & C_{\alpha,\beta_2} & C_{\alpha,\alpha} \end{bmatrix}$$

with

$$
\begin{aligned}
C_{\beta_1,\beta_1} &= \mathbf{z}_l\mathbf{z}_l^T \Big[1+\exp(-v_i^k\eta_l/k)(v_i/T_R-1) \\
&\quad +\Big(\Gamma[1/k,v_i^k\eta_l/k]-\Gamma[1/k]\Big)/\Big((\eta_l/k)^{1/k}kT_R\Big)\Big] \\
C_{\beta_1,\beta_2} &= 0 \\
C_{\beta_1,\alpha} &= \mathbf{z}_l\mathbf{w}_l^T \Big[1+\exp(-v_j^k\eta_l/k)(v_j/T_R-1) \\
&\quad +\Big(\Gamma[1/k,v_j^k\eta_l/k]-\Gamma[1/k]\Big)/\Big((\eta_l/k)^{1/k}kT_R\Big)\Big] \\
C_{\beta_2,\beta_1} &= \mathbf{z}_l\mathbf{z}_l^T \Big[\exp(-v_j^k\eta_l/k)(v_j/T_R-1)-\exp(-v_i^k\eta_l/k)(v_i/T_R-1) \\
&\quad +\Big(\Gamma[1/k,v_j^k\eta_l/k]-\Gamma[1/k,v_j^k\eta_l/k]\Big)/\Big((\eta_l/k)^{1/k}kT_R\Big)\Big] \\
C_{\beta_2,\beta_2} &= \mathbf{z}_l\mathbf{z}_l^T \Big[-\exp(-v_j^k\eta_l/k)(v_j/T_R-1) \\
&\quad +\Big(\Gamma[1/k,T_R^k\eta_l/k]-\Gamma[1/k,v_j^k\eta_l/k]\Big)/\Big((\eta_l/k)^{1/k}kT_R\Big)\Big] \\
C_{\beta_2,\alpha} &= \mathbf{z}_l\mathbf{w}_l^T \Big[-\exp(-v_i^k\eta_l/k)(v_i/T_R-1) \\
&\quad +\Big(\Gamma[1/k,T_R^k\eta_l/k]-\Gamma[1/k,v_i^k\eta_l/k]\Big)/\Big((\eta_l/k)^{1/k}kT_R\Big)\Big] \\
C_{\alpha,\beta_1} &= \mathbf{w}_l\mathbf{z}_l^T \Big[1+\exp(-v_i^k\eta_l/k)(v_i/T_R-1) \\
&\quad +\Big(\Gamma[1/k,v_i^k\eta_l/k]-\Gamma[1/k]\Big)/\Big((\eta_l/k)^{1/k}kT_R\Big)\Big] \\
C_{\alpha,\beta_2} &= \mathbf{w}_l\mathbf{z}_l^T \Big[-\exp(-v_j^k\eta_l/k)(v_j/T_R-1) \\
&\quad +\Big(\Gamma[1/k,T_R^k\eta_l/k]-\Gamma[1/k,v_j^k\eta_l/k]\Big)/\Big((\eta_l/k)^{1/k}kT_R\Big)\Big] \\
C_{\alpha,\alpha} &= \mathbf{w}_l\mathbf{z}_l^T \Big[1-\exp(-T_R^k\eta_l/k) \\
&\quad +\Big(\Gamma[1/k,T_R^k\eta_l/k]-\Gamma[1/k]\Big)/\Big((\eta_l/k)^{1/k}kT_R\Big)\Big]
\end{aligned}
$$

where $\eta_l = \exp(\mathbf{z}_l^T\boldsymbol{\beta}_0+\mathbf{w}_l^T\boldsymbol{\alpha}_0)$, $\Gamma[z]=\int_0^\infty t^{z-1}e^{-t}\,dt$ is the gamma function, and $\Gamma[z,a]=\int_a^\infty t^{z-1}e^{-t}\,dt$ is the incomplete gamma function.

5.3.1 Solving for Q

Now that we have C_{ij} and therefore $I_{\boldsymbol{\theta_0}}(v)$, we solve for $\hat{\mathbf{h}}_0$ and $\hat{\mathbf{h}}$ from (5.4). Under the alternative $\mathbf{h}$ is unrestricted, therefore $\hat{\mathbf{h}}=\mathbf{X}$ (with $\mathbf{X}=(\mathbf{X}_{\beta_1}^T,\mathbf{X}_{\beta_2}^T,\mathbf{X}_\alpha^T)^T$ from Section 5.1 making the second term in (5.4) equal to 0. Under the null hypothesis, $\mathbf{h}_0$ is restricted to be $(\mathbf{h}_{00}^T,\mathbf{h}_{00}^T,\mathbf{h}_\alpha^T)^T$, where the third component is unrestricted. Then $\hat{\mathbf{h}}_0=(\hat{\mathbf{h}}_{00}^T,\hat{\mathbf{h}}_{00}^T,\mathbf{X}_\alpha^T)^T=(\hat{\mathbf{h}}_{00}^T I_{d,2,1}^T,\mathbf{X}_\alpha^T)^T$. From linear algebra: $\hat{\mathbf{h}}_{00}=(I_{d,2,1}^T I_{\boldsymbol{\beta_0}} I_{d,2,1})^{-1} I_{d,2,1}^T I_{\boldsymbol{\beta_0}}\mathbf{X}_\beta$. Then plugging in $\hat{\mathbf{h}}_{00}$ into the left term in (5.4) and factoring out $\mathbf{X}_\beta$, we get $\mathbf{X}_\beta^T\mathbf{Q}\mathbf{X}_\beta$ as in Theorem 5.1.1. It is clear that $I_{d,2,1}^T\mathbf{Q}=\mathbf{0}$ making $\mathbf{Q}$ rank d.

5.3.2 Special Case: α known

If α is known then $\mathbf{X} = \mathbf{X}_\beta$ and $I_{\boldsymbol{\theta_0}} = I_{\boldsymbol{\beta_0}}$. Then $\mathbf{Q}I_{\boldsymbol{\theta_0}}^{-1}$ is idempotent, which implies for each $\nu \in [a,b]$: $\mathbf{X}_\beta^T\mathbf{Q}\mathbf{X}_\beta$ is a χ^2_d random variable with d degrees of freedom (Graybill, 1976). Thus $\mathbf{X}_\beta^T\mathbf{Q}\mathbf{X}_\beta$ is a χ^2_d process over $\nu \in [a,b]$ as in Corollary 5.1.2.

5.3.3 Special Case: β_1 and β_2 are scalars

The case of β_1 and β_2 as scalars is very similar to the results for (4.2) in Section 4.3.2. The matrix $I_{d,2,1}$ simplifies to the one vector $\mathbf{1} = (1,1)^T$. $\mathbf{Q}$ is a 2×2 matrix with $\mathbf{1}^T\mathbf{Q} = \mathbf{0}$. Furthermore $\{Q\}_{11} = |I_{\boldsymbol{\beta_0}}|/\mathbf{1}^T I_{\boldsymbol{\beta_0}}\mathbf{1} = 1/((I^{11}+I^{22})-(I^{12}+I^{21}))$, where $I_i^{lk} = \{I_{\boldsymbol{\beta_0}}(\nu_i)^{-1}\}_{l,k}$ and $|I_{\boldsymbol{\beta_0}}|$ is the determinant of the matrix $I_{\boldsymbol{\beta_0}}$. Then $2\Lambda(\nu) = \{Q\}_{11}(X_{\beta_1} - X_{\beta_2})^2 = Z^2(\nu)$ as in Corollary 5.1.3. If α is known, then $Var(Z(\nu_i), Z(\nu_j)) = 1$ and $Z^2(\nu)$ is a χ^2_1 process.

5.4 Type I Error and Power: CGD data revisited

We revisit the chronic granulomatous disease (CGD) data from Section 3.3. From exploratory analysis, we were able to determine that $k = 2$ was a reasonable assumption. We then tested for a change-point $\nu \in [0.05, 0.95]$ in the control data group. The likelihood ratio test statistic of 37.79 ($\hat{\nu} = 0.072$) exceeded the 5% and 1% critical values obtained from Corollary 3.1.4. In Chapter 3, we chose $a = 0.05$ and $b = 0.95$ arbitrarily based on the plotted data. We also used a plug-in estimate of θ_0 based on observed censoring proportion. In the context of the problem, these two decisions were reasonable, but for this Chapter we take a more general approach.

First we will use the maximum likelihood estimate $\hat{\boldsymbol{\theta}}_0$. Second we will compare the choice $a = 0.05$ and $b = 0.95$ from 90% coverage of the interval of possible values for $y \in [0,1]$, with an alternative of choosing a narrower interval based on the 80% coverage of the control group data with $a = 0.057$ as the 10% quantile and $b = 0.748$ as the 90% quantile. The choice of $[a,b]$ based on quantiles (as is done in Chapter 4) should ensure more stable estimation and faster convergence, since for reasonable quantiles there will be enough data points before and after any potential change-point ν.

Lastly, our new methodology allows us to incorporate covariate information such as age and sex. In the next section, we will compare critical values obtained from Theorem 3.1.2 and Corollary 3.1.4 based only on the control data to critical values from Theorem 5.1.1 and Corollary 5.1.3 including the treatment group data. Finally, we compare these values to critical values obtained using control and treatment data with age and sex covariate information via Corollary 5.1.3.

5.4.1 Critical Values

We obtain critical values via Corollary 3.1.4 by inputting a, b, and $\hat{\theta}_0$ and solving for c with a numerical root finder. These values are displayed as vertical red lines in Figures 5.1 and 5.2. To generate the other critical values we create a grid of 1000 v values spread evenly from $a = 0.05$ to $b = 0.95$. We then generate realizations of the corresponding quadratic processes (Z^2 or $\mathbf{X}^T\mathbf{QX}$) and find the maximum over the grid. We repeat this process 10,000 times and record the 50%, 95%, and 99% quantiles. We repeat this whole procedure 10 times to get a rough distribution of the quantiles for each method. Figures 5.1 shows the results when we allow v to be over the entire grid. Figure 5.2 shows the results when we restrict v to the narrower grid of $a = 0.057$ and $b = 0.748$.

In both Figures 5.1 and 5.2 , the critical values with and without covariate information match very closely. The simulated values also match reasonably well to the those obtained by Corollary 3.1.4. This is not surprising due to the binary nature of z_i. When only considering treatment and control, the likelihood can be separated into terms containing the treatment and those for the control. Then the exact profile likelihood ratio which treats the baseline treatment parameter α as a nuisance is exactly equivalent to removing all treatment data and testing just the control data using Theorem 3.1.2 or Corollary 3.1.4. In addition, the quadratic processes Z^2 and $\mathbf{X}^T\mathbf{QX}$ are identically distributed, so critical values generated from either one should be equivalent (Although using $\mathbf{X}^T\mathbf{QX}$ is much slower). What is interesting is that the covariates age and sex have a minimal impact on the critical values. Their presence forces Z^2 to no longer be a χ^2_1 process. But for this CGD data set, although the $Var(Z)$ is not exactly 1, it is quite close to 1 to around four decimal places.

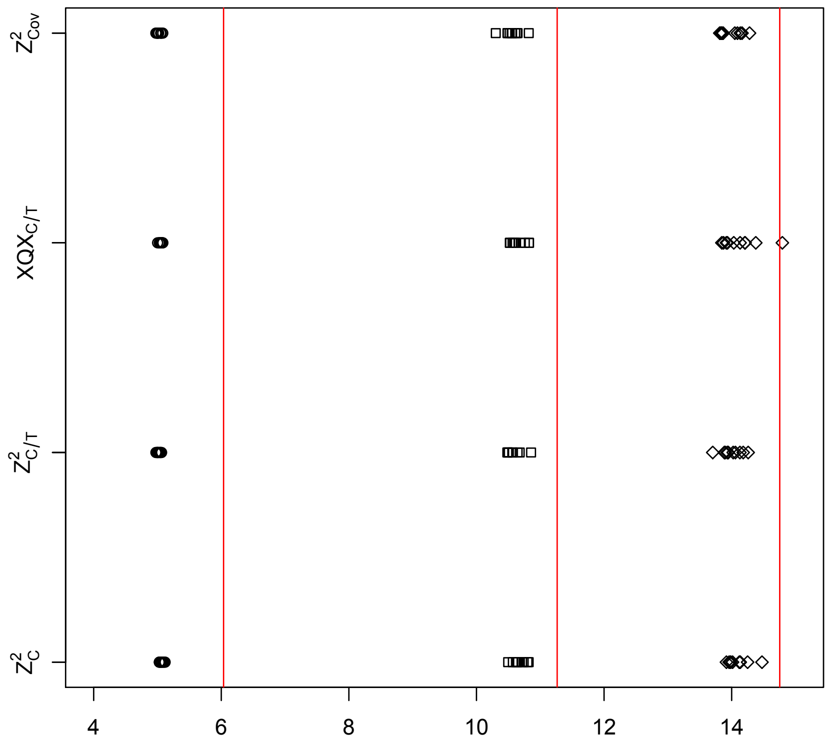

Figure 5.1: Critical values for CGD data with $a = 0.05$ and $b = 0.95$. Quantiles at 50% (circle), 95% (square), and 99% (diamond) for quadratic processes. From top to bottom: Critical values from Corollary 5.1.3 accounting for covariates, critical values from Corollary 5.1.3 using only treatment and control, critical values from Theorem 5.1.1 using only treatment and control, and critical values from Theorem 3.1.2 using only control data. Each point represents quantiles from 10000 realizations of each process. This was replicated 10 times for each setting. The vertical lines represent the approximation from Corollary 3.1.4.

5.4.2 Type I Error and Power

To assess type I error and power, data were simulated based on the maximum likelihood estimates under the null $\hat{\beta}_0 = 1.538$ and $\hat{\boldsymbol{\alpha}}_0 = (-0.150, -0.432, -0.448)^T$ and alternative hypotheses $\hat{\beta}_1 = 4.564$, $\hat{\beta}_2 = 1.044$, and $\hat{\boldsymbol{\alpha}} = (-0.268, -0.388, -0.282)^T$. Covariates $\mathbf{z}$ and $\mathbf{w}$ were resampled from the 108 observations in the CGD data, thus accounting for the treatment, age, and sex. Likelihood ratio test statistics were calculated and compared to the critical values obtained from Corollary 5.1.3. This was repeated 1000 times for sample sizes of 100, 200, and 500. Type I error was calculated as the percentage of times that the test statistic exceeded the 5% and 1% critical values when the data was generated under null hypothesis. Power was calculated as the percentage of times that the test statistic exceeded the 5% and 1% critical values when the data was generated under alternative hypothesis.

These rates obtained by using Corollary 5.1.3 are then compared to rates from by Davies' approximation (1987), which allows us to approximate p-values for the supremum of χ^2 processes. While the result of Davies only strictly applies to χ^2 processes, our results for Section 5.4.1 suggest that Z^2 can be reasonably approximated by a χ_1^2 process, at least for the CGD data set.

The results for the Type I error simulations are recorded in Table 5.1. For the wider interval, convergence to nominal error rates seems to occur for a sample size of around 500. Rates for our method (LAN) are more or less comparable to those for Davies. For the narrower interval based on the quantiles of the control group, we see convergence at a sample size of 100. Furthermore our method is within a standard error of the nominal levels at each sample size. In contrast the method of Davies appears to converge to lower than nominal values by 500 samples. Power simulations based on $\hat{\theta}$ are displayed in Table 5.2. Power is high and appear to be slightly higher for the Davies method at smaller samples. Otherwise power levels are essentially the same. The close to nominal type I error combined with high power demonstrate that our test using both the LAN and Davies methods is appropriate and useful for this CGD data set, particularly when we choose the narrower, quantile-based values for a and b.

5.5 Conclusions

In this chapter, we have applied the techniques introduced in Chapter 4 to generalize the results of Chapter 3. In the process, we have shown that the LAN expansion and weak convergence results still hold in the presence of covariates and censoring. Our

Table 5.1: Type I error simulations for CGD data using Corollary 5.1.3 (LAN) and Davies (1987)

$[a,b]=[0.05,0.95]$					$[a,b]=[0.057,0.748]$			
Sample Size	100	200	500		Sample Size	100	200	500
s.e. 5%	0.022	0.015	0.010		s.e. 5%	0.022	0.015	0.010
s.e. 1%	0.010	0.007	0.004		s.e. 1%	0.010	0.007	0.004
LAN 5%	0.096	0.060	0.055		LAN 5%	0.047	0.035	0.043
LAN 1%	0.067	0.036	0.013		LAN 1%	0.012	0.007	0.006
Davies 5%	0.111	0.058	0.039		Davies 5%	0.053	0.038	0.020
Davies 1%	0.078	0.036	0.011		Davies 1%	0.019	0.010	0.003

Table 5.2: Power simulations for CGD data using Corollary 5.1.3 (LAN) and Davies (1987)

$[a,b]=[0.05,0.95]$					$[a,b]=[0.057,0.748]$			
Sample Size	100	200	500		Sample Size	100	200	500
LAN 5%	0.934	0.998	1.000		LAN 5%	0.950	0.998	1.000
LAN 1%	0.892	0.998	1.000		LAN 1%	0.895	0.998	1.000
Davies 5%	0.961	0.980	1.000		Davies 5%	0.961	0.998	1.000
Davies 1%	0.916	0.980	1.000		Davies 1%	0.911	0.998	1.000

simulation of critical values displays the many options we have including asymptotic approximations (Corollary 3.1.4), simulations based on univariate Gaussian processes (Corollary 5.1.3), and those based on multivariate Gaussian processes (Theorem 5.1.1). When all options are available, results are essentially be the same. However, this flexibility allows us to perform a Likelihood Ratio Test for a change-point even when no simplification for Theorem 5.1.1 can be found. Furthermore, our use of Davies' approximation (1987) suggests that this relatively easy to implement method is effective when χ^2 assumptions are met, and may still work when assumptions are slightly violated.

5.6 Supplemental: Derivatives

We include the first and second derivatives evaluated at $\boldsymbol{\theta}=\boldsymbol{\theta}_0$. From these, we can construct envelope functions $M(x,\delta,\mathbf{z},\mathbf{w})$ and $F(x,\delta,\mathbf{z},\mathbf{w})$ and calculate C_{ijl} via integration. Let $\eta=\exp(\mathbf{z}^T\boldsymbol{\beta}_0+\mathbf{w}^T\boldsymbol{\alpha}_0)$.

$$\dot{l}_v(\boldsymbol{\theta_0}; x, \delta, \mathbf{z}, \mathbf{w})|_{\boldsymbol{\beta}_1=\boldsymbol{\beta}_2=\boldsymbol{\beta}_0, \boldsymbol{\alpha}=\boldsymbol{\alpha}_0}$$

$$\begin{aligned}
\partial l_v(\boldsymbol{\theta_0}; x, \delta, \mathbf{z}, \mathbf{w})/\partial \boldsymbol{\beta}_1 &= \mathbf{z}\left(\delta - \eta x^k/k\right)\mathbf{1}_{\{x<v\}} + \mathbf{z}\left(-\eta v^k/k\right)\mathbf{1}_{\{x\geq v\}} \\
\partial l_v(\boldsymbol{\theta_0}; x, \delta, \mathbf{z}, \mathbf{w})/\partial \boldsymbol{\beta}_2 &= \mathbf{z}\left(\delta - \eta(x^k - v^k)/k\right)\mathbf{1}_{\{x\geq v\}} \\
\partial l_v(\boldsymbol{\theta_0}; x, \delta, \mathbf{z}, \mathbf{w})/\partial \boldsymbol{\alpha} &= \mathbf{w}\left(\delta - \eta x^k/k\right)
\end{aligned}$$

$$\partial \dot{l}_v(\boldsymbol{\theta_0}; x, \delta, \mathbf{z}, \mathbf{w})/\partial v|_{\boldsymbol{\beta}_1=\boldsymbol{\beta}_2=\boldsymbol{\beta}_0, \boldsymbol{\alpha}=\boldsymbol{\alpha}_0}$$

$$\begin{aligned}
\partial^2 l_v(\boldsymbol{\theta_0}; x, \delta, \mathbf{z}, \mathbf{w})/\partial v \partial \boldsymbol{\beta}_1 &= -\mathbf{z}\eta v^{k-1}\mathbf{1}_{\{x\geq v\}} \\
\partial^2 l_v(\boldsymbol{\theta_0}; x, \delta, \mathbf{z}, \mathbf{w})/\partial v \partial \boldsymbol{\beta}_2 &= \mathbf{z}\eta v^{k-1}\mathbf{1}_{\{x\geq v\}} \\
\partial^2 l_v(\boldsymbol{\theta_0}; x, \delta, \mathbf{z}, \mathbf{w})/\partial v \partial \boldsymbol{\alpha} &= 0
\end{aligned}$$

$$\ddot{l}_v(\boldsymbol{\theta_0}; x, \delta, \mathbf{z}, \mathbf{w})|_{\boldsymbol{\beta}_1=\boldsymbol{\beta}_2=\boldsymbol{\beta}_0, \boldsymbol{\alpha}=\boldsymbol{\alpha}_0}$$

$$\begin{aligned}
\partial^2 l_v(\boldsymbol{\theta_0}; x, \delta, \mathbf{z}, \mathbf{w})/\partial \boldsymbol{\beta}_1^T \partial \boldsymbol{\beta}_1 &= \mathbf{z}\mathbf{z}^T\left(-\eta x^k/k\right)\mathbf{1}_{\{x<v\}} + \mathbf{z}\mathbf{z}^T\left(-\eta v^k/k\right)\mathbf{1}_{\{x\geq v\}} \\
\partial^2 l_v(\boldsymbol{\theta_0}; x, \delta, \mathbf{z}, \mathbf{w})/\partial \boldsymbol{\beta}_2^T \partial \boldsymbol{\beta}_1 &= 0 \\
\partial^2 l_v(\boldsymbol{\theta_0}; x, \delta, \mathbf{z}, \mathbf{w})/\partial \boldsymbol{\alpha}^T \partial \boldsymbol{\beta}_1 &= \mathbf{z}\mathbf{w}^T\left(-\eta x^k/k\right)\mathbf{1}_{\{x<v\}} + \mathbf{z}\mathbf{w}^T\left(-\eta v^k/k\right)\mathbf{1}_{\{x\geq v\}} \\
\partial^2 l_v(\boldsymbol{\theta_0}; x, \delta, \mathbf{z}, \mathbf{w})/\partial \boldsymbol{\beta}_2^T \partial \boldsymbol{\beta}_2 &= \mathbf{z}\mathbf{z}^T\left(-\eta(x^k - v^k)/k\right)\mathbf{1}_{\{x\geq v\}} \\
\partial^2 l_v(\boldsymbol{\theta_0}; x, \delta, \mathbf{z}, \mathbf{w})/\partial \boldsymbol{\alpha}^T \partial \boldsymbol{\beta}_2 &= \mathbf{w}\mathbf{z}^T\left(-\eta(x^k - v^k)/k\right)\mathbf{1}_{\{x\geq v\}} \\
\partial^2 l_v(\boldsymbol{\theta_0}; x, \delta, \mathbf{z}, \mathbf{w})/\partial \boldsymbol{\alpha}^T \partial \boldsymbol{\alpha} &= -\mathbf{w}\mathbf{w}^T \eta x^k/k
\end{aligned}$$

$$\partial \ddot{l}_v(\boldsymbol{\theta_0}; x, \delta, \mathbf{z}, \mathbf{w})/\partial v|_{\boldsymbol{\beta}_1=\boldsymbol{\beta}_2=\boldsymbol{\beta}_0, \boldsymbol{\alpha}=\boldsymbol{\alpha}_0}$$

$$\begin{aligned}
\partial^2 l_v(\boldsymbol{\theta_0}; x, \delta, \mathbf{z}, \mathbf{w})/\partial v \partial \boldsymbol{\beta}_1^T \partial \boldsymbol{\beta}_1 &= \mathbf{z}\mathbf{z}^T\left(-\eta v^{k-1}\right)\mathbf{1}_{\{x\geq v\}} \\
\partial^2 l_v(\boldsymbol{\theta_0}; x, \delta, \mathbf{z}, \mathbf{w})/\partial v \partial \boldsymbol{\beta}_2^T \partial \boldsymbol{\beta}_1 &= 0 \\
\partial^2 l_v(\boldsymbol{\theta_0}; x, \delta, \mathbf{z}, \mathbf{w})/\partial v \partial \boldsymbol{\alpha}^T \partial \boldsymbol{\beta}_1 &= \mathbf{z}\mathbf{w}^T\left(-\eta v^{k-1}\right)\mathbf{1}_{\{x\geq v\}} \\
\partial^2 l_v(\boldsymbol{\theta_0}; x, \delta, \mathbf{z}, \mathbf{w})/\partial v \boldsymbol{\beta}_2^T \partial \boldsymbol{\beta}_2 &= \mathbf{z}\mathbf{z}^T\left(\eta v^{k-1}\right)\mathbf{1}_{\{x\geq v\}} \\
\partial^2 l_v(\boldsymbol{\theta_0}; x, \delta, \mathbf{z}, \mathbf{w})/\partial v \boldsymbol{\alpha}^T \partial \boldsymbol{\beta}_2 &= \mathbf{w}\mathbf{z}^T\left(\eta v^{k-1}\right)\mathbf{1}_{\{x\geq v\}} \\
\partial^2 l_v(\boldsymbol{\theta_0}; x, \delta, \mathbf{z}, \mathbf{w})/\partial v \boldsymbol{\alpha}^T \partial \boldsymbol{\alpha} &= 0
\end{aligned}$$

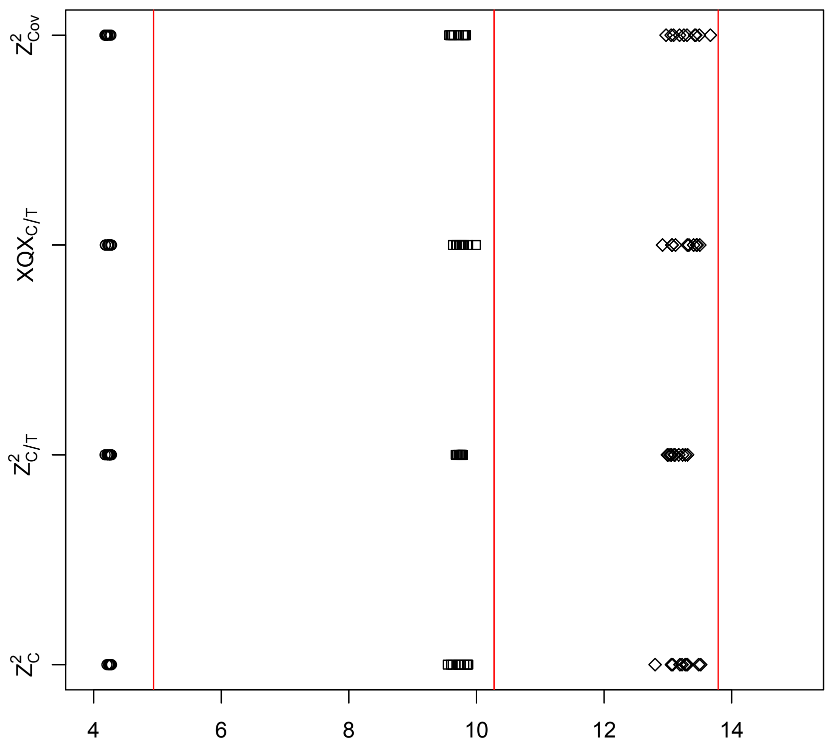

Figure 5.2: Critical values for CGD data with $a = 0.057$ and b $= 0.748$, the 10% and 90% observed quantiles for the control data. Quantiles at 50% (circle), 95% (square), and 99% (diamond) for limiting quadratic processes. From top to bottom: Critical values from Corollary 5.1.3 accounting for covariates, critical values from Corollary 5.1.3 using only treatment and control, critical values from Theorem 5.1.1 using only treatment and control, and critical values from Theorem 3.1.2 using only control data. Each point represents quantiles from 10000 realizations of each process. This was replicated 10 times for each setting. The vertical lines represent the approximation from Corollary 3.1.4.

CHAPTER 6

CONCLUDING REMARKS

In the preceding chapters, we have demonstrated how to devise and implement asymptotic tests for the presence of a nuisance change-point for hazard models. These test are irregular due to the presence of the change-point only under the alternative. One way to deal with this is to treat the test statistic as a process with respect to the change-point, then base analysis on the supremum of this process over a range of potential change-points. We have outlined a methodology for using the likelihood ratio test statistic. However, similar use of the partial likelihood seems appropriate. While current methods in the literature have restricted the number of covariate coefficients that can change to one (Chapter 2), our method applies generally to any number. We anticipate future work to confirm these assertions.

If we consider the broader motivations of model selection, we quickly see more opportunities. In Chapter 5, we developed tests for change-points in the presence of covariates, but model selection in classic regression involves comparing full and reduced models in terms of the covariate information. Likewise we can imagine the need to test for the inclusion of a covariate such as age given not only other covariates such as gender, but also the presence of a change-point. Although the change-point is now a parameter under both the null and alternative hypothesis, our work here leads us to doubt any appeal to a simple χ^2 asymptotic distribution.

Model selection is a broad arena going far beyond hypothesis testing. Other methods are available, particularly in the Bayesian framework. These techniques include comparing posterior model probabilities, stochastically choosing models through reversible-jump Metropolis-Hastings (Green, 1995) and Stochastic Search Variable Selection (George and McCulloch, 1993, 1997), and ratio sampling for Bayes Factors and posterior ratios (Meng and Wong, 1996; Gelman and Meng, 1998). Most of these methods have been applied to exponential family distributions, so care and ingenuity may be needed when applying them to hazard models. Additionally, the irregularities induced by our change-point should make implementation of these methods more challenging. Perhaps our LAN expansion methods can be used here as well.

Taking a step back from hazard models, we may consider applying some of the methods reviewed here to other problems involving nuisance parameters. For example Ritz and Skovgaard (2005) develop a score test for the presence of a nuisance parameter for curved exponential families. Their motivating example is the test of a random intercept model for normal data versus an alternative that has serial correlation. The overall method and asymptotic results fit in well with the literature in Chapter 2 and our own work.

APPENDIX **A**

THE SUP DISTRIBUTION OF AN ORNSTEIN-UHLENBECK PROCESS

As an alternative to the approximation from Corollary 3.1.4 (from Leadbetter et al. (1983)), we can instead use the results from Durbin (1985) which provide an approximation for either $P[\sup_{a \le v \le b} |Z(v)| > c]$ or $P[\sup_{a^* \le t \le b^*} |W(t)| > c]$. The question then becomes, which approximation is better and should we use the original parameterization or the stationary one? This exploration leads us to some interesting results.

Our three main findings are the following: (1) for an Ornstein-Uhlenbeck process with unit variance, the approximations from Leadbetter et al. (1983) and Durbin (1985) are identical. We demonstrate with an example. (2) The approximation from Durbin is invariant to our reparameterization. In other words, using $Z(v)$ or $W(t)$ gives the same results. We can show that this holds for any strictly monotonic reparameterization of an Ornstein-Uhlenbeck process with unit variance. (3) This suggests that we might find a reparameterization of a non-stationary $Z(v)$ such that it becomes an Ornstein-Uhlenbeck process $W(t)$. Conditions from Durbin (1985) are not quite strong enough, so we add one more.

A.1 The Leadbetter et al. (1983) Approximation

Assume we have a Guassian Process $Z(t)$ with stationary covariance $\rho(s,t)$ for $s \le t$. Define $r(\tau) = \rho(0,\tau)$. Define the following condition for $r(\tau)$:

$$r(\tau) = 1 - C|\tau|^{\alpha} + o(|\tau|^{\alpha}) \text{ as } \tau \to \infty \text{ and } 0 < \alpha \le 2. \tag{A.1}$$

Theorem A.1.1 (Tail Approximation Theorem 12.2.9) *If $r(\tau)$ satisfies (A.1) then for each fixed $h > 0$ such that $\sup_{\epsilon \le t \le h} r(t) = \delta_\epsilon < 1$ for all $\epsilon > 0$,*

$$\lim_{u\to\infty} \frac{1}{u^{2/\alpha}\phi(u)/u} P\{M(h) > u\} = hC^{1/\alpha}H_\alpha$$

where ϕ is the standard normal density, $M(h) = \sup_{0\le t\le h}\{Z(t)\}$, and H_α is a constant depending on α.

It turns out that $H_1 = 1$ and $H_2 = \frac{1}{\sqrt{2\pi}}$, but other values of α lead to a complicated formula.

We now consider the the zero-mean Gaussian process $W(t)$ with $\rho(s,t) = \exp[-\theta(t-s)]$ for finite $\theta > 0$. Then $W(t)$ is an Ornstein-Uhlenbeck process, which is stationary. From Remark 12.2.10 in Leadbetter et al. (1983) we see that $r(\tau) = 1 - C|\tau| + o(|\tau|)$ with $\alpha = 1$ and $C = \theta$. Then

$$\begin{aligned} P\{\sup_{a\le t\le b} W(t) > u\} &\approx (bC^{1/\alpha}H_\alpha - aC^{1/\alpha}H_\alpha)u^{2/\alpha}\phi(u)/u \\ &= (b-a)C^{1/\alpha}H_\alpha u^{2/\alpha}\phi(u)/u \\ &= (b-a)\tfrac{\theta u}{\sqrt{2\pi}}e^{-u^2/2}. \end{aligned}$$

Similarly, $P\{\sup_{a\le t\le b} |W(t)| > u\} \approx (b-a)\dfrac{2\theta u}{\sqrt{2\pi}}e^{-u^2/2}$. Setting $\theta = \frac{1}{2}$ we obtain the approximation in Corollary 3.1.4.

A.2 The Durbin (1985) Approximation

We now outline the conditions and results of Durbin's approximation: $y(\tau)$ is a continuous Gaussian process with covariance function $\rho(\sigma,\tau)$ for $0 \le \sigma \le \tau$. We assume $E[y(\tau)] = 0$.

Theorem A.2.1 (Durbin's Theorem) *The first passage density of $y(t)$ to the (upper) boundary $a(t)$ is $p(t) = b(t)f(t)$. where*

$$b(t) = \lim_{s\uparrow t}(t-s)^{-1}E[I(s,y)(a(s)-y(s))|y(t) = a(t)]$$

and

$$f(t) = [2\pi\rho(t,t)]^{-1/2}e^{-\frac{1}{2}a^2(t)/\rho(t,t)}$$

$I(s,y)$ is the indicator that the sample path does not cross the boundary prior to s.

Durbin proves this theorem holds for conditions A1 - A3:

A1. The boundary function $a(\tau)$ is continuous in $0 \le \tau < t$ and is left differentiable at t.

A2. The covariance function $\rho(\sigma,\tau)$ is positive definite and has continuous first-order partial derivatives on the set $\{(\sigma,\tau) : 0 \le \sigma \le \tau \le t\}$ where appropriate left or right derivatives are taken at $\sigma = \tau$, $\sigma = 0$, and $\tau = t$.

A3. The variance of the increment $y(t) - y(s)$ satisfies the condition

$$\lim_{s \uparrow t} (t-s)^{-1} V[y(t) - y(s)] = \lambda_t$$

where $0 < \lambda_t < \infty$.

Durbin notes that A3 is equivalent to (see correction Durbin 1988 for typographical errors):

$$\lim_{s \uparrow t} \left[\frac{\partial \rho(s,t)}{\partial s} - \frac{\partial \rho(s,t)}{\partial t} \right] = \lambda_t$$

Durbin provides an approximation to $p(t)$: $p_1(t) = b_1(t) f(t)$ where

$$\begin{aligned} b_1(t) &= \lim_{s \uparrow t} (t-s)^{-1} E[a(s) - y(s) | y(t) = a(t)] \\ &= \tfrac{a(t)}{\rho(t,t)} \tfrac{\partial \rho(\sigma,t)}{\partial t} |_{\sigma = t} - a'(t) \end{aligned}$$

The approximation $p_1(t)$ is our focus.

A.2.1 The Ornstein-Uhlenbeck process

We consider an Ornstein-Uhlenbeck process $W(t)$ with $\rho(s,t) = \exp[-\theta(t-s)]$ for finite $\theta > 0$. Clearly, $\rho(t,t) = 1$. We choose $a(t) = c$, a constant. Then conditions A1 and A2 are clearly satisfied. Let us consider condition A3:

$$\begin{aligned} \lim_{s \uparrow t} \left[\frac{\partial \rho(s,t)}{\partial s} - \frac{\partial \rho(s,t)}{\partial t} \right] &= \lim_{s \uparrow t} \left[\theta e^{-\theta(t-s)} - \left(-\theta e^{-\theta(t-s)} \right) \right] \\ &= \lim_{s \uparrow t} \left[2\theta e^{-\theta(t-s)} \right] = 2\theta > 0 \end{aligned}$$

Thus conditions A1-A3 are satisfied.

Constructing the approximation $p_1(t)$ we get:

$$f(t) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}c^2}$$

and

$$b_1(t) = \frac{c}{1}\theta e^{-\theta(t-s)}|_{s=t} - 0 = c\theta.$$

Then

$$\begin{aligned} P[\sup_{a^* \le t \le b^*} |W(t)| > c] &= 2P[\sup_{a^* \le t \le b^*} W(t) > c] \\ &\approx 2\int_{a*}^{b*} b_1(t) f(t) dt \\ &= 2\tfrac{c\theta}{\sqrt{2\pi}} e^{-\frac{1}{2}c^2}(b^* - a^*) \\ &= \tfrac{\sqrt{2}c\theta}{\sqrt{\pi}} e^{-\frac{1}{2}c^2}(b^* - a^*) \end{aligned}$$

which is the same as Corollary 3.1.4.

A.2.2 Invariance to monotonic reparameterization

Suppose a reparameterization $g(t) = \mu$ exits such that $g(t)$ is strictly monotonic and continuously differentiable. Then $Z(\mu) = W(g^{-1}(\mu))$ and $\rho_Z(\nu,\mu) = \rho_W(g^{-1}(\nu), g^{-1}(\mu)) = \exp\left[-\theta|g^{-1}(\mu) - g^{-1}(\nu)|\right]$. Then conditions A1-A3 hold and $\lambda_t = 2\theta\left|\frac{dg^{-1}(\mu)}{d\mu}\right|$. Furthermore $f(\mu)$ remains the same, and $b_1(\mu) = c\theta\left|\frac{dg^{-1}(\mu)}{d\mu}\right|$. Then

$$\begin{aligned} P[\sup_{a \le t \le b} W(t) > c] &= P[\sup_{a^* \le \mu \le b^*} Z(\mu) > c] \\ &\approx \left|\int_{g(a)}^{g(b)} b_1(\mu) f(\mu) d\mu\right| \\ &= \tfrac{c\theta}{\sqrt{2\pi}} e^{-\frac{1}{2}c^2}\left|\int_{g(a)}^{g(b)} \tfrac{dg^{-1}(\mu)}{d\mu} d\mu\right| \\ &= \tfrac{c\theta}{\sqrt{2\pi}} e^{-\frac{1}{2}c^2}\left|g^{-1}(g(b)) - g^{-1}(g(a))\right| \\ &= \tfrac{c\theta}{\sqrt{2\pi}} e^{-\frac{1}{2}c^2}(b-a) \end{aligned}$$

Thus our approximation is invariant to the specific choice of $g(t)$ if we have an Ornstein-Uhlenbeck process. Similarly if we start with a possibly non-stationary Gaussian process $Z(\mu)$ with $\rho(\nu,\nu) = 1$ and can find a monotonic reparameterization $g^{-1}(\mu)$to an Ornstein-Uhlenbeck process (as in Chapter 3), then Durbin's approximation will be exactly the same for either parameterization.

A.2.3 Finding a reparameterization

Suppose that we have a zero-mean, unit-variance Gaussian process $Z(\mu)$ with known covariance structure $\rho(\nu,\mu)$. If we want to re-express our process $Z(\mu)$ an as Ornstein-

Uhlenbeck process $W(t)$ then, the discussion above suggest that a continuously differentiable and strictly monotonic transformation $g^{-1}(\mu)$ may be obtained by the following:

$$g^{-1}(\mu) = \frac{1}{\theta}\int_{\{m\leq\mu\}} \frac{\partial\rho(v,m)}{\partial v}|_{v=m}\,dm$$

for $v \leq m$.

In order for this $g^{-1}(\mu)$ to be continuously differentiable and strictly monotonic, $\frac{\partial\rho(v,m)}{\partial v}|_{v=m}$ needs to be continuous and strictly positive or strictly negative. Then $g^{-1}(\mu)$ has the properties above and $Z(g(t))$ is an Ornstein-Uhlenbeck process. We note that the existence of such a $g^{-1}(\mu)$ implies that conditions A2 and A3 are satisfied with $\lambda_t = \left|\frac{\partial\rho(v,\mu)}{\partial v}|_{v=\mu}\right|$.

A.2.4 Example from the Exponential Hazard Function

Taking the general setting from Chapter 3 and restricting $k=1$, $T=1$, and $\theta=1$, we get the following:

$$\rho(v,\mu) = \sqrt{\frac{v(e^{\mu}-e\mu)}{\mu(e^{v}-ev)}}$$

for $v \leq \mu$.

$$\frac{\partial\rho(v,\mu)}{\partial v}|_{v=\mu} = \frac{e^{\mu}(\mu-1)}{2\mu(e\mu-e^{\mu})}$$

which is well-defined and strictly negative for $0 < a \leq \mu \leq b < 1$ (Bounded away from 0 and 1).

Then applying our results from Section A.2.3:

$$\begin{aligned} g^{-1}(\mu) &= \tfrac{1}{\theta}\int_0^{\mu} \tfrac{e^{m}(m-1)}{2m(em-e^{m})}\,dm \\ &= \tfrac{1}{2\theta}\log\left(\tfrac{\mu}{e\mu-e^{\mu}}\right) \\ &= \tfrac{1}{2\theta}\log\left(\tfrac{g_2(\mu)}{g_2(0)-g_2(\mu)}\right) \\ &= t \end{aligned}$$

Then

$$\mu = g_2^{-1}\left(\frac{g_2(0)e^{2\theta t}}{1+e^{2\theta t}}\right)$$

Choosing $\theta = \frac{1}{2}$ we get the transformation from Corollary 3.1.3.

APPENDIX **B**

CGD DATA: MODEL FITTING AND PARAMETER INFERENCE

While the focus of this work has been on the testing of models for model selection, model fitting and inference is obviously important. Although we have not derived any new results, we demonstrate some common techniques for model fitting and parameter-based inference using the CGD data set from Chapters 3 and 5. In particular we compare a Bayesian Markov Chain Monte Carlo approach to a bootstrap-Maximum Likelihood method. Both methods are fairly well established. Our implementation here is effective, but may be overly simplistic.

In the CGD data set, we are interested in the placebo controlled trial of gamma interferon in chronic granulomatous disease. The response is time to first infection (days). The covariates we consider are treatment/control, gender, and age. About 70-80% of the data is censored. By looking at the quantile-quantile plots of the treatment and control data, the times to infection seem to have a Rayleigh distribution (Weibull distribution with shape parameter equal to 2). We wish to see if the treatment is effective in extending time to infection and to examine the impacts of the covariates. This data is also interesting because it has patients entering at different days τ_i, and right censoring occurring at the end of the study $T = 321$ days. We refer to this as staggered-entry and right-censoring.

Rayleigh(λ) data has a linear hazard function, $h(t) = \theta t = \frac{2}{\lambda^2} t$. Using the θ parameterization with the control data in Chapter 3 we tested the alternative hypothesis: $h(t) = \theta_1 1_{\{t<\nu\}} t + \theta_2 1_{\{t \geq \nu\}} t$ for some unknown time ν; versus the null hypothesis:

$h(t) = \theta_0 t$. Similarly we retested in Chapter 5 using covariate information and a modified parameterization. Both tests yielded a large test statistic for the data; larger than the corresponding 1% critical value. This suggests that we should fit the alternative model and incorporate the uncertainty of the change-point into our model inferences.

B.1 Methods

For this analysis we will use a parameterization more in line with generalized linear models, focusing on the expected value. This is slightly different from the notation used in Chapter 5 but is still equivalent. For Rayleigh$(t_i;\lambda_i)$ distributed data, $E[t_i] = \frac{\sqrt{\pi}}{2}\lambda_i \propto \lambda_i$. Our covariates $\mathbf{z}$ are treatment(presence/absence), age, and gender. Since $\lambda_i > 0$, we'll model using a log link function: $\log(\lambda_i) = \mathbf{z}_i^T \boldsymbol{\beta}$. We first fit through Bayesian methods using Metropolis Sampling. We then bootstrap by resampling the data and finding the maximum likelihood estimates for each sample. Both methods will provide us with point and interval estimates for our parameters.

B.1.1 The Model

In our model we assume the alternative hypothesis: a change occurs for control patients at an unknown time $\nu \in [a, b]$, where $T > b$ is the end of the study. Then the hazard function is

$$h(t_i) = \begin{cases} 2(\lambda_{1,i})^{-2} t_i, & t_i < \nu \\ 2(\lambda_{2,i})^{-2} t_i, & t_i \geq \nu \end{cases}$$

where $\lambda_{1,i}$ and $\lambda_{2,i}$ are related to the covariate information by

$$\begin{aligned} \log(\lambda_{1,i}) &= \beta_1 + \beta_2 1_{\{Trt=1\}} + \beta_3 Age + \beta_4 Gender \\ \log(\lambda_{2,i}) &= \beta_1 + \beta_2 1_{\{Trt=1\}} + \beta_3 Age + \beta_4 Gender + \beta_{CP} 1_{\{t_i \geq \nu\}} 1_{\{Trt=0\}} \end{aligned} \tag{B.1}$$

We are still observe staggered entry at time τ_i and right censoring at time T, so our data are (x_i, δ_i), where $x_i = \min(t_i, T - \tau_i)$ and $\delta_i = 1_{\{t_i \leq T - \tau_i\}}$. Then the likelihood becomes

$$
\begin{aligned}
L(y|\lambda_1,\lambda_2) = & \prod_{i=1}^n \left(\frac{2y_i}{\lambda_{2,i}^2}\right)^{c_i} e^{-\left(\frac{\nu^2}{\lambda_{1,i}^2}+\frac{y_i^2-\nu^2}{\lambda_{2,i}^2}\right)} \\
= & \prod_{i=1}^n \left(2y_i e^{-2\log(\lambda_{2,i})}\right)^{c_i} \exp\left(-\nu^2 e^{-2\log(\lambda_{1,i})} - (y_i^2-\nu^2)e^{-2\log(\lambda_{2,i})}\right)
\end{aligned}
$$

B.1.2 Bayesian Fitting

In Bayesian analysis, we are interested in the posterior distribution of the parameters $\pi(\tilde{\beta},\phi,\nu|x,\delta,z)$. While closed form solutions are not available, we can use Gibbs Sampling, which samples from the full conditional distributions $\pi(\tilde{\beta}_j,|\tilde{\beta}_{-j},\phi,\nu,x,\delta,z)$, $\pi(\phi,|\tilde{\beta},\nu x,\delta,z)$, and $\pi(\nu,|\tilde{\beta},\phi x,\delta,z)$. This procedure eventually converges and samples from the full joint posterior distribution. If we are unable to sample one of the full conditional distributions directly, we can the use a Metropolis sampler which is a type of rejection sampler.

We select the following proper priors.

$$
\begin{aligned}
\tilde{\beta} &\sim N(0,\phi^{-1}I_4) \\
\phi &\sim Gam(\gamma_1,\gamma_2) \\
\nu &\sim U(a,b)
\end{aligned}
$$

To evaluate our model we use a Gibbs step to sample $\pi(\phi|-)$ and a Metropolis step to sample $\pi(\tilde{\beta}|-)$ and $\pi(\nu|-)$. We choose $\gamma_1=\gamma_2=0.1$ for ϕ and a proposal density $N(\beta_j^{i-1},0.1)$ for each β_j. We choose a proposal density $N(\nu^{i-1},\sqrt{5})$ for ν and choose $[a,b]=[0.057,0.748]\times T$, which is consistent with our findings in Chapter 5.

B.1.3 Bootstrap Fitting

In our bootstrap approach, we resample with replacement the original data with the corresponding covariate structure intact. For each sample, we maximize the likelihood and store the parameters. Since the likelihood is irregular (and thus unstable) due to the presence of the change-point, we maximize over a grid of potential ν and choose the parameters that correspond to the largest likelihood. This is very much the same technique that has been used in Chapters 3, 4, and 5. The only real difference is that the samples are not randomly generated from a null distribution, but are instead resampled from the original data. To save on computation time, we use the observed data points (which are integers) and a small neighborhood around them (± 0.5) as potential ν values instead of conducting a larger grid search. This works well in practice and is consistent with the claim by Worsley (1988) that values of ν which maximize the likelihood are close to the uncensored observations.

B.2 Results & Interpretation

The Bayesian sampler was run for 10,000 iterations. Trace plots of the parameters showed convergence after a few hundred iterations, so a burn-in of 1000 was used (figures not included). The bootstrap method was run for 1000 iterations, because it uses a computationally intensive non-linear optimizer to maximize the likelihood. Table B.1 compares the estimates from both methods. The bootstrap intervals are very wide, making it difficult to conclude much. In contrast the Bayesian intervals are narrow enough for many not to contain zero, suggesting that covariates have an impact. Maximum likelihood estimates from the original data (not included) match up reasonably well with the bootstrap values. However, if we treat the estimated change-point $\hat{\nu}$ as fixed, the approximate intervals for the other parameters (assuming asymptotic normality of the estimates) are much narrower than our Bayesian intervals.

Although the bootstrap distribution of the parameters seem too wide in general, the method does match up well with the Bayesian samples in terms of the change-point ν. Figure $B.1$ shows a high density of posterior samples around observations at 19 and 23 days. The bootstrap estimates were also mostly concentrated at these observations (Figure $B.2$). The major difference is that posterior samples converged to this area of the parameter space, while the bootstrap samples sometimes led to estimates of ν that were far away. It is possible that the bootstrap method could be improved by making some adjustment to the search grid for ν and by increasing the number of iterations. However, preliminary exploration into these adjustments shows that little is gained.

If we now focus on the Bayesian estimates, we see that interval for β_{CP} is well away from 0, suggesting that we do in fact have a change-point of the form (B.1), which is consistent with our test results from Chapters 3 and 5. The positive estimate means that the slope of the hazard function decreases abruptly after ν for patients not receiving the treatment. The presence of this change-point term actually increases the effect of the treatment, implying a larger difference in hazard function slopes of patients (and a larger difference in the expected time to infection) under the treatment versus the control. From Table B.1 it seems that treatment and gender are important, and age might have a weak effect.

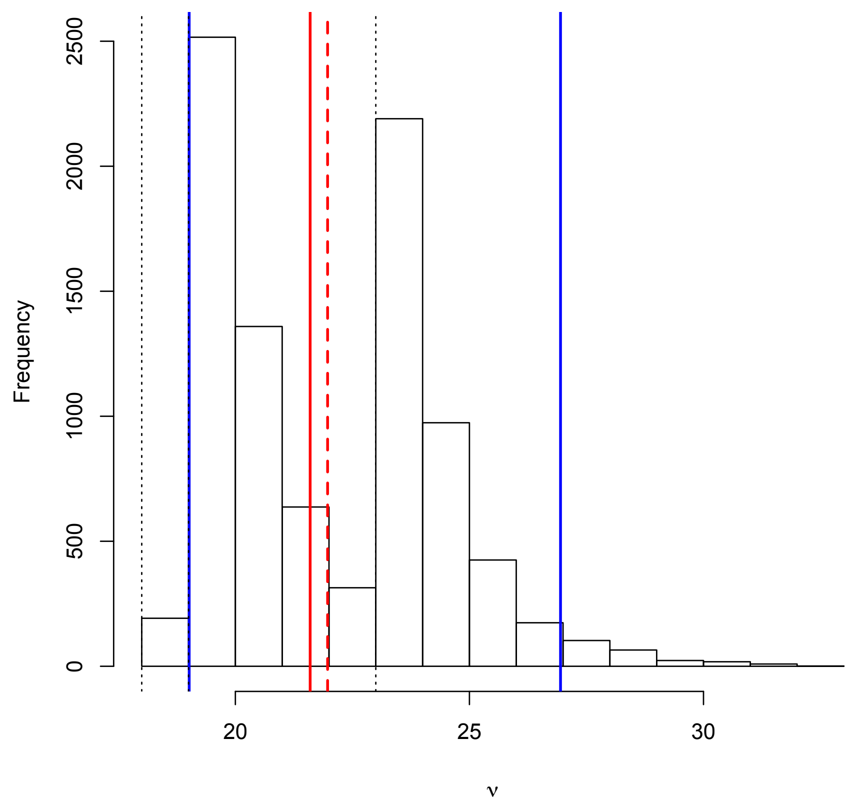

Figure B.1: Histogram for posterior sample of ν. Solid red line is the median and dashed red line is the mean. Blue lines are the 95% interval, and dotted black lines are the observed data.

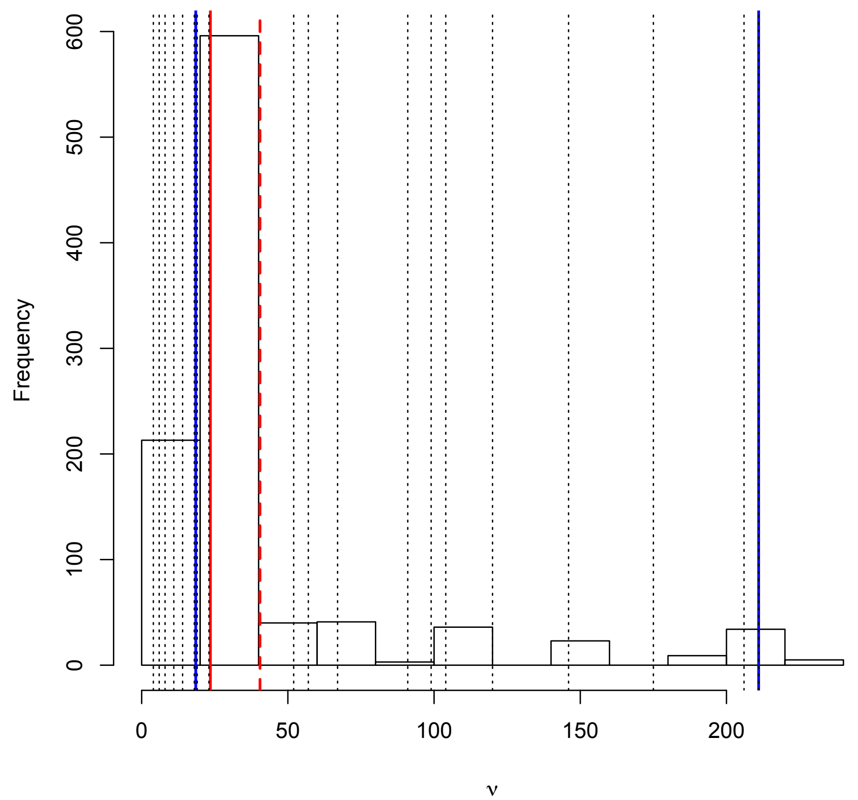

Figure B.2: Histogram for bootstrap sample of $\hat{\nu}$. Solid red line is the median and dashed red line is the mean. Blue lines are the 95% interval, and dotted black lines are the uncensored control group events.

Table B.1: Estimates for Bayesian and Bootstrap Analysis

Bayes Estimates				
Parameters	Mean	Median	95% Interval	
intercept	3.45	3.45	3.09	3.83
treatment	1.71	1.71	1.31	2.15
age	0.20	0.20	-0.01	0.42
gender	0.86	0.86	0.45	1.24
BetaCP	1.39	1.39	0.98	1.84
Phi	25.88	22.49	4.58	65.48
Nu	21.97	21.60	19.01	26.95

Bootstrap Estimates				
Parameters	Mean	Median	95% Interval	
intercept	4.64	4.38	3.83	6.34
treatment	1.37	1.54	-0.22	2.12
age	0.17	0.15	-0.08	0.48
gender	-0.06	0.00	-0.70	0.54
BetaCP	1.72	1.77	-0.72	2.45
Nu	40.5	23.5	18.5	211

Bibliography

[1] Achar, J. A. and Bolfarine, H. (1989). Constant hazard against a change-point alternative: A Bayesian approach with censored data. *Communications in Statistics. Theory and Methods*, 18:3801–3819.

[2] Alexander, K. S., Dudley, R. M., Gaenssler, P., Philipp, W., Pollard, D., Pyke, R., and Stute, W. (1984). Discussion of the paper of professors Giné and Zinn. *The Annals of Probability*, 12:990–998.

[3] Boos, D. D. (1979). A differential for L-statistics. *The Annals of Statistics*, 7(5):955–959.

[4] Breslow, N. and Crowley, J. (1974). A large sample study of the life table and product limit estimates under random censorship. *The Annals of Statistics*, 2:437–453.

[5] Casellas, J. (2007). Bayesian inference in a piecewise Weibull proportional hazards model with unknown change points. *Journal of Animal Breeding and Genetics*, 124:176–184.

[6] Cox, D. R. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society: Series B*, 34:187–220.

[7] Cox, D. R. (1975). Partial likelihood. *Biometrika*, 62:269–276.

[8] Csörgő, M. and Horváth, L. (1997). *Limit Theorems in Change-Point Analysis*. John Wiley & Sons.

[9] Davies, R. B. (1977). Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika*, 64:247–254.

[10] Davies, R. B. (1987). Hypothesis testing when a nuisance parameter is present only under the alternatives. *Biometrika*, 74:33–43.

[11] Dudley, R. M. (1999). *Uniform Central Limit Theorems*. Cambridge University Press.

[12] Duggins, J. W. (2010). *Parametric Resampling Methods for Retrospective Change-point Analysis*. PhD thesis, Virginia Polytechnic Institute and State University.

[13] Dupuy, J.-F. (2006). Estimation in a change-point hazard regression model. *Statistics & Probability Letters*, 76:182–190.

[14] Dupuy, J.-F. (2009). Detecting change in a hazard regression model with right-censoring. *Journal of Statistical Planning and Inference*, 139:1578–1586.

[15] Durbin, J. (1985). The first-passage density of a continuous Gaussian process to a general boundary. *Journal of Applied Probability*, 22:99–122.

[16] Durbin, J. (1988). Correction: The first-passage density of a continuous Gaussian process to a general boundary. *Journal of Applied Probability*, 25:840.

[17] Ferger, D. (2007). The maximum likelihood method with estimated nuisance parameters in hazard rate models with discontinuities. *Statistica Sinica*, 17:1091–1114.

[18] Fleming, T. R. and Garrington, D. P. (1991). *Counting Processes and Survival Analysis*. John Wiley & Sons.

[19] Gelman, A. and Meng, X.-L. (1998). Simulating normalizing constants: From importance sampling to bridge sample to path sampling. *Statistical Science*, 13:163–185.

[20] George, E. I. and McCulloch, R. E. (1993). Variable selection via Gibbs sampling. *Journal of the American Statistical Association*, 88:881–889.

[21] George, E. I. and McCulloch, R. E. (1997). Approaches for Bayesian variable selection. *Statistica Sinica*, 7:339–373.

[22] Ghosh, S. K. and Ebrahimi, N. (2008). Bayesian analysis of change-point hazard rate problem. *Journal of Statistical Theory and Practice*, 2:523–533.

[23] Giné, E. and Zinn, J. (1984). Some limit theorems for empirical processes. *The Annals of Probability*, 12:929–989.

[24] Goodman, M. S., Li, Y., and Tiwari, R. C. (2006). Survival analysis with change point hazard functions. *Harvard University Biostatistics Working Paper Series*, 40.

[25] Graybill, F. A. (1976). *Theory and Application of the Linear Model*. Duxbury Press.

[26] Green, P. J. (1995). Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika*, 82:711–732.

[27] Groenewald, P. and Schoeman, A. (2001). Bayesian estimation in change-point hazard rate models. *South African Statistical Journal*, 35:95–121.

[28] Henderson, R. (1990). A problem with the likelihood ratio test for a change-point hazard rate model. *Biometrika*, 77:835–843.

[29] Kim, D.-Y., Woodroofe, M., and Wu, Y. (2004). Testing for a change in the hazard rate with staggered entry. *Communications in Statistics. Theory and Methods*, 33:2041–2058.

[30] Kosorok, M. R. and Song, R. (2007). Inference under right censoring for transformation models with a change-point based on a covariate threshold. *The Annals of Statistics*, 35:957–989.

[31] Leadbetter, M. R., Lindgren, G., and Rootzén, H. (1983). *Extremes and Related Properties of Random Sequences and Processes*. Springer, New York.

[32] Liang, K.-Y., Self, S. G., and Liu, X. (1990). The Cox proportional hazards model with change point: An epidemiologic application. *Biometrics*, pages 783–793.

[33] Liu, M., Lu, W., and Shao, Y. (2008). A Monte Carlo approach for change-point detection in the cox proportional hazards model. *Statistics in Medicine*, 27:3894–3909.

[34] Loader, C. R. (1991). Inference for a hazard rate change point. *Biometrika*, 78:749–757.

[35] Luo, X., Turnbull, B. W., and Clark, L. C. (1997). Likelihood ratio tests for a change-point with survival data. *Biometrika*, 84:555–565.

[36] Matthews, D. E. and Farewell, V. T. (1982). On testing for a constant hazard against a change-point alternative. *Biometrics*, 38:463–468.

[37] Matthews, D. E. and Farewell, V. T. (1985). On a singularity in the likelihood for a change-point hazard rate model. *Biometrika*, 72:703–704.

[38] Matthews, D. E., Farewell, V. T., and Pyke, R. (1985). Asymptotic score-statistic processes and tests for constant hazard against a change-point alternative. *The Annals of Statistics*, 13:583–591.

[39] Meeker, W. Q. and Escobar, L. A. (1998). *Statistical Methods for Reliability Data*. John Wiley & Sons.

[40] Meng, X.-L. and Wong, W. H. (1996). Simulating ratios of normalizing constants via a simple density: A theoretical exploration. *Statistica Sinica*, 6:831–860.

[41] Na, M. H., Jeon, J., and Park, D. H. (2005). Testing whether failure rate changes its trend with unknown change points. *Journal of Statistic Planning and Inference*, 129:317–325.

[42] Na, M. H. and Lee, S. (2003). A family of IDMRL tests with unknown turning point. *Statistics*, 37(5):457–462.

[43] Nguyen, H. T., Rogers, G. S., and Walker, E. A. (1984). Estimation in change-point hazard rate models. *Biometrika*, 71:299–304.

[44] O'Hagan, A. (1997). Properties of intrinsic and fractional Bayes factors. *Test*, 6:101–118.

[45] O'Quigley, J. (2008). *Proportional Hazards Regression*. Springer.

[46] O'Quigley, J. and Natarajan, L. (2004). Erosion of regression effect in a survival study. *Biometrics*, pages 344–351.

[47] Park, D. H. (1988). Testing whether failure rate changes its trend. *IEEE Transactions in Reliability*, 37:375–378.

[48] Pham, D. T. and Nguyen, H. T. (1990). Strong consistency of the maximum likelihood estimator in the change point hazard rate model. *Statistics*, 21:203–216.

[49] Pham, T. D. and Nguyen, H. T. (1993). Bootstrapping the change-point of a hazard rate. *Annals of the Institute of Statistical Mathematics*, 45:331–340.

[50] Prentice, R. L. and Self, S. G. (1983). Asymptotic distribution theory for Cox-type regression models with general relative risk form. *The Annals of Statistics*, 11:804–813.

[51] Qin, J. and Sun, J. (1997). Statistical analysis of right-censored failure-time data with partially specified hazard rates. *The Canadian Journal of Statistics*, 25:325–336.

[52] Ritz, C. and Skovgaard, I. M. (2005). Likelihood ratio tests in curved exponential families with nuisance parameters present only under the alternative. *Biometrika*, 92:507–517.

[53] Siegmund, D. (1985). *Sequential Analysis: Tests and Confidence Intervals*. Springer-Verlag.

[54] Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and van der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B*, 64:583–639.

[55] van der Vaart, A. W. (1998). *Asymptotic Statistics*. Cambridge University Press.

[56] van der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes*. Springer.

[57] Worsley, K. T. (1988). Exact percentage points of the likelihood-ratio test for a change-point hazard-rate model. *Biometrics*, 44:259–263.

[58] Xu, R. and Adak, S. (2002). Survival analysis with time-varying regression effects using a tree-based approach. *Biometrics*, 58:305–315.

[59] Yao, Y. C. (1986). Maximum likelihood estimation in hazard rate model with a change point. *Communications in Statistics. Theory and Methods*, 15:2455–2466.

[60] Zhao, X., Wu, X., and Zhou, X. (2009). A change-point model for survival data with long-term survivors. *Statistica Sinica*, 19:377–390.