NON-PARAMETRIC AND SEMI-PARAMETRIC ESTIMATION IN FORWARD AND BACKWARD RECURRENCE TIME DATA

Pourab Roy

A dissertation submitted to the faculty at the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Biostatistics in the Gillings School of Global Public Health.

Chapel Hill
2015

Approved by:

Michael R. Kosorok

Jason P. Fine

Chirayath Suchindran

David Couper

Donglin Zeng

Stephen Cole

© 2015
Pourab Roy
ALL RIGHTS RESERVED

ABSTRACT

Pourab Roy: Non-parametric and Semi-parametric Estimation in Forward and Backward Recurrence Time Data
(Under the direction of Michael R. Kosorok and Jason P. Fine)

In prevalent cohort survival studies where subjects are recruited at a cross-section and followed prospectively in time, the observed event times are length-biased and further follow a multiplicative censoring scheme. For such studies there is an associated initiation time which may be unknown. In this case we only observe the time from sampling to the event of interest. This is the forward recurrence time. Further in such cases standard left-truncation survival analysis methods are not applicable. In other scenarios like current duration studies, the time of the initiating event may be known but there is no subsequent follow-up after sampling. Here we observe the backward recurrence times. In presence of covariates, the proportional hazards model may not be applicable to forward and backward recurrence time data. However, due to the invariance of the accelerated failure time model under length bias and cross-sectional sampling, it can serve as a useful alternative. In particular, existing estimators for the regression parameter like the ordinary least squares and Tsiatis log rank estimators may be valid. The problem however is that these estimators are based on the conditional distribution of the time variable given the covariates. Under length bias sampling, the covariate distribution is functionally dependent on the regression parameter. Thus a "naive" analysis conditioning on the covariates may result in information loss. We show that if the covariate distribution is left completely unspecified then there is no loss of information under a conditional analysis. We also perform simulation studies

to compare our method to the existing methods for forward and backward recurrence time data analysis. Finally, we analyze time-to-pregnancy data comparing our method to ordinary least squares regression.

Next, we show the connection between k-monotone densities and forward and backward recurrence time data. We show that if we start with a k-monotone density, the corresponding recurrence time density is (k+1)-monotone. So, to use k-monotone density estimation for forward recurrence time data, we develop an algorithm for consistent estimation of a k-monotone density under right censoring. We determine the rate of convergence and asymptotic distribution of the proposed estimator. We look at the viability of the estimator under some simulation settings and also apply it to the ARIC data.

Finally, we look at the effect of recurrence time on competing risks. We determine the recurrence time subdistributions and also develop an algorithm for estimating the original subdistributions. We show the consistency and determine the asymptotic distribution of the proposed estimator. We look at simulation results to determine the efficacy of the estimator and also a data application for the ARIC data.

ACKNOWLEDGEMENTS

I would like to express my profound gratitude to my teacher and advisor Professor Michael Kosorok for guiding me through this research. His invaluable suggestions and constant encouragement have been a great source of inspiration for me. The many mentoring sessions with Prof. Kosorok that I had the privilege of experiencing, helped unfold myriad perspectives on the subject which I would never have grasped on my own. My foray into academics has been an immensely enriching experience due to the way in which Professor Kosorok gradually drew me into his intellectual ambit. I am really grateful for the time he spared to advise me every step of the way and for the infinite patience with which he dealt with my lapses.

I would also like to thank my co-advisor Professor Jason Fine. He has helped stimulate my interest in the subject in so many ways. During the discussions that he held with me I often found solutions to problems that I had been grappling with for days, solutions which seemed so obvious after he had pointed them out to me. The academic dialogues initiated by him were simply fascinating and left a deep imprint on my mind. Prof. Fine has always been very kind to me and has forgiven me for all my mistakes. I would like to thank my committee members: Dr. Chirayath Suchindran, Dr. David Couper, Dr. Stephen Cole and Dr. Donglin Zeng for their insightful knowledge, attention to technical detail and overall support for my projects. I have been extremely fortunate in having them as my mentors.

I would like to say thank you to my parents who have been pillars of support. They have always set the bar high and impelled me to pursue lofty academic goals. Without their prodding I would not have been writing this today.

I have innumerable friends who have contributed in various ways in making this venture a happy and meaningful one. Be it class and campus support or simply help in dispersing the blues, they have never let me feel that I was so far away from home. Sayan Dasgupta, my senior by a year has been my counsel and source of support in all kinds of situations. The camaraderie that I shared with Anjishnu, Agniva, JD, Shalini, Sebastian, Ludmila, Anze, Brad and my friends at the table tennis club have made life within and outside the university a particularly memorable one. I would also like to take his opportunity to thank Poulami Maitra, the other "pea" in my pod, who has always supported me through the good times and the bad.

Finally I would like to dedicate this thesis to my late grandparents, who have always believed in me and encouraged me to try my best in every endeavor.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1: INTRODUCTION

A prevalent cohort consists of subjects who have experienced an initiating event, like disease onset, prior to their entry to the study and who are followed forward in time until another (terminating) event, like death or symptom development. Sampling may be achieved in a small cross-section. In some prevalent cohort study the onset time may be unobservable as in HIV sero-prevalence studies (Brookmeyer and Gail 1987) where the time of infection to HIV is unknown and interest lies in the follow-up time from enrollment to AIDS. Such prevalent cohorts do not provide information on the time between the initiating event and the terminating event T, but only provide partial information in terms of the forward recurrence time T_f, the time from sampling to the terminating event time.

In other scenarios, the time of the initiating event may be known but there may not be any subsequent follow-up after cross-sectional sampling. This is known as the current duration study design, which is encountered, for example, in time to pregnancy surveys. In Keiding et al. (2002), the authors show that the distribution of the times from initiating attempt to cross-sectional sampling for couples that are currently attempting to get pregnant, identifies the distribution of the realized time to pregnancy or unsuccessful end of attempt. Another similar study based on current durations is Yamaguchi's mover-stayer study (Yamaguchi 2003). Such studies provide information on the backward recurrence time T_b.

Thus, we find that the forward and backward recurrence time data are going to be length-biased. This is because if the actual time to event (the distance between the

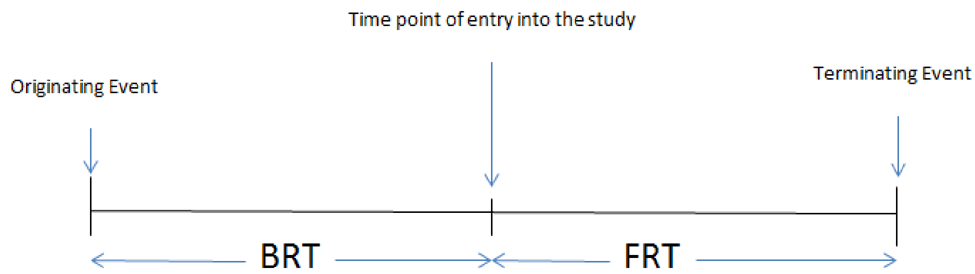


Figure 1.1: Pictorial Representation of FRT and BRT

originating event and the terminating event) was shorter than the cross-section time, then that data point would never be included in the study. Thus, longer times are favored in the sample. Hence, it is length biased.

1.1 Estimation of the Regression Parameter in the AFT model

In both prevalent cohort and current duration study designs, only subjects who have experienced the initiating event prior to sampling, but have not yet experienced the terminating event can be sampled. Thus both the forward and the backward recurrence times are length biased i.e. the sample is biased towards larger values of T. One way to model this bias (Cox (1969), Vardi (1982)) is to sample proportionally to length, i.e., if F_T is the distribution of T then the length-biased version T_{LB} has a distribution given by

$$F_{LB}(t) = \frac{\int_0^t u dF_T(u)}{\mu_T}, \quad t \geq 0,$$

where $\mu_T = \int_0^\infty u dF_T(u)$.

Further, if it can be assumed that the incidence of the disease follows a stationary Poisson process then the cross-sectional sampling time is distributed uniformly between the onset time and the terminating time (Cox (1969), Van Es et al. (2000), Keiding et al. (2002)). Thus $T_f = T_{LB}V$, where V is uniform(0,1). It follows that if $S_T = 1 - F_T$ is the survival function of T then for both T_f and T_b (commonly denoted as $\tilde{T}$ here), the density $g_{\tilde{T}}$ is given by

$$g_{\tilde{T}}(t) = \frac{S_T}{\mu_T}. \tag{1.1}$$

Another interesting way to look at it is the sum of the forward and backward recurrence time data. This sum yields the total length-biased time(T_{LB}) which is known as Stirling's interval. Given the total length-biased data, the backward recurrence time follows a Uniform$(0, T_{LB})$ distribution. So, one can start with the length-biased data and work backwards to estimate the recurrence time data. If both T_f and T_b are observed then standard left-truncated techniques apply. For example, consider the Canadian Health and Aging study of Dementia (Asgharian et al. 2002), where a cohort of elderly subjects were followed from diagnosis until death or end of study. Ages at onset were also ascertained from the subjects' caregivers. There has been extensive work towards estimating the length-biased distribution. Two common approaches found in the literature for left-truncated data, are the conditional and the unconditional likelihood approaches, where the conditioning is applied to the onset times. In the unconditional approach it is assumed that the distribution of the left truncation times are uniformly distributed, which holds under a certain stationarity condition discussed below. In the conditional approach one simply assumes the truncation distribution to be degenerate at the observed truncation times. In the unconditional approach, Vardi (1982) derived the non parametric maximum likelihood estimator (NPMLE) for the length-biased distribution arising out of prevalent cohorts. Asgharian and Wolfson

(Asgharian et al. 2002), (Asgharian and Wolfson 2005) derived the NPMLE for the unbiased incident-case survival function obtained from length-biased prevalent cohort data. The conditional perspective has been investigated by Wang et. al. (Wang et al. (1986), Wang (1991), Wang et al. (1993)). Wang et al. (1986) showed that when the truncating distribution is left completely unspecified, then there is little loss of information when the likelihood is conditioned on the truncation times.

In the presence of covariates, a popular semiparametric model is the proportional hazards (PH) model (Cox 1972) given by

$$\lambda_{T|Z}(t) = e^{\theta' Z}\lambda(t),$$

where, $\lambda_{T|Z}$ is the hazard function of T given the covariate vector Z and λ is an unspecified baseline hazard function. Here the density of T is given by $e^{\theta' z}\lambda(t)e^{-e^{\theta' z}\Lambda(t)}$, where, Λ is the cumulative hazard satisfying $\Lambda(0) = 0$. For the uncensored case, estimates for (θ, λ) can be derived using the partial likelihood based on risk sets (Cox 1972). For right censored data Tsiatis (1981) derives consistency and asymptotic normality of the maximum partial likelihood estimator for θ. Klaassen (1989) showed that the estimator is also semiparametric efficient. Details can also be found in Bickel et al. (1993), Murphy and Van der Vaart (2000) and Van der Vaart (1998).

For length-biased data arising out of left-truncation, Wang (1996) derives a consistent and asymptotically normal estimator for θ based on a modification of the risk set in order to adjust for the length-bias. Wang defines indicator variables $\Delta_j(t_i)$ for $t_i \leq t_j$, which equals 1 with probability t_i/t_j and 0 with probability $1-t_i/t_j$ and defines the modified risk set as

$$R_i^* = \{j : t_i \leq t_j, \Delta_j(t_i) = 1\}.$$

This allows the individuals in the risk set R_i^* to have the population risk set structure, i.e., the conditional probability of the ith subject to die at time t_i given that there is a death from the risk set R_i^* is given by

$$\frac{e^{\theta' z_i}}{\sum_{j \in R_i^*} e^{\theta' z_j}}.$$

Thus estimation of θ can be carried out by maximizing the above (pseudo) partial likelihood.

Under cross-sectional sampling of length-biased forward recurrence times, it is not clear how to modify the risk set in order to adjust for the selection bias. Brookmeyer and Gail (1987) discuss the different directions of the bias that might result from fitting a naive proportional hazards model to this kind of data arising out of prevalent cohort studies, where, the onset time of disease is unobservable.

The major difficulties in carrying out semiparametric inference for θ under the PH model and using forward or backward recurrence time data are:

1. Since the Cox model specifies the covariate effect on the hazard function and not the time variable itself, the effect of length-biased cross-sectional sampling on the covariate distribution is intractable. We shall see below that for the accelerated failure time model, this is not a problem.

2. Even if a naive analysis conditioned on Z is carried out, the derivation of the efficient score and information seems difficult. Consider the simplest case of uncensored backward recurrence times where we observe $Y = (T_b, Z)$. The log-likelihood for the naive analysis conditional on the covariates is given by

$$l_{\theta,\Lambda}(t, z) = -e^{\theta' z}\Lambda(t) - log \int e^{-e^{\theta' z}\Lambda(t)} dt.$$

The ordinary score for θ is

$$\dot{l}_{\theta,\Lambda}(t,z) = -ze^{\theta' z}\left\{\Lambda(t) - \frac{\int e^{-e^{\theta' z}\Lambda(t)}\Lambda(t)dt}{\int e^{-e^{\theta' z}\Lambda(t)dt}}\right\}.$$

Consider the parametric path $\eta \mapsto \Lambda_\eta(t) = \int_0^t (1+\eta h)d\Lambda(s)$, where $h \in L_2(\Lambda)$ and $|\eta| \approx 0$. Replacing Λ by Λ_η in the log-likelihood and differentiating at $\eta = 0$ we get the score for the nuisance parameter Λ as:

$$A_{\theta,\Lambda}h = e^{\theta' z}\left\{\int_0^t h(s)d\Lambda(s) + \frac{\int e^{-e^{\theta' z}\Lambda(t)}\int_0^t h(s)d\Lambda(s)dt}{\int e^{-e^{\theta' z}\Lambda(t)}dt}\right\}.$$

Deriving the adjoint operator $A^*_{\theta,\Lambda}$ (Murphy and Van der Vaart 2000) here is not straight forward as in the usual Cox model. As far as we know, a systematic study of forward and backward recurrence times under the proportional hazards assumption for the core time variable has yet to be undertaken.

A useful alternative to the PH model to model survival time T in the presence of covariates Z is the Accelerated Failure Time (AFT) model. Under the AFT model

$$T = e^{\theta' Z}U. \tag{1.2}$$

Here T is the failure time measured from the time of some initiating event like birth and disease onset, θ is a $p \times 1$ regression parameter, Z is a $p \times 1$ covariate vector with density h and U a non-negative random variable independent of Z with density g, survival function S and hazard $\lambda(u) = g(u)/S(u)$.

The problem of estimating θ in presence of the nuisance parameters g and h have been taken up by many authors. Two popular approaches are least-squares and rank based procedures. Bickel et al. (1993) derived the semiparametric efficient score and the efficient information for estimating θ when g and h are unspecified and given Z,

$\log T - \theta' Z$ is assumed independent of Z and the censoring variable C. Estimation of θ based on the least squares approach has been studied by Miller (1976), James and Buckley (1979) and Koul et al. (1981). Linear rank test based procedures using the partial likelihood score have been developed by Tsiatis (1990), Ritov (1990), Wei et al. (1990), Ying (1993), Fygenson and Ritov (1994) and Jin et al. (2003). For various submodels, these estimators are efficient. For example, the Buckley-James estimator is efficient when the true error density is standard normal while Tsiatis' linear rank test based estimator with unit weights is efficient for a class of extreme value error distributions. The latter is fully efficient if the weight function adaptively estimates λ'/λ, where λ is the hazard function corresponding to the error distribution.

Another important recent development in estimating θ in a tobit model and the one that we will build upon in section 3 for the random right censored case is Cosslett's (Cosslett (2004)) asymptotically efficient estimator via a smoothed self-consistency equation. Without covariates and right censoring, nonparametric estimation of the monotone baseline density g can be achieved using the Grenander (1956) estimator for a monotone density. For the right censoring case this is generalized by the Denby-Vardi (Denbi and Vardi (1986)) NPMLE. In the uncensored case, Woodroofe and Sun (1993) pointed out a technical difficulty in using the actual likelihood. They show that the Denby-Vardi NPMLE is inconsistent near the origin in the sense that $\hat{g}(0+)$ does converge in probability but the limit is strictly greater than $g(0+)$ almost surely. They propose using a penalized likelihood

$$L_\alpha(g) = \sum_{i=1}^{n} \log g(u_i) - n\alpha g(0+),$$

where g varies over a class of decreasing left-continuous densities and $\alpha > 0$ is a smoothing parameter. Keiding et. al. (2002) discuss that another way to avoid this inconsistency near zero is to estimate the survival function of T conditioned on $T > t_0$

for some small $t_0 > 0$ and then estimate $S(t)/S(t_0)$ by $\hat{g}(t)/\hat{g}(t_0)$.

1.2 2-monotone Density Estimation Under Censoring

The main motivation for this problem arises from the connection between recurrence times and k-monotone densities. A density is said to be k-monotone if $(-1)^j g^{(j)}$ is non-negative, non-increasing and convex for $j = 1(1)k-2$. Estimation of functions restricted by monotonicity or other inequality constraints has received much attention. Estimation of monotone regression and density functions goes has been done by Grenander (1956). Asymptotic distribution theory for monotone regression estimators was established by Brunk (1970), and for monotone density estimators by Prakasa Rao (1969). The asymptotic theory for monotone regression function estimators was reexamined by Wright (1981), and the asymptotic theory for monotone density estimators was reexamined by Groeneboom (1985). The "universal component" of the limit distribution in these problems is the distribution of the location of the maximum of two-sided Brownian motion minus a parabola. Groeneboom (1988) examined this distribution and other aspects of the limiting Gaussian problem with canonical monotone function $f_0(t) = 2t$ in great detail. Groeneboom (1985) provided an algorithm for computing this distribution, and this algorithm has recently been implemented by Groeneboom et al. (2001).

The first work on convex density function estimation has been done by Anevski (1994), who was motivated by some problems involving the migration of birds discussed by Hampel (1987). Jongbloed (2001) established lower bounds for minimax rates of convergence, and established rates of convergence for a "sieved maximum likelihood estimator". Finally, a least squares estimator as well as a non-parametric maximum likelihood estimator for 2-monotone densities were established by Groeneboom et al. (2001) which were further modified by Balabdaoui and Wellner (2007) to correct for the consistency near 0.

The least squares (LS) estimator $\tilde{f}_n$ of a convex decreasing density function f_0 is defined as a minimizer of the criterion function

$$Q_n(f) = \frac{1}{2}\int f(x)^2 dx - \int f(x) d\mathbb{F}_n(x),$$

over $\mathcal{K}$, the class of convex and decreasing nonnegative functions on $[0, \infty)$.

Our aim for this section is to adapt the existing methods for censored observations by utilizing the fact that the decreasing density assumption leads to well-behaved properties of the estimator.

1.3 Recurrence Time Density Estimation for Competing Risks

In this section, our main goal is to study nonparametric estimation for forward and backward recurrence time data with competing risks in the absence of covariates and censoring. The set-up is as follows. We analyze a system that can fail from K competing risks, where $K \in \mathbb{N}$ is fixed. The random variables of interest are (X, Y), where $X \in \mathbb{R}$ is the failure time of the system, and $Y \in \{1, \ldots, K\}$ is the corresponding failure cause. We cannot observe (X, Y) directly. Rather, we observe the corresponding recurrence time failure $T \in \mathbb{R}$. This means that at time T, we observe that the failure occurred and we also observe the failure cause Y . Such data can arise naturally in cross-sectional studies with several failure causes.

The Kaplan-Meier estimator can easily be generalized to include competing risks. Let $t_{j1} < t_{j2} < \cdots < t_{jk_j}$ denote the k_j distinct failure times for failures of type j. Let n_{ji} denote the number of subjects at risk just before t_{ji} and let d_{ji} denote the number of deaths due to cause j at time t_{ji}. Then the same arguments used to derive the usual

K-M estimator lead to

$$\hat{S}_j(t) = \prod_{i:t_{ji}<t} (1 - \frac{d_{ji}}{n_{ji}}).$$

It is interesting to note that $\hat{S}_j(t)$ is exactly the same as the standard K-M estimator that one would obtain if all failures of type other than j were treated as censored cases. If there are no ties between different types of failure, then

$$\hat{S}(t) = \prod_{j=1}^{K} \hat{S}_j(t),$$

so the K-M estimator of the overall survival is the product of the K-M estimators of the cause-specific survivor-like functions.

The Nelson-Aalen estimator of the cause-specific cumulative hazard is

$$\hat{\Lambda}_j(t) = \sum_{i:t_{ji}<t} \frac{d_{ji}}{n_{ji}},$$

and corresponds to an estimate of the cause-specific hazard $\lambda_j(t)$ that takes the value d_{ji}/n_{ji} at t_{ji} and 0 elsewhere. One can also exponentiate the negative of the Nelson-Aalen integrated hazard to obtain an alternative estimator of the cause-specific survivor-like function $S_j(t)$.

A non-parametric maximum likelihood estimator of $F_j(t)$ was proposed by Aalen (1976) and can be thought of as a special case of the Aalen-Johansen theory of estimation for time-inhomogenous Markov processes (Aalen and Johansen 1978). The estimator, known as the Aalen-Johansen estimator is given by

$$F_j(t) = \sum_{i:t_{ji}\leq t} \hat{S}(t_{j-1}) \left(\frac{d_{ji}}{n_{ji}} \right).$$

Our aim is to use this Aalen-Johansen estimator to find an estimate of $F_j(t)$ under the added restriction that $f_j(t)$ is decreasing. For this, we also look at estimation under shape restrictions. Without covariates and right censoring, nonparametric estimation of the monotone baseline density g can be achieved using the Grenander (1956) estimator for a monotone density. The estimate is arrived at by obtaining the least concave majorant of the estimate without any shape restrictions. It can be shown that the convergence rate of the Grenander estimator in the general case is $n^{1/3}$ and its asymptotic distribution is basically a Brownian motion with parabolic drift.

1.4 Overview of the dissertation

In Chapter 2 we look at regression parameter estimation in the AFT model for recurrence time data. The problem however is that these estimators are based on the conditional distribution of the time variable given the covariates. Under length bias sampling, the covariate distribution is functionally dependent on the regression parameter. Thus a "naive" analysis conditioning on the covariates may result in information loss. We show that if the covariate distribution is left completely unspecified then there is no loss of information under a conditional analysis in section 2.5. We also derive a semiparametric asymptotically efficient estimator for the regression parameter in section 2.6 and show its efficacy under simulated data settings (Section 2.7) as well as actual backward recurrence time data (Section 2.8).

Next, we look at k-monotone density estimation in Chapter 3 and prove the fact that if the original density is k-monotone, the corresponding recurrence time density is (k+1)-monotone. So, we develop an algorithm for the estimation of k-monotone densities in the presence of right censoring (Section 3.3). We show the consistency of the estimator and determine its asymptotic distribution in Sections 3.4 and 3.5 respectively.

Finally, we look at recurrence time density estimation in the presence of competing risks in Chapter 4. We develop an algorithm for the estimation of the subdistributions in Section 4.3.2 using the Aalen-Johansen estimator and the Grenander Estimator. Next we look at the consistency and asymptotic properties of the estimator in sections 4.4 and 4.5. Finally, we look at some simulation results and an analysis of the Atherosclerosis Risk in Communities (ARIC) Study data in section 4.7. We finish the report with a summary of the results and look at some future research topics in Section 5.

CHAPTER 2: REGRESSION PARAMETER ESTIMATION IN THE AFT MODEL

In prevalent cohort survival studies where subjects are recruited at a cross-section and followed prospectively in time, the observed event times are length-biased and further follow a multiplicative censoring scheme. For such studies there is an associated initiation time which may be unknown. In this case we only observe the time from sampling to the event of interest. This is the forward recurrence time. Further, in such cases, standard left-truncation survival analysis methods are not applicable.In other scenarios like current duration studies, the time of the initiating event may be known but there is no subsequent follow-up after sampling. Here we observe the backward recurrence times.

In the presence of covariates, the proportional hazards model may not be applicable to forward and backward recurrence time data. However, due to the invariance of the accelerated failure time model under length bias and cross-sectional sampling, it can serve as an useful alternative. In particular, existing estimators for the regression parameter like the ordinary least square and Tsiatis' log rank estimators may be valid. The problem however is that these estimators are based on the conditional distribution of the time variable given the covariates. Under length bias sampling, the covariate distribution is functionally dependent on the regression parameter. Thus a "naive" analysis conditioning on the covariates may result in information loss. We show that if the covariate distribution is left completely unspecified then there is no loss of information under a conditional analysis in section 2.5. We also derive a semiparametric asymptotically efficient estimator for the regression parameter in section 2.6 and show its efficacy under simulated data settings (Section 2.7) as well as actual backward recurrence time data (Section 2.8).

Next, we look at k-monotone density estimation and prove the fact that if the original density is k-monotone, the corresponding recurrence time density is (k+1)-monotone. So, we develop an algorithm for the estimation of k-monotone densities in the presence of right censoring (Section 3.3). We show the consistency of the estimator and determine its asymptotic distribution in Sections 3.4 and 3.5 respectively.

2.1 FRT and BRT

A prevalent cohort consists of subjects who have experienced an initiating event, like disease onset, prior to their entry to the study and who are followed forward in time until another (terminating) event, like death or symptom development. Sampling may be achieved in a small cross-section. In some prevalent cohort study the onset time may be unobservable as in HIV sero-prevalence studies (Brookmeyer and Gail 1987) where the time of infection to HIV is unknown and interest lies in the follow-up time from enrollment to AIDS. Such prevalent cohorts do not provide information on the time between the initiating event and the terminating event T, but only provide partial information in terms of the forward recurrence time T_f, the time from sampling to the terminating event time.

In other scenarios, the time of the initiating event may be known but there may not be any subsequent follow-up after cross-sectional sampling. This is known as the current duration study design, which is encountered, for example, in time to pregnancy surveys. In Keiding et al. (2002), the authors show that the distribution of the times from initiating attempt to cross-sectional sampling for couples that are currently attempting to get pregnant, identifies the distribution of the realized time to pregnancy or unsuccessful end of attempt. Another similar study based on current durations is Yamaguchi's mover-stayer study (Yamaguchi 2003). Such studies provide information on the backward recurrence time T_b.

In both prevalent cohort and current duration study designs, only subjects who have experienced the initiating event prior to sampling, but have not yet experienced the terminating event can be sampled. Thus both the forward and the backward recurrence times are length biased i.e. the sample is biased towards larger values of T. One way to model this bias (Cox (1969), Vardi (1982)) is to sample proportionally to length, i.e., if F_T is the distribution of T then the length-biased version T_{LB} has a distribution given by

$$F_{LB}(t) = \frac{\int_0^t u dF_T(u)}{\mu_T}, \quad t \geq 0,$$

where $\mu_T = \int_0^\infty u dF_T(u)$.

Further, if it can be assumed that the incidence of the disease follows a stationary Poisson process then the cross-sectional sampling time is distributed uniformly between the onset time and the terminating time (Cox (1969), Van Es et al. (2000), Keiding et al. (2002)). Thus $T_f = T_{LB}V$, where V is uniform(0,1). It follows that if $S_T = 1 - F_T$ is the survival function of T then for both T_f and T_b (commonly denoted as $\tilde{T}$ here), the density $g_{\tilde{T}}$ is given by

$$g_{\tilde{T}}(t) = \frac{S_T}{\mu_T}. \tag{2.1}$$

If both T_f and T_b are observed then standard left-truncated techniques apply. For example, the Canadian Health and Aging study of Dementia (Asgharian et al. 2002), where a cohort of elderly subjects were followed from diagnosis until death or end of study. Ages at onset were also ascertained from the subjects' caregivers. There has been extensive work towards estimating the length-biased distribution. Two common approaches found in the literature for left-truncated data, are the conditional and the unconditional likelihood approaches, where the conditioning is applied to the onset times. In the unconditional approach it is assumed that the distribution of the left truncation times are uniformly distributed, which holds under a certain stationarity condition discussed below. In the conditional approach one simply assumes the truncation distribution to be degenerate at the observed truncation times. In the unconditional approach, Vardi (1982) derived the non parametric maximum likelihood estimator (NPMLE) for the length-biased distribution arising out of prevalent cohorts. Asgharian and Wolfson (Asgharian et al. 2002), (Asgharian and Wolfson 2005) derived the NPMLE for the unbiased incident-case survival function obtained from length-biased prevalent cohort data. The conditional perspective has been investigated by Wang et. al. (Wang et al. 1986), (Wang 1991), (Wang et al. 1993). Wang et al. (1986) showed that when the truncating distribution is left completely unspecified, then there is little loss of information when the likelihood is conditioned on the truncation times.

2.2 Cox Proportional Hazards Model

In the presence of covariates, a popular semiparametric model is the proportional hazards (PH) model (Cox 1972) given by

$$\lambda_{T|Z}(t) = e^{\theta' Z}\lambda(t),$$

where, $\lambda_{T|Z}$ is the hazard function of T given the covariate vector Z and λ is an unspecified baseline hazard function. Here the density of T is given by $e^{\theta' z}\lambda(t)e^{-e^{\theta' z}\Lambda(t)}$, where, Λ is the cumulative hazard satisfying $\Lambda(0) = 0$. For the uncensored case, estimates for (θ, λ) can be derived using the partial likelihood based on risk sets (Cox 1972). For right censored data Tsiatis (1981) derives consistency and asymptotic normality of the maximum partial likelihood estimator for θ. Klaassen (1989) showed that the estimator is also semiparametric efficient. Details can also be found in Bickel et al. (1993), Murphy and Van der Vaart (2000) and Van der Vaart (1998).

If we assume a PH model for the core T, then by (2.1), under length-biased and cross-sectional sampling, the conditional density of the forward or the backward recurrence time $\tilde{T}$, given Z, is given by

$$g_{\tilde{T}|Z=z}(t) = \frac{e^{-e^{\theta' z}\Lambda(t)}}{\int e^{-e^{\theta' z}\Lambda(t)}}$$

and the conditional hazard is given by

$$\lambda_{\tilde{T}|Z=z}(t) = \frac{e^{-e^{\theta' z}\Lambda(t)}}{\int_t^\infty e^{-e^{\theta' z}\Lambda(u)} du}$$

Thus going from T to $\tilde{T}$ the proportional hazard structure is lost unless, either the baseline hazard is constant or when T given Z follows a Pareto distribution (Van Es et al. 2000). Note that the former case of an exponential distribution can be obtained as a limiting case of the Pareto distribution. Thus usual techniques of estimating θ using the PH model will not apply in general for forward and backward recurrence times. Furthermore, a naive PH model based analysis on $\tilde{T}$ may produce biased estimates.

2.3 Length-Biased Data

For length-biased data arising out of left-truncation, Wang (1996) derives a consistent and asymptotically normal estimator for θ based on a modification of the risk set in order to adjust for the length-bias. Wang defines indicator variables $\Delta_j(t_i)$ for $t_i \leq t_j$, which equals 1 with probability t_i/t_j and 0 with probability $1 - t_i/t_j$ and defines the modified risk set as

$$R_i^* = \{j : t_i \leq t_j, \Delta_j(t_i) = 1\}.$$

This allows the individuals in the risk set R_i^* to have the population risk set structure, i.e., the conditional probability of the ith subject to die at time t_i given that there is a death from the risk set R_i^* is given by

$$\frac{e^{\theta' z_i}}{\sum_{j \in R_i^*} e^{\theta' z_j}}.$$

Thus estimation of θ can be carried out by maximizing the above (pseudo) partial likelihood.

Under cross-sectional sampling of length-biased forward recurrence times, it is not clear how to modify the risk set in order to adjust for the selection bias. Brookmeyer and Gail (1987) discuss the different directions of the bias that might result from fitting a naive proportional hazards model to this kind of data arising out of prevalent cohort studies, where, the onset time of disease is unobservable.

The major difficulties in carrying out a semiparametric inference for θ under the PH model and using forward or backward recurrence time data are:

1. Since the Cox model specifies the covariate effect on the hazard function and not the time variable itself, the effect of length-biased cross-sectional sampling on the covariate distribution is intractable. We shall see below that for the accelerated failure time model, this is not a problem.

2. Even if a naive analysis conditioned on Z is carried out, the derivation of the efficient score and information seems difficult. Consider the simplest case of uncensored backward recurrence times where we observe $Y = (T_b, Z)$. The log-likelihood for the

naive analysis conditional on the covariates is given by

$$l_{\theta,\Lambda}(t,z) = -e^{\theta' z}\Lambda(t) - log\int e^{-e^{\theta' z}\Lambda(t)}dt.$$

The ordinary score for θ is

$$\dot{l}_{\theta,\Lambda}(t,z) = -ze^{\theta' z}\left\{\Lambda(t) - \frac{\int e^{-e^{\theta' z}\Lambda(t)}\Lambda(t)dt}{\int e^{-e^{\theta' z}\Lambda(t)dt}}\right\}.$$

Consider the parametric path $\eta \mapsto \Lambda_\eta(t) = \int_0^t (1+\eta h)d\Lambda(s)$, where $h \in L_2(\Lambda)$ and $|\eta| \approx 0$. Replacing Λ by Λ_η in the log-likelihood and differentiating at $\eta = 0$ we get the score for the nuisance parameter Λ as:

$$A_{\theta,\Lambda}h = e^{\theta' z}\left\{\int_0^t h(s)d\Lambda(s) + \frac{\int e^{-e^{\theta' z}\Lambda(t)}\int_0^t h(s)d\Lambda(s)dt}{\int e^{-e^{\theta' z}\Lambda(t)}dt}\right\}.$$

Deriving the adjoint operator $A^*_{\theta,\Lambda}$ (Murphy and Van der Vaart 2000) here is not straight forward as in the usual Cox model. As far as we know, a systematic study of forward and backward recurrence times under the proportional hazards assumption for the core time variable has yet to be undertaken.

2.4 Accelerated Failure Time Model

A useful alternative to the PH model to model survival time T in the presence of covariates Z is the Accelerated Failure Time (AFT) model. Under the AFT model

$$T = e^{\theta' Z}U. \tag{2.2}$$

Here T is the failure time measured from the time of some initiating event like birth and disease onset, θ is a $p \times 1$ regression parameter, Z is a $p \times 1$ covariate vector with density h and U a non-negative random variable independent of Z with density g, survival function S and hazard $\lambda(u) = g(u)/S(u)$.

The problem of estimating θ in presence of the nuisance parameters g and h have been taken up by many authors. Two popular approaches are least-squares and rank based procedures. Bickel et al. (1993) derived the semiparametric efficient score and the efficient information for estimating θ when g and h are unspecified and given Z, $\log T - \theta' Z$ is assumed independent of Z and the censoring variable C. Estimation of θ based on the least squares approach has been studied by Miller (1976), James and Buckley (1979) and Koul et al. (1981). Linear rank test based procedures using the partial likelihood score have been developed by Tsiatis (1990), Ritov (1990), Wei et al. (1990), Ying (1993), Fygenson and Ritov (1994) and Jin et al. (2003). For various submodels, these estimators are efficient. For example, the Buckley-James estimator is efficient when the true error density is standard normal while Tsiatis' linear rank test based estimator with unit weights is efficient for a class of extreme value error distributions. The latter is fully efficient if the weight function adaptively estimates λ'/λ, where λ is the hazard function corresponding to the error distribution.

Another important recent development in estimating θ in a tobit model and the one that we will build upon in section 3 for the random right censored case is Cosslett's (Cosslett (2004)) asymptotically efficient estimator via a smoothed self-consistency equation.

Without covariates and right censoring, nonparametric estimation of the monotone baseline density g can be achieved using the Grenander (1956) estimator for a monotone density. For the right censoring case this is generalized by the Denby-Vardi (Denbi and Vardi (1986)) NPMLE. In the uncensored case, Woodroofe and Sun (1993) pointed out a technical difficulty in using the actual likelihood. They show that the Denby-Vardi NPMLE is inconsistent near the origin in the sense that $\hat{g}(0+)$ does converge in probability but the limit is strictly greater than $g(0+)$ almost surely. They propose using a penalized likelihood

$$L_\alpha(g) = \sum_{i=1}^{n} \log g(u_i) - n\alpha g(0+),$$

where g varies over a class of decreasing left-continuous densities and $\alpha > 0$ is a smoothing parameter. Keiding et. al. (2002) discuss that another way to avoid this inconsistency near zero is to estimate the survival function of T conditioned on $T > t_0$ for some small $t_0 > 0$ and then estimate $S(t)/S(t_0)$ by $\hat{g}(t)/\hat{g}(t_0)$.

Under the above assumptions and an application of (2.1) we obtain the joint distribution of $(\tilde{T}, \tilde{Z})$ as

$$f_{\tilde{T},\tilde{Z}}(t,z) = \frac{e^{-\theta' z} S(e^{-\theta' z} t)}{\mu_g} \times \frac{e^{\theta' z} h(z)}{\int e^{\theta' z} h(z) dz}. \tag{2.3}$$

Thus if T follows the AFT model in (2.2) then, $\tilde{T}$ follows a AFT model given by

$$\tilde{T} = e^{\theta' \tilde{Z}} \tilde{U}, \tag{2.4}$$

where $\tilde{Z}$ has a density of the form $h_{\tilde{Z},\theta}(z) = e^{\theta' z} h(z) / \int e^{\theta' z} h(z) dz$ and $\tilde{U}$ has a monotone density given by

$$g_{\tilde{U}}(u) = \frac{S(u)}{\int_0^\infty S(v) dv}.$$

Thus, the resulting AFT model for $\tilde{T}$ has the same covariate effect but a different baseline distribution.

For uncensored backward recurrence times, Klaassen et al. (Klaasen et al. (2004)) derived the semiparametric efficient score and information and proved existence of an efficient estimator for estimating θ in (2.4) when h, the core covariate distribution is either known or known to have zero mean. Their results show that an unconditional (on

$\tilde{Z}$) approach, under these restrictive assumptions, result in a gain in semiparametric information for estimating θ in (2.4). In the next section, we generalize these results to a completely unspecified core covariate distribution and to possibly right censored forward recurrence times. We show that when the core covariate distribution h is completely unspecified, there is no gain in information under an unconditional analysis.

Since the AFT model assumption is preserved under length-biased cross-sectional sampling, a natural question is that whether estimators for θ based on observing $(T \wedge C, Z, I\{T \leq C\})$, where C is a censoring variable independent of T given Z and $\delta = I\{T \leq C\}$, are also valid when based on observing $(\tilde{T} \wedge \tilde{C}, \tilde{Z}, \delta)$. In particular, whether an efficient estimator under (2.2) is also efficient under (2.4). Here $\tilde{C}$ is the censoring variable corresponding to $\tilde{T}$ and $\delta = I\{\tilde{T} \leq \tilde{C}\}$. We assume that $\tilde{T}$ and $\tilde{C}$ are independent given $\tilde{Z}$. The issue here is that length-biased cross-sectional sampling results in the density $\tilde{Z}$ to be functionally dependent on θ and thus might contain information about θ unlike the usual set-up where the covariates are ancillary for the regression parameter θ.

In section 2.5 we derive the semiparametric efficient score and information for estimating θ based on the data $(\tilde{T}_i \wedge \tilde{C}_i, \tilde{Z}_i, \tilde{\delta}_i),\ i = 1, \cdots, n$, while leaving the covariate distribution h completely unspecified. In section 2.6 we derive an asymptotically semiparametric efficient estimator for θ. Results from numerical studies are presented in section 2.7 and results from the data analysis are shown in section 2.8.

2.5 Efficient Score and Estimation

We first take up the calculation of the efficient score and information for the forward recurrence time (T_f), subject to right censoring. Let $\theta \in \Theta$, where Θ is a compact set in $\Re^k$. Let θ_0 be the true value of the regression parameter and suppose that θ_0 belongs in the interior of Θ. For fixed but arbitrary θ, we define our semiparametric model in terms of the distribution of $U(\theta) = e^{-\theta'\tilde{Z}}T_f = e^{-(\theta-\theta_0)'\tilde{Z}}\tilde{U}$ and the corresponding censored variable $U^c(\theta) = e^{-\theta'\tilde{Z}}\tilde{C}$. The conditional distribution of $U(\theta)$ given $\tilde{Z} = z$ is

$$g_{U(\theta)}(u) = \frac{e^{(\theta-\theta_0)'z}S(e^{(\theta-\theta_0)'z}u)}{\int S(v)dv},$$

while the conditional hazard is given by

$$\lambda_{U(\theta)}(u) = \frac{S(e^{(\theta-\theta_0)'z}u)}{\int_u^\infty S(e^{(\theta-\theta_0)'z}w)dw}.$$

Note that given $\tilde{Z}$ the distribution of $U(\theta)$ is monotonic. We now make the following assumptions:

A1: T_f and $\tilde{C}$ are independent given $\tilde{Z}$.

A2: $\mu_g = \int S(v)dv < \infty$.

A3: $E_g U^2 \lambda(U) = \int u^2 g^2 S^{-1}(u)du < \infty$.

Remark 1. *The independent censoring assumption is valid here as $\tilde{C}$ is also measured from the time of sampling. Assumption (A3) is needed to ensure that the density of $U(\theta)$ has finite Fisher information for location.*

Let $\mathcal{G}$ be the class of density functions on $\Re^+$ and $\mathcal{H}$ be a class of density function on $\Re^k$. The semiparametric model for the core AFT model in (2.2) is given by

$$\mathcal{P}^* = \{P_{\theta,g,h} : \ \theta \in \Theta, g \in \mathcal{G}, h \in \mathcal{H}\},$$

where, the distribution $P_{\theta,g,h}$ has a density with respect to an absolutely continuous measure μ given by

$$\frac{dP_{\theta,g,h}}{d\mu}(t) = e^{(\theta-\theta_0)'z}g(e^{(\theta-\theta_0)'z}t)h(z).$$

For the AFT model for forward recurrence times the semiparametric model is

$$\mathcal{P} = \left\{ P_{\theta, S_g, h} : \ \theta \in \Theta, g \in \mathcal{G}, h \in \mathcal{H}' \right\}, \tag{2.5}$$

where $S_g(u) = \int_u^\infty g(v)dv$ for $g \in \mathcal{G}$, and

$$\mathcal{H}' = \left\{ h : h \in \mathcal{H}, \int e^{\theta' z} h(z) d(z) < \infty, \int z^2 e^{\theta' z} h(z) dz < \infty, \theta \in \Theta \right\}.$$

We assume that Θ is a compact subset of $\mathcal{R}^k$. Further, $P_{\theta, S_g, h}$ is dominated by an absolutely continuous measure μ with density

$$\frac{d}{d\mu} P_{\theta, S_g, h} = \frac{e^{(\theta-\theta_0)' z} S_g(e^{(\theta-\theta_0)' z} u)}{\int S_g(v) dv} \times \frac{e^{\theta' z} h(z)}{\int e^{\theta' z} h(z) dz}.$$

Define $\mathcal{S} = \{S_g : g \in \mathcal{G}\}$. Let the true distribution be $P_0 = P_{\theta_0, S_0, h_0}$ with $S_0 = S_{g_0}$. Define the submodels

$$\begin{aligned} \mathcal{P}_\theta &= \{P_{\theta, S_0, h_0} : \theta \in \Theta\}, \\ \mathcal{P}_S &= \{P_{\theta_0, S, h_0} : S \in \mathcal{S}\} \text{ and} \\ \mathcal{P}_h &= \{P_{\theta_0, S_0, h} : h \in \mathcal{H}'\}. \end{aligned}$$

Let $\dot{\mathcal{P}}_\theta$, $\dot{\mathcal{P}}_S$ and $\dot{\mathcal{P}}_h$ be the respective tangent spaces for $\mathcal{P}_\theta, \mathcal{P}_S$ and $\mathcal{P}_h$ at $P_0 = P_{\theta_0, S_0, h_0}$. Let $\dot{l}_\theta$ be the ordinary score for θ when S and h are fixed. Then the efficient score function $\tilde{l}_\theta \in (L_2^0(P_0))^k$ for θ in the full model $\mathcal{P}$ at P_0 is $\tilde{l}_\theta = \dot{l}_\theta - \Pi_0(\dot{l}_\theta | \dot{\mathcal{P}}_S + \dot{\mathcal{P}}_h)$, where $\Pi_0(l|\mathcal{Q})$ denotes the orthogonal projection of l onto the linear span of $\mathcal{Q}$ (Bickel et al. (1993)).

The next lemma helps identify $\dot{\mathcal{P}}_S$, i.e., the tangent space for the nuisance parameter corresponding to the decreasing density function of the forward recurrence times, as a dense set in the maximal tangent space $L_2^0(S)$.

Lemma 2. *Consider the semiparamatric model $\mathcal{P} = \{P_g : g \in \mathcal{G}\}$, where the distribution P_g has density $p_g(u) = S_g / \int S_g$ and $\mathcal{G}$ is a collection of densities on $\mathcal{R}^+$. Let $\dot{\mathcal{G}}_g$ and $\dot{\mathcal{P}}_g$ be the tangent sets for the models $\mathcal{G}$ and $\mathcal{P}$ respectively at g. If A_g is the score operator mapping tangents in $\dot{\mathcal{G}}_g$ to $\dot{\mathcal{P}}_g$ then, $A_g\dot{\mathcal{G}}_g$ is dense in the maximal tangent set $L_2^0(S)$ for $\mathcal{P}$.*

The proof of this lemma is given in Appendix A.

Theorem 3. *Suppose that the covariate vector $\tilde{Z}$ is almost surely bounded. Then under (A1)–(A3) and with $\phi(u) = 1 - ug(u)/S(u)$ and*

$$M(t) = I\{U(\theta) \leq t\} - \int_0^t I\{U(\theta) > s\}\lambda_{U(\theta)}(s)ds, \tag{2.6}$$

the ordinary score for θ at $\theta = \theta_0$ is

$$\dot{l}_{\theta_0} = z\int_0^{U^c(\theta_0)} R\phi(s)dM(s) - (z - E\tilde{Z}), \tag{2.7}$$

the tangent space $\dot{\mathcal{P}}_S$ for S is $\{\dot{l}_S b : b \in L_2^0(S)\}$ where the score operator $\dot{l}_S$ for S is given by

$$\dot{l}_S b = \int_0^{U^c(\theta_0)} Rb(s)dM(s), \tag{2.8}$$

the tangent space for h is $\{b : b \in L_2(h), \int b(z)e^{\theta_0' z}h(z)dz = 0\}$, and the efficient score for θ at $\theta = \theta_0$ is

$$\tilde{l}_{\theta,S} = \int_0^{U^c(\theta_0)} (z - E\{\tilde{Z}|U^c(\theta_0) \geq s\})R\phi(s)dM(s), \tag{2.9}$$

where for $a \in L_2^0(S)$,

$$Ra(t) = a(t) - \frac{\int_t^\infty a(u)S(u)du}{\int_t^\infty S(u)du}.$$

The proof of this theorem is attached in Appendix A.

Remark 4. *The efficient score is free of h. Thus for estimating θ efficiently we do not need to estimate the covariate distribution. Thus we do not need a separate identifiability condition for h like the mean-zero condition assumed in Klaasen et al. (2004) and can be left completely unspecified.*

Remark 5. *The efficient information is given by*

$$\tilde{I}_{\theta_0} = E \int_0^{U^c(\theta_0)} D(\tilde{Z}, C, \theta_0, s) D(\tilde{Z}, C, \theta_0, s)' (R\phi)^2(s) dF_{U(\theta_0)}(s), \tag{2.10}$$

where $D(\tilde{Z}, C, \theta_0, s) = (\tilde{Z} - E\{\tilde{Z}|\ U^c(\theta_0) \geq s\})$ and a' denotes the transpose of the vector a.

The efficient score and the information here are similar to the ones in Bickel et al. (1993) for the censored regression problem based on (T, C, Z) except that in the latter case $\phi(u) = 1 + ug'(u)/g(u)$ while in our case $\phi = 1 - ug(u)/S(u)$. The main similarity is that in both situations, the efficient scores do not use information in the marginal distribution of the covariates. The reason behind this is the fact that $\dot{\mathcal{P}}_S \perp \dot{\mathcal{P}}_h$. Thus we can assume that the covariates distribution is degenerate at the observed values and carry out a conditional analysis. Moreover, a conditional analysis can also be carried out for the FRT case without loss of information when the core covariate distribution is completely unspecified. In this case, an efficient estimator based on the core incident cases will also be efficient for the FRT case. However, the efficiency bounds may be different. In general there is a loss of information going from the core to the FRT model. The loss is minimal when the core density is also non-increasing. This is true for example in the exponential case. For the uncensored Weibull density the relative information for the BRT case to the core case is $V(\tilde{Z})/2V(Z)$. Here $V(Z)$ and $V(\tilde{Z})$ are the variances of the core and the observed covariates respectively. In the next chapter we construct an asymptotically efficient estimator for θ which could be applied to both these problems.

Remark 6. *In terms of the residuals* $\epsilon(\theta_0) = \log U(\theta_0) = \log T_f - \theta_0' Z$ *with density* f*, distribution* F *and hazard function* $\lambda_f = f/(1-F)$*, the efficient score at* $\theta = \theta_0$ *is*

$$\tilde{l}_{\theta_0,f} = \int_{-\infty}^{\epsilon^c(\theta_0)} (Z - E\{Z|\epsilon^c(\theta_0) \geq s\}) R\phi(s) dM(s), \tag{2.11}$$

where $\phi = f'/f$*,* $M(t) = I\{\epsilon(\theta_0) \leq t\} - \int_{-\infty}^{t} I\{\epsilon(\theta_0) > s\}\lambda_f(s)ds$ *and* $R\phi(t) = \phi(t) - E\{\phi(\epsilon(\theta_0))|\epsilon(\theta_0) > t\}$*. Note that* $f(\epsilon) = e^{\epsilon} g(e^{\epsilon})$ *when observing* T*, while* $f(\epsilon) = e^{\epsilon} S(e^{\epsilon}) / \int S_g(v)dv$ *when observing* T_f*.*

We obtain the efficient score for the backward recurrence time case as a cor:

Corollary 7. *Let* $U(\theta) = e^{-\theta' z} T_b$ *and suppose that the covariate vector* $\tilde{Z}$ *is almost surely bounded and* $f_{U(\theta)}$ *has finite Fisher information for location. Then with* $\lambda(u) = g(u)/S(u)$*, the efficient score for estimating* θ *at* $\theta = \theta_0$ *is*

$$\tilde{l}_{\theta_0,\lambda} = (\tilde{Z} - E\tilde{Z})[1 - U(\theta_0)\lambda(U(\theta_0))] \tag{2.12}$$

and the efficient information is given by

$$E[\tilde{l}_{\theta_0} \tilde{l}_{\theta_0}'] = E(\tilde{Z} - E\tilde{Z})(\tilde{Z} - E\tilde{Z})' E[1 - U(\theta_0)\lambda(U(\theta_0))]^2. \tag{2.13}$$

Proof. The backward recurrence times are uncensored. Thus we take $U^c(\theta) = \infty$, $M(t) = I\{U(\theta) \leq t\}$ and $R\phi = \phi$ in (10) and (12) to get the desired results. $\square$

Remark 8. *For the case when* h *is assumed known, Klaassen et. al. (Klaasen et al. (2004)) derive the semiparametric efficient score and information. The efficient score* *is given by* $-(z - E\tilde{Z})u\lambda(u)$, *while the efficient information is* $Var(\tilde{Z})E[U\lambda(U)]^2$. *Klaassen et. al. also derive the efficient information under the assumption that the core covariate distribution has mean zero. In these cases there is a gain in information. Such gains are possible by projecting the ordinary score to restricted nuisance tangent spaces.*

2.6 Asymptotically Efficient Estimator

In this section we derive a semiparametric asymptotically efficient estimator of the regression parameter θ in the AFT model based on Severini and Wong's (Severini and Wong (1992)) profile likelihood approach used to estimate the euclidean parameter in presence of a nuisance parameter. Here one identifies a parametric submodel belonging to the nuisance space that passes through the true parameter point and is least-favorable in the sense of having the least Fisher's information among all parametric submodels (Severini and Wong (1992), Stein (1956)). The idea of estimating the euclidean parameter is based on estimating a least-favorable curve and then maximizing the corresponding likelihood to obtain M-estimators. For the tobit model, Cosslett(Cosslett (2004)) derives an estimator for the regression parameter based on a smoothed self consistent estimate for the distribution of the errors. We adapt this estimator for the randomly right censored linear regression model. The smoothed self-consistent equation for the NPMLE of the survival function of the errors leads to a least favorable submodel for the hazard function. The corresponding estimated log-likelihood is maximized to get a semiparametric efficient estimator.

For arbitrarily fixed θ define the uncensored residuals by $\epsilon_\theta = \log T - \theta' Z$ and the censored residuals by $\epsilon_\theta^c = \log C - \theta' Z$. Also define $e \equiv e_\theta = \epsilon_\theta \wedge \epsilon_\theta^c$, where $x \wedge y$ denotes the minimum of x and y. The data consists of n independent realizations of

$Y = (e, \delta, Z)$. The model is

$$\mathcal{P} = \{P_{\theta,\lambda} : \theta \in \Theta, \lambda \in \mathit{\Lambda}\}, \tag{2.14}$$

where $\Theta \subset \Re$ and $\mathit{\Lambda} = \{\lambda : \mathcal{X} \mapsto \Re\}$ with further specification made below. Under $P_{\theta,\lambda}$, ϵ_θ given $Z = z$ has hazard function λ. Let (θ_0, λ_0) be the true value of the parameter (θ, λ). Let f_0 and S_0 be the density and the survival functions corresponding to λ_0. Under P_{θ,λ_0}, ϵ_θ has density $f_0(\,\cdot + (\theta - \theta_0)'z)$ and survival function $S_0(\,\cdot + (\theta - \theta_0)'z)$. The efficient score and information for estimating θ is given by Proposition 4.6.1 in BKRW (Bickel et al. (1993)). For the forward and the backward recurrence time considered in section 2, the efficient score and information is given in (2.11). Let $\zeta_{\theta,z}$ be the survival function of ϵ_θ^c given $Z = z$. We make the following additional assumptions:

(C1) The covariate vector Z is bounded almost surely with density h.

(C2) θ_0 belongs to the interior of an open and bounded set $\Theta \subset \Re^k$. Along with (C1) this gives $\alpha \equiv \operatorname{ess.sup}_{\theta \in \Theta} |(\theta - \theta_0)'Z| < \infty$.

(C3) $\tau = \sup_t\{t : Pr[C > \exp\{t + \alpha\} \mid Z] > 0\}$ exists and is finite and further $S_0(\tau + \alpha) > 0$.

Remark 9. *While conditions (C1) and (C2) are standard, condition (C3) is related to the standard end of study assumption made on the distribution of T and C in right-censored data settings. The analysis is thus restricted to an interval $e \in (-\infty, \tau]$.*

Let $\gamma_{\theta,z}(t) \equiv \exp\{-\int_{-\infty}^t \lambda_0(s + (\theta - \theta_0)'z)ds\}\zeta_{\theta,z}(t)$ denote the at-risk probability function given Z and under P_{θ,λ_0}. Note that under assumption (C1)–(C3), $\gamma_{\theta,z}(t) > 0$ for all $t \leq \tau$ and $\theta \in \Theta$.

A standard approach for constructing asymptotically efficient estimators for euclidean parameters in presence of a nuisance parameters is the profile likelihood approach of Severini and Wong (Severini and Wong (1992)) where the nuisance parameter is replaced by a suitable consistent estimate (like the NPMLE) in the efficient estimating equation for the euclidean parameter.

The likelihood $p(y;\theta,\lambda)$ under $P_{\theta,\lambda}$ is given by

$$p(y;\theta,\lambda) = \{\lambda(e_\theta)\}^\delta \exp\{-\int_{-\infty}^{e_\theta} \lambda(u)du\}.$$

Define the Kulback-Lieber distance as,

$$\kappa(\theta,\lambda) = -\int \log p(y;\theta,\lambda)p(y;\theta_0,\lambda_0)dy.$$

After an application of Fubini's theorem and integration by parts on the integral term in $p(y;\theta,\lambda)$ we get

$$\kappa(\theta,\lambda) = -\int\int \left\{\lambda_0\Big(t+(\theta-\theta_0'z)\Big)\log\lambda(t)-\lambda(t)\right\}\gamma_{\theta_0,z}(t)h(z)dzdt.$$

For Severini and Wong's profile likelihood approach to work one needs to identify a least-favorable parametric submodel which is essentially a smooth curve $\theta \mapsto \lambda_\theta$ through (θ_0,λ_0), with $\lambda_{\theta_0} = \lambda_0$ such that for any other submodel $\lambda_{1\theta}$ also satisfying $\lambda_{1\theta_0} = \lambda_0$, we have,

$$-E_0\frac{d^2}{d\theta^2}\log p(Y_1;\theta,\lambda_\theta)\Big|_{\theta=\theta_0} \leq -E_0\frac{d^2}{d\theta^2}\log p(Y_1;\theta,\lambda_{1\theta})\Big|_{\theta=\theta_0},$$

where E_0 is the expectation under P_{θ_0,λ_0}. Under model identifiability, it is sufficient to show that for arbitrarily fixed θ, $\kappa(\theta,\lambda_\theta) < \kappa(\theta,\lambda_{1\theta})$ for any $\lambda_{1\theta} \neq \lambda_\theta$ satisfying $\lambda_{1\theta_0} = \lambda_0$.

Let $L_n(\theta, \lambda_\theta) \equiv \sum \log p(y_i; \theta, \lambda_\theta)$ be the log-likelihood obtained by replacing the nuisance parameter λ by the least-favorable submodel λ_θ. Let $\hat{\lambda}_\theta$ be an estimator of λ_θ which converges to λ_θ at a rate faster than $n^{-1/4}$. Then under regularity conditions the maximizer $\hat{\theta}_n$ of $L_n(\theta, \hat{\lambda}_\theta)$ is consistent and asymptotic normal with the asymptotic variance equal to the efficient information (Severini and Wong (1992)).

Cosslett (Cosslett (2004)) derives an efficient estimator for θ in the tobit model by solving the score equation based on the smoothed self-consistent estimator for the error density. Below we derive the smoothed self-consistent estimator $\hat{\lambda}_\theta$ for the right-censored regression problem and show that it converges to a least-favorable submodel denoted by λ_θ at a rate $n^{-\nu_1}$, where $\nu_1 > 1/4$. For consistency and asymptotic normality we also need that $\hat{\lambda}'_\theta$ goes to λ'_θ at a rate $n^{-\nu_2}$, where $\nu_1 + \nu_2 \geq 1/2$. Here $\lambda'_\theta \equiv d\lambda_\theta/d\theta$ and $\hat{\lambda}'_\theta = d\hat{\lambda}_\theta/d\theta$ are the total derivatives taken with respect to θ.

Let $\mathbb{G}_{n,\theta}(t) \equiv n^{-1} \sum I\{e_j(\theta) \leq t\}$ and $\mathbb{F}_{n,\theta}(t) \equiv n^{-1} \sum \delta_j I\{e_j(\theta) \leq t\}$ denote the empirical distribution functions of the observed residuals and the observed uncensored residuals respectively. Efron's (Efron (1967)) self consistent equation for an estimator $\bar{S}$ for the survival function of the uncensored residuals for right censored is given by

$$\bar{S}(t) = \int_t^\infty d\mathbb{G}_{n,\theta}(u) + \bar{S}(t) \int_{-\infty}^t \frac{1}{\bar{S}(v)} d(\mathbb{G}_{n,\theta} - \mathbb{F}_{n,\theta})(u).$$

It is well known that the Kaplan-Meier survival function satisfies the above equation.

In general, smoothing of a function $f(v_i)$ of an observation v_i can be achieved by replacing it with $\tilde{f}(v_i) = h_n^{-1} \int f(u) K\left(h_n^{-1}(u - v_i)\right) du$, with a suitable kernel K and bandwidth h_n. Thus the self consistency equation after smoothing becomes

$$\begin{aligned}\tilde{S}(t) \quad &= \quad \int \bar{K}\left(\frac{t-v}{h_n}\right) d\mathbb{G}_{n,\theta}(v) + \\ &\quad \tilde{S}(t) \int_{-\infty}^t \frac{\int \frac{1}{h_n} K\left(\frac{u-v}{h_n}\right) d(\mathbb{G}_{n,\theta} - \mathbb{F}_{n,\theta})(v)}{\tilde{S}(u)} du,\end{aligned}$$

where $\bar{K}(u) = \int_u^\infty K(v)dv$ and $\tilde{S}$ is the smoothed version of $\bar{S}$. The above integral equation is linear in $1/\tilde{S}$ with an explicit solution

$$\hat{S} \equiv \hat{S}_\theta(e) = \exp\left\{-\int_{-\infty}^{e} \frac{\hat{g}_{n,\theta}(v)}{\hat{G}_{n,\theta}(v)}dv\right\}, \tag{2.15}$$

where

$$\hat{g}_{n,\theta}(t) = h_n^{-1} \int K\left(h_n^{-1}(t-v)\right) d\mathbb{F}_{n,\theta}(v)$$

and

$$\hat{G}_{n,\theta}(t) = \int \bar{K}\left(h_n^{-1}(t-v)\right) d\mathbb{G}_{n,\theta}(v).$$

The integral in (2.15) always exists since by construction the right-hand tail of $\hat{S}$ decreases at the same rate as the tail of the kernel function. The density estimator corresponding to $\hat{S}$ in (2.15) is

$$\hat{f}_\theta(t) = \frac{\hat{g}_{n,\theta}(t)}{\hat{G}_{n,\theta}(t)} \exp\left\{-\int_{-\infty}^{t} \frac{\hat{g}_{n,\theta}(v)}{\hat{G}_{n,\theta}(v)}dv\right\}. \tag{2.16}$$

The next set of lemmas establish the convergence of $\hat{g}_n$ and $\hat{G}_n$. Define,

$$g_\theta(t) = \int f_0\left(t + (\theta - \theta_0)'z\right) \zeta_{\theta,z}(t) h(z)dz, \tag{2.17}$$

$$G_\theta(t) = \int S_0\left(t + (\theta - \theta_0)'z\right) \zeta_{\theta,z}(t) h(z)dz \tag{2.18}$$

and

$$\lambda_\theta(t) = \frac{g_\theta(t)}{G_\theta(t)} = \frac{\int \lambda_0(t + (\theta-\theta_0)'z)\gamma_{\theta,z}(t)h(z)dz}{\int \gamma_{\theta,z}(t)h(z)dz}. \tag{2.19}$$

While $g_\theta(\cdot)$ is the conditional density of the observed residuals e_θ given $\delta = 1$, times $P\{\delta = 1\}$, $G_\theta(\cdot)$ is the unconditional survival function of e_θ. Similar "estimable" functions were considered towards establishing the asymptotic properties of the Buckley and James estimator in Ritov (1990).

Consider the submodel $\theta \mapsto \lambda_\theta$ given by (2.19). Note that $\lambda_{\theta_0} = \lambda_0$. Consider the parametric submodel $\lambda_\nu = \lambda(1 + \nu a)$, where $a \in L_2(f_0)$. Then,

$$\frac{d\kappa(\theta, \lambda_\nu)}{d\nu}\Big|_{\nu=0} = -\int\int a(t)\left\{\lambda_0(t + (\theta - \theta_0)'z) - \lambda(t)\right\}\gamma_{\theta_0,z}(t)h(z)dzdt.$$

Thus, for arbitrarily fixed $\theta \in \Theta$, $\kappa(\theta, \lambda)$ is minimized at $\lambda = \lambda_\theta$. Thus λ_θ is least-favorable. Also note that

$$\begin{aligned}
&\kappa(\theta, \lambda_\theta) - \kappa(\theta_0, \lambda_0) \\
&= -\int\int \left\{\log\frac{\lambda_0(t + (\theta - \theta_0)'z)}{\lambda_0(t)} + 1 - \frac{\lambda_0(t + (\theta - \theta_0)'z)}{\lambda_0(t)}\right\} \\
&\qquad\quad \times \lambda_0(t)\gamma_{\theta_0,z}(t)h(z)dzdt \\
&\geq 0,
\end{aligned}$$

where the equality holds if and only if $\theta = \theta_0$. Thus if we define $m(\theta) \equiv E_{\theta_0,\lambda_0}\log(p_{\theta,\lambda_\theta}/p_{\theta_0,\lambda_0})$ then $m(\theta)$ is maximized at $\theta = \theta_0$. Let $\hat{\lambda}_\theta \equiv \hat{g}_{n,\theta}/\hat{G}_{n,\theta}$. For each n, the log-likelihood is given by

$$L_n(\theta, \lambda_\theta) = \sum_{i=1}^{n}\left\{\delta_i \log\lambda_\theta(e_i) - \int_{-\infty}^{e_i}\lambda_\theta(u)du\right\}.$$

Define $\hat{\theta} \equiv \hat{\theta}_n$ to be an element of Θ satisfying

$$L_n(\hat{\theta}, \hat{\lambda}_{\hat{\theta}}) = \sup_{\theta\in\Theta} L_n(\theta, \hat{\lambda}_\theta). \tag{2.20}$$

In order to prove consistency, asymptotic normality and efficiency of $\hat{\theta}$ we need uniform convergence of $\hat{\lambda}_\theta$ to λ_θ and $\hat{\lambda}'_\theta$ to λ'_θ at appropriate rates (Severini and Wong (1992)). For the purpose of M-estimation, we only to need to establish this uniform convergence on $(-\infty, \tau]$. The following regularity conditions and conditions on the kernel are needed to establish the asymptotics:

(R1) f_0, f_0' and f_0'' are bounded.

(R2) The functions g_θ, g_θ', g_θ'', G_θ, G_θ', G_θ'' are continuous and uniformly bounded in $\theta \in \Theta$. Here, $g_\theta' \equiv dg_\theta/d\theta$ and the other derivatives are defined similarly.

(R3) $E_{\theta_0,\lambda_0}|\epsilon_\theta|^p < \infty$, for some $p > 4$.

(R4) λ_θ and $\log \lambda_\theta$ have two continuous derivatives. Further, these functions and their derivatives are bounded by integrable functions for all $\theta \in \Theta$.

(K1) K is a bounded, differentiable and symmetric function satisfying $\int K = 1$, $\int u^2 K(u)du < \infty$, $\int [K'(u)]^2 du < \infty$ and $\int [K''(u)]^2 du < \infty$.

(K2) The bandwidth h_n satisfies $h_n = n^{-\beta}$, where $1/8 \leq \beta < 1/5$.

Remark 10. *While conditions (R1)–(R3) are standard, condition (R4) cannot be expressed in a more straightforward way in terms of the underlying true error hazard. It is worthwhile to note that (R4) holds in case the error density is normal and with bounded covariates.*

Lemma 11. *Under assumptions (C1) and (C2), the classes* $\mathcal{F} = \{\delta I[e_\theta \leq t] : \theta \in \Theta,\ t \in \mathbb{R}\}$, $\mathcal{G} = \{ZI[e_\theta \leq t] : \theta \in \Theta,\ t \in \mathbb{R}\}$ *and* $\mathcal{H} = \{\delta Z I[e_\theta \leq t] : \theta \in \Theta,\ t \in \mathbb{R}\}$ *are* P*-Donsker.*

Proof. $\{I[e_\theta \leq t] : \theta \in \Theta,\ t \in \mathbb{R}\}$ is P-Donsker since a finite-dimensional vector space of measurable functions is P-Donsker. Let $G(y)$ be a uniformly bounded function then by the preservation thm $\{G(y)I[e_\theta \leq t]\}$ is P-Donsker. □

Lemma 12. *Under (C1), (C2), (R1)–(R4) and (K1) we have,*

(a) $|\hat{g}_{n,\theta}(t) - g_\theta(t)| = O_p(n^{-1/2})O_p(h_n^{-1}) + O_p(h_n^2)$,

(b) $|\hat{G}_{n,\theta}(t) - G_\theta(t)| = O_p(h_n^2)$,

(c) $|\hat{g}'_{n,\theta}(t) - g'_\theta(t)| = O_p(n^{-1/2})O_p(h_n^{-2})$ *and*

(d) $|\hat{G}'_{n,\theta}(t) - G'_\theta(t)| = O_p(h_n^2)$

uniformly in t *and* $\theta \in \Theta$,

The proof of this lemma is given in Appendix A.

Lemma 13. *Under (C3) and conditions (R1)–(R4) and (K1),*

(a) $\sup_t |\hat{\lambda}_{n,\theta}(t) - \lambda_\theta(t)| = O_p(n^{-1/2})O_p(h_n^{-1}) + O_p(h_n^2)$

(b) $\sup_t |\hat{\lambda}'_{n,\theta}(t) - \lambda'_\theta(t)| = O_p(n^{-1/2})O_p(h_n^{-2})$

uniformly in $t \in I$ *and* $\theta \in \Theta$.

Proof. Note that under (C3), (R2) and (K1),

$$\begin{aligned}
&\sup_{t\in I} |\hat{\lambda}_\theta(t) - \lambda_\theta(t)| \\
&\le \sup_{t\in I}\left(\left|\frac{\hat{g}_{n,\theta}(t) - g_\theta(t)}{G_\theta(t)}\right| + |\hat{g}_{n,\theta}(t)|\left|\frac{\hat{G}_{n,\theta}(t) - G_\theta(t)}{\hat{G}_{n,\theta}(t)G_\theta(t)}\right|\right) \\
&\le G_\theta(\tau)^{-1}\sup_t |\hat{g}_{n,\theta}(t) - g_\theta(t)| \\
&\quad + \left(\sup_t |\hat{g}_{n,\theta}(t) - g_\theta(t)| + \sup_t |g_\theta(t)|\right) G(\tau)^{-1} \\
&\qquad \cdot \left(G(\tau) - \sup_t |\hat{G}_{n,\theta}(t) - G_\theta(t)|\right)^{-1} \sup_t |\hat{G}_{n,\theta}(t) - G_\theta(t)|.
\end{aligned}$$

By lemma 3, the dominating term in the preceding line is $O_p(|\hat{g}_\theta - g_\theta|)$. Thus part (a) follows. A similar inequality for $|\hat{\lambda}'_\theta - \lambda'_\theta|$ shows that the dominating term is $O_p(|\hat{g}'_\theta - g'_\theta|)$ which proves part (b). $\square$

Remark 14. *Note that along with (K2), lem 4 gives the convergence rates of order $n^{-\nu_1}$ with $\nu_1 \geq 1/4$, for $\hat{\lambda}_\theta \to \lambda_\theta$ and of order $n^{-\nu_2}$ with $\nu_1 + \nu_2 \geq 1/2$, for $\hat{\lambda}'_\theta \to \lambda'_\theta$, uniformly in $\theta \in \Theta$ and $t < \tau$.*

Define $L_n(\theta, \hat{\lambda}_\theta) \equiv \sum\{\delta \log \hat{\lambda}_\theta(e_\theta) - \int_{-\infty}^{e_\theta} \hat{\lambda}_\theta(u)du\}$ then $\hat{\theta}_n$ be the maximizer of $\theta \mapsto L_n(\theta, \hat{\lambda}_\theta)$. In order to show that $\hat{\theta}_n$ is consistent for θ, we need to show that $n^{-1}L_n(\theta, \hat{\lambda}_\theta) - n^{-1}L_n(\theta, \lambda_\theta)$ converges to zero uniformly in θ. For isolated terms in the left tail like an isolated $e_{(1)}$, $\hat{\lambda}_\theta(e_{(1)})$ is of order $O(1/nh_n)$, thus we need to use a trimming factor to avoid any contribution of such terms to $L_n(\theta, \hat{\lambda}_\theta)$ or to $L_n(\theta, \lambda_\theta)$.

Consider the trimming function $\tau(\cdot)$, given by

$$\tau(u) = \begin{cases} 1 & \text{for } u \geq 1 \\ \psi(u) & \text{for } 0 < u < 1 \\ 0 & \text{for } d \leq 0, \end{cases}$$

where, in order to make τ twice differentiable with continuous derivatives we choose a smooth bridge function $\psi(\cdot)$ which is twice differentiable with $\psi(0) = \psi'(0) = \psi'(1) = \psi''(0) = \psi''(1) = 0$ and $\psi(1) = 1$. Also we let $b_n \downarrow 0$ at a suitable rate and define $\tau_1(\lambda) \equiv \tau(b_n^{-1}(\lambda - b_n))$ and

$$L_n^*(\theta, \hat{\lambda}_\theta) \equiv n^{-1} \sum \{\delta \log \hat{\lambda}_\theta(e_\theta)\tau_1(\hat{\lambda}_\theta(e_\theta)) - \int_{-\infty}^{e_\theta} \hat{\lambda}_\theta(u)du\}.$$

Similarly define $L_n^*(\theta, \lambda_\theta)$.

Lemma 15. *Under assumptions (C1)–(C4) and conditions (R1)–(R4) and (K1) and (K2), if the trimming rate $b_n = n^{-\alpha}$ with $0 < \alpha < 1/4$, then, $\left| n^{-1}L_n^*(\theta, \hat{\lambda}_\theta) - n^{-1}L_n(\theta, \lambda_\theta) \right| \xrightarrow{\mathrm{P}} 0$, uniformly in $\theta \in \Theta$.*

Proof.

$$
\begin{aligned}
& L_n^*(\theta, \hat{\lambda}_\theta) - L_n^*(\theta, \lambda_\theta) \\
& = n^{-1} \sum \delta \left[\tau_1(\lambda_\theta(e_\theta)) \log \frac{\hat{\lambda}_\theta(e_\theta)}{\lambda_\theta(e_\theta)} + \log \hat{\lambda}_\theta(e_\theta) \{ \hat{\lambda}_\theta(e_\theta) - \lambda_\theta(e_\theta) \} \tau_1'(\hat{\gamma}) \right] \\
& \quad - n^{-1} \sum \int_{-\infty}^{e_\theta} \{ \hat{\lambda}_\theta(u) - \lambda_\theta(u) \} du.
\end{aligned}
$$

Now consider $\tau(\lambda_\theta) \log(\hat{\lambda}_\theta / \lambda_\theta) = \tau(\lambda_\theta) \log[1 + \lambda_\theta^{-1}(\hat{\lambda}_\theta - \lambda_\theta)]$. This is zero when $\lambda_\theta \leq b_n$. Thus the worst possible case is when $\lambda_\theta = b_n$. Now $\tau(b_n) \log[1 + b_n^{-1} O_p(n^{-1/4})] \to 0$ as $n \to \infty$ if $b_n = n^{-\alpha}$ with $0 < \alpha < 1/4$.

The second term $\tau'(\gamma)(\hat{\lambda}_\theta - \lambda_\theta) \log \hat{\lambda}_\theta = 0$ if $\gamma < b_n$ or $\gamma > 2b_n$. In the worst possible case we have $\gamma = \hat{\lambda}_\theta + \delta_n = b_n$ where $\delta_n = n^{-\beta}$ with $\beta \geq 1/4$. Then $\log \hat{\lambda}_\theta = \log b_n + \log(1 - \delta_n / b_n)$. Thus $\tau'(\gamma)(\hat{\lambda}_\theta - \lambda_\theta) \log \hat{\lambda}_\theta = \tau'(b_n) O_p(n^{-1/4})[\log b_n + \log(1 - \delta_n / b_n)] \to 0$ as $n \to \infty$ since $\delta_n / b_n \to 0$.

A similar Taylor's expansion inside the integral term yields,

$$
\int_{-\infty}^{e_\theta} \tau(\lambda_\theta)(\hat{\lambda}_\theta - \lambda_\theta)(u) du \leq O_p(n^{-1/4})[n^{-1} \sum e_\theta - m_n],
$$

where $m_n = sup\{t : \lambda_\theta(t) \leq b_n\}$. Note that for exponential and sub-exponential tails, $m_n \downarrow -\infty$ at a rate slower than b_n. Also, since, $E|e_\theta| < \infty$, the integral term goes to zero. Thus $L_n^*(\theta, \hat{\lambda}_\theta) - L_n^*(\theta, \lambda_\theta) \to 0$. Also, since $|L_n^*(\theta, \lambda_\theta)| \leq |L_n(\theta, \lambda_\theta)|$, under regularity condition (R4), the bounded convergence thm gives

$$
L_n^*(\theta, \lambda_\theta) - L_n(\theta, \lambda_\theta) \to E[L_n^*(\theta, \lambda_\theta) - L_n(\theta, \lambda_\theta)].
$$

Also, $L_n^*(\theta, \lambda_\theta)$ converges to $L_n(\theta, \lambda_\theta)$ point-wise as $n \to \infty$. Thus another use of bounded convergence gives $E[L_n^*(\theta, \lambda_\theta) - L_n(\theta, \lambda_\theta)] \to 0$. $\square$

Theorem 16. *If $\hat{\theta}_n$ is the maximizer of the log-likelihood as in (2.20) then under (C1)–(C4), (R1)–(R4) and (K1) and (K2), $\hat{\theta}_n \xrightarrow{\mathrm{P}} \theta_0$ as $n \to \infty$.*

Proof. Let $m(\theta) \equiv E_0 L_n(\theta, \lambda_\theta)$. In section 3.3, we saw that $m(\theta)$ is maximized at θ_0. Also by condition (R4) and assumption (A3) we have that $L_n(\theta, \lambda_\theta)$ is a sum of n bounded terms and thus by the weak law of large numbers $n^{-1}L_n(\theta, \lambda) \to m(\theta)$ for every $\theta \in \Theta$. Under condition (R3) and (R4) and the compactness of Θ and an application of the argmax thm (cor 3.2.2 in van der Vaart and Wellner (1996)), we have $\hat{\theta}_n \xrightarrow{\mathrm{P}} \theta_0$. $\square$

For asymptotic normality and efficiency, we use the profile likelihood theory in Murphy and Van der Vaart (Murphy and Van der Vaart (2000)). thm 1 in Murphy and Van der Vaart (2000) provide sufficient conditions under which the log profile likelihood $\log pl_n(\theta)$ admits the following asymptotic expansion:

$$\begin{aligned}
&\log pl_n(\theta_0) + (\tilde{\theta}_n - \theta_0)' \sum_{i=1}^{n} \tilde{l}_0(Y_i) - \frac{1}{2} n(\tilde{\theta}_n - \theta_0)' \tilde{I}_0 (\tilde{\theta}_n - \theta_0)' \\
&+ o_p(\sqrt{n}||\tilde{\theta}_n - \theta_0|| + 1)^2,
\end{aligned}$$

for any consistent estimator $\tilde{\theta}_n$. Here $\tilde{I}_0$ is the efficient information matrix assumed to be non-singular and $\tilde{l}_0$ is the efficient score function. The above expansion implies that the maximum likelihood estimator is asymptotically linear to the efficient score which implies that the maximum likelihood estimator is asymptotically normal with covariance matrix $\tilde{I}_0^{-1}$ (Murphy and Van der Vaart (2000)). The theory suggests that the profile likelihood can be treated as an ordinary likelihood at least in the asymptotic sense. Further the above expansion implies that the curvature of the log profile likelihood can serve as an estimate for the efficient information matrix. Thus for asymptotic normality and efficiency we only need to check the conditions listed in thm 1 in Murphy and Van der Vaart (2000).

Theorem 17. $\sqrt{n}\left(\hat{\theta}_n - \theta_0\right) \xrightarrow{\text{D}} N(0, \tilde{I}_0^{-1})$, *where* $\tilde{I}_0$ *is the efficient information matrix.*

Proof. Here we adapt to the notation in Murphy and Van der Vaart (Murphy and Van der Vaart (2000)) and apply their main thm. For η in a neighborhood of θ_0 define the map $\eta \mapsto \lambda_\eta(\theta, \lambda)$ by

$$\lambda_\eta(\theta, \lambda)(t) = \frac{\int \lambda(t + (\eta - \theta)'z)\gamma_{\eta,z}(t)h(z)dz}{\int \gamma_{\eta,z}(t)h(z)dz},$$

where $\gamma_{\eta,z}$ is defined in section 3.2. Also define the log-likelihood as

$$\eta \mapsto l(\eta, \theta, \lambda) = \log p(y; \eta, \lambda_\eta(\theta, \lambda))$$

Let $\dot{l}(\eta, \theta, \lambda)$ denote the derivative of $\eta \mapsto l(\eta, \theta, \lambda)$. Note that $\lambda_\theta(\theta, \lambda) = \lambda$ for every θ and λ in the parameter space and that $\dot{l}(\theta_0, \theta_0, \lambda_0) = \tilde{l}_{\theta_0,\lambda_0}$, where $\tilde{l}$ is the efficient score given in (2.11). Thus following (Murphy and Van der Vaart (2000)), the above submodel is least favorable at (θ_0, λ_0). Under the regularity conditions it can be shown that in a neighborhood $\mathcal{V}$ around $(\theta_0, \theta_0, \lambda_0)$, $\{\dot{l}(\eta, \theta, \lambda) : (\eta, \theta, \lambda) \in \mathcal{V}\}$ is P_0-Donsker with square integrable envelope and $\{\ddot{l}(\eta, \theta, \lambda) : (\eta, \theta, \lambda) \in \mathcal{V}\}$ is P_0-Glivenko-Cantelli. In order to apply theorem 1 in Murphy and Van der Vaart (2000), we need to check the "no-bias" condition given by

$$E_0\dot{l}(\theta_0, \hat{\theta}_n, \hat{\lambda}_{\hat{\theta}_n}) = o_p(|\hat{\theta}_n - \theta_0| + n^{-1/2}).$$

Consider writing $E_0\dot{l}(\theta_0, \theta_0, \lambda)$ as

$$\begin{aligned} & E_0\left\{\frac{p_{\theta_0,\lambda_0} - p_{\theta_0,\lambda}}{p_{\theta_0,\lambda_0}}\left(\dot{l}(\theta_0, \theta_0, \lambda) - \dot{l}(\theta_0, \theta_0, \lambda_0)\right)\right\} \\ & \quad - E_0\dot{l}(\theta_0, \theta_0, \lambda_0)\left\{\frac{p_{\theta_0,\lambda} - p_{\theta_0,\lambda_0}}{p_{\theta_0,\lambda_0}} - A_0(\lambda - \lambda_0)\right\}, \end{aligned} \tag{2.21}$$

where, A_0 is the score operator for λ at (θ_0, λ_0) and is given by

$$A_0 h = \int_{-\infty}^{e(\theta_0)} Rh(s)dM(s),$$

and R and the martingale M are as in rem 5. Note that $E_0 \tilde{l}_0 A_0 h = 0$ for every $h \in L_2(f_0)$ by the orthogonality property of the efficient score. Since $\lambda \mapsto p_{\theta_0,\lambda}$ is twice differentiable and $\lambda \mapsto \dot{l}(\theta_0, \theta_0, \lambda)$ is differentiable at λ_0, taking a first order Taylor's expansion in the first term in (2.21) and a second order Taylor's expansion in the second term of (2.21) around λ_0 we see that the expression in (2.21) is of order $O_p(||\lambda - \lambda_0||^2)$. Thus following the discussion in Murphy and Van der Vaart (2000) it is sufficient to have

$$||\hat{\lambda}_{\hat{\theta}_n} - \lambda_0|| = O_p(|\hat{\theta}_n - \theta_0|) + o_p(n^{-1/4}),$$

for the "no-bias" condition to hold.

Since $\hat{\theta}_n$ is consistent for θ_0 and $\theta \mapsto \hat{\lambda}_\theta$ is differentiable, by lemma 4, we have the desired result. □

2.7 Simulation Studies

As our method is unconditional on the covariate distribution, it is a special case of the model used in the paper by Zeng and Lin (2007) (since we assume that the covariates are constant over time). So, we use their profile likelihood approach to estimate θ and compare it with Klassen's mean zero approach and also the known covariate structure approach. We consider only 1 covariate Z $\sim Unif(-1, 1)$. So, $\tilde{Z}$ has density given by $\frac{\theta e^{\theta z}}{e^{\theta} - e^{-\theta}}$, where $-1 \leq z \leq 1$. We take different values of θ and assume that the error distribution is standard normal ie U is lognormal. Then, we use all three methods to estimate θ. For the profile-likelihood approach, we use the gaussian kernel and a bandwith of $h_n = Qn^{-1/5}$ where Q is the interquartile range of the data. We consider

1000 replicates and look at the mean bias and variance in estimating θ. We also look at what happens when the covariate distribution is misspecified.For this, we consider Z$\sim$ $(x, 1)$ for some choice of x. So, $\tilde{Z}$ has density given by $\frac{\theta e^{\theta z}}{e^{\theta} - e^{-\theta x}}$ where x$\leq$z$\leq$1. We take the values $x = -0.9$ and -0.8 and compare both the mean 0 and known covariate distribution, assuming Z $\sim Unif(-1, 1)$.We take θ=1 for these simulations. We consider 1000 replicates in this case as well. The results are given in Table 1.

Table 2.1: Estimates for the Backward Recurrence Time Data

Parameters	Sample Size	Profile Likelihood Approach			Vanishing Mean		Known Covariance	
		Bias	SE	CP	Bias	SE	Bias	SE
$\theta = 1$	100	-0.0325	0.258	0.944	-0.1100	0.230	-0.0513	0.201
	200	0.0057	0.188	0.956	-0.0515	0.161	-0.0013	0.137
	400	0.0096	0.133	0.951	0.0022	0.111	0.0005	0.099
$\theta = 0.5$	100	0.0277	0.202	0.958	-0.1645	0.190	-0.0201	0.142
	200	0.0156	0.188	0.951	-0.0333	0.134	0.0068	0.095
	400	0.0099	0.117	0.955	0.0154	0.092	0.0044	0.067
$\theta = 2$	100	0.0099	0.299	0.939	0.0266	0.553	0.0274	0.366
	200	0.0028	0.203	0.957	0.0553	0.284	0.0043	0.216
	400	-0.0045	0.169	0.945	0.0122	0.198	-0.0029	0.105
$x = -0.9$	100	-0.0287	0.254	0.952	-0.0548	0.250	0.0306	0.222
$\theta = 1$	200	0.0094	0.181	0.941	0.1812	0.171	0.1154	0.138
	400	0.0015	0.124	0.946	0.1545	0.114	0.2133	0.094
$x = -0.8$	100	0.0226	0.269	0.933	0.3351	0.270	0.1945	0.211
$\theta = 1$	200	-0.0063	0.169	0.949	0.4895	0.184	0.2338	0.142
	400	0.0034	0.108	0.944	0.3039	0.127	0.2781	0.099

Thus, we find that the estimates obtained using our methods are quite comparable to the special case where the covariance structure is known, though Klassen' methods have lesser variance. However, their estimates are very sensitive to model specification. On the other hand, our naive analysis yields unbiased estimates in both cases. The variance estimators accurately reflect the actual variance, while the confidence intervals have correct coverage probabilities.

2.8 Data Analysis

For the data analysis, we use a subset of the data used by Keiding et al (2012). It is a backward recurrence time data on the time to pregnancy obtained from a large French telephone survey. Women were eligible if they were between 18-44 years old, were living with a male partner and did not use any method to avoid pregnancy. We consider only nulliparous women who had not initiated any fertility treatment. The response variable was the current duration of unprotected intercourse, which is the time elapsed from the start of unprotected intercourse and the interview.

The estimates obtained for the covariates along with the 95 % confidence intervals are given in Table No 2. We note that the naive estimator can accurately determine the effect of the covariates and is comparable with the ordinary least squares results. Thus, we find that the "naive" estimator works really well and usually has better variance than the least squares estimator. Hence, we can conclude that this method works really well in estimating the regression parameter in the AFT model.

Table 2.2: Estimates for time ratios and the CI for nulliparous women

Covariate	No	Semiparametric AFT Time Ratio	OLS Time Ratio
Tobacco Consumption at recruitment			
Non-Smokers	159	1	1
Smokers	92	1.20(0.75,1.78)	1.04(0.70,1.53)
Age at recruitment			
0-17	3	7.50(1.50,38.0)	7.32(1.29,41.4)
18-24	50	2.00(1.20,3.41)	2.08(1.24,3.49)
25-29	93	1	1
30-34	62	1.00(0.61,1.74)	1.01(0.63,1.64)
35-39	41	1.10(0.61,2.02)	0.93(0.54,1.62)
40-44	2	0.13(0.01,1.17)	0.13(0.02,1.10)
Frequency of Sexual Intercourse			
$<$1 per month	0		
1-3 per month	44	2.20(1.20,3.89)	2.18(1.27,3.71)
1-2 per week	109	1.20(0.78,1.92)	1.23(0.81,1.86)
$\geq$3 per week	98	1	1
Menstrual Cycle Length			
$<$27 days	53	1	1
27-29 days	110	0.90(0.52,1.55)	0.86(0.52,1.41)
$\geq$30 days	88	1.10(0.62,1.81)	0.99(0.59,1.63)

CHAPTER 3: TWO-MONOTONE DENSITY ESTIMATION IN THE PRESENCE OF RIGHT CENSORING

3.1 Introduction

A density function g on $\mathbb{R}^+$ is monotone (or 1-monotone) if it is nonincreasing. It is 2-monotone if it is nonincreasing and convex, and k-monotone for k ≥ 3 if and only if $(-1)^j g^{(j)}$ is nonnegative, nonincreasing and convex for j=1(1)k-2. Figure 3.3 shows an estimate of a 2-monotone density for the standard exponential distribution for a sample size of 100, under no censoring.

Estimation of functions restricted by monotonicity or other inequality constraints has received much attention. Estimation of monotone regression and density functions goes has been done by Grenander (1956). Asymptotic distribution theory for monotone regression estimators was established by Brunk (1970), and for monotone density estimators by Prakasa Rao (1969). The asymptotic theory for monotone regression function estimators was reexamined by Wright (1981), and the asymptotic theory for monotone density estimators was reexamined by Groeneboom (1985). The "universal component" of the limiting distribution in these problems is the distribution of the location of the maximum of two-sided Brownian motion minus a parabola. Groeneboom (1988) examined this distribution and other aspects of the limit Gaussian problem with canonical monotone function $f_0(t) = 2t$ in great detail. Groeneboom (1985) provided an algorithm for computing this distribution, and this algorithm has recently been implemented by Groeneboom et al. (2001).

The first work for convex density function estimation has been done by Anevski

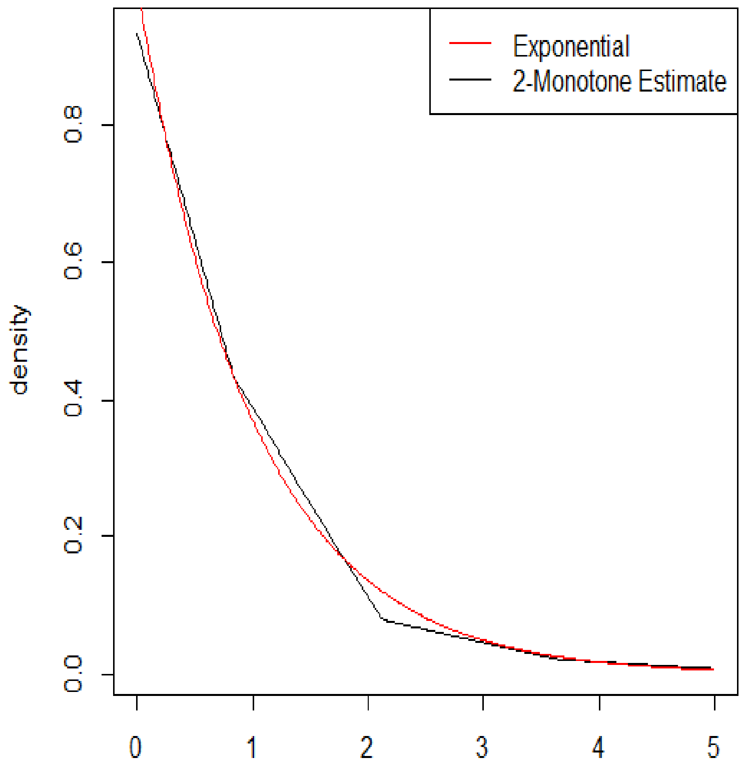


Figure 3.1: Example of a 2-Monotone Density Estimate

(1994), who was motivated by some problems involving the migration of birds discussed by Hampel (1987). Jongbloed established lower bounds for minimax rates of convergence and established rates of convergence for a"sieved maximum likelihood estimator". Finally, a least squares estimator as well as a non-parametric maximum likelihood estimator for 2-monotone densities were established by Groeneboom et al. (2001) which were further modified by Balabdaoui and Wellner (2007) to correct for the consistency near 0.

The least squares (LS) estimator $\tilde{f}_n$ of a convex decreasing density function f_0 is defined as a minimizer of the criterion function

$$Q_n(f) = \frac{1}{2}\int f(x)^2 dx - \int f(x) d\mathbb{F}_n(x),$$

over $\mathcal{K}$, the class of convex and decreasing nonnegative functions on $[0, \infty)$. Here $\mathbb{F}_n$ is the empirical distribution function of the sample. The definition ofQ_n is motivated by the fact that if F_n had density f_n with respect to Lebesgue measure, then the least squares criterion would be

$$\begin{aligned} \frac{1}{2}\int (f(x) - f_n(x))^2 dx &= \frac{1}{2}\int f(x)^2 dx - \int f(x) fn(x) dx + \frac{1}{2} f_n(x)^2 dx \\ &= \frac{1}{2}\int f(x)^2 dx - \int f(x) d\mathbb{F}_n(x) + \int f_n(x)^2 dx, \end{aligned} \tag{3.1}$$

where the last (really undefined) term does not depend on the unknown f with respect to which we seek to minimize the criterion. Note that $\mathcal{C}$, the class of convex and decreasing density functions on $[0,\infty)$, is the subclass of $\mathcal{K}$ consisting of functions with integral 1. The next two lemmas, taken from Groeneboom's paper (Groeneboom et al. 2001) with a slight change in notation, help characterize the LSE for 2-monotone densities.

Lemma 18. *There exists a unique $\tilde{f}_n \in \mathcal{K}$ that minimizes Q_n over $\mathcal{K}$. This solution is piecewise linear and has at most one change of slope between two successive observations $X_{(i)}$ and $X_{(i+1)}$ and no changes of slope at observation points. The first change of slope is to the right of the first order statistic and the last change of slope, which is also the right endpoint of the support of $\tilde{f}_n$, is to the right of the largest order statistic.*

Lemma 19. *Let Y_n be defined by*

$$Y_n(x) = \int_0^x F_n(t) dt, \qquad x \geq 0.$$

Then the piecewise linear function $\tilde{f}_n \in \mathcal{K}$ *minimizes* Q_n *over* $\mathcal{K}$ *if and only if the following conditions are satisfied for* $\tilde{f}_n$ *and its second integral* $\tilde{H}_n = \int_{0<t<u<x} \tilde{f}_n(t)dtdu$*:*

$$\tilde{H}_n(x) \geq Y_n(x), \quad \forall x \geq 0,$$

and

$$\tilde{H}_n(x) = Y_n(x), \quad \text{if } \tilde{f}_n'(x+) > \tilde{f}_n'(x-)$$

For $g \in \mathcal{C}$, the convex subset of $\mathcal{K}$ corresponding to convex and decreasing densities on $[0,\infty)$, define the minus loglikelihood function by

$$-\int \log\, g(x)d\mathbb{F}_n(x), \quad g \in \mathcal{C},$$

and the nonparametric maximum likelihood estimator as the minimizer of this function over $\mathcal{C}$. To relax the constraint $\int g(x)dx = 1$ and get a criterion function to minimize over all of $\mathcal{K}$, we define

$$\psi_n(g) = -\int \log\, g(x)d\mathbb{F}_n(x) + \int g(x)dx, \quad g \in \mathcal{K}.$$

The next lemma from Groeneboom et al. (2001) characterize the MLE estimators for 2-monotone densities.

Lemma 20. *The MLE f, exists and is unique. It is a piecewise linear function and has at most one change of slope in each interval between successive observations. It is also the unique minimizer of* ψ_n *over* $\mathcal{K}$*.*

Thus, for n = 1, the MLE is a function on $[0,\infty)$ which only changes slope at the endpoint of its support. Denoting this point by θ, the observation by x_1, we find that the maximum likelihood estimator corresponds to $\theta = 2x_1$, which differs from the least squares estimator. However, it can be shown that asymptotically, the two estimators are equivalent. The maximum likelihood estimator is characterized as follows: Let $G_n(t,f) = \int_0^t f(u)^{-1} d\mathbb{F}_n(u)$ and let $H_n(t,f) = \int_0^t G_n(u,f)\ du$. Then, the MLE estimator $\hat{H}_n = H_n(t,\hat{f})$ satisfies

$$\hat{H}_n(t) \leq \frac{1}{2}t^2, \quad \forall\ t \geq 0,$$

and

$$\hat{H}_n(t) = \frac{1}{2}t^2, \quad \text{if} \hat{f}_n'(t-) < \hat{f}_n'(t+).$$

This characterization also implies the fact that the estimator can have at most one change of slope between successive observations. So, the MLE estimator can be treated as a sort of "envelope" function whereas the LSE estimator can be treated as an "invelope" function (a term coined by Groeneboom et al.). Both are a kind of "derivative" of the empirical distribution function, just like the Grenander estimator of a decreasing density. Also, the MLE $\hat{f}_n$ solves a sort of a weighted least squares problem with "self-induced" weights.

An example of an invelope function is shown in figure 3.2, and it is taken from Groeneboom, Jongbloed and Wellner's (2001) paper. The solid line Y_n represents the integral of F while the dashed line $\tilde{H}_n$ represents the estimate for a sample of size 20 for the density $f(x) = 3(1-x)^2$, $0 \leq x \leq 1$. The LSE and MLE estimators are uniformly consistent on closed intervals bounded away from 0. More formally, if $X_1, X_2, \ldots$ are i.i.d. observations from $f_0 \in \mathcal{C}$, then for each $c > 0$,

$$sup_{c \leq x < \infty} |\tilde{f}_n(x) - f_0(x)| \rightarrow_{as} 0$$

and

$$sup_{c \leq x < \infty} |\hat{f}_n(x) - f_0(x)| \rightarrow_{as} 0.$$

Although the estimates are inconsistent at 0, Balabdaoui and Wellner (2007) later showed that $\tilde{f}_n(n^{-\alpha}) \to_{as} f_0(0)$, for any $\alpha \in (0, 1/3)$. The convergence rates of both $\tilde{f}_n(t)$ and $\hat{f}_n(t)$ are found to be $n^{2/5}$, while the convergence rate of $\tilde{f}'_n(t)$ and $\hat{f}'_n(t)$ are $n^{1/5}$, for any time point $t > 0$.

3.2 k-monotonicity and Recurrence Times

The result connecting k-monotonicity and the recurrence time structure is given in the following lemma.

Lemma 21. *If a density is k-monotone, then the corresponding recurrence time density is (k+1)-monotone for any value of k>1.*

Proof. The proof of the lemma hinges on the fact that

$$g_{\tilde{T}}(t) = \frac{S_T(t)}{\mu_T}.$$

Thus, $(-1)^j g_{\tilde{T}}^{(j)}(t) = (-1)^{(j-1)} \frac{f_T^{(j-1)}(t)}{\mu_T}$, which is convex and decreasing for j=1(1)k-1. So, the recurrence time density is (k+1)-monotone.

$\square$

Thus, for any given density, the corresponding forward and backward recurrence time data becomes 2-monotone. So, we need a shape constraint for the recurrence time data estimate. Although a lot of work has been done on k-monotone densities, no-one has looked at it under right-censoring. In the next section, we propose two possible estimators for right-censored 2-monotone density estimates based on the LSE and MLE estimates in the censored case.

3.3 Algorithm

We now make the following assumptions for our model:

A1: $\tilde{T}$ and $\tilde{C}$ are independent.

A2: $\mu_f = \int S(v)dv < \infty$.

A3: $E_{\tilde{g}}(T^2) < \infty$.

The first condition basically states that the censoring time is independent of the recurrence time. The second and the third assumptions yield finite first and second mon=ments for the recurrence time. These two conditions can be ensured by assuming $g_0(0+) < \tau < \infty$.

Let us assume that $y_1, \ldots, y_k$ are the uncensored observations and $z_{k+1}, \ldots, z_n$ are the censored observations for some forward recurrence time data. Now, we assume that we have censored forward recurrence time data from some decreasing density. Then, to estimate the original density, we first estimate the corresponding two-monotone density $g_{\tilde{T}}$. After estimating $g_{\tilde{T}}$, we look at the estimation of S_T by using the fact that $S_T(t) = g_{\tilde{T}}(t)/g_{\tilde{T}}(0+)$. Now, although the estimate at g_{0+} under shape restriction is inconsistent, Balabdaoui and Wellner (2007) have shown that the estimate of $g_{\tilde{T}}(n^{-\alpha})$ converges almost surely to $g(0+)$ as n goes to infinity for $0 < \alpha < 1/4$.

We want to estimate the distribution of the data keeping in mind that it is 2-monotone. Our algorithm is as follows:

1. Fit a parametric least squares model to the data.

2. Obtain estimates of the censored observations $(\hat{z_{k+1}}, \ldots \hat{z_n})$ using conditional expectation.

3. Obtain non-parametric least squares estimator of the 2-monotone density using $y_1, \ldots, y_k, \tilde{z}_{k+1}, \ldots \tilde{z}_n$ (using the least squares estimates characterized by Groeneboom et al.).

4. Repeat steps 2-3 until the estimate converges.

5. Use $\widehat{g}(t)/\widehat{g}(n^{-1/5})$ to estimate $S(t)$, the survival function of the original density.

Instead of using the least squares based estimates, we can use MLE in Steps 1 and 3 to yield the following algorithm :

1. Fit a parametric maximum likelihood model to the data.

2. Obtain estimates of the censored observations $(\hat{z_{k+1}}, \ldots \hat{z_n})$ using conditional expectation.

3. Obtain non-parametric maximum likelihood estimates of the 2-monotone density using $y_1, \ldots, y_k, \hat{z}_{k+1}, \ldots \hat{z}_n$ (using the least squares estimates characterized by Groeneboom et al.).

4. Repeat steps 2-3 until the estimate converges.

5. Use $\widehat{g}(t)/\widehat{g}(n^{-1/5})$ to estimate $S(t)$, the survival function of the original density.

3.4 Consistency

Although we have two possible algorithms, we will prove the consistency and asymptotic results using the least squares based estimator ($\tilde{g}_n$) only. The maximum likelihood based estimator is asymptotically equivalent and its results will follow similarly. Let us assume that $\tilde{g}_{nk}$ is the estimated density at the k-th iteration for a sample size of n. Then, the consistency of $\tilde{g}_n$ is given by the following theorem:

Theorem 22. *Suppose that $y_1, \ldots, y_k$ are the uncensored observations and $z_{k+1}, \ldots, z_n$ are the censored observations from i.i.d. random variables with density $f_0 \in \mathcal{C}$. Then, the least squares estimate is uniformly consistent on closed intervals bounded away from 0, i.e., for each c>0, we have w.p. 1, $sup_{c \le x < \infty} |\tilde{g}_n(x) - f_0(x)| \to 0$.*

The outline of the proof of this theorem is given in Appendix 5.3. It is well known that the Grenander estimator of a bounded decreasing density on [0,∞) is inconsistent at zero. A similar result holds for $\tilde{g}_n$. If we assume that both $X_{(1)}$ and $X_{(2)}$ are uncensored, then from its characterization, we can see that

$$\text{Lim inf } P\left(\tilde{g}_n(0) \ge 2f_0(0)\right) > 0.$$

However, using similar arguments as in Balabdaoui, we can show that $\tilde{g}_n(n^{-\alpha}) \to_{as} g_{\tilde{T}}(0)$, for $\alpha \in (0, 1/3)$. We will use $\alpha = 1/5$ for our calculations.

3.5 Convergence Rate and Asymptotic Distribution

We will use lemma 4.1 from Groeneboom to determine the convergence rate. The lemma is stated below:

Lemma 23. *Let x_0 be a point where f_0 is continuous and has a strictly positive second derivative. Let ξ_n be a sequence of numbers converging to x_0 and define $\tau_n^- = max\{t < \xi_n : \tilde{g}_n'(t-) < \tilde{g}_n'(t+)\}$ and $\tau_n^+ = min\{t > \xi_n : \tilde{g}_n'(t-) < \tilde{g}_n'(t+)\}$.*

Then, $\tau_n^+ - \tau_n^- = O_p(n^{-1/5})$.

Using this lemma, we can prove the following theorem,

Theorem 24. *Under the setup of lemma 4.1 of Groeneboom, for each $M > 0$, we have*

$$sup_{|t| \le M} |\tilde{g}_n(x_0 + n^{-1/5}t) - f_0(x_0) - n^{-1/5} t f_0'(x_0)| = O_p(n^{-2/5})$$

and

$$sup_{|t|\leq M}|\tilde{g}_n^{'}(x_0+n^{-1/5}t)-f_0^{'}(x_0)| = O_p(n^{-1/5}).$$

The proof of the theorem is exactly the same and hence is omitted. Finally, we can determine the asymptotic distribution from of the estimates from the following theorem:

Theorem 25. *Under suitable conditions, the least squares estimates obtained asymptotically converge to the following distribution:*

$$n^{2/5}c_1(f_0,\theta)(\tilde{g_n}(x_0)-f_0(x_0)) \rightarrow_d H^{''}(0),$$

$$n^{1/5}c_2(f_0,\theta)(\tilde{g_n^{'}}(x_0)-f_0^{'}(x_0)) \rightarrow_d H^{(3)}(0),$$

where $c_1(f_0,\theta)=(\frac{24}{v(\theta)f_0^2(x_0)f_0^{''}(x_0)})^{1/5}$ *and* $c_2(f_0,\theta)=(\frac{24^3}{v^{1/2}(\theta)f_0(x_0)f_0^{''}(x_0)^3})^{1/5}$*, and where* H *is same as the one defined in section 3.1.*

The proof of this theorem is given in Appendix 5.3.

Remark 26. *We are interested in the survival function of the original density ie* S_T*. Now,* $g_{\tilde{T}}=\frac{S_T}{\mu_T}$ *and*

$$S_T(t)=\frac{g_{\tilde{T}}(t)}{g_{\tilde{T}}(0)}.$$

But, the estimates we obtained are inconsistent at 0. However, it can be shown that $\tilde{g}_n(n^{-\alpha}) \rightarrow_{as} g_{\tilde{T}}(0)$*, for* $\alpha \in (0,1/3)$*. We use* $\alpha = 1/5$ *for our calculations.*

Thus, we find that the algorithm yields a consistent estimator for a 2-monotone density. The results can also easily be extended for a k-monotone density. So, this method can be used to estimate the original density from recurrence time data in the absence of covariates.

3.6 Determination of Monotonicity

An important concern is the determination of the monotonicity of the data obtained if we do not know that beforehand. For this, our idea is to obtain k-monotone density estimates to the data for various choices of k. Then, we use some sort of distance criteria (like maximum absolute deviance) to look at the difference between the consecutive estimates. Finally, we aim to use some sort of change-point analysis to determine the best possible value of k. For example, we have simulated data of sample size 100 from a standard exponential distribution and then from a 5-monotone density. Then we obtained the k-monotone least square density estimates for different choices of k. The results are given in figures 3.3 and 3.4.

In figure 3.3, we are looking at the maximum absolute deviation from the actual exponential density. Thus, we find that as the monotonicity increases, the maximum absolute deviation decreases, i.e. we get a better estimator. This is expected, as the exponential distribution is k-monotone for all k. In figure 3.4, we are looking at the difference of the maximum absolute deviations. So, we find that the deviations decrease at first, and then there is an increase at 5-6 and then it decreases again. So, basically after 5, it starts estimating a different quantity and although the deviations decrease, the estimation is biased.

3.7 Simulation Studies

Thus, we have shown that we can obtain consistent and asymptotically efficient estimators for two-monotone densities in the presence of censoring. Also, we have shown that the least squares based estimator and the maximum likelihood based estimator yield the same asymptotic results. The logical next step is to look at the performance of these estimators under both simulation settings as well as for data obtained from various studies.

For simulation studies, we generate data from the exponential distribution (for various choices of the parameter θ) and then use a uniform density (Uniform$(0, k)$ for various choices of k) to right censor the data. In general, it is seen that as the sample size increases, we have a better fit, i.e. there is a decrease in the maximum absolute difference between the estimated density and the target density. Also, we have a better fit as the censoring proportion decreases. The following figures (3.5 and 3.6) show a special case of the estimation when the sample size is 100 and the censoring proportion is 30%.

The figures show that the estimation is very similar to the 2-monotone density estimation in the absence of censoring. Finally, we look at the maximum absolute deviations for the estimation of an exponential distribution for various choices of n and p. The maximum absolute deviation is calculated over $(0.2, x_{(n)})$.This simulation is repeated 1000 times and the mean value of the maximum deviations is reported in table 3.1.

Table 3.1: Maximum Absolute Deviations for Various Choices of n and p

n	100	200	500
p=0.3	0.235	0.207	0.147
p=0.2	0.211	0.193	0.121
p=0.1	0.200	0.178	0.099

3.8 Data Analysis

We have also obtained permission to use data from the ARIC study. The data is on a population-based cohort of people between ages 44 and 66 and has their age at visit 1 as well as their time of death/censoring. The data comes from a cohort where participants are entered as and when they come into the study. Also, the participants are followed till December 31, 2012, unless lost to follow-up before that. The dataset provides us with a perfect opportunity to use our estimators for length-biased right-censored observations. We have a total of 14,255 participants in the dataset, out of which around 9552 are censored. So, the censoring percentage is around 67% and is quite high. We also have some other covariates which we ignore for this study. The covariates are race (White and African American), indicator of diabetes at baseline,indicator of hypertension at baseline, HDL and LDL levels at baseline in mg/d: and education level. Diabetes is defined as fasting glucose $\geq$ 126 mg/dL, non-fasting glucose $\geq$ 200 mg/dL, self report of diagnosis of diabetes by a physician, or use of diabetic medication in the preceding two weeks. Hypertension is defined as systolic blood pressure $\geq$ 140, diastolic blood pressure $\geq$ 90, or use of anti-hypertensive medications in the previous two weeks. Education level takes the following values :1 = Grade school or 0 years education, 2=High school, but no degree, 3=High school graduate, 4=Vocational school, 5= College and 6= Graduate school or Professional school. Before doing the actual analysis, we look at the summary statistics with respect to the covariates, to see whether there is any difference in the two groups. The results are summarized in table 3.2. This shows that the two groups are quite homogenous with respect to these factors.The distribution of age at entry for the two groups is shown in figure 3.7.

Thus, we find that the two groups are not symmetric with respect to age. Younger people have a higher chance of being censored. Thus, age should be treated as an important factor when we include covariates in the data. Finally, we provide a density estimate in figure 3.8 and compare our method with the parametric approach, assuming the Weibull distribution and the two-parameter gamma distribution for the original density. The knots for the estimated density are $(0.055, 0.190, 0.620, 15, 16, 35, 40, 51, 120)$ with corresponding weights $(0.0037, 0.02, 0.0089, 0.11, 0.13, 0.10, 0.11, 0.45, 0.04, 0.02)$. Now, although the three estimates are close together, we notice that they are vastly different

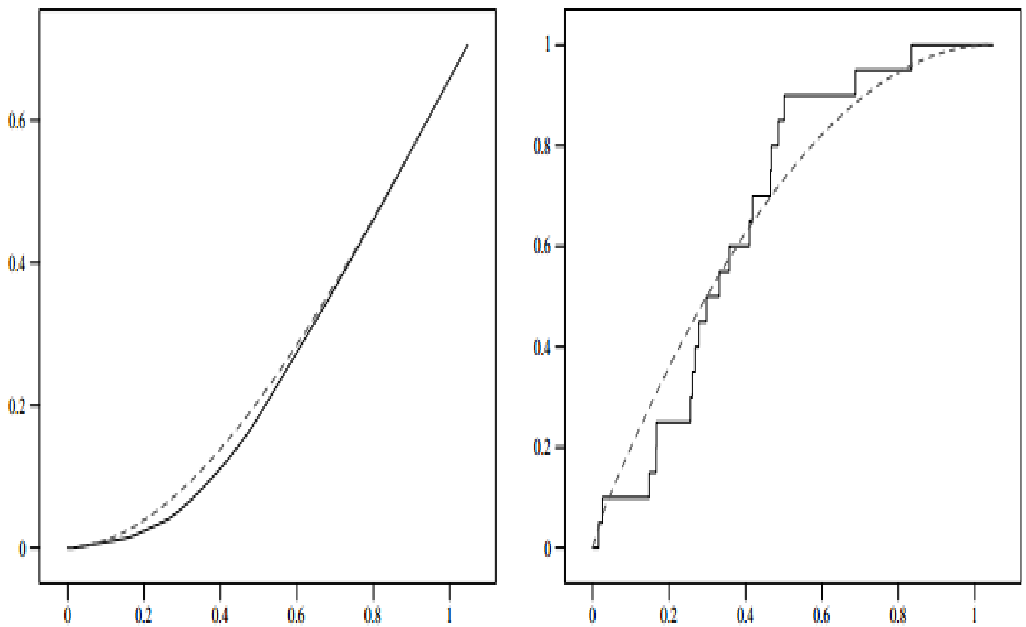
Table 3.2: Summary Statistics for the ARIC Data in the Presence of Censoring

Value	Event (Death)		Unobserved (Censored)	
	Frequency	Percentage	Frequency	Percentage
Race				
1	1510	32.10	2320	24.20
0	3190	67.90	7240	75.80
Hypertension				
1	2530	54.10	6930	72.80
0	2150	45.90	2590	27.20
Diabetes				
1	3760	80.50	8840	93.30
0	908	19.50	639	6.74
Education				
1	680	14.50	593	6.22
2	856	18.20	1110	11.70
3	1440	30.70	3180	33.40
4	368	7.84	831	8.71
5	1000	21.40	2700	28.30
6	346	7.37	1120	11.80

from the empirical distribution of the uncensored observations. In fact, the distribution estimates are less than the empirical distribution of the uncensored observations. The reason for this becomes apparent, when we look at the histogram of the observed death times in figure 3.9. We started with the assumption that the original density was decreasing, which made the forward recurrence time density decreasing and convex. However, it is apparent from the histogram that this condition is not satisfied by the observed data. Hence, even though the estimate obtained by our method is quite close to the estimate obtained by the other two parametric methods, they fail to properly identify the underlying structure of the data. Also, the fact that the censoring percentage is so high, is another reason why the estimates may be a bit unstable.

Figure 3.2: Invelope Function

P. GROENEBOOM, G. JONGBLOED AND J. A. WELLNER



Solid: Y_n, *dashed*: $\tilde{H}_n$.

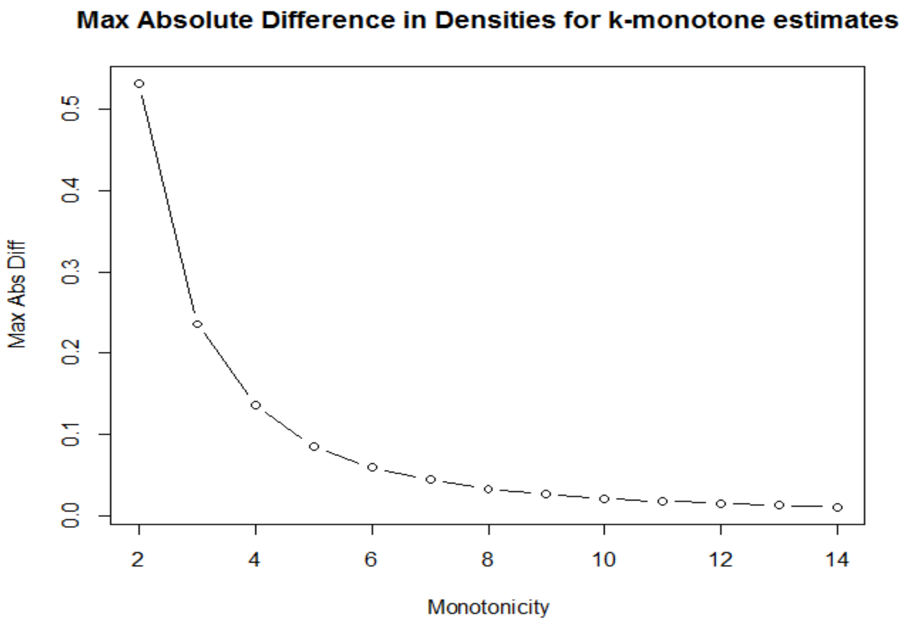


Figure 3.3: k-monotone Density Estimates for the Exponential Distribution.

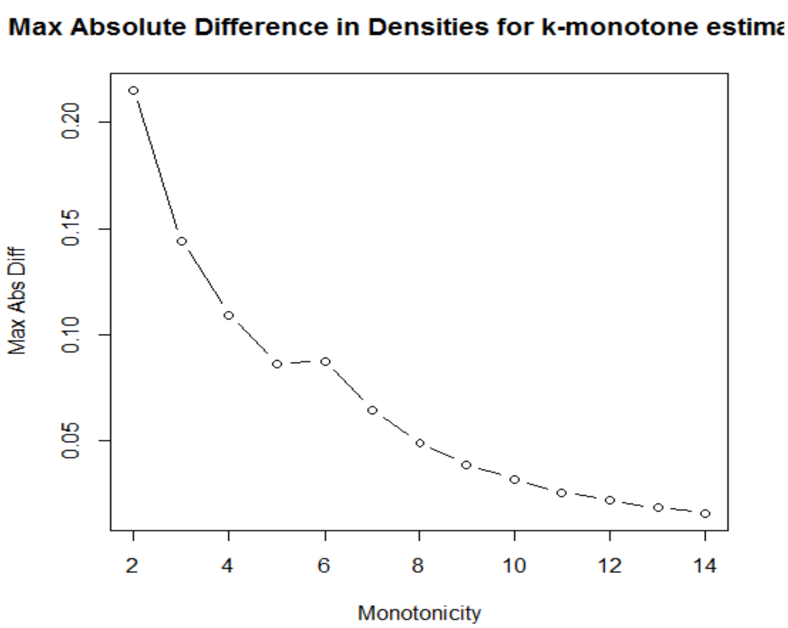


Figure 3.4: Obtaining k-monotone Density Estimates for a 5-Monotone Density.

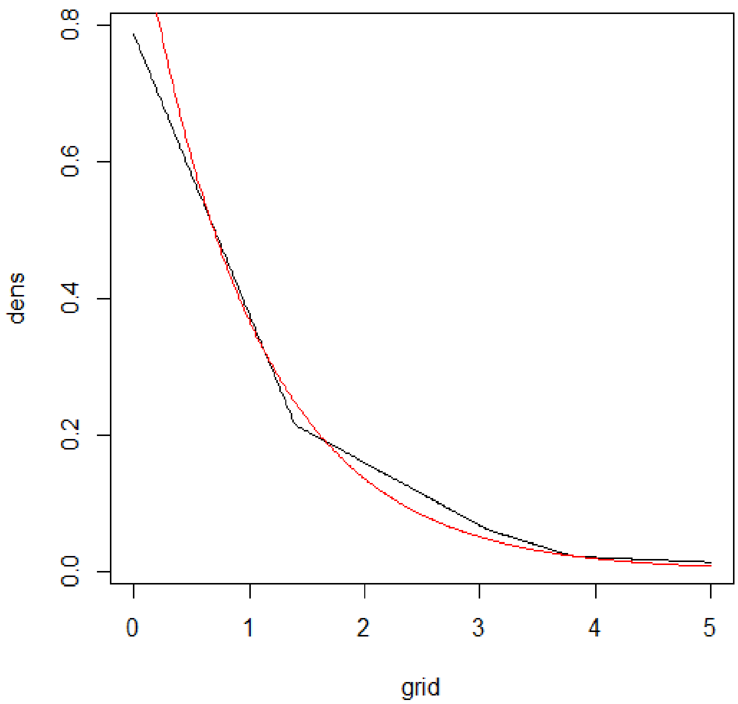


Figure 3.5: Density Estimation, $p = 0.3$

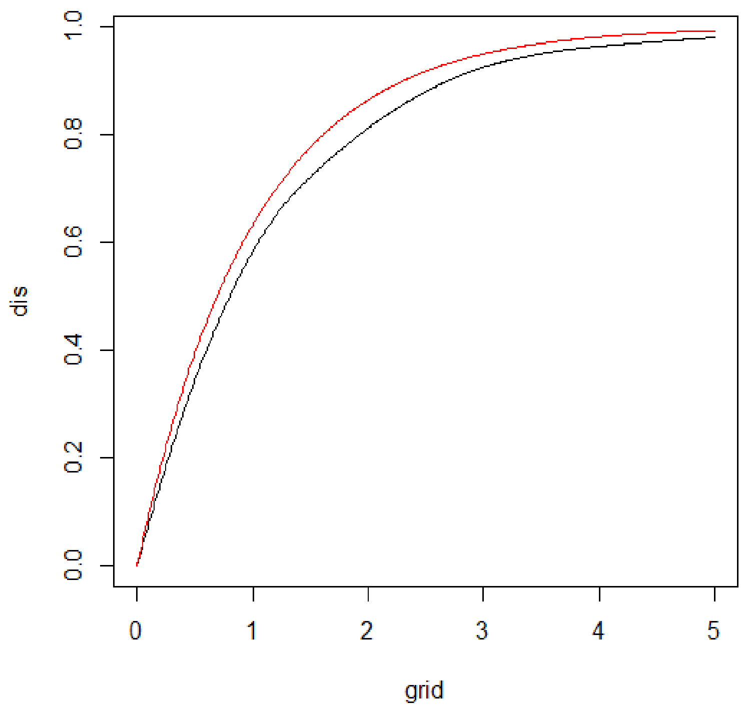


Figure 3.6: Distribution Estimation, $p = 0.3$

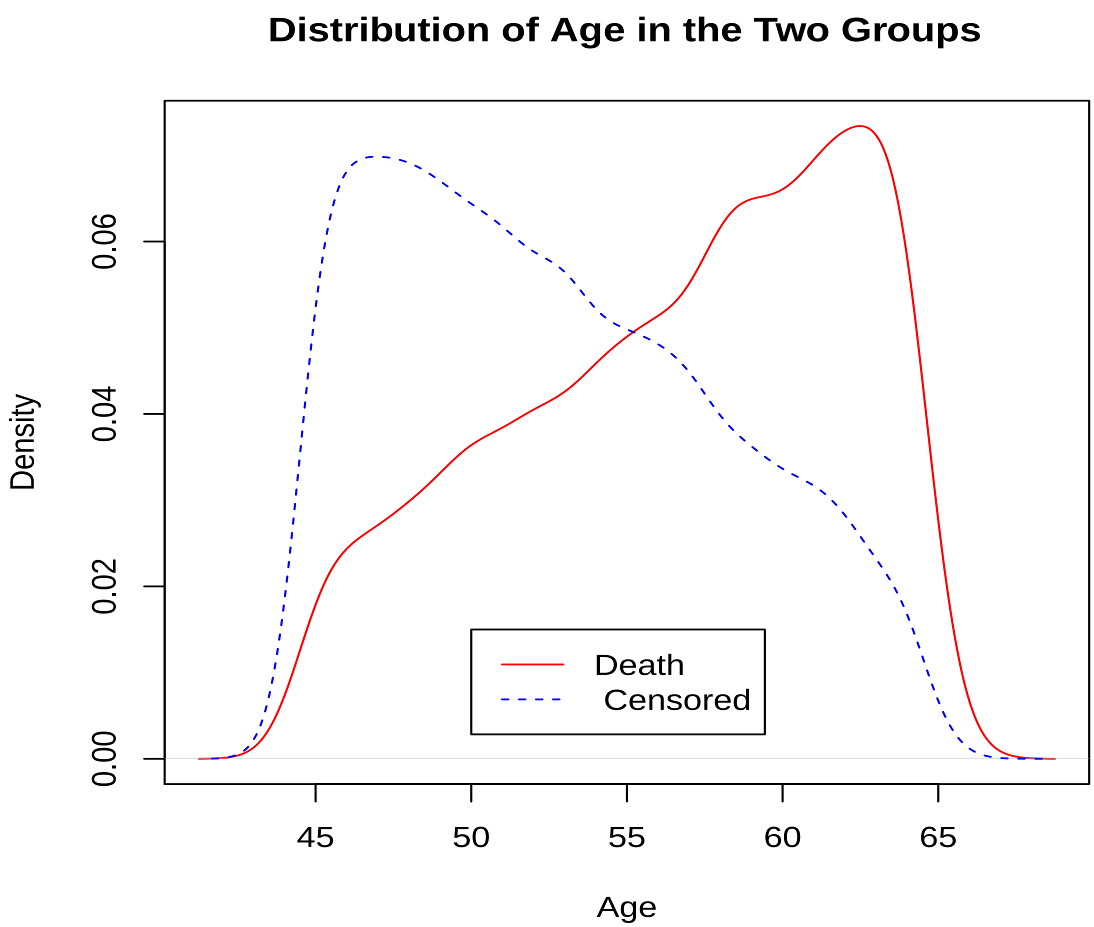


Figure 3.7: Distribution of Age in the Two Groups.

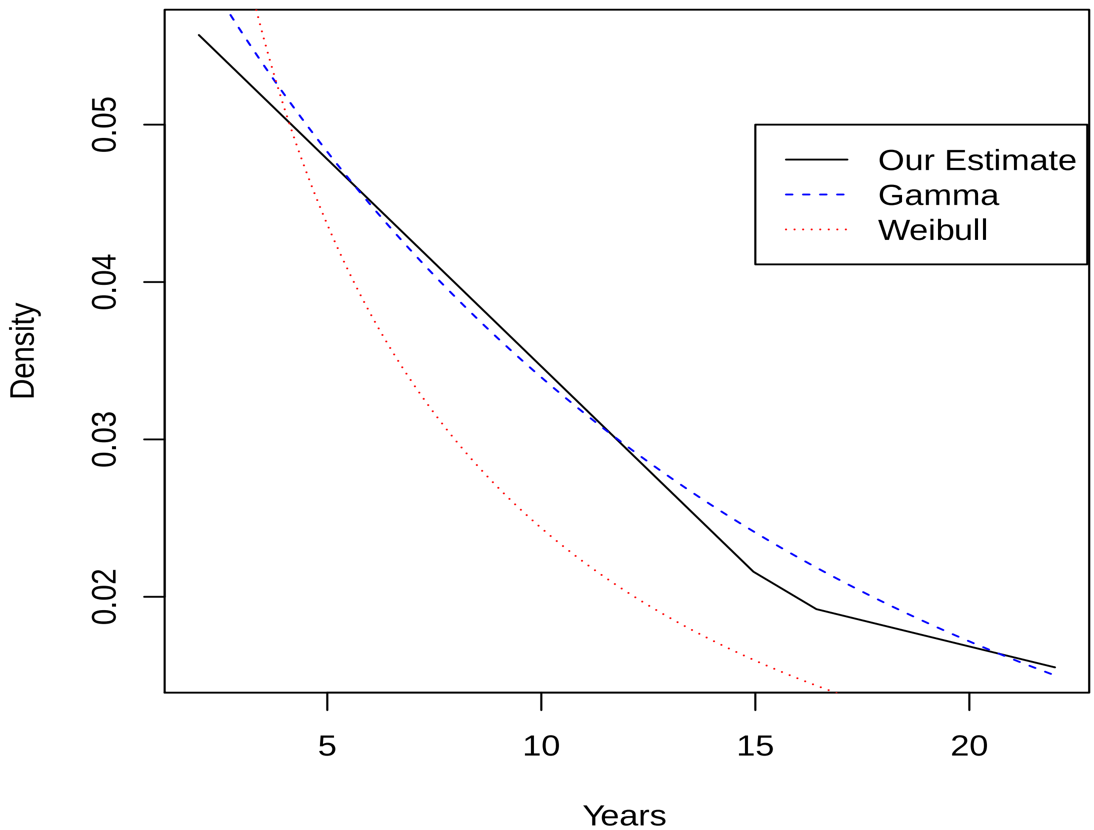


Figure 3.8: Density Estimates for the ARIC data in the Presence of Censoring.

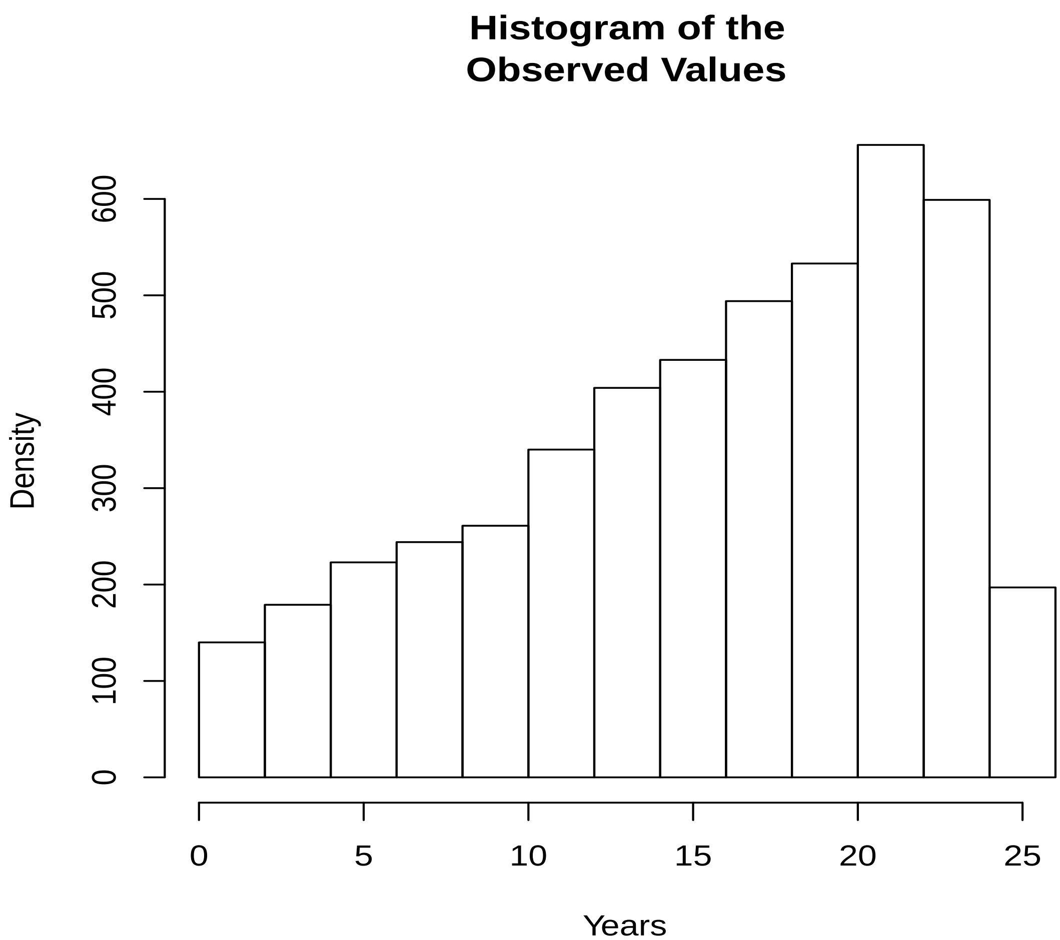


Figure 3.9: Historam of the uncensored observations in the ARIC data.

CHAPTER 4: ESTIMATION OF FRT AND BRT IN COMPETING RISKS

4.1 Introduction

In prevalent cohort survival studies where subjects are recruited at a cross-section and followed prospectively in time, the observed event times are length-biased and further follow a multiplicative censoring scheme. For such studies there is an associated initiation time which may be unknown. In this case we only observe the time from sampling to the event of interest. This is the forward recurrence time. Further, in such cases, standard left-truncation survival analysis methods are not applicable. In other scenarios, such as current duration studies, the time of the initiating event may be known but there is no subsequent follow-up after sampling. Here we observe the backward recurrence times.

In this section, our main goal is to study nonparametric estimation for forward and backward recurrence time data with competing risks in the absence of covariates and censoring. The set-up is as follows. We analyze a system that can fail from K competing risks, where $K \in \mathbb{N}$ and is fixed and finite. The random variables of interest are (X, Y), where $X \in \mathbb{R}$ is the failure time of the system, and $Y \in \{1, \ldots, K\}$ is the corresponding failure cause. We cannot observe (X, Y) directly. Rather, we observe the corresponding recurrence time failure $T \in \mathbb{R}$. This means that at time T, we observe that the failure occurred and we also observe the failure cause Y. Such data can arise naturally in cross-sectional studies with several failure causes. For example, if we look at development of AIDS from the onset of HIV, any other disease arising from HIV may be treated as a competing risk.

The cause-specific hazard, $h_k(t)$, is the instantaneous risk of dying from a particular cause k given that the subject is still alive at time t:

$$h_k(t) = Lim_{\delta t \to 0}\{\frac{P(t \leq T \leq t + \delta t, K = k|T \geq t)}{\delta t}\}$$

and

$$S_k(t) = \exp(-\int_0^t h_k(u)\ du).$$

$S_k(t)$ may be treated as the survival when k is the only cause of failure present. The overall survival is obtained by

$$S(t) = \prod_{i=1}^{K} S_k(t).$$

Cause specific hazards are modelled using the Cox PH model or fitting parametric models taking into account various time-dependent effects.

Another novel approach is to use subdistribution functions (Fine and Gray 1999). The subdistribution hazard, $g_k(t)$, is the instantaneous risk of dying from a particular cause k given that the subject has not died from cause k. So,

$$g_k(t) = Lim_{\delta t \to 0}\{\frac{P(t \leq T \leq t + \delta t, K = k|T \geq t \mathrm{or} T \leq t, K \neq k)}{\delta t}\},$$

and the subsurvival function is given by

$$Q_k(t) = \exp(-\int_0^t g_k(u)\ du).$$

For example, suppose there are two causes of failure for a bulb. The first cause $X_1 \sim \text{Exp}(1)$ and the second cause $X_2 \sim \text{Exp}(1.5)$. Then, the subdistributions of X_1 and X_2 are obtained by $F_i(t) = \int_0^t S(u)h_i(u)du$, for $i = 1, 2$. The subbdistributions are also shown in figure 4.1.

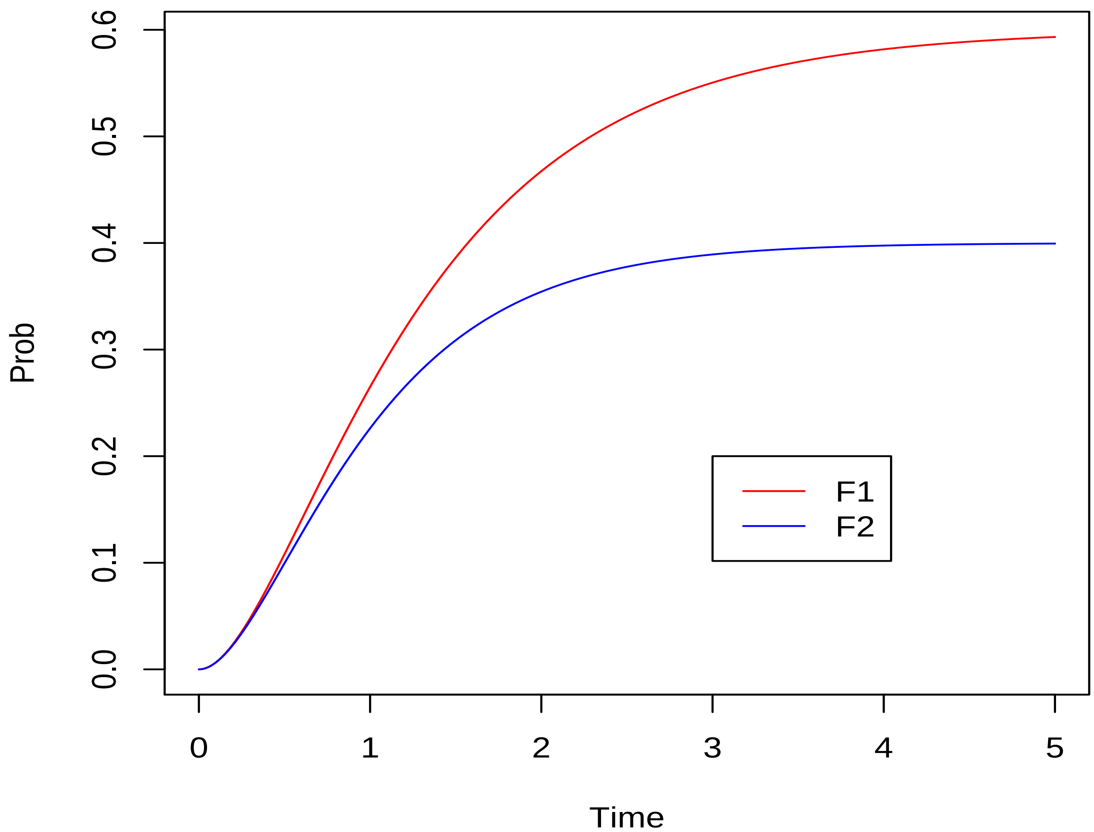


Figure 4.1: Example of Subdistributions

We want to study the nonparametric estimation of the sub-distribution functions $G_{01}, \ldots, G_{0K}$, where $G_{0k}(s) = P(X \leq s, Y = k)$, $k = 1, \ldots, K$. We will show in Section 4.2 that this leads to estimation under a shape constraint. In section 4.3 we provide an estimator for the density. We establish the consistency and convergence rate in 4.4 and the limiting distribution in section 4.5. Finally, we look at the performance of the estimator under a simulation set-up as well as a data analysis in section 4.7.

4.2 Density under Recurrence Times

The following notation will be used throughout this chapter. The observed data is denoted by (T, Δ), where T is the observation time and $\Delta = j$ is the cause of failure, $j = 1 \dots K$. Let (T_i, Δ_i), $i = 1, \dots, n$, be n i.i.d. observations of (T, Δ). The order statistics of $T_1, \dots, T_n$ are denoted by $T_{(1)}, \dots, T_{(n)}$. Furthermore, F_j are the different subdistributions of T, $\mathbb{F}_{nj}$ are the empirical subdistributions of the observed T_i, $i = 1, \dots, n$.

In both forward and backward recurrence time studies, only subjects who have experienced the initiating event prior to sampling, but have not yet experienced the terminating event, can be sampled. Thus, both the forward and the backward recurrence times are length biased i.e., the sample is biased towards larger values of T. One way to model this bias (Cox (1969), Vardi (1982)) is to sample proportionally to length, i.e., if F_T is the distribution of T then the length-biased version T_{LB} has a distribution given by

$$F_{LB}(t) = \frac{\int_0^t u dF_T(u)}{\mu_T}, \quad t \geq 0,$$

where $\mu_T = \int_0^\infty u dF_T(u)$.

Further, if it can be assumed that the incidence of the disease follows a stationary Poisson process then the cross-sectional sampling time is distributed uniformly between the onset time and the terminating time (Cox (1969), Van Es et al. (2000), Keiding et al. (2002)). Thus $T_f = T_{LB}V$, where V is uniform(0,1). It follows that if $S_T = 1 - F_T$ is the survival function of T then for both T_f and T_b (commonly denoted as $\tilde{T}$ here), the density $g_{\tilde{T}}$ is given by

$$g_{\tilde{T}}(t) = \frac{S_T(t)}{\mu_T}. \tag{4.1}$$

So, in the presence of competing risks

$$g_{\tilde{T}}(t) = \frac{\prod \mu_i}{\mu} \prod_{i=1}^{K} \frac{S_i(t)}{\mu_i(t)},$$

where μ_i is the mean of the cause specific failure distributions. For now, let us assume that there are two underlying competing risks X_1 and X_2 with corresponding densities l_1 and l_2 respectively. Hence, we observe T=min(X_1, X_2). Let us assume that the corresponding subdensities are f_1 and f_2. Then, the corresponding density in the length biased case is given by the following lemma.

Lemma 27. *Under the setup given above, the recurrence time densities are given by* $v_1(x) = Q_1(x)/\mu$ *and* $v_2(x) = Q_2(x)/\mu$*, where* $Q_i(x) = \int_x^\infty f_i(t)dt$*, for i=1,2 and* μ *is the mean of the original distribution* $f = f_1 + f_2$*.*

Proof. Let Y_l and Z_l denote the corresponding length-biased distributions for X_1 and X_2. Then, we can adapt the proof in (Dauxois et al. 2014) for the uncensored case to show that

$G_1(y_l) = \frac{1}{\mu} \int_0^{y_l} x l_1(x) \bar{L}_2(x) dx.$

Hence, $g_1(y_l) = \frac{1}{\mu} y_l f_1(y_l)$.

Similarly, $g_2(z_l) = \frac{1}{\mu} z_l f_2(z_l)$.

Now, let $\tilde{X}_1$ and $\tilde{X}_2$ be the corresponding recurrence time densities.

Then, we have, $\tilde{X}_1 = Y_l U$, where $U \sim Uniform(0,1)$.

Hence, $f(\tilde{x} \mid y_l) = 1/y_l$, $0 \leq \tilde{x} \leq y_l$.

So, $v_1(x) = \int_x^\infty \frac{1}{y_l} \frac{1}{\mu} f_1(y_l) dy_l = \frac{1}{\mu} \int_x^\infty f_1(y_l) dy_l = Q_1(x)/\mu$.

Similarly, we can obtain that $v_2(x) = Q_2(x)/\mu$. □

The previous result can be easily extended to multiple competing risks, by assuming that the other competing risks together form a separate larger risk and using the previous calculation. The result is summarized in the next corollary:

Corollary 28. *Under the previous setup if we have K competing risks with subdensities* $g_1, \ldots, g_K$*, the recurrence time densities are given by* $v_i(x) = Q_i(x)/\mu$*, where* $Q_i(x) = \int_x^\infty f_i(t)dt$*, for i=1,2 and* μ *is the mean of the original distribution i.e.,* $\mu = \int_0^\infty S(u)du$*.*

Thus, we find that there is an additional shape restriction on the observed subdensities. The recurrence time subdensities are all decreasing. We start by looking at some parametric estimation under different scenarios in 4.6. Then, we formally define an estimator in section 4.3.

4.3 Non-Parametric Estimation

4.3.1 Previous Work

The Kaplan-Meier estimator can easily be generalized to include competing risks. Let $t_{j1} < t_{j2} < \cdots < t_{jk_j}$ denote the k_j distinct failure times for failures of type j. Let n_{ji} denote the number of subjects at risk just before t_{ji} and let d_{ji} denote the number of deaths due to cause j at time t_{ji}. Then the same arguments used to derive the usual K-M estimator lead to

$$\hat{S}_j(t) = \prod_{i:t_{ji}<t} \left(1 - \frac{d_{ji}}{n_{ji}}\right).$$

It is interesting to note that $\hat{S}_j(t)$ is exactly the same as the standard K-M estimator that one would obtain if all failures of type other than j were treated as censored cases. If there are no ties between different types of failure, then

$$\hat{S}(t) = \prod_{j=1}^{K} \hat{S}_j(t),$$

so the K-M estimator of the overall survival is the product of the K-M estimators of the cause-specific survivor-like functions.

The Nelson-Aalen estimator of the cause-specific cumulative hazard is

$$\hat{\Lambda}_j(t) = \sum_{i:t_{ji}<t} \frac{d_{ji}}{n_{ji}},$$

and corresponds to an estimate of the cause-specific hazard $\lambda_j(t)$ that takes the value d_{ji}/n_{ji} at t_{ji} and 0 elsewhere. One can also exponentiate the negative of the Nelson-Aalen integrated hazard to obtain an alternative estimator of the cause-specific survivor-like function $S_j(t)$. A non-parametric maximum likelihood estimator of $F_j(t)$ was proposed by (Aalen 1976) and can be thought of as a special case of the Aalen-Johansen theory of estimation for time-inhomogenous Markov processes (Aalen and Johansen 1978). The estimator, known as the Aalen-Johansen estimator is given by

$$F_j(t) = \sum_{i:t_{ji}<t} \hat{S}(t_{ji-1}) \frac{d_{ji}}{n_{ji}}.$$

Our aim is to use this Aalen-Johansen estimator to find an estimate of $F_j(t)$ under the added restriction that $f_j(t)$ is decreasing.

4.3.2 Methods

We now make the following assumptions for our model:

A1: $\left(\tilde{T}, \Delta\right)$ and $\tilde{C}$ are independent.

A2: $\mu_f = \int S(v)dv < \infty$.

A3: $E_{\tilde{g}}(T^2) < \infty$.

The first condition basically states that the censoring time is independent of the recurrence time. The second and the third assumptions yield finite first and second moments for the recurrence time. These two conditions can be ensured by assuming

$g_0(0+) < \tau < \infty$. Although we need finite moments for the subdistributions as well, since $f(t) = f_1(t) + \cdots + f_K(t)$, the above conditions ensure that all the moments exist for the subdistributions. The algorithm is as follows:

- We obtain the unrestricted NPMLE of the subdistriutions, i.e., the Aalen Johansen estimator.
- We use the LCM (least concave majorant of the estimated subdistributions) as the estimate under the decreasing density assumption ($\tilde{\mathbb{F}}_j$).
- The derivative of the subdistributions gives us the decreasing subdensities ($\tilde{f}_j$).
- We estimate the overall density by $f = \sum_{i=1}^{K} \tilde{f}_j$.
- We use $\hat{Q}_j = \tilde{f}_j / f(n^{-\alpha})$.

An important thing to note is that under no censoring, the Aalen-Johansen estimator reduces to calculating the empirical subdistribution functions for the estimates, i.e.,

$$\mathbb{F}_j(t) = \frac{1}{n} \sum_{i=1}^{n} I(T_i \leq t, \Delta = j).$$

This fact will be used in the following sections to prove consistency and determine the rate of convergence and the asymptotic distributions.

4.4 Consistency and Convergence Rate

Let us denote the LCM obtained by the notation $\tilde{\mathbb{F}}_j(t)$ We first prove that the estimators $\tilde{\mathbb{F}}_j$ based on the Aalen-Johansen estimator are consistent. The proof is based on a minor modification of Marshall's lemma. Let $|| \,.\, ||_a^b$ denote the supremum norm over the interval $[a, b]$. For concave F, we have Marshall's lemma.

Lemma 29. *If F is concave, then for any $0 < \tau < \tau_H$,*

$$|| \tilde{\mathbb{F}}_{jn} - F_j ||_0^\tau \leq || \mathbb{F}_{jn} - F_j ||_0^{\tau_n},$$

where $\tau_n = inf\ \{t : \tilde{f}_{jn}(t+) < \tilde{f}_{jn}(t)\}$.

Proof. We can see that

$$|| \tilde{\mathbb{F}}_{jn} - F_j ||_0^\tau \leq || \tilde{\mathbb{F}}_{jn} - F_j ||_0^{\tau_n} \leq || \mathbb{F}_{jn} - F_j ||_0^{\tau_n},$$

where the second inequality follows from Marshall's lemma.

□

Theorem 30. *Under the given conditions, $\tilde{f}_{jn}(t) \to_{as} f_j(t)$ for all $j = 1, \ldots, K$ and for all $t > 0$.*

Proof. Fix $0 < \delta < t$, and note that by definition of $\tilde{f}_{jn}$,

$$\frac{\tilde{\mathbb{F}}_{jn}(t+\delta) - \tilde{\mathbb{F}}_{jn}(t)}{\delta} \leq \tilde{f}_{jn}(t) \leq \frac{\tilde{\mathbb{F}}_{jn}(t) - \tilde{\mathbb{F}}_{jn}(t-\delta)}{\delta}.$$

By Marshall's lemma, the upper and lower bounds converge almost surely to $\delta^{-1}(f_j(t) - f_j(t-\delta))$ and $\delta^{-1}(f_j(t+\delta) - f_j(t))$, respectively. By the assumptions on f_j and the arbitrariness of δ, we obtain $\tilde{f}_{jn} \to^{as} f_j(t)$.

□

The strong consistency of $\tilde{\mathbb{F}}_{jn}$ follows directly from the above lemma. To determine the rate of convergence, we need to perform an interesting inverse transformation of the problem that will also be useful for obtaining the weak limiting distribution. Define the stochastic process $\{\hat{s}_n(a) : a > 0\}$ by $\hat{s}_n(a) = argmax_{s \geq 0}\{\mathbb{F}_{jn}(s) - as\}$, where the largest value is selected when multiple maximizers exist.

Theorem 31. *The rate of convergence of $\tilde{f}_{jn}$ to f_j is $n^{1/3}$, for $j = 1, \ldots, K$.*

The proof of this theorem is outlined in Appendix 5.3.

4.5 Asymptotic Distribution

The main theorem regarding the asymptotic distribution is given below:

Theorem 32. *Suppose that f_j is as described above and is also continuous. Then, we can conclude that for any t bounded away from 0,*

$$n^{1/3} \mid f_j(t) f^{j(1)}(t) \mid (\tilde{f}_{jn}(t) - f_j(t)) \to_d 2\mathbb{C},$$

where $\mathbb{C}$ denotes the Chernoff distribution and $f^{j(1)}$ is the derivative of f_j.

The main proof of this theorem is given in Appendix 5.3.

Remark 33. *To determine the asymptotic covariance between $\tilde{f}_{jn}(t)$ and $\tilde{f}_{kn}$, we note that in the proof of the asymptotic distribution, we have a martingale M_n. The martingales for the two processes are going to be dependent because $\mathbb{F}_n^j$ and $\mathbb{F}_n^k$ have asymptotic covariance $-F_j F_k$. This can be simplified to yield the required covariance.*

4.6 Simulations for Parametric Estimation

For parametric estimations, we look at the Weibull and 2-parameter Gamma densities. We assume that there are two competing risks and no censoring. We follow the approach given in Jeong and Fine (2006). The idea is to model the cause-specific hazard instead of the whole survival function. For two competing risks, we can write $\lambda(t \mid \Psi) = \lambda_1(t \mid \Psi_1) + \lambda_1(t \mid \Psi_2)$. Then, the survival function becomes

$$S(t \mid \Psi) = \int_0^t \exp\{-\lambda_1(u \mid \Psi_1) - \lambda_1(\mid \Psi_2)\} du.$$

Thus, $S(t \mid \Psi) = S_1(t \mid \Psi_1)S_2(t \mid \Psi_2)$, where S_1 and S_2 may be thought of as two pseudo survival functions. The idea is to model these pseudosurvival functions, so that we know the cause specific hazard in each case. For our simulations, we assume that the original density has pseudosurvival functions S_1 and S_2, which follow a Weibull distribution with parameters (1,1) and (2,2) respectively. We simulate recurrence times for these observations for a sample size of 100 and estimate the mean and the variance of the parameters. The simulations are repeated 1000 times and the results are summarized in table 4.1. The above simulation setup is repeated for a sample size of 500. Next, we assume that the pseudosurvival functions follows Gamma densities with parameters (1,1) and (2,2), under the same setup and perform the simulations in exactly the same way. All the results are summarized in Table 4.1.

Table 4.1: Estimated Parameters for the Simulated Subsurvival Functions

Parameters	n=100			n=500		
	Bias	MSE	CP	Bias	MSE	CP
Weibull						
α_1	.0846	0.2827	0.944	0.0100	.0991	0.953
β_1	.0727	0.3210	0.953	0.0002	0.1440	0.951
α_2	.2419	0.7585	0.939	−0.2337	0.2727	0.942
β_2	.0323	0.2743	0.955	0.0203	0.1149	0.952
Gamma						
α_1	0.0521	0.2341	0.951	0.0092	0.0995	0.954
β_1	0.0326	0.2579	0.951	0.0018	0.0932	0.949
α_2	0.2840	0.5386	0.943	0.1032	0.1763	0.947
β_2	0.1064	0.2483	0.947	0.0102	0.0982	0.951

Thus, we find that the estimates obtained using our methods are quite close to the actual parameter values, and that the mean bias decreases as the sample size increases. The variance estimators accurately reflect the actual variance, while the confidence intervals have correct coverage probabilities. This validates our calculations for the recurrence time densities.

4.7 Data Analysis

We will utilize data from the ARIC study. The data is on a population-based cohort between ages 44 and 66 and has their age at visit 1 as well as their time of death/censoring. The data comes from a cohort where participants are entered as and when they come into the study. Also, the participants are followed till December 31, 2012, unless lost to follow-up before that. The dataset provides us with a perfect opportunity to use our estimators for competing risks data. We have a total of 14,255 people in the dataset, out of which around 9552 are censored. So, the censoring percentage is around 67% and is quite high. We also have some other covariates which we ignore for the study at this time. The covariates are race (White and African American), indicator of diabetes at baseline,indicator of hypertension at baseline, HDL and LDL levels at baseline in mg/dL and education level. Diabetes is defined as fasting glucose $\geq$ 126 mg/dL, non-fasting glucose $\geq$ 200 mg/dL, self report of diagnosis of diabetes by a physician, or use of diabetic medication in the preceding two weeks. Hypertension is defined as systolic blood pressure $\geq$ 140, diastolic blood pressure $\geq$ 90, or use of anti-hypertensive medications in the previous two weeks. Education level takes the following values: 1 = Grade school or 0 years education, 2=High school, but no degree, 3=High school graduate, 4=Vocational school, 5= College and 6= Graduate school or Professional school. The competing risks we are interested in are incident stroke, heart failure, heart attack and death. The variable for heart attack actually looks at incident myocardial infarction (MI, i.e., a heart attack), fatal CHD, silect MI detected by ECG, or coronary revascularization procedure and is the most general heart attack related variable in ARIC and used in most manuscripts. Since a large portion of the data is actually censored, we are going to lump the death and censored observations into one variable. So, our main inference is going to be about the other three variables. Before doing the actual analysis, we look at the summary statistics with respect to the covariates, to see whether there is any difference between the different groups. The results are summarized in 4.2. Thus, we find that the covariates are quite similar in the different competing risks groups.

Table 4.2: Summary Statistics for the ARIC Data in the presence of Competing Risks

Value	Stroke		Heart Failure		Heart Attack		Death/Cens	
	Frequency	Percentage	Frequency	Percentage	Frequency	Percentage	Frequency	Percentage
Race								
1	352	40.50	618	35.30	288	25.10	2570	24.50
2	517	59.50	1140	64.70	861	74.90	7920	75.50
Hypertension								
1	406	46.90	896	51.20	629	55.10	7530	72.10
2	459	53.10	855	48.80	512	44.90	2910	27.90
Diabetes								
1	688	80.00	1320	76.30	930	81.70	9650	92.80
2	172	20.00	412	23.70	208	18.30	754	7.25
Education								
1	134	15.40	259	14.80	136	11.80	744	7.11
2	157	18.10	356	20.40	188	16.40	1270	12.10
3	256	29.50	517	29.60	368	32.10	3490	33.30
4	74	8.52	112	6.40	89	7.75	924	8.83
5	175	20.10	383	21.90	279	24.30	2860	27.30
6	73	8.40	122	6.98	88	7.67	1190	11.30

We estimate the subdistribution functions using our method, and the hazard specific cumulative incidence function approach from the Jeong and Fine (2006) paper. We estimate the parametric models using Weibull and 2-parameter gamma as the pseudo-survival functions. The results are given in figures 4.2, 4.3 and 4.4. We also look at the subdensity estimates for our estimator in 4.5. Thus, we see that the three estimators yield comparable subdistribution functions. Therefore, our method is comparable to the parametric models and yield well-behaved nonparametric estimators.

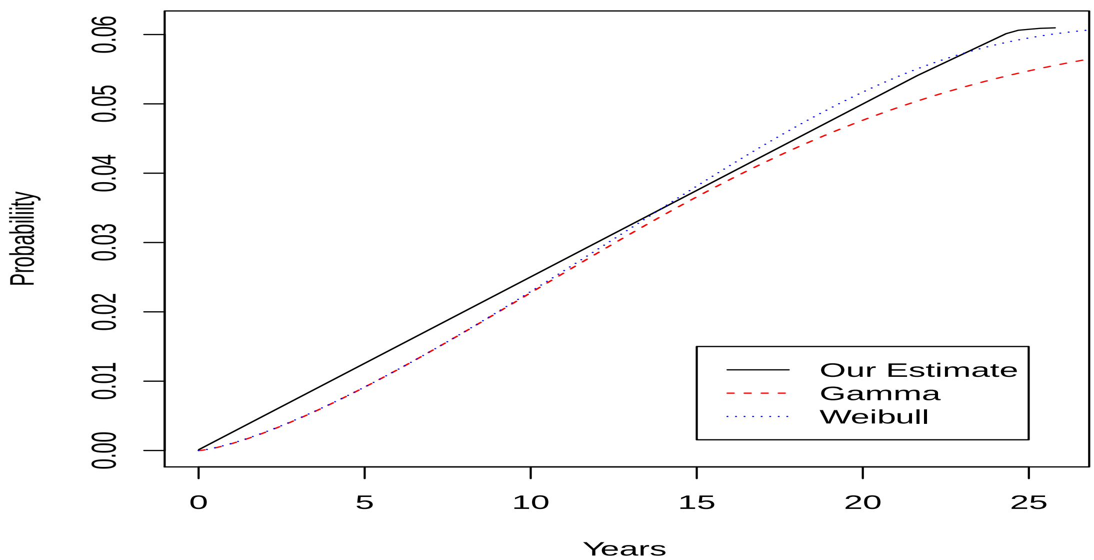


Figure 4.2: Estimates of the Subdistribution of Stroke for the ARIC Data.

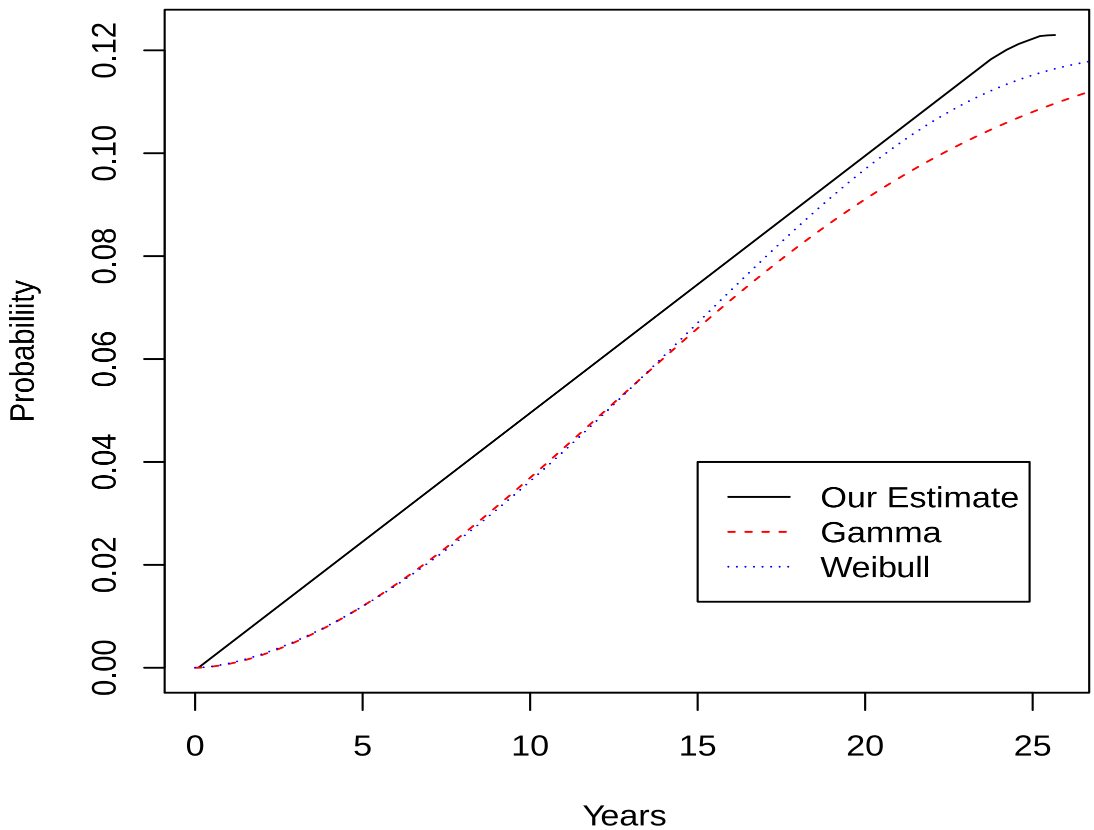


Figure 4.3: Estimates of the Subdistribution of Heart Failure for the ARIC Data.

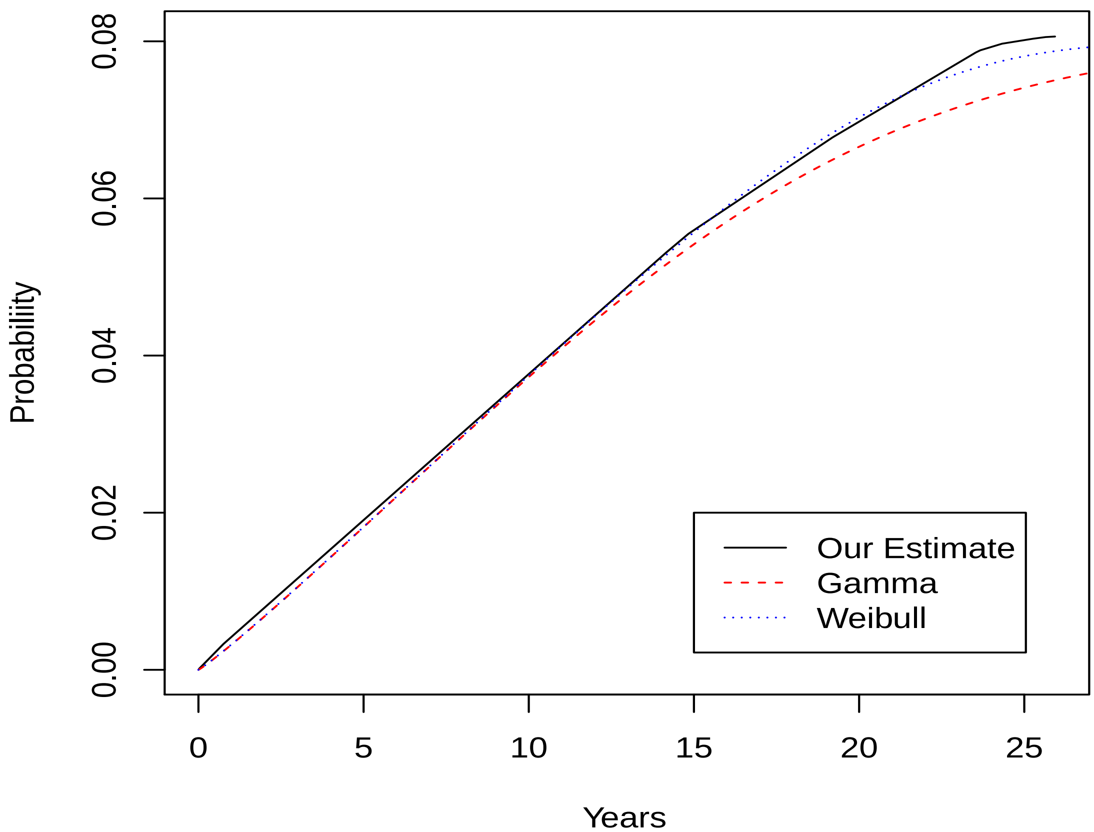


Figure 4.4: Estimates of the Subdistribution of MI for the ARIC Data.

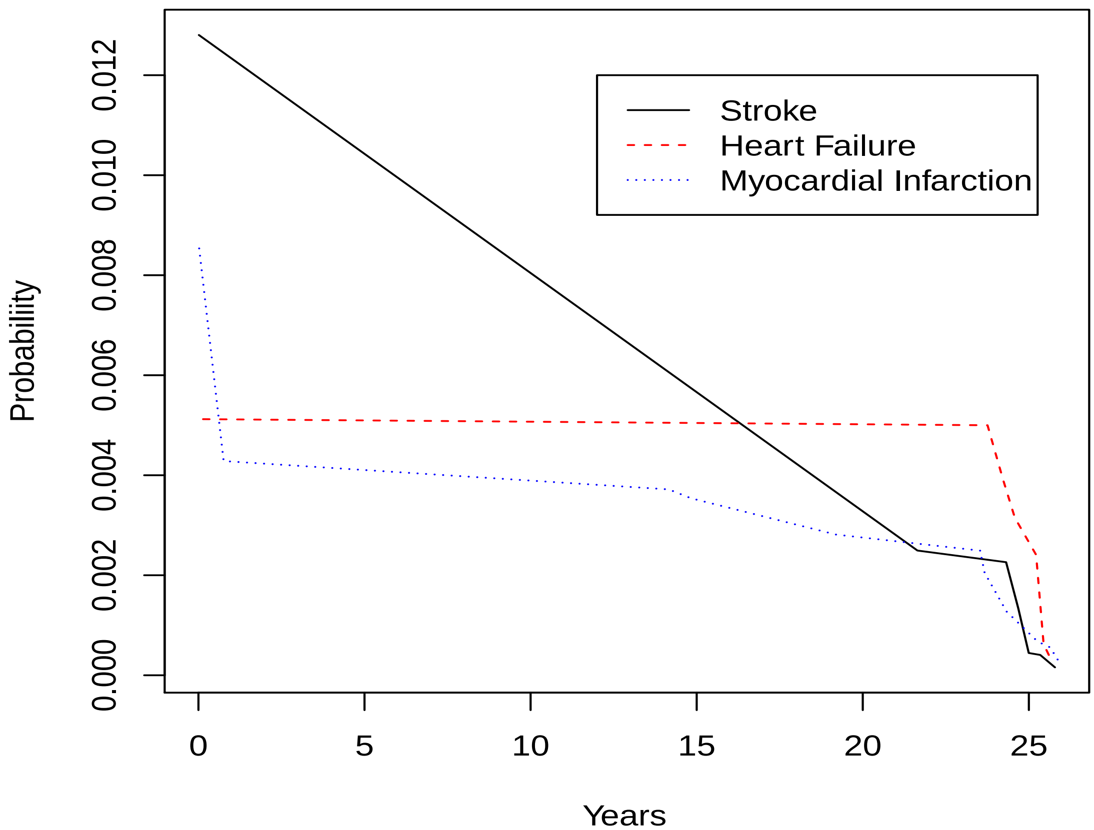


Figure 4.5: Subdensity Estimates for the three competing risks based on our method.

CHAPTER 5: DISCUSSIONS AND FUTURE PROJECTS

In this dissertation, we focused on three different nonparametric and semiparametric methods used in recurrence time data estimation. The first focused on the estimation of the AFT model. The second and third both looked at non-parametric estimation in the absence of covariates. The second was focussed on estimation in the presence of censoring, while the third looked at estimation in the presence of competing risks, when there is no censoring.

5.1 AFT model for Recurrence Time

Accelerated failure time models have become an important alternative to the popularly used Cox Proportional Hazard model in the analysis of censored data. AFT models are particularly useful when we are interested in studying the effects of covariates on duration. Different approaches to estimating covariate effect on duration or its logarithmic transform in the presence of censoring are the least squares type estimators, the rank based estimators, minimum distance estimators and modified Buckley-James estimators. In the study of length-biased duration data collected at a cross-section, AFT models become particularly useful in light of the invariance property that they are preserved with the same covariate effect, when shifting from the case of fully observed durations (for example, disease onset to death) to the corresponding partially observed backward or forward recurrence times arising out of prevalent cohort and current durations study designs. This invariance is quite intuitive: linearity between the time variable and the covariates is preserved under length-biased cross-sectional sampling. Also, the correspondence between the underlying survival function and the survival function for the forward or backward recurrence time under the stationarity of the incidence process, allows for estimation of the former. In view of this, a semiparametric approach is necessary to enable estimation of the regression parameter as well as the underlying survival function. For the purpose of efficiently estimating the regression parameter, one needs to identify the corresponding semiparametric efficient score. This is crucial since in backward and the forward recurrence time cases, the covariate distribution is functionally dependent on the regression parameter and thus bringing into question the validity and efficiency of a naive (based on not considering the effect of length-biased cross-sectional sampling) analysis observed forward or backward recurrence times. The efficient score function does not change if we assume that the covariate distribution is degenerate at the observed values and thereby validating the use of a naive estimator for the regression parameter based on conditioning on the covariates. However, for the purpose of estimating the core survival function, a naive analysis is not valid in general.

It is of both practical and theoretical interest to investigate such invariance properties for other known statistical models. In particular, we saw that the proportional hazard model $\lambda(t \mid z) = e^{-\theta' z}\lambda_0(t)$ is not preserved under length-biased cross-sectional sampling, except when the core time variable has a Pareto density, which includes the exponential distribution as a limiting case. For the special case of the exponential distribution, the PH model coincides with the AFT model. A future research interest is to investigate semiparametric linear transformation models given by$T = g(\theta' Z)U$, where U has a known density but the transformation m is unknown and needs to be estimated.

Another important topic to look at is time-dependent accelerated failure time model. Here, we assume that we observe the failure time T and the covariate history $\bar{Z}(T)$, i.e., the covariate history upto time T. In the time-dependent model, there is a latent variable U, which is related to the observation $(T, \bar{Z}(T))$ as

$$U = \int_0^T exp\{\phi_0 \bar{Z}(t)\},$$

where ϕ_0 is an unknown parameter (Robins 1992). The AFT structure may not remain if we look at the corresponding recurrence time model for the time-dependent data. So, this is an interesting area for future research.

5.2 Censored 2-Monotone Data

In Chapter 3, we have successfully developed an algorithm to determine the non-parametric two-monotone density estimate. The main purpose behind doing this was because recurrence time data for decreasing densities are going to be two-monotone. We have also shown that if the original density is k-monotone, then the recurrence time density is $(k+1)$-monotone. So, it would seem that we should be able to extend our results to general k-monotone density estimation using the results of Balabdaoui and Wellner (2007). The main theory for uncensored data is to find the knots τ_n (where the $k-1$th derivative changes slope), and fit separate $(k-1)$-degree splines between two successive knots. It can be shown that the estimates thus obtained are consistent and have a convergence rate of $n^{(k-1)/(2k+1)}$. A possible idea is to use the same algorithm, altering the estimator for the 2-monotone part with the k-monotone part. Hence, following the notation of Chapter 3, one possible algorithm may be as follows:

1. Fit a parametric least squares model to the data.

2. Obtain the estimates of the censored observations $(\hat{z_{k+1}}, \ldots \hat{z_n})$ using conditional expectation.

3. Obtain the non-parametric least squares estimator of the k-monotone density using $y_1, \ldots, y_k, \tilde{z}_{k+1}, \ldots \tilde{z}_n$ (using the least squares estimates characterized by Balabdaoui and Wellner (2007)).

4. Repeat steps 2-3 until the estimate converges.

5. Use $\widehat{g}(t)/\widehat{g}(n^{-1/5})$ to estimate $S(t)$, the survival function of the original density.

However, the asymptotic properties of the estimator should be carefully explored, because the rates of convergence calculations require different techniques.

Another issue that we have avoided is inference for the two-monotone density estimator. There are a number of problems with using the Chernoff distribution ($\mathbb{C}$) for inference. The first problem is that the density of $\mathbb{C}$ does not have a closed form (Groeneboom 1988). Also, the normalizing constant can be difficult to estimate. Also, it has been shown by Kosorok (2008a), that the nonparametric bootstrap is inconsistent for pointwise inference (i.e., inference for $f(t)$ at a given value of $t \in [0, 1]$). Hence, it is very difficult to obtain valid uniform confidence bands for the Grenander estimator. It appears as if establishing the uniform rate,which seems to be $n^{1/3}(log\ n)^{-1/3}$, is not too hard in comparison to establishing distributional convergence. Under some simplifying assumptions, it can be shown that the extremal limiting distribution may yield an extreme value distribution in the limit after standardization. Establishing this, however, seems to be difficult without results for convergence of empirical processes over non-compact index sets.

Finally, we can also look at the estimation of two-monotone densities under censoring in the presence of covariates. As far as we know, very little work has been published on this topic. Some work on regression has been done by Balabdaoui and Wellner (2007), but no one has looked at the results under censoring, which might be an interesting topic.

5.3 Competing Risks

Finally, we worked on recurrence time density estimation in the presence of competing risks. We developed an algorithm for estimation and showed its consistency and asymptotic properties.

Competing risks are said to be involved, if a patient may suffer from a number of different mutually exclusive risk factors, such as death from different causes. For competing risks, we obtain the cause specific hazard as follows:

We would like to extend the results that we obtained to include covariates in the model. The cause-specific hazard, $h_k(t \mid Z)$, is the instantaneous risk of dying from a particular cause k given that the subject is still alive at time t with covariates Z:

$$h_k(t \mid Z) = Lim_{\delta t \to 0}\{\frac{P(t \leq T \leq t + \delta t, K = k \mid T \geq t, Z = z)}{\delta t}\},$$

and

$$S_k(t \mid Z) = exp(-\int_0^t h_k(u \mid Z)\ du).$$

Cause specific hazards are modelled using the Cox PH model or fitting parametric models taking into account various time-dependent effects.

Another novel approach is to use subdistribution functions (Fine and Gray 1999). The subdistribution hazard, $g_k(t \mid Z)$, is the instantaneous risk of dying from a particular cause k given that the subject has not died from the cause k. Thus,

$$h_k(t \mid Z) = Lim_{\delta t \to 0}\{\frac{P(t \leq T \leq t + \delta t, K = k | T \geq t \text{ or } T \leq t, K \neq k \text{ and } Z = z)}{\delta t}\},$$

and the subsurvival function is given by

$$Q_k(t \mid Z) = \exp(-\int_0^t g_k(u \mid Z)\ du).$$

Our future aim will be to extend our current results to the Fine and Grey model for recurrence time data.

The issue of inference also remains open for this particular problem, as this is also based on the Chernoff distribution and the Grenander estimator. Under some simplifying assumptions, it can be shown that the extremal limiting distribution may yield an extreme value distribution in the limit after standardization. Establishing this, however, seems to be very difficult, as mentioned previously. It is also important to determine whether this can be accomplished without imposing assumptions so strong that the primacy of the Grenander is lost.

APPENDIX A: TECHNICAL DETAILS FOR CHAPTER 2

Here we give some results with their proofs for Chapter 2

Here is the proof of the first lemma. The satement of the lemma is :-

Consider the semiparamatric model $\mathcal{P} = \{P_g : g \in \mathcal{G}\}$, where the distribution P_g has density $p_g(u) = S_g / \int S_g$ and $\mathcal{G}$ is a collection of densities on $\mathcal{R}^+$. Let $\dot{\mathcal{G}}_g$ and $\dot{\mathcal{P}}_g$ be the tangent sets for the models $\mathcal{G}$ and $\mathcal{P}$ respectively at g. If A_g is the score operator mapping tangents in $\dot{\mathcal{G}}_g$ to $\dot{\mathcal{P}}_g$ then, $A_g\dot{\mathcal{G}}_g$ is dense in the maximal tangent set $L_2^0(S)$ for $\mathcal{P}$.

Proof. Let S be the survival function corresponding to g. Consider the following parametric path through g:

$$\eta \mapsto g_\eta = \frac{\psi(\eta a)g}{\int \psi(\eta a)g},$$

where $\psi : \mathcal{R} \mapsto \mathcal{R}^+$ is a bounded, continuously differentiable with bounded derivative ψ' satisfying $\psi(0) = \psi'(0) = 1$ and $a \in L_2^0(g)$. Note that $L_2^0(g)$ is the maximal nonparametric tangent set for $\mathcal{G}$ while $L_2^0(S)$ is the maximal tangent set for $\mathcal{P}$. The corresponding parametric submodel for p_g is

$$p_{g_\eta}(u) = \frac{\int_u^\infty \psi(\eta a)(v)g(v)dv}{\int_0^\infty \int_u^\infty \psi(\eta a)(v)g(v)dvdu}.$$

Thus the tangent set $\dot{\mathcal{P}}_g$ is given by the functions

$$\begin{aligned} A_g a(u) &= \frac{\int_u^\infty a(v)g(v)dv}{S(u)} - \int_0^\infty \frac{\int_u^\infty a(v)g(v)dv}{\int_0^\infty S(u)du} du \\ &\equiv b(u). \end{aligned}$$

Note that $b \in L_2^0(S)$, $b(\infty) = 0$ and $b(0) = \int_0^\infty va(v)g(v)dv$. Thus $\dot{\mathcal{G}}_g = \{a : a \in L_2^0(g),\ \int va(v)g(v)dv < \infty\}$ and $\dot{\mathcal{P}}_g = \{b : b \in L_2^0(S),\ b(0) < \infty,\ b(\infty) = 0\}$. Let A_g^*

be the adjoint (Bickel et al. (1993)) for A_g, then for $b \in \dot{\mathcal{P}}_g$ and $a \in \dot{\mathcal{G}}_g$ we must have

$$\langle A_g a,\ b \rangle_{L_2^0(S)} = \langle a,\ A_g^* b \rangle_{L_2^0(g)},$$

where $\langle \cdot, \cdot \rangle_{\mathcal{L}}$ denotes the inner product on the space $\mathcal{L}$ (Van der Vaart (1998), Bickel et al. (1993)). This follows from considering $\dot{\mathcal{G}}_g$ and $\dot{\mathcal{P}}_g$ as dual normed spaces. Now for $b \in L_2^0(S)$,

$$\begin{aligned} \langle A_g a,\ b \rangle_{L_2^0(S)} &= \int_0^\infty \int_u^\infty a(v) g(v) b(u) dv du \\ &= \int_0^\infty \int_0^v b(u) a(v) g(v) du dv. \end{aligned}$$

Thus, $A_g^* b(v) = \int_0^v b(u) du$ ($\in L_2^0(g)$). Note that $A_g^* b(v) = 0$ gives $b(v) = 0$. This implies that $N(A_g^*) = 0$, where the kernel $N(A_g^*) = \{b : A_g^* b = 0\}$. $N(A_g^*)$ is also the orthocomplement of the range set $R(A_g) = A_g \dot{\mathcal{G}}_g$ and thus $A_g \dot{\mathcal{G}}_g$ is dense in $L_2^0(S)$. Thus the closure of $A_g \dot{\mathcal{G}}_g$ is equal to $L_2^0(S)$. $\square$

We also look at the proof of the main theorem, whose statement is :-

Suppose that the covariate vector $\tilde{Z}$ is almost surely bounded. Then under (A1)–(A3) and with $\phi(u) = 1 - ug(u)/S(u)$ and

$$M(t) = I\{U(\theta) \leq t\} - \int_0^t I\{U(\theta) > s\} \lambda_{U(\theta)}(s) ds, \tag{5.1}$$

the ordinary score for θ at $\theta = \theta_0$ is

$$\dot{l}_{\theta_0} = z \int_0^{U^c(\theta_0)} R\phi(s) dM(s) - (z - E\tilde{Z}), \tag{5.2}$$

the tangent space $\dot{\mathcal{P}}_S$ for S is $\{\dot{l}_S b : b \in L_2^0(S)\}$ where the score operator $\dot{l}_S$ for S is given by

$$\dot{l}_S b = \int_0^{U^c(\theta_0)} Rb(s)dM(s), \tag{5.3}$$

the tangent space for h is $\{b : b \in L_2(h), \int b(z)e^{\theta_0' z}h(z)dz = 0\}$, and the efficient score for θ at $\theta = \theta_0$ is

$$\tilde{l}_{\theta,S} = \int_0^{U^c(\theta_0)} (z - E\{\tilde{Z}|U^c(\theta_0) \geq s\})R\phi(s)dM(s), \tag{5.4}$$

where for $a \in L_2^0(S)$,

$$Ra(t) = a(t) - \frac{\int_t^\infty a(u)S(u)du}{\int_t^\infty S(u)du}.$$

Proof. The likelihood for 1 observation $(U_i \wedge U_i^c, \delta_i, \tilde{Z}_i)$ is given by

$$l(\theta) = \left\{g_{U(\theta)}(U_i)\right\}^{\delta_i} \left\{\int_{U_i^c}^\infty g_{U(\theta)}(u)du\right\}^{1-\delta_i} h_{Z,\theta}(\tilde{Z}_i).$$

Taking log and differentiating with respect to θ we get the ordinary score for θ

$$\begin{aligned} \dot{l}_\theta &= \tilde{Z}i\left\{\delta_i\phi(e^{(\theta-\theta_0)'\tilde{Z}_i}U_i) + (1-\delta_i)E\left[\phi(e^{(\theta-\theta_0)'\tilde{Z}}U(\theta))|U(\theta) > U_i^c\right]\right\} \\ &\quad -(\tilde{Z}_i - E\tilde{Z}). \end{aligned}$$

The desired expression in (5.2) for the ordinary score function for θ can be derived by noting that the first term on the right hand side of the last equation is a Doob's martingale (Bickel et al. (1993)) and can be expressed in terms of the counting process martingale in (5.1) using proposition A.3.6 in Bickel et al. (1993).

From lemma 1 we conclude that the tangent space $\dot{\mathcal{P}}_S$ for S can be taken to be the maximal tangent space $L_2^0(S)$ and thus (5.3) follows from (5.2). In order to find $\Pi_0(\dot{l}_{\theta_0}|\dot{\mathcal{P}}_S) = \dot{l}_S b^*$ we find $b^* \in L_2^0(S)$ such that $\dot{l}_{\theta_0} - \dot{l}_S b^* \perp \dot{l}_S b \; \forall \; b \in L_2^0(S)$. That is $E\left\{\left(\dot{l}_{\theta_0} - \dot{l}_S b^*\right)\dot{l}_S b\right\} = 0$. Note that $\dot{l}_{\theta_0} - \dot{l}_S b^* = \int_{-\infty}^{U^c(\theta_0)}(\tilde{Z}R\phi - Rb^*)dM(s) - (\tilde{Z} - E\tilde{Z})$.

Conditioning on $\tilde{Z}$ and $U^c(\theta_0)$ and using the fact that $U(\theta_0)$ is distributed independently of $\tilde{Z}$ and $U^c(\theta_0)$ we get

$$
\begin{aligned}
& E\left\{\left(\dot{l}_{\theta_0} - \dot{l}_S b^*\right)\dot{l}_S b\right\} \\
&= EE\left\{(\dot{l}_{\theta_0} - \dot{l}_S b^*)\dot{l}_S b \Big| \tilde{Z}, U^c(\theta_0)\right\} \\
&= EE\left\{\int_0^{U^c(\theta_0)} (\tilde{Z}R\phi(s) - Rb^*(s))Rb(s)I\{U(\theta_0) \geq s\}\lambda_{U(\theta_0)}(s)ds \Big| \tilde{Z}, U^c(\theta_0)\right\} \\
&= E\left\{\int_0^{U^c(\theta_0)} (\tilde{Z}R\phi(s) - Rb^*(s))Rb(s)dF_{U(\theta_0)}(s)\right\} \\
&= \int \left\{E(\tilde{Z}I\{U^c(\theta_0) \geq s\})R\phi(s) - EI\{U^c(\theta_0) \geq s\}Rb^*(s)\right\} Rb(s)dF_{U(\theta_0)}(s).
\end{aligned}
$$

The second equality above is obtained by using the result that if $Y_i = \int f_i dM, i = 1, 2,$ then

$$
EY_1Y_2 = E\int f_1 f_2 d\langle M, M\rangle = E\int f_1(s)f_2(s)I\{U(\theta_0) \geq s\}d\Lambda(s).
$$

Thus $E\left\{\left(\dot{l}_{\theta_0} - \dot{l}_S b^*\right)\dot{l}_S b\right\} = 0$ for all $b \in L_2^0(S)$ if

$$
Rb^*(s) = \frac{E(\tilde{Z}I\{U^c(\theta_0) \geq s\})}{EI\{U^c(\theta_0) \geq s\}}R\phi(s) = E\{\tilde{Z}|\ U^c(\theta_0) \geq s\}R\phi(s).
$$

Thus the projection of $\dot{l}_\theta$ on $\dot{\mathcal{P}}_s$ is given by

$$
\Pi_0(\dot{l}_{\theta_0}|\dot{\mathcal{P}}_S) = \int_0^{U^c(\theta_0)} E\{\tilde{Z}|\ U^c(\theta_0) \geq s\}R\phi(s)dM(s). \tag{5.5}
$$

Now for finding $\dot{\mathcal{P}}_h$ for $h \in \mathcal{H}'$, we consider the parametric path $\eta \mapsto h_\eta = (1+\eta a)h$, where $a \in L_2^0(h)$. The score operator for h is given by

$$
\dot{l}_h a = a - \frac{\int e^{\theta_0' z}a(z)h(z)dz}{\int e^{\theta_0' z}h(z)dz} \equiv b(z),
$$

for $a \in L_2^0(h)$. Note that $b \in \left\{b : \int b(z)e^{\theta_0' z}h(z)dz = 0\right\}$. If h is unrestricted then the tangent space can be taken to be the ortho-complement of the linear span of $e^{\theta_0' z}$, i.e., $[e^{\theta_0' z}]^{\perp}$ in $L_2(h)$. Since $U(\theta_0)$ is distributed independently of Z and $U^c(\theta_0)$, $E_0\{b(\tilde{Z})\dot{l}_S b'\} = 0$ for any $b \in [e^{\theta_0' z}]^{\perp}$ and $\dot{l}_S b' \in \dot{\mathcal{P}}_S$, i.e., $\dot{\mathcal{P}}_S \perp \dot{\mathcal{P}}_h$. Since $(z - E\tilde{Z}) \in [e^{\theta_0' z}]^{\perp}$, we get

$$\Pi_0(\dot{l}_{\theta_0} |\ [e^{\theta_0' z}]^{\perp}) = -(z - E\tilde{Z}). \tag{5.6}$$

Now subtracting (5.5) and (5.6) from (5.2) we get from the efficient score (5.4). $\square$

Next we give an outline of the proof of 12. The satement of the lemma is

Under (C1), (C2), (R1)–(R4) and (K1) we have,

(a) $|\hat{g}_{n,\theta}(t) - g_{\theta}(t)| = O_p(n^{-1/2})O_p(h_n^{-1}) + O_p(h_n^2)$

(b) $|\hat{G}_{n,\theta}(t) - G_{\theta}(t)| = O_p(h_n^2)$

(c) $|\hat{g}'_{n,\theta}(t) - g'_{\theta}(t)| = O_p(n^{-1/2})O_p(h_n^{-2})$

(d) $|\hat{G}'_{n,\theta}(t) - G'_{\theta}(t)| = O_p(h_n^2)$

uniformly in t and $\theta \in \Theta$,

Proof. Note that $E_{\theta,\lambda}\hat{g}_{n,\theta}(t) = \frac{1}{h_n}\int K\left(h_n^{-1}(t-u)\right)dG_{0,\theta}(t)$ for all $t \in \Re$, where, $G_{0,\theta}(t) \equiv \int_{-\infty}^{t}\int f_0(v + (\theta - \theta_0)'z)\zeta_{\theta,z}(v)h(z)dzdv$. Thus the error term is given by

$$\begin{aligned}
\Big|\hat{g}_{n,\theta}(t) - E_{\theta,\lambda}\hat{g}_{n,\theta}(t)\Big| &= \Big|\int \frac{1}{h_n}K\left(\frac{t-u}{h_n}\right)\{d\mathbb{G}_{n,\theta}(t) - dG_{0,\theta}(t)\}\Big| \\
&= \Big|\int \frac{1}{h_n^2}K'\left(\frac{t-u}{h_n}\right)\{\mathbb{G}_{n,\theta}(t) - G_{0,\theta}(t)\}dt\Big| \\
&\le \Big\|G_{n,\theta} - G_{0,\theta}\Big\|_{\infty}\Big|\frac{1}{h_n}\int K'(v)dv\Big|.
\end{aligned}$$

By lemma 2 the first term in the preceding line is $O_p(n^{-1/2})$ and by condition (K1) the second term is $O(h_n^{-1})$. The bias term is given by

$$
\begin{aligned}
&\Big|E_{\theta,\lambda}\hat{g}_{n,\theta}(t) - g_\theta(t)\Big| \\
&= \Big|\int \frac{1}{h_n}K\left(\frac{t-u}{h_n}\right) g_\theta(u)du - g_\theta(t)\Big| \\
&= \Big|\int \frac{1}{h_n}K\left(\frac{t-u}{h_n}\right) \{g_\theta(t) + (t-u)g'_\theta(t) + \frac{1}{2}(t-u)^2 g''_\theta(t_u)\}du - g_\theta(t)\Big| \\
&\le \frac{1}{2}\|g''_\theta\|_\infty\ h_n^2 \int v^2 K(v)dv.
\end{aligned}
$$

By (R1) and (K1), the bias term is $O_p(h_n^2)$, thus proving (a).

Also,

$$
\hat{g}'_{n,\theta}(t) = -z\int \frac{1}{h_n}K'\left(\frac{t-u}{h_n}\right) dG_{n,\theta}(u) + \int \frac{1}{h_n^2}K'\left(\frac{t-u}{h_n}\right) dH_{n,\theta}(u),
$$

where $H_{n,\theta}(t) = n^{-1}\sum \delta Z I\{e_\theta \le t\}$. Let $H_{0,\theta}(t) \equiv E_{\theta,\lambda}H_{n,\theta}(t) = \int_{-\infty}^{t}\int z f_0(u + (\theta - \theta_0)'z)\zeta_{\theta,z}(t)h(z)dzdu$. Then the error term for $\hat{g}'_{n,\theta}$ is

$$
\begin{aligned}
\Big|\hat{g}'_{n,\theta}(t) - E_{\theta,\lambda}\hat{g}'_{n,\theta}(t)\Big| \quad &\le \quad |z|\Big|\int \frac{1}{h_n^2}K'\left(\frac{t-u}{h_n}\right)\{d\mathbb{G}_{n,\theta}(t) - dG_{0,\theta}(t)\}\Big| \\
&\quad + \Big|\int \frac{1}{h_n^2}K'\left(\frac{t-u}{h_n}\right)\{d\mathbb{H}_{n,\theta}(t) - dH_{0,\theta}(t)\}\Big|.
\end{aligned}
$$

Another use of Integration-by-parts, lemma 2 and condition (K1) in both the terms in the preceding statement yield

$$
\Big|\hat{g}'_{n,\theta}(t) - E_{\theta,\lambda}\hat{g}'_{n,\theta}(t)\Big| = O_p(n^{-1/2})O_p(h_n^{-2}).
$$

The bias term for $\hat{g}'_{n,\theta}$ after Integration-by-parts is

$$
\Big|E_{\theta,\lambda}\hat{g}'_{n,\theta}(t) - g'_\theta(t)\Big| \quad \le \quad |z|\Big|\int \frac{1}{h_n}K\left(\frac{t-u}{h_n}\right) g_\theta^{(1)}(u)du - g_\theta^{(1)}(t)\Big|
$$

$$+\left| \int \frac{1}{h_n} K\left(\frac{t-u}{h_n}\right) h_\theta^{(1)}(u)du - g_\theta^{(1)}(t)\right|,$$

where $h_\theta(t) \equiv \int z f_0(t + (\theta - \theta_0)'z)\zeta_{\theta,z}(t)h(z)dz$, $g_\theta^{(1)}(t) \equiv \partial\theta g_\theta(t)/\partial t$ and $h_\theta^{(1)}(t) \equiv \partial\theta h_\theta(t)/\partial t$. After another Taylor's expansion, application of lemma 3.4.1, and by conditions (R1) and (A1), we obtain

$$\left| E_{\theta,\lambda}\hat{g}'_{n,\theta}(t) - g'_\theta(t)\right| = O_p(h_n^2),$$

which proves part (c). Parts (b) and (d) can be proved similarly. $\square$

APPENDIX B: TECHNICAL DETAILS FOR CHAPTER 3

Here we give you some results along with their proofs for Chapter 3

The first theorem we look at is 22. The statement of the theorem is :-

Suppose that $y_1, \ldots, y_k$ are the uncensored observations and $z_{k+1}, \ldots, z_n$ are the censored observations from iid random variables with density $f_0 \in \mathcal{C}$. Then, the least squares estimate is uniformly consistent on closed intervals bounded away from 0 ie for each c>0, we have wp 1,

$sup_{c \leq x < \infty} |\tilde{g}_n(x) - f_0(x)| \to 0.$

Proof. The key step in the proof is showing that $\tilde{g}_n$ is uniformly bounded and then follows from the proof by Groeneboom. We let τ_{nk} denote the set of locations of change of slope of $H_{nk}^{''}$ as defined in Section 2.6. First assume that $f_0(0) < \infty$. Fix $\delta > 0$, such that $[0, \delta]$ is contained in the interior of the support of f_0, and let $\tau_{nk,1}$ e be the last point of change of slope in $(0, \delta]$, or zero if there is no such point. Also, the last point of change of slope is to the right of $X_{(n)}$, we may assume that there exists a point of change of slope $\tau_{nk,2}$ strictly to the right of δ. Now, using the fact that $\tilde{g}_n(\tau_n) < \tilde{g}_n(\delta/2) < i/\delta$, we will follow the proof in Groeneboom to show that $\tilde{g}_n(\tau_{n,1})$ is uniformly bounded. So, we can use a Helley argument to show that it has a convergent subsequence and hence by Theorem 3.1 of Groeneboom, the convergence is to f_0.

□

Next, we look at 25. The statement of the theorem is :

Under suitable conditions, the least squares estimates obtained asymptotically converge to the following distribution:

$$n^{2/5} c_1(f_0, \theta)(\tilde{g_n}(x_0) - f_0(x_0)) \to_d H^{''}(0) \text{ and}$$

$$n^{1/5} c_2(f_0, \theta)(\tilde{g_n^{'}}(x_0) - f_0^{'}(x_0)) \to_d H^{(3)}(0),$$

where $c_1(f_0,\theta) = (\frac{24}{v(\theta)f_0^2(x_0)f_0^{''}(x_0)})^{1/5}$ and $c_2(f_0,\theta) = (\frac{24^3}{v^{1/2}(\theta)f_0(x_0)f_0^{''}(x_0)^3})^{1/5}$

where H is same as the one defined in section 3.1.

Proof. Define the local Y_n-process by

$$\begin{aligned}
\tilde{Y}_n^{loc}(t) &= n^{4/5}\int_{x_0}^{x_0+n^{-1/5}t}\{\mathbb{G}_n(\nu)-\mathbb{G}_n(x_0)-\int_{x_0}^{\nu}(f_0(x_0+(u-x_0)f_0^{'}(x_0))du\}d\nu\\
&= n^{4/5}\int_{x_0}^{x_0+n^{-1/5}t}\{\mathbb{G}_n(\nu)-\mathbb{G}_n(x_0)-(F_0(\nu)-F_0(x_0))\}d\nu\\
&\quad +n^{4/5}\int_{x_0}^{x_0+n^{-1/5}t}\tfrac{1}{6}f_0^{''}(x_0)(\nu-x_0)^3d\nu\ +o(1)\\
&=_d n^{3/10}\int_{x_0}^{x_0+n^{-1/5}t}\{\mathbb{U}_n(F_0(\nu))-\mathbb{U}_n(F_0(x_0))\}d\nu+\tfrac{1}{24}f_0^{''}(x_0)t^4+o(1)\\
&\Rightarrow \sqrt{v^{1/2}(\theta)f_0(x_0)}\int_0^t W(s)\ ds\ +\tfrac{1}{24}f_0^{''}(x_0)t^4
\end{aligned}$$

where $v(\theta)$ is the variance on $\mathbb{G}_n$ under censoring percentage θ. Plugging in this value in theorem 6.1 of Groeneboom, we get the desired result.

□

APPENDIX C: TECHNICAL DETAILS FOR CHAPTER 4

Here we give some theorems along with their proofs for Chapter 4

We begin with the proof of Theorem 31:

The rate of convergence of $\tilde{f}_{jn}$ to f^j is $n^{1/3}$, for $j = 1, \ldots, K$.

Proof. The function $\hat{s}_n$ described above is a sort of inverse of the function $\tilde{f}_{jn}$ in the sense that $\tilde{f}_{jn} \leq a$ if and only if $\hat{s}_n(a) \leq t$ for every $t \geq 0$ and $a > 0$. Hence,

$$P(n^{1/3}(\tilde{f}_{jn}(t) - f^j(t)) \leq x) = P(\hat{s}_n(f^j(t) + xn^{-1/3} \leq t),$$

and the desired rate and weak convergence result can be deduced from the argmax values of $x \to \hat{s}_n(f^j(t) + xn^{-1/3})$. Applying the change of variable $s \to t + g$ in the definition of $\hat{s}_n$, we obtain

$$\hat{s}_n(f^j(t) + xn^{-1/3}) - t = argmax_{g>-t}\{\tilde{\mathbb{F}}_{jn}(t+g) - (f^j(t) + xn^{-1/3})(t+g)\}.$$

Thus, the probability on the LHS is precisely $P(\hat{g}_n \leq 0)$, where $\hat{g}_n$ is the argmax above. Now, by the previous argmax expression combined with the fact that the location of the maximum of a function does not change when the function is shifted vertically, we have $\hat{g}_n \equiv argmax\{g > -t\}\{M_n(g) \equiv \tilde{\mathbb{F}}_{jn}(t+g) - \tilde{\mathbb{F}}_{jn}(t) - f^j(t)g - xgn^{-1/3}\}$.

Now, $\hat{g}_n = O_P(1)$ and $M_n(g) \to M(g) \equiv F^j(t+g) - F^j(t) - f^j(t)g$ uniformly on compact sets. So, we have, $\hat{g}_n = o_P(1)$. We now utilize Theorem 14.4 of Kosorok (2008b) to obtain the rate for $\hat{g}_n$, under the L_1 norm. Since $M_n(0) = M(0) = 0$, we obtain that $M(g) \leq -g^2$, and by using Theorem 11.2 of Kosorok (2008b), we obtain

$$E^* sup_{|g|<\delta}\sqrt{n} \mid M_n(g) - M(g) \mid \leq \phi_n(\delta) \equiv \delta^{1/2} + \sqrt{n}\delta n^{-1/3}.$$

Clearly, $\phi_n(\delta)/\delta^\alpha$ is decreasing in δ for $\alpha = 3/2$. Since $n^{2/3}\phi_n(n^{-1/3}) = n^{1/2} + n^{1/6}n{-}1/3 = O(n^{1/2})$, Theorem 14.4 of Kosorok (2008b) yields $n^{1/3}\hat{g}_n = O_P(1)$.

□

The next theorem we look at is theorem 32: Suppose that F^j is as described above and is also continuous. Then, we can conclude that for any t bounded away from 0,

$$n^{1/3} \mid f^j(t)f^{j(1)}(t) \mid (\tilde{f}_{jn}(t) - f^j(t)) = 2\mathbb{C},$$

where $\mathbb{C}$ denotes the Chernoff distribution and $f^{j(1)}$ is the derivative of f^j.

Proof. Let $\hat{h}_n = n^{1/3}\hat{g}_n$. Since the maximum of a function does not change when the function is multiplied by a constant, we have that $\hat{h}_n$ is the argmax of the process

$$\begin{aligned} h \to n^{2/3}M_n(n{-}1/3h) = n^{2/3}(\mathbb{P}_n - P)(1\{X \le t + hn^{-1/3}, cause = j\} \\ -1\{X \le t, cause = j\}) + n^{2/3}[F^j(t + hn^{-1/3}) - F^j(t) - f^j(t)hn{-}1/3] - xh. \end{aligned} \tag{5.7}$$

Let $0 < V < \infty$ and applying Theorem 11.20 of Kosorok (2008b) to the sequence of classes $\mathcal{F}_n^j = \{n^{1/6}1\{X \le t + hn^{-1/3}, cause = j\} - 1\{X \le t, cause = j\} : -V \le h \le V\}$ with envelope sequence $F_n^j = n^{1/6}1t - Vn^{-1/3} \le X \le t + Vn^{-1/3}$, to obtain that the process on the right side converges in $l^\infty(-V, V)$ to $h \to \mathbb{H}(h) \equiv \sqrt{f^j(t)}\mathbb{Z}(h) + f^{j(1)}(t)h^2 - xh$, where $\mathbb{Z}$ is a two-sided Brownian motion originating at zero (two independent Brownian motions starting at zero, one going to the right of zero and the other going to the left). From the proof of 31, we obtain that $\hat{h}_n = O_P(1)$. Also, since $\mathbb{H}$ is continuous with a unique maximum, the argmax theorem now yields by the arbitrariness of V that $\hat{h}_n \to \hat{h}$, where $\hat{h} = argmax\ \mathbb{H}$. $\hat{h}$ can be further simplified to

$$\mid \frac{4f^j(t)}{f^{j(1)}(t)^2} \mid^{1/3} argmax\{\mathbb{Z}(h) - h^2\} + \frac{x}{f^{j(1)}(t)}$$

Now,

$$P(\mid \frac{4f^j(t)}{f^{j(1)}(t)^2} \mid^{1/3} argmax\{\mathbb{Z}(h) - h^2\} + \frac{x}{f^{j(1)}(t)} \leq 0)$$

$$= P(4 \mid f^j(t) f^{j(1)}(t) \mid^{1/3} argmax\{\mathbb{Z} - h^2\} \leq x)$$

which yields

$$n^{1/3} \mid f^j(t) f^{j(1)}(t) \mid (\tilde{f}_{jn}(t) - f^j(t)) = 2\mathbb{C}.$$

$\square$

REFERENCES

Aalen, O. (1976), "Nonparametric Estimation in connection with multiple decrement models," *Scandinavian Journal of Statistics*, 3, 15–27.

Aalen, O. and Johansen, S. (1978), "An empirical transition matrix for nonhomogenous Markov chain based on censored observations," *Scandinavian Journal of Statistics*, 5, 141–150.

Anevski, D. (1994), "Estimating the derivative of a convex density," *Technical Report 1994:8, Dept.Mathematical Statistics, Univ. Lund.*

Asgharian, M., M'lan, C. E., and Wolfson, D. B. (2002), "Length-biased sampling with right censoring: an unconditional approach," *J. Amer. Statist. Assoc.*, 97, 201–209.

Asgharian, M. and Wolfson, D. B. (2005), "Asymptotic behavior of the unconditional NPMLE of the length-biased survivor function from right censored pevelant cohort data," *Ann. Statis.*, 33, 2109–2131.

Balabdaoui, F. and Wellner, J. (2007), "Estimation of a k-monotone Density: Limit Distribution Theory and the Spline Connection," *Annals of Stats.*, 35, 25362564.

Bickel, P.J.and Klaassen, C., Ritov, Y., and Wellner, J. (1993), "Efficient and adaptive estimation for semiparametric models," *Springer-Verlag, New York.*

Brookmeyer, R. and Gail, M. H. (1987), "Biases in prevelant cohorts," *Biometrics*, 43, 739–749.

Brunk, H. D. (1970), *Estimation of isotonic regression. In Nonparametric Techniques in Statistical Inference*, Cambridge Univ. Press.

Cosslett, S. R. (2004), "Efficient semiparametric estimation of censored and truncated regression via a smoothed self-consistency equation," *Econometrica*, 72, 1277–1293.

Cox, D. R. (1969), """Some sampling problems in technology" in new developments in survey sampling," *Johnson and Smith. Wiley, New York.*

— (1972), "Regression models and life tables (with discussion)," *J. R. Statist. Soc.*, B 34, 187–220.

Dauxois, J., Guilloux, A., and Kirmani, S. (2014), "Estimation in a Competing Risks Proportional Hazards Model Under Length-biased Sampling With Censoring," *Lifetime Data Analysis, Springer Verlag (Germany)*, 20.

Denbi, L. and Vardi, Y. (1986), "The survival curve with decreasing density," *Technometrics*, 28, 359–367.

Efron, B. (1967), "The two sample problem with censored data," *in proceedings of the fifth Berkeley symposium on mathematical statistics and probability,4,Berkeley: University of California press.*

Fine, J. and Gray, R. (1999), "A Proportional Hazards Model for the Subdistribution of a Competing Risk," *Jour. of Amer. Stat. Assoc.*, 94, 496–509.

Fygenson, M. and Ritov, Y. (1994), "Monotone estimating equations for censored data," *Annals of Statistics*, 22, 732–746.

Grenander, U. (1956), "On the theory of mortality measurements, Part II," *Skand. Akt.*, 39, 123–153.

Groeneboom, P. (1985), "Estimating a monotone density," *Proceedings of the Berkeley Conference in Honor of Jerzy Neyman and Jack Kiefer (L. M. Le Cam and R. A. Olshen, eds.)*, 2, 539–554.

— (1988), "Brownian motion with a parabolic drift and Airy functions." *Probab.Theory Related Fields*, 81, 79–109.

Groeneboom, P., Jongbloed, G., and Wellner, J. A. (2001), "Estimation of a Convex Function: Characterizations and Asymptotic Theory," *Annals of Stat.*, 1653–1698.

Hampel, F. R. (1987), "Design, modelling and analysis of some biological datasets," *Design, Data and Analysis, by Some Friends of Cuthbert Daniel (C. L. Mallows, ed.)*, 111–115.

James, I. and Buckley, J. (1979), "Linear regression with censored data," *Biometrika*, 66, 429–436.

Jeong, J. and Fine, J. (2006), "Direct parametric inference for the cumulative incidence function," *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 55, 187200.

Jin, Z., Lin, D. Y., Wei, L. J., and Z., Y. (2003), "Rank-based inference for the accelerated failure time model," *Biometrika*, 90, 341–353.

Keiding, N., Kvist, K., Hartvig, H., Tvede, M., and Juul, S. (2002), "Estimating time to pregnancy from current durations in a cross-sectional sample," *Biostatistics*, 3, 565–578.

Klaasen, C. A. J., Mokveld, P. J., and Van Es, B. (2004), "Efficient estimation in the accelerated failure time model under cross-sectional sampling," *Math. Preprint Ser. 03-06, University of Amsterdam.*

Klaassen, C. A. J. (1989), "Efficient estimation in the Cox model for survival data," *Proc. fourth Prague symp. Asympt. Statist.*, 313–319.

Kosorok, M. (2008a), "Bootstrapping the Grenander estimator," *IMS Collections*, 1, 282–292.

Kosorok, M. R. (2008b), *Introduction to empirical processes and semiparametric inference*, Springer.

Koul, H., Susurla, V., and Van Ryzin, J. (1981), "Regression analysis with randomly right-censored data," *Ann. Statis.*, 9, 1276–1288.

Miller, R. G. (1976), "Least squares regression with censored data," *Biometrika*, 63, 449–464.

Murphy, S. A. and Van der Vaart, A. W. (2000), "On profile likelihood," *J. Amer. Statist. Assoc.*, 95, 449–485.

Prakasa Rao, B. L. S. (1969), "Estimation of a unimodal density." *Sankhyd Ser. A*, 31, 23–36.

Ritov, Y. (1990), "Estimation in a linear regression model with censored data," *Ann. Statis.*, 18, 303–328.

Severini, T. A. and Wong, W. H. (1992), "Profile likelihood and conditionally parametric models," *Ann. Statis.*, 20, 1768–1802.

Stein, C. (1956), "Efficient nonparametric testing and estimation," *Proc. third Berkeley symp. math. statist. probab.*, 1, 187–196.

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

— (1990), "Estimating regression parameters using linear rank tests for censored data," *Ann. Statis.*, 18, 354–372.

Van der Vaart, A. W. (1998), "Asymptotic Statistics," *Cambridge, New York.*

van der Vaart, A. W. and Wellner, J. A. (1996), *Weak convergence and Empirical processes: With applications in statistics*, Springer-Verlag, New York.

Van Es, B., Klaasen, C. A. J., and Oudshoorn, K. (2000), "Survival analysis under cross section sampling: length bias and multiplicative censoring," *J. Statist. Plan. Infer.*, 91, 295–312.

Vardi, Y. (1982), "Nonparametric estimation in the presence of length bias," *Ann. Statis.*, 10, 616–620.

Wang, M. (1996), "Hazards regression analysis for length-biased data," *Biometrika*, 83, 343–354.

Wang, M., Jewell, N. P., and Tsai, W.-Y. (1986), "Asymptotic properties of the product limit estimate under random truncation," *Ann. Statist.*, 14, 1597–1605.

Wang, M.-C. (1991), "Nonparametric estimation from cross-secional survival data," *J. Amer. Statist. Assoc.*, 86, 130–143.

Wang, M.-C., R., B., and Jewell, N. P. (1993), "Statistical models for prevalent cohort data," *Biometrics*, 49, 1–11.

Wei, L. J., Ying, Z. L., and Lin, D. Y. (1990), "Linear regression analysis of censored survival data based on rank tests," *Biometrika*, 77, 845–851.

Woodroofe, M. and Sun, J. (1993), "A penalized maximum likelihood estimate of $f(0+)$ when f is non-increasing," *Stat. Sinica*, 3, 501–515.

Wright, F. T. (1981), "The asymptotic behavior of monotone regression estimates," *Ann. Statist.*, 9, 443–448.

Yamaguchi, K. (2003), "Accelerated failure time mover stayer regression models for the analysis of last episode data," *Sociol. Method.*, 33, 81–110.

Ying, Z. (1993), "a large sample study of rank estimation for censored regression data," *Ann. Statist.*, 21, 76–99.

Zeng, D. and Lin, D. (2007), "Efficient Estimation for the Accelerated Failure Time Model," *Journal of the American Statistical Association*, 102, 1387–1396.