Estimation of survival of left truncated and right censored data under increasing hazard

Russell Shinohara

Master of Science

Mathematics and Statistics

McGill University

Montreal, Quebec

June 2007

A thesis submitted to McGill University in partial fulfillment of the requirements of the degree of Master of Science

Library and
Archives Canada

Published Heritage
Branch

395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque et
Archives Canada

Direction du
Patrimoine de l'édition

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence
ISBN: 978-0-494-38435-0
Our file Notre référence
ISBN: 978-0-494-38435-0

ACKNOWLEDGEMENTS

My advisor, Prof. Masoud Asgharian, has been an incredible source of guidance and aid during the course of my program. His sage advice mixed with in with his warm manners and seemingly eternal patience have made him a delight to work with. I would also like to express my appreciation to Prof. David Wolfson and Prof. Russell Steele who introduced me to the field of statistics and taught me many valuable lessons through their courses. Finally, I would like to thank Pierre-David Letourneau, a close friend of mine, for his help with the translation of my abstract.

On a personal note, I would like to thank my parents, Phyllis Granoff and Koichi Shinohara, for their encouragement through the progress of seeking my degree. I would also like to thank my brother, Mark Colarusso. My girlfriend, Debika Biswal, has been invaluable to me through many stressful times during my studies.

ABSTRACT

When subjects are recruited through a cross-sectional survey they have already experienced the initiation of the event of interest, say the onset of a disease. This method of recruitment results in the fact that subjects with longer duration of the disease have a higher chance of being selected. It follows that censoring in such a case is not non-informative. The application of standard techniques for right-censored data thus introduces a bias to the analysis; this is referred to as length-bias. This paper examines the case where the subjects are assumed to enter the study at a uniform rate, allowing for the analysis in a more efficient unconditional manner. In particular, a new method for unconditional analysis is developed based on the framework of a conditional estimator. This new method is then applied to the several data sets and compared with the conditional technique of Tsai [23].

ABRÉGÉ

Lorsque plusieurs sujets d'étude sont identifiés et recrutés auprès d'une étude transversale, ils ont pour la plupart déjà subi l'amorce de l'événement d'intérêt, c'est-à-dire la contraction de la maladie. Conséquemment, par l'application d'une telle méthode de recrutement, les sujets qui survivent durant une période plus longue ont à la fois plus de chances d'être sélectionnés. Il s'en suit que la troncature, dans un tel cas, n'est point non-informative. L'application des techniques généralement utilisées afin d'obtenir des données bien tronquées introduit ainsi une source de biais dans l'analyse ; nous y référrerons comme biais de durée. Cette thèse a pour but d'examiner la situation générale où les sujets entreraient dans l'étude à un taux constant, permettant ainsi à l'analyse de se montrer plus fiable en étant inconditionnelle. Plus particulièrement, une méthode nouvelle en rapport avec l'analyse inconditionnelle et basée sur un estimateur conditionnel sera développée. Cette nouvelle méthode sera appliquée à maints ensembles de données et les résultats seront ensuite comparée avec ceux obtenus pas la méthode conditionnelle de Tsai[23].

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER 1
Survival Analysis and the Classical Setting

1.1 Introduction

The purpose of survival analysis is to examine the behavior of a population that experiences 'failures' over time. This failure could be a part in an automobile wearing out or a subject in a clinical trial dying. Let $X_1, \dots, X_n$ denote the failure times of these 'individuals'. We assume that the X_i are independent and we denote the *distribution function* $F(x) = P[X_i \leq x]$ for any i. We further define the *survival function* $S(x) = 1 - F(x) = P[X_i > x]$. The focus of a survival analysis to estimate $S(x)$ (or equivalently, $F(x)$). At this point, it is convenient to consider separately the cases of discrete and continuous failure times; that is, whether the failures can only occur at a fixed finite (or countable) number of points or they occur continuously though time. In either case, the *hazard function* can be defined as:

$$\lambda(x) = \lim_{h \to 0} \frac{P[x \leq X_i < x + h | X_i \geq x]}{h} \tag{1.1}$$

From here, it can be noted that for a continuous failure time distribution, the following identities hold:

$$\lambda(x) = \frac{f(x)}{S(x)} = -\frac{d \ln[S(x)]}{dx} \text{ and} \tag{1.2}$$

$$S(x) = \exp[-\Lambda(x)] = \exp\left[-\int_0^x \lambda(u) du\right] \tag{1.3}$$

where $f(x) = \frac{dF(x)}{dx}$ is the *density* corresponding to $F(x)$ and $\Lambda(x) = \int_0^x \lambda(u)du$ is the *cumulative hazard function* as implicitly defined above.

1.2 Right Censoring

Oftentimes, the failure process cannot be observed entirely. How would one treat an individual in a study who leaves the study for an unrelated reason? The only observed information in this case would be the fact that the individual left the study and the time at which he did. To remove such individuals, who are deemed *censored*, could bias the analysis drastically. Consider a subject who lives unusually long in a study; he is far more likely to be lost to follow up due to administrative problems than a subject who dies towards the beginning of the study. In fact, the censored case contains a significant amount of information about the failure distribution. Removing the censored cases would hence result in an underestimation of survival.

The question thus arises of how to deal with censored data. There is a fundamental difference between a failure and a censoring, yet one cannot discard censored cases. We shall introduce standard counting process methodologies to deal with such a case as in [16], but first let us establish a more rigorous notation for the right censored case. Let X_i be the failure times as above and C_i be the censoring times. Further, let $Y_i = \min(X_i, C_i)$ and $\Delta_i = I[X_i \leq C_i]$. It is assumed that the $X_1, X_2, \ldots, X_n$ are independent of $C_1, C_2, \ldots, C_n$ and the C_i are independent of each other. Now, let

$$N_i(t) = I[T_i \leq t, \Delta = 1] \text{ and } \tag{1.4}$$

$$N(t) = \sum_{i=1}^{n} N_i(t) = \sum_{t_i \leq t} \delta_i \tag{1.5}$$

We note that both of these processes are *counting processes* in that they are non-zero right-continuous stochastic processes with constant jumps of size 1, $N(0) = 0$, and $N(t) < \infty$ almost surely. Let us further define $\mathcal{F}_t$ to be the *history* of the processes up to and including time t. Under the independence of the X_i and C_i, we have that:

$$\begin{aligned} &P[t \leq Y_i \leq t + dt, \Delta_i = 1|\mathcal{F}_{t-}] \\ &\quad = P[t \leq X_i \leq t + dt, C_i > t + dt|X_i \geq t, C_i \geq t] \qquad (1.6) \\ &\quad = [\lambda(t)dt]I[T_i \geq t] \qquad (1.7) \end{aligned}$$

where $\mathcal{F}_{t-}$ denotes the history up to t. Let $dN(t) = N[(t + dt)^{-}] - N(t-)$ be the change in the process $N(t)$ over the interval $[t, t + dt)$. Then:

$$E[dN(t)|\mathcal{F}_{t-}] = E[\#\{i : t \leq X_i \leq t + dt, C_i > t + dt|\mathcal{F}_{t-}\}] \quad = R(t)\lambda(t)dt \tag{1.8}$$

where $R(t)$ is the number of subjects such that $Y_i \geq t$. The *intensity* and *cumulative intensity processes* are defined as $i(t) = R(t)\lambda(t)$ and $I(t) = \int_0^t i(s)ds$. Since $E[N(t)|\mathcal{F}_{t-}] = E[I(t)|\mathcal{F}_{t-}] = I(t)$, it follows that $M(t) = N(t) - I(t)$ is a martingale; it is thus dubbed the *counting process martingale.* Note that $N(t)$ is a non-decreasing step function and $I(t)$ is a predictable smooth process (the *compensator* process).

Hence, in the expression:

$$\frac{dN(t)}{R(t)} = \lambda(t)dt + \frac{dM(t)}{R(t)} \tag{1.9}$$

$dM(t)/R(t)$ can be considered noise (since $M(t)$ can be considered the difference of the counting process and its compensator), and hence integrating yields the Nelson-Aalen estimator of cumulative hazard (defining $0/0 = 0$ for convenience):

$$\hat{\Lambda}(t) = \int_0^t \frac{I[R(u) > 0]}{R(u)} dN(u) \tag{1.10}$$

$$= \int_0^t I[R(u) > 0]\lambda(u)du + \int_0^t \frac{I[R(u) > 0]}{R(u)} dM(u) \tag{1.11}$$

Note that $\int_0^t \frac{I[R(u)>0]}{R(u)} dM(u)$ is also a martingale, as it is the integral of a predictable process. Finally, an estimate of survival can be written down:

$$\hat{S}(t) = \prod_{j=0}^{t} [1 - d\hat{\Lambda}(t)] \tag{1.12}$$

$$= \prod_{j=0}^{t} \left(1 - \frac{dN(t)}{R(t)}\right) \tag{1.13}$$

is known as the Kaplan-Meier estimator of survival. Many important properties are available about this estimator (see [16]), including the fact that it is the nonparametric maximum likelihood estimator. Although this counting process approach is effective and concise, it is not available when the assumption of independent censoring fails to hold which is often the case in prevalent cohort studies.

CHAPTER 2
Left Truncation, Right Censoring, and Biased Sampling

2.1 Introduction and Preliminaries

When subjects are recruited through a cross-sectional survey with follow-up they have already experienced the initiation of the event of interest, say the onset of a disease. These subjects are termed prevalent cases as opposed to incident cases, those who have yet to develop the disease. Subjects who experience a shorter duration of the disease are less likely to be recruited into the study (see Wicksell (1925) [29], Goldsmith (1976) [11], and Cox (1969) [8]), and thus this method of recruitment favours subjects with longer duration of the disease. In other words, a sample taken in this fashion is not a representative sample from the population of interest. The complexity of the analysis of these data is furthered by the loss to follow-up of some subjects. This censoring in prevalent cohort data is informative in that the censoring time (the time when a subject is dropped from the study) and the survival time in the study are positively correlated. The application of standard techniques for right-censored data thus introduces a bias to the analysis (see Asgharian, M'Lan, and Wolfson (2002) [3]).

The framework of the problem is a random triple (T, X, C) for each subject. T is referred to as the truncation time; the time from disease onset until a subject is admitted to the study and under observation. X and C are the times until death (or another failure) and censoring (loss to follow-up) from onset of the disease. It is assumed that (T, C) is independent of the failure times X. Further, let $Y = \min(X, C)$ be the observed failure time and Δ be the censoring indicator (1 if $X \leq C$ and 0 otherwise). The observed data is thus (T, Y, Δ). In the past, such data has been analyzed either conditionally or unconditionally on $T < C$. This conditioning results in the loss of information. This paper examines the case where the subjects are assumed to enter the study at a uniform rate, allowing for the analysis in a more efficient unconditional manner.

The most obvious method of estimating survival in the left truncated and right censored case stems from the Kaplan-Meier estimator (1.13). As studied first by Lynden-Bell (1971) [18], then Woodroofe (1985) [31], and finally by Tsai, Jewell, and Wang (1987) [24], the modified Kaplan-Meier estimator becomes:

$$\hat{S}(x) = \begin{cases} \sum_{x_i < x} (R(x_j) - d_j)/R(x_j) & \text{for } x < x_{(1)} \\ 1 & \text{elsewhere} \end{cases} \tag{2.1}$$

where $x_{(1)}$ is the smallest of the failure times, and $R(x_j)$ and d_j are the number of subjects in the risk set and the number of failures at x_j. See Wang [28] for a comparison of this product limit estimator to other methods described below.

In the context of likelihood-based estimation of hazard, a natural simplification involves the assumption of non-decreasingness of hazard. This is often the case such as in the investigation of mortality resulting from a disease that affects primarily elderly patients. Theory relating to the case of non-decreasingness can also easily be converted to the case of non-increasingness and these two situations can be fused to address so-called U-shaped hazards. We thus assume, for the remainder of this paper, the hazard to be non-decreasing (herein, we use "increasing" for convenience).

2.2 Left truncated and Right Censored Data

The conditional approach introduced by Lynden-Bell [18], studied further by Woodroofe [31], Tsai, Jewell, and Wang [24] in the 1980's, and completed by Wang [28] in the early 1990's is a modified version of Gill's martingale approach. It is most efficient only when the left truncation induced by a cross-sectional sampling is assumed to have a completely unknown distribution. In many interesting cases, however, the truncation distribution is known at least up to some unknown parameters such as in Wicksell [29], Fisher [10], Patil and Rao [21], Asgharian et al. [3], and de Una-Alvarez [9]. It has been shown that under such circumstances the conditional approach is not most efficient. Following an unconditional approach, Asgharian et al. [3] found the most efficient estimator of the survival function.

Finding the most efficient estimator becomes considerably more complicated when, in addition to the information about the truncation distribution, a priori information about the form of the hazard of failure is also available.

The seminal work of Grenander [12] (1956) is the first to address this problem. Marshall and Proschan [19] and Rao [22] systematically studied distributional properties of Grenander's estimator. Barlow et al. [7] presents a comprehensive study of estimation under monotone restrictions. Padgett and Wei [20] were the first to address estimation under monotone restriction when observations are subject to right censoring. Huang and Wellner [14], Wellner [13], and Banerjee [6] have studied the distributional properties of Padgett and Wei's estimator. Tsai [23] considered estimation under monotone constraints when observations are subject to both left truncation and right censoring. Inspired by the previous work on the subject, Tsai used a conditional approach essentially conditioning away the left truncation and therefore circumventing complexities due to informative censoring. Tsai thus found the conditional nonparametric maximum likelihood estimator (NPMLE) of the survival function.

In many interesting applications, including the Channing House data analyzed in Tsai [23], there are reasons to believe that the left truncation distribution follows some specific form. The uniform distribution which corresponds to the problem of length-biased sampling, also known as stock sampling in labour force studies, plays a pivotal role among these forms. This is the form that Wicksell [29] originally introduced and justified using the sampling mechanism. It is also the limiting case of Fisher's model [10]. The uniform distribution also seems a reasonable model in labour force studies when we use a cross-sectional survey to determine the distribution of spell times when the society is in economic equilibrium (see Lancaster (1990) [17] and de Una-Alvanez [9]).

In medical applications studying disease duration when subjects are recruited through a cross-sectional survey, the uniform truncation model might be feasible if the incidence of the disease is stationary. The assumption of uniform truncation is also referred to as the "stationarity assumption". There are many diseases for which such an assumption is reasonable (see Asgharian et al. [5], Addona and Wolfson [2], and Addona et al. [1] (2007)).

2.3 Length-Biased Sampling

Length-biased sampling is not confined to our particular setup nor even biomedical problems. It has arisen in the past in fields ranging from applied physics to industrial statistics. Wicksell [29] was one of the first to note this phenomenon in 1925 in the study of Germ Centers (small spherical bodies or follicles) in a spleen. The spleen is cut cross-sectionally and the images of these bodies are observed. Wicksell notes that:

> "a random section will contain a relatively greater number of actually large follicles, than of small ones, because the former more frequently will come within reach of the section plain."

Goldsmith [11] observed the length-bias problem in a similar scenario in 1967. His observations consisted of thin slices of matter containing observable particles. The quantity of interest was the distribution of the sizes of these particles. It is clear that this problem is similar to that of Wicksell [29].

Yet another important milestone in the recognition of length-bias was in Cox's 1969 [8] discussion of issues in sampling theory. It was addressed in the context of sampling textile fibers. Consider a group of fibers of differing lengths assembled parallel to each other. If a cut were to be made perpendicularly to the axis of orientation, a naive estimate from this information would be subject to bias. In the problem of cross-sectional sampling of a prevalent cohort, the cohort is cut in time rather than space, but statistically these problems are identical.

The above discussion is by no means an exhaustive list. The problem of length bias has also been recognized by Zelen and Feinleib [32] in the context of disease screening. Further investigation was also conducted by Vardi [25] [26] [27] and by Asgharian et al. [3] and Asgharian and Wolfson [4].

In this thesis, we study the nonparametric maximum likelihood estimator of the survival function under the stationarity assumption and under monotone restrictions on the hazard of failure when observations are subject to right censoring. We derive the nonparametric maximum likelihood estimator for left truncated and right censored data under monotone constraints and the stationarity assumption in the next chapter.

CHAPTER 3
Estimation under Monotone Constraint

3.1 Background and Preliminaries

The first likelihood-based estimate of non-decreasing hazard was described in Grenander (1956) [12]. The case examined was that of a continuous failure time distribution on (a, ∞) with neither censoring nor truncation. Letting $\lambda(x) = f(x)/(1 - F(x))$ denote the hazard (where f and F denote the density and cumulative distribution functions for X respectively), the likelihood l_x thus must be maximized subject to

$$\log l_x = \log l_a - \int_a^x \lambda(x)dx \quad \text{and} \tag{3.1}$$

$$\frac{d^2 \log l_x}{dx^2} = -\frac{d\lambda}{dx} \leq 0 \tag{3.2}$$

Now, fixing the points $x_1, x_2, \ldots, x_n$ and the hazard at these points $\lambda_i = \lambda(x_i)$ the likelihood can be seen to be:

$$L = f(x_1)f(x_2)\ldots f(x_n) \tag{3.3}$$

$$= \lambda_1 e^{-\int_a^{x_1} \lambda(x)dx} \lambda_2 e^{-\int_a^{x_2} \lambda(x)dx} \ldots \lambda_n e^{-\int_a^{x_n} \lambda(x)dx} \tag{3.4}$$

It can thus easily be seen that the likelihood is maximized whenever

$$\int_a^{x_n} l_x^* \lambda(x)dx = \int_a^{x_1} \lambda(x)dx + \int_a^{x_2} \lambda(x)dx + \ldots + \int_a^{x_n} \lambda(x)dx \tag{3.5}$$

is minimized under the condition of non-decreasingness of $\lambda(x)$. Grenander suggests that this can be be achieved with respect to choice of $\lambda(x)$ as the step function:

$$\begin{aligned}
\lambda(x) &= 0 & \text{for } 0 \le x < x_1 \\
\lambda(x) &= \lambda_1 & \text{for } x_1 \le x < x_2 \\
&\vdots & \qquad (3.6) \\
\lambda(x) &= \lambda_{n-1} & \text{for } x_{n-1} \le x < x_n
\end{aligned}$$

This can also be seen as estimating the likelihood by a polygonal figure with exponential arcs as sides joined at the points x_i. Grenander then fixed $\int_a^{x_n} l_x^* \lambda(x)dx$ and maximized the product of the λ_i using his previously established scheme to complete the specification of his estimator.

Marshall and Proschan (1964) [19] examined this case in more detail. They formalized the increasing failure rate (IFR) assumption of (3.2); they defined F to be IFR if the support of F is a (possibly infinite) closed interval (α, β) and (3.2) holds on $[\alpha, \beta)$. Furthermore, Marshall and Proschan [19] showed that if F is IFR and $F(x_0) < \infty$, then F is absolutely continuous on $(-\infty, x_0)$. The likelihood (from (3.4), setting $a = -\infty$) to be considered was:

$$\log(L(F)) = \sum_{i=1}^{n} \log \lambda(x_i) + \sum_{i=1}^{n} \int_{-\infty}^{x_i} \lambda(x)dx \qquad (3.7)$$

Another interesting discussion in [19] relates to the definition of maximum likelihood itself. Let $\mathcal{F}$ denote the class of failure distributions with IFR. The trouble with maximizing the likelihood (3.7) directly in $\mathcal{F}$ is that $f(x_n)$ can be chosen to be arbitrarily large. It is thus natural to consider the subclass $\mathcal{F}^M$ of distributions whose corresponding failure rate λ is bounded above by M. The maximum likelihood estimator is then defined to be the limiting estimator as M tends to infinity.

Let λ^* denote the estimator from (3.6) (clearly a member of $\mathcal{F}^M$). It is clear that by IFR, $\lambda^*(x) \leq \lambda(x)$ and $\int_{\infty}^{x_i} \lambda(x)dx \geq \int_{\infty}^{x_i} \lambda^*(x)dx$ and thus $L(F) \leq L(F^*)$ where F^* denotes the failure distribution corresponding to λ^*. Hence, in order to maximize (3.7) it suffices to maximize:

$$\log(L(F^*)) = \sum_{1}^{n} \log \lambda(x_i) - \sum_{1}^{n-1} (n-i)(x_{i+1} - x_i)\lambda(x_i) \tag{3.8}$$

under the assumption of non-decreasing bounded (by M) hazard. By appealing to an optimization scheme in [12] and letting $M \to \infty$, the maximum likelihood estimator was determined to be:

$$\hat{\lambda}_i = \min_{v \geq i+1} \min_{u \leq i} [v-u][(n-u)(X_{u+1} - X_u) + \ldots + (n-v+1)(X_v - X_{v-1})]^{-1} \tag{3.9}$$

$$\text{for } i = 1, \ldots, n-1$$

Marshall and Proschan [19] further established the consistency of their estimate and investigated the cases of decreasing hazard and discrete versions of the problem. Details of the derivation and proof of consistency are also presented carefully in Barlow et al. [7]. Rao (1970) [22] studied the asymptotic weak behavior of the estimator.

3.2 Estimation under Monotone Constraints with Right Censored Data

The first modification of Grenander's estimator was produced by Padgett and Wei in 1980 [20] to allow for right censorship. That is, the observable quantities are $Y = \min(X, C)$ and $\Delta = I[X \leq C]$ where X denotes the true failure time and C denotes a censoring time. It is assumed that the $X_1, X_2, \ldots, X_n$ are independent of $C_1, C_2, \ldots, C_n$ and the C_i are independent of each other. For convenience, a was chosen to be 0 in (3.1) (that is, failure times are non-negative). The log-likelihood then becomes:

$$\log(L(F)) = \sum_1^n \delta_i \log \lambda(x_i) + \sum_1^n \int_0^{x_i} \lambda(x) dx \tag{3.10}$$

$$\leq \sum_1^n \delta_i \log \lambda(x_i) - \sum_1^{n-1} (n-i)(x_{i+1} - x_i)\lambda(x_i) = \log(L(F^*)) \tag{3.11}$$

following a similar argument to that of Marshall and Proschan [19]. This can also be maximized in a similar fashion yielding an estimator with a similar minimax form. Furthermore, Padgett and Wei [20] suggested an estimator of survival based on the increasing hazard estimate:

$$\hat{S}(t) = \exp\left\{-\int_0^t \hat{\lambda}(x)dx\right\} = \exp\left[-\sum_{i:w_i \leq t} \hat{\lambda}_i\{\min(t, w_{i+1}) - w_i\}\right] \quad \text{for } t \geq 0 \tag{3.12}$$

where $w_1, \ldots, w_m$ are the distinct exactly observed (uncensored) failure times. Note the smoothness of this estimate.

3.3 Estimation under Monotone Constraints with Right Censored and Left Truncated Data: Conditional Approach

The advancement that is most pertinent to the current investigation was that of Tsai (1988) [23]. The left-truncated and right censored problem described in the setup section was exactly addressed, conditionally on $T < C$. At the time, the product-limit estimator for right censored and left truncated data of Tsai, Jewell, and Wang [24] was the most widely accepted estimator for survival: defining $d(y) = \sum_1^n I[y_i = y, \delta_i = 1]$ and $R(y) = \sum_1^n I[t_i \leq y \leq y_i]$ as the risk set, the product-limit estimator is:

$$\hat{S}_{PL} = \prod_{y_j \leq y} \left\{ 1 - \frac{d(y_j)}{R(y_j)} \right\} \tag{3.13}$$

Tsai thus suggested yet another modification to Grenander's estimator to exploit IFR in such a case.

The log likelihood of the left truncated right censored case can easily be written down as:

$$\begin{aligned} \log(L) &= \sum_{i=1}^{n} \delta_i \log \lambda(y_i) - \sum_{i=1}^{n} \int_{t_i}^{y_i} \lambda(x)dx & (3.14) \\ &\leq \sum_{i=1}^{n} \delta_i \log \lambda(y_i) - \sum_{i=1}^{n} \{(z_{l_i+1} - z_{l_i})\lambda(z_{l_i}) + \\ &\quad (z_{l_i+2} - z_{l_i+1})\lambda(z_{l_i+1}) + \ldots + (z_{\mu_i} - z_{\mu_i-1})\lambda(z_{\mu_i-1}\} & (3.15) \end{aligned}$$

where l_i and μ_i are defined such that $z_{l_i} = t_i$ and $z_{\mu_i} = y_i$. By the argument of Marshall and Proschan [19], it remains to maximize

$$\log(L^*) = \sum_{i=1}^{n} \delta_i \log \lambda(y_i) - \sum_{j=1}^{k-1} R(z_j+)(z_{j+1} - z_j)\lambda(z_j) \tag{3.16}$$

$$= \sum_{i=1}^{k-1} d(z_i) \log \lambda(z_i) - R(z_i+)(z_{i+1} - z_i)\lambda(z_i) \tag{3.17}$$

where $z_1, z_2, \ldots, z_k$ are the distinct order statistics of $t_1, \ldots, t_n, y_1, \ldots, y_n$. From Barlow et al. [7], the estimator maximizing (3.14) is:

$$\hat{\lambda}(y) = \sum_{j=1}^{k-1} \hat{\lambda}_j \mathbf{1}[z_j \leq y < z_{j+1}] + \hat{\lambda}_k(y)\mathbf{1}[y \geq z_k] \tag{3.18}$$

$$\text{where } \hat{\lambda}_j = \min_{s \leq j} \max_{t \geq j} \frac{\sum_{l=s}^{t} d(z_l)}{\sum_{l=s}^{t} R(z_l+)(z_{l+1} - z_l)} \tag{3.19}$$

where, $\hat{\lambda}_k$ is either infinite if $d(z_k) > 0$ or set to $\hat{\lambda}_{k-1}$ for convenience. The survival estimate (3.12) was then suggested for a smooth estimated survival under IFR. To the best of our knowledge, no distributional results are yet available about Tsai's estimator.

3.4 Unconditional Approach

Unconditional methodologies for estimating survival were pioneered by Vardi (1989) [27] in the context of multiplicative censoring. Given a sample of data $Z_1, Z_2, \ldots, Z_n$ and $X_1, X_2, \ldots, X_n$ from a distribution G say, only $Y_i = Z_i U_i$ and the X_i are observed where the U_i are independent uniform $(0, 1)$ random variables. For the Y_i, this is equivalent to the so-called stationarity in truncation assumption;

patients enter the study uniformly through time.

The likelihood for this problem is:

$$L(G) = \prod_{i=1}^{m} dG(x_i) \prod_{i=1}^{n} \int_{z \geq y_i} \frac{1}{z} dG(z) \tag{3.20}$$

Directly from here, one can argue that a maximizing distribution G puts all of its mass on the points $\mathcal{X} = \{x_1, \ldots, x_m, y_1, \ldots, y_n\}$: let there be a point outside of $\mathcal{X}$ that has positive mass, x. Moving that mass to the point in $\mathcal{X}$ which is closest but less than x would result in an increase in the likelihood. It can also easily be seen that any mass to the left of $\min \mathcal{X}$ could be shifted to $\min \mathcal{X}$. Hence, the problem of maximizing (3.20) is a discrete problem. Vardi also allowed for ties in his discrete likelihood:

$$L(p) = \prod_{j=1}^{h} p_j^{\xi_j} \left(\sum_{k=j}^{h} \frac{1}{t_k} p_k \right)^{\zeta_j} \tag{3.21}$$

where $t_1, \ldots, t_h$ are the discrete failure times, $\xi_j = \sum_{i=1}^{m} I(x_i = t_j)$, and $\zeta_j = \sum_{i=1}^{n} I(y_i = t_j)$. It thus remains to maximize (3.21) with respect to $p = (p_1, \ldots, p_h)$ subject to $p_i \geq 0$ and $\sum p_i = 1$. This can be achieved through the EM algorithm, with $z_1, z_2, \ldots, z_n, x_1, x_2, \ldots, x_n$ as the 'complete data' and $x_1, \ldots, x_m, y_1, \ldots, y_n$ as the 'incomplete data'. Further details can be found in [27], and a comparison of Vardi's estimator to the product-limit estimator can be found in Wang (1991) [28].

It is important to note that Vardi's estimator approximates the length-biased survival function S_{LB}. It is thus necessary to consider the problem of estimating the unbiased survival function S_u given S_{LB}. Fortunately, from Cox [8] and Asgharian, M'Lan and Wolfson (2002) [3]:

$$S_u(y) = \frac{\int_y^\infty \frac{1}{x} dS_{LB}(x)}{\int_0^\infty \frac{1}{x} dS_{LB}(x)} \tag{3.22}$$

and hence by the invariance property of maximum likelihood estimators

$$\hat{S}_u(y) = \frac{\int_y^\infty \frac{1}{x} d\hat{S}_{LB}(x)}{\int_0^\infty \frac{1}{x} d\hat{S}_{LB}(x)} \tag{3.23}$$

as suggested by Asgharian et al. [3]. Asymptotic properties of the estimator are also discussed in [3].

A closer investigation of the length-biased version of Vardi's estimator can be found in Asgharian and Wolfson (2005) [4]. The development of the likelihood requires several preparations. The first is a reparametrization of the observed data. Let A_i, R_i, and C_i denote the current-age, residual lifetime, and residual censoring times of the the i-th subject. More concretely, A_i can be thought of as time from onset of a disease until recruitment into a study, R_i be the time from recruitment to death, and so on. The observed information is thus $(A_i, \min(R_i, C_i), \delta_i)$. The first thing to note in this setup is that censoring is in fact informative. Indeed, since for an observed subject $X = A + R$, $Y = A + C$, and $A|X$ is uniform $(0, X)$ we have that $\text{Cov}(X, Y) > 0$ except in trivial cases. This further supports the need for a new estimator since it disqualifies the Kaplan-Mayer estimator as a maximum likelihood estimator.

At this point, it is informative to note that there is a strong connection between the setup described in the previous paragraph and renewal theory. The joint density of (A, R) can thus easily be seen to be:

$$f_{A,R}(a,r) = \frac{f_{X'}(a+r)}{\mu_{X'}} I[a > 0, r > 0] \tag{3.24}$$

and from Cox [8] (or directly from renewal theory), the length-biased density of the failure time is:

$$g(x) = \frac{x f_U(x)}{\mu_U} \tag{3.25}$$

where the subscript U indicates the unbiased version, namely those corresponding to X.

The likelihood can thus be written down as:

$$\begin{aligned}
L &= \left(\prod_{i\in\mathcal{UC}} f_{A,R}(a_i, r_i)\right)\left(\prod_{j\in\mathcal{C}} dP(a_j, R_j \geq c_j)\right) && (3.26)\\
&= \left(\prod_{i\in\mathcal{UC}} dG(x_i)\right)\left(\prod_{j\in\mathcal{C}} \int_{c_j\leq r} \frac{f_U(a_j + r)}{\mu_U} dr\right) && (3.27)\\
&= \left(\prod_{i\in\mathcal{UC}} dG(x_i)\right)\left(\prod_{j\in\mathcal{C}} \int_{y_j\leq z} z^{-1} dG(z)\right) && (3.28)\\
&= \prod_{i=1}^{k} \left[(dG(x_i))^{\delta_i} \left(\int_{y_i\leq r} z^{-1} dG(z)\right)^{1-\delta_i}\right] && (3.29)
\end{aligned}$$

The estimator from (3.23) can thus be seen as the maximization of (3.29). Asgharian and Wolfson [4] show strong uniform consistency, convergence in distribution to a Gaussian process, and asymptotic efficiency of this estimator.

3.5 Maximum Likelihood Derivation

There is a striking similarity of the likelihood of Asgharian and Wolfson [4] to that of Tsai [23]. With the hope of finding an analytical form for the maximum likelihood estimator in this case, consider the likelihood (3.29):

$$L = \prod_{i=1}^{n} (dG(y_i))^{\delta_i} \left(\int_{y_i \leq z} z^{-1} dG(z) \right)^{1-\delta_i} \tag{3.30}$$

$$= \prod_{i=1}^{n} (g(y_i))^{\delta_i} \left(\int_{y_i \leq z} z^{-1} g(z) dz \right)^{1-\delta_i} \tag{3.31}$$

$$= \prod_{i=1}^{n} \left(\frac{y_i \cdot f_u(y_i)}{\mu_u} \right)^{\delta_i} \left(\int_{y_i \leq z} z^{-1} \frac{z f_u(z)}{\mu_u} dz \right)^{1-\delta_i} \tag{3.32}$$

$$= \prod_{i=1}^{n} \left(\frac{{y_i}^{\delta_i} \cdot f_u(y_i)^{\delta_i}}{\mu_u^{\delta_i}} \right) \left(\int_{y_i \leq z} z^{-1} \frac{z f_u(z)}{\mu_u} dz \right)^{1-\delta_i} \tag{3.33}$$

$$= \prod_{i=1}^{n} \left(\frac{{y_i}^{\delta_i} \cdot f_u(y_i)^{\delta_i}}{\mu_u} \right) \left(\int_{y_i \leq z} f_u(z) dx \right)^{1-\delta_i} \tag{3.34}$$

$$\propto \prod_{i=1}^{n} \frac{f_u(y_i)^{\delta_i} S_u(y_i)^{1-\delta_i}}{\mu_u} \tag{3.35}$$

but, $\mu_u = \int_0^\infty S_u(t)dt$ and $\lambda_u(y) = \frac{f_u(y)}{S_u(y)}$, so

$$= \prod_{i=1}^{n} \frac{f_u(y_i)^{\delta_i} S_u(y_i)^{1-\delta_i}}{\int_0^\infty S_u(t)dt} \tag{3.36}$$

$$= \prod_{i=1}^{n} \frac{\lambda_u(y_i)^{\delta_i} S_u(y_i)}{\int_0^\infty S_u(t)dt} \tag{3.37}$$

Now, $\int_o^\infty S_u(t)dt$ can only be contributed to at $\{z_i, i = 1...k\} = \{t_j,\ j = 1, ..., n\} \cup \{y_j,\ j = 1, ..., n\}$ (without loss of generality, we assume that the z_i's are ordered and let $z_0 = 0$), so:

$$= \prod_{i=1}^{n} \frac{\lambda_u(y_i)^{\delta_i} S_u(y_i)}{\sum_{j=1}^{k} S_u(z_j)(z_j - z_{j-1})} \tag{3.38}$$

$$= \prod_{i=1}^{n} \frac{\lambda_u(y_i)^{\delta_i} S_u(y_i)}{S_u(t_i)} \frac{S_u(t_i)}{\sum_{j=1}^{k} S_u(z_j)(z_j - z_{j-1})} \tag{3.39}$$

We recognize this as the likelihood from Tsai [23], except multiplied by a_i:

$$= \prod_{i=1}^{n} \frac{\lambda_u(y_i)^{\delta_i} S_u(y_i)}{S_u(t_i)} a_i \tag{3.40}$$

where $a_i = \frac{S_u(t_i)}{\sum_{j=1}^{k} S_u(z_j)(z_j - z_{j-1})}$. Hence,

$$\log(L) = \sum_{i=1}^{n} \delta_i \log \lambda(y_i) - \sum_{i=1}^{n} \int_{t_i}^{y_i} \lambda(u)du + \sum_{i=1}^{n} \log(a_i) \tag{3.41}$$

Letting $R(y) = \sum_{i=1}^{n} \mathbf{1}[t_i \leq y \leq y_i]$ and $d(y) = \sum_{i=1}^{n} \mathbf{1}[y_i = y, \delta_i = 1]$, by an argument in Tsai [23] and Marshall and Proschan [19], it remains to maximize

$$\log(L^*) = \sum_{i=1}^{n} \delta_i \log \lambda(y_i) - \sum_{j=1}^{k-1} R(z_j+)(z_{j+1} - z_j)\lambda(z_j) + \sum_{i=1}^{n} \log(a_i) \tag{3.42}$$

$$= \sum_{i=1}^{k-1} d(z_i) \log \lambda(z_i) - R(z_i+)(z_{i+1} - z_i)\lambda(z_i) + a_i^*(z_i) \tag{3.43}$$

where $a^*(z_j) = \sum_{i=1}^{n} \log(a_i) I[z_j = t_i]$. Setting $n_i = d(z_i)$, $T_i = R(z_i+)(z_{i+1} - z_i)$, and $\lambda_i = \lambda(z_j)$, we recognize this resembling the application at the end of Example 1.8 in Barlow et al [7]. However, if we choose $T_i = R(z_i+)(z_{i+1} - z_i) - a^*(z_i)/\lambda(z_i)$

and n_i and λ_i as above, then we get

$$\hat{\lambda}(y) = \sum_{j=1}^{k-1} \hat{\lambda}_j \mathbf{1}[z_j \leq y < z_{j+1}] + \hat{\lambda}_k(y)\mathbf{1}[y \geq z_k] \tag{3.44}$$

$$\text{where } \hat{\lambda}_j = \min_{s \leq j} \max_{t \geq j} \frac{\sum_{l=s}^{t} d(z_l)}{\sum_{l=s}^{t} R(z_l+)(z_{l+1} - z_l) - a^*(z_l)/\lambda(z_l)} \tag{3.45}$$

where, as in Tsai, $\hat{\lambda}_k$ is either infinite if $d(z_k) > 0$ or set to $\hat{\lambda}_{k-1}$ for convenience.

There is one more complication that must be dealt with, however, before we can find a closed form for the nonparametric maximum likelihood estimator for lambda. This is that the a_i's depend on the unknown underlying survival function, S. In order to deal with this, we use an estimate of $S(z_i)$ at each step iteration. The algorithm thus becomes:

1. Choose an increasing failure rate starting vector, say $\hat{\lambda}^{(0)}$. We suggest a uniform distribution of mass on each of the z_i's (an increasing hazard).
2. For $i = 1, \ldots, k-1$ do
   (a) Let

$$\hat{\lambda}_j^{(i)} = \min_{s \leq j} \max_{t \geq j} \frac{\sum_{l=s}^{t} d(z_l)}{\sum_{l=s}^{t} R(z_l+)(z_{l+1} - z_l) - a^*_{(i-1)}(z_l)/\hat{\lambda}^{(i-1)}(z_l)} \tag{3.46}$$

   (b) Then, let

$$\hat{\lambda}^{(i)}(y) = \sum_{j=1}^{k-1} \hat{\lambda}_j^{(i)}[z_j \leq y < z_{j+1}] + \hat{\lambda}_k(y)\mathbf{1}[y \geq z_k] \tag{3.47}$$

3.6 Numerical Complications

The difficulty with the proposed estimator is its sensitivity to numerical error. To start with, the results of the method are extremely sensitive to the scale of the problem. The term from where this phenomenon originates is the denominator of the minimax in (3.49). Any error in $\hat{\lambda}^{(i-1)}$, especially when $\hat{\lambda}^{(i-1)}$ is small, propagates to the next iteration. This results in the estimator having difficulties jumping and the estimate being near zero for most of the study. In fact, for scenarios in which hazard is increasing, it is common for the hazard to be negligible at the beginning of the study.

In order to deal with the problem of division by zero, a simple truncation of the hazard from below was applied (namely, $\hat{\lambda}(z) = 0$ was not allowed). Although this avoided division by zero, this did not completely solve the issues brought about by numerical error. For $\hat{\lambda}(z_l)$ small, the term $a^*_{(i-1)}(z_l)/\hat{\lambda}^{(i-1)}(z_l)$ can be very large. On the other hand, the scale of the problem or the size of the risk set can result in the term $R(z_l+)(z_{l+1} - z_l)$ being large, hence giving rise to the addition of a small number to a large number. The opposite case is also possible; either of these scenarios are detrimental to the performance of the estimate.

The solution proposed is to add a fixed quantity to the likelihood (3.41). The likelihood becomes

$$\log(L) = \sum_{i=1}^{n} \delta_i \log \lambda(y_i) - \sum_{i=1}^{n} \int_{t_i}^{y_i} \lambda(u)du + \sum_{i=1}^{n} [\log(a_i) + b] \qquad (3.48)$$

where b is a fixed scaling parameter. The derivation, if followed directly from above, leads to the modified estimator:

1. Choose an increasing failure rate starting vector, say $\hat{\lambda}^{(0)}$. We suggest a uniform distribution of mass on each of the z_i's (an increasing hazard).
2. For $i = 1, \ldots, k-1$ do
   (a) Let

   $$\hat{\lambda}_j^{(i)} = \min_{s \leq j} \max_{t \geq j} \frac{\sum_{l=s}^{t} d(z_l)}{\sum_{l=s}^{t} R(z_l+)(z_{l+1} - z_l) - (b^*_{(i-1)}(z_l))/\hat{\lambda}^{(i-1)}(z_l)} \qquad (3.49)$$

   where $b^*_{(i-1)}(z_l) = \sum_{j=1}^{n}(\log(a_j^{(i-1)})+b)I[z_l = t_j]$, and $a_j^{(i-1)}$ is our estimate of a_j at the $i - 1^{st}$ iteration.

   (b) Then, let

   $$\hat{\lambda}^{(i)}(y) = \sum_{j=1}^{k-1} \hat{\lambda}_j^{(i)}[z_j \leq y < z_{j+1}] + \hat{\lambda}_k(y)\mathbf{1}[y \geq z_k] \qquad (3.50)$$

This method was then further modified to allow for the scaling of the likelihood at each iteration according to the last iteration completed. Furthermore, the first iteration is run twice: first to find a scaling scheme and then to find an estimate of hazard. During both of these steps, the original stab at the hazard is used in the iteration as the previous estimate.

CHAPTER 4
Simulations and Applications

4.1 Application: Simulated Weibull Data

The first application of the method was to simulated Weibull data ($n = 30$) with a variety of parameters. The method was coded in C++ using the Gnu Scientific Library and was determined to converge in each of the cases after 10 iterations. The truncation time was set after the failure time simulation according to multiplicative censoring requirements. A fixed time after truncation censoring was applied; that is $C = T + c$ where c is deterministic and was set to ensure a reasonable amount of censoring in the sample. See figures 4–1 to 4–9 for the results. Legends are the same throughout. As can be noted from these plots, although the hazard seems to be estimated better in some cases by Tsai's estimator, the proposed estimator does better when estimating survival, often the true quantity of interest. It is important to note that due to numerical instabilities, it is not always immediately evident that our estimate of survival is closer to the true value. It is sometimes necessary to re-scale the method many times before satisfactory results are yielded. An alternative method for analyzing this so-called type I censoring scenario is described by de Una-Alvarez (2004) [9].

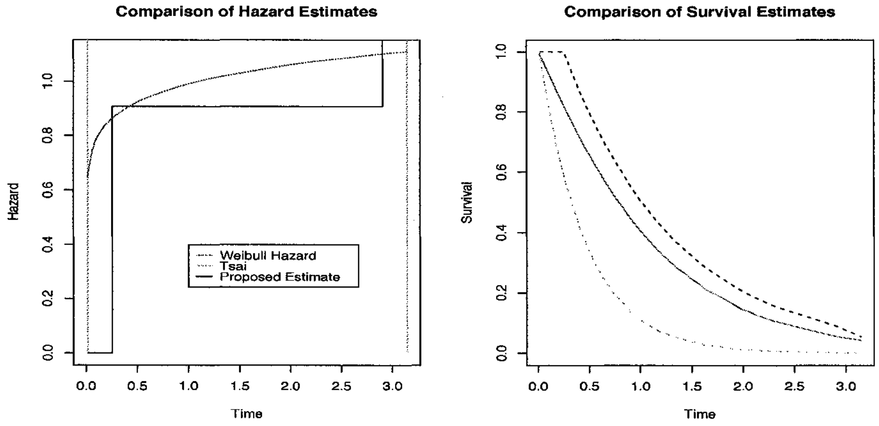

Figure 4–1: comparison of results from a Weibull(1.1,1.1)

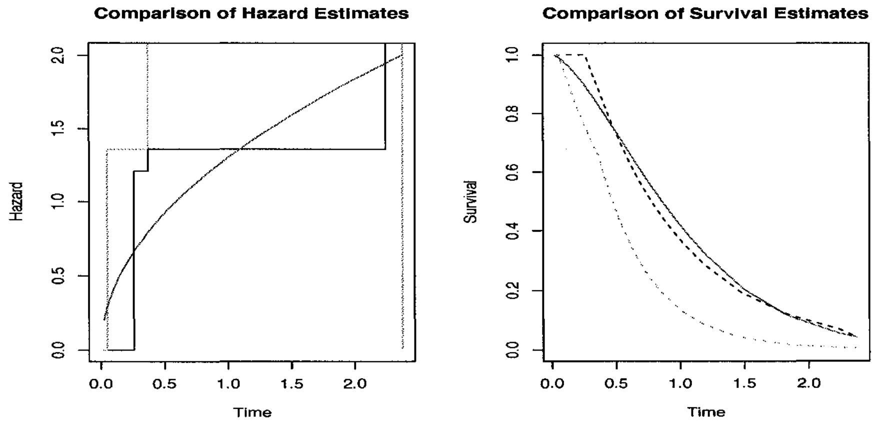

Figure 4–2: comparison of results from a Weibull(1.1,1.5)

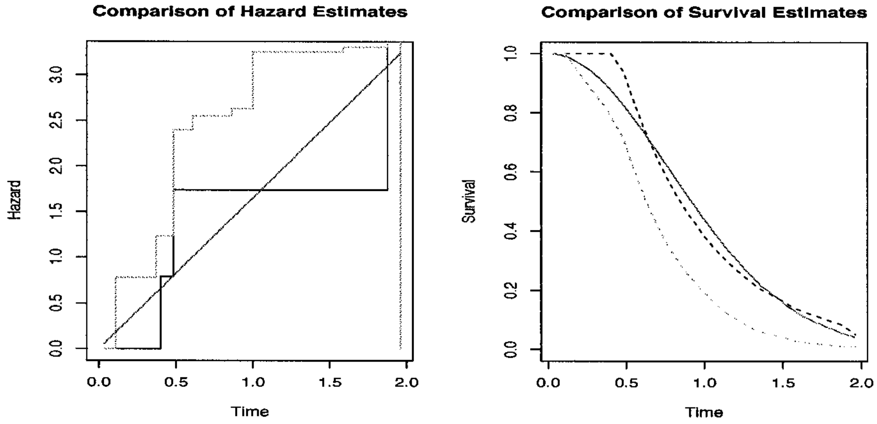

Figure 4–3: comparison of results from a Weibull(1.1,2)

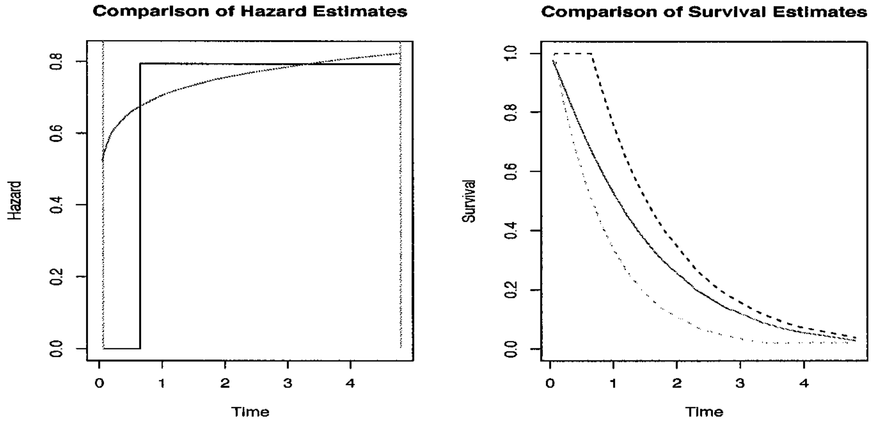

Figure 4–4: comparison of results from a Weibull(1.5,1.1)

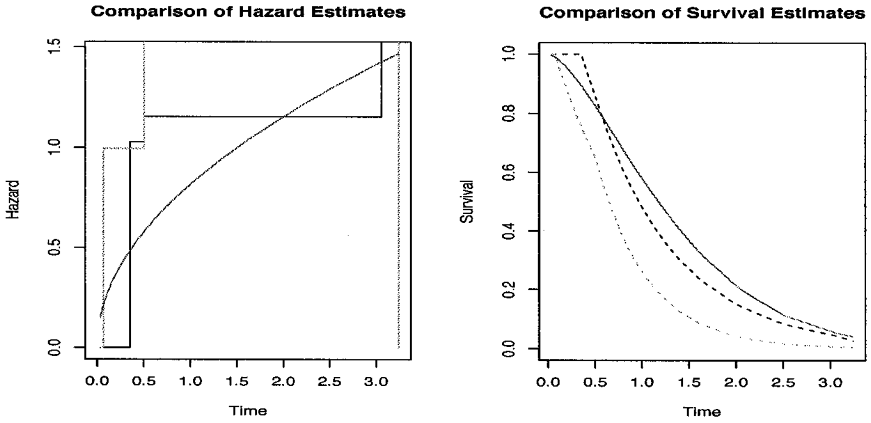

Figure 4–5: comparison of results from a Weibull(1.5,1.5)

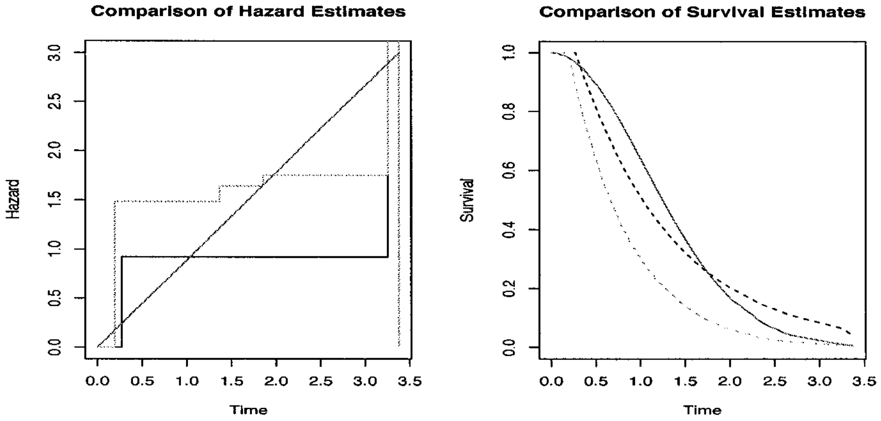

Figure 4–6: comparison of results from a Weibull(1.5,2)

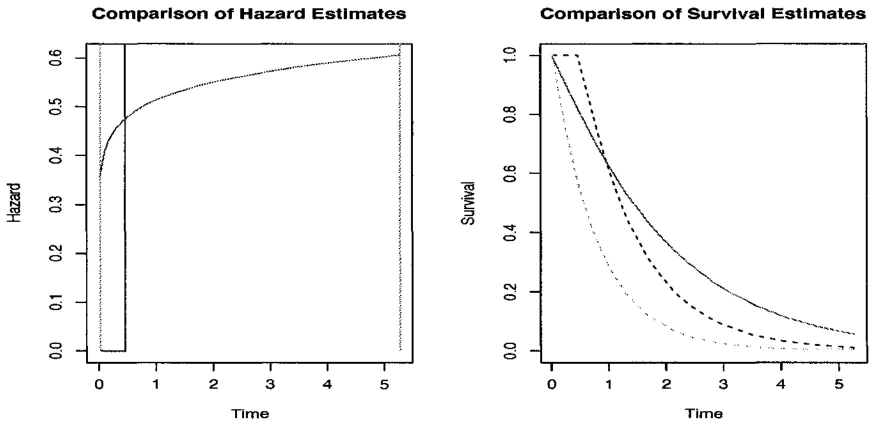

Figure 4–7: comparison of results from a Weibull(2,1.1)

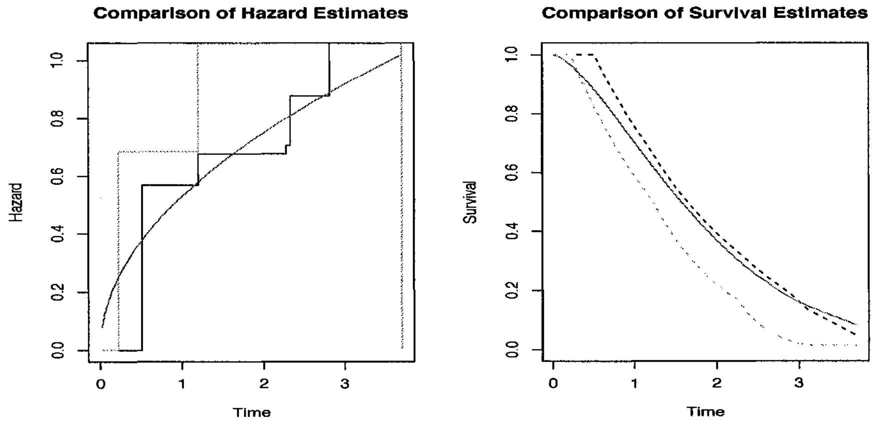

Figure 4–8: comparison of results from a Weibull(2,1.5)

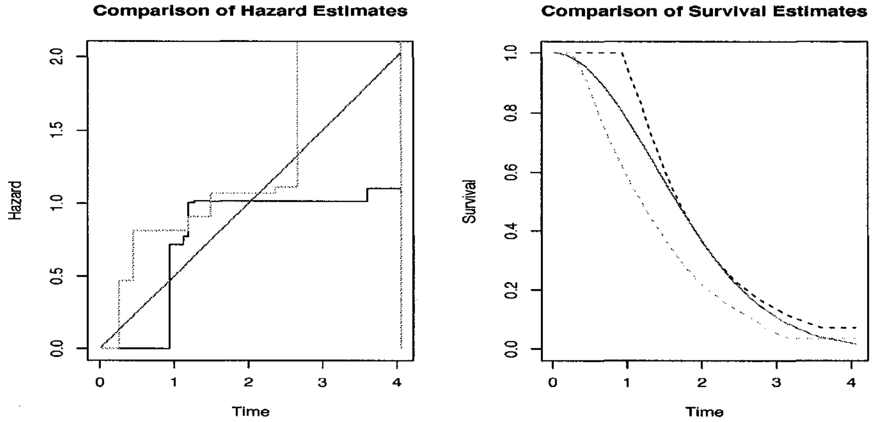

Figure 4–9: comparison of results from a Weibull(2,2)

4.2 Application: Channing House Data

The method was also applied to data given in Hyde [15] concerning survival times (in months) of elderly men in the Channing House retirement home situated in Palo Alto, California. Of these 97 men, 46 died and 55 left the study. The remaining 46 were alive at the end of the study. The risk set of the data is very small around the first two failures, and hence these individuals were not included in the analysis. A comparison of results of the proposed method, Tsai's method, and the product-limit estimator can be seen in figure 4–10.

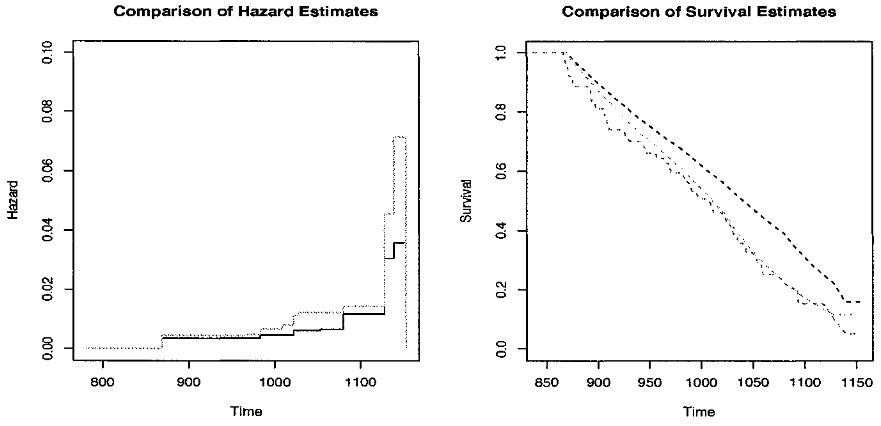

Figure 4–10: comparison of results from the Channing House data set

4.3 Application: The Canadian Study of Health and Aging

Data concerning dementia was collected in the Canadian Study of Health and Aging (CSHA). In 1991, 10 263 Canadians over sixty-five were accepted into the study. 821 of these were determined to have one of possible Alzheimer's disease, probable Alzheimer's disease, or vascular dementia. This screening phase was deemed CSHA-1. The second phase, CSHA-2, took place approximately five years later when the 821 cases were re-examined and survival and censoring times were recorded. The purpose of the analysis was to examine the effects of dementia on survival, as in [30]. Subsets of various sizes and the full data set were selected and were analyzed using Tsai's and the proposed method. From the results, figures 4–11 to 4–14, it can be seen that the proposed estimator converges to a reasonable estimate at smaller sample sizes than Tsai's. This is attributed to the added efficiency resulting from the incorporation of the assumption of stationarity in truncation. It is also clear that for large sample sizes (see 4–16), the estimates are very close supporting a conjecture of consistency of the proposed estimate.

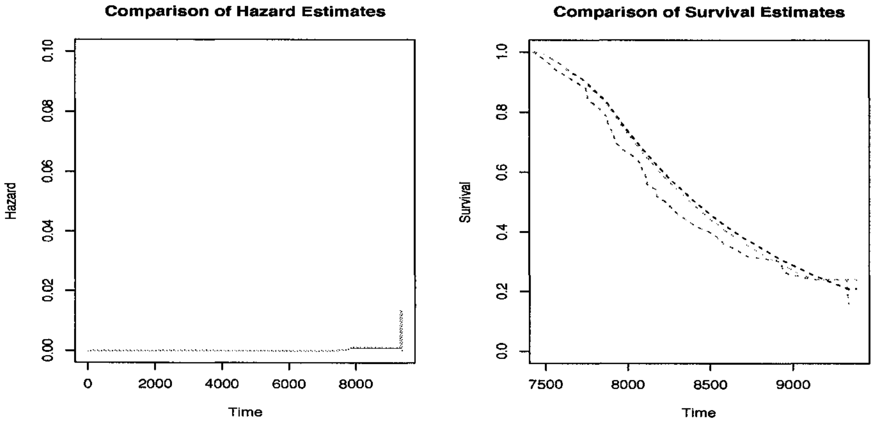

Figure 4–11: comparison of results from a subset of the CSHA data set (n=50)

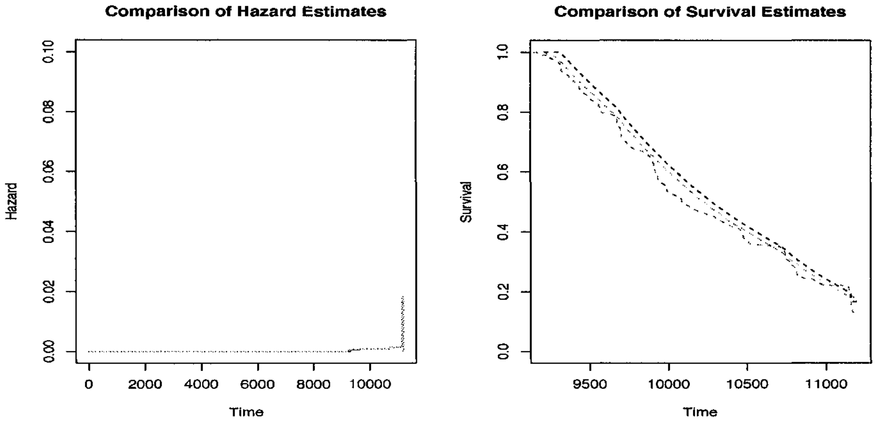

Figure 4–12: comparison of results from a subset of the CSHA data set (n=100)

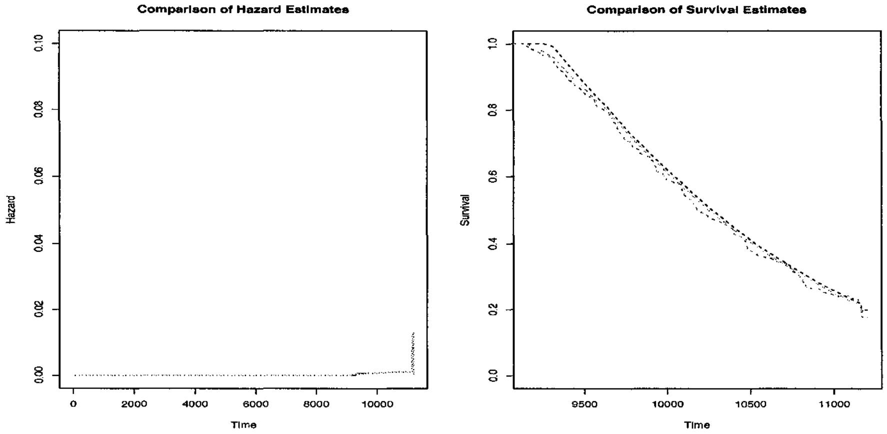

Figure 4–13: comparison of results from a subset of the CSHA data set (n=200)

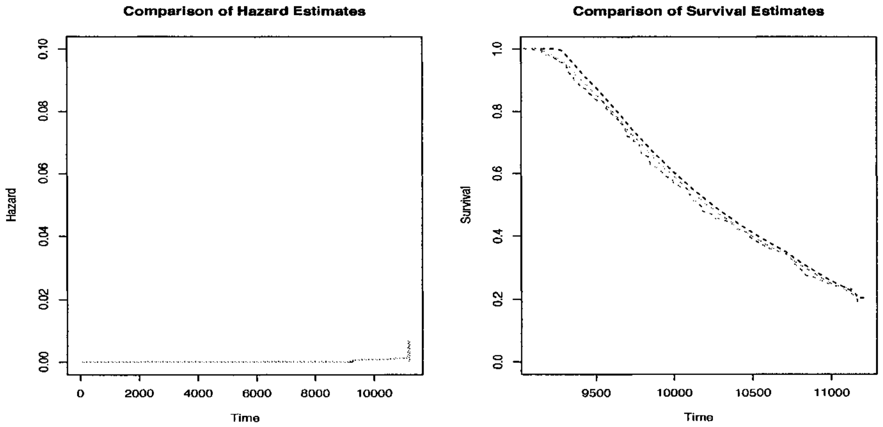

Figure 4–14: comparison of results from a subset of the CSHA data set (n=300)

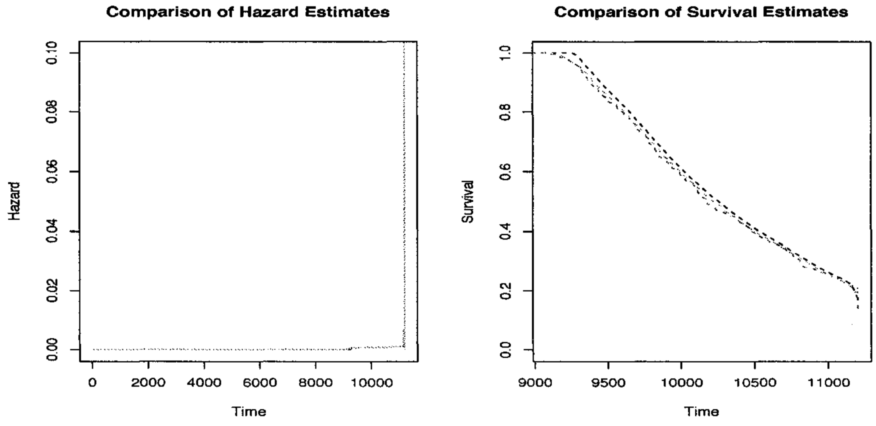

Figure 4–15: comparison of results from a subset of the CSHA data set (n=500)

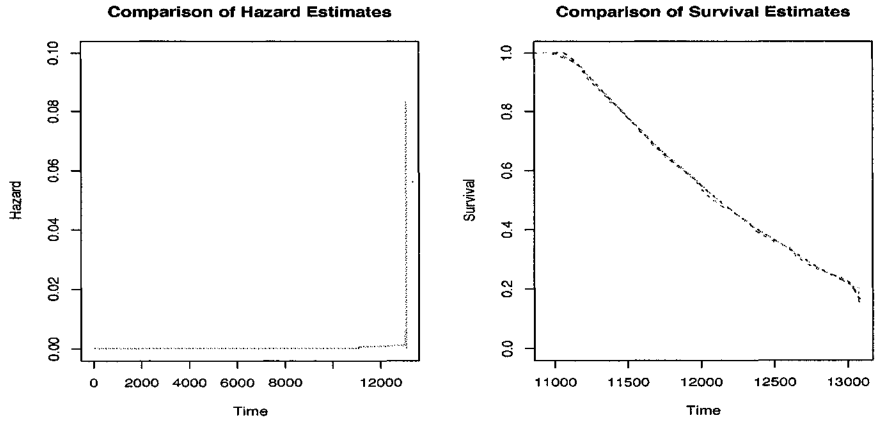

Figure 4–16: comparison of results from full CSHA data set

CHAPTER 5
Conclusion and Future Directions

5.1 Conclusion

The method proposed, although it clearly has many advantages including improved effciency, does have several disadvantages when compared to Tsai's estimator. The first and foremost is the robustness of Tsai's estimator to deviations from the stationarity assumption. When stationarity is a reasonable assumption, such as those mentioned in previous chapters, one gains more efficienct by taking the estimator introduced in this thesis.

The second disadvantage of this method is its sensitivity to numerical error. If there are two observations that are very close or a large gap in the data, the method may fail. There is also an inherent problem with the scaling scheme proposed above. We suggest scaling a_i^* in order for $R(z_l+)(z_{l+1} - z_l)$ to be on the same scale as $a_{(i-1)}^*(z_l)/\hat{\lambda}^{(i-1)}(z_l)$. Unfortunately, as λ is often far from constant, it is sometimes not possible to make the above quantities comparable throughout the entire data set. Although these challenges are not theoretically relevant, they can have a quite a big impact on the estimator.

The final limitation to note is the relative computational inefficiency of the suggested algorithm when compared to past methods. Tsai's method required only one ‘iteration’; the suggested method requires many more. The calculation of the a_i^* is also cumbersome at each iteration.

5.2 Future Directions

In order to deal with the problems with numerical error, one might have tried two suggestions if not for time considerations. The first is to set observations very close to each other as ties. Although this would result in a loss of information, controlling the tolerance could prevent the numerical problems while minimizing the loss of information. The second suggestion involves the substitution of $b = \sum_{i=1}^n b_i$ where each b_i can be controlled to allow for the scaling of $a^*_{(i-1)}(z_l)/\hat{\lambda}^{(i-1)}(z_l)$ at each point independently or in sections, thus alleviating the local nature of scaling.

The main goal of future studies should be to establish the consistency of the estimator. Tsai [23] was able to show the consistency of his estimator without trouble; however, the implicit definition of our case adds a new level of complexity. A bootstrap would also be useful for comparing the efficiency of the proposed method with that of past methods. Distributional results would also clearly be of interest. Wellner and Groeneboom [13] have written extensively on this kind of estimation and a careful study of their literature could prove fruitful.

Appendix: Code

```
#include<gsl/gsl_vector.h>

#include<gsl/gsl_matrix.h>

#include<gsl/gsl_rng.h>

#include<gsl/gsl_randist.h>

#include<gsl/gsl_sort_vector.h>

#include<gsl/gsl_permutation.h>

#include<gsl/gsl_permute_vector.h>


#include<stdio.h>

#include<math.h>

#include <iostream>

using namespace std;


const double epsilon = 1E-50;

const double tol = 1E-5;


//find the signum of a double, with signum(0):=1

double signum(double x) {

if (x < 0)
```

```
return(-1);
else
return(1);
}

//use logs to multiply two numbers
double logmult(double x, double y) {
if ((x<epsilon) || (y<epsilon))
return(x*y);
else {
double sign = signum(x)*signum(y);
return(sign*exp(log(fabs(x))+log(fabs(y))));
}
}

//use logs to divide two numbers
double logdiv(double x, double y) {
if ((x<epsilon) || (y<epsilon))
return(x/y);
else {
double sign = signum(x)*signum(y);
return(sign*exp(log(fabs(x))-log(fabs(y))));
}
```

```
}

//the number of exactly observed failures at this point
int d(double y,gsl_vector *yy,gsl_vector *ddelta, int n, int k) {
int temp = 0;
for (int i=0;i<n;i++) {
if ( (fabs(gsl_vector_get(yy,i)-y)<epsilon)
&& (fabs(gsl_vector_get(ddelta,i)-1.0)<epsilon) )
temp++;
}
return(temp);
}

//the number of subjects in the risk set at this time
int R(double y,gsl_vector *tt,gsl_vector *yy, int n, int k) {
int temp = 0;
for (int i=0;i<n;i++) {
if ( ((gsl_vector_get(tt,i)<=y) && (y<=gsl_vector_get(yy,i)))
|| (fabs(y-gsl_vector_get(yy,i))<epsilon)
|| (fabs(y-gsl_vector_get(tt,i))<epsilon) ) {
temp++;
}
}
```

```
return(temp);

}



//estimation of survival for the a_i's using

the estimator from Padgett & Wei (1980)

double S(double z,gsl_vector *llambdac,gsl_vector *ddelta, gsl_vector *yy,

gsl_vector *tt, gsl_vector *zz,gsl_vector *ztemp, gsl_vector *tinz,

gsl_vector *dd, int n, int k,gsl_vector *exactfails,int numexactfails) {

double temp = 0.0;


int i=0;

while ((i<(numexactfails-1))

&(gsl_vector_get(zz,(int)gsl_vector_get(exactfails,i))<=z)) {

temp += gsl_vector_get(dd,(int)gsl_vector_get(exactfails,i))

*gsl_vector_get(llambdac,(int)gsl_vector_get(exactfails,i))

*(min(z,gsl_vector_get(zz,(int)gsl_vector_get(exactfails,i+1)))

-gsl_vector_get(zz,(int)gsl_vector_get(exactfails,i)));

i++;

}


return(exp(-temp));

}
```

```
//create vectors of the d's, and R's at the z_i's
void makedrvec(gsl_vector *yy,gsl_vector *tt,gsl_vector *ddelta,
gsl_vector *zz,gsl_vector *ztemp, gsl_vector *tinz, gsl_vector *dd,
gsl_vector *RR,FILE *f, int n, int k) {
for (int i=0;i<k;i++) {
gsl_vector_set(dd,i,(double)d(gsl_vector_get(zz,i),yy,ddelta,n,k));
gsl_vector_set(RR,i,(double)R(gsl_vector_get(zz,i)+10*epsilon,tt,yy,n,k));
}
}

//create vectors of the S's at the z_i's
void makesvec(gsl_vector *llambdac,gsl_vector *ddelta,gsl_vector *zz,
gsl_vector *ztemp, gsl_vector *tinz,gsl_vector *yy,gsl_vector *tt,
gsl_vector *SS, gsl_vector *dd, int n, int k,
gsl_vector *exactfails,int numexactfails) {

for (int i=0;i<k;i++)
gsl_vector_set(SS,i,S(gsl_vector_get(zz,i),llambdac,ddelta,
yy,tt,zz,ztemp,tinz,dd,n,k,exactfails,numexactfails));
}

//calculate a-star (including the log)
```

```
double astar(double z,gsl_vector *zz,gsl_vector *ztemp,

gsl_vector *tinz, gsl_vector *tt, gsl_vector *yy, gsl_vector *ddelta,

gsl_vector *llambdac, gsl_vector *dd,gsl_vector *RR,gsl_vector *SS,

gsl_permutation *p,gsl_vector *pp,FILE *f, int n, int k, double astarscale,

gsl_vector *exactfails,int numexactfails) {



//find the t corresponding to z

int place = -999;

for (int i=0;i<n;i++)

if (fabs(gsl_vector_get(tt,i)-z)<epsilon) place=i;


//output the quotient

if (place == -999)

return 1E-300;

else {


//make the numerator

double numer = 0;

numer = logmult(S(z,llambdac,ddelta,yy,tt,zz,ztemp,tinz,

dd,n,k,exactfails,numexactfails),exp(astarscale));


//make the denominator
double denom = 1*gsl_vector_get(zz,0); // the zero-th term
for (int i=1;i<k;i++)
denom = denom + logmult(gsl_vector_get(SS,i),
(gsl_vector_get(zz,i)-gsl_vector_get(zz,i-1)));

double retval;
if (fabs(denom)>(1/tol)*fabs(numer))
retval=epsilon;
else
if (fabs(denom)<(tol)*fabs(numer))
retval=1/epsilon;
else
retval=logdiv(numer,denom);

if (retval <= epsilon) {
retval = epsilon;
return(-1/epsilon);
} else {
return(log(retval));
}

}
```

```
}

//make a-star terms (including the log)
double makeastarvec(gsl_vector *astarvec,gsl_vector *zz,gsl_vector *ztemp,
gsl_vector *tinz, gsl_vector *tt, gsl_vector *yy, gsl_vector *ddelta,
gsl_vector *llambdac, gsl_vector *dd,gsl_vector *RR,gsl_vector *SS,
gsl_permutation *p,gsl_vector *pp,FILE *f, int n, int k, double astarscale,
gsl_vector *exactfails,int numexactfails) {

for (int i=0;i<n;i++)
gsl_vector_set(astarvec,i,astar(gsl_vector_get(tt,i),zz,
ztemp,tinz,tt,yy,ddelta,llambdac,dd,RR,SS,p,pp,f,n,k,
astarscale,exactfails,numexactfails));
}

//make the terms to minimax
double terms(gsl_vector *zz,gsl_vector *ztemp, gsl_vector *tinz,
int ss, int qq, gsl_vector *yy, gsl_vector *tt, gsl_vector *ddelta,
gsl_vector *llambdac, gsl_vector *dd,gsl_vector *RR,
gsl_vector *SS,gsl_permutation *p,gsl_vector *pp,
FILE *f, int n, int k, double astarscale,gsl_vector *exactfails,
int numexactfails, gsl_vector *astarvec) {
```

```
double numer = 0;

double denom = 0;

double astartemp = 0;


for (int l=ss;l<=qq;l++) {

numer +=  (double)gsl_vector_get(dd,l);

double Rtemp = ((double)gsl_vector_get(RR,l));

if (gsl_vector_get(tinz,l)>0) {

astartemp = gsl_vector_get(astarvec,(int)gsl_vector_get(tinz,l));

} else

astartemp = 1E-300;


if ((fabs(astartemp) > tol*epsilon)

&& (fabs(gsl_vector_get(llambdac,l)) > tol*epsilon)) {

denom += logmult(Rtemp,(gsl_vector_get(zz,l+1)-gsl_vector_get(zz,l)))

-logdiv(astartemp,gsl_vector_get(llambdac,l));

-log(gsl_vector_get(llambdac,l))));

} else {

if ((fabs(gsl_vector_get(llambdac,l)) <= tol*epsilon)

&& (fabs(astartemp) > epsilon)) {

denom  = 999/epsilon;

break;

} else {
```

```
if ((fabs(gsl_vector_get(llambdac,l)) > epsilon)
&& (fabs(astartemp) <= tol*epsilon)) {
denom += logmult(Rtemp,(gsl_vector_get(zz,l+1)-gsl_vector_get(zz,l)));
} else {
denom += logmult(Rtemp,(gsl_vector_get(zz,l+1)-gsl_vector_get(zz,l)))
-logdiv(astartemp,gsl_vector_get(llambdac,l));
}
}
}
}


double retval;
if (fabs(denom)>(1/tol)*fabs(numer)) {
retval=epsilon;
} else {
if (fabs(denom)<(tol)*fabs(numer)) {
retval=1/epsilon;
} else {
retval=logdiv(numer,denom);
}
}
```

```
return(retval);

}

//make the matrix of terms to minimax
void makemat(gsl_vector *zz,gsl_vector *ztemp, gsl_vector *tinz,
gsl_vector *yy, gsl_vector *tt, gsl_vector *ddelta, gsl_vector *llambdac,
gsl_matrix *tmat, gsl_vector *dd,gsl_vector *RR,gsl_vector *SS,
gsl_permutation *p,gsl_vector *pp,FILE *f, int n, int k, double astarscale,
gsl_vector *exactfails, int numexactfails, gsl_vector *astarvec) {

for (int i=0;i<(k-1);i++)
for (int j=0;j<(k-1);j++)
gsl_matrix_set(tmat,i,j,99999.0);

for (int s=0;s<(k-1);s++)
for (int q=s;q<(k-1);q++) {
gsl_matrix_set(tmat,s,q,terms(zz,ztemp,tinz,s,q,yy,tt,
ddelta,llambdac,dd,RR,SS,p,pp,f,n,k,astarscale,
exactfails,numexactfails,astarvec));
if ((gsl_matrix_get(tmat,s,q))<=0)
gsl_matrix_set(tmat,s,q,epsilon);
}
```

```
}

void maximin(gsl_vector *zz,gsl_vector *ztemp, gsl_vector *tinz,
gsl_vector *yy, gsl_vector *tt, gsl_vector *ddelta, gsl_matrix *llambda,
gsl_vector *dd,gsl_vector *RR,gsl_vector *SS,gsl_permutation *p,
gsl_vector *pp,FILE *f,int n, int k, int maxiter, double astarscale,
gsl_vector *exactfails, int numexactfails) {

//make the original stab at the hazard
for (int i=0;i<(k);i++)
gsl_matrix_set(llambda,i,0,(1.0/(k-i)));

gsl_vector *llambdac = gsl_vector_alloc(k);
gsl_matrix *tmat = gsl_matrix_calloc(k-1,k-1);
gsl_vector *mins = gsl_vector_calloc(k);
gsl_vector *tomin = gsl_vector_calloc(k);
gsl_vector *tomax = gsl_vector_calloc(k);
gsl_vector *temp = gsl_vector_calloc(k);
gsl_vector *astarvals = gsl_vector_calloc(k);
gsl_vector *astarvalsc = gsl_vector_calloc(k);
gsl_vector *astarcomp = gsl_vector_calloc(k);
gsl_vector *astarvec = gsl_vector_calloc(n);

short initial = 0;
int startscale = 0;
int endscale = 15;
double scalefrom = 0.0;
double hazscale = 10*tol;
double scaleto = 0.0;

for (int i=1;i<(maxiter);i++) {
cout << "Iteration:  i = " << i << endl;
gsl_matrix_get_col(llambdac,llambda,i-1);
makesvec(llambdac,ddelta,zz,ztemp,tinz,yy,tt,SS,
dd,n,k,exactfails,numexactfails);
makeastarvec(astarvec,zz,ztemp,tinz,tt,yy,ddelta,
llambdac,dd,RR,SS,p,pp,f,n,k,astarscale,
exactfails,numexactfails);
makemat(zz,ztemp,tinz,yy,tt,ddelta,llambdac,tmat,dd,RR,
SS,p,pp,f,n,k,astarscale,exactfails,numexactfails,astarvec);


for (int j=0;j<(k-1);j++) {
gsl_vector_set_all(mins,-99999.9);
gsl_vector_set_all(tomax,-99999.9);

gsl_vector_set_all(tomin,99999.9);

for (int r=0;r<=j;r++) {

gsl_vector_set_all(tomin,99999.9);

for (int s=j;s<(k-1);s++)

gsl_vector_set(tomin,s,gsl_matrix_get(tmat,r,s));

gsl_vector_set(mins,r,gsl_vector_min(tomin));

}


gsl_matrix_set(llambda,j,i,gsl_vector_max(mins));

}




// adjust scaling

scaleto = 0.0;

for (int j=0;j<(k-1);j++) {

gsl_vector_set(astarcomp,j,logmult(gsl_vector_get(RR,j),

(gsl_vector_get(zz,j+1)-gsl_vector_get(zz,j))));

if ((j>=startscale)&&(j<=endscale))

scaleto+=gsl_vector_get(astarcomp,j);

}

scalefrom = 0.0;
```

```
hazscale = 0.0;

int l = 0;

for (int j=0;j<k;j++) {

gsl_vector_set(astarvalsc,j,logdiv(astar(gsl_vector_get(zz,j),

zz,ztemp,tinz,tt,yy,ddelta,llambdac,dd,RR,SS,p,pp,f,n,k,

astarscale,exactfails,numexactfails),gsl_matrix_get(llambda,j,i-1)));

gsl_vector_set(astarvals,j,astar(gsl_vector_get(zz,j),zz,ztemp,tinz,tt,

yy,ddelta,llambdac,dd,RR,SS,p,pp,f,n,k,astarscale,exactfails,numexactfails));


if (((fabs(gsl_vector_get(astarvalsc,j))>epsilon)

&&(fabs(gsl_vector_get(astarvalsc,j))<1/epsilon))

&&(fabs(gsl_matrix_get(llambda,j,i-1))>epsilon)) {

if ((j>=startscale)&&(j<=endscale)) {

scalefrom+=gsl_vector_get(astarvalsc,j);

hazscale+=gsl_matrix_get(llambda,j,i-1);

l++;

}

}

}


scaleto *= 1;

if (l>0) {

cout << "scalefrom = " << scalefrom << ";
```

```
scaleto = " << scaleto << "; hazscale = "
<< hazscale << "; l = " << l <<  endl;
astarscale += (scaleto-scalefrom)/l*(hazscale/l);
cout << "astarscale set to " << astarscale << endl;
} else cout << "Scaling scheme continued..." << endl;

if (i==3) {
FILE *ffastar = fopen("astar.txt","w");
gsl_vector_fprintf(ffastar,astarvals,"%f");
fclose(ffastar);
FILE *ffastarc = fopen("astarc.txt","w");
gsl_vector_fprintf(ffastarc,astarvalsc,"%f");
fclose(ffastarc);

FILE *ffother = fopen("other.txt","w");
gsl_vector_fprintf(ffother,astarcomp,"%f");
fclose(ffother);
}

if ((i==1)&&(initial==0)) {
initial = 1;
i--;
cout << "Pre-run complete" << endl;
```

```
}

}

//clean up
gsl_vector_free(astarcomp);
gsl_vector_free(astarvals);
gsl_vector_free(astarvalsc);
gsl_vector_free(astarvec);
gsl_vector_free(mins);
gsl_vector_free(tomin);
gsl_vector_free(tomax);
gsl_vector_free(llambdac);
gsl_matrix_free(tmat);
gsl_vector_free(temp);
}

// Weibull hazard
double wbhaz(double z, double a, double b) {
return(b/a*pow((z/a),(b-1)));
}

int main(int argc, char *argv[]) {
```

```
const int n = atoi(argv[1]);

const int k = atoi(argv[2]);

const double a = atof(argv[3]);

const double b = atof(argv[4]);

const double datascale = atof(argv[5]);

double astarscale = atof(argv[6]);

const int maxiter = atoi(argv[7]);

const int seed = atoi(argv[8])+2;


FILE *f = fopen ("out.txt","wt");

gsl_matrix *lambda = gsl_matrix_calloc(k,maxiter);

gsl_vector *x = gsl_vector_calloc(n);

gsl_vector *y = gsl_vector_calloc(n);

gsl_vector *u = gsl_vector_calloc(n);

gsl_vector *z = gsl_vector_calloc(k);

gsl_vector *c = gsl_vector_calloc(n);

gsl_vector *t = gsl_vector_calloc(n);

gsl_vector *tind = gsl_vector_calloc(n);

gsl_vector *dd = gsl_vector_calloc(k);

gsl_vector *RR = gsl_vector_calloc(k);

gsl_vector *SS = gsl_vector_calloc(k);

gsl_vector *delta = gsl_vector_calloc(n);

gsl_vector *finalhaz = gsl_vector_calloc(k);
gsl_vector *truehaz = gsl_vector_calloc(k);
gsl_vector *ztemp = gsl_vector_calloc(2*n);
gsl_vector *dtemp = gsl_vector_calloc(2*n);
gsl_vector *tinz = gsl_vector_calloc(k);

const gsl_rng_type * T;
gsl_rng_env_setup();
T = gsl_rng_default;
gsl_rng *r = new gsl_rng;
r = gsl_rng_alloc(T);
gsl_rng_set(r,seed);

//read y,t, and delta from file
FILE *h = fopen("t.txt","r");
gsl_vector_fscanf(h,t);
fclose(h);

FILE *g = fopen("y.txt","r");
gsl_vector_fscanf(g,y);
fclose(g);

FILE *ff = fopen("delta.txt","r");
gsl_vector_fscanf(ff,delta);
```

```
fclose(ff);

//combine the t's and y's to make the z's
for (int i=0;i<n;i++) {
gsl_vector_set(ztemp,i,gsl_vector_get(t,i));
gsl_vector_set(ztemp,i+n,gsl_vector_get(y,i));
gsl_vector_set(dtemp,i+n,gsl_vector_get(delta,i));
}
gsl_permutation *p = gsl_permutation_calloc(2*n);
gsl_vector *pp = gsl_vector_calloc(2*n);
gsl_sort_vector_index(p,ztemp);
gsl_sort_vector(ztemp);
gsl_permute_vector(p,dtemp);

int l=1;
gsl_vector_set(pp,0,0);
gsl_vector_set(z,0,gsl_vector_get(ztemp,0));
gsl_vector_set(dd,0,gsl_vector_get(dd,0));
for (int i=1;i<2*n;i++) {
if (fabs(gsl_vector_get(ztemp,i)-gsl_vector_get(ztemp,i-1))>epsilon) {
gsl_vector_set(z,l,gsl_vector_get(ztemp,i));
gsl_vector_set(dd,l,gsl_vector_get(dtemp,i));
gsl_vector_set(pp,l,i);
l++;

} else {

gsl_vector_set(dd,l-1,gsl_vector_get(dtemp,i)+gsl_vector_get(dd,l-1));

}

}




//make the vector tinz

int i=0;

int j=0;

while (j<(2*n)) {

if (fabs(gsl_vector_get(z,i)-gsl_vector_get(ztemp,j))>tol)

i++;

if (gsl_permutation_get(p,j)<n)

gsl_vector_set(tinz,i,gsl_permutation_get(p,j));

j++;

}


cout << "tinz= ";

for (int i=0;i<k;i++)

cout << gsl_vector_get(tinz,i) << " ";

cout << endl;
cout << "z= ";
for (int i=0;i<k;i++)
cout << gsl_vector_get(z,i) << " ";
cout << endl;

cout << "pp= ";
for (int i=0;i<(2*n);i++)
cout << gsl_vector_get(pp,i) << " ";
cout << endl;



cout << "dd= ";
int ll=0;
for (int i=0;i<k;i++){
cout << gsl_vector_get(dd,i) << " ";
if (gsl_vector_get(dd,i)>epsilon) ll++;
}
cout << endl;
cout << "numexactfails = " << ll << endl;
int numexactfails = ll;

//for the weibull case:
FILE *fft = fopen("wbhaz.txt","w");
```

```
for (int i=0;i<k;i++)

gsl_vector_set(truehaz,i,wbhaz(gsl_vector_get(z,i),a,b));

gsl_vector_fprintf(fft,truehaz,"%f");

fclose(fft);


//scale the data

  gsl_vector_scale(z,datascale);

  gsl_vector_scale(t,datascale);

  gsl_vector_scale(y,datascale);




//do the algorithm


makedrvec(y,t,delta,z,ztemp,tinz,dd,RR,f,n,k);

gsl_vector *exactfails = gsl_vector_calloc(numexactfails);


j=0;

for (int i=0;i<k;i++) {

if (gsl_vector_get(dd,i)>epsilon) {

gsl_vector_set(exactfails,j,i);

j++;

}

}
```

```
cout << "exactfails= ";
for (int i=0;i<numexactfails;i++)
cout << gsl_vector_get(exactfails,i) << " ";
cout << endl;

maximin(z,ztemp,tinz,y,t,delta,lambda,dd,RR,SS,
p,pp,f,n,k,maxiter,astarscale,exactfails,numexactfails);

FILE *ffmat = fopen("lambda_mat.txt","w");
gsl_matrix_fprintf(ffmat,lambda,"%f");
fclose(ffmat);

FILE *fd = fopen("d.txt","w");
gsl_vector_fprintf(fd,dd,"%f");
fclose(fd);

FILE *ffhaz = fopen("final.txt","w");
gsl_matrix_get_col(finalhaz,lambda,maxiter-1);
gsl_vector_fprintf(ffhaz,finalhaz,"%f");
fclose(ffhaz);
FILE *ffr = fopen("R.txt","w");
gsl_vector_fprintf(ffr,RR,"%f");
```

```
fclose(ffr);


FILE *ffs = fopen("S.txt","w");

gsl_vector_fprintf(ffs,SS,"%f");

fclose(ffs);


gsl_vector_free(tinz);

gsl_vector_free(ztemp);

gsl_vector_free(dtemp);

gsl_vector_free(exactfails);

gsl_matrix_free(lambda);

gsl_vector_free(finalhaz);

gsl_vector_free(truehaz);

gsl_vector_free(x);

gsl_vector_free(y);

gsl_vector_free(u);

gsl_vector_free(z);

gsl_vector_free(c);

gsl_vector_free(t);

gsl_vector_free(tind);

gsl_vector_free(dd);

gsl_vector_free(RR);

gsl_vector_free(SS);
gsl_vector_free(delta);

gsl_permutation_free(p);

gsl_vector_free(pp);

gsl_rng_free(r);

fclose(f);

return(1);

}
```

References

[1] V. Addona, M. Asgharian, and D. B. Wolfson. Nonparametric maximum likelihood estimation of the incidence rate using data from a prevalent cohort study with follow-up. Submitted, 2007.

[2] V. Addona and D. B. Wolfson. A formal test for the stationarity of the incidence rate using data from a prevalent cohort study with follow-up. *Lifetime Data Anal.*, 12(3):267–284, Sep 2006.

[3] M. Asgharian, C. E. M'Lan, and D. B. Wolfson. Length-biased sampling with right censoring: An unconditional approach. *Journal of the American Statistical Association*, 97(457):201–209, 2002.

[4] M. Asgharian and D. B. Wolfson. Asymptotic behavior of the unconditional npmle of the length-biased survivor function from right censored prevalent cohort data. *The Annals of Statistics*, 33(5), 2005.

[5] M. Asgharian, D. B. Wolfson, and Xun Zhang. Checking stationarity of the incidence rate using prevalent cohort survival data. *Statistics in Medicine*, 25:1751–1767, 2006.

[6] M. Banerjee. Estimating monotone, unimodal and u-shaped failure rates using asymptotic pivots. *to appear in Statistica Sinica*, 2007.

[7] R.E. Barlow, D.J. Bartholomew, J. M Bremner, and H. D. Brunk. *Statistical Inference under Order Restrictions: The Theory and Application of Isotonic Regression.* John Wiley and Sons, 1972.

[8] D. R. Cox. Some sampling problems in technology. In *New Developments in Survey Sampling.* Wiley Interscience, 1969.

[9] J. de Una-Alvarez. Nonparametric estimation under length-biased sampling and type i censoring: a moment based approach. *Annals of the Institute of Statistical Mathematics*, 56(4):667–681, 2004.

[10] R. A. Fisher. The sampling distribution of some statistics obtained from nonlinear equations. *The Annals of Eugenics*, 9:238–249, 1039.

[11] P. L. Goldsmith. The calculation of true particle size distributions from the sizes observed in a thin slice. *British Journal of Applied Physics*, 18:813–830, 1967.

[12] U. Grenander. On the theory of mortality meaurement, part ii. *Skandinavisk Aktuarietidskrift*, pages 71–153, 1956.

[13] P. Groeneboom and J. Wellner. *Information Bounds and Nonparametric Maximum Likelihood Estimation.* Birkhauser, 1992.

[14] J. Huang and J. Wellner. Estimation of a monotone density or monotone hazard under random censoring. *Scandinavian Journal of Statistics. Theory and Applications*, 22(1):3–33, 1995.

[15] J. Hyde. Testing survival under right censoring and left truncation. *Biometrika*, 64:225–230, 1977.

[16] J. Klein and M. Moeschberger. *Survival Analysis.* Springer, 1997.

[17] T. Lancaster. *The Econometric Analysis of Transition Data.* Cambridge University Press, 1990.

[18] D. Lynden-Bell. A method of allowing for known observational selection in small samples applied to 3cr quasars. *Monthly Notices of the Royal Astronomical Society*, 155:95–118, 1971.

[19] A. W. Marshall and F. Proschan. Maximum likelihood estimation for distributions with monotone failure rate. *The Annals of Mathematical Statistics*, 36(1):69–77, 1965.

[20] W. J. Padgett and L. J. Wei. Maximum likelihood estimation of a distribution function with increasing failure rate based on censored observations. *Biometrika*, 67(2):470–474, 1980.

[21] G. P. Patil and C. R. Rao. Weighted distributions and size-biased sampling with applications to wildlife populations and human families. *Biometrics*, 34(2):179–189, 1978.

[22] B. L. S. P. Rao. Estimation for distributions with monotone failure rate. *The Annals of Mathematical Statistics*, 41(2):507–519, 1970.

[23] W-Y. Tsai. Estimation of the survival function with increasing failure rate based on left truncated and right censored data. *Biometrika*, 7(2):319–324, 1988.

[24] W-Y. Tsai, N. P. Jewell, and M-C. Wang. A note on the product-limit estimator under right censoring and left truncation. *Biometrika*, 7(4):883–886, 1987.

[25] Y. Vardi. Nonparametric estimation in the presence of length bias. *The Annals of Statistics*, 10:616–620, 1982.

[26] Y. Vardi. Empirical distributions in selection bias models (with discussion). *The Annals of Statistics*, 13:178–205, 1985.

[27] Y. Vardi. Multiplicative censoring, renewal processes, deconvolution and decreasing density: Nonparametric estimation. *Biometrika*, 76(4):751–761, 1989.

[28] M-C. Wang. Nonparametric estimation from cross-sectional survival data. *Journal of the American Statistical Association*, 86(413):130–143, 1991.

[29] S.D. Wicksell. The corpuscle problem: a mathematical study of a biometric problem. *Biometrika*, 17:84–99, 1925.

[30] C. Wolfson, D. B. Wolfson, M. Asgharian, C. E. M'Lan, T. Ostbye, K. Rockwood, and D. B. Hogan. A reevaluation of the duration of survival after the onset of dementia. *New England Journal of Medicine*, 344:1111–1116, 2001.

[31] M. Woodroofe. Estimating a distribution function with truncated data. *The Annals of Statistics*, 13(1):163–177, 1985.

[32] M. Zelen and M. Feinleib. On the theory of screening for chronic diseases. *Biometrika*, 56(3):601–614, 1969.