ESTIMATION OF THE BASELINE HAZARD FUNCTION IN COX'S REGRESSION MODEL UNDER ORDER RESTRICTION

By

DAEHYUN CHUNG

A DISSERTATION PRESENTED TO THE GRADUATE SCHOOL
OF THE UNIVERSITY OF FLORIDA IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

UNIVERSITY OF FLORIDA

1992

To My Wife
Gyehee

ACKNOWLEDGEMENTS

I would like to express sincerest gratitude to my advisor, Dr. M. Chang for his advice, guidance, encouragement, review and constructive criticisms in manuscript preparation.

Thanks and appreciation are extended to Dr. P. V. Rao, Dr. J. Shuster, and Dr. L. Chen who as members of the supervisory committee provided suggestion towards the improvement of the quality of research, manuscript review.

I would also like to thank Dr. B. Presnell for his careful perusal of this manuscript and constructive suggestion towards the improvement of the equality of the manuscript.

My special thanks and appreciation are due to Dr. Shuster and Dr. Cantor for providing me financial support during the last semester, without which my Ph.D. program would have been seriously disturbed.

Finally I can never repay a debt to my parents, wife, daughter and son for the love and understanding they have given me.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Abstract of Dissertation Presented to the Graduate School
of the University of Florida in Partial Fulfillment of the
Requirements for the Degree of Doctor of Philosophy

ESTIMATION OF THE BASELINE HAZARD FUNCTION IN
COX'S REGRESSION MODEL UNDER ORDER RESTRICTION

By

Daehyun Chung

May 1992

Chairman: Myron Chang
Major Department: Statistics

This dissertation is devoted to exploring the problem of inference about the baseline hazard function of Cox's regression model, especially when the baseline hazard function is assumed to be monotonic.

Assuming monotonicity of the baseline hazard function will improve the efficiency of estimation of the baseline hazard function in Cox's regression model. The isotonic regression method is applied to find the isotonic estimator of the baseline hazard function. The maximum likelihood estimation of parameters with order restriction is closely related to the problem of isotonic regression.

The test for the monotonicity of the baseline hazard function is discussed for random censoring model. The strong consistency of the isotonic estimator of the baseline hazard function is shown. To improve the maximum likelihood estimator of the baseline hazard function, when there are censored observations, we consider an alternative using the concept of the window. The asymptotic distribution of the isotonic window estimator of the baseline hazard function is obtained for fixed time t.

CHAPTER 1
INTRODUCTION

1.1 Review

Recent statistical research deals extensively with the methods for the analysis of survival data derived from laboratory studies of animals or clinical studies of humans. The proportional hazard model proposed by Cox (1972) is an important tool for analyzing such data. We intend to explore the problem of inference about the baseline hazard function in Cox's model. Survival data are different from the data that are gathered in conventional studies, because survival data often include censoring times, which precludes exact determination of the key dependent variable, survival time.

In medical studies, experimenters frequently face censored data in clinical trials for chronic diseases. Some patients may withdraw from the study; others may die for nonrelated reasons. Still others may be alive at last contact. For lost patients, survival times are at least as large as the elapsed time of entry and the time they withdraw or die for non-related reasons. For patients still alive, survival times are at least as large as the time from entry to the time of the end of the study. These observations, either withdrawal or failure for competing reasons during the study, are defined to be censored observations. Loosely speaking, a censored observation contains only partial information about the random variable of interest.

In survival analysis, we usually encounter three types of censoring. Type I censoring occurs when we have a fixed censoring time, so that an observation is uncensored only if failure occurs before the fixed censoring time.

Because of financial constraints, we may only observe first r observations out of n possible observations. In other words, observation ceases after the rth failure. This type of censoring is defined as type II censoring.

Both type I and II censorings arise frequently in engineering sciences. For instance, we use a batch of electrical bulbs as experimental units. We turn on all electrical bulbs simultaneously, when we start an experiment, to investigate how long they last on the average. Since it may take an extremely long time for some bulbs to burn out, we usually cannot wait until all light bulbs burn out. We may be forced to stop the experiment at a prespecified fixed time or the time when a prespecified fraction of all light bulbs have burned out. The first case is classified as type I censoring, while the second case is an example of type II censoring.

The most general type of censoring is random censoring. Random censoring occurs, unlike type I and II censoring, when censoring times of individuals are treated as random variables from an unknown distribution.

Random censoring arises in medical applications with animal studies or clinical trials. In clinical trials, each patient may join the study at different times. Our concern is to measure how long they survive, while we treat them with one or several therapies. But we may lose patients for reasons unrelated to the factors being studied. For example, a cancer patient may move and never report back to the clinic center, or he may refuse to continue receiving designated treatments which he considers unsatisfactory. Another example occurs when a cancer patient dies in a car accident. The cause of death is not cancer, but an accident which is not related to our study objectives . We only know that the individual survived until the car accident. In this dissertation, we shall concentrate on random censoring.

With random censoring, we make the following basic assumption: The censoring mechanisms are "noninformative." In other words, the censoring time is independent of the survival time.

One of the principal problems in survival analysis is that of developing methods for exploring the association between failure times and explanatory variables. For example, a clinical study is designed to compare several treatment programs in terms of the failure times. The explanatory variables would include indicator components for treatment as well as other prognostic factors. The Cox regression model is a conventional technique for investigating the relationship between survival time and covariates.

The hazard or failure rate function is conceptually simple and is a specialized way of representing the distribution of the failure times. The hazard function gives the risk of failure at any time t, given that the individual has not failed prior to time t. Cox (1972) suggests the following model which presumes that covariates affect the hazard function multiplicatively. Let Z be a row vector of p covariates. Then the Cox model satisfies that

$$\lambda(t; z) = \lambda_0(t) \exp(z\beta), \tag{1.1}$$

where $\lambda_0(t)$ is an unspecified function of time and β is a p–dimensional column vector of parameters. This model, though largely nonparametric, permits the estimation of β and leads to estimates of survival functions of the Kaplan and Meier type (1958), when covariates are present in the data.

Since the ratio of the hazard functions corresponding to any two different z-values is constant over t, (1.1) is often called a proportional hazard model. The factor $\exp(z\beta)$ describes the instantaneous risk of failure for an individual with covariate z relative to that at a standard value $z = 0$. Since $\lambda_0(t)$ gives the hazard for an individual under the standard condition $z = 0$, $\lambda_0(t)$ is called the baseline hazard function.

One of the attractive features of the model (1.1) is that the nuisance function $\lambda_0(t)$ can be removed completely from inferences about β (Cox, 1975). Another advantage is that the covariate information on different individuals is easily incorporated into (1.1).

Several assumptions on $\lambda_0(t)$ are possible in the analysis of the model (1.1). The simplest one is to assume $\lambda_0(t)$ is constant. This is equivalent to assuming an underlying exponential distribution. The next simplest case is to assume that a family of hazard functions has two unknown parameters. The Weibull distribution is an example of a two parameter family of hazard functions. A weaker assumption is that $\lambda_0(t)$ is arbitrary but monotonic increasing or decreasing in t. In certain situations, it is reasonable to expect that the failure rate will increase monotonically or, at least over a certain interval of time. For certain electronic components, manufacturing defects tend to cause failure early in life, so that the failure rate may be higher during the initial period of age. This is the case when a decreasing failure rate can be expected. In many physical situations, the object does become more likely to fail as it ages. Examples of these are moving parts, human beings past youth and so on. In such cases, one would expect an increasing failure rate.

A main problem of considerable interest is the inference about the regression parameters, allowing the baseline hazard function to be arbitrary. The conditional likelihood approach suggested by Cox(1972) is a pioneering method leading to inference about the regression parameter β.

Cox writes: "Suppose then $\lambda_0(t)$ is arbitrary. No information can be contributed about β by the time intervals in which no failure occurs because the components $\lambda_0(t)$ might conceivably be identically zero in such intervals. We therefore argue conditionally on the set of instants at which failures occur; in discrete time, we shall condition also on the observed multiplicities. Once we require a method of analysis holding for all $\lambda_0(t)$, consideration of this conditional distribution seems inevitable." He treats his conditional likelihood as an ordinary likelihood, so that he finds maximum likelihood estimators and their asymptotic distribution for the regression parameter β.

The method of marginal likelihood was developed by Kalbfleish and Prentice (1973), for the analysis of the regression parameters in model (1.1). The order statistics and the rank statistics of observed failure times are the focus of their discussion. They consider the group G of differentiable, strictly monotone increasing transformations of $(0, \infty)$ onto $(0, \infty)$. They argue that the estimation problem for the regression parameters based on rank statistics is invariant under the group G of transformations on the failure time t. The group G acts transitively on the order statistics, while leaving the rank statistics invariant. Only the rank statistics can carry information about the regression parameters when $\lambda_0(t)$ is completely unknown. That is, the rank statistics are marginally sufficient for the estimation of the regression parameters. The marginal likelihood of the regression parameters is proportional to the probability that the rank vector should be observed from the marginal distribution of the ranks. For censored data, the marginal likelihood becomes more complicated if the number of ties is large, but the computation can be simplified by using an approximation suggested by Breslow (1974). For uncensored data, the marginal likelihood is identical to the conditional likelihood.

The partial likelihood approach to inferences about the regression parameters which gives essentially equivalent results to those given by marginal likelihood is described by Cox (1975). The partial likelihood is useful especially when it is appreciably simpler than the full likelihood, as for example, when it involves only the parameters of interest and no nuisance parameters. A reduction of dimensionality, when we have many nuisance parameters, is possible by using partial likelihood. This approach is especially can be fruitful when $\lambda_0(t)$ is assumed to be an unknown arbitrary function, to be treated as a nuisance function. Cox (1975) shows that the marginal likelihood derived by Kalbfleish and Prentice (1973) is equivalent to a partial likelihood.

1.2 The Problem

All of the arguments above are considered with the assumption that $\lambda_0(t)$ is completely unspecified. If we have additional information on $\lambda_0(t)$, for example, monotonicity or constancy of $\lambda_0(t)$, since information should be useful for inference of the regression parameters in (1.1). From a practical point of view, we can have empirical or prior information about the hazard rate without taking into account the effect of covariates. If experimental units are like light bulbs, machine tools, or car engines, we are concerned about the way in which the items in question wear out. It is reasonable to assume that the failure rate of aging items will tend to increase, when we do not consider the effect of the covariates.

The efficiency of inference about the regression parameters under various assumptions about $\lambda_0(t)$ is referred to as a "major outstanding problem" by Cox (1972). Many researchers have attempted to answer the above problem. Meshalkin and Kagan (1972) showed that knowledge of the baseline hazard function is helpful in reducing the asymptotic variances of the estimates of β in the model (1.1) by 10 to 20 %. They assume that the baseline hazard function has an exponential form of a linear function of t. Efron (1977) argues that if the class of nuisance functions is large, then the inferences about the regression parameters based on partial or marginal likelihood are asymptotically equal to those based on all the data. He also carries out the calculation of an information matrix which shows that Cox's partial likelihood has full asymptotic efficiency under mild conditions. Oakes (1977) also deals with the same problem from a different point of view. Efron (1977) and Oakes (1977) use different parametrizations of the baseline hazard function. In Efron's formulation, the baseline hazard function may depend on the regression parameters as well as the nuisance parameters. The baseline hazard function is assumed to be either known completely or known up to a multiplicative constant by Oakes. Explicit formulas for the asymptotic variances of the estimates of β are derived informally and compared. Oakes also concludes that the amount of information lost through a lack of knowledge of the baseline hazard function in any specified data set is usually small.

Inferences about the baseline hazard function $\lambda_0(t)$ are also an important part of survival analysis, because $\lambda_0(t)$ reveals the survival pattern. Breslow (1974) suggests that $\lambda_0(t)$ be approximated by a step function which has discontinuities at observed failure times. He considers the joint likelihood function of $\lambda_0(t)$ and β and derives the maximum likelihood estimators of $\lambda_0(t)$ and β. Beyond Breslow's paper, there is, however, little discussion on inferences about $\lambda_0(t)$ in the literature.

For the one population problem, i.e., for $\beta \equiv 0$ in (1.1), many results concerning inferences about hazard functions under order restrictions are available in literature. Grenander (1956) was the first to use the concept of the greatest convex minorant to estimate the failure rate under the assumption that the failure rate is monotonically increasing. The various problems of inference under order restrictions are discussed by Barlow et al. (1972) and Robertson et al. (1988), who deal with a wide class of extremum problems whose solutions are provided by isotonic regression. They discuss the problems of estimating monotone failure functions and of prove strong consistency of the isotonic estimates of monotone failure functions. They examine tests for exponentiality against monotone failure rate alternatives in situations with type II censoring and random censoring.

We intend to extend the results suggested by Barlow and coworkers (1972) to estimate the baseline hazard function under the assumption that it is increasing in time t. and to test whether the baseline hazard function is constant or increasing in t.

1.3 Preview

The objective of this dissertation is how to estimate the baseline hazard function, $\lambda_0(t)$, in the Cox regression model, under the assumption that $\lambda_0(t)$ is increasing or decreasing monotonically and examine the consistency and normality of its estimator. We are also concerned about the test for the monotonicity of the baseline hazard function.

Chapter 2 contains tests for constant failure rate versus increasing failure rate. We develop a procedure for Type II censoring and then extend it to random censoring with covariates.

In chapter 3, we focus on estimating the baseline hazard function by a step function. Isotonic regression is adapted to find the estimators by solving likelihood equations under order restrictions on parameters.

In chapter 4, we prove the strong consistency of the estimate of $\lambda_0(t)$ using the strong consistency of estimator for β obtained by partial likelihood.

In chapter 5, we deal with the problem of improving the maximum likelihood estimator of $\lambda_0(t)$ by considering windows. The asymptotic distribution of the isotonic estimator of $\lambda_0(t)$ is found. Finally, the optimal size of the window is derived by minimizing the mean square error of its estimator.

Chapter 6 contains the results of simulation study which show the superiority of the isotonic estimator over the maximum likelihood estimator of $\lambda_0(t)$ under the monotonicity assumption.

CHAPTER 2
TEST FOR MONOTONICITY OF THE BASELINE HAZARD FUNCTION

2.1 Introduction

When it is known a priori that the baseline hazard function is an increasing function, that information can be used to find a better estimates of the baseline hazard function. Further, when no information about the baseline hazard function is available, it is of interest to check whether it is monotone increase or not, before we estimate the baseline hazard function.

In this chapter, we will develop methods by which we can test the hypothesis of constant versus increasing baseline hazard function. That is,

$$H_0 : \lambda_0(t) \text{ is constant}$$

versus

$$H_1 : \lambda_0(t) \text{ is increasing.} \tag{2.1}$$

The cumulative total time on test statistic is a fundamental tool used to develop the proposed test procedure.

The problem of testing the hypotheses (2.1) without the covariates (i.e., $\beta \equiv 0$ in (1.1)) is reviewed in section 2. In section 3, we extend the concept of total time to test the hypotheses (2.1) under Cox's regression model with random censoring. In section 4, we illustrate a graphical method of testing the hypotheses (2.1).

2.2 Test for Isotonicity without Covariates.

We consider the problem of testing the hypotheses (2.1) where $\lambda_0(t)$ is given by $\frac{f(t)}{1-F(t)}$. This problem is equivalent to a test of the hypothesis that F is an exponential distribution against the increasing failure rate alternative, since under the null hypotheses (2.1), the failure time T has an exponential distribution with mean $\frac{1}{\lambda}$.

Bickel and Doksum (1969) discuss the problem of testing for the given hypothesis (2.1) when a sample of complete observations is available. Their test is based on the ranks of the normalized spacings between ordered observations. Their results are extended by Barlow and Doksum (1972) to the case where we have Type II censoring. The total time on test statistic is considered as a key to develop the test for the hypotheses (2.1).

In this section, we summarize the results in Barlow and Doksum.

Suppose a study continues until the kth failure time occurs and at that time all surviving individuals are assumed to be censored. We obtain the first k ordered observations out of a sample of n individuals:

$$Y_{u(1)} \leq Y_{u(2)} \leq \cdots \leq Y_{u(k)}, \qquad (1 \leq k \leq n)$$

Let $D_{n:i} = (n-i+1)(Y_{u(i)} - Y_{u(i-1)})$, $i = 1, \cdots, n$, be the normalized sample spacings, where $Y_{u(0)} = 0$. It is well known that when the failure time has an exponential distribution with parameter λ, the normalized spacings $D_{n:i}$ are independent and exponentially distributed random variables with mean $\frac{1}{\lambda}$.

We have the following theorem that forms the basis for many tests for exponentiality versus increasing (or decreasing) failure rate.

<u>*Theorem 2.2.1*</u> *If F has an increasing failure rate, then the $D_{n:i} = (n-i+1)(Y_{u(i)} - Y_{u(i-1)})$ $i = 1, \cdots, n$, is stochastically decreasing in i $(= 1, 2, \cdots, n)$ for fixed n.*

<u>*Proof of Theorem 2.2.1*</u> *See Breslow et al.(1972)*

Under the alternative hypothesis of increasing failure rate, Theorem 2.2.1 implies that

$$D_{n:1} \underset{\text{st}}{\geq} D_{n:2} \underset{\text{st}}{\geq} \cdots \underset{\text{st}}{\geq} D_{n:n},$$

while under the null hypothesis of constant failure rate the above theorem implies that

$$\Pr(D_{n:i} < D_{n:j}) = \frac{1}{2} \qquad i \neq j$$

where $\underset{\text{st}}{\geq}$ implies stochastic ordering. We have zero slope in the linear regression of the $D_{n:n-i}$ on the value i under exponentiality of the distribution F. Since $D_{n:i}$ tends to be larger than $D_{n:j}$ for $i < j$, the slope in the linear regression is positive under the assumption of increasing failure rate.

When the slope is nonzero, it is, as usual, sensitive to change in scales. Therefore it is desirable to make the statistic scale invariant by dividing the slope by the average of $D_{n:i}$, i.e., $\frac{1}{n}\sum_{i=1}^{n} D_{n:i}$.

The slope in the regression of the $D_{n:n-i}$ on i is linearly related to the sum of the areas of triangles formed by $(0,0), (i,0)$, and $(i, D_{n:n-i})$ for $i = 0, \cdots, (n-1)$. Let us define

$$V_n \stackrel{\text{def}}{=} \frac{1}{n}\sum_{i=0}^{n-1} i D_{n:n-i} \Big/ \frac{1}{n}\sum_{i=1}^{n} D_{n:i}$$

$$= \frac{1}{n}\sum_{i=1}^{n}(i-1)D_{n:n-i+1}\Big/\frac{1}{n}\sum_{i=1}^{n}D_{n:i}.$$

We can rewrite V_n as

$$\begin{aligned}
V_n &= \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{i-1}D_{n:n-i+1}\Big/\frac{1}{n}\sum_{i=1}^{n}D_{n:i} \\
&= \frac{1}{n}\sum_{j=1}^{n-1}\sum_{i=j+1}^{n}D_{n:n-i+1}\Big/\frac{1}{n}\sum_{i=1}^{n}D_{n:i} \\
&= \frac{1}{n}\sum_{j=1}^{n-1}\sum_{i=1}^{n-j}D_{n:i}\Big/\frac{1}{n}\sum_{i=1}^{n}D_{n:i} \\
&= \frac{1}{n}\sum_{k=n-1}^{1}\sum_{i=1}^{k}D_{n:i}\Big/\frac{1}{n}\sum_{i=1}^{n}D_{n:i} \\
&= \frac{1}{n}\sum_{k=1}^{n-1}\sum_{i=1}^{k}D_{n:i}\Big/\frac{1}{n}\sum_{i=1}^{n}D_{n:i}. \qquad (2.2)
\end{aligned}$$

The term $T_n(Y_{u(i)}) = \sum_{j=1}^{i} D_{n:j}$ in (2.2) is recognized as the total time on test statistic up to the ith failure. If n individuals are placed on test, when testing commences, then n individuals survive up to time $Y_{u(1)}$, $(n-1)$ individuals survive through the interval $[Y_{u(1)}, Y_{u(2)})$, etc. In general, $(n-i+1)$ individuals survive through the interval $[Y_{u(i-1)}, Y_{u(i)})$. Hence

$$\begin{aligned}
T_n(Y_{u(i)}) &= \sum_{j=1}^{i} D_{n:j} \\
&= nY_{u(1)} + (n-1)(Y_{u(2)} - Y_{u(1)}) + \cdots + (n-i+1)(Y_{u(i)} - Y_{u(i-1)})
\end{aligned}$$

is interpreted as the total time on test statistic, which is the sum of times for individuals who survived up to the ith failure. The following definition will be needed to develop a test for the hypotheses given in (2.1).

Definition. Given the first k $(1 \leq k \leq n)$ ordered observations out of n individuals,

$$
\begin{aligned}
V_k &\stackrel{\text{def}}{=} \frac{1}{k}\sum_{i=1}^{k-1}\sum_{j=1}^{i} D_{n:j} \Big/ \frac{1}{k}\sum_{i=1}^{k} D_{n:i} \\
&= \sum_{i=1}^{k-1} \frac{T_n(Y_{u(i)})}{T_n(Y_{u(k)})}
\end{aligned}
\tag{2.3}
$$

is called the *cumulative total time on test statistic,* where $D_{n:i} = (n - i + 1)(Y_{u(i)} - Y_{u(i-1)})$ for $i = 1, \cdots, n$.

Under the alternative hypothesis that $\lambda(t)$ is increasing, Theorem 2.2.1 shows that V_k tends to be large by the fact that $D_{n:j}$ is stochastically decreasing in j. We may not accept the null hypothesis that $\lambda(t)$ is constant when V_k is fairly large. Hence the statistic V_k $(1 \leq k \leq n)$ provides a good tool for determining whether the failure rate is indeed constant or increasing.

In order to perform the test based on V_k we need to find its null distribution. We shall use the following well-known theorem to find the statistic which is stochastically equivalent to V_k and whose distribution is known.

Theorem 2.2.2 If the failure times have an exponential distribution with parameter λ, then

$$
V_k \underset{\text{st}}{=} \sum_{j=1}^{k-1} U_j \tag{2.4}
$$

where U_j $(j = 1, \cdots, k-1)$ are independent uniform random variables on (0,1).

To prove this theorem we need to prove the following lemmas.

<u>Lemma 2.2.1</u> Conditional on $\sum_{i=1}^{n} D_{n:i} = m$, $D_{n:1}, \cdots, D_{n:n-1}$ *have a uniform distribution over the area:*

$$d_i \geq 0, \qquad i = 1, \cdots, n-1, \qquad \sum_{i=1}^{n-1} d_i \leq m,$$

under the null hypothesis that $\lambda_0(t)$ *is constant.*

<u>Proof of Lemma 2.2.1</u> See Appendix A.

<u>Lemma 2.2.2</u> Let $X_i = \frac{D_{n:i}}{\sum_{j=1}^{n} D_{n:j}}$, $\quad i = 1, \cdots, n-1$. *Then under the null hypothesis that* $\lambda_0(t)$ *is constant, and conditional on* $\sum D_{n:i} = m, X_1, \cdots, X_{n-1}$ *have a uniform distribution over the area:*

$$x_i \geq 0 \qquad i = 1, \cdots, n-1 \qquad x_1 + \cdots + x_{n-1} \leq 1. \tag{2.5}$$

<u>Proof of Lemma 2.2.2</u> It is seen that Lemma 2.2.2 is a consequence of Lemmma 2.2.1 by a scale change.

<u>Proof of Theorem 2.2.2</u> This theorem is stated by Barlow(1972), but no proof of this theorem is given. Hence, we have provided a proof in Appendix B.

Theorem 2.2.2 demonstrates that V_k can be considered as sum of $(k-1)$ i.i.d. uniform random variables on (0,1). Hence it is possible to compute $C_{k,(1-\alpha)}$ such that

$$\begin{aligned} \alpha &= \Pr[\text{Reject } H_0 | H_0 \text{ is true}] \\ &= \Pr[V_k > C_{k,(1-\alpha)} | \lambda(t) \text{is constant}]. \end{aligned}$$

Table 2.1. Percentiles $C_{k,1-\alpha}$ of the Cumulative Total Time on Test Statistic, V_k under H_0

k-1	$1-\alpha$				
	0.900	0.9500	0.975	0.990	0.995
2	1.533	1.684	1.776	1.859	1.900
3	2.157	2.331	2.469	2.609	2.689
4	2.753	2.953	3.120	3.300	3.411
5	3.339	3.565	3.754	3.963	4.097
6	3.917	4.166	4.367	4.610	4.762
7	4.489	4.759	4.988	5.244	5.413
8	5.056	5.346	5.592	5.869	6.053
9	5.619	5.927	6.189	6.487	6.683
10	6.178	6.504	6.781	7.097	7.307
11	6.735	7.077	7.369	7.702	7.924
12	7.289	7.647	7.953	8.302	8.535

k= number of failures observed in data.

Table 2.2. Percentiles $C_{k,\alpha}$ of the Cumulative Total Time on Test Statistic, V_k under H_0

k-1	α				
	0.100	0.0500	0.125	0.010	0.005
2	0.447	0.316	0.224	0.141	0.100
3	0.843	0.669	0.531	0.391	0.311
4	1.247	1.047	0.880	0.700	0.589
5	1.661	1.435	1.246	1.037	0.903
6	2.083	1.834	1.633	1.390	1.238
7	2.511	2.241	2.012	1.756	1.587
8	2.944	2.645	2.408	2.131	1.947
9	3.381	3.073	2.811	2.513	2.317
10	3.822	3.496	3.219	2.903	2.693
11	4.265	3.293	3.631	3.298	3.076
12	4.711	4.353	4.047	3.698	3.465

k=number of failures observed in data.

From Theorem 2.2.2, it follows that the null distribution of V_k is approximately normal with mean $\frac{1}{2}(k-1)$ and variance $\frac{1}{12}(k-1)$ for large k. By standardizing V_k, it follows that

$$\frac{V_k - \frac{1}{2}(k-1)}{\sqrt{\frac{(k-1)}{12}}} = \{12(k-1)\}^{\frac{1}{2}}[(k-1)^{-1}V_k - \frac{1}{2}]$$

converges to a $N(0,1)$ random variable under the null hypothesis that $\lambda(t)$ is constant. To perform the test we use critical numbers $C_{k,(1-\alpha)}$ which are tabulated by Barlow and Proschan (1968) for small k (see Table 2.1). For example, if $1-\alpha = .90$ we look in row $k-1 = 4$ and $C_{4,0.90} = 2.753$ in Table 2.1 and 2.2 . If the observed value of V_k greater than the number 2.753, it is concluded that the true distribution has an increasing failure rate at the 10% significance level. If the observed value of V_k is less than 1.247, we can make the opposite conclusion such that the true distribution has a decreasing failure rate with a 10% significance level.

Example (Barlow, 1972). In Table 2.3, we list the times between air- conditioner failures on selected aircraft. After roughly 2000 hours of service the planes received major overhauls: the failure interval containing major overhaul is omitted from the listing since the length of that failure interval may have been affected by the overhaul.

We wish to determine if the intervals between failures have an exponential distribution or if there is a wearout trend as the equipment ages. In the event that there is a wearout trend, maintenance should be scheduled according to equipment age rather than the present policy.

The V_k associated with the data in Table 2.3 are given in Table 2.4. Since the sample size for plane 7908 exceeds the range of Table 2.1, we can use the fact that for large k, and under H_0

$$Z = \{12(k-1)\}^{\frac{1}{2}}[(k-1)^{-1}V_k - \frac{1}{2}]$$

Table 2.3. Interval Between Failures of Air Conditioning Equipment on Jet Aircraft

K	aircraft				
	7907	7908	7915	7916	8044
1	194	413	359	50	487
2	15	14	9	254	18
3	41	58	12	5	100
4	29	37	270	283	7
5	33	100	603	35	98
6	181	65	3	12	5
7		9	104		85
8		169	2		91
9		447	436		43
10		184			230
11		36			3
12		201			130
13		118			
14		34			
15		31			
16		18			
17		18			
18		67			
19		57			
20		62			
21		7			
22		22			
23		34			

Major overhaul before the 14th observation

Table 2.4. Statistics and Conclusion

plane	sample size k	statistic V_k	conclusion
7907	6	$V_6 = 2.243$	Exponential at the 10 % level
7908	23	$V_{23} = 8.829$ Z=-1.607	Exponential at the 10 % level
7915	9	$V_9 = 2.80$	Decreasing failure rate at the 10 % level
7916	6	$V_6 = 1.67$	Exponential at the 10 % level
8044	12	$V_{12} = 4.22$	Exponential at the 10 % level

is approximately normally distributed with mean 0 and variance 1.

Using the fact that

$$V_k \underset{st}{=} U_1 + \cdots + U_{k-1}$$

is symmetric about $\frac{(k-1)}{2}$, i.e.,

$$\Pr(V_k - \frac{k-1}{2} \geq x) = \Pr(V_k - \frac{k-1}{2} \leq -x)$$

we can obtain the lower critical numbers for V_k. Those numbers are given in Table 2.2. If V_k is less than the lower critical number, we conclude that the data are from a distribution with a decreasing failure rate. For plane 7915 we obtain V_9=2.80 which is less than 2.944 in Table 2.2. Hence, we conclude at the 10% significant level that the failure times of the air-conditioner of plane 7915 have a decreasing failure rate.

2.3 Test for Isotonicity with Covariates.

We would like to extend the method introduced in Section 1 to test the monotonicity of the baseline hazard function in Cox's regression model. Our concern is to determine whether the baseline hazard rate is constant or increasing, regardless of the values of the covariates. We assume that the observations are subject to random censoring. Barlow and Proschan (1969) deal with the same problem for samples subject to random censoring when no covariates are available. It does not substantially complicate matters to develop the method to test the hypothesis about the baseline hazard function when covariates are available.

Suppose n individuals are put on test at time $t = 0$. Among n individuals, we assume that k individuals failed and the remaining $(n-k)$ individuals are censored. Let $Y_{u(1)} < Y_{u(2)} < \cdots < Y_{u(k)}$ be the ordered failure times with corresponding covariates $Z_{(1)}, Z_{(2)}, \cdots, Z_{(k)}$. Suppose that m_i individuals with covariates $Z_{(i1)}, \cdots, Z_{(im_i)}$ are censored in the interval $[Y_{u(i)}, Y_{u(i+1)})$, for $i = 0, 1, \cdots, k$, where $Y_{u(0)} = 0$ and $Y_{u(k+1)} = \infty$.

The set of actual survival times for the n individuals can be characterized by

$$y_{u(1)} < y_{u(2)} < \cdots < y_{u(k)} \qquad y_{u(i)} \leq y_{c(i1)} \cdots < y_{c(im_i)}$$

where $y_{c(i1)} \cdots y_{c(im_i)}$ are the failure times associated with individuals censored in the interval $[Y_{u(i)}, Y_{u(i+1)})$.

Now let $h(y_{u(i)})$ denote the conditional probability that $Y_{u(i)} < Y_{c(i1)}, \cdots, Y_{c(im_i)}$, given $Y_{u(i)} = y_{u(i)}, i = 0, 1, \cdots, k$. Then (see Kalbfleisch & Prentice, 1980)

$$h(y_{u(i)}) = \Pr[Y_{u(i)} < Y_{c(i1)}, \cdots, Y_{c(im_i)} | Y_{u(i)} = y_{u(i)}]$$

$$= b\prod_{l=0}^{m_i} \exp[-\int_0^{y_{u(i)}} \lambda_0(u)\exp(z_{(il)}\beta)]du$$

$$= \exp[-\sum_{l=1}^{m_i} \exp(z_{(il)}\beta)\int_0^{y_{u(i)}} \lambda_0(u)du] \qquad i = 0, 1, \cdots, k. \qquad (2.6)$$

Define

$$n(y_{u(i)}) = \sum_{j\in R(y_{u(i)})} \exp(z_j\beta)$$

$$= \sum_{j=i}^{k}[\exp(z_{(j)}\beta) + \sum_{l=1}^{m_j}\exp(z_{(jl)}\beta)] \qquad i = 0, 1, \cdots, k. \qquad (2.7)$$

where $R(t)$ is the risk set prior to t.

Note that if $\beta = 0$, then $(y_{u(i)} - y_{u(i-1)})(n(y_{u(i-1)}) - 1)$ is the total time on test between the $(i-1)$st and the ith observed failures,

Theorem 2.3.1 *Let*

$$U_i = \int_{Y_{u(i-1)}}^{Y_{u(i)}} n(t)\lambda_0(t)dt \qquad i = 1, \cdots, k,$$

where $n(t) = \sum_{l\in R(t)} \exp(z_l\beta)$ *and* $Y_{u(0)} = 0$. *Then* $U_i, i = 1, \cdots, k$ *are independently distributed with density* $\exp(-u)$.

Proof of Theorem 2.3.1 *Let*

$$S_0(t) = \int_0^t n(x)\lambda_0(x)dx.$$

$S_0(t)$ *is well defined up to the first observed failure,* $y_{u(1)}$, *since* $n(t)$ *depends only upon* *the numbers of the observed censored individuals less than $y_{u(1)}$ which are greater than t. By definition*

$$U_1 = \int_0^{Y_{u(1)}} n(x)\lambda_0(x)dx = S_0(Y_{u(1)}).$$

Now to show that U_1 has density $\exp(-u_1)$, *we compute*

$$\begin{aligned}
\Pr(U_1 > u_1) &= \Pr(S_0(Y_{u(1)}) > u_1) \\
&= \Pr(Y_{u(1)} > S_0^{-1}(u_1)) \\
&= \exp[-\sum_{j=1}^{k}\left(\exp(z_{(j)}\beta) + \sum_{l=1}^{m_j}\exp(z_{jl}\beta)\right)\int_0^{S_0^{-1}(u_1)}\lambda_0(x)dx] \\
&= \exp[-\int_0^{S_0^{-1}(u_1)} n(x)\lambda_0(x)dx] \\
&= \exp[-S_0(S_0^{-1}(u_1))] \\
&= \exp(-u_1),
\end{aligned}$$

using (2.6). Next we will show U_2 is independent of U_1 and also exponentially distributed with mean 1. Let

$$U_2 = \int_{Y_{u(1)}}^{Y_{u(2)}} n(t)\lambda_0(t)\,dt \qquad \textit{and} \qquad S_{x_1}(t) = \int_{x_1}^{t} n(x)\lambda_0(x)\,dx.$$

Then the conditional probability that the U_2 is greater than u_2 given the first failure occurs at x_1 is

$$\begin{aligned}
\Pr(U_2 > u_2)|Y_{u(1)} = x_1) &= \Pr(S_{x_1}(Y_{u(2)}) > u_2|Y_{u(1)} = x_1) \\
&= \Pr(Y_{u(2)} > S_{x_1}^{-1}(u_2)|Y_{u(1)} = x_1) \\
&= \exp[-S_{x_1}(S_{x_1}^{-1}(u_2))] \\
&= \exp[-u_2].
\end{aligned}$$

Thus U_2 is independent of U_1 and also exponentially distributed with mean 1. Continuing in this manner by conditioning on previous events, we prove that U_i for $i = 1, \cdots, k$, are independent and distributed exponentially with mean 1.

Since we wish to test the null hypothesis that the baseline hazard function is constant, we put $\lambda_0(t) = \lambda$. Then by a simple transformation, we note that

$$U_i = \int_{Y_{u(i-1)}}^{Y_{u(i)}} n(u)du \qquad i = 1, \cdots, k$$

are independent exponentially distributed with mean $\frac{1}{\lambda}$.

<u>Theorem 2.3.2</u> *Let us define V_k by*

$$\begin{aligned} V_k &= \frac{\sum_{i=1}^{k-1} \int_0^{Y_{u(i)}} n(u)du}{\int_0^{Y_{u(k)}} n(u)du} \\ &\underset{st}{=} \sum_{i=1}^{k-1} \frac{\sum_{j=1}^{i} U_j}{\sum_{i=1}^{k} U_i} \end{aligned}$$

where U_i's are i.i.d. random variables with exponential distributions with parameter 1.

Then under H_0, V_k is distributed as the sum of $(k-1)$ independent uniform random variables over (0, 1) when β in $n(u)$ is known.

<u>Proof of Theorem 2.3.2</u> *The proof follows immediately from Theorem 2.2.2.*

The next theorem implies that V_k is reasonable test statistic for the given hypothesis.

Theorem 2.3.3 If $\lambda_0(u)$ is increasing in $u \geq 0$ and $n(u) \geq 0$ for $u \geq 0$, then

$$V_k \underset{\text{st}}{\geq} \sum_{i=1}^{k-1} \frac{\sum_{j=1}^{i} U_j}{\sum_{j=1}^{k} U_j}$$

where $U_1, U_2, \cdots, U_k$ are independently distributed as exponential random variables with mean 1.

Proof of Theorem 2.3.3 Since $n(u) \geq 0$ and $\lambda_0(u)$ is increasing in u, we have

$$\int_0^t \lambda_0(u)n(u)du \leq \lambda_0(t)\int_0^t n(u)du$$

which implies $\frac{\int_0^t \lambda_0(u)n(u)du}{\int_0^t n(u)du}$ is increasing in $t \geq 0$. Hence for $i = 1, \cdots, k$ we have

$$\frac{\int_0^{Y_{u(i)}} \lambda_0(u)n(u)du}{\int_0^{Y_{u(i)}} n(u)du} \leq \frac{\int_0^{Y_{u(k)}} \lambda_0(u)n(u)du}{\int_0^{Y_{u(k)}} n(u)du}$$

which is equivalent to

$$\frac{\int_0^{Y_{u(i)}} \lambda_0(u)n(u)du}{\int_0^{Y_{u(k)}} \lambda_0(u)n(u)du} \leq \frac{\int_0^{Y_{u(i)}} n(u)du}{\int_0^{Y_{u(k)}} n(u)du}$$

i.e.,

$$\frac{\sum_{i=1}^{k-1}\int_0^{Y_{u(i)}} \lambda_0(u)n(u)du}{\int_0^{Y_{u(k)}} \lambda_0(u)n(u)du} \leq \frac{\sum_{i=1}^{k-1}\int_0^{Y_{u(i)}} n(u)du}{\int_0^{Y_{u(k)}} n(u)du}$$

Since by Theorem 2.3.1,

$$\sum_{i=1}^{k-1} \frac{\sum_{j=1}^{i} U_j}{\sum_{j=1}^{k-1} U_j} \stackrel{\text{d}}{=} \frac{\sum_{i=1}^{k-1}\int_0^{Y_{u(i)}} \lambda_0(u)n(u)du}{\int_0^{Y_{u(k)}} \lambda_0(u)n(u)du}$$

the proof is complete.

We reject the null hypothesis that $\lambda_0(t)$ is constant if the value of V_k exceeds the critical number given in Table 2.1. In practice, we must replace β by the maximum likelihood estimator of β from marginal likelihood, to obtain an asymptotically valid the test.

Example. Consider a data set generated by using the following proportional hazard model,

$$\lambda_0(t; z) = 2t \exp(2z),$$

where z is given under H_0, covariate values. (See Section 6.2.) Suppose we have two groups of patients, e.g., male and female. We are interested in testing whether or not the hazard rate is increasing, regardless of sex. The data set generated is shown below.

Male :	0.0248	0.0361*	0.0452*	0.0821	0.125*	0.1489*
	0.1596	0.2008*	0.2017	0.2144*	0.2352*	0.2469*
	0.2749	0.3017	0.3045*	0.3164*	0.3189	0.3797*
	0.4907	0.6531				
Female :	0.0511	0.0611	0.0625	0.0766	0.0768	0.2216
	0.2294	0.2604	0.2984	0.3149	0.3662	0.5187
	0.7587	0.7906	0.8859	0.9317	0.9442	1.0004
	1.0810	1.5018				

* indicates censored observations.

We would like to see if we obtain the same result using $\hat{\beta} = 1.932$ as the result using $\beta = 2.0$. When $\beta = 2.0$, $V_k = 17.79$, while when $\beta = 1.932$, $V_k = 17.70$. We obtain z=2.48 by standardizing V_k. Hence, we make the same conclusion that the baseline hazard rate is increasing at the 5% significance level. Therefore, this example supports the validity of replacement of β with $\hat{\beta}$ obtained from the marginal likelihood to perform the test.

2.4 A graphical Method to Test Isotonicity with Covariates

We intend to develop a graphical procedure which allows us to visually examine whether or not the baseline hazard function is constant.

When the baseline hazard is constant, the corresponding failure time distribution , F_0, is an exponential distribution. Epstein(1960) introduced a graphical procedure for checking to see if the underlying distribution is really exponential. He plots $y = \log(1/(1-\hat{F}(t))$ against t where the cumulative distribution $F(t)$ is

$$F(t) = \begin{cases} 0 & t < 0 \\ 1 - \exp(-\frac{t}{\theta}) & t \geq 0 \end{cases}$$

assuming that $\theta > 0$.

If the failure rate is increasing, it is not difficult to see that $-\log(1 - F(t))$ is convex on $(0, \infty)$. To extend the idea suggested by Epstein to test for the baseline hazard rate in Cox's regression model, we estimate the survival function of failure time T, given $Z = z$.

Turning to our problem, note that survival function of the failure time T given $Z = z$, is given as

$$S(t; z) = S_0(t)^{\exp(z\beta)}$$

where $S_0(t)$ is an arbitrary survival function. First we consider the calculation of the nonparametric maximum likelihood estimate of $S_0(t)$ (Kalbfleish & Prentice, 1980).

The probability that an individual with covariate z fails at $y_{u(i)}$ is

$$S_0(y_{u(i)})^{\exp(z_{(i)}\beta)} - S_0(y_{u(i)} + 0)^{\exp(z_{(i)}\beta)}$$

where $S_0(y_{u(i)} + 0)$ is a right limit of $y_{u(i)}$. We assume that the contribution to the likelihood of a survival time censored at t is

$$S_0(t+0)^{\exp(z\beta)}.$$

In effect, the observed censoring time, t, tells us only that the observed failure time is greater than t. Thus we obtain the likelihood function

$$\mathcal{L} = \prod_{i=0}^{k} [\{S_0(y_{u(i)})^{\exp(z_{(i)}\beta)} - S_0(y_{u(i)}+0)^{\exp(z_{(i)}\beta)}\} \prod_{j=1}^{m_i} S_0(y_{c(ij)})^{\exp(z_{(ij)}\beta)}]. \quad (2.8)$$

It is clear that

$$S_0(t) = S_0(y_{u(i)})$$

for $y_{u(i)} < t \leq y_{u(i+1)}$, in order not to make $\mathcal{L} = 0$. In other words, the solution is a discrete distribution.

Let $1-\alpha_i$ be the hazard rate at $y_{u(i)}$, i.e.,

$$1-\alpha_i = \Pr(T = y_{u(i)} | T \geq y_{u(i)}) \qquad i = 1, \cdots, k.$$

Then we have

$$\begin{aligned} \Pr(T \geq y_{u(i)}) &= S_0(y_{u(i)}) \\ &= S_0(y_{u(i-1)}+0) \\ &= \prod_{j=0}^{i-1} \alpha_j \qquad i = 1, \cdots, k \end{aligned} \quad (2.9)$$

where $\alpha_0 = 1$. Note that the LHS of (2.9) is equivalent to

$$\exp[-\int_0^{y_{u(i)}} \lambda_0(u)du]$$

by definition of survival function $S_0(u)$. That is, we have

$$\exp[-\int_0^{y_{u(i)}} \lambda_0(u)du] = \prod_{j=0}^{i-1} \alpha_j \qquad\qquad i = 1, \cdots, k$$

Taking logarithm of both sides, we obtain

$$-\sum_{j=0}^{i-1} \log \alpha_j = \int_0^{y_{u(i)}} \lambda_0(u)du.$$

When $\lambda_0(u)$ is constant, we can see that $\sum_{j=0}^{i-1} \log \alpha_j$ is a linear function of $y_{u(i)}$, while it is a convex function of $y_{u(i)}$ when the λ_0 is increasing.

Since the α_j's are unknown, as an asymptotic approximation we set $\alpha_j = \hat{\alpha}_j$, the maximum likelihood estimators of the α_j's, $j = 1, \cdots .i-1$. In order to obtain the maximum likelihood estimators of the α_j's, the likelihood function (2.8) can be rewritten as

$$\begin{aligned}
\mathcal{L} &= \prod_{i=1}^{k}[\{(\prod_{j=0}^{i-1}\alpha_j)^{\exp(z_{(i)}\beta)} - (\prod_{j=0}^{i}\alpha_j)^{\exp(z_{(i)}\beta)}\}\prod_{l=1}^{m_i}(\prod_{j=0}^{i}\alpha_j)^{\exp(z_{(il)}\beta)}] \\
&= \prod_{i=1}^{k}\{(\prod_{j=0}^{i-1}\alpha_j)^{\exp(z_{(i)}\beta)}(1-\alpha_i{}^{\exp(z_{(i)}\beta)})\prod_{l=1}^{m_i}(\prod_{j=0}^{i}\alpha_j)^{\exp(z_{(il)}\beta)}\} \\
&= \prod_{i=1}^{k}(1-\alpha_i{}^{\exp(z_{(i)}\beta)})\prod_{i=1}^{k}\{(\prod_{j=0}^{i}\alpha_i)^{\{\exp(z_{(i)}\beta)+\sum_{l=1}^{m_i}\exp(z_{(il)}\beta)\}}\} \Big/ \prod_{i=1}^{k}\alpha_i{}^{\exp(z_{(i)}\beta)} \\
&= \prod_{i=1}^{k}(1-\alpha_i{}^{\exp(z_{(i)}\beta)})\prod_{j=1}^{k}\{\alpha_j{}^{\sum_{i=j}^{k}\exp(z_{(i)}\beta)+\sum_{l=1}^{m_i}\exp(z_{(il)}\beta)}\}\prod_{i=1}^{k}\alpha_i{}^{\exp(z_{(i)}\beta)} \\
&= \prod_{i=1}^{k}(1-\alpha_i{}^{\exp(z_{(i)}\beta)})\prod_{j=1}^{k}\{\alpha_j{}^{\sum_{i=j}^{k}\{\exp(z_{(i)}\beta)+\sum_{l=1}^{m_i}\exp(z_{(il)}\beta)\}-\exp(z_{(j)}\beta)}\}
\end{aligned}$$

$$= \prod_{i=1}^{k}(1-\alpha_i^{\exp(z_{(i)}\beta)})\alpha_i^{\sum_{R(y_{u(i)})}\exp(z_{(l)}\beta)-\exp(z_{(i)}\beta)}, \tag{2.10}$$

by using (2.9). We now replace β with $\hat{\beta}$ which is estimated from its marginal likelihood and then maximize (2.10) with respect to $\alpha_1, \cdots, \alpha_k$. Differentiating the logarithm of (2.10) with respect to α_i, we obtain the normal equation,

$$\frac{-\exp(z_{(i)}\beta)\hat{\alpha}_i^{\exp(z_{(i)}\beta)}}{1-\hat{\alpha}_i^{\exp(z_{(i)}\beta)}} + \sum_{R(y_{u(i)})} \exp(z_{(l)}\beta) - \exp(z_{(i)}\beta) = 0. \tag{2.11}$$

We can obtain the maximum likelihood estimate of α_i as a solution to (2.11), i.e.,

$$\hat{\alpha}_i = (1 - \frac{\exp(z_{(i)}\beta)}{\sum_{l\in R(y_{u(i)})}\exp(z_l\beta)})^{\exp(-z_{(i)}\beta)}.$$

Note that an iteration method is required to obtain the maximum likelihood estimators of α_i when we have multiple individuals falling at $y_{(i)}$.

In practice, we simply plot $-\sum_{j=0}^{i-1}\log\hat{\alpha}_j$ against $y_{u(i)}$, $i = 1, \cdots, k$, to apply the graphical method. If the baseline hazard function is constant, then the plot should be roughly linear while if the baseline hazard rate is monotone increasing, then the plot will tend to be a convex curve.

CHAPTER 3
ESTIMATION OF THE BASELINE HAZARD FUNCTION

3.1 Notation.

In the remainder of this dissertation we assume that the baseline hazard function, $\lambda_0(t)$, in Cox's regression model increases monotonically. Breslow(1974) approximates the baseline hazard function by a step function with discontinuities at each observed failure time. However the maximum likelihood estimates of the baseline hazard function are inefficient, since they do not take into account the monotonicity of $\lambda_0(t)$.

Let n be the number of individuals in the sample. We shall define a random variable, representing failure time as T where observed failure time is t. Let Z be a row vector of s measured covariates. Let $T_1, \cdots, T_n$ and $C_1, \cdots, C_n$ be the independent random variables of failure times and censoring times, respectively. We observe $(Y_1, \delta_1), \cdots, (Y_n, \delta_n)$ where

$$Y_i = T_i \wedge C_i, \qquad \delta_i = I(T_i \leq C_i)$$

with corresponding covariates $Z_1, \cdots, Z_n$. Define $Y_{u(i)}$ to be the ith order statistic from uncensored failure times for $i = 1, \ldots, k$ $(k \leq n)$. Let $Z_{u(i)}$ represent the corresponding covariate of $Y_{u(i)}$.

As defined in (1.1), the proportional hazard model is given by

$$\lambda(t; z) = \lambda_0(t) \exp(z\beta).$$

The survival function is defined by

$$S(t; z) = \exp[-\int_0^t \lambda_0(u) \exp(z\beta) du]$$

and the density function by

$$f(t;z) = \lambda(t;z)S(t;z)$$

given that $Z = z$.

Breslow(1974) obtains the maximum likelihood estimator of $\lambda_0(t)$ without the assumption of increasing failure rate. In this chapter we derive the maximum likelihood estimator of $\lambda_0(t)$ under the assumption of increasing failure rate. We approximate the baseline hazard function by a nondecreasing step function with discontinuities at each observed failure time; that is,

$$\lambda_0(t) = \lambda_i \qquad y_{u(i-1)} < t \leq y_{u(i)} \tag{3.1}$$

where $\lambda_i \leq \lambda_{i+1}$ for $i = 1, \cdots, k,$, $\lambda_0 = 0$, $y_{u(0)} = 0$, and $y_{u(k+1)} = \infty$. For estimation purposes, we assume that an individual censored in the interval $[y_{u(i-1)}, y_{u(i)})$ is censored at $y_{u(i-1)}$. This approach is similar to Breslow(1974).

The likelihood corresponding to the observations described at the beginning of this section is

$$\mathcal{L}_0 = \prod_{i=1}^{k}[\lambda_0^{d_i}(y_{u(i)})\exp(s_{u(i)}\beta)\exp\{-\int_0^{y_{u(i)}}\lambda_0(u)du \sum_{l\in H(y_{u(i)})}\exp(z_l\beta)\}]$$

where d_i is the number of individuals who failed at $y_{u(i)}$ and $s_{u(i)}$ is the sum of covariates of individuals failing at $y_{u(i)}$. $H(y_{u(i)})$ is the set of labels attached to the individuals who either fail or are censored observations at $y_{u(i)}$. particular way, for example,

Using (3.1), $\mathcal{L}_0$ reduces to

$$\mathcal{L}_1 = \prod_{i=1}^{k}[\lambda_i^{d_i}\exp(s_{u(i)}\beta)\exp\{-\lambda_i(y_{u(i)} - y_{u(i-1)})\sum_{l\in R(y_{u(i)})}\exp(z_l\beta)\}]$$

where $R(y_{u(i)})$ is the risk set prior to $y_{u(i)}$.

To maximize $\mathcal{L}_1$, we consider the logarithm of $\mathcal{L}_1$, denoted by $\mathcal{L}_2$,

$$\mathcal{L}_2 = \sum_{i=1}^{k}\{d_i \log \lambda_i + s_{u(i)}\beta - \lambda_i(y_{u(i)} - y_{u(i-1)}) \sum_{l\in R(y_{u(i)})} \exp(z_l\beta)\}. \qquad (3.2)$$

In the remainder of this chapter, the term "maximum likelihood estimator" of $\lambda_0(t)$ will be referred to the solution $\hat{\lambda}_i, i = 1, \cdots, k$, which maximizes $\mathcal{L}_2$ subject to $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_k$. In next section, we apply the isotonic regression method to find the maximum likelihood estimator of $\lambda_0(t)$. In the following chapter, we shall replace β, by the the marginal likelihood estimator of β to find maximum likelihood estimator of $\lambda_0(t)$.

3.2 Isotonic Regression

We shall introduce isotonic regression through an example. The importance of isotonic regression can be illustrated by the classical two sample problem, in which the mean response for two treatments are compared. If treatment 1 is a standard crop treatment, while treatment 2 is an experimental treatment (standard plus fertilizer), then it may be possible to assert that $\mu_2 \geq \mu_1$, where μ_i is the population mean amount of produce per acre with applying treatment i ($i = 1, 2$). Suppose we want to test the hypothesis $\mu_1 = \mu_2$. Since significance comes from an observed difference, say $\bar{y_1} - \bar{y_2}$, we perform a one-sided test. If we ignore the ordering assumption, then we would use the standard two-sample t test for the equality of two treatments. It is well known that the one-sided test is significantly more powerful than the two sided test. Since the two-sided test makes no use of the prior information that $\mu_1 \leq \mu_2$. By taking such information into account, the one-sided test gives higher power to test the difference between the treatments. Isotonic regression analysis provides a method of using the ordering assumption about the parameters.

We need the following definitions to introduce the isotonic regression.

Let X be the finite set $\{x_1, x_2, \cdots, x_k\}$ with the simple order $x_1 \leq x_2 \leq \cdots \leq x_k$ and ω be a positive weight function. A function f, on X, is *isotonic* if $f(x_1) \leq f(x_2) \leq \cdots \leq f(x_k)$

Suppose g is a given function on X. A function g^* on X is an *isotonic regression* of g with weights ω if and only if g^* is isotonic and g^* minimizes

$$\sum_{x \in X} [g(x) - f(x)]^2 \omega(x) \tag{3.3}$$

in the class of all isotonic functions f on X.

Robertson et al. (1988) in the following theorem argues that for a finite set of X, 'isotonic' estimators reduce error in the sense by the following theorem.

Theorem 3.2.1 Suppose we have a quasi-order on a finite set of X. If $\hat{\theta}$ is any function on X and if $\hat{\theta}^*$ is the isotonic estimator of θ with weights ω, then

$$\sum_{x \in X} \Psi[\hat{\theta}^*(x) - \theta(x)]\omega(x) \leq \sum_{x \in X} \Psi[\hat{\theta}(x) - \theta(x)]\omega(x)$$

for any convex function Ψ on $(-\infty, \infty)$ and any isotonic function θ on X.

Proof of Theorem 3.2.1 See Robertson et al. (1988, p41).

Theorem 3.2.1 states that the isotonic estimator $\hat{\lambda}_0^*$ of λ_0 reduces error in a number of ways as seen by taking $\Psi(t) = |t|^p$, $p \geq 1$. For example, with $p = 1$, we can see that $\hat{\lambda}_0^*$ has less total absolute error and with $p = \infty$, $\hat{\lambda}_0^*$ has less maximum absolute error.

Let us consider a graphical interpretation of (3.3).

In Figure 3.1, $(g-f)^2$ can be interpreted as the excess of the rise in the graph of the function $\Psi(u)=u^2$ from f to g over the rise of its tangent line at f. Clearly, we see

$$d=(g-f)^2=g^2-[f^2+(g-f)2f]=\Psi(g)-\Psi(f)-(g-f)\psi(f)$$

where $\psi(f)$ is the derivative of $\Psi(u)$ at f. This excess is nonnegative for every convex function $\Psi(u)$, whether $f<g$ or $g<f$. It is strictly positive if $\Psi(u)$ is strictly convex and $f\neq g$.

Let us generalize the square error measure (3.3), replacing $\Psi(u)=u^2$ by any convex function. Let $\Psi(u)$ be a convex function, which is finite on an open interval I containing the range of function g and infinite elsewhere. Denote the discrepancy of $\Psi(f-g)$ by

$$\Delta_\Psi(g(x),f(x))=\begin{cases}\Psi(g)-\Psi(f)-(g-f)\psi(f) & f(x),g(x)\in I\\ \infty & \text{otherwise,}\end{cases}\tag{3.4}$$

where $\psi(f)$ is the derivative of $\Psi(u)$ at f. If $\Psi(u)$ does not have a derivative at f, then $\psi(f)$ denotes any number between the left and right derivative at f. From (3.4), it can be seen that

$$\Delta_\Psi(r,t)=\Delta_\Psi(r,s)+\Delta_\Psi(s,t)+(r-s)[\psi(s)-\psi(t)]$$

if r,s and t are in the domain of Ψ.

<u>*Theorem 3.2.2*</u> *Let Ψ be a convex function which is finite in an open interval I, containing the range of the function g and infinite elsewhere. If g^* is the isotonic regression of g, f is isotonic on X, and the range of f is contained in I, then*

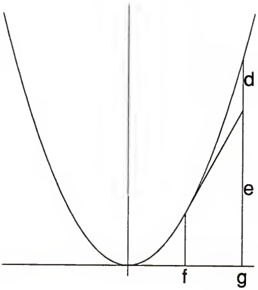

Figure 3.1. Graphical interpretation of the square error measure of discrepancy The graph of $\Phi(u) = u^2$ is drawn.

$$\sum_x \Delta_\Psi[g(x), f(x)]\omega(x) \geq \sum_x \Delta_\Psi[g(x), g^*(x)]\omega(x) + \sum_x \Delta_\Psi[g^*(x), f(x)]\omega(x).$$

Consequently g^* minimizes

$$\sum_x \Delta_\Psi[g(x), f(x)]\omega(x)$$

in the class of isotonic f with range in I and maximizes

$$\sum_x \{\Psi[f(x)] + [g(x) - f(x)]\psi[f(x)]\}\omega(x). \tag{3.5}$$

The maximizing (minimizing) function is unique if $\Psi(x)$ is strictly convex.

<u>Proof of Theorem 3.2.2</u> *See Barlow et al.(1972).*

<u>Corollary 3.2.1</u> *Let $\psi_1, \psi_2, \cdots, \psi_p$ be arbitrary real valued functions and let $h_1, h_2, \cdots, h_m$ be isotonic functions on X. Then g^* minimizes*

$$\sum_x \Delta_\Psi(g, f)\omega(x)$$

in the class of isotonic functions f with range in I satisfying any or all of the side conditions

$$\sum_x [g(x) - f(x)]\psi_j[f(x)]\omega(x) = 0 \qquad j = 1, 2, \cdots, p \tag{3.6}$$

$$\sum_x f(x)h_j(x)\omega(x) \geq \sum_x g(x)h_j(x)\omega(x) \qquad j = 1, 2, \cdots, m$$

Theorem 3.2.2 and Corollary 3.2.1 can be used to show that the isotonic regression provides a solution for a wide variety of estimation problems with order restriction in which the objective function is other than least squares. The effort for solving these problems is focused on finding the appropriate choice of the function $\Psi(x)$.

Example. (Barlow et al., 1972) Suppose that for each of various levels x of a stimulus (e.g., dose of insecticide) the probability of a response (e.g., death of the insect) is $\mu(x)$. We would like to estimate $\mu(x)$, known to be nondecreasing in x. If X is the finite set of stimulus levels, it is simply ordered by the dosage levels. Suppose for $x \in X$, there are $m(x)$ independent trials at stimulus level x, $a(x)$ responses occur, and $\bar{y}(x) = a(x)/m(x)$ is the average number of responses per trial.

If $m(x)$ is large for x, the ratios $\bar{y}(x)$ can be expected to be in increasing order that are natural estimates of the probability $\mu(x)$. But if some consecutive ratios $\bar{y}(x)$ have reversed ordering, another estimator would be required. The isotonic regression of $\bar{y}$ with weights $m(x)$ is an obvious candidate.

Let $b(x) = m(x) - a(x)$ denote the number of nonresponses among $m(x)$ trials at stimulus level x. If $f(x)$ denotes an arbitrary function bounded between 0 and 1 on X, the likelihood at f of the sample is

$$\prod_{x \in X} [f(x)]^{a(x)} [1 - f(x)]^{b(x)}$$

and the negative log-likelihood can be written as

$$-\sum_{x \in X} \Big\{ \bar{y}(x) \log f(x) + [1 - \bar{y}(x)] \log[1 - f(x)] \Big\} m(x) \tag{3.7}$$

Thus the solution of the problem of maximum likelihood estimation of μ is the function that minimizes (3.7) over the class of isotonic functions on X.

Consider the following convex function.

$$\begin{aligned}\Psi(u) &= u\log u + (1-u)\log(1-u), \qquad o < u < 1 \\ \Psi(0) &= 0, \\ \Psi(1) &= 0.\end{aligned} \tag{3.8}$$

Then,

$$\Delta(f,g) = g\log g + (1-g)\log(1-g) - g\log f + (1-g)\log(1-f). \tag{3.9}$$

Noting that the first two terms on the right of (3.9) do not involve f, the problem of finding maximum likelihood estimator of $\mu(x)$ is equivalent to finding f which minimizes the discrepancy determined by convex function (3.8). Theorem 3.2.2 states that $\bar{y}^*(x)$, the isotonic regression of $\bar{y}(x)$ minimizes

$$\sum_{x\in X}\Big\{\bar{y}(x)\log\bar{y}(x) + [1-\bar{y}(x)]\log[1-\bar{y}(x)]$$
$$-\bar{y}(x)\log f(x) + [1-\bar{y}(x)]\log[1-f(x)]\Big\}m(x),$$

or $\sum_{x\in X}\Delta(\bar{y}(x), f(x))m(x)$.

Recall that our problem is to maximize (3.2)

$$\mathcal{L}_2 = \sum_{i=1}^{k}\{\log\lambda_i + S_{u(i)}\beta - \lambda_i(y_{u(i)} - y_{u(i-1)})\sum_{l\in R(y_{u(i)})}\exp(z_l\beta)\}, \tag{3.10}$$

subject to $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_k$, assuming that all failures are distinct, i.e., $d_i = 1$. To obtain the maximum likelihood estimator of λ_i, $i = 1, \cdots, k$, by applying the previous theorem and corollary, we define $\Psi(u) = u \log u$. Then we obtain

$$\Delta_\Psi(g, f) = g \log g - g \log f - (g - f).$$

Since $g \log g$ does not depend on f, Theorem 3.2.2 implies that the isotonic regression g^* of g maximizes

$$\sum_x [g(x) \log f(x) + g(x) - f(x)]\omega(x), \tag{3.11}$$

in the class of positive isotonic functions. By Corollary 3.2.1, with $\psi = 1$, g^* also maximizes

$$\sum_x [g(x) \log f(x) - f(x)]\omega(x), \tag{3.12}$$

in the class of positive isotonic functions f satisfying

$$\sum_x [g(x) - f(x)]\omega(x) = 0. \tag{3.13}$$

Since $s_{u(i)}$ is independent of λ_i, the problem of (3.10) is equivalent to maximizing

$$\sum_{i=1}^{k} \left(\frac{1}{(y_{u(i)} - y_{u(i-1)}) \sum_{l \in R(y_{u(i)})} \exp(z_l\beta)} \log \lambda_i \quad - \quad \lambda_i \right) (y_{u(i)} - y_{u(i-1)}) \sum_{l \in R(y_{u(i)})} \exp(z_l\beta) \tag{3.14}$$

subject to $\lambda_1 \le \lambda_2 \le \cdots \le \lambda_k$.

Let $X = \{1, 2, \cdots, k\}$, $f(i) = \lambda_i$,

$$g(i) = \frac{1}{(y_{u(i)} - y_{u(i-1)}) \sum_{l \in R(y_{u(i)})} \exp(z_l \beta)}$$

and

$$\omega(i) = (y_{u(i)} - y_{u(i-1)}) \sum_{l \in R(y_{u(i)})} \exp(z_l \beta)$$

$i = 1, \cdots, k$. by substitution, we can see that the expression (3.12) is same as (3.14).

It can be seen that the solution $\hat{\lambda}_1, \hat{\lambda}_2, \cdots, \hat{\lambda}_k$, that maximizes (3.14) satisfies

$$\sum_{i=1}^{k} \Big(\frac{1}{(y_{u(i)} - y_{u(i-1)}) \sum_{l \in R(y_{u(i)})} \exp(z_l \beta)} - \hat{\lambda}_i \Big) (y_{u(i)} - y_{u(i-1)}) \sum_{l \in R(y_{u(i)})} \exp(z_l \beta) = 0. \quad (3.15)$$

If $\hat{\lambda}_1 \leq \hat{\lambda}_2 \leq \cdots \leq \hat{\lambda}_k$, and $\rho > 0$, then we have $\rho\hat{\lambda}_1 \leq \rho\hat{\lambda}_2 \leq \cdots \leq \rho\hat{\lambda}_k$. It is easily seen that

$$\sum_{i=1}^{k} \Big(\frac{1}{(y_{u(i)} - y_{u(i-1)}) \sum_{l \in R(y_{u(i)})} \exp(z_l \beta)} \log \rho\hat{\lambda}_i - \rho\hat{\lambda}_i \Big) (y_{u(i)} - y_{u(i-1)}) \sum_{l \in R(y_{u(i)})} \exp(z_l \beta) \quad (3.16)$$

achieves its maximum as a function of ρ at $\rho = 1$. Substituting 1 for ρ in (3.16) yields (3.14). On setting its derivative at $\rho = 1$ to zero, we obtain (3.15), which implies the the solution $\hat{\lambda}, \hat{\lambda}_2, \cdots, \hat{\lambda}_k$, satisfies the condition (3.13).

We can derive that the isotonic regression g^* is the maximum likelihood estimator of λ_i, $i = 1, \cdots, k$, where $\lambda_0(t) = \lambda_i$ for $y_{u(i-1)} \leq t < y_{u(i)}$, since g^* maximizes (3.12), subject to (3.13) and hence it also maximizes (3.14) subject to (3.15) and $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_k$.

Now we consider the graphical interpretation of the isotonic regression g^*. Assuming the ordering $x_1 < x_2 < \cdots < x_k$, plot the points $P_j = (W_j, G_j)$, in the Cartesian plane where

$$G_j = \sum_{i=1}^{j} g(x_i)\omega(x_i)$$

and

$$W_j = \sum_{i=1}^{j} \omega(x_i). \qquad \text{for } j = 1, \cdots, k$$

Let $P_0 = (0,0)$. These points form the cumulative sum diagram (CSD) of the given function g with weight ω. The slope of the chord joining P_{i-1} to $P_j\,(i \leq j)$ represents the weighted average

$$Av\{x_i, x_{i+1}, \cdots, x_j\} = \sum_{r=i}^{j} g(x_r)\omega(x_r) / \sum_{r=i}^{j} \omega(x_r).$$

It is clear to see that $g(x_j)$, $j = 1, \cdots, k$, is the slope of segment joining P_{j-1} to P_j.

It is well known that the greatest convex minorant (GCM) of the CSD is the graph of the supremum of all convex functions whose graphs lie below the CSD. Let us next consider the graphical method for GCM. First draw a line for which the entire CSD lies on or above it. If it intersects in more than one point, then the segment joining its leftmost and rightmost intersections becomes a part of the graph of GCM. The GCM is made up of such segments. Graphically the GCM is the path along which a taut string lies if it joins P_0 and P_k and is constrained to lie below the CSD. The value of the isotonic regression g^* at a point x_j is just the slope of the GCM at the points P_j^* with abscissa

$$\sum_{i=1}^{j} \omega(x_i).$$

Table 3.1 above and Figure 3.2 below clarify the concepts.

Table 3.1. Example of CSD and GCM

j	$\omega(x_j)$	W_j	$g(x_j)$	G_j	G_j^*	g_j^*
1	1	1	-2	-2	-2	-2
2	2	3	5/2	3	-8/5	1/5
3	3	6	-4/3	-1	-1	1/5
4	2	8	1	1	1	1

Remark: $W_j = \sum_{i=1}^{j} \omega(x_j)$ $G_j = \sum_{i=1}^{j} g(x_j)\omega(x_j)$ $G_j^* = \sum_{i=1}^{j} g^*(x_j)\omega(x_j)$ $j = 1, 2, 3, 4.$

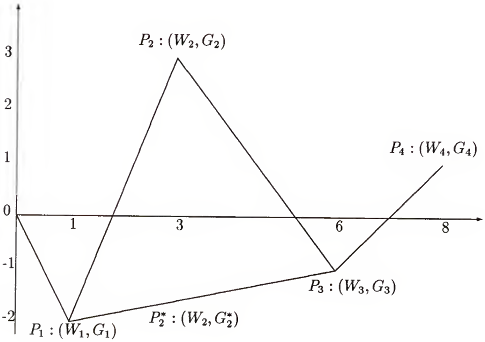

Figure 3.2. Example of CSD and GCM
Slope at P of CSD: $G_j - G_{j-1}/W_j - W_{j-1} = g(x_j)$;
Slope at P_j^* of GCM: $G_j^* - G_{j-1}^*/W_j - W_{j-1} = g(x_j), j = 1, 2, 3, 4$

By Utilizing the fact that the isotonic regression can be represented graphically as the slope of the GCM (Barlow et al., 1972), we obtain the formula of the isotonic regression of g,

$$g^*(x_i) = \max_{s\le i} \min_{t\ge i} Av(s,t)$$

where

$$Av(s,t) = \sum_{r=s}^{t} g(x_r)\omega(x_r) \Big/ \sum_{r=s}^{t} \omega(x_r).$$

Since we have

$$g = \frac{1}{(y_{u(i)} - y_{u(i-1)}) \sum_{l\in R(y_{u(i)})} \exp(z_l\beta)}$$

and

$$\omega = (y_{u(i)} - y_{u(i-1)}) \sum_{l\in R(y_{u(i)})} \exp(z_l\beta),$$

we obtain the isotonic regression, which is equivalent to the maximum likelihood estimate of λ_i,

$$\hat{\lambda}_i^* = \max_{s\le i} \min_{t\ge i} \frac{t-s+1}{\sum_s^t (y_{u(j)} - y_{u(j-1)}) \sum_{l\in R(y_{u(j)})} \exp(z_l\beta)}.$$

The function g and ω are dependent upon the unknown regression parameters, β, which must be replaced by a constant estimator. Typically, we use the value $\hat{\beta}$ which is obtained from the marginal likelihood by a Newton-Raphson iteration.

CHAPTER 4
CONSISTENCY OF THE ISOTONIC ESTIMATOR

4.1 Notation and Assumption

In this chapter, we shall prove the strong consistency of the isotonic regression estimator of the baseline hazard rate, for $x_0 \in (y_{u(i)}, y_{u(i+1)}]$, $i = 0, 1, \cdots, k$.

$$\hat{\lambda}_i(x_0) = \max_{s \le i} \min_{t \ge i} \frac{t - s + 1}{\sum_s^t (y_{u(j)} - y_{u(j-1)}) \sum_{l \in R(y_{u(j)})} \exp(z_l \hat{\beta})} \tag{4.1}$$

where $\hat{\beta}$ is the estimator of the regression parameter by marginal likelihood. For simplicity, we shall assume all failures are distint and that the covariate z is single valued.

We shall use the notation and formulas as defined by Tsiatis (1981).

Let the covariate Z be a random variable with density $q(z)$ and distribution $Q(z)$. We assume that $Q(z)$ has compact support, i.e., there exists z_0 such that $\Pr(0 \le Z \le Z_0) = 1$. The Cox regression model links the distribution of the failure time to the covariates Z. We assume that T_0 is the time when the study ends. So $\Pr(T \le T_0) = 1$. Let $\mu(t|z)$ denote the hazard function of the censored distributions, given $Z = z$. It follows that the conditional probability of surviving until time t without being censored, given that $Z = z$, is given by

$$\begin{aligned} H(t|z) &= \Pr(T \ge t|z) \\ &= \exp[-\int_0^t \{\lambda_0(x)\exp(z\beta) + \mu(x|z)\}dx]. \end{aligned} \tag{4.2}$$

The conditional probability of surviving until t without being censored and eventually dying before being censored, given $Z = z$, is

$$\begin{aligned} F(t|z) &= \Pr(T \geq t, \delta = 1|z) \\ &= \int_t^{T_0} \{\lambda_0(x)\exp(z\beta)H(x|z)\}dx, \end{aligned} \tag{4.3}$$

Furthermore the probability of surviving until time t without being censored and eventually dying before being censored is

$$\begin{aligned} F(t) &= \Pr(T \geq t, \delta = 1) \\ &= \int F(t|z)q(z)dz \\ &= \int\int_t^{T_0} \lambda_0(x)\exp(z\beta)H(x|z)dxq(z)dz. \end{aligned} \tag{4.4}$$

We assume that $F(t)$ is continuously differentiable and has an inverse function. The derivative of $F(t)$ is

$$\begin{aligned} \frac{dF(t)}{dt} &= -\int \lambda_0(t)\exp(z\beta)H(t|z)dQ(z) \\ &= -\lambda_0(t)\int \exp(z\beta)H(t|z)dQ(z). \end{aligned} \tag{4.5}$$

For $g(z)$, a continuous function of z, define

$$E(g(z),t) = E[g(Z)I_{[T\geq t]}]$$

$$\begin{aligned} &= E[g(Z)P(T \geq t|Z)] \\ &= \int g(z)H(t|z)dQ(z) \end{aligned} \tag{4.6}$$

and

$$\begin{aligned} E_1(g(z),t) &= E[g(Z)I_{[T\geq t,\delta=1]}] \\ &= E[g(Z)P(T \geq t, \delta = 1|Z)] \\ &= E[g(Z)F(t|Z)] \\ &= \int \int_t^{T_0} g(z)\lambda_0(x)\exp(z\beta)H(x|z)dxdQ(z). \end{aligned}$$

By differentiation of $E_1(g(z),t)$ with respect to t, we obtain

$$\frac{d}{dt}E_1[g(z),t] = -\lambda_0(t)\int g(z)\exp(z\beta)H(t|z)dQ(z).$$

Using (4.5) and (4.6), we further obtain

$$\lambda_0(t) = -\frac{dF(t)}{dt}\Big/E(\exp(z\beta),t). \tag{4.7}$$

We can define the usual empirical estimates of $F(x)$ and $E(\exp(z\beta),x)$ by

$$F_n(x) = \frac{1}{n}\sum_{i=1}^{n} I_{[T_i\geq x,\delta_i=1]}$$

and

$$E_n(\exp(z\beta), x) = \frac{1}{n} \sum_{i \in R(x)} \exp(z_i\beta) = \int_{\{z:T>x\}} \exp(z\beta) dQ_n(z) \quad (4.8)$$

respectively where Q_n is the empirical distribution of Z. Further we define that

$$E_n(\exp(z\hat{\beta}), x) = \frac{1}{n} \sum_{i=1}^{n} \exp(z_i\hat{\beta}) I_{[T_i \geq x]}$$

and

$$K_\xi(x) = \int_\xi^x E(\exp(z\beta), u) du.$$

4.2 Consistency of Isotonic Estimator of $\lambda_0(t)$

The objective of this subsection is to show, for fixed x_0, $\hat{\lambda}_i(x_0)$ is a consistent estimator of $\lambda_0(x_0)$ where

$$\hat{\lambda}_i(x_0) = \max_{s \leq i} \min_{t \geq i} \frac{t - s + 1}{\sum_s^t (y_{u(j)} - y_{u(j-1)}) \sum_{l \in R(y_{u(j)})} \exp(z_l \hat{\beta})}, \quad (4.9)$$

for $x_0 \in (y_{u(i)}, y_{u(i+1)}]$. We can rewrite (4.1) for $x_0 \in (y_{u(i)}, y_{u(i+1)}]$,

$$\hat{\lambda}_i(x_0) = \max_{s \leq i} \min_{t \geq i} \frac{F_n(y_{u(s-1)}) - F_n(y_{u(t)})}{\hat{K}_\xi(y_{u(t)}) - \hat{K}_\xi(y_{u(s-1)})} \quad (4.10)$$

where

$$\hat{K}_\xi(x) = \int_\xi^x E_n(\exp(z\hat{\beta}), u)du$$

for $y_{u(1)} \le x < y_{u(n)}$ and $y_{u(1)} \le \xi < y_{u(n)}$.

Next we shall prove, for any fixed x_0, such that $0 < F(x_0) < 1$,

$$\lambda(x_0{}^-) \le \liminf_{n\to\infty} \hat{\lambda}_i(x_0) \le \limsup_{n\to\infty} \hat{\lambda}_i(x_0) \le \lambda(x_0{}^+), \tag{4.11}$$

where $x_0 \in (y_{u(i)}, y_{u(i+1)}]$ and i depends on the sample.

First we must prove in the following lemma and corollary

$$\sup_{0<\xi<T_0} |\hat{K}_\xi(x) - K_\xi(x)| \to 0 \qquad as \qquad n \to \infty. \tag{4.12}$$

The following lemma implies (4.12).

Lemma 4.2.1 Let $\hat{\beta}$ be the maximum likelihood estimator of β obtained from marginal likelihood. It follows that

$$\sup_{0<u<T_0} |E_n(\exp(z\hat{\beta}), u) - E(\exp(z\beta), u)| \to 0 \qquad \text{a.s.} \tag{4.13}$$

where $\Pr(T \le T_0) = 1$.

Proof of Lemma 4.2.1 Note that from the triangle inequality

$$\begin{aligned}
\sup_{0<u<T_0} & \mid E_n(\exp(z\hat{\beta}), u) - E(\exp(z\beta), u) \mid \\
&= \sup_{0<u<T_0} |E_n(\exp(z\hat{\beta}), u) - E_n(\exp(z\beta), u) + E_n(\exp(z\beta), u) - E(\exp(z\beta), u)| \\
&\le \sup_{0<u<T_0} \{|E_n(\exp(z\hat{\beta}), u) - E_n(\exp(z\beta), u)| \\
&\qquad + |E_n(\exp(z\beta), u) - E(\exp(z\beta), u)|\}.
\end{aligned} \tag{4.14}$$

Therefore it suffices to show that

$$\sup_{0<u<T_0} |E_n(\exp(z\hat{\beta}), u) - E_n(\exp(z\beta), u)| \to 0 \qquad \text{a.s.} \tag{4.15}$$

and

$$\sup_{0<u<T_0} |E_n(\exp(z\beta), u) - E(\exp(z\beta), u)| \to 0 \qquad \text{a.s.} \tag{4.16}$$

In terms of integral forms of $E_n(\exp(z\hat{\beta}), u)$ *and* $E_n(\exp(z\beta), u)$, *(4.15) can be rewritten as*

$$\begin{aligned}
\sup_{0<u<T_0} \quad & | \quad \int_{\{z:T\geq u\}} \exp(z\hat{\beta}) dQ_n(z) - \int_{\{z:T\geq u\}} \exp(z\beta) dQ_n(z)| \\
& = \sup_{0<u<T_0} | \int_{\{z:T\geq u\}} \{\exp(z\hat{\beta}) - \exp(z\beta)\} dQ_n(z)| \\
& \leq \sup_{0<u<T_0} \int_{\{z:T\geq u\}} |\exp(z\hat{\beta}) - \exp(z\beta)| dQ_n(z) \\
& \leq \int_0^{Z_0} |\exp(z\hat{\beta}) - \exp(z\beta)| dQ_n(z),
\end{aligned} \tag{4.17}$$

where Z_0 *and 0 are assumed to the upper bound and the low bound of the random variable* z *respectively. Since* $\hat{\beta}$ *converges almost surely to* β, $|\exp(z\hat{\beta}) - \exp(z\beta)|$ *is bounded and converges to* 0 *almost surely. Moreover* dQ_n *is a bounded measure. Hence, (4.15) converges almost surely to* 0.

Tsiatis (1981, Appendix 1) shows that the Glivenko-Cantelli lemma (Chow & Teicher, 1978, p 260) can be applied to prove that (4.16) converges to 0 *almost surely,* *since* $E_n(\exp(z\beta), u)$ *and* $E(\exp(z\beta), u)$ *are bounded by assumption and nondecreasing in u. This completes the proof of Lemma 4.2.1.*

Corollary 4.2.1 *Under the conditions of Lemma 4.2.1,*

$$\sup_{0<\xi<T_0} |\hat{K}_\xi(x) - K_\xi(x)| \to 0 \qquad as \qquad n \to \infty.$$

Theorem 4.2.1 *For* $\hat{\lambda}_i$ *defined by (4.10) and for every fixed* x_0 *with* $0 < F(x_0) < 1$,

$$\lambda(x_0^-) \le \liminf_{n\to\infty} \hat{\lambda}_i(x_0) \le \limsup_{n\to\infty} \hat{\lambda}_i(x_0) \le \lambda(x_0^+). \tag{4.18}$$

Proof of Theorem 4.2.1 *To prove the first part of (4.18), let* ξ *be an arbitrary point such that*

$$F^{-1}(1) < \xi < x_0.$$

converging to ξ *as It follows from (4.10) that*

$$\begin{aligned} \hat{\lambda}_i(x_0) &= \inf_{x_0 \le y_{u(t)}} \sup_{y_{u(s)} \le x_0} \frac{F_n(y_{u(s-1)}) - F_n(y_{u(t)})}{\hat{K}_\xi(y_{u(t)}) - \hat{K}_\xi(y_{u(s-1)})} \\ &\ge \inf_{x_0 \le x} \frac{F_n(\xi) - F_n(x)}{\hat{K}_\xi(x) - \hat{K}_\xi(\xi)}. \end{aligned} \tag{4.19}$$

Since F_n converges to F, Corollary 4.2.1 implies

$$\begin{aligned}
\liminf_{n\to\infty} \hat{\lambda}_i(x_0) &\geq \inf_{x_0 \leq x} \frac{F(\xi) - F(x)}{K_\xi(x)} \\
&\geq \inf_{\xi < x_0 \leq x} \frac{F(\xi) - F(x)}{F(\xi) - F(x)} \lambda_0(\xi) \\
&= \lambda_0(\xi) \qquad (4.20)
\end{aligned}$$

where the second inequality follows from (a) the monotonicity of λ_0 and F and (b)

$$\begin{aligned}
K_\xi(x) &= \int_\xi^x E(\exp(z\beta), u) du \\
&= -\int_\xi^x \frac{1}{\lambda_0(u)} dF(u) \\
&\leq \frac{1}{\lambda_0(\xi)} (F(\xi) - F(x)) \qquad (4.21)
\end{aligned}$$

Since ξ is an arbitrary number which is less than x_0, we have

$$\liminf_{n\to\infty} \hat{\lambda}_i(x_0) \geq \lambda_0(x_0{}^-) \qquad (4.22)$$

To prove the second part of (4.18), let ξ be an arbitrary point such that

$$x_0 < \xi < F^{-1}(0).$$

We have from (4.10),

$$\hat{\lambda}_i(x_0) = \inf_{x_0 \le y_{u(t)}} \sup_{y_{u(s)} \le x_0} \frac{F_n(y_{u(s-1)}) - F_n(y_{u(t)})}{\hat{K}_\xi(y_{u(t)}) - \hat{K}_\xi(y_{u(s-1)})}$$

$$\le \sup_{x \le x_0} \frac{F_n(x) - F_n(\xi)}{\hat{K}_\xi(\xi) - \hat{K}_\xi(x)} \quad (4.23)$$

Since F_n *converges to* F *as* $n \to \infty$, *Lemma 4.2.1 implies*

$$\limsup_{n\to\infty} \hat{\lambda}_i(x_0) \le \sup_{x \le x_0} \frac{F(x) - F(\xi)}{-K_\xi(x)}$$

$$\le \sup_{x \le x_0 < \xi} \frac{F(x) - F(\xi)}{F(x) - F(\xi)} \lambda_0(\xi)$$

$$= \lambda_0(\xi) \quad (4.24)$$

where the second inequality follows from (a) the monotonicity of λ_0 *and* F *and (b)*

$$-K_\xi(x) = \int_\xi^x -E(\exp(z\beta), u)du$$

$$= \int_\xi^x \frac{1}{\lambda_0(u)} dF(u)$$

$$\ge \frac{1}{\lambda_0(\xi)}(F(x) - F(\xi)).$$

Since ξ is an arbitrary number which is greater than x_0, we have

$$\limsup_{n\to\infty} \hat{\lambda}_i(x_0) \leq \lambda_0(x_0{}^+). \tag{4.25}$$

By combining (4.22) and (4.25), it follows that the isotonic estimator of the baseline hazard function is strongly consistent, provided that $\lambda_0(x)$ is continuous:

$$\lim_{n\to\infty} \hat{\lambda}_i(x_0) = \lambda_0(x_0) \qquad a.s.$$

where $\hat{\lambda}_i(x_0)$ is defined in (4.7).

CHAPTER 5
THE ISOTONIC ESTIMATOR BASED ON THE WINDOW

5.1 Introduction

In Chapter 3, we considered estimators of the baseline hazard rate based on order statistics. In practical situations, the observed data are frequently grouped into intervals. In other words, we are able to count only the number of failures and number of censored observations within specified intervals. Hence it is important to consider estimators corresponding to grouped data rather than maximum likelihood estimator of the baseline hazard rate based on order statistics. A general class of isotonized fixed and random "window" estimators of failure rate are proposed and discussed by Barlow and van Zwet (1969). They showed that the window estimators with appropriate window size have higher asymptotic efficiency than the unrestricted maximum likelihood estimator. Intuitively it seems reasonable that an improved estimator can be obtained by forcing the estimator to be monotone. In our case, we might expect that the estimator based on the window has the property that it is closer to $\lambda_0(t)$ than the maximum likelihood estimator, by the criterion of mean square error. We are also interested in determining an appropriate size of window to optimize the maximum likelihood estimator.

Note that the maximum likelihood estimator of $\lambda_0(t)$ based on the ordered sample $\{y_{u(1)} < y_{u(2)} < \cdots < y_{u(k)}\}$ is

$$\hat{\lambda}_0(x) = \frac{1}{(y_{u(i)} - y_{u(i-1)}) \sum_{l \in R(y_{u(i)})} \exp(z_l \hat{\beta})} \qquad i = 1, \cdots, k$$

for $y_{u(i-1)} < x \leq y_{u(i)}$.

For each n, let us define a grid on $(0,\infty)$ over a finite or an infinite sequence $0 = t_{n,0} < t_{n,1} \cdots < t_{n,i} < \cdots$. In each window $[t_{n,j}, t_{n,j+1})$, a point $x_{n,j}$, denoted by x_j for simplicity, is chosen.

If $y_{u(1)} \leq t_{n,i} \leq x < t_{n,i+1} \leq y_{u(k)}$, $\lambda_0(x)$ is estimated by

$$\hat{\lambda}_0^*(x) = \frac{F_n(t_{n,i}) - F_n(t_{n,i+1})}{(t_{n,i+1} - t_{n,i})E_n(\exp(z\hat{\beta}, x)} \tag{5.1}$$

where F_n and $E_n(\exp(z\hat{\beta}), x)$ are defined in Section 4.1.

We next obtain

$$\hat{\lambda}_I(x) = \min_{s\geq i+1} \max_{r\leq i} \sum_{j=r}^{s-1} \frac{\hat{\lambda}_0^*(x_j)\omega(x_j)}{\sum_{j=r}^{s-1}\omega(x_j)}$$

for $t_{n,i} \leq x < t_{n,i+1}$ where the with weights $\omega(x_j)$ are given by

$$\omega(x_j) = (t_{n,j+1} - t_{n,j})E_n(\exp(z\hat{\beta}), x_j).$$

$\hat{\lambda}_I(x)$ is called the "isotonic estimator of $\hat{\lambda}_0^*(x)$ with respect to the discrete measure ω.

With $\hat{\lambda}_0^*(x)$ for the initial estimator and weights

$$\omega(x_j) = (t_{n,j+1} - t_{n,j})E_n(\exp(z\hat{\beta}), x_j)$$

we obtain the isotonic estimator of $\lambda_0(x)$ for $t_{n,i} \leq x < t_{n,i+1}$ by

$$\hat{\lambda}_I^*(x) = \min_{s\geq i+1} \max_{r\leq i} \frac{F_n(t_{n,r}) - F_n(t_{n,s})}{\sum_{j=r}^{s-1}(t_{n,j+1} - t_{n,j})E_n(\exp(z\hat{\beta}), x_j)}.$$

5.2 Asymptotic Distribution of the Basic Estimators.

We will assume that the size of the window is related to the sample size, as $t_{n,i+1} - t_{n,i} = cn^{-\alpha}$ for $c > 0$ and $0 < \alpha < 1$, in order to derive asymptotic approximations. Although in practice, we may be interested in simultaneous estimation of $\lambda_0(x)$ at all points, we shall concentrate on the asymptotic behavior of $\hat{\lambda}_0^*(x)$ where x is considered fixed. For mathematical convenience, it will be assumed that

$$x_i = \frac{t_{n,i+1} + t_{n,i}}{2}$$

that is, x_i is the midpoint of a grid spacing. In this chapter, the main objective is to find the asymptotic distribution of

$$\hat{\lambda}_I^*(x) = \min_{s \geq i+1} \max_{r \leq i} \frac{F_n(t_{n,r}) - F_n(t_{n,s})}{\sum_{j=r}^{s-1}(t_{n,j+1} - t_{n,j}) E_n(\exp(z\hat{\beta}), x_j)} \tag{5.2}$$

where $\hat{\beta}$ is a maximum likelihood estimator of β from marginal likelihood.

Before considering the asymptotic properties of $\hat{\lambda}_I^*(x)$, we must first derive the asymptotic distribution of $\hat{\lambda}_0^*(x)$ in (5.1).

For given time x, we define the empirical distribution $Q_n(z)$ of Z by

$$Q_n(z) = \frac{1}{n} \sum_{i=1}^{n} I_{[Z_i \leq z, T_i \geq x]}$$

From the definitions of pages 46–48, Chapter 4, and the above definition, we have the following.

$$\begin{aligned} E(\exp(z\beta), x) &= \int \exp(z\beta) H(x|z) dQ(z) \\ E_n(\exp(z\beta), x) &= \frac{1}{n} \sum_{i=1}^{n} \exp(z_i\beta) I_{[T_i \geq x]} \\ &= \int \exp(z\beta) dQ_n(z) \end{aligned}$$

and similarly

$$\begin{aligned} E_n(\exp(z\hat{\beta}), x) &= \frac{1}{n}\sum_{i=1}^{n} \exp(z_i\hat{\beta}) I_{[T_i \geq x]} \\ &= \int \exp(z\hat{\beta}) dQ_n(z) \end{aligned}$$

In order to apply the method developed by von Misses (1964) to derive the asymptotic results of $\hat{\lambda}_0^*(x)$, the following definition is needed.

$$\begin{aligned} H(\epsilon, \eta) &= \Big\{ \int \exp[z(\epsilon\hat{\beta} + (1-\epsilon)\beta)] \\ &\qquad [\eta dQ_n(z) + (1-\eta) H(x|z) dQ(z)] \Big\}^{-1} \end{aligned} \tag{5.3}$$

Substituting $\epsilon = \eta = 1$ and $\epsilon = \eta = 0$ in (5.3) respectively, we note that

$$H(1,1) = \frac{1}{E_n(\exp(z\hat{\beta}), x)} \tag{5.4}$$

and

$$H(0,0) = \frac{1}{E(\exp(z\beta), x)}. \tag{5.5}$$

A *Taylor* expansion of $H(\epsilon, \eta)$ at $(0,0)$ is utilized to obtain the form which approximates $\hat{\lambda}_0^*(x)$ (Serfling 1980, chapter 6).

Putting $\epsilon = \eta = 1$ in a *Taylor* expansion of $H(\cdot, \cdot)$ yields

$$H(1,1) = H(0,0) + H'_\epsilon(0,0) + H'_\eta(0,0) + h.o.t \tag{5.6}$$

where "h.o.t" means higher order terms. Next, we consider the terms of $H'_\epsilon(0,0)$ and $H'_\eta(0,0)$;

$$\begin{aligned} H'_\epsilon(0,0) &= -\frac{(\hat\beta-\beta)\int z\exp(z\beta)H(x|z)dQ(z)}{\{\int \exp(z\beta)H(x|z)dQ(z)\}^2} \\ &= -\frac{(\hat\beta-\beta)E(z\exp(z\beta),x)}{E^2(\exp(z\beta),x)} \end{aligned} \tag{5.7}$$

$$\begin{aligned} H'_\eta(0,0) &= -\frac{\int \exp(z\beta)\{dQ_n(z)-H(x|z)Q(z)\}}{\{\int \exp(z\beta)H(x|z)dQ(z)\}^2} \\ &= -\frac{E_n(\exp(z\beta),x)-E(\exp(z\beta),x)}{E^2(\exp(z\beta),x)} \end{aligned} \tag{5.8}$$

Hence, from (5.4) to (5.8), we have

$$\begin{aligned} \frac{1}{E_n(\exp(z\hat\beta),x)} &= \frac{1}{E(\exp(z\beta),x)} - (\hat\beta-\beta)\frac{E(z\exp(z\beta),x)}{E^2(\exp(z\beta),x)} \\ &- \frac{E_n(\exp(z\beta),x)-E(\exp(z\beta),x)}{E^2(\exp(z\beta),x)} + h.o.t \end{aligned} \tag{5.9}$$

From (5.1), we can rewrite $\hat\lambda_0^*(x)$ as

$$\begin{aligned} \hat\lambda_0^*(x) &= \frac{F_n(t_{n,i})-F_n(t_{n,i+1})}{(t_{n,i+1}-t_{n,i})E(\exp(z\beta),x)}\Big\{1-(\hat\beta-\beta)\frac{E(z\exp(z\beta),x)}{E(\exp(z\beta),x)} \\ &-\frac{E_n(\exp(z\beta),x)-E(\exp(z\beta),x)}{E(\exp(z\beta),x)} + h.o.t\Big\}. \end{aligned} \tag{5.10}$$

It is straightforward to show that the high order term is of order $O_p(n^{-1})$, since $(\hat{\beta}-\beta) = O_p(n^{\frac{-1}{2}})$ (Tsiatis, 1981) and $(E_n - E) = O_p(n^{\frac{-1}{2}})$ by the strong law of large numbers (Chow & Teicher, 1978). We shall show in Lemma 5.2.2 that the second and third terms of (5.10) are asymptotically negligible, so that we only need to consider the asymptotic properties of the first term of (5.10). Let us define

$$Y_n \stackrel{\text{def}}{=} \frac{F_n(t_{n,i}) - F_n(t_{n,i+1})}{(t_{n,i+1} - t_{n,i})E(\exp(z\beta), x)}. \tag{5.11}$$

Let us next consider the mean of Y_n. Since $E[F_n(x)] = F(x)$ for fixed x,

$$\begin{aligned} E[Y_n] &= E\Big[\frac{F_n(t_{n,i}) - F_n(t_{n,i+1})}{(t_{n,i+1} - t_{n,i})E(\exp(z\beta), x)}\Big] \\ &= \frac{F(t_{n,i}) - F(t_{n,i+1})}{(t_{n,i+1} - t_{n,i})E(\exp(z\beta), x)}. \end{aligned} \tag{5.12}$$

Utilization of the *Taylor* expansion of $F(t_{n,i+1})$ at x where $x = \frac{t_{n,i}+t_{n,i+1}}{2}$, yields

$$\begin{aligned} F(t_{n,i+1}) &= F(x) + (t_{n,i+1} - x)F^{(1)}(x) + \frac{1}{2}(t_{n,i+1} - x)^2 F^{(2)}(x) \\ &\quad + \frac{1}{3!}(t_{n,i+1} - x)^3 F^{(3)}(x) + \frac{1}{4!}(t_{n,i+1} - x)^4 F^{(4)}(x^*) \end{aligned}$$

where x^* is between x and $t_{n,i+1}$. Similarly we obtain

$$\begin{aligned} F(t_{n,i}) &= F(x) + (t_{n,} - x)F^{(1)}(x) + \frac{1}{2}(t_{n,i} - x)^2 F^{(2)}(x) \\ &\quad + \frac{1}{3!}(t_{n,i} - x)^3 F^{(3)}(x) + \frac{1}{4!}(t_{n,i} - x)^4 F^{(4)}(x^{**}) \end{aligned}$$

where x^{**} is between $t_{n,i}$ and x. Hence $F(t_{n,i}) - F(t_{n,i+1})$ reduces to

$$F(t_{n,i}) - F(t_{n,i+1}) = -F^{(1)}(x)cn^{-\alpha} + \frac{1}{24}c^3n^{-3\alpha}F^{(3)}(x) + qn^{-4\alpha} \qquad (5.13)$$

assuming $|F^{(4)}(\cdot)| < \infty$ where q is some constant. Hence, from (5.11) and (5.12) we obtain the asymptotic mean of Y_n for fixed x,

$$E\left[\frac{F_n(t_{n,i}) - F_n(t_{n,i+1})}{cn^{-\alpha}E(\exp(z\beta),x)}\right] \approx \frac{-F^{(1)}(x)}{E(\exp(z\beta),x)} + \frac{1}{24}\frac{c^2n^{-2\alpha}F^{(3)}(x)}{E(\exp(z\beta),x)} + O(n^{-3\alpha}). \qquad (5.14)$$

Note that from (4.7), the first term in the right hand side of (5.13) is $\lambda_0(x)$.

Next, we find the asymptotic variance of Y_n, using the variance formula for a binomial random variable.

$$\begin{aligned} \mathrm{Var}[Y_n] &= \mathrm{Var}\left[\frac{F_n(t_{n,i}) - F_n(t_{n,i+1})}{(t_{n,i+1} - t_{n,i})E(\exp(z\beta),x)}\right] \\ &= \frac{1}{c^2n^{-2\alpha}E^2(\exp(z\beta),x)}\frac{1}{n}\{F(t_{n,i}) - F(t_{n,i+1})\} \\ &\qquad \{1 - (F(t_{n,i}) - F(t_{n,i+1}))\} \end{aligned} \qquad (5.15)$$

which is simplified, using (5.12), to the following;

$$\begin{aligned} \mathrm{Var}[Y_n] &= \frac{1}{c^2n^{-2\alpha+1}E^2(\exp(z\beta),x)}\{[F(t_{n,i}) - F(t_{n,i+1})] - [F(t_{n,i}) - F(t_{n,i+1})]^2\} \\ &= \frac{1}{c^2n^{-2\alpha+1}E^2(\exp(z\beta),x)}(-cn^{-\alpha}F^{(1)}(x) + \frac{1}{24}c^2n^{-2\alpha}F^{(3)}(x) + Bn^{-3\alpha}) \end{aligned}$$

$$= \frac{-F^{(1)}(x)}{cE^2(\exp(z\beta), x)} n^{-1+\alpha} + O(n^{-1}) \tag{5.16}$$

where B is a positive constant.

The asymptotic distribution of Y_n is presented in the next lemma.

Lemma 5.2.1 *Let us assume that $\lambda_0(x)$ is continuous and differentiable and that $F^{(3)}(x)$ exists. If $t_{n,i+1} - t_{n,i} = cn^{-\alpha}$ for $c < 0$ and $0 < \alpha < 1$, then*

$$[\frac{c}{-F^{(1)}(x)}]^{\frac{1}{2}} E(\exp(z\beta), x) n^{(\frac{1-\alpha}{2})} \left(Y_n - \lambda_0(x) - \frac{c^2 n^{-2\alpha} F^{(3)}(x)}{24E(\exp(z\beta), x)}\right)$$

has an asymptotic standard normal distribution for $\frac{1}{7} < \alpha \leq \frac{1}{5}$.

If $\frac{1}{5} < \alpha < 1$, then

$$[\frac{c}{-F^{(1)}(x)}]^{\frac{1}{2}} E(\exp(z\beta), x) n^{(\frac{1-\alpha}{2})} (Y_n - \lambda_0(x))$$

has an asymptotic standard normal distribution.

Proof of Lemma 5.2.1 *Let us define*

$$X_j = I_{[t_{n,i} \leq T_j < t_{n,i+1}, \delta_j = 1]}$$

for $j = 1, \cdots, n$. It follow that

$$\begin{aligned} F_n(t_{n,i}) - F_n(t_{n,i+1}) &= \frac{1}{n} \sum_{j=1}^{n} I_{[t_{n,i} \leq T_j < t_{n,i+1}, \delta_j = 1]} \\ &\stackrel{\text{def}}{=} \bar{X}_n \end{aligned}$$

Noting that X_j's are i.i.d. Bernoulli random variables with parameter $F(t_{n,i}) - F(t_{n,i+1})$, It follows that $n(F_n(t_{n,i}) - F_n(t_{n,i+1}))$ is distributed as a binomial random variable with parameters n and $F(t_{n,i}) - F(t_{n,i+1})$. To prove the lemma, we shall prove that

$$Z_n \stackrel{\text{def}}{=} \frac{\frac{F_n(t_{n,i})-F_n(t_{n,i+1})}{cn^{-\alpha}E(\exp(z\beta),x)} - \frac{F(t_{n,i})-F(t_{n,i+1})}{cn^{-\alpha}E(\exp(z\beta),x)}}{\sqrt{\frac{(F(t_{n,i})-F(t_{n,i+1}))(1-F(t_{n,i})+F(t_{n,i+1}))}{c^2n^{1-2\alpha}E^2(\exp(z\beta),x)}}}$$

has an asymptotic standard normal distribution. To prove the asymptotic normality of Z_n, we show that the moment generating function of Z_n converges to the moment generating function of the standard normal distribution.

$$\begin{aligned}
M_{Z_n}(t) &= E[\exp(tZ_n)] \\
&= E[\exp(t\frac{\frac{\bar{X}_n}{cn^{-\alpha}E(\exp(z\beta),x)} - \frac{F(t_{n,i})-F(t_{n,i+1})}{cn^{-\alpha}E(\exp(z\beta),x)}}{\sqrt{\frac{F(t_{n,i})-F(t_{n,i+1}))(1-F(t_{n,i})+F(t_{n,i+1}))}{c^2n^{1-2\alpha}E^2(\exp(z\beta),x)}}})] \\
&= E[\exp(t\frac{\frac{1}{n}\sum_{j=1}^{n}(\frac{X_j}{cn^{-\alpha}E(\exp(z\beta),x)} - \frac{F(t_{n,i})-F(t_{n,i+1})}{cn^{-\alpha}E(\exp(z\beta),x)})}{\sqrt{\frac{F(t_{n,i})-F(t_{n,i+1}))(1-F(t_{n,i})+F(t_{n,i+1}))}{c^2n^{1-2\alpha}E^2(\exp(z\beta),x)}}})] \\
&= E[\exp(t\prod_{j=1}^{n}\frac{X_j-(F(t_{n,i})-F(t_{n,i+1}))}{\sqrt{n(F(t_{n,i})-F(t_{n,i+1}))(1-F(t_{n,i})+F(t_{n,i+1}))}})] \\
&= E[\exp(t\frac{X_1-(F(t_{n,i})-F(t_{n,i+1}))}{\sqrt{n(F(t_{n,i})-F(t_{n,i+1}))(1-F(t_{n,i})+F(t_{n,i+1}))}})]^n \\
&= \{(F(t_{n,i})-F(t_{n,i+1})) \\
&\quad \exp(t\frac{1-(F(t_{n,i})-F(t_{n,i+1}))}{\sqrt{n(F(t_{n,i})-F(t_{n,i+1}))(1-F(t_{n,i})+F(t_{n,i+1}))}}) \\
&\quad +(1-(F(t_{n,i})-F(t_{n,i+1}))) \\
&\quad \exp(t\frac{-(F(t_{n,i})-F(t_{n,i+1}))}{\sqrt{n(F(t_{n,i})-F(t_{n,i+1}))(1-F(t_{n,i})+F(t_{n,i+1}))}})\}^n
\end{aligned}$$

$$
\begin{aligned}
&= \{(F(t_{n,i}) - F(t_{n,i+1})) \\
&\quad (1 + t\frac{1 - (F(t_{n,i}) - F(t_{n,i+1}))}{\sqrt{n(F(t_{n,i}) - F(t_{n,i+1}))(1 - F(t_{n,i}) + F(t_{n,i+1}))}} \\
&\quad + \frac{t^2}{2}\{\frac{1 - (F(t_{n,i}) - F(t_{n,i+1}))}{\sqrt{n(F(t_{n,i}) - F(t_{n,i+1}))(1 - F(t_{n,i}) + F(t_{n,i+1}))}}\}^2 + o(n^{-1}) \\
&\quad + (1 - (F(t_{n,i}) - F(t_{n,i+1})) \\
&\quad (1 + t\frac{-(F(t_{n,i}) - F(t_{n,i+1}))}{\sqrt{n(F(t_{n,i}) - F(t_{n,i+1}))(1 - F(t_{n,i}) + F(t_{n,i+1}))}} \\
&\quad + \frac{t^2}{2}\{\frac{-(F(t_{n,i}) - F(t_{n,i+1}))}{\sqrt{n(F(t_{n,i}) - F(t_{n,i+1}))(1 - F(t_{n,i}) + F(t_{n,i+1}))}}\}^2 + o(n^{-1}))\}^n
\end{aligned}
$$

$$
\begin{aligned}
&= \{F(t_{n,i}) - F(t_{n,i+1}) \\
&\quad + \frac{t^2}{2}\frac{(F(t_{n,i}) - F(t_{n,i+1}))(1 - F(t_{n,i}) + F(t_{n,i+1}))^2}{n(F(t_{n,i}) - F(t_{n,i+1}))(1 - F(t_{n,i}) + F(t_{n,i+1}))} + o(n^{-1}) \\
&\quad + (1 - (F(t_{n,i}) - F(t_{n,i+1})) \\
&\quad + \frac{t^2}{2}\frac{(1 - F(t_{n,i}) + F(t_{n,i+1}))F(t_{n,i}) - F(t_{n,i+1})^2}{n(F(t_{n,i}) - F(t_{n,i+1}))(1 - F(t_{n,i}) + F(t_{n,i+1}))} + o(n^{-1})\}^n
\end{aligned}
$$

$$
= \left(1 + \frac{t^2}{2n} + o(n^{-1})\right)^n \overset{n\to\infty}{\longrightarrow} \exp(\frac{t^2}{2})
$$

Hence from Levy continuity theorem (Chow & Teicher, 1978), it follows that the standardized first term of (5.10)

$$
\frac{\frac{F_n(t_{n,i})-F_n(t_{n,i+1})}{cn^{-\alpha}E(\exp(z\beta),x)} - \frac{F(t_{n,i})-F(t_{n,i+1})}{cn^{-\alpha}E(\exp(z\beta),x)}}{\sqrt{\frac{F(t_{n,i})-F(t_{n,i+1})(1-F(t_{n,i})+F(t_{n,i+1}))}{c^2n^{1-2\alpha}E^2(\exp(z\beta),x)}}}
$$

has an asymptotic standard normal distribution.

To simplify the notation, the following four terms are defined

$$
A_n = \lambda_0(x) + \frac{1}{24}\frac{c^2n^{-2\alpha}F^{(3)}(x)}{E(\exp(z\beta),x)} + O(n^{-3\alpha}),
$$

$$
\begin{aligned}
B_n &= \frac{F(t_{n,i}) - F(t_{n,i+1})}{cn^{-\alpha} E(\exp(z\beta), x)}, \\
V_n^2 &= \frac{(F(t_{n,i}) - F(t_{n,i+1}))(1 - F(t_{n,i}) + F(t_{n,i+1}))}{c^2 n^{1-2\alpha} E^2(\exp(z\beta), x)}, \\
U_n^2 &= \frac{F^{(1)}(x)}{cE^2(\exp(z\beta), x)} n^{-1+\alpha}.
\end{aligned}
$$

From the asymptotic result obtained above, we have

$$
\frac{Y_n - B_n}{V_n} \xrightarrow{\text{d}} N(0,1).
$$

Since we aim to find the asymptotic distribution of $\frac{Y_n - A_n}{U_n}$*, we rewrite*

$$
\frac{Y_n - A_n}{U_n} = \frac{Y_n - B_n + B_n - A_n}{U_n} = \frac{Y_n - B_n}{V_n}\frac{V_n}{U_n} + \frac{B_n - A_n}{U_n}.
$$

Note that

$$
\lim_{n\to\infty} \frac{V_n^2}{U_n^2} = \lim_{n\to\infty} \frac{\frac{F(t_{n,i}) - F(t_{n,i+1})(1 - F(t_{n,i}) + F(t_{n,i+1}))}{c^2 n^{1-2\alpha} E^2(\exp(z\beta), x)}}{\frac{F^{(1)}(x)}{cE^2(\exp(z\beta), x)} n^{-1+\alpha}} = 1
$$

and when $\frac{1}{7} < \alpha < 1$*, following from (4.7) and (5.12),*

$$
\lim_{n\to\infty} \frac{B_n - A_n}{U_n}
$$

$$= - \lim_{n\to\infty} \frac{\lambda_0(x) + \frac{1}{24}\frac{c^2 n^{-2\alpha} F^{(3)}(x)}{E(\exp(z\beta),x)} + O(n^{-3\alpha}) - \frac{F(t_{n,i}) - F(t_{n,i+1})}{cn^{-\alpha}E(\exp(z\beta),x)}}{\sqrt{\frac{F^{(1)}(x)}{cE^2(\exp(z\beta),x)} n^{-1+\alpha}}}$$

$$= 0.$$

Therefore it follows that

$\frac{Y_n - A_n}{U_n}$ *is asymptotically distributed as* $N(0,1)$. *That is, when* $\frac{1}{7} < \alpha < 1$

$$[\frac{c}{-F^{(1)}(x)}]^{\frac{1}{2}} E(\exp(z\beta),x) n^{(\frac{1-\alpha}{2})}\Big(Y_n - \lambda_0(x) - \frac{c^2 n^{-2\alpha} F^{(3)}(x)}{24E(\exp(z\beta),x)} - O(n^{-3\alpha})\Big)$$

has an asymptotic standard normal distribution. For $\frac{1}{7} < \alpha \leq \frac{1}{5}$, $O(n^{-3\alpha})$ *is asymptotically negligible and for* $\frac{1}{5} < \alpha < 1$, $\frac{1}{24}\frac{c^2 n^{-2\alpha} F^{(3)}(x)}{E(\exp(z\beta),x)}$ *is asymptotically negligible because of the factor* $n^{\frac{1-\alpha}{2}}$. *This completes the proof of the lemma.*

Next, we prove that let us show the second and third terms in $\hat{\lambda}_0^*(x)$ in (5.10) are asymptotically negligible. In other words, we shall show that they converge to 0 in probability.

<u>*Lemma 5.2.2*</u> *For* $0 < \alpha < 1$, *The summation of the second and third terms of* $\hat{\lambda}_0^*(x)$ *which are multiplied by* $n^{(\frac{1-\alpha}{2})}$,

$$n^{(\frac{1-\alpha}{2})} \frac{F_n(t_{n,i}) - F_n(t_{n,i+1})}{(t_{n,i+1} - t_{n,i}) E^2(\exp(z\beta),x)}$$

$$\{E(z\exp(z\beta),x)(\hat{\beta} - \beta) + (E_n(\exp(z\beta),x) - E(\exp(z\beta),x))\},$$

converges to 0 in probability.

Proof of Lemma 5.2.2 From (5.13), (5.15) and Chebyshev's theorem,

$$Y_n = \frac{F_n(t_{n,i}) - F_n(t_{n,i+1})}{(t_{n,i+1} - t_{n,i})E(\exp(z\beta), x)}$$

converges to $\frac{F^{(1)}(x)}{E(\exp(z\beta),x)}$ in probability. Since we assume that

$$E(z\exp(z\beta), x) < \infty,$$

it suffices to show that $n^{\frac{1-\alpha}{2}}(\hat{\beta}-\beta)$ and $n^{\frac{1-\alpha}{2}}(E_n(\exp(z\beta), x)-E(\exp(z\beta), x))$ converge to 0 in probability. It is well known that $\sqrt{n}(\hat{\beta}-\beta) = O_p(1)$ and $\sqrt{n}(E_n(\exp(z\beta), x)-E(\exp(z\beta), x)) = O_p(1)$, so that

$$n^{\frac{-\alpha}{2}}\sqrt{n}(\hat{\beta} - \beta) = n^{\frac{-\alpha}{2}}O_p(1) \overset{n\to\infty}{\longrightarrow} 0,$$

and

$$n^{\frac{-\alpha}{2}}\sqrt{n}(E_n(\exp(z\beta), x) - E(\exp(z\beta), x)) = n^{\frac{-\alpha}{2}}O_p(1) \overset{n\to\infty}{\longrightarrow} 0.$$

This completes the proof of the lemma.

Theorem 5.2.1 The standardized form of $\hat{\lambda}_0^*(x)$,

$$[\frac{c}{-F^{(1)}(x)}]^{\frac{1}{2}}E(\exp(z\beta), x)n^{(\frac{1-\alpha}{2})}\Big(\hat{\lambda}_0^*(x) - \lambda_0(x) - \frac{c^2n^{-2\alpha}F^{(3)}(x)}{24E(\exp(z\beta), x)}\Big)$$

has an asymptotic standard normal distribution for $\frac{1}{7} < \alpha \leq \frac{1}{5}$.

If we have $\frac{1}{5} < \alpha < 1$, it follows that

$$[\frac{c}{-F^{(1)}(x)}]^{\frac{1}{2}}E(\exp(z\beta), x)n^{(\frac{1-\alpha}{2})}(\hat{\lambda}_0^*(x) - \lambda_0(x))$$

has an asymptotic standard normal distribution.

<u>*Proof of Theorem 5.2.1*</u> *Since we proved in Lemma 5.2.1. that the normalized first term of* $\hat{\lambda}_0^*(x)$ *is asymptotically distributed as a standard normal random variable and since we proved in Lemma 5.2.2 that the second and third terms of (5.10) with the factor* $n^{(\frac{1-\alpha}{2})}$ *converge to 0 in probability, Slutsky's theorem (Chow & Teicher, 1978) implies that* $\hat{\lambda}_0^*(x)$ *and* $\frac{F_n(t_{n,i})-F_n(t_{n,i+1})}{(t_{n,i+1}-t_{n,i})E(\exp(z\beta),x)}$ *have same asymptotic distribution. Thus the conclusions in the theorem follow from Lemma 5.2.1.*

5.3 Asymptotic Distribution of Isotonic Estimators Based on the Window

In this section, we shall derive the asymptotic distribution of the isotonic regression of $\hat{\lambda}_0^*(x)$ for the wide window case. (i.e., grid spacing of the form $cn^{-\alpha}$ where $0 < \alpha < \frac{1}{3}$.) The wide window case is important, since the isotonic estimator of $\lambda_0(x)$ has an asymptotic normal distribution in this case.

Barlow and van Zwet (1971) prove that this result for the case without covariates. The basic estimator for failure rate based on ordered observations and the isotonic estimator based on the window are asymptotically equivalent. Similarly we shall prove that $\hat{\lambda}_0^*(x)$ and $\hat{\lambda}_I^*(x)$ are asymptotically equivalent under mild regularity conditions.

<u>*Theorem 5.3.1*</u> *If*

(i) $\lambda_0(x) = \frac{-F^{(1)}(x)}{E(\exp(z\beta),x)}$ *is strictly increasing in* $x \geq 0$;

(ii) $\lambda_0(x)$ *is continuously differentiable and* $F^{(3)}(\cdot)$ *exists in a neighborhood of* x;

(iii) $t_{n,i+1} - t_{n,i} = cn^{-\alpha}$ *and* $0 < \alpha < \frac{1}{3}$,

it follows that

$$\lim_{n\to\infty} \Pr[\hat{\lambda}_0^*(x) \neq \hat{\lambda}_I^*(x)] = 0. \tag{5.17}$$

Prior to proving this result, three lemmas are needed.

<u>*Lemma 5.3.1*</u> *For* $0 < \alpha < \frac{1}{3}$ *and arbitrary numbers* $\epsilon > 0$, *and* $\delta > 0$, *and* $0 < B < T_0$,

$$\Pr[\sup_{0\le x\le B} |\hat{\lambda}_0^*(x) - \lambda_0(x)| > \epsilon] < \delta \qquad \textit{for large} n$$

where

$$\hat{\lambda}_0^*(x) = \frac{F_n(t_{n,i}) - F_n(t_{n,i+1})}{(t_{n,i+1} - t_{n,i})E_n(\exp(z\hat{\beta}), x)}$$

<u>*Proof of Lemma 5.3.1*</u> *Using (5.10), we have for arbitrary* x,

$$\begin{aligned}
&\hat{\lambda}_0^*(x) - \lambda_0(x) \\
&= \frac{F_n(t_{n,i}) - F_n(t_{n,i+1})}{(t_{n,i+1} - t_{n,i})E_n(\exp(z\hat{\beta}), x)} - \frac{F(t_{n,i}) - F(t_{n,i+1})}{(t_{n,i+1} - t_{n,i})E(\exp(z\beta), x)} \\
&\quad + \frac{F(t_{n,i}) - F(t_{n,i+1})}{(t_{n,i+1} - t_{n,i})E(\exp(z\beta), x)} - \lambda_0(x) \\
&= \frac{F_n(t_{n,i}) - F_n(t_{n,i+1})}{(t_{n,i+1} - t_{n,i})E(\exp(z\beta), x)}\Big\{1 - (\hat{\beta} - \beta)\frac{E(z\exp(z\beta), x)}{E(\exp(z\beta), x)} \\
&\quad - \frac{E_n(\exp(z\beta), x) - E(\exp(z\beta), x)}{E(\exp(z\beta), x)} + h.o.t.\Big\} \\
&\quad - \frac{F(t_{n,i}) - F(t_{n,i+1})}{(t_{n,i+1} - t_{n,i})E(\exp(z\beta), x)} + \frac{F(t_{n,i}) - F(t_{n,i+1})}{(t_{n,i+1} - t_{n,i})E(\exp(z\beta), x)} \\
&\quad - \lambda_0(x) \\
&= \frac{1}{(t_{n,i+1} - t_{n,i})E(\exp(z\beta), x)}\Big\{[F_n(t_{n,i}) - F(t_{n,i})] - [F_n(t_{n,i+1}) - F(t_{n,i+1})]\Big\}
\end{aligned}$$

$$+\Big\{\frac{1}{(t_{n,i+1}-t_{n,i})E(\exp(z\beta),x)}\{-F^{(1)}(x)cn^{-\alpha}$$
$$+\frac{1}{2}(r^2-(1-r)^2)c^2n^{-2\alpha}F^{(2)}(x)+\frac{(r^3-(1-r)^3)c^3n^{-3\alpha}F^{(3)}(x)}{6}\}$$
$$-\lambda_0(x)\Big\}+\frac{F_n(t_{n,i})-F_n(t_{n,i+1})}{(t_{n,i+1}-t_{n,i})E(\exp(z\beta),x)}(\hat{\beta}-\beta)\frac{E(z\exp(z\beta),x)}{E(\exp(z\beta),x)}$$
$$+\frac{F_n(t_{n,i})-F_n(t_{n,i+1})}{(t_{n,i+1}-t_{n,i})E(\exp(z\beta),x)}\frac{E_n(\exp(z\beta),x)-E(\exp(z\beta),x)}{E(\exp(z\beta),x)}+h.o.t$$

$$= I+II+III+IV+h.o.t \qquad (5.18)$$

where $|t_{n,i}-x| = rcn^{-\alpha}$ *for* $0<r<1$.

Let us consider the first part (I) of (5.18);

$$\Pr[\sup_{0\le x\le B}\frac{1}{(t_{n,i+1}-t_{n,i})E(\exp(z\beta),x)}|[F_n(t_{n,i})-F(t_{n,i})]-[F_n(t_{n,i+1})-F(t_{n,i+1})]|>\epsilon]$$

$$= \Pr[\sup_{0\le x\le B}\frac{1}{(t_{n,i+1}-t_{n,i})E(\exp(z\beta),x)}\Big\{|F_n(t_{n,i})-F(t_{n,i})|+|F_n(t_{n,i+1})-F(t_{n,i+1})|\Big\}>\epsilon]$$

$$\le \Pr[\sup_{0\le x\le B}\frac{1}{cE(\exp(z\beta),x)}n^{\alpha}|F_n(t_{n,i})-F(t_{n,i})|>\frac{1}{2}\epsilon]$$
$$+\Pr[\sup_{0\le x\le B}\frac{1}{cE(\exp(z\beta),x)}n^{\alpha}|F_n(t_{n,i+1})-F(t_{n,i+1})|>\frac{1}{2}\epsilon]$$

$$< \delta \quad \text{for large } n. \qquad (5.19)$$

The last line follows, since

(i) $E(\exp(z\beta),x)$ *is bounded on [0,B],*

(ii) $\Pr[\sup_{\{0\leq x\leq B\}} n^{\alpha}|F_n(t_{n,i+1}) - F(t_{n,i+1})| > \epsilon] < \delta$ *for large* n *and* $0 < \alpha < \frac{1}{3}$.

It is straightforward to show that the second part of (5.18) goes to 0 as n goes to ∞. *Let us consider the third part of (5.18);*

$$\Pr[\sup_{0\leq x\leq B} |\frac{F_n(t_{n,i}) - F_n(t_{n,i+1})}{(t_{n,i+1} - t_{n,i})E(\exp(z\beta), x)}(\hat{\beta} - \beta)\frac{E(z\exp(z\beta), x)}{E(\exp(z\beta), x)}| > \epsilon]$$

$$\leq \Pr[2c^{-1}n^{\alpha}(\hat{\beta} - \beta) \sup_{0\leq x\leq B} \frac{E(z\exp(z\beta), x)}{E^2(\exp(z\beta), x)} > \epsilon]$$

$$< \delta. \qquad (5.20)$$

The last line follows since $E(z\exp(z\beta), x)$ *and* $E(\exp(z\beta), x)$ *are bounded on [0,B] and* $n^{\alpha}(\hat{\beta} - \beta)$ *goes to* 0 *as* n *goes to* ∞ *when* $0 < \alpha < \frac{1}{3}$.
Finally, consider the fourth part of (5.18);

$$\Pr[\sup_{0\leq x\leq B} |\frac{F_n(t_{n,i}) - F_n(t_{n,i+1})}{(t_{n,i+1} - t_{n,i})E(\exp(z\beta), x)} \frac{E_n(\exp(z\beta), x) - E(\exp(z\beta), x)}{E(\exp(z\beta), x)}| > \epsilon]$$

$$< \Pr[2c^{-1} \sup_{0\leq x\leq B} |\frac{n^{\alpha}\{E_n(\exp(z\beta), x) - E(\exp(z\beta), x)\}}{E^2(\exp(z\beta), x)}| > \epsilon]$$

$$< \delta. \qquad (5.21)$$

The last line follows from the fact that $n^{\frac{1}{2}}\{E_n(\exp(z\beta), x) - E(\exp(z\beta), x)\}$ *converges to a Gaussian process (Tsiatis, 1981). It was shown in Section (5.2) that the h.o.t is negligible. Combining (5.19) (5.20) and (5.21), the proof of the lemma is complete.*

Let us define $\hat{\lambda}_I^{*[A,B]}(x)$ for any A and B $(0 \le A < x < B \le T_0)$.

$$\begin{aligned}\hat{\lambda}_I^{*[A,B]}(x) &\stackrel{\text{def}}{=} \min_{\substack{s \ge i+1 \\ t_{n,s} \le B}} \max_{\substack{r \le i \\ t_{n,r} \ge A}} \frac{F_n(t_{n,s}) - F_n(t_{n,r})}{\sum_{j=r}^{s-1}(t_{n,j+1} - t_{n,j})E_n(\exp(z\hat{\beta}), x_j)} \\ &= \min_{\substack{s \ge i+1 \\ t_{n,s} \le B}}{}' \max_{\substack{r \le i \\ t_{n,r} \ge A}} \sum_{j=r}^{s-1} \frac{\hat{\lambda}_0^*(x_j)\omega(x_j)}{\sum_{j=r}^{s-1}\omega(x_j)}.\end{aligned}$$

It is clear that

(i) $\hat{\lambda}_I^*(x) = \hat{\lambda}_I^{*[0,T_0]}(x)$

(ii) $\hat{\lambda}_I^*(x) \le \hat{\lambda}_I^{*[0,B]}(x)$

(iii) $\hat{\lambda}_I^*(x) \ge \hat{\lambda}_I^{*[A,T_0]}(x)$.

Next, we prove the following lemma, which allows us to consider only a bounded range of time for proving Theorem 5.3.1.

<u>*Lemma 5.3.2*</u> *For a fixed x and any A and B, $0 \le A < x < B \le T_0$,*

$$\lim_{n\to\infty} \Pr[\hat{\lambda}_I^*(x) \ne \hat{\lambda}_I^{*[A,B]}(x)] = 0.$$

<u>*Proof of Lemma 5.3.2*</u> *The proof of this lemma consists of proving the following two results.*

$$\lim_{n\to\infty} \Pr[\hat{\lambda}_I^*(x) \ne \hat{\lambda}_I^{*[0,B]}(x)] = 0 \tag{5.22}$$

and

$$\lim_{n\to\infty} \Pr[\hat{\lambda}_I^{*[0,B]}(x) \ne \hat{\lambda}_I^{*[A,B]}(x)] = 0. \tag{5.23}$$

Since the proofs of (5.22) and (5.23) are similar, we shall prove (5.22). In other words, we need to prove that for any $\delta_1 > 0$, there exists N such that whenever $n > N$,

$$\Pr[\hat{\lambda}_I^*(x) \neq \hat{\lambda}_I^{*[0,B]}(x)] < \delta_1. \tag{5.24}$$

If $B = T_0$, then $\hat{\lambda}_0^*(x) = \hat{\lambda}_I^{*[0,B]}(x)$ and the proof is completed. Assume $B < T_0$. Let us divide $[x, T_0]$ into

$$[x, T_0] = \underbrace{[x, C_1)}_{I_1} \cup \underbrace{[C_1, C_2)}_{I_2} \cup \underbrace{[C_2, B)}_{I_3} \cup \underbrace{[B, B_1)}_{I_4} \cup \underbrace{[B_1, T_0]}_{I_5}$$

where x is a fixed point and B_1 is close to T_0. Let C_1 and C_2 be any two points satisfying

$$\lambda_0(C_2) - \lambda_0(C_1) > 2\epsilon \tag{5.25}$$

and

$$\lambda_0(B) - \lambda_0(C_2) > \delta + \epsilon \tag{5.26}$$

for some $\delta > 0$.

From Lemma 5.3.1, for any ϵ and $\delta > 0$, there exists N such that whenever $n > N$

$$\Pr[\sup_{0<x<B_1} |\hat{\lambda}_0^*(x) - \lambda_0(x)| < \epsilon] > 1 - \frac{\delta}{3}.$$

Denote the subset of all sample points satisfying

$$\sup_{0<x<B_1} |\hat{\lambda}_0^*(x) - \lambda_0(x)| < \epsilon$$

as Ω_1. Then $\Pr(\Omega_1) > 1 - \frac{\delta}{3}$ if $n > N_1$.

Define

$$I_i(x) = \begin{cases} 1 & \text{if } x \in I_i \\ 0 & \text{otherwise} \end{cases}$$

for $i = 1, \cdots, 5$.

Denote the subset of all sample points satisfying

$$\frac{\sum_{j=i}^{s-1} \omega(x_j) I_4(x)}{\sum_{j=i}^{s-1} \omega(x_j)} > W > 0$$

for some $W > 0$ *as* Ω_2. *Then* $\Pr(\Omega_2) > 1 - \frac{\delta}{3}$ *if* $n > N_2$.

Denote the subset of all sample points satisfying

$$\frac{\sum_{j=i}^{s-1} \omega(x_j) I_5(x)}{\sum_{j=i}^{s-1} \omega(x_j)} < \frac{\epsilon}{J}$$

as Ω_3. *Then* $\Pr(\Omega_3) > 1 - \frac{\delta}{3}$ *if* $n > N_3$.

Now we focus our discussion upon fixed sample points in Ω *such that when* $n > \max\{N_1, N_2, N_3\}$, $\Pr(\Omega) > 1 - \delta$ *where* $\Omega = \Omega_1 \cap \Omega_2 \cap \Omega_3$.

Since we note

$$\hat{\lambda}_I^*(x) \leq \hat{\lambda}_I^{*[0,B]}(x),$$

it suffices to show

$$\hat{\lambda}_I^*(x) \geq \hat{\lambda}_I^{*[0,B]}(x)$$

for fixed x. *If*

$$\min_{\substack{s \geq i+1 \\ t_{n,s} > B}} \frac{\sum_{j=i}^{s-1} \hat{\lambda}_0^*(x_j)\omega(x_j)}{\sum_{j=r}^{s-1} \omega(x_j)} > \min_{\substack{s \geq i+1 \\ t_{n,s} \leq C_1}} \frac{\sum_{j=i}^{s-1} \hat{\lambda}_0^*(x_j)\omega(x_j)}{\sum_{j=r}^{s-1} \omega(x_j)}, \tag{5.27}$$

it follows from the fact that C_1 *is arbitrary point less than* B *that*

$$\min_{s \geq i+1} \frac{\sum_{j=i}^{s-1} \hat{\lambda}_0^*(x_j)\omega(x_j)}{\sum_{j=r}^{s-1} \omega(x_j)} = \min_{\substack{s \geq i+1 \\ t_{n,s} \leq B}} \frac{\sum_{j=i}^{s-1} \hat{\lambda}_0^*(x_j)\omega(x_j)}{\sum_{j=r}^{s-1} \omega(x_j)} \tag{5.28}$$

Next, using Lemma 5.3.1, the monotonicity of $\lambda_0(x)$, (5.25), and (5.26) we observe that

$$\begin{aligned} J &\stackrel{\text{def}}{=} \frac{\sum_{j=i}^{s-1} \hat{\lambda}_0^*(x_j)\omega(x_j)I_1(x)}{\sum_{j=i}^{s-1} \omega(x_j)I_1(x)} \\ &\leq \frac{\sum_{j=i}^{s-1} \lambda_0(x_j)\omega(x_j)I_1(x)}{\sum_{j=i}^{s-1} \omega(x_j)I_1(x)} + \epsilon \\ &\leq \lambda_0(C_1) + \epsilon, \end{aligned} \tag{5.29}$$

$$\begin{aligned} \frac{\sum_{j=i}^{s-1} \hat{\lambda}_0^*(x_j)\omega(x_j)I_2(x)}{\sum_{j=i}^{s-1} \omega(x_j)I_2(x)} &\geq \frac{\sum_{j=i}^{s-1} \lambda_0(x_j)\omega(x_j)I_2(x)}{\sum_{j=i}^{s-1} \omega(x_j)I_2(x)} - \epsilon \\ &\geq \lambda_0(C_1) - \epsilon \\ &> J - 2\epsilon, \end{aligned} \tag{5.30}$$

$$\begin{aligned} \frac{\sum_{j=i}^{s-1} \hat{\lambda}_0^*(x_j)\omega(x_j)I_3(x)}{\sum_{j=i}^{s-1} \omega(x_j)I_3(x)} &\geq \frac{\sum_{j=i}^{s-1} \lambda_0(x_j)\omega(x_j)I_3(x)}{\sum_{j=i}^{s-1} \omega(x_j)I_3(x)} \\ &\geq \lambda_0(C_2) - \epsilon \\ &\geq \lambda_0(C_1) + \epsilon > J, \end{aligned} \tag{5.31}$$

$$\begin{aligned} \frac{\sum_{j=i}^{s-1} \hat{\lambda}_0^*(x_j)\omega(x_j)I_4(x)}{\sum_{j=i}^{s-1} \omega(x_j)I_4(x)} &\geq \frac{\sum_{j=i}^{s-1} \lambda_0(x_j)\omega(x_j)I_4(x)}{\sum_{j=i}^{s-1} \omega(x_j)I_4(x)} - \epsilon \\ &\geq \lambda_0(B_1) - \epsilon \\ &\geq \lambda_0(C_2) + \delta \\ &\geq \lambda_0(C_1) + 2\epsilon + \delta \\ &\geq J + \epsilon + \delta. \end{aligned} \tag{5.32}$$

We also observe that

$$\begin{aligned}\frac{\sum_{j=i}^{s-1}\omega(x_j)I_4(x)}{\sum_{j=i}^{s-1}\omega(x_j)} &= \frac{t_{n,s}-B}{t_{n,s}-t_{n,i}}\frac{\sum_{j=i}^{s-1}\omega(x_j)I_4(x)/(t_{n,s}-B)}{\sum_{j=i}^{s-1}\omega(x_j)/(t_{n,s}-t_{n,i})}\\ &= \frac{t_{n,s}-B}{t_{n,s}-t_{n,i}}r\\ &> W \qquad (5.33)\end{aligned}$$

for large n and constants $r>0$ and $W>0$, and since B_1 is close to T_0, there exists N such that for $n>N$, we have

$$\begin{aligned}\frac{\sum_{j=i}^{s-1}\omega(x_j)I_5(x)}{\sum_{j=i}^{s-1}\omega(x_j)} &= \frac{(t_{n,s}-B_1)}{(t_{n,s}-t_{n,i})}\frac{\sum_{j=i}^{s-1}\omega(x_j)I_5(x)/(t_{n,s}-B_1)}{\sum_{j=i}^{s-1}\omega(x_j)/(t_{n,s}-t_{n,i})}\\ &< \frac{(t_{n,s}-B_1)}{(t_{n,s}-t_{n,i})}r\\ &< \frac{\epsilon}{J} \qquad (5.34)\end{aligned}$$

for some constant $r>0$.

To prove (5.27), using (5.29)- (5.34) we observe that for arbitrary $\epsilon>0$ and some $\delta>0$

$$\begin{aligned}\frac{\sum_{j=i}^{s-1}\hat\lambda_0^*(x_j)\omega(x_j)}{\sum_{j=i}^{s-1}\omega(x_j)} &= \frac{\sum_{j=i}^{s-1}\omega(x_j)I_1(x)}{\sum_{j=i}^{s-1}\omega(x_j)}\frac{\sum_{j=i}^{s-1}\hat\lambda_0^*(x_j)\omega(x_j)I_1(x)}{\sum_{j=i}^{s-1}\omega(x_j)I_1(x)}\\ &+\frac{\sum_{j=i}^{s-1}\omega(x_j)I_2(x)}{\sum_{j=i}^{s-1}\omega(x_j)}\frac{\sum_{j=i}^{s-1}\hat\lambda_0^*(x_j)\omega(x_j)I_2(x)}{\sum_{j=i}^{s-1}\omega(x_j)I_2(x)}\\ &+\frac{\sum_{j=i}^{s-1}\omega(x_j)I_3(x)}{\sum_{j=i}^{s-1}\omega(x_j)}\frac{\sum_{j=i}^{s-1}\hat\lambda_0^*(x_j)\omega(x_j)I_3(x)}{\sum_{j=i}^{s-1}\omega(x_j)I_3(x)}\\ &+\frac{\sum_{j=i}^{s-1}\omega(x_j)I_4(x)}{\sum_{j=i}^{s-1}\omega(x_j)}\frac{\sum_{j=i}^{s-1}\hat\lambda_0^*(x_j)\omega(x_j)I_4(x)}{\sum_{j=i}^{s-1}\omega(x_j)I_4(x)}\\ &+\frac{\sum_{j=i}^{s-1}\omega(x_j)I_5(x)}{\sum_{j=i}^{s-1}\omega(x_j)}\frac{\sum\hat\lambda_0^*(x_j)\omega(x_j)I_5(x)}{\sum_{j=i}^{s-1}\omega(x_j)I_5(x)}\\ &\geq \frac{\sum_{j=i}^{s-1}\omega(x_j)I_1(x)}{\sum_{j=i}^{s-1}\omega(x_j)}J\end{aligned}$$

$$
\begin{aligned}
&+\frac{\sum_{j=i}^{s-1}\omega(x_j)I_2(x)}{\sum_{j=i}^{s-1}\omega(x_j)}(J-2\epsilon)\\
&+\frac{\sum_{j=i}^{s-1}\omega(x_j)I_3(x)}{\sum_{j=i}^{s-1}\omega(x_j)}J\\
&+\frac{\sum_{j=i}^{s-1}\omega(x_j)I_4(x)}{\sum_{j=i}^{s-1}\omega(x_j)}(J+\delta+\epsilon)\\
\geq\ & J-2\epsilon+\frac{\sum_{j=i}^{s-1}\omega(x_j)I_4(x)}{\sum_{j=i}^{s-1}\omega(x_j)}(\epsilon+\delta)-\frac{\sum_{j=i}^{s-1}\omega(x_j)I_5(x)}{\sum_{j=i}^{s-1}\omega(x_j)}J\\
&> J-3\epsilon+W(\delta)\\
&> J \qquad (5.35)
\end{aligned}
$$

for $\delta > \frac{3\epsilon}{W}$. Hence (5.27), which implies (5.24), holds.

It is awkward to deal with $\hat{\lambda}_I^*(x)$ to prove Theorem 5.3.1 since $E_n(\exp(z\hat{\beta}), x)$ is in the denominator of $\hat{\lambda}_I^*(x)$. Hence we shall therefore define an alternative estimator which has the same asymptotic properties and is easier to handle.

For simplicity, let us define

$$\omega_0(x_j) = (t_{n,j+1} - t_{n,j})E(\exp(z\beta), x_j)$$

and

$$
\begin{aligned}
\tilde{\lambda}_I^{*[A,B]}(x) &= \min_{\substack{s\geq i+1\\ t_{n,s}\leq B}}\ \max_{\substack{r\leq i\\ t_{n,r}\geq A}} \frac{F_n(t_{n,s}) - F_n(t_{n,r})}{\sum_{j=r}^{s-1}(t_{n,j+1}-t_{n,j})E(\exp(z\beta), x_j)}\\
&= \min_{\substack{s\geq i+1\\ t_{n,s}\leq B}}\ \max_{\substack{r\leq i\\ t_{n,r}\geq A}} \frac{F_n(t_{n,s}) - F_n(t_{n,r})}{\sum_{j=r}^{s-1}\omega_0(x_j)}\\
&= \min_{\substack{s\geq i+1\\ t_{n,s}\leq B}}\ \max_{\substack{r\leq i\\ t_{n,r}\geq A}} \frac{\sum_{j=r}^{s-1}Y_n(x_j)\omega_0(x_j)}{\sum_{j=r}^{s-1}\omega_0(x_j)},
\end{aligned}
$$

From (5.11), we define

$$Y_n = Y_n(x) = \frac{F_n(t_{n,i}) - F_n(t_{n,i+1})}{(t_{n,j+1} - t_{n,j})E(\exp(z\beta), x_j)}.$$

Our first objective is to show that Theorem 5.2.1 holds with $\tilde{\lambda}_I^{*[A,B]}(x)$ instead of $\hat{\lambda}_I^*(x)$. We have shown that $Y_n(x)$ has the same asymptotic distribution as $\hat{\lambda}_0^*(x)$ in Lemma 5.2.1. If we can show that for $0 < \alpha < \frac{1}{3}$

$$\sup_{A \le x_r < x < x_s \le B} n^{\frac{1-\alpha}{2}} \left| \frac{F_n(t_{n,s}) - F_n(t_{n,r})}{\sum_{j=r}^{s-1} \omega(x_j)} - \frac{F_n(t_{n,s}) - F_n(t_{n,r})}{\sum_{j=r}^{s-1} \omega_0(x_j)} \right| \to 0 \tag{5.36}$$

in probability as n goes to ∞, then it suffices to show

$$\lim_{n \to 0} \Pr[\tilde{\lambda}_I^{*[A,B]}(x) \ne Y_n(x)] = 0$$

to prove Theorem 5.3.1. The above would follow because

$$\begin{aligned} \tilde{\lambda}_I^{*[A,B]}(x) &= \min_{\substack{s \ge i+1 \\ t_{n,s} \le B}} \max_{\substack{r \le i \\ t_{n,r} \ge A}} \frac{F_n(t_{n,s}) - F_n(t_{n,r})}{\sum_{j=r}^{s-1} \omega_0(x_j)} \\ &= \min_{\substack{s \ge i+1 \\ t_{n,s} \le B}} \max_{\substack{r \le i \\ t_{n,r} \ge A}} \left[\frac{F_n(t_{n,s}) - F_n(t_{n,r})}{\sum_{j=r}^{s-1} \omega(x_j)} \right. \\ &\quad \left. + \left\{ \frac{F_n(t_{n,s}) - F_n(t_{n,r})}{\sum_{j=r}^{s-1} \omega_0(x_j)} - \frac{F_n(t_{n,s}) - F_n(t_{n,r})}{\sum_{j=r}^{s-1} \omega(x_j)} \right\} \right] \end{aligned} \tag{5.37}$$

Therefore, (5.40) would imply that

$$n^{\frac{1-\alpha}{2}} |\hat{\lambda}_I^{*[A,B]}(x) - \tilde{\lambda}_I^{*[A,B]}(x)| \to 0 \tag{5.38}$$

in probability as n goes to ∞.

Lemma 5.3.3 *For* $0 < \alpha < \frac{1}{3}$

$$\sup_{A \le x_r < x < x_s \le B} n^{\frac{1-\alpha}{2}} \left| \frac{F_n(t_{n,s}) - F_n(t_{n,r})}{\sum_{j=r}^{s-1} \omega(x_j)} - \frac{F_n(t_{n,s}) - F_n(t_{n,r})}{\sum_{j=r}^{s-1} \omega_0(x_j)} \right| \to 0 \tag{5.39}$$

in probability as n goes to ∞.

Proof of Lemma 5.3.3 *Note that*

$$n^{\frac{1-\alpha}{2}} \left| \frac{F_n(t_{n,s}) - F_n(t_{n,r})}{\sum_{j=r}^{s-1} \omega(x_j)} - \frac{F_n(t_{n,s}) - F_n(t_{n,r})}{\sum_{j=r}^{s-1} \omega_0(x_j)} \right|$$

$$= n^{\frac{1-\alpha}{2}} |F_n(t_{n,s}) - F_n(t_{n,r})| \frac{|\sum_{j=r}^{s-1} \omega(x_j) - \sum_{j=r}^{s-1} \omega_0(x_j)|}{\sum_{j=r}^{s-1} \omega(x_j) \sum_{j=r}^{s-1} \omega_0(x_j)}$$

$$\le n^{\frac{1-\alpha}{2}} |F_n(t_{n,s}) - F_n(t_{n,r}) - F(t_{n,s}) + F(t_{n,r})| \frac{|\sum_{j=r}^{s-1} \omega(x_j) - \sum_{j=r}^{s-1} \omega_0(x_j)|}{\sum_{j=r}^{s-1} \omega(x_j) \sum_{j=r}^{s-1} \omega_0(x_j)}$$
$$+ n^{\frac{1-\alpha}{2}} |F(t_{n,s}) - F(t_{n,r})| \frac{|\sum_{j=r}^{s-1} \omega(x_j) - \sum_{j=r}^{s-1} \omega_0(x_j)|}{\sum_{j=r}^{s-1} \omega(x_j) \sum_{j=r}^{s-1} \omega_0(x_j)}$$

$$= I + II$$

Since $\sqrt{n}(E_n(\exp(z\hat{\beta}), x) - E(\exp(z\beta), x))$ *converges to a Gaussian process (Tsiatis, 1981),*

$$\left| \frac{1}{s-r} \sum_{j=r}^{s-1} \sqrt{n}[E_n(\exp(z\hat{\beta}), x) - E(\exp(z\beta), x)] \right|$$

$$\le \sup_{A \le x \le B} |\sqrt{n}|E_n(\exp(z\hat{\beta}), x) - E(\exp(z\beta), x)|$$

$$= O_p(1). \tag{5.40}$$

Note that since

$$\sum_{j=r}^{s-1} \omega(x_j) \to \sum_{j=r}^{s-1} \omega_0(x_j) \tag{5.41}$$

in probability as n goes to ∞ as (3.11), where $d \leq E(\exp(z\beta), x) \leq D$ for positive constants d and D, it follows that

$$cn^{-\alpha}(s-r)d \leq \sum_{j=r}^{s-1} \omega_0(x_j) \leq cn^{-\alpha}(s-r)D$$

and

$$cn^{-\alpha}(s-r)d \leq \sum_{j=r}^{s-1} \omega(x_j) \leq cn^{-\alpha}(s-r)D.$$

Next, using (5.40) we observe that

$$\begin{aligned} |\sum_{j=r}^{s-1} \omega(x_j) - \sum_{j=r}^{s-1} \omega_0(x_j)| &= cn^{-\alpha-\frac{1}{2}} |\sum_{j=r}^{s-1} \sqrt{n}(E_n(\exp(z\hat{\beta}), x) - E(\exp(z\beta), x))| \\ &= cn^{-\alpha-\frac{1}{2}}(s-r)|\frac{1}{s-r}\sum_{j=r}^{s-1} \sqrt{n}(E_n - E)| \\ &= c(s-r)n^{-\alpha-\frac{1}{2}} O_p(1) \end{aligned} \tag{5.42}$$

and $\sqrt{n}(F_n(x) - F(x)) = O_p(1)$.

Using the above results, the first term (I) becomes

$$\begin{aligned} (I) &\leq n^{\frac{-\alpha}{2}} O_p(1) \frac{c(s-r)n^{-\alpha-\frac{1}{2}} O_p(1)}{c^2(s-r)^2 n^{-2\alpha} d^2} \\ &= n^{\frac{\alpha-1}{2}} \frac{O_p(1)}{c(s-r)d^2} \\ &\leq n^{\frac{\alpha-1}{2}} \frac{O_p(1)}{cd^2}, \end{aligned} \tag{5.43}$$

which converges to 0 in probability as n goes to ∞.

$$F(t_{n,s}) - F(t_{n,r}) = (t_{n,s} - t_{n,r})|F^{(1)}(t^*)|$$

where t^* is between $t_{n,s}$ and $t_{n,r}$ and

$$t_{n,s} - t_{n,r} = (s-r)cn^{-\alpha},$$

Similarly, we show that the last term (II) converges to 0 in probability as follows;

$$\begin{aligned}(II) &\leq n^{\frac{-\alpha+1}{2}}(s-r)cn^{-\alpha}|F^{(1)}(t^*)|\frac{c(s-r)n^{-\alpha-\frac{1}{2}}O_p(1)}{(s-r)^2d^2c^2n^{-2\alpha}} \\ &= n^{\frac{-\alpha}{2}}|F^{(1)}(t^*)|O_p(1) \qquad (5.44)\end{aligned}$$

which converges to 0 in probability as n goes to ∞.

Using the previous lemmas we shall prove for fixed x,

$$\lim_{n\to\infty} \Pr[Y_n(x) \neq \tilde{\lambda}_I^{*[A,B]}(x)\} = 0$$

Proof of Theorem 5.3.1 *Throughout the proof of Theorem 5.3.1, we assume the range of time is bounded, since we have proven that the isotonic estimators of $\lambda_0(x)$ are equivalent in situations where the range of time is either finite or infinite. For fixed $x_i \in [t_{n,i}, t_{n,i+1})$ we define the isotonic estimator,*

$$\tilde{\lambda}_I^{*[A,B]}(x_i) = \min_{\substack{s\geq i+1 \\ t_{n,s}\leq B}} \max_{\substack{r\leq i \\ t_{n,r}\geq A}} \frac{\sum_{j=r}^{s-1} Y_n(x_j)\omega_0(x_j)}{\sum_{j=r}^{s-1}\omega_0(x_j)}.$$

Since

$$\left\{ \min_{\substack{s \geq i+1 \\ t_{n,s} \leq B}} \max_{\substack{r \leq i \\ t_{n,r} \geq A}} \frac{\sum_{j=r}^{s-1} Y_n(x_j)\omega_0(x_j)}{\sum_{j=r}^{s-1} \omega_0(x_j)} \neq Y_n(x_i) \right\}$$

implies either

$$\left\{ \exists\ m > 1 \ni I_{[A \leq t_{n,r} < t_{n,s} \leq B]} \frac{\sum_{j=i}^{m+i} Y_n(x_j)\omega_0(x_j)}{\sum_{j=i}^{m+i} \omega_0(x_j)} < Y_n(x_i) \right\}$$

or

$$\left\{ \exists\ m > 1 \ni I_{[A \leq t_{n,r} < t_{n,s} \leq B]} \frac{\sum_{j=i-m}^{i} Y_n(x_j)\omega_0(x_j)}{\sum_{j=i-m}^{i} \omega_0(x_j)} > Y_n(x_i) \right\}$$

or both, it suffices to show

$$\lim_{n \to \infty} \Pr \left\{ \exists\ m > 1 \ni I_{[A \leq t_{n,r} < t_{n,s} \leq B]} \frac{\sum_{j=i}^{m+i} Y_n(x_j)\omega_0(x_j)}{\sum_{j=i}^{m+i} \omega_0(x_j)} < Y_n(x_i) \right\} = 0 \qquad (5.45)$$

and

$$\lim_{n \to \infty} \Pr \left\{ \exists\ m > 1 \ni I_{[A \leq t_{n,r} < t_{n,s} \leq B]} \frac{\sum_{j=i-m}^{i} Y_n(x_j)\omega_0(x_j)}{\sum_{j=i-m}^{i} \omega_0(x_j)} > Y_n(x_i) \right\} = 0 \qquad (5.46)$$

Due to the similarity of the proofs of (5.44) and (5.45), we shall only derive (5.44).

Since

$$\left\{ \exists\ m > 1 \ni I_{[A \leq t_{n,r} < t_{n,s} \leq B]} \frac{\sum_{j=i}^{m+i} Y_n(x_j)\omega_0(x_j)}{\sum_{j=i}^{m+i} \omega_0(x_j)} < Y_n(x_i) \right\}$$
$$= \bigcup_{m=1} \left\{ I_{[A \leq t_{n,r} < t_{n,s} \leq B]} \frac{\sum_{j=i}^{m+i} Y_n(x_j)\omega_0(x_j)}{\sum_{j=i}^{m+i} \omega_0(x_j)} < Y_n(x_i) \right\},$$

we shall show

$$\lim_{n\to\infty}\sum_{m=1}^{\infty}\Pr\{I_{[A\le t_{n,r}<t_{n,s}\le B]}\sum_{j=i}^{m+i}\frac{Y_n(x_j)\omega_0(x_j)}{\sum_{j=i}^{m+i}\omega_0(x_j)}<Y_n(x_i)\}=0. \qquad (5.47)$$

Let us define

$$p_j=\frac{\omega_0(x_j)}{\sum_{j=i}^{m+i}\omega_0(x_j)}.$$

By Chebyshev's theorem, it follows that

$$\begin{aligned}\Pr\{\sum_{j=i}^{m+i}\frac{Y_n(x_j)\omega_0(x_j)}{\sum_{j=i}^{m+i}\omega_0(x_j)}<Y_n(x_i)\} &= \Pr\{\sum_{j=i}^{m+i}Y_n(x_j)p_j<Y_n(x_i)\}\\ &= \Pr\{\sum_{j=i}^{m+i}(Y_n(x_j)-Y_n(x_i))p_j<0\}\\ &\le \frac{Var[\sum_{j=i}^{m+i}p_j(Y_n(x_j)-Y_n(x_i))]}{\{\sum_{j=i}^{m+i}p_jE[Y_n(x_j)-Y_n(x_i)]\}^2} \qquad (5.48)\end{aligned}$$

First, let us simplify $Var[\sum_{j=i}^{m+i}p_j(Y_n(x_j)-Y_n(x_i))]$.

It is well-known that

$$Var[\sum_{j=i}^{m+i}p_j(Y_n(x_j)-Y_n(x_i))]\le 3(Var[\sum_{j=i}^{m+i}p_jY_n(x_j)]+Var[\sum_{j=i}^{m+i}p_jY_n(x_i)]).$$

$Var[\sum_{j=i}^{m+i}p_jY_n(x_j)]$ *is simplified as*

$$\begin{aligned}Var[\sum_{j=i}^{m+i}p_jY_n(x_j)] &= Var[\sum_{j=i}^{m+i}\frac{\omega_0(x_j)}{\sum_{j=i}^{m+i}\omega_0(x_j)}\frac{F_n(t_{n,j})-F_n(t_{n,j+1})}{\omega_0(x_j)}]\\ &= Var[\frac{F_n(t_{n,i})-F_n(t_{n,m+i})}{\sum_{j=i}^{m+i}\omega_0(x_j)}]\end{aligned}$$

$$\leq \{\frac{1}{\sum_{j=i}^{m+i}\omega_0(x_j)}\}^2\frac{1}{n}\{F(t_{n,i}) - F(t_{n,m+i})\}$$

$$\leq \frac{1}{n(m+1)^2 d^2 n^{-2\alpha} c^2} mcn^{-\alpha}|F^{(1)}(x*)|$$

$$\leq \frac{c_1}{n^{1-\alpha}} \tag{5.49}$$

for some constant c_1 *where* $d \leq E(\exp(z\beta), x) \leq D$ *for some positive constants* d *and* D*. Similarly* $Var[\sum_{j=i}^{m+i} p_j Y_n(x_i)]$ *is also simplified from (5.15) as*

$$\begin{aligned} Var[\sum_{j=i}^{m+i} p_j Y_n(x_i)] &= (\sum_{j=i}^{m+i} p_j)^2 Var[Y_n(x_i)] \\ &\leq \frac{c_2}{n^{1-\alpha}}. \end{aligned} \tag{5.50}$$

Using (5.13), and noting

$$P_{m+j}^2 = \left\{\frac{\omega_0(x_{m+j})}{\sum_{j=i}^{m+i}\omega_0(x_j)}\right\}^2 \geq \frac{B}{m^2}$$

for some constant $B > 0$ *and denoting [ ] the greatest integer less than the quantity within the bracket, it follows that*

$$\begin{aligned} \{\sum_{j=i}^{m+i} p_j E[Y_n(x_j) - Y_n(x_i)]\}^2 &\geq \{\sum_{j=i+[\frac{m}{2}]}^{m+i} p_j E[Y_n(x_j) - Y_n(x_i)]\}^2 \\ &= P_{m+j}^2 E^2[Y_n(x_{i+[\frac{m}{2}]}) - Y_n(x_i)]([\frac{m}{2}])^2 \end{aligned}$$

$$\geq \frac{B}{m^2} n^{-2\alpha} c_3^2 n^{-2\alpha} m^2 ([\frac{m}{2}])^2, \qquad (5.51)$$

for some constant, c_3.

Hence by (5.48), (5.49) and (5.50),

$$\Pr\{\sum_{j=i}^{m+i} \frac{Y_n(x_j)\omega_0(x_j)}{\sum_{j=i}^{m+i}\omega_0(x_j)} < Y_n(x_i)\} \leq \frac{c_4}{n^{1-3\alpha} m^2}$$

for some constant c_4. *It follows that for* $0 < \alpha < \frac{1}{3}$

$$\lim_{n\to\infty} \sum_{m=1}^{\infty} \Pr\{\sum_{j=i}^{m+i} \frac{Y_n(x_j)\omega_0(x_j)}{\sum_{j=i}^{m+i}\omega_0(x_j)} < Y_n(x_i)\} = 0.$$

Hence it follows that the asymptotic distributions of $\hat{\lambda}_I^*(x)$ and $\hat{\lambda}_0^*(x)$ are identical. The following result is an immediate consequence of Theorem 5.3.1.

<u>*Theorem 5.3.2*</u> *Under the same assumptions as in Theorem 5.3.1, for* $\frac{1}{5} < \alpha < \frac{1}{3}$

$$[\frac{-c}{F^{(1)}(x)}]^{\frac{1}{2}} E(\exp(z\beta), x) n^{\frac{1-\alpha}{2}} (\hat{\lambda}_I^*(x) - \lambda_0(x))$$

and for $\frac{1}{7} < \alpha \leq \frac{1}{5}$

$$[\frac{-c}{F^{(1)}(x)}]^{\frac{1}{2}} E(\exp(z\beta), x) n^{\frac{1-\alpha}{2}} \left(\hat{\lambda}_I^*(x) - \lambda_0(x) - \frac{c^2 n^{-2\alpha} F^{(3)}(x)}{24 E(\exp(z\beta), x)}\right)$$

are asymptotically distributed as $N(0,1)$.

<u>*Proof of Theorem 5.3.2*</u> *This is an immediate consequence of Theorems 5.2.1 and 5.3.1.*

5.4 Determination of Window Size

When we discuss estimation based on windows, the following question arises, "What is the optimal size of the window?" The answer determines how many windows we can have in order to obtain the smoothness of the baseline hazard function, taking into account the mean square error of estimates of the baseline hazard function. The recommendation on window size is made in terms of the mean square error (Barlow et al., 1972).

Parzen(1961) discusses the problem for choosing the size of the window in density estimation. In his paper, he uses the following estimate of the density, $f(x)$,

$$f_n(x) = \frac{F_n(x+h) - F_n(x-h)}{2h}$$

where h is a suitably chosen positive number. After reviewing the statistical properties of $f_n(x)$, especially the mean square error of $f_n(x)$, he finds the value of h which minimizes the mean square error for a fixed value of n. From (5.13) and (5.15) we can see that the mean square error of $\hat{\lambda}_I^*(x)$ in estimating $\lambda_0(x)$ is

$$\begin{aligned} \text{MSE}[\hat{\lambda}_I^*(x)] &= \text{Var}[\hat{\lambda}_0^*(x)] + \text{Bias}^2[\hat{\lambda}_I^*(x)] \\ &\approx \frac{F^{(1)}(x)}{cE^2(\exp(z\beta), x)} n^{\alpha-1} + O(n^{-4\alpha}) \qquad (5.52) \end{aligned}$$

If $\text{MSE}^{\frac{1}{2}}[\hat{\lambda}_I^*(x)]$ is asymptotically smaller than the window size $O(n^{-\alpha})$, then it is intuitively clear that asymptotically $\hat{\lambda}_I^*(t_{n,i}) < \hat{\lambda}_I^*(t_{n,i+1})$. This follows because $|\lambda_0(t_{n,i+1}) - \lambda_0(t_{n,i})| = O(n^{-\alpha})$ (when $\lambda_0(x)$ has a positive first derivative in a neighborhood of x).

Clearly, $\mathrm{MSE}^{\frac{1}{2}}[\hat{\lambda}_I^*(x)] = O(n^{-\alpha})$ if $\alpha < \frac{1}{3}$. This follows from

$$\sigma[\hat{\lambda}_I^*(x)] = \mathrm{Var}^{\frac{1}{2}}[\hat{\lambda}_I^*(x)] = O(n^{\frac{\alpha-1}{2}}) = O(n^{-\alpha})$$

provided $\frac{\alpha-1}{2} < -\alpha$, or $\alpha < \frac{1}{3}$.

Returning to our problem, we can show the mean square error of $\hat{\lambda}_I^*(x)$ is minimized when we choose $\alpha = \frac{1}{5}$. To see this, recall from (5.13) and (5.15) that the mean square error is approximately

$$\mathrm{MSE}[\hat{\lambda}_I^*(x)] \approx \frac{\lambda_0^2(x)n^{\alpha-1}}{cF^{(1)}(x)} + \left[\frac{c^2F^{(3)}(x)}{24E(\exp(z\beta), x)}\right]^2 n^{-4\alpha}$$

The minimum is achieved by choosing $\alpha = \frac{1}{5}$, where MSE is treated as a function of α. Unfortunately, the optimal size of the window still depends on the value of c. To solve this problem, if we choose $t_{n;i} = y_{u(i)}$ by the principle of maximum likelihood, the order statistics from a sample size of n , we obtain the same estimator as one derived in Chapter 3. We modify the window size by choosing $t_{n;i} = y_{u([in^{\frac{4}{5}}])}$ where [ ] denotes the greatest integer of the quantity within the brackets. this yields

$$y_{u([(i+1)n^{\frac{4}{5}}])} - y_{u([in^{\frac{4}{5}}])} = O_p(\frac{n^{\frac{4}{5}}}{n}) = O_p(n^{\frac{-1}{5}}),$$

so that the recommended requirement that the mean square error as a function of α is minimized at $\alpha = \frac{1}{5}$ is satisfied.

CHAPTER 6
SIMULATION AND CONCLUSION

Suppose we have a single binary covariate model (two different groups of cancer patients e.g., male and female), and we are interested in the instantaneous failure rate at a given time conditional upon survival up to this given time. In terms of Cox's regression model, it is enough to consider the regression parameter and the baseline hazard function. This chapter contrasts the three different estimators of the baseline hazard function which are found in Chapter 3 and Chapter 5.

6.1 Estimators

There are three estimators of $\lambda_0(t)$ for fixed t. We assume that $\lambda_0(t)$ increases monotonically, and $\hat{\beta}$ is the maximum likelihood estimator of β obtained from marginal likelihood.

(i) The first estimator of $\lambda_0(t)$, denoted by E_1, is obtained using Breslow's approach (1974) which approximates an increasing function by a nondecreasing step function. The joint likelihood of β and the λ_i's is used. The maximum likelihood estimator of $\lambda_0(t)$ is

$$\hat{\lambda}_i = \frac{1}{(y_{u(j)} - y_{u(j-1)}) \sum_{l \in R(y_{u(j)})} \exp(z_l \hat{\beta})}. \tag{6.1}$$

(ii) The second estimator of $\lambda_0(t)$, denoted by E_2, is derived using the isotonic regression method with ordered failure observations. The isotonic estimator of $\lambda_0(t)$ is

$$\hat{\lambda}_0^* = \max_{s \le i} \min_{t \ge i} (t - s + 1) \Big/ \sum_{s}^{t} (y_{u(j)} - y_{u(j-1)}) \sum_{l \in R(y_{u(j)})} \exp(z_l \hat{\beta}). \tag{6.2}$$

(iii) The final estimator of $\lambda_0(t)$, denoted by E_3, is also derived in a similar fashion to the second estimator, except that a window is used with ordered failure observations. The isotonic estimator based on the window is

$$\hat{\lambda}_I^*(x) = \min_{s \geq i+1} \max_{r \leq i} \frac{F_n(t_{n,r}) - F_n(t_{n,s})}{\sum_{j=r}^{j=s-1}(t_{n,j+1} - t_{n,j})E_n(\exp(z\hat{\beta}), x_j)} \qquad (6.3)$$

where $E_n(\exp(z\hat{\beta}), x) = \frac{1}{n}\sum_{i=1}^{n} \exp(z_i\hat{\beta}) I_{[T_i \geq x]}$.

6.2 Procedure for the Simulation

The outline of the simulation is as follows:

(1) Generate four sets of 50 random numbers from a uniform distribution on (0,1), and call them $U1_i, U2_i, U3_i$ and $U4_i$ for $i = 1, \cdots, 50$ (see Table 6.1).

(2) Obtain the survival and censor data $Y1_i, Y2_i, Y3_i$ and $Y4_i$ by converting $U1_i, U2_i, U3_i$ and $U4_i$ using $Yk_i = F^{-1}(Uk_i)$ for $k = 1, \cdots, 4$, where

$$F(x) = \exp(-x^2e^2) \qquad \text{for } U1_i$$

$$F(x) = \exp(-x) \qquad \text{for } U2_i$$

$$F(x) = \exp(-x^2) \qquad \text{for } U3_i$$

and

$$F(x) = \exp(-2x) \qquad \text{for } U4_i \qquad i = 1, \cdots, 50$$

respectively.

Note that Cox's regression model is given as $\lambda(x; z) = 2x\exp(2z)$. When $F(x) = \exp(-x^2e^2)$, the corresponding Cox regression model is $\lambda(x; 1) = 2x\exp(2)$. When $F(x) = \exp(-x^2)$, the corresponding Cox regression model is $\lambda(x; 0) = 2x$. The baseline hazard function is $2t$ which increases monotonically in t.

(3) Form Group 1 with $Y1_i$ in such a way that censored data is created comparing $Y1_i$ with $Y2_i$. If $Y1_i > Y2_i$, then $Y1_i$ is a censored observation, otherwise $Y1_i$ is an uncensored observation. Similarly form Group 2 with $Y3_i$ and $Y4_i$.

(4) Obtain the maximum likelihood estimator E_1 of $\lambda_0(t)$ using (6.1) and the isotonic estimator E_2 based on the ordered observations using (6.2).

(5) Obtain the isotonic estimator E_3 with the optimal size of window using (6.3). The optimal size of window is determined adjusting the order statistics so that

$$\text{size of the } i \text{ th window } = y_{u([(i)\frac{n^{\frac{4}{5}}}{c}])} - y_{u([(i-1)\frac{n^{\frac{4}{5}}}{c}])}$$

for appropriate $c > 0$. After completing 100 pilot simulations with the values of $c = 36, 33, 30, 27$ and 24, we found that the mean square error is minimized at $c = 36$. Since the sample size of each simulation is 100, the window size is $n^{\frac{4}{5}}/36 = 1.105$. This implies that most of observations must be used to obtain the isotonic estimator which gives the minimum mean square error.

(6) Repeat the steps (1) to (5) 1000 times with $c = 36$. Noting that all estimators are functions of time t, obtain the three estimators at $p(i)$ where $p(i)$ is defined

$$p(i) = \sqrt{\log(0.95 - 0.05(i-1))}$$

for $i = 1, \cdots, 19$. In other words, $p(i)$ is the $100(0.95 - 0.05(i-1))$ percentile of the baseline survival function

$$S_0(t) = \exp(-t^2).$$

(7) Find the mean square error of E_1, E_2 and E_3. We define the mean square error by

$$MSE[\theta] = \frac{1}{1000}\sum_{j=1}^{1000}(E_{ij}(t) - \theta)^2 \qquad \text{for} \qquad i = 1, 2, 3,$$

where the values of t are $p(i)'s$. Note that $\lambda_0(t) = 2t$ is assumed.

(8) Find the relative efficiencies of E_2 and E_3 versus E_1. We define the relative efficiency as the ratio of mean square errors of two different estimators for fixed $p(i)$. E_2 and E_3 are the isotonic estimators while E_1 is the maximum likelihood estimator. Table 6.1 gives the relative efficiencies of the isotonic estimators as compared to the maximum likelihood estimator. (Also, see Figure 6.1.)

(9) To investigate the relative efficiencies of E_2 and E_3 over E_1 for extended cases, let us assume that for $r > 1$

$$\lambda_0(x) = rx^{r-1}$$

so that the baseline survival function is

$$S_0(x) = \exp(-x^r).$$

(10) Repeat the steps (2) and (3) using

$$F(x) = \exp(-x^r e^2) \qquad \text{for } U1_i$$

$$F(x) = \exp(-x) \qquad \text{for } U2_i$$

$$F(x) = \exp(-x^r) \qquad \text{for } U3_i$$

and

$$F(x) = \exp(-2x) \qquad \text{for } U4_i$$

(11) Find the maximum likelihood estimator and the isotonic estimator with data generated by the step (10).

(12) As in step (6), obtain the three estimators of $\lambda_0(t)$ at fixed $p(i)$ which is defined by

$$p(i) = \{\log(0.95 - 0.05(i-1))\}^{\frac{1}{r}}$$

for $i = 1, \cdots, 19$ and $r > 1$.

(13) Repeat the steps (7) and (8) 1000 times with $r = 1.0, 1.5, 2.5$ and 3.0.

(14) The relative efficiencies E_2 to E_1 and E_3 to E_1 for given values of r are presented in terms of the probabilities that an individual survives up to time t. Hence we can see the validity of the relative efficiency measure as the survival probabilities vary. When $i = 1$, the relative efficiency is important because the probability of survival to time $p(1)$ is 95%. When $i = 19$, the efficiency is not meaningful, since the probability that an individual survives up to time $p(19)$ is only 5%. (See Tables 6.2–6.5 and Figures 6.2–6.5.)

6.3 Conclusion

The sum of mean square errors over t values is applied to determine the better estimators. We discuss the general findings of the simulation in this section.

The main problem is to determine the optimal size of the window which yields the minimum of mean square errors of E_3 for fixed t values. (Refer to Section 5.4.) We assume that for $0 < \alpha < 1$ and some positive constant $c > 0$,

$$t_{n,i+1} - t_{n,i} = cn^{-\alpha} \tag{6.4}$$

to derive E_3. It turns out that the window size must be proportional to $n^{-\frac{1}{5}}$ where n is the sample size. Because the constant c is unknown, we have to modify the optimal size of the window which depends on the ordered observations.

Using definition (6.3), we consider the isotonic estimators based on the window only when the right limit of the last window is less than the largest failure time. We do not have estimators between the right limit of the last window and the largest failure time, where the isotonic estimators E_3 are not defined. We are interested in three estimators of $\lambda_0(t)$ when $\lambda_0(t)$ increases monotonically. We anticipate that E_2 and E_3 are better estimators than E_1, since E_2 and E_3 are obtained using the assumption that $\lambda_0(t)$ increases monotonically. We define the relative efficiency of E_i versus E_1 by

$$R_i = \frac{\text{sum of squared errors of } E_1}{\text{sum of squared errors of } E_i}$$

for $i = 2, 3$. The relative efficiency is presented in terms of the probabilities that an individual survives up to time t. Hence we can see the validity of the efficiency as the survival probability varies. For example, when $i = 19$, the efficiency draws little attention, since the probability that one survives up to time $p(19)$ is only 5%.

The isotonic regression method for estimating $\lambda_0(t)$ using the window outperforms the maximum likelihood estimator over whole range, when $\lambda_0(t)$ is assumed to be an increasing function. But the isotonic regression method based upon the ordered failure observations is better than the maximum likelihood estimator up to the time when an individual can survive with probability more than 0.5, while it is just as good as the maximum likelihood estimator when the probability of survival is at most 0.5 in terms of the baseline survival function. The isotonic regression method with a window is more efficient than the isotonic regression method without a window when the survival chance of an individual is at most 0.5, while the isotonic regression method with a window does not show any significant improvement over the isotonic regression method without a window when the survival chance of an individual is greater than 0.5.

We conclude that among the three estimators of the baseline hazard function of Cox's regression model, the isotonic regression method with a window is the most efficient estimator for the whole range.

Table 6.2. Relative Efficiencies of E_2 to E_1 and E_3 to E_1 when r=1.0

Time	R_1	R_2	No. of Estimators
1	7.69198	0.62580	1000
2	53.20185	5.86470	1000
3	34.00603	4.76669	1000
4	3.96048	0.59525	1000
5	5.14114	0.73327	1000
6	1.91799	0.29948	1000
7	2.03008	0.36398	1000
8	3.59762	0.92791	1000
9	3.77598	1.15719	1000
10	2.11634	0.71342	999
11	1.82748	0.76406	995
12	0.96547	0.55186	964
13	0.83177	0.69434	896
14	0.84174	0.68594	756
15	0.64364	0.26801	562
16	0.96873	2.97194	332
17	0.89445	0.15094	163
18	0.17965	1.4E-02	50
19	0.14717	2.5E-03	6

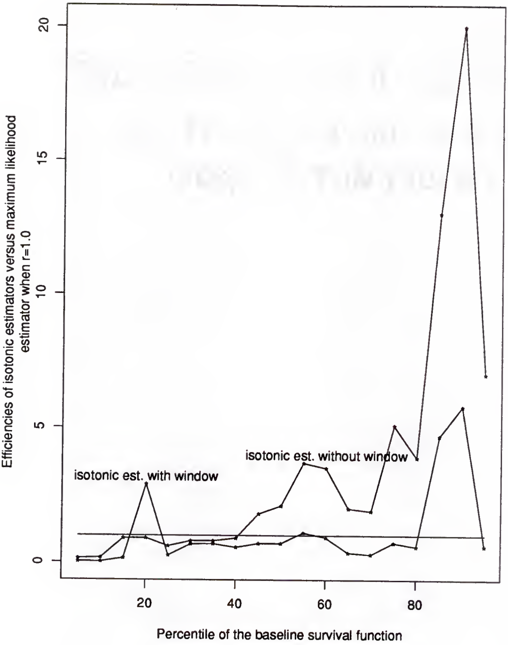

Figure 6.1. Efficiencies of E_2 and E_3 versus E_1 when $r = 1.0$

Table 6.2. Relative Efficiencies of E_2 to E_1 and E_3 to E_1 when r=1.5

Time	R_1	R_2	No. of Estimators
1	1.93105	0.51494	1000
2	2.84304	1.00692	1000
3	5.45463	2.62717	1000
4	7.35513	4.36878	1000
5	5.53266	3.75795	1000
6	3.87954	3.31796	1000
7	2.84534	2.80851	1000
8	7.21880	7.90580	1000
9	1.92854	3.22761	1000
10	8.94882	29.75328	998
11	2.99180	8.50156	990
12	1.14700	4.21029	954
13	0.72456	4.32069	905
14	1.08323	7.25389	805
15	0.94693	22.04793	664
16	0.74733	3.67969	491
17	1.02179	18.34358	314
18	0.81902	1.54799	151
19	1.70782	0.24051	44

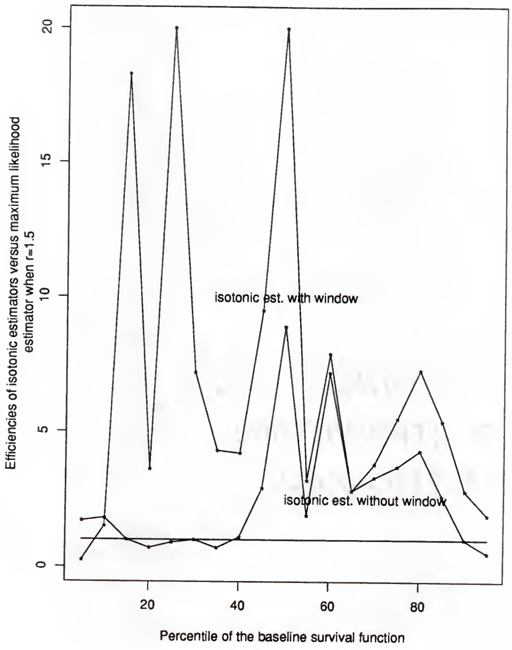

Figure 6.2. Efficiencies of E_2 and E_3 versus E_1 when $r = 1.5$

Table 6.3. Relative Efficiencies of E_2 to E_1 and E_3 to E_1 when r=2.0

Time	R_1	R_2	No. of Estimators
1	12.19570	4.14622	1000
2	2.67036	1.05323	1000
3	3.36746	1.58910	1000
4	2.11647	1.30644	1000
5	2.57971	1.83293	1000
6	2.75637	2.44174	1000
7	6.74682	7.19189	1000
8	21.93844	28.44507	999
9	2.64356	4.48147	999
10	0.95675	2.65322	996
11	1.20568	4.18544	984
12	1.04331	3.78065	950
13	0.82456	4.42452	908
14	1.08746	4.60336	823
15	0.72260	6.67253	713
16	0.94320	6.09248	557
17	1.04300	27.19504	414
18	0.92408	1.28166	231
19	2.68491	0.52304	81

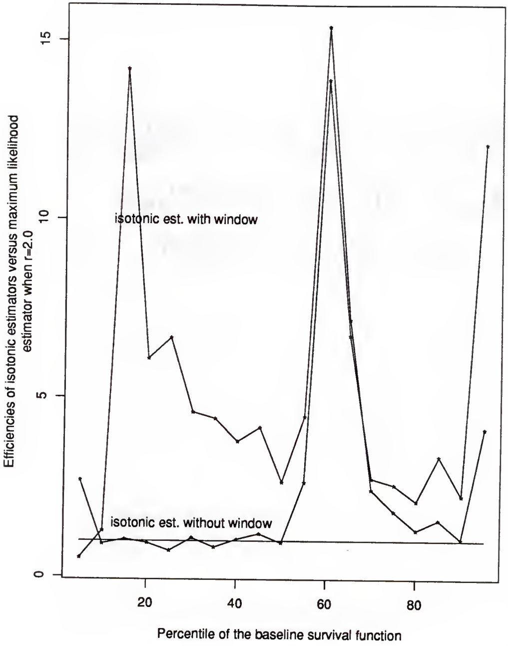

Figure 6.3. Efficiencies of E_2 and E_3 versus E_1 when $r = 2.0$

Table 6.4. Relative Efficiencies of E_2 to E_1 and E_3 to E_1 when r=2.5

Time	R_1	R_2	No. of Estimators
1	4.24194	1.51224	1000
2	2.46923	1.17791	1000
3	4.12568	2.74376	1000
4	2.02903	1.51935	1000
5	13.99279	12.75204	1000
6	20.28677	20.33314	1000
7	3.81523	4.481471	1000
8	5.13366	7.35062	999
9	2.27329	5.34532	999
10	1.42994	6.60936	995
11	1.52489	6.82602	980
12	1.27592	5.57158	948
13	1.02940	23.24822	911
14	0.84960	4.03035	835
15	0.92992	9.11620	735
16	1.10174	12.22548	604
17	1.02400	33.20784	465
18	0.94614	1.09369	290
19	1.20355	2.40673	104

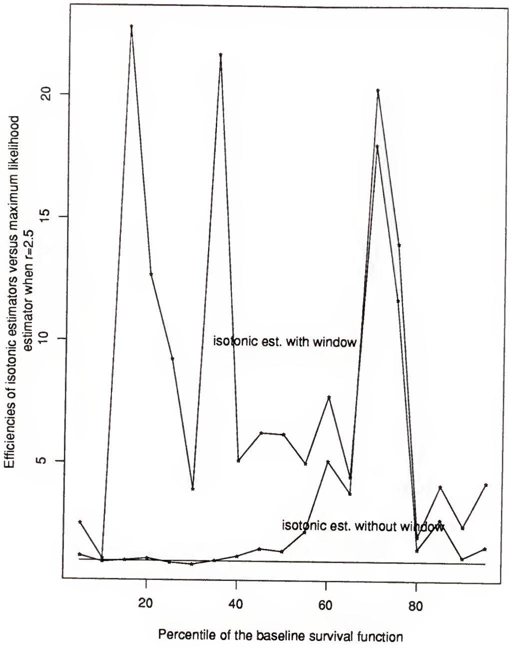

Figure 6.4. Efficiencies of E_2 and E_3 versus E_1 when $r = 2.5$.

Table 6.5. Relative Efficiencies of E_2 to E_1 and E_3 to E_1 when r=3.0

Time	R_1	R_2	No. of Estimators
1	2.46000	1.11768	1000
2	4.58195	2.55058	1000
3	4.03790	3.91666	1000
4	3.81653	4.15003	1000
5	3.00217	3.24918	1000
6	2.07541	2.30397	1000
7	8.53029	10.95093	1000
8	3.38755	6.22937	999
9	2.30834	7.19948	999
10	1.39705	12.99846	993
11	1.14255	6.43423	976
12	1.37777	7.00777	946
13	0.94358	10.39849	912
14	1.62100	12.27069	845
15	0.74918	7.36184	748
16	0.97037	17.39330	635
17	1.02220	37.40968	498
18	0.99905	0.95536	315
19	1.49304	1.39411	127

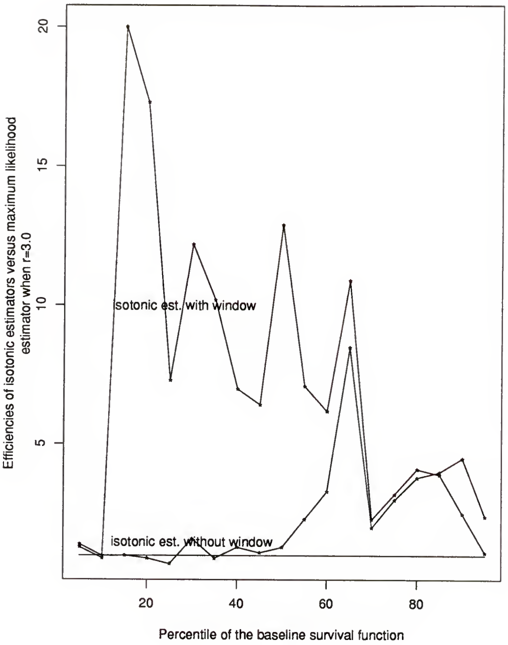

Figure 6.5. Efficiencies of E_2 and E_3 versus E_1 when $r = 3.0$.

APPENDIX A
PROOF OF LEMMA 2.2.1

Lemma 2.2.1

Conditional on $\sum_{i=1}^{n} D_{n;i} = m$, $D_{n;1}, \cdots, D_{n;n-1}$ have a uniform distribution over the area:

$$d_i \geq 0, \qquad i = 1, \cdots, n-1, \qquad \sum_{i=1}^{n-1} d_i \leq m, \tag{A.1}$$

under the null hypothesis that $\lambda_0(t)$ is constant.

Proof of Lemma 2.2.1

It is well known that $D_{n;i}$ $(i = 1, \cdots, n)$ are independent and distributed exponentially with mean $\frac{1}{\lambda}$, ie.,

$$f_{D_{n;i}}(d_i) = \lambda \exp(-\lambda d_i) \qquad i = 1, \cdots, n.$$

Therefore $\sum D_{n;i}$ has a gamma distribution with parameters n and λ, ie.,

$$f_{\sum D_{n;i}}(d) = \frac{1}{\Gamma(n)} \lambda^n d^{n-1} exp(-\lambda d)$$

It is easy to see that the conditional distribution of $D_{n;1}, \cdots, D_{n;n-1}$ given $\sum_{i=1}^{n} D_{n;i} = m$, is

$$f_{D_{n;1}, \cdots, D_{n;n-1} | \sum D_{n;i} = m}(d_1, \cdots, d_{n-1})$$

$$= \frac{f_{D_{n;1}, \cdots, D_{n;n}}(d_1, \cdots, d_{n-1}, m - d_1 - \cdots - d_{n-1})}{f_{\sum D_{n;i}}(m)}$$

$$= \frac{\left[\prod_{i=1}^{n-1} \lambda \exp(-\lambda d_i)\right] \lambda \exp(-\lambda(m - d_1 - \cdots - d_{n-1}))}{\frac{\lambda^n m^{n-1} \exp(-\lambda m)}{(n-1)!}}$$

$$= \frac{(n-1)!}{m^{n-1}}.$$

Lemma 2.2.2

Let $X_i = \frac{D_{n;i}}{\sum_{j=1}^{n} D_{n;j}}$ $i = 1, \cdots, n-1$. Then conditional on $\sum D_{n;i} = m, X_1, \cdots, X_{n-1}$ have a uniform distribution over the area:

$$x_i \geq 0 \qquad i = 1, \cdots, n-1 \qquad x_1 + \cdots + x_{n-1} \leq 1. \tag{A.2}$$

Proof of Lemma 2.2.2

It is seen that Lemma 2.2.2 is a consequence of Lemmma 2.2.1 by a scale change.

APPENDIX B
PROOF OF THEOREM 2.2.2

Theorem 2.2.2

If failure time has an exponential distribution with parameter λ,

$$V_k \underset{\text{st}}{=} \sum_{j=1}^{k-1} U_j \tag{B.1}$$

where V_k is defined in (2.3) and U_j are independent uniform random variables on (0,1) for $j = 1, \cdots, k-1$.

Proof of Theorem 2.2.2

The proof of Theorem 2.2.2 will be limited to the case $k = n$. It can be seen, from the structure of V_k in (2.3) and Theorem 2.2.1 that the proof of Theorem 2.2.2 for $k < n$ is the same as that for $k = n$. It is easy to see that by definition of V_n in (2.2)

$$V_n = (n-1)X_1 + (n-2)X_2 + \cdots + X_{n-1}.$$

Denote the distribution function of the random variable in the right hand side of (B.1) with $k = n$ by G. To finish the proof we shall prove that conditional on $\sum D_{n:i}$, V_n has the same distribution G.

The moment generating function of the random variables of the right hand side of (B.1) with $k = n$ is

$$E[\exp\{t(U_1 + \cdots + U_{n-1})\}] = [\frac{\exp(t) - 1}{t}]^{n-1} \tag{B.2}$$

As we have shown in Lemma 2.2.2, conditional on $\sum D_{n:i}$, $X_1 \cdots X_{n-1}$ have a uniform distribution over the simplex (A.2). In order to prove that conditional on $\sum D_{n:i}$, V_n has the same distribution G, we may assume that $X_1 \cdots X_{n-1}$ have a uniform distribution over the simplex (A.2) and show that $V_n = (n-1)X_1 + (n-2)X_2 + \cdots + X_{n-1}$ has the same moment generating function as (B.1). That is, we must show

$$E[\exp tV_n] = [\frac{\exp(t) - 1}{t}]^{n-1} \tag{B.3}$$

When $n = 2$, (B.3) clearly holds.

By an induction argument, suppose (B.3) holds for $n-1$. We will show that it holds for n. Next we compute

$$\begin{aligned}
E[\exp tV_n] &= E[\exp t\{(n-1)X_1 + (n-2)X_2 + \cdots + X_{n-1})\}] \\
&= EE[\exp t\{(n-1)X_1 + (n-2)X_2 + \cdots + X_{n-1})\}|X_1] \\
&= \int_0^1 E[\exp t\{(n-1)X_1 + (n-2)X_2 + \cdots + X_{n-1})\}|X_1 = x_1] \\
&\quad (1-x_1)^{n-2}(n-1)dx_1 \\
&= \int_0^1 \exp[t(n-1)x_1]E[\exp t\{(n-2)X_2 + \cdots + X_{n-1})\}|X_1] \\
&\quad (1-x_1)^{n-2}(n-1)dx_1 \\
&= \int_0^1 \exp[t(n-1)x_1]E[\exp(t(1-x_1))\{(n-2)\frac{X_2}{(1-x_1)} \\
&\quad + \cdots + \frac{X_{n-1}}{(1-x_1)})\}|X_1](1-x_1)^{n-2}(n-1)dx_1.
\end{aligned}$$

It can be shown that conditional on $X_1 = x_1, \frac{X_2}{(1-x_1)}, \cdots, \frac{X_{n-1}}{(1-x_1)}$ have a uniform distribution over the simplex;

$$y_i \geq 0, \quad i = 2, \cdots, n-1, \qquad \sum_{i=2}^{n-1} y_i \leq 1.$$

By the induction assumption, we have that conditional on $X_1 = x_1$, the moment generating function of $(n-2)\frac{X_2}{(1-x_1)} + \cdots + \frac{X_{n-1}}{(1-x_1)}$ is $[\frac{\exp(t)-1}{t}]^{n-2}$.

Hence, we obtain

$$\begin{aligned}
E[\exp tV_n] &= \int_0^1 \exp[t(n-1)x_1][\frac{\exp[(1-x_1)t]-1}{(1-x_1)t}]^{n-2}(1-x_1)^{n-2}(n-1)dx_1 \\
&= \int_0^1 \exp(tx_1)[\frac{\exp(t)-\exp(x_1 t)}{t}]^{n-2}(n-1)dx_1 \\
&= -[\frac{\exp(t)-\exp(x_1 t)}{t}]^{n-1}\Big|_{x_1=0}^{x_1=1} \\
&= [\frac{\exp(t)-1}{t}]^{n-1}.
\end{aligned}$$

This completes the proof.

REFERENCES

Barlow, R. E., Barthlow, D. M., Bremner, J. M. & Brunk, H. D. (1972). *Statistical Inference Under Order Restrictions*. New York: Wiely.

Barlow, R. E. & Doksum, K. A. (1972). "Isotonic test for convex orderings." *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability* 1, 293–323.

Barlow, R. E. & Proschan, F. (1969). " A note on tests for monotone failure rate based on incomplete data." *Annals of Mathematical Statistics* 40, 595–600.

Barlow, R. E. & van Zwet, W. R. (1969). "Asymptotic properties of isotonic estimators for the generalized failure rate function. Part II: Asymptotic distribution." *Operations Research Center Report ORC* 69 –10, University of California Berkeley.

Bickel, P. J. & Doksum, K. A. (1969). "Tests for monotone failure rate based on normalized spacings." *Annals of Mathematical Statistics* 40, 1216–1235.

Breslow, N. (1974). "Covariance analysis of censored survival data." *Biometrics* 30, 89–99.

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

Chow, Y. S. & Teicher, H. (1978). "*Probability Theory*. New York: Springer–Verlag.

Cox, D. R. (1972). "Regression models and life tables (with discussion)". *Journal of the Royal Statistical Society* 34, 187–220.

Cox, D. R. (1975). "Partial likelihood." *Biometrika* 62, 269–276.

Efron, B. (1977). "The efficiency of Cox's likelihood function for the censored data." *Journal of the American Statistical Association* 72, 557–565.

Epstein, B. (1960). "Tests for the validity of the assumption that the underlying distribution of life is exponential, Part I." *Technometrics* 2, 83–101.

Grenander, U. (1956). "On the theory mortality measurement, Part II." *Skand. Akt.* 39, 125–153.

Kalbfleisch, J. D. & Prentice, R. L. (1973). "Marginal likelihoods based on Cox's regression and life model." *Biometrika* 60, 267–278.

Kalbfleisch, J. D. & Prentice, R. L. (1980). *The Statistical Analysis of Failure Time Data.* New York: Wiley.

Kaplan, E. L., & Meier, P. (1958). "Nonparametric estimation form incomplete observations." *Journal of the American Statistical Association* 53, 457–481.

Marshall, A. W. & Proschan F. (1965). "Maximum likelihood estimation for distributions with monotone failure rate." *Annals of Mathematical Statistics* 36, 69–77.

Meshalkin. L. D. & Kagan, A. R. (1972). Contribution to the discussion on the paper of D. R. Cox (1972).

Oakes. O. (1977). "The asymptotic information in censored survival data." *Biometrika* 64, 441–448.

Parzen, E. (1962). "On estimation of a probability density and mode." *Annals of Mathematical Statistics* 33, 1065–1076.

Proschan, F. & Pyke, R. (1967). *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, 3, 293–311.

Rao, C. R. (1965). *Linear statistical inference and its applications.* New York: Wiley.

Robertson, T., Wright, F. T., & Dystra, R. L. (1988). *Order Restricted Statistical Inference.* New York: Wiely.

Serfling, R. J. (1980). *Approximation Theorems of Mathematical Statistics.* New York: Wiely.

Tsiatis, A. A. (1981). "A large sample study of Cox's regression model." *Annals of Statistics* 9, 93–108.

von Misses, R. (1964). *Mathematical Theory of Probability and Statistics.* New York: Wiely.

BIOGRAPHICAL SKETCH

Daehyun Chung was born on September 11, 1957, in South Korea. He has been a resident of Gainesville, Florida, since 1986. He was awarded a Bachelor of Science degree in mathematics in 1979 and a Master of Statistics degree in 1981, both from the Korea University. Before coming to United States, He had been an officer and teacher in the Department of Mathematics of the Korean Naval Academy for military duty. He also received a master's degree in applied statistics from Bowing Green State University in Bowling Green, Ohio, in 1986. Since then he has been working toward the Ph.D. in statistics from the University of Florida, while serving as a statistical analyst and consultant for the Division of Biostatistics, IFAS, and a teaching assistant.

Mr. Chung had married Gyehee Choi in 1981 and has a daughter, Hyunju, and a son, Yearncharn.

After graduation, Mr. Chung looks forward to teaching and doing research in statistics.

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Myron N. Chang, Chairman
Associate Professor of Statistics

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Jonathan J. Shuster
Professor of Statistics

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Pejaver V. Rao
Professor of Statistics

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Li-Chein Shen
Associate Professor of Mathematics

This dissertation was submitted to the Graduate Faculty of the Department of Statistics in the College of Liberal Arts and Sciences and to the Graduate School and was accepted as partial fulfillment of the requirements for the degree of Doctor of Philosophy.

May 1992

____

Dean, Graduate School