Inference on time-to-event distribution from retrospective data with imperfect recall

Sedigheh Mirzaei Salehabadi

Thesis submitted to the Indian Statistical Institute
in partial fulfillment of the requirements
for the award of the degree of
Doctor of Philosophy

March, 2015

Indian Statistical Institute
203, B.T. Road, Kolkata, India.

To my Mother,

brother and sisters,

and the lovely memories of
my father and my eldest brother

Acknowledgements

First and foremost, I praise Almighty Allah, the most merciful, for providing me the opportunity to step in one of the most beautiful areas in the world of science, Statistics. I am extremely lucky that I got the opportunity to pursue research in statistics in Indian Statistical Institute (ISI). My tentative steps on this path has become firmer with support from many people to whom I would like to express my sincerest gratitude.

This thesis in its current form, would not have been possible without the guidance, patience and continued support of my thesis advisor, Prof. Debasis Sengupta. I would like to sincerely thank him for his warm encouragement, thoughtful guidance, sound advice, excellent teaching, good company, and lots of good ideas. It has been an honor for me to work with such a caring and affectionate researcher like him.

The work included in this thesis is based on data collected through the project 'Physical growth, body composition and nutritional status of the Bengal school aged children, adolescents, and young adults of Calcutta, India: Effects of socioeconomic factors on secular trends', funded by the Neys Van Hoogstraten Foundation of the Netherlands and run at the Biological Anthropology Unit of ISI. I would like to express my deepest appreciation to Prof. Parasmani Dasgupta, the project leader, for making the data available for this thesis and also for his helpful discussion throughout my research work.

Prof. Probal Chaudhuri and Prof. Mausumi Bose, helped me to adjust to the academic environment in India, where I had been a newcomer. I wish to express my deepest appreciation and heartfelt thanks to these truly great teachers and wonderful persons. I would like to acknowledge them not only for being excellent teachers, but also for their constructive suggestions, criticism and valuable advice. Their constant support through all stages of my stay in India and excellent care will always remain appreciated.

Staying In India will always be an unforgettable memory since I had many wonderful experiences. Describing my visit of India is impossible to sum up in few lines. The best I can do is to borrow of few lines from a master of words.

> "This is indeed India; the land of dreams and romance, of fabulous wealth and fabulous poverty, of splendor and rags, of palaces and hovels, of famine and pestilence, of genii and giants and Aladdin lamps, of tigers and elephants, the cobra and the jungle, the country of a thousand nations and a hundred tongues, of a thousand religions and two million gods, cradle of the human race, birthplace of human speech, mother of history, grandmother of legend,

great-grandmother of tradition, whose yesterdays bear date with the mouldering antiquities of the rest of the nations, the one sole country under the sun that is endowed with an imperishable interest for alien prince and alien peasant, for lettered and ignorant, wise and fool, rich and poor, bond and free, the one land that all men desire to see, and having seen once, by even a glimpse, would not give that glimpse for the shows of all the rest of the globe combined."

–Mark Twain, Following the Equator, 1897.

I arrived in India a little apprehensive about my stay, but all my concerns vanished in due course of time with the assistance of some wonderful people. On the day of my arrival in India, I was warmly received by Dr. Farkhondeh Sajadi in Delhi airport, and this gave me the strength to stay in India. I wish to extend my deepest appreciation to her. I was warmly welcomed by Mr. Debasis De in Kolkata airport, and later on by Prof. Aditya Bagchi (Dean of Studies) and Prof. Goutam Mukherjee (professor-in-charge of Theoretical Statistics and Mathematics Division) who remained helpful throughout my stay at ISI. I convey my thanks to all of them.

I would like to express my thanks to Prof. Tapas Samanta, the Head of the Applied Statistics Unit for his support and kindness. I would like to express my genuine appreciation to Prof. Anup Dewanji for his excellent assistance, moral support and his numerous advice that were always helpful.

I am immensely indebted to other teachers who taught me in ISI, Kolkata, including Prof. Alok Goswami, Late Prof. Somesh Chandra Bagchi, Prof. Sumitra Purkayastha and Dr. Shreela Gangopadhyay. Special thanks are also due to Prof. B.V. Rao, with whom I did not take any course but was lucky to interact during my stay at ISI Kolkata. He has helped me on different occasions.

I would like to thank Prof. Ratan Dasgupta for his support and for providing a quick accommodation in ladies hostel. I want to express my sincere gratitude to Prof. Sarbani Palit and her family and Prof. Kumud R. Sarkar and his family. Many times, they did not let me to feel alone. Their moral support and genuine concern of my well being is highly appreciated.

My time in ISI was made enjoyable by many friends who became a part of my life. Though it is impossible to thank everyone individually, I would like to express a few words of gratitude to my best friends. With the help of Fouzia, Richa, Rituparna, Debasmita, Suchismita and Aparajita I could adjust to hostel life. I thank all of them. I am grateful to Anshu, Arunangshu, Ayan, Bipul, Bijaya, Debasish, Farhad, Ganesh, Jane, Koushik, Mannu, Mickel, Monika, Moumita, Mujtaba, Munmun, Oyedeji, Prosenjit, Purvasha, Rajat, Raju, Ritwik, Roshni, Roya, Sanjay, Shashank, Shereen, Souminda, Srimoyee, Sudipta, Suparna, Swagata, Trijit and some ISEC trainees for being such wonderful friends. Special thanks go to Swagata and her family; I am grateful for their help and support during my stay in India. My sincere appreciation is extended to Buddha, Dibyendu, Farkhondeh, Kaushik, Minerva, Neena, Palash, Prajamitra, Prasenjit, Radhe, Shashank and Subhajit not only for their friendship but also for the excellent discussions throughout my research life at ISI. The last three years of hostel life, became enjoyable with my friend, Qaiser. I am grateful for the time spent with her and for her excellent help and accompany. The last two years of my staying in India, became unforgettable with my friend Ejaz. Together, we have shared so many special moments, contributing to a common story. My genuine thanks go to him. I express special gratitude to my Iranian friend, Behzad who has been a faraway company for me during my stay away from home. Hereby, I would like to thank him again for everything. I would also like to thank my friend Negar and her family in Iran for her constant help. Although I have lived far from my family, communications with friends provided the right emotional atmosphere for me.

I would like to thank our Dean of Studies, Prof. P. Bandyopadhyay and director Prof. Bimal Kumar Roy for providing necessary facilities for my Ph.D. in the institute. I am thankful to all the staff members of ISI Kolkata, especially to the office staff of Applied Statistics Unit and Theoretical Statistics and Mathematics Unit, for their help and generosity. I take this opportunity to sincerely acknowledge the Indian Statistical Institute for the financial support it provided me. I was honored to be an ISI research fellow.

Last but by no means the least, I would like to pay high regards to my parents. They have given me their unambiguous support throughout as always, for which my mere expression of gratitude does not suffice. I whole-heartedly thank and appreciate them for their love, encouragement, help, tremendous patience and spiritual support in all aspects of my life. Many thanks are due to them for their faith in me and allowing me to pursue my dreams. Though I lost my lovely father in the middle of my research life, his memory and advices are always the best support for me. Words are not enough to express my deep sense of gratitude towards my beloved sisters and brothers, who have also been great friends. I want to express my gratitude to my eldest sister Masumeh, for filling the gap of my absence by being with my mother. I would like to thank her and her family. I would like to pay my deepest appreciation to my lovely sister Iran for understanding me so well and helping me so much. My heartfelt thanks go to her and her lovely family for their constant support. In spite of the irreparable loss of my beloved brother Ali at the last stage of my research work, I cannot forget his excellent support. His greatness and dedication always inspired me to stay in the right path. I am unable to find words to thank Fatemeh, my lovely sister. Leaving home and coming to India to study, would have been impossible without her whole-hearted support and help. She has provided assistance in numerous ways. Finally, I wish to express my deepest appreciation and heartfelt thanks to my brother and best ever friend, Mohammad, for his constant moral support. He and his family were a faraway company for me during my stay away from home. My deepest thanks go to them.

March, 2015

Sedigheh Mirzaei Salehabadi

List of abbreviations, notations and symbols

AFT	Accelerated failure time	4
$AMLE$	Approximate NPMLE	53
CHF	Cumulative hazard function	4
$CRLB$	Cramer-Rao Lower Bound	24
$F(T_i-)$	$\lim_{x\to 0^-} F(T+x)$	42
I_E	Indicator function of event E	12
MLE	Maximum likelihood estimator	6
MSE	Mean square error	24
$NPMLE$	Non-parametric maximum likelihood estimator	3
PL	Product limit	3
$Stdev$	Standard deviation	24

Contents

Chapter 1

Introduction

1.1 Preamble

Time-to-event data arises from measurements of time till the occurrence of an event of interest. Such data are common in the fields of biology, epidemiology, public health, medical research, economics and industry. The event of interest can be the death of a human being (Klein and Moeschberger, 2003), failure of a machine (Zhiguo et al., 2007), onset of menarche in adolescent and young adult females (Bergsten-Brucefors, 1976; Chumlea et al., 2003; Mirzaei, Sengupta and Das, 2015), onset (or relapse) of a disease (Klein and Moeschberger, 2003), dental development (Demirjian, Goldstien and Tanner, 1973; Eveleth and Tanner, 1990), breast development (Cameron, 2002; Aksglaede et al., 2009), beginning of a criminal career (Hosmer et.al., 2008), marriage or birth of the first child (Allison, 1982), end of a work career (LeClere, 2005), end of a strike (Hosmer, Lemeshow and May, 2008), discontinuation of breast-feeding (Clements et al., 1997), healing a wound (Nelson et al., 2004) and so on. The time-to-event can be measured in days, weeks, years, etc.

Data on time-to-event are variously described as duration data, survival data, lifetime data or failure time data, even though the event of interest is not necessarily failure or death. Models and methods for the collection and analysis of such data comprise the field of survival analysis.

While traditional parametric and nonparametric methods of inference are sometimes used in the analysis of survival data, special methods are often needed because of the pattern of incompleteness in such data. Incomplete observations contain only partial information about the random variable of interest, i.e., we do not know the exact times-to-event in all the cases. Some typical forms of incompleteness in survival data are truncation, grouping and censoring. Truncation occurs when certain individuals are screened out of the study in such a way that individual instances of screening are not observable even though the screening criterion is known. Grouping happens when one is able to observe the number of events occurring in certain specified time intervals, rather than the exact times of those occurrences. Censoring occurs when one has knowledge of either the actual time of occurrence or an interval containing it.

In many clinical and epidemiological studies, a subject may be observed only up to a certain time, and there may be no follow up after that. In some cases, a subject may be observed after the event of interest has already taken place. In case the event is not known to have taken place, one would know that it has happened after the date of the last observation. In case the event is found to have taken place by the date of first observation, one would know that it has happened at an earlier time. These are instances of censoring from the right and from the left, respectively. When a subject is not continuously monitored, it is also possible that one would only know an interval of time, when the event of interest has taken place. This is called interval censoring. Left-, right- or interval-censored data can occur along with uncensored (complete) data also.

When some data are censored, it makes conceptual sense to work with an unobserved or notional time-to-event, for the purpose of modeling. Thus, for both censored and uncensored data, one can invoke an underlying time-to-event, which may be regarded as a sample from a distribution. In survival analysis, it is common to describe this distribution through the corresponding survival function or the hazard rate. The survival function at a given time describes the probability of the subject surviving (or not experiencing the event) up to that time. The hazard function at a certain time gives the probability per unit time of that event occurring in a short interval immediately afterwards, given that the event has not happened up to that specified time. Once the survival function or the hazard rate has been specified, one can derive from it other quantities of interest (e.g., median survival time, mean remaining life at a given time, cumulative hazard function).

There are many parametric models for the probability distribution of the time-to-event. Popular ones include the exponential, Weibull, lognormal, gamma, Gompertz, log-logistic, Pareto, generalized gamma and so on. Once a parametric model for the time-to-event has been chosen, standard techniques for parametric inference become applicable. However, one has to judiciously use the information from both censored and uncensored observations. As an extension of likelihood based methods for complete data, there have been similar methods for various types of censored (or otherwise incomplete) data. Usual large-sample properties of many likelihood based techniques have been shown to hold for incomplete-data likelihood also, under appropriate conditions (Lawless, 1982). Since the validity of the assumed parametric model needs empirical support, goodness-of-fit tests for such parametric models are important. Modifications of these tests for censored data have also been proposed (Lawless, 2003, Chapter 10). One can find a good summary of these methods in Lee and Wang (2003) and Kalbfleisch and Prentice (2002).

While parametric models may be unavoidable for short data, nonparametric methods have been used to estimate the duration distribution when the sample size is not very small. The traditional life table method (Lawless, 2003) for computing conditional survival probabilities is a nonparametric technique for grouped data. Using this estimator as a starting point, Kaplan and Meier (1958) developed the product limit (PL) estimator of survival function for randomly right censored data and showed that this estimator is, in fact, the nonparametric maximum likelihood estimator (NPMLE) of that function for the said type of data. Since then, a wealth of nonparametric methods have enriched the field of survival analysis. Since the exact analysis of performance of these inferential methods is often difficult, large sample asymptotic properties are studied. The presence of censoring is a challenge to analysis of this kind. Formal asymptotic analysis of many of the nonparametric methods received a boost with the advent of the counting process theory involving martingales, and the Nelson-Aalen estimator of the cumulative hazard function (CHF) of a distribution (Aalen, 1975). In the case of interval censoring, Turnbull (1976) proposed an NPMLE and a computational algorithm for the same, by using Efron's idea of self-consistency (Efron, 1967). Good summaries of these methods are available in Miller (1981), Klein and Moeschberger (2003) and Lawless (2003).

Until this point we have only discussed duration data arising from a population presumed to be homogeneous . Many practical situations involve heterogeneous populations, and it is important to consider the relation of time-to-event to factors causing that heterogeneity. The effect of covariates on time-to-event distribution has been a matter of much interest. A general way to incorporate covariates is through regression models, in which the dependence of time-to-event on concomitant variables is explicitly recognized. The relative risk regression model proposed by Cox (1972) has become the workhorse of regression analysis for censored data. It is one of the models that deal with covariates parametrically, while keeping a nonparametric flavour as far as the baseline distribution is concerned. In this sense, the Cox regression model is a semi-parametric model. It makes fewer assumptions than a completely parametric model, but more assumptions than a model that would ignore any linkage between distributions underlying different homogeneous groups. The method of analysis based on partial likelihood, proposed by Cox (1972), can accommodate right-censoring, which is common in survival data, and left-truncation, which arises when there are delayed entries in a cohort (Breslow et al., 1983). Other regression models for survival data with covariates include the accelerated failure time (AFT) model (Wei, 1992), the additive hazard regression model (Klein and Moeschberger, 2003), the proportional odds ratio model (Dabrowska and Doksum, 1988), The non-proportional hazard model (Vonta, 1996) and so on. A summary of these methods is available in Hosmer, Lemeshow and May (2008). There has also been some work on more general regression models for survival data, such as single index regression models (Chaudhuri, 2007), various generalizations of the Cox model (Bagdonavicius and Nikulin, 2003), models with random effect/frailty (Wienke, 2010) and accelerated intensity frailty model (Liu, Lu and Zhang, 2014). Methods of analysis for these models are generally developed under the assumption that the available data are either complete or randomly censored from the right. Finkelstein, Moore and Schoenfeld (1993) discussed methods for fitting a discrete proportional hazards model for the case where the data are interval-censored or right-truncated. Alioum and Commenges (1996) discussed a method for fitting a proportional hazards model for interval censored data based on Turnbull's estimator.

1.2 Observational studies on time-to-event

In order to estimate the time-to-event distribution, one would ideally like to observe a number of individuals continuously or periodically until the occurrence of the landmark event (Korn et al.,1997; McKay et al.,1998). In a longitudinal study, exposure status of subjects are recorded at multiple follow-up visits. This can partially alleviate any distortion created by faded memory. If, at every inspection, one only keeps record of whether the event of interest has happened, then the resulting observation amounts to interval censored data. Longitudinal studies also provide one the opportunity to observe individual patterns of change in quantitative variables. However, researchers often opt for cross-sectional studies in order to save time and cost.

Cross-sectional studies can produce dichotomous data on the current status of an individual (whether or not the landmark event has occurred till the day of observation), which is sometimes described as status quo data (Teilmann et al., 2009). A binary data regression model, such as the probit or the logistic model with time as the covariate, is often used for estimating the probability distribution function (Hediger and Stine, 1987; Nakamura, 1991; Betensky, 2000; Dunson and Dinse, 2002; Cook and Lawless, 2007). It is also possible to estimate the distribution non-parametrically, by regarding the current status data as either left or right censored observations, depending on whether or not the event has occurred at the time of observation. Left and right censored data are both special cases of interval censored data (Keiding et al., 1996).

In some cross-sectional studies, a subject is asked to recall the time of the landmark event, in case it has already taken place. Such retrospective data are usually incomplete (Roberts, 1994; Padez, 2003). In many cases (e.g., when the event has not happened or the subject cannot recall when it had happened) one can specify only a range for the requisite time. Thus, data arising from retrospective studies are also interval-censored.

It may be noted that current status data is the simplest form of interval censored data. Here, the time-to-event is either left-censored or right-censored, and there is no case with complete data or censoring from both sides. This special type of censoring is sometimes referred to as Case I interval censoring. Data arising from longitudinal studies with multiple inspection times give rise to censoring from either side or both sides, but there is no case of complete data. This type of censoring is referred to as Case II interval censoring. Recall based data with possibility of partial or no-recall, arising from retrospective cross-sectional studies lead to the most general form of interval censored data, including possible instances of complete data and data censored from left, right or both sides. This type of censoring is called mixed interval censoring (Sun, 2006, Chapter 2).

While drawing inference from interval censored data, one typically assumes the censoring to be non-informative. This means that there is a notional non-observation window that is independent of the event being observed. In this situation, one can use a likelihood that is adjusted for this type of censoring. This likelihood would lead to the maximum likelihood estimator (MLE) under a parametric model or the nonparametric maximum likelihood estimator (NPMLE) proposed by Turnbull (1976).

When current status data are obtained with observation times independent of the corresponding time-to-event, the resultant interval censoring may be regarded as non-informative. Thus, the general form of the likelihood for such data can be used, together with the corresponding parametric estimators or the nonparametric (Turnbull) estimator. In the case of recall data however, the non-observation window is likely to depend on the time of occurrence of the event. Consequently, the underlying censoring mechanism in this set-up is likely to depend on the time-to-event, thereby making the censoring informative. This is because of the fact that memory generally fades with time. As an example, for two post-menrcheal subjects interviewed at the same age, the one with more recent onset of menarche is more likely to remember the date. Therefore, parametric methods based on the likelihood for (non-informatively) interval censored data would not be applicable to incomplete data arising from recall inadequacy. The Turnbull estimator is also not meant for informatively censored data.

This difficulty is a significant one. On the one hand, we have current status data that is suitable for application of general methods for interval censored data, but suffers from shortage of information because of coarse grouping. On the other hand, we have retrospective data, which is more informative, but is not amenable to the application of those general methods.

There have been several approaches to handle informative censoring for various types of data, and the models and methods proposed there are specific to the emergent mechanism of censoring. Finkelstein et al. (2002) and Kaciroti et al. (2012) attempted multivariate modeling of the duration of interest and the censoring/inspection times. Scharfstein and Robins (2002) considered right censored data while inducing dependence between the time-to-event and the censoring time through prognostic factors/covariates. Tanaka and Rao (2005) also considered informative right-censoring where these two times were combined in a competing risk set-up. The commonality in all these models is that censoring is assumed to occur through duration variables that have the same origin of measurements as that of the duration of interest.

In respect of retrospective data, the above models are not suited for dealing with censoring that occurs because of the inability of the subject to recall the date of occurrence. This is because of the fact that the question of remembering or forgetting is relevant only after the event of interest has taken place, and generally has no meaning before that event (and in particular, at the origin of time measurements). Therefore, notional times of censoring measured from a common origin of durations would not be meaningful in this situation.

Thus, a different type of model is needed for retrospective data on time-to-event. There is also a need for developing new methodology to estimate the time-to-event distribution, parametrically and/or nonparametrically, which will be able to harness the incremental information contained in recall data as compared to status data.

1.3 Matters investigated in this thesis

In this thesis, we propose a new model for time-to-event data with informative censoring arising from a recall based retrospective study, under the assumption that a subject, who has been interviewed after the occurrence of the event of interest, either recalls the date perfectly or cannot recall it all.

In Chapter 2, we introduce the model for retrospectively collected time-to-event data. Under this model, the time of observation is assumed to be independent of the time-to-event, and the recall probability is regarded as a function of the time gap between the event and the observation. We show how one can perform likelihood based parametric inference on the basis of this model and study properties of the maximum likelihood estimator (MLE).

In Chapter 3, we derive the nonparametric maximum likelihood estimator of the survival function of the time-to-event under the model mentioned above. We examine computational as well as asymptotic issues. A computationally simpler estimator is also studied.

In Chapter 4, we consider regression under Cox's model for recall data with covariates, arising from retrospective studies. We develop a likelihood based estimator of the regression coefficients as well as the baseline survival function, under the model assumed in Chapter 2.

In Chapter 5, we summarize the main contributions of this thesis, and indicate a few directions of possible future work.

The methods developed here have been motivated by a recent anthropometric study conducted by the Biological Anthropology Unit of the Indian Statistical Institute in and around the city of Kolkata from 2005 to 2011 (ISI, 2012, p.108). In this retrospective data set, over four thousand randomly selected individuals, aged between 7 and 21 years, were surveyed. The subjects were interviewed on or around their birthdays. The data set on female subjects contains age, menarcheal status, age at menarche (if recalled), and some other information. For this data set, the landmark event is the onset of menarche, which is sometimes recalled as an exact date, sometimes recalled as being within a range of possible dates and sometimes not recalled at all. In order to minimize errors in recall, which has been recognized as a problematic issue with recall data (Rabe-Hesketh, Yang and Pickles, 2001; Wen and Chen, 2014), we regarded a date as not recalled at all even when a range of possible dates was recalled. In this sense, out of the total of 2195 female subjects interviewed, 977 individuals had menarche but 'could not recall' the date of onset, 443 individuals had menarche and recalled the exact date of its onset, while 775 individuals did not have menarche till the age at interview. The methods developed in Chapters 2-4 have been applied to this data set to demonstrate their usage.

The contents of Chapters 2-4 are based on Mirzaei, Sengupta and Das (2015), Mirzaei and Sengupta (2015a) and Mirzaei and Sengupta (2015b) respectively.

Chapter 2

Parametric estimation of time-to-event distribution

2.1 Introduction

In some retrospective studies, a subject is asked to recall the time of occurrence of a landmark event, in case it has already occurred. This kind of data can be incomplete, as the event may not have happened till the time of observation, or the time of its occurrence may not be recalled. In the latter case, the possible range of dates may depend on the time of occurrence of the event as discussed in Chapter 1, and consequently the underlying censoring mechanism may be informative. In this chapter, we develop a new model suitable for this specific type of informative interval censoring, and develop a parametric approach for estimating the distribution of the time-to-event. We demonstrate that the new approach produces more precise estimates than what can be achieved through current status data, and avoids the problems of bias and inconsistency that are encountered when one uses recall data without recognizing the informativeness of the censoring.

The standard large sample results for the properties of a maximum likelihood estimator (MLE) based on complete data does not automatically extend to the case of censored data. Different adjustments are needed for different types of censoring. The results applicable to random right censoring are available in Kalbfleisch and Prentice (2002, Chapter 3). For large sample properties of parametric MLEs arising from non-informatively interval censored data, one can see Sun (2006, Chapter 2). We establish in this chapter consistency and asymptotic normality of the MLE under the chosen model for informatively interval censored data.

In Section 2.2, we introduce the new model that relates the underlying time-to-event with actual observations in a retrospective study. The corresponding likelihood is also presented. This model is able to handle, as special cases, data arising from current status monitoring, non-informative interval censoring and random right censoring. In Section 2.3 we discuss an alternative formulation of the model introduced in Section 2.2. In Section 2.4, we derive some asymptotic results for the MLE based on the likelihood obtained in Section 2.2. In Section 2.5, we demonstrate theoretically the advantage of the new estimator over MLE based on current status data and retrospective data wrongly assumed to be non-informatively interval censored. In Section 2.6, we present a simulation study to investigate the finite-sample properties of the estimates. In Section 2.7, we discuss the issue of model adequacy and present a graphical technique for this purpose. In Section 2.8, we illustrate the methods through the analysis of the real data set mentioned in Section 1.3. Some concluding remarks are provided in Section 2.9.

2.2 Model

Let the time-to-event of n subjects, $T_i, (i = 1, 2, \ldots, n)$ be samples from the distribution F_θ, where θ is a vector of parameters. The i^{th} subject is visited at time S_i. It is assumed that the S_i's are samples from another distribution and are independent of the T_i's.

In the case of status quo data, one observes $(S_i, \delta_i), (i = 1, 2, \ldots, n)$ where $\delta_i = I_{(T_i \leqslant S_i)}$, the indicator of the event $(T_i \leqslant S_i)$. The likelihood is

$$\prod_{i=1}^{n} [F_\theta(S_i)]^{\delta_i} [\bar{F}_\theta(S_i)]^{1-\delta_i}, \tag{2.1}$$

where $\bar{F}_\theta(S_i) = 1 - F_\theta(S_i)$.

In a retrospective study, the subject may not recall clearly the date of the event. Here, we ignore the possibility of the subject recalling an approximate date, and regard such occurrence as a non-recall event. Let ε_i be the indicator of the ith subject recalling the exact time of his/her landmark event. Note that whenever $\delta_i = 1$ and $\varepsilon_i = 0$, it is known that $T_i < S_i$. If the underlying censoring mechanism is presumed to be non-informative, then the likelihood is

$$\prod_{i=1}^{n} \left[(F_\theta(S_i))^{1-\varepsilon_i}(f_\theta(T_i))^{\varepsilon_i}\right]^{\delta_i} [\bar{F}_\theta(S_i)]^{1-\delta_i}, \tag{2.2}$$

where f_θ is the probability density function corresponding to the distribution F_θ.

It has been pointed out in the Chapter 1 that non-informativeness of censoring is difficult to justify in the present context. The non-recall probability may depend on the observation time and the time-to-event, and may be expressed as a function π defined over the domain $\{(t, s) : 0 < t < s\}$ by the equation

$$\pi(t, s) = P(\varepsilon_i = 0 | S_i = s, T_i = t).$$

There would be three cases for an individual i, with different contributions to the likelihood.

CASE (i) When $\delta_i = 0$ (the event has not occurred till the time of observation), the contribution of the individual to the likelihood is $\bar{F}_\theta(S_i)$.

CASE (ii): When $\delta_i \varepsilon_i = 1$ (the event has occurred and the subject can remember the time), the contribution of the individual to the likelihood is $f_\theta(T_i)\{1 - \pi(T_i, S_i)\}$.

CASE (iii): When $\delta_i(1 - \varepsilon_i) = 1$ (the event has occurred but the subject cannot recall the time), the contribution of the individual to the likelihood is $\int_0^{S_i} f_\theta(u)\pi(u, S_i)du$.

The likelihood according to this model is

$$\prod_{i=1}^{n} \left[\left(\int_0^{S_i} f_\theta(u)\pi(u, S_i)du\right)^{1-\varepsilon_i} [f_\theta(T_i)(1 - \pi(T_i, S_i))]^{\varepsilon_i}\right]^{\delta_i} [\bar{F}_\theta(S_i)]^{1-\delta_i}. \tag{2.3}$$

We presume, for the sake of simplicity, that the non-recall probability depends only on the time elapsed since the event of interest, $S_i - T_i$. In other words, we model π as

$$\pi(t,s) = \pi_\eta(s-t), \qquad 0 < t < s, \tag{2.4}$$

where π_η is a family of functions indexed by the parameter η. According to this model, the likelihood is

$$\prod_{i=1}^{n}\left[\left(\int_0^{S_i} f_\theta(u)\pi_\eta(S_i-u)du\right)^{1-\varepsilon_i}[f_\theta(T_i)(1-\pi_\eta(S_i-T_i))]^{\varepsilon_i}\right]^{\delta_i}[\bar{F}_\theta(S_i)]^{1-\delta_i}. \tag{2.5}$$

The MLE based on the above likelihood is expected to harness the information in the recall data without making unrealistic assumptions about censoring. The parameter η, which can be a vector, would have to be regarded as a nuisance parameter in the present context.

In an unpublished technical report, Stine and Small (1986) had used MLE based on a special case of the above likelihood, where π_η is presumed to be a piecewise constant function. They did not study the statistical properties of the estimator.

Here, the informativeness of the censoring mechanism is captured through the function π_η. When π_η is a constant, the likelihood (2.5) becomes a constant multiple of the likelihood (2.2). As a further special case, if $\pi_\eta = 1$, the likelihood (2.5) reduces to the likelihood (2.1). On the other hand, when $\pi_\eta = 0$, the likelihood reduces to

$$\prod_{i=1}^{n}[f_\theta(T_i)]^{\delta_i}[\bar{F}_\theta(S_i)]^{1-\delta_i}, \tag{2.6}$$

which is the same as the likelihood for randomly right-censored prospective data obtained from continuous monitoring. Thus, the likelihood (2.5) is based on a model that is more general than the usual censoring models.

2.3 An alternative formulation

There can be an alternative modeling of the recalled event time involving another (possibly notional) duration. There could be an underlying distribution (F_θ) for the time till the occurrence of the event of interest, and another distribution (π_η) for the time from that occurrence to the forgetting of the date. The latter may in fact be a sub-distribution function, with some mass at infinity. Suppose U_i is the unobservable time that the ith subject would take to forget the epoch of his/her landmark event, having distribution π_η. We assume that, for $i = 1, \ldots, n$, the distribution of the triplet (T_i, S_i, U_i) is the product of its one-dimensional marginals and the triplet for the different individuals are independent of one another. Note that S_i is always observed, U_i is never observed, and T_i is observed only when $T_i \leq S_i \leq T_i + U_i$. The observables δ_i and ε_i can be expressed in terms of these random variables as the indicators of $T_i \leq S_i$ and $U_i + T_i > S_i$, respectively, the latter being defined only when $\delta_i = 1$. It follows that, given $S_i = s$ and $T_i = t$ with $s > t$, the non-recall probability depends on the time elapsed since the landmark event as

$$P(\varepsilon_i = 0 | T_i = t, S_i = s) = P(U_i \leq S_i - T_i | T_i = t, S_i = s) = \pi_\eta(s - t). \qquad (2.7)$$

In this formulation also, there would be three cases for determining the contribution of individual i to the likelihood.

CASE (i) When $S_i < T_i$ (neither event has occurred till the date of observation), the contribution of the individual to likelihood is $\bar{F}_\theta(S_i)$.

CASE (ii): When $T_i \leq S_i < T_i + U_i$ (only the first event has occurred), the contribution of the individual to likelihood is $f_\theta(T_i)\{1 - \pi_\eta(S_i - T_i)\}$.

CASE (iii): When $S_i \geq T_i + U_i$, (both the events have occurred) the contribution of the individual to likelihood is $\int_0^{S_i} f_\theta(u)\pi_\eta(S_i - u)du$.

It can be seen that these contributions also lead to the likelihood (2.5). In fact, the above formulation provides an interpretation of the 'forgetting function' π_η (introduced in the previous section) as the distribution function of the time to the forgetting event, measured from the date of occurrence of the main event. This interpretation holds when π_η is non-decreasing, while the general formulation of Section 2.2 remains applicable even when π_η does not have this property.

2.4 Large sample properties

The factors in the product likelihood (2.5) have different forms in different cases. For example, T_i is used only when $\delta_i = 1$ and $\varepsilon_i = 1$. In order for the standard asymptotic results to be applicable, each factor of this likelihood has to be expressed as the density of some random vector in a suitable probability space.

We have already assumed that the T_i's (time-to-event) are samples from the distribution F_θ and the S_i's (ages on interview date) are samples from another distribution. Let G be the common distribution of the S_i's. Let

$$V_i = (S_i - T_i)\varepsilon_i\delta_i, \tag{2.8}$$

where ε_i and δ_i are as defined in Section 2.2. Note that the vector

$$Y_i = (S_i, V_i, \delta_i) \tag{2.9}$$

is observed in all cases, and contains all the requisite information.

We now show that the i^{th} factor in the product likelihood (2.5) is in fact proportional to the density of Y_i. We prove this result below, after dropping the subscript i for simplicity. The dominating probability measure used for defining this density is $\mu = \vartheta_1 \times \vartheta_2 \times \vartheta_3$ where ϑ_1 is the measure with respect to which G has a density (e.g., the counting or the Lebesgue measure, depending if G is discrete or continuous, respectively) ϑ_2 is the sum of the counting and the Lebesgue measures, and ϑ_3 is the counting measure (Ash, 2000).

Theorem 2.1. *The density of $Y = (S, V, \delta)$ with respect to the measure μ is*

$$h(s, v, \delta) = \begin{cases} g(s)\bar{F}_\theta(s) & \text{if } v = 0 \text{ and } \delta = 0, \\ g(s)\int_0^s f_\theta(u)\pi_\eta(s-u)du & \text{if } v = 0 \text{ and } \delta = 1, \\ g(s)f_\theta(s-v)(1-\pi_\eta(v)) & \text{if } v > 0 \text{ and } \delta = 1, \\ 0 & \text{otherwise}, \end{cases} \tag{2.10}$$

where g is the density of G with respect to the measure ϑ_1.

Proof. The density in the first two cases can be obtained by considering the corresponding probability masses:

$$\begin{aligned} h(s,0,0) &= P(V=0,\delta=0|S=s)g(s) \\ &= P(T>s|S=s)g(s) = \bar{F}_\theta(s)g(s); \\ h(s,0,1) &= E_T[h(s,0,1)|T] \\ &= E_T[P(S>T|S=s,T)g(s)\pi_\eta(s-T)] \\ &= \int_0^s g(s)\pi_\eta(s-u)f_\theta(u)du. \end{aligned}$$

In the third case, the density can be derived as the derivative of a probability,

$$\begin{aligned} h(s,v,1) &= g(s)\frac{\partial P(V<v,\delta=1|S=s)}{\partial v} \\ &= g(s)\lim_{h\to 0}\frac{P(v<V\leqslant v+h,\delta=1|S=s)}{h} \\ &= g(s)\lim_{h\to 0}\frac{P(v<V\leqslant v+h|S=s)}{h} \\ &= g(s)\lim_{h\to 0}\frac{P(v<s-T\leqslant v+h,T<s,\varepsilon=1)}{h} \\ &= g(s)\lim_{h\to 0}\frac{P(s-v-h<T\leqslant s-v,\varepsilon=1)}{h} \\ &= g(s)\lim_{h\to 0}\frac{E_T[P(\varepsilon=1|T)I_{(s-v-h<T\leq s-v)}]}{h} \\ &= g(s)\lim_{h\to 0}\frac{\int_{s-v-h}^{s-v} f_\theta(u)(1-\pi_\eta(s-u))du}{h} \\ &= g(s)f_\theta(s-v)(1-\pi_\eta(v)). \end{aligned}$$

□

The likelihood (2.5) can be written in terms of S_i, V_i and δ_i as

$$\begin{aligned} &\prod_{i=1}^n \left[\left(\int_0^{S_i} f_\theta(u)\pi_\eta(S_i-u)du\right)^{I(V_i=0)} [f_\theta(S_i-V_i)(1-\pi_\eta(V_i))]^{I(V_i>0)}\right]^{\delta_i} \\ &\quad\times [\bar{F}_\theta(S_i)]^{1-\delta_i} \\ &= \frac{\prod_{i=1}^n h(S_i,V_i,\delta_i)}{\prod_{i=1}^n g(S_i)}. \end{aligned} \tag{2.11}$$

The numerator is a product of densities of the form (2.10), while the denominator does not contain any information about θ. This likelihood can also be interpreted as a product of conditional densities of (V_i, δ_i) given S_i, for $i = 1, 2, \ldots, n$. Further, this conditional likelihood is free from g, i.e., inference for θ can proceed by ignoring any parameter of g.

Once the likelihood (2.5) is identified as a product of densities, standard results for consistency and asymptotic normality of the MLE become applicable. We would look for conditions on the original variables S_i, T_i and ε_i, which completely determine the observable triplet (S_i, V_i, δ_i). Since the likelihood involves only the conditional density of (V_i, δ_i) given S_i, it suffices to look for conditions on the distribution of (T_i, ε_i) only. Specifically, the conditions would involve the density f_θ, the density of T_i, and the function π_η, which defines the conditional density of the binary random variable ε_i given T_i and S_i.

We need the following conditions in order to establish consistency of the MLE.

(C1) The parameter θ is identifiable with respect to the family of densities f_θ of the time-to-event, and the parameter η is identifiable with respect to the family of functions π_η representing non-recall probability. In other words, $\theta_1 \neq \theta_2$ implies $f_{\theta_1} \neq f_{\theta_2}$, and $\eta_1 \neq \eta_2$ implies $\pi_{\eta_1} \neq \pi_{\eta_2}$.

(C2) The parameter spaces for θ and η are open.

(C3) The sets $A_1 = \{t : f_\theta(t) > 0\}$ and $A_2 = \{v : \pi_\eta(v) > 0\}$ are independent of θ and η respectively.

(C4) The function $f_\theta(t)$ is differentiable with respect to θ for all t such that the derivative is absolutely bounded by an integrable function $h_1(t)$, and the function $\pi_\eta(v)$ is differentiable with respect to η for all v such that the derivative is absolutely bounded by an integrable function $h_2(v)$.

Theorem 2.2. *Let for $i = 1, 2, \ldots, n$, T_i and S_i be samples from distributions F_θ and G, respectively, where F_θ has density f_θ, and $\delta_i = I_{(T_i < S_i)}$. Let $\varepsilon_1, \ldots, \varepsilon_n$ be independent binary random variables with*

$$P(\varepsilon_i = 0 | S_i = s, T_i = t, t < s) = \pi_\eta(s - t).$$

Let f_θ and π_η satisfy Conditions (C1)-(C4), Then there exists a sequence $(\hat{\theta}_n, \hat{\eta}_n)$

of local maxima of the likelihood function (2.5) *which is consistent, i.e.,*

$$\begin{pmatrix} \hat{\theta}_n \\ \hat{\eta}_n \end{pmatrix} \xrightarrow{P} \begin{pmatrix} \theta \\ \eta \end{pmatrix} \qquad \textit{for all} \quad \theta \textit{ and } \eta \tag{2.12}$$

Proof. Let Y_i be constructed from $(T_i, S_i, \varepsilon_i)$ in the manner described in (2.9). Note that Y_i is observable even when $(T_i, S_i, \varepsilon_i)$ is not. We have shown in Theorem 2.1 that the Y_i's are samples from a distribution. Their joint density is proportional to the likelihood (2.5), which can be written in terms of Y_i's alone, and the constant of proportionality is the marginal density of S_i's (free of the parameters). It can be easily seen that the hypothesis of the present theorem and Conditions (C1)-(C3) imply that the distribution of Y_1 satisfies Conditions C1-C4 of Theorem 7.1.1 of Lehman (1999). Condition (C4) implies that the quantities $\int_0^s \frac{\partial}{\partial\theta} f_\theta(u)\pi_\eta(s-u)du$ and $\int_0^s f_\theta(u)\frac{\partial}{\partial\eta}\pi_\eta(s-u)du$ are well defined, and are the derivatives of the conditional density of (V_i, δ_i) given S_i with respect to θ and η, respectively, in the case $v = 0$ and $\delta = 1$. It is easier to establish the corresponding implications in the other cases, which lead to the fulfillment of Condition C5 of Theorem 7.1.1 of Lehman (1999). The statement of the theorem follows. □

The log-likelihood obtained from (2.5) can be written as

$$\ell(\theta, \eta) = \sum_{i=1}^{n} \Bigg[\delta_i(1-\varepsilon_i)\log\left(\int_0^{S_i} f_\theta(u)\pi(S_i - u)du\right) + \delta_i\varepsilon_i \log\left(f_\theta(T_i)(1-\pi(S_i - T_i))\right) + (1-\delta_i)\log\left(\bar{F}_\theta(S_i)\right)\Bigg]. \tag{2.13}$$

In order to establish asymptotic normality of the MLE, the following additional conditions on $\ell(\theta, \eta)$ are needed.

(C5) Second partial derivatives of $\ell(\theta, \eta)$ with respect to θ and η exist and are continuous, and may be passed under the integral sign in $\int \ell(\theta, \eta)d\mu$.

(C6) The elements of the matrix

$$A(\theta, \eta) = \begin{bmatrix} \frac{\partial^2}{\partial\theta\partial\theta^T}\ell(\theta,\eta) & \frac{\partial^2}{\partial\theta\partial\eta^T}\ell(\theta,\eta) \\ \frac{\partial^2}{\partial\eta\partial\theta^T}\ell(\theta,\eta) & \frac{\partial^2}{\partial\eta\partial\eta^T}\ell(\theta,\eta) \end{bmatrix}$$

are bounded in absolute value, uniformly in some neighborhood of the true value of the parameter (θ, η), by some function $K(x)$ such that $E_{(\theta_0, \eta_0)} K(X) < \infty$.

(C7) The Fisher information matrix of a single sample

$$I(\theta, \eta) = \frac{1}{n} E \begin{bmatrix} \left(\frac{\partial}{\partial \theta} \ell(\theta, \eta)\right) \left(\frac{\partial}{\partial \theta} \ell(\theta, \eta)\right)^T & \left(\frac{\partial}{\partial \theta} \ell(\theta, \eta)\right)\left(\frac{\partial}{\partial \eta} \ell(\theta, \eta)\right)^T \\ \left(\frac{\partial}{\partial \eta} \ell(\theta, \eta)\right)\left(\frac{\partial}{\partial \theta} \ell(\theta, \eta)\right)^T & \left(\frac{\partial}{\partial \eta} \ell(\theta, \eta)\right) \left(\frac{\partial}{\partial \eta} \ell(\theta, \eta)\right)^T \end{bmatrix}$$

is non-singular.

Theorem 2.3. *Let for $i = 1, 2, \ldots, n$, T_i and S_i be samples from distributions F_θ and G respectively, where F_θ has density f_θ, and $\delta_i = I_{(T_i < S_i)}$. Let $\varepsilon_1, \ldots, \varepsilon_n$ be independent binary random variables with*

$$P(\varepsilon_i = 0 | S_i = s, T_i = t, t < s) = \pi_\eta(s - t).$$

Let f_θ and π_η satisfy Conditions (C1)-(C7), Then any consistent sequence of roots $(\hat{\theta}_n, \hat{\eta}_n)$ of the likelihood equation obtained from (2.5) satisfies

$$\sqrt{n} \left(\begin{pmatrix} \hat{\theta}_n \\ \hat{\eta}_n \end{pmatrix} - \begin{pmatrix} \theta \\ \eta \end{pmatrix} \right) \xrightarrow{L} N\left(0, I^{-1}(\theta, \eta)\right). \tag{2.14}$$

Proof. Let Y_i be defined as in (2.9). From Theorem 2.1, $Y_1, \ldots, Y_n$ are i.i.d. samples from a distribution. The assumption of the present theorem, together with Conditions (C1)-(C2), imply that the distribution of Y_1 satisfies Conditions (1) and (5) of Theorem 18 of Ferguson (1996). It is also easy to see that Conditions (C5)-(C7) imply that the distribution of Y_1 satisfies Conditions (2)-(4) of Theorem 18 of Ferguson (1996). The proof follows. □

2.5 Theoretical comparison of estimates

2.5.1. Bias of MLE based on conventional likelihood

If one ignores the informative nature of censoring, then the likelihood (2.2) would appear to be appropriate. We now show that an MLE based on that likelihood may be inconsistent under the general censoring model of Section 2.2. Inconsistency is established if the bias can be shown not to go to zero as the sample size goes to infinity. As the MLE based on (2.2) is not generally available in closed form, we avoid computing the asymptotic bias, and compute instead the expected value of the score function obtained from the likelihood (2.2), computed under the general model.

Let $f_\theta(t) = \frac{1}{\theta}e^{-\frac{t}{\theta}}$ and $\pi_\eta(u) = 1 - e^{-\frac{u}{\eta}}$. The derivative of the log-likelihood obtained from (2.2) with respect to θ is

$$\sum_{i=1}^{n}\left[\delta_i(1-\varepsilon_i)\left(\frac{\frac{s_i}{\theta^2}e^{\frac{-s_i}{\theta}}}{1-e^{\frac{-s_i}{\theta}}}\right)+\delta_i\varepsilon_i\left(\frac{-1}{\theta}+\frac{t_i}{\theta^2}\right)+(1-\delta_i)\frac{s_i}{\theta^2}\right]. \qquad (2.15)$$

The expectation of (2.15) with respect to the general model of Section 2.2 is n times

$$E_S\left[\frac{S}{\theta^2}\bar{F}_\theta(S)+\int\left(\frac{-S}{\theta}+\frac{t}{\theta^2}\right)(1-\pi_\eta(S-t))f_\theta(t)dt+\frac{\frac{S}{\theta^2}e^{\frac{-S}{\theta}}}{1-e^{\frac{-S}{\theta}}}\int\pi_\eta(S-t)f_\theta(t)dt\right].$$

In the further special case $\eta = \theta$, the above expression reduces to

$$E_S\left[\frac{1}{2\theta}\frac{\frac{S}{\theta}e^{-\frac{S}{\theta}}}{1-e^{-\frac{S}{\theta}}}\left(2-2e^{-\frac{S}{\theta}}-\frac{S}{\theta}-\frac{S}{\theta}e^{-\frac{S}{\theta}}\right)\right].$$

For the expectation to be equal to zero, the function in square brackets should be orthogonal to the probability function of S, which would not hold in general. One can design infinitely many distributions of S, which would violate this condition. If the expected value of the score function obtained from (2.2) is not zero, the asymptotic bias of the corresponding 'MLE' is also not zero.

2.5.2. *Additional information from recall data*

In order to identify the additional information arising from recall data, we return to the expression of the likelihood in terms of the joint density of (S, V, δ). We presume that the distribution of S does not involve any unknown parameter. Then the joint density of the observed triplet can be written as

$$h_{\theta,\eta}(s,v,\delta) = f_\theta(s,\delta)f_{\theta,\eta}(v|s,\delta).$$

Thus, the log-likelihood for a single sample is

$$\log(h_{\theta,\eta}(s,v,\delta)) = \log(f_\theta(s,\delta)) + \log(f_{\theta,\eta}(v|s,\delta)),$$

and consequently, information for the two parameters is of the form

$$I_R(\theta,\eta) = I_S(\theta,\eta) + I_A(\theta,\eta), \tag{2.16}$$

where the matrices I_R, I_S and I_A are the information arising from recall data, status quo data and recall data conditioned on status quo data, respectively.

Since the likelihood of status quo data is free from η, $I_S(\theta,\eta)$ is a function of θ alone, and can be written as

$$I_S(\theta,\eta) = \begin{bmatrix} I_1 & 0 \\ 0 & 0 \end{bmatrix},$$

where

$$I_1 = -E\left[\frac{\partial^2}{\partial\theta\partial\theta^T}\log(f_\theta(s,\delta))\right].$$

On the other hand, the additional information obtain from the recall data is

$$I_A(\theta,\eta) = \begin{bmatrix} I_2 & I_3 \\ I_3^T & I_4 \end{bmatrix},$$

where

$$I_2 = -E\left[\frac{\partial^2}{\partial\theta\partial\theta^T}\log(f_{\theta,\eta}(v|s,\delta))\right],$$

$$I_3 = -E\left[\frac{\partial^2}{\partial\theta\partial\eta^T}\log(f_{\theta,\eta}(v|s,\delta))\right],$$

$$I_4 = -E\left[\frac{\partial^2}{\partial\eta\partial\eta^T}\log(f_{\theta,\eta}(v|s,\delta))\right].$$

In particular, the additional information of θ, the parameter of interest, is

$$I_2 - I_3 I_4^{-1} I_3^T.$$

When η is known, the additional information reduces to I_2.

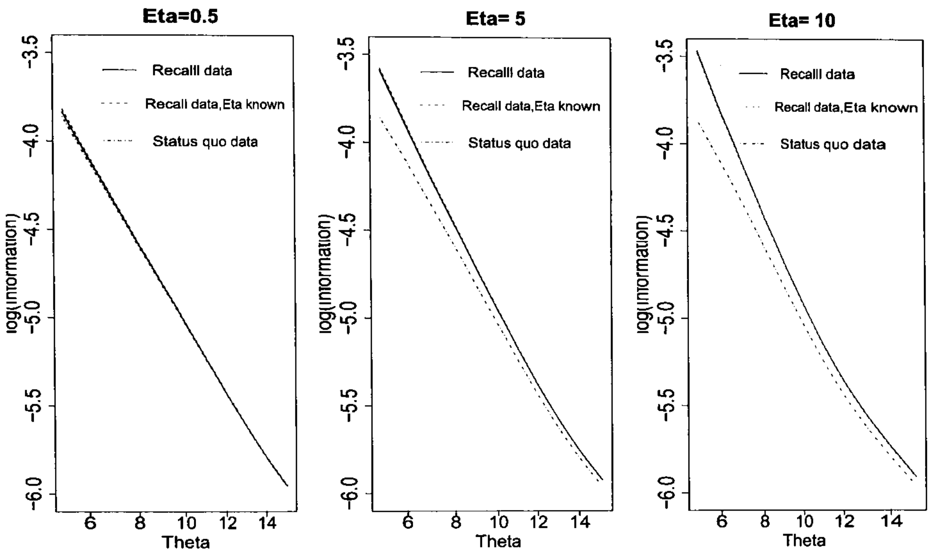

Fig 2.1: Log of information based on recall data and status quo likelihoods.

As an example, consider the special case, where $f_\theta(t) = \frac{1}{\theta}e^{-\frac{t}{\theta}}$ and $\pi_\eta(v) = 1 - e^{-v/\eta}$. Figure 2.1 shows plots of the log of information arising from status quo data (I_1), from recall data $(I_1 + I_2 - I_3 I_4^{-1} I_3^T)$ and from recall data with known η, $(I_1 + I_2)$, for different values of η and a range of values of θ. It can be seen that, when η is large, there is a considerable gap between the first two, while there is not much gap between the second and the third curves. Thus, in this case, the price for not knowing the nuisance parameter η is minimal compared to the gain from recall data. On the other hand, for a small value of η (i.e., time-to-event forgotten quickly), recall data does not augment the information noticeably. When the gap is substantial, the MLE based on status quo data is expected to have poor efficiency with respect to the MLE computed from (2.5).

2.6 Simulation results

For the purpose of simulation, we generate sample times-to-event from the Weibull distribution with shape and scale parameters α and β, respectively. Thus, $\theta =$

(α, β). Further, we assume that the 'age at interview' follows the discrete uniform distribution over [7,21] and that

$$\pi_\eta(x) = 1 - e^{-\frac{x}{\eta}}. \tag{2.17}$$

We use the following values of the parameters.

(i) $\alpha = 11$, $\beta = 13$ and $\eta = 3$,

(ii) $\alpha = 10$, $\beta = 12$ and $\eta = 5$.

The two choices correspond to median ages of the landmark event of about 11.57 and 12.58 years, and inter-quantile ranges of about 1.78 and 1.80 years, respectively. These choices are in line with the data analytic example given in Section 2.8. The mean times to forget are chosen as 3 and 5 years, respectively.

We compare the performance of the MLE's based on the status quo likelihood (2.1), the interval censoring likelihood (2.2) and the recall data likelihood (2.5) for our model. Computation of MLE's in all the cases is done through numerical optimization of the likelihood using the 'Quasi-Newton' method (Nocedal and Wright, 2006).

We run 1000 simulations for each of the above combinations of parameters, for sample sizes $n = 50$, 500 and 1000.

Table 2.1 shows the bias, the standard deviation (Stdev), the mean squared error (MSE) and the Cramer-Rao Lower Bound (CRLB) for the MLE's of the three parameters based on the three likelihoods, for the combination of parameter values in cases (i) and (ii).

In both cases it is found that the bias for the MLE based on interval censoring likelihood stabilizes around a positive constant when the sample size increases. Bias of the MLE from status quo data reduces as the sample size increases. The standard deviation of the MLE based on our model is smaller than that based on status quo data, and is also in line with the Cramer-Rao lower bound – particularly when the sample size is large.

In order to check the robustness of the proposed method against departure from the assumed form of the non-recall probability function π_η, we use the following

TABLE 2.1
Bias, Stdev, MSE and CRLB of estimated parameters in case (i) $\alpha = 11$, $\beta = 13$, $\eta = 3$ *and case (ii)* $\alpha = 10$, $\beta = 12$,$\eta = 5$

Estimator	Property	Case	$n = 50$			$n = 500$			$n = 1000$		
			α	β	η	α	β	η	α	β	η
MLE from Status quo	Bias	(i)	7.302	–0.103	–	0.647	0.002	–	0.499	0.001	–
	Stdev		13.67	0.519	–	1.361	0.149	–	0.923	0.104	–
	MSE		239.9	0.280	–	2.271	0.022	–	1.101	0.011	–
	CRLB		14.135	0.217	–	1.672	0.021	–	0.850	0.010	–
MLE from Interval censoring	Bias		2.788	0.193	–	1.410	0.223	–	1.332	0.220	–
	Stdev		4.528	0.318	–	0.854	0.102	–	0.589	0.069	
	MSE		28.26	0.138	–	2.721	0.062	–	2.122	0.053	–
	CRLB		1.933	0.043	–	0.255	0.004	–	0.126	0.003	–
Proposed MLE	Bias		1.563	–0.016	0.058	0.325	0.008	–0.005	0.250	0.006	-0.0002
	Stdev		4.100	0.310	0.875	0.777	0.100	0.242	0.545	0.069	0.169
	MSE		19.239	0.096	0.769	0.709	0.010	0.058	0.360	0.005	0.028
	CRLB		5.266	0.043	0.619	0.566	0.004	0.062	0.288	0.004	0.030
MLE from Status quo	Bias	(ii)	8.589	–0.047	–	1.083	0.048	–	0.938	0.032	
	Stdev		9.529	0.507	–	1.239	0.148	–	0.873	0.100	
	MSE		164.6	0.259	–	2.707	0.024	–	1.641	0.011	
	CRLB		68.83	0.121	–	1.571	0.019	–	0.757	0.010	
MLE from Interval censoring	Bias		2.391	0.198	–	1.369	0.217	–	1.317	0.213	–
	Stdev		2.667	0.287	–	0.614	0.088	–	0.431	0.061	–
	MSE		12.83	0.121	–	2.250	0.055		1.919	0.049	
	CRLB		2.327	0.027	–	0.195	0.003	–	0.096	0.003	
Proposed MLE	Bias		1.581	0.093	0.166	0.619	0.046	0.009	0.570	0.042	0.003
	Stdev		2.500	0.281	1.392	0.585	0.086	0.391	0.411	0.060	0.284
	MSE		8.747	0.088	1.963	0.726	0.009	0.152	0.493	0.005	0.081
	CRLB		4.366	0.014	7.062	0.369	0.002	0.693	0.166	0.003	0.080

non-recall function for data generation.

$$\pi_\eta(x) = 0.05 I_{(0<x\leq 2.5)} + 0.35 I_{(2.5<x\leq 4.5)} + 0.95 I_{(4.5<x<\infty)}. \quad (2.18)$$

We generate the data from two different models.

(iii) The 'time-to-event' from Weibull distribution with parameters $\alpha = 11$ and $\beta = 13$ and the π_η function defined in (2.18),

(iv) The 'time-to-event' from Weibull distribution with parameters $\alpha = 10$ and $\beta = 12$ and the π_η function defined in (2.18).

We run 1000 simulations for each of the above combinations of parameters, for sample sizes $n = 50$, 500 and 1000. Table 2.2 shows the performance of MLE's based on the status quo likelihood (2.1), the interval censoring likelihood (2.2) and the recall data likelihood (2.5) based on the incorrect model (2.17). We compute the bias, the standard deviation and the MSE for the MLE's of the parameters of interest, based on the three likelihoods, for the combination of parameter values in cases (iii) and (iv).

TABLE 2.2

Bias, Stdev and MSE of estimated parameters in case (iii) $\alpha = 11$ and $\beta = 13$ and the π_η function defined in (2.18) and case (iv) $\alpha = 10$ and $\beta = 12$ and the π_η function defined in (2.18)

Estimator	Property	Case	$n = 50$		$n = 500$		$n = 1000$	
			α	β	α	β	α	β
MLE from Status quo	Bias	(iii)	8.666	-0.100	0.692	-0.009	0.484	0.005
	Stdev		14.945	0.515	1.330	0.152	0.928	0.102
	MSE		298.2	0.275	2.259	0.023	1.095	0.010
MLE from Interval censoring	Bias		2.544	0.262	1.401	0.254	1.311	0.237
	Stdev		3.221	0.306	0.734	0.096	0.547	0.068
	MSE		16.839	0.162	2.502	0.074	2.019	0.061
Proposed MLE	Bias		2.014	0.117	0.899	0.111	0.812	0.101
	Stdev		3.091	0.293	0.706	0.093	0.523	0.066
	MSE		13.602	0.099	1.308	0.021	0.933	0.014
MLE from Status quo	Bias	(iv)	9.039	-0.066	1.158	0.045	0.881	0.040
	Stdev		14.846	0.514	1.292	0.151	0.814	0.103
	MSE		301.9	0.268	3.010	0.024	1.439	0.012
MLE from Interval censoring	Bias		2.644	0.287	1.598	0.286	1.486	0.269
	Stdev		2.677	0.301	0.675	0.096	0.463	0.066
	MSE		14.15	0.173	3.011	0.091	2.423	0.077
Proposed MLE	Bias		2.169	0.351	1.147	0.142	1.038	0.140
	Stdev		2.583	0.289	0.655	0.092	0.450	0.063
	MSE		11.37	0.207	1.745	0.029	1.281	0.024

In both cases, the MSE of the MLEs based on our method is generally smaller than the same obtained from the two other methods, but somewhat larger than the MSE reported in Table 2.1.

We now check the robustness of the method against the basic assumption that the non-recall probability function depends only on the time since the landmark event. In view of the possibility that some subjects having had that event in early age may remember the date even after a long time, we consider the alternative form of the non-recall probability function as follows.

$$\pi(S,T) = \begin{cases} 0.5\left(1 - e^{-\frac{S-T}{\eta}}\right) & \text{if } T < 9, \\ \left(1 - e^{-\frac{S-T}{\eta}}\right) & \text{if } T \geq 9. \end{cases} \tag{2.19}$$

Under the above model, those who had that event in very early ages would remember it more often, making these cases account for a larger share of exact recall cases, as compared to the model (2.17).

We generate data from two different models.

(v) The 'time-to-event' from Weibull distribution with parameters $\alpha = 11$ and $\beta = 13$, and the π function defined in (2.19) when $\eta = 3$,

(vi) The 'time-to-event' from Weibull distribution with parameters $\alpha = 10$ and $\beta = 12$, $\eta = 5$ and the π function defined in (2.19) when $\eta = 5$.

TABLE 2.3
Bias, Stdev and MSE of estimated parameters of interest in case (v) $\alpha = 11$ and $\beta = 13$, and the π function defined in (2.19) when $\eta = 3$

Sample size	Property	MLE	α	β	$P(T < 9)$
$n = 50$	Bias	Status quo	8.306	–0.112	0.001
		Interval censoring	2.348	0.175	–0.005
		Proposed model	1.313	–0.035	0.003
	Stdev	Status quo	15.46	0.511	0.026
		Interval censoring	4.852	0.341	0.012
		Proposed model	4.951	0.335	0.019
	MSE	Status quo	304.9	0.273	0.0007
		Interval censoring	29.21	0.147	0.0002
		Proposed model	26.21	0.114	0.0003
$n = 500$	Bias	Status quo	2.449	-0.006	0.0004
		Interval censoring	0.789	0.206	-0.006
		Proposed model	–0.209	–0.023	0.002
	Stdev	Status quo	1.349	0.147	0.008
		Interval censoring	0.806	0.103	0.004
		Proposed model	0.743	0.103	0.005
	MSE	Status quo	1.878	0.022	0.00006
		Interval censoring	1.272	0.053	0.00005
		Proposed model	0.595	0.011	0.00004
$n = 1000$	Bias	Status quo	0.136	0.0002	–0.0001
		Interval censoring	0.764	0.206	–0.006
		Proposed model	–0.191	–0.019	0.002
	Stdev	Status quo	0.854	0.101	0.005
		Interval censoring	0.546	0.072	0.002
		Proposed model	0.498	0.071	0.003
	MSE	Status quo	0.747	0.010	0.00003
		Interval censoring	0.883	0.048	0.00004
		Proposed model	0.284	0.005	0.00001

We run 1000 simulations for each of the above combinations of parameters, for sample sizes $n = 50$, 500 and 1000. Tables 2.3 and 2.4 show the performance of MLE's based on the status quo likelihood (2.1), the interval censoring likelihood (2.2) and the recall data likelihood (2.5) under the model (2.17), for the combination of parameter values in cases (v) and (vi) respectively. In addition to the original parameters α and β, we consider the derived parameter $P(T < 9)$ representing the probability of having the event very early, which is expected to be overestimated when the exponential model (2.17) is assumed instead of (2.19). We compute the bias, the standard deviation and the MSE for the MLE's based on the three likelihoods.

TABLE 2.4
Bias, Stdev and MSE of estimated parameters of interest in case (vi) $\alpha = 10$ and $\beta = 12$, $\eta = 5$ and the π function defined in (2.19) when $\eta = 5$

Sample size	Property	MLE	α	β	$P(T < 9)$
n=50	Bias	Status quo	7.425	−0.103	−0.002
		Interval censoring	1.211	0.222	−0.012
		Proposed model	0.558	−0.046	0.004
	Stdev	Status quo	14.59	0.519	0.058
		Interval censoring	2.220	0.294	0.025
		Proposed model	2.094	0.291	0.032
	MSE	Status quo	267.7	0.280	0.003
		Interval censoring	6.391	0.136	0.0007
		Proposed model	4.694	0.087	0.001
n=500	Bias	Status quo	0.286	0.023	−0.001
		Interval censoring	0.528	0.155	−0.012
		Proposed model	−0.135	−0.022	0.003
	Stdev	Status quo	1.288	0.151	0.019
		Interval censoring	0.806	0.103	0.004
		Proposed model	0.540	0.089	0.010
	MSE	Status quo	1.739	0.023	0.0004
		Interval censoring	0.603	0.032	0.0002
		Proposed model	0.309	0.008	0.0001
n=1000	Bias	Status quo	0.129	0.010	−0.001
		Interval censoring	0.478	0.155	−0.012
		Proposed model	−0.088	−0.021	0.003
	Stdev	Status quo	0.845	0.102	0.013
		Interval censoring	0.401	0.064	0.006
		Proposed model	0.383	0.062	0.007
	MSE	Status quo	0.731	0.010	0.0002
		Interval censoring	0.389	0.028	0.0002
		Proposed model	0.154	0.004	0.00006

In both cases, the bias, the standard deviation and the MSE of the MLEs based on our method is smaller than the same, computed from the two other methods. Further, the proposed estimator of $P(T < 9)$ is found to have a positive bias as expected. The amount of bias is not very large. Performances of the MLEs of α and β are in line with that reported in Table 2.1, where there was no specification error in the non-recall probability function.

2.7 Adequacy of the model

In order to check how well the assumed parametric model actually fits the data, one can use the chi-square goodness of fit test (Gibbons and Chakraborti, 2003). For this purpose, the data may be transformed to the trivariate vector $Y = (S, V, \delta)$, and the support of the joint distribution of this vector may be appropriately partitioned, depending on the availability of data. An example is given in the next section.

Modeling of the non-recall function can be a critical issue. There would be a trade off between a flexible model with many parameters (nuisance parameters in the present context) on the one hand, and a parsimonious but restrictive model on the other. The following exploratory technique may be used as a guideline for selecting the functional form of the non-recall probability π. Assume π has the form

$$\pi(x) = b_1 I_{(x_1 < x \leq x_2)} + b_2 I_{(x_2 < x \leq x_3)} + \ldots + b_k I_{(x_k < x < \infty)}, \tag{2.20}$$

where k is large integer, $x_1, x_2, \ldots, x_k$ are a chosen set of time-points in increasing order and $b_1, b_2, \ldots, b_k$ are unspecified parameters taking values in the range $[0, 1]$. In view of (2.20), the likelihood (2.5) reduces to

$$L = \prod_{i=1}^{n} \Bigg[\left\{ \sum_{l=1}^{k} b_l (F_\theta(S_i - x_l) - F_\theta(S_i - x_{l+1})) \right\}^{1-\varepsilon_i} \left\{ f_\theta(T_i) \left(1 - \sum_{l=1}^{k} b_l I(S_i - x_{l+1} < T_i \leq S_i - x_l) \right) \right\}^{\varepsilon_i} \Bigg]^{\delta_i} [\bar{F}_\theta(S_i)]^{1-\delta_i}. \tag{2.21}$$

If the distribution of T is known, one can obtain the MLE of the parameters $b_1, b_2, \ldots, b_k$. The Hessian matrix with respect to the b_l's ($l = 1, 2, \ldots, k$) can easily be shown to be nonnegative definite. Therefore, there is a unique maximum of the likelihood function for these parameters. One can use Newton–Raphson iterative steps to determine the conditional MLE of the piecewise constant function π, for any given F_θ. While using a parametric form π_η, one can first estimate the MLEs $\hat{\theta}$ and $\hat{\eta}$ and then compare the plot of $\pi_{\hat{\eta}}$ with the plot of the conditional MLE of the piecewise constant version of π with large k, with F_θ held fixed at $F_{\hat{\theta}}$. This graphical comparison can be used to judge the suitability of the function π_η.

2.8 An example

For the data set described in Subsection 1.3, the landmark event is the onset of menarche in young and adolescent females. We used the Weibull model for menarcheal age and the exponential model for non-recall probability, as in the previous section, and used the three different methods mentioned in that section to estimate the parameters as well as the median of age at menarche. Table 2.5 gives a summary of the findings. Figure 2.2 shows the plot of the survival functions corresponding to the three sets of estimates.

The median estimated from our method is close to the median estimated from the status quo likelihood, but the confidence interval based on our estimate is narrower. The standard errors of the distributional parameters are also smaller. It is seen that the median estimated from the interval censoring likelihood, which ignores the informative nature of censoring, is quite different from the other two estimates. The corresponding 95% confidence interval does not have any overlap with other two confidence intervals. The survival functions estimated from the three models, shown in Figure 2.2, also show that the MLE based on interval censoring likelihood is very different from the other two MLE's. This occurrence may be attributed to the bias of this MLE, which is expected even when the sample size is large (see Sections 4.1 and 5).

Figure 2.3 shows the loci of upper and lower confidence limits for the probability of no menarche based on status quo MLE and recall data MLE. The latter pair of limits correspond to a narrower interval for any given age.

TABLE 2.5
Estimated parameters and median age at menarch from different methods for real data

Estimator	Estimate (standard error) α	β	η	Median	95% Confidence Interval of Median
MLE from status quo	10.74 (0.320)	12.17 (0.005)		11.76	(11.62,11.90)
MLE from interval censoring	11.80 (0.061)	12.65 (0.001)		12.25	(12.20,12.30)
MLE from proposed model	10.19 (0.090)	12.21 (0.001)	3.47 (0.140)	11.78	(11.72,11.84)

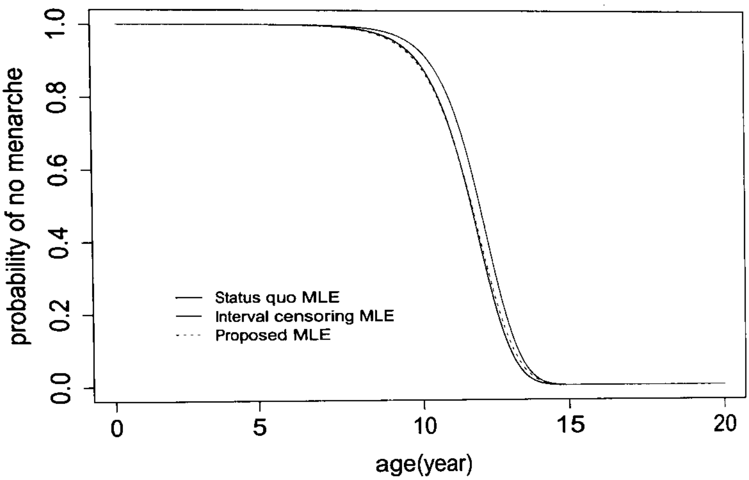

Fig 2.2: Survival plots for real data based on three methods.

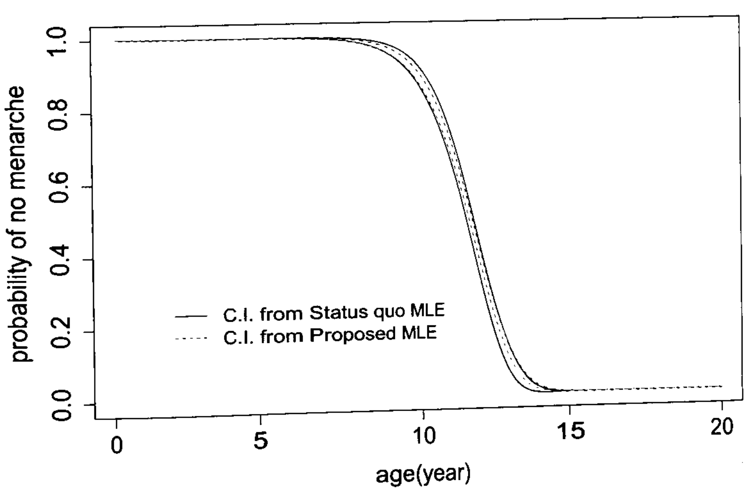

Fig 2.3: Confidence Interval for Probability of no Menarche based on two methods.

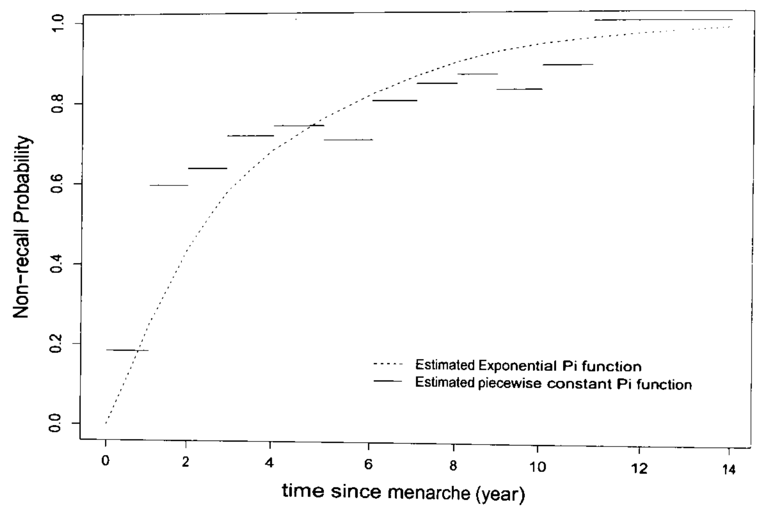

Fig 2.4: Plots of exponential and piecewise constant MLE of π.

In order to check how well the assumed parametric model fits the data, we use the chi-square goodness of fit test, by categorizing the triplet (S, V, δ) as follows.

The range of S is split into the two sets $\{7, 8, 9, 10, 11\}$ and $\{12, 13, 14, 15, 16, 17, 18, 19, 20, 21\}$;

the range of V is split into the three sets $\{0\}$, $(0, 1.5]$ and $(1.5, 11]$;

the range of δ has two points, 0 and 1, in any case.

The combinations of these categories produce 12 bins, while, there are three parameters to estimate. Thus, the null distribution should be χ^2 with 8 degrees of freedom. The p–value of the test statistics for the given data happens to be 0.11. Therefore, the model can be said to be appropriate.

As we mentioned in the last section, one can check adequacy of the functional form of π_η by comparing $\pi_{\hat{\eta}}$ with the conditional MLE of a piecewise constant function (2.20). We use segments of one year duration for this analysis. Note that, for the given data, the largest value of $S_i - T_i$ in a perfectly recalled case happens to be 10.88 years. With F chosen as Weibull and α and β fixed at the values reported in Table 2.4, we obtain the conditional MLE of the values of π in the different segments. Whenever $x_l \geqslant 11$, the likelihood (2.21) is an increasing function of b_l, and is maximized at $b_l = 1$. Therefore, the maximization is needed with respect to $b_1, \ldots, b_{11}$ only. Figure 2.4 shows the plot of the exponential $\pi_{\hat{\eta}}$ and the conditional MLE of the piecewise constant π in the range 0 to 14 years. The two plots are found to be close to each other. This supports the choice of the exponential form of π_η.

2.9 Concluding remarks

The thrust of this chapter has been to offer a realistic model for recall data on time-to-event, so that informative censoring can be handled. As the MLE obtained from the usual interval censoring likelihood is not consistent, and the MLE under the proposed model has performance in accordance with theoretical analysis , the latter should be preferred.

While the development of Section 2.4 has been on the basis of the joint distribution of the observable vector (S, V, δ), where V is as defined in (2.8), one could have alternatively worked with the conditional distribution of (V, δ), given S. The conditional density, which is proportional to (2.10), would have led to the same MLE. Also, the asymptotic results would continue to hold, with the additional condition that as the sample size goes to infinity, the proportion of observations with any particular value of the observation time goes to a fraction in the interval (0,1).

In many situations, data come from a heterogeneous population. Consider a subject having time of occurrence of the landmark event T_i, which is a single sample from a distribution $F_i(t;\theta)$ with density $f_i(t;\theta)$ and compact support, for $i = 1, \ldots, n$. Let these subjects be interviewed at times $S_1, \ldots, S_n$, respectively. Then the likelihood (conditional on the times of observation) is

$$\prod_{i=1}^{n}[\bar{F}_i(S_i;\theta)]^{1-\delta_i}\left[\{f_i(T_i;\theta)(1-\pi_\eta(S_i-T_i))\}^{\varepsilon_i}\left(\int_0^{S_i} f_i(u;\theta)\pi_\eta(S_i-u)du\right)^{1-\varepsilon_i}\right]^{\delta_i}. \tag{2.22}$$

The heterogeneity may be attributed to a set of covariates. Let Z_i be the r-dimensional vector of covariates for the ith subject, assumed to be independent of the censoring mechanism and the observation time. The distribution of T_i would depend on Z_i. Under an additive regression model, the distribution of T_i for an individual with covariate vector Z_i is

$$F_i(t;\theta) = F(t;\theta,\beta^T Z_i), \tag{2.23}$$

where β is the regression parameter. Substitution of (2.23) in (2.22) produces the conditional likelihood, given times of observation and covariate values. As for the properties of the MLE, note that in this case, instead of the triplet $Y_i = (S_i, V_i, \delta_i)$ considered in Section 2.4, we have the 4-tuple $Y_i = (S_i, V_i, \delta_i, Z_i)$ which contains all the information. The maximization of the product of the joint distribution of the Y_i's with respect to θ, η and β is equivalent to maximizing the said conditional likelihood with respect to the same. Therefore, we can use the standard asymptotic results for the MLE's.

There may also be heterogeneity in the function π for different individuals, possibly explained through covariates. Appropriate parametric modeling of the dependence on covariates can be done, and the ensuing inference problem can be handled in a similar manner.

As mentioned in Section 1.3, the recalled time-to-event can sometimes be erroneous. Skinner and Humphreys (1999), while working with data without instances of non-recall, has modeled erroneously recalled time-to-event as $t_i' = t_i k_i$, where t_i is the correct time-to-event and k_i is a multiplicative error of recall that is independent of t_i. Since k_i's are unobservable, they have used a mixed-effects regression model to account for erroneous recalls. A similar adjustment may be made in the term $f_i(T_i; \theta)$ of the likelihood (2.22), so that the situation of error in exact recall can be handled.

The data set analyzed here includes some cases, where the respondents recalled the event time with some uncertainty, that is, in the form of a range of dates. In the present analysis, we have treated these cases as instances of no recall. A more sophisticated modeling of partial recall may be contemplated. This issue is discussed in Chapter 5.

The estimation and the regression problems can also be considered without any distributional assumption. We consider the issues of non-parametric estimation and semi-parametric regression in Chapters 3 and 4, respectively.

Chapter 3

Nonparametric estimation of time-to-event distribution

3.1 Introduction

Consider recall based interval censored data on the time of occurrence of a landmark event, arising from a retrospective study. We have seen in Chapter 2 how the time-to-event distribution can be estimated parametrically in this situation. If no distributional assumption is made, it is tempting to use the likelihood for interval censored data, leading to the NPMLE obtained by Turnbull (1976). In fact, there are instances when the Turnbull estimator has actually been used for this type of data, where the object of study had been the distribution of age at reaching a developmental landmark (Aksglaede et al., 2009). However, as we have also noted in Section 1.2, the censoring mechanism for such data is likely to be informative. It may be recalled that the Turnbull estimator is not meant for informatively censored data, and its consistency in this situation is not guaranteed. The model (2.5) introduced in Chapter 2 provides a realistic framework for handling the special type of interval censored data in the present context. Therefore, a more meaningful non-parametric alternative to the Turnbull estimator would be the NPMLE obtained by maximizing (2.5), without using any distributional assumption. This is exactly what we propose to do in this chapter.

As in the parametric case, the performance benchmarks for the proposed estimation problem would be set by nonparametric maximizers of the current status likelihood (2.1) and the non-informative interval censoring likelihood (2.2), the latter being known as the Turnbull estimator.

In Section 3.2, we check the identifiability of the parameter of interest, namely the distribution function. In Section 3.3, the likelihood (2.5) is simplified, after a specific form of the non-recall function is assumed. In Section 3.4, we derive the NPMLE under the model, establish its existence and asymptotic uniqueness. In Section 3.5, we provide a self-consistency algorithm for computing the NPMLE. In Section 3.6, we present a computationally simpler alternative estimator that is asymptotically equivalent to the NPMLE. In Section 3.7, we discuss how the variance of the NPMLE can be estimated. In Section 3.8, we show that both the NPMLE and its approximation are consistent estimators of the underlying distribution under general conditions. Results of Monte Carlo simulations and an illustrative data analysis are reported in Sections 3.9 and 3.10, respectively. Some concluding remarks are provided in Section 3.11.

3.2 Identifiability of time-to-event distribution

Before embarking on developing a method of estimation, we need to visit the issue of identifiability of the function of interest. We rewrite the joint density of the observables given in (2.10), without assuming any particular family of distributions for the underlying time-to-event, as follows.

$$h(s,v,\delta) = \begin{cases} g(s)\bar{F}(s) & \text{if } v = 0 \text{ and } \delta = 0, \\ g(s)\int_0^s f(u)\pi_\eta(s-u)du & \text{if } v = 0 \text{ and } \delta = 1, \\ g(s)f(s-v)(1-\pi_\eta(v)) & \text{if } v > 0 \text{ and } \delta = 1, \\ 0 & \text{otherwise.} \end{cases} \tag{3.1}$$

Here, g is the density of G (distribution of observation times), f is the density of the time-to-event (corresponding to the distribution function F) and π_η is the non-recall probability expressed as a function of the time elapsed since the occurrence of the event. Note that G can be a continuous, discrete or mixed distribution, and g represents its Radon-Nikodym derivative with respect to an appropriate dominating measure. The parameter of interest is the function F. We address the question as to whether the functions F, π_η and G are identifiable from h, in the next theorem.

Theorem 3.1. *(a) The distribution G is completely identifiable from h.*

(b) If G has an absolutely continuous component over the support of F, then π_η and F are identifiable from h.

(c) If G has probability mass only over the space of integers and the function π_η comes from a family $\mathcal{P}$ satisfying the condition: '$\pi_1, \pi_2 \in \mathcal{P}$ implies that $(1-\pi_2)/(1-\pi_1)$ is not periodic with period one', then π_η and F are identifiable from h.

Proof. (a) We have, from (3.1) (with $v > 0$ and $\delta = 1$),

$$h(s, v, 1) = g(s)f(s-v)(1-\pi_\eta(v)),$$

that is,

$$1 - \pi_\eta(v) = \frac{h(s,v,1)}{g(s)f(s-v)} \qquad \forall s, v \text{ s.t. } v < s. \tag{3.2}$$

By substituting the above expression in (3.1) for $v = 0$ and $\delta = 1$ and simplifying the equation, we have

$$F(s) = \frac{h(s,0,1) + \int_0^s h(s, s-u, 1)du}{g(s)}. \tag{3.3}$$

By substituting the above expression of $F(s)$ in (3.1) with $v = 0$ and $\delta = 0$, we obtain

$$g(s) = h(s,0,0) + h(s,0,1) + \int_0^s h(s, s-u, 1)du. \tag{3.4}$$

The above identity holds over the support of G irrespective of whether G is a discrete, continuous or mixed distribution. The identifiability of G follows.

(b) By substituting (3.4) in (3.3), we have

$$F(s) = \frac{h(s,0,1) + \int_0^s h(s, s-u, 1)du}{h(s,0,0) + h(s,0,1) + \int_0^s h(s, s-u, 1)du}. \tag{3.5}$$

If G has an absolutely continuous component over the support of F, for every s and all real valued $v < s$, we have from (3.2),

$$\pi_\eta(v) = 1 - \frac{h(s,v,1)}{g(s)f(s-v)}. \tag{3.6}$$

Thus, (3.6) together with (3.4) and (3.5) identifies F and π_η completely.

(c) For the sake of contradiction, let us assume there are two pairs of choices of f and π_η, say (f_1, π_1) and (f_2, π_2), such that their substitution in the right hand side of (3.1) produces the same function. If we follow the steps leading to (3.2) for these two pairs of functions, then we have, for all integers s and all $v < s$,

$$f_1(s-v)(1-\pi_1(v)) = f_2(s-v)(1-\pi_2(v)).$$

Hence,

$$\frac{f_1(v)}{f_2(v)} = \frac{1-\pi_2(s-v)}{1-\pi_1(s-v)} \qquad \forall s, v \text{ s.t. } v < s. \tag{3.7}$$

Since the above identity holds for all integers s, we can write

$$\frac{1-\pi_2(s-v)}{1-\pi_1(s-v)} = \frac{1-\pi_2(1-v)}{1-\pi_1(1-v)} \quad \text{for all integer } s \text{ and all } v < s. \tag{3.8}$$

The above equation implies that the function $(1-\pi_1)/(1-\pi_2)$ is periodic over the relevant domain with period 1, which contradicts the assumption. Therefore, the pair (f, π_η) is uniquely defined for any given h.

□

The following example shows that if G is a discrete distribution over the set of integers and yet the condition given in part (c) of the above theorem does not hold, then f may not be identifiable from h.

Example 1. Let $\pi_1 = 0.5$ and π_2 be a periodic function with period one, defined over the interval (0,1] by the equation $\pi_2(v) = v$, and . Let $f_2(v) = 1/(t_{max} - t_{min})$ for $v \in [t_{min}, t_{max}]$, and $f_1 = f_2((1-\pi_2)/(1-\pi_1))$, defined over the same interval. Let g be any probability mass function defined over the space of positive integers. It may be verified that either of the triplets of functions (g, f_1, π_1) and (g, f_2, π_2), when substituted in (3.1), produce the following h:

$$h(1, v, \delta) = \begin{cases} g(s)(1 - \frac{s}{t_{max} - t_{min}}) & \text{if } v = 0 \text{ and } \delta = 0, \\ g(s)\frac{0.5s}{t_{max} - t_{min}} & \text{if } v = 0 \text{ and } \delta = 1, \\ g(s)\frac{1 - (v - [v])}{t_{max} - t_{min}} & \text{if } v > 0 \text{ and } \delta = 1, \\ 0 & \text{otherwise.} \end{cases}$$

The above counterexample involves a forgetting function, which is not a monotonically non-decreasing function. In Section (2.9), we have observed that it is possible to interpret π_η as a distribution function in a restrictive model, where the probability of forgetting increases with time. Specifically, π_η would then be the distribution of the time taken to forget the event. We now present another counterexample to show that F can be non-identifiable even when π_η is a continuous distribution function.

Example 2. Let π_1 be the distribution function of the uniform distribution over $[t_{min}, t_{max}]$, f_2 be the density of the same distribution,

$$f_1(t) = \left(1 + (-1)^{[2t+1/2]} \cdot \frac{\left(t - \frac{[2t+1/2]}{2}\right)}{t_{max}}\right) \cdot \left(\frac{1}{t_{max} - t_{min}}\right)$$

and $\pi_2(t) = 1 - (1 - \pi_1(t))f_1(t)/f_2(t)$. Let g be any probability mass function over the space of positive integers. It can be verified that either of triplets of functions (g, f_1, π_1) and (g, f_2, π_2), produce the same h. In this example, π_1 and π_2 can be interpreted as distribution functions of the forgetting time. Figures 3.1 and 3.2 show the plots of the two density functions of time-to-event and the corresponding forgetting time distribution functions, in the special case when $t_{min} = 8$ and $t_{max} = 16$.

We now proceed with the problem of estimation, after assuming that either of the conditions given in part (b) and (c) of Theorem 3.1 are satisfied.

3.3 Reduction of the problem

It is known that nonparametric maximization of the likelihood (2.6) leads to the Kaplan-Meier estimator (Kaplan and Meier, 1958), while maximization of (2.2)

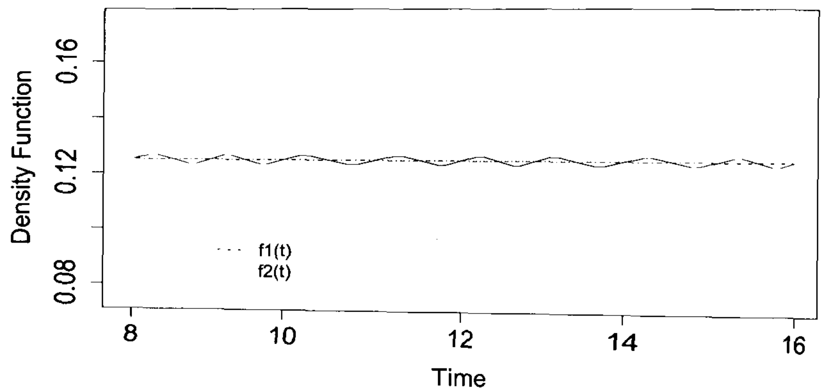

Fig 3.1: Density functions of time-to-event.

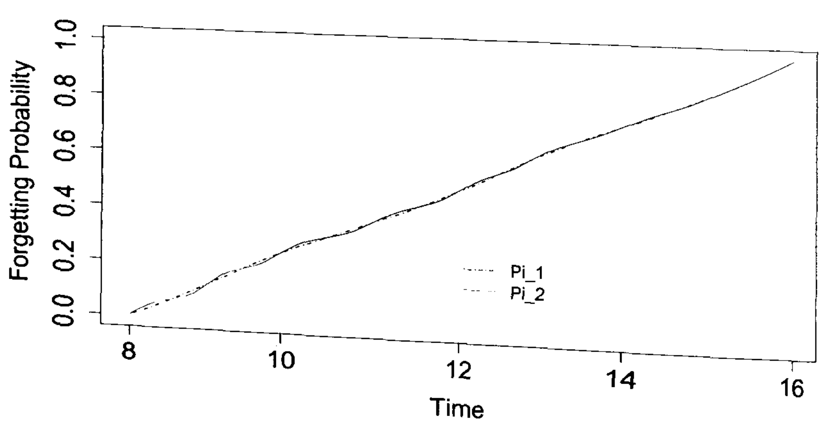

Fig 3.2: Forgetting probability functions.

or (2.1) produces the Turnbull estimator (Turnbull, 1976) or a special case of it. On the other hand, the likelihood (2.5) is difficult to maximize because of the integral contained in the expression. In order to simplify it, we assume that the function π_η in (2.5) is piecewise constant, having the form

$$\pi_\eta(x) = b_1 I_{(x_1<x\leq x_2)} + b_2 I_{(x_2<x\leq x_3)} + \ldots + b_k I_{(x_k<x<\infty)}, \tag{3.9}$$

where $0 = x_1 < x_2 < \cdots < x_k$; $0 < b_1, b_2, \ldots, b_k \leq 1$. We also assume that k and $x_1, x_2, \ldots, x_k$ are known, while $b_1, b_2, \ldots, b_k$ constitute the unknown parameter η. Note that it is possible to constrain the elements of the vector parameter η to be in increasing order, so that π_η is a non-decreasing function. Such a choice would correspond to the general perception that memory fades with time. However, we do not use this constraint here.

When (3.9) holds, and the parametric functions F_θ, $\bar{F}_\theta$ and f_θ are replaced by the general distribution F, its survival function $\bar{F}$ and its density f, respectively, the likelihood (2.5) reduces to

$$L = \prod_{i=1}^{n} [\bar{F}(S_i)]^{1-\delta_i} \left[\left\{ f(T_i) \left(1 - \sum_{l=1}^{k} b_l I_{(W_{l+1}(S_i)<T_i\leq W_l(S_i))} \right) \right\}^{\varepsilon_i} \cdot \left\{ \sum_{l=1}^{k} b_l \big(F(W_l(S_i)) - F(W_{l+1}(S_i)) \big) \right\}^{1-\varepsilon_i} \right]^{\delta_i}, \tag{3.10}$$

where $W_l(S_i) = (S_i - x_l) \vee t_{min}$ for $l = 1, \ldots, k$ and $W_{k+1}(S_i) = t_{min}, i = 1, 2, \ldots, n$. Note that

$$W_{k+1}(S_i) \leq W_k(S_i) \leq W_{k-1}(S_i) \leq \cdots \leq W_1(S_i) \leq t_{max}, \tag{3.11}$$

Depending on the value of S_i, some of the above inequalities may in fact be equalities. Specifically, if l is an index such that $S_i - x_{l+1} \leq t_{min} < S_i - x_l$, then $t_{min} = W_{k+1}(S_i) = \cdots = W_{l+1}(S_i)$. Further, if l is an index such that $S_i - x_{l+1} < t_{max} \leq S_i - x_l$, then $W_l(S_i) = \cdots = W_1(S_i)$. The remaining equalities would be strict.

Anticipating point masses at T_i whenever $\delta_i\varepsilon_i = 1$, the likelihood (3.10) can be rewritten as

$$L = \prod_{i=1}^{n} [\bar{F}(S_i)]^{1-\delta_i} \left[\left\{ \sum_{l=1}^{k} b_l \left(F(W_l(S_i)) - F(W_{l+1}(S_i)) \right) \right\}^{1-\varepsilon_i} \right.$$
$$\left. \cdot \left\{ \left(F(T_i) - F(T_i-) \right) \left(1 - \sum_{l=1}^{k} b_l I_{(W_{l+1}(S_i) < T_i \le W_l(S_i))} \right) \right\}^{\varepsilon_i} \right]^{\delta_i}. \quad (3.12)$$

The simple form of the above likelihood paves the way for estimation.

3.4 Nonparametric MLE

The likelihood (3.12) involves probabilities assigned to intervals of the type $[t, t_{max}]$ or $(t, t_{max}]$, as per the baseline probability distribution. Since these intervals have overlap, we try to write them as unions of some disjoint intervals. Let $\mathcal{I}_1$, $\mathcal{I}_2$ and $\mathcal{I}_3$ be sets of indices i (between 1 and n) that satisfy the conditions $\delta_i = 0$, $\delta_i\varepsilon_i = 1$ and $\delta_i(1 - \varepsilon_i) = 1$, respectively. Consider the intervals

$$\begin{aligned}
A_i &= (S_i, t_{max}] && \text{for } i \in \mathcal{I}_1; \\
A_i &= [T_i, t_{max}] && \text{for } i \in \mathcal{I}_2; \\
A'_i &= (T_i, t_{max}] && \text{for } i \in \mathcal{I}_2; \\
A_{il} &= \begin{cases} (W_l(S_i), t_{max}], & l = 1, \ldots, k, \\ [W_l(S_i), t_{max}], & l = k+1, \end{cases} && \text{for } i \in \mathcal{I}_2 \cup \mathcal{I}_3.
\end{aligned} \quad (3.13)$$

and the sets

$$\begin{aligned}
\mathcal{A}_1 &= \{A_i : \ i \in \mathcal{I}_1\}; \\
\mathcal{A}_2 &= \{A_i \setminus A'_i : \ i \in \mathcal{I}_2\}; \\
\mathcal{A}_3 &= \{A'_i : \ i \in \mathcal{I}_2\}; \\
\mathcal{A}_4 &= \{A_{i(l+1)} \setminus A_{il} : \ 1 \le l \le k \text{ and } i \in \mathcal{I}_3\}.
\end{aligned} \quad (3.14)$$

As the underlying distribution F is absolutely continuous, the elements of $\mathcal{A}_2$ and $\mathcal{A}_3$ are distinct with probability 1. Let n_i be the cardinality of $\mathcal{I}_i$, $i = 1, 2, 3$. We arrange the singleton elements of $\mathcal{A}_2$ in increasing order, and denote them as $B_1, B_2, \ldots, B_{n_2}$. We also arrange the elements of $\mathcal{A}_3$ in the corresponding order and denote them as $B_{n_2+1}, B_{n_2+2}, \ldots, B_{2n_2}$. We then collect the unique elements of $\mathcal{A}_1 \cup \mathcal{A}_4$ that are distinct from $B_1, B_2, \ldots, B_{2n_2}$, and denote them as

$B_{2n_2+1}, B_{2n_2+2}, \ldots, B_M$. Observe that the collection $B_1, B_2, \ldots, B_M$ consists of the distinct elements of $\mathcal{A}_1 \cup \mathcal{A}_2 \cup \mathcal{A}_3 \cup \mathcal{A}_4$, arranged in a particular order. Denote the non-empty subsets of the index set $\{1, 2, \ldots, M\}$ by $s_1, s_2, \ldots, s_{2^M-1}$. Define

$$I_r = \left\{ \bigcap_{i \in s_r} B_i \right\} \bigcap \left\{ \bigcap_{i \notin s_r} B_i^c \right\} \quad \text{for } r = 1, 2, \ldots, 2^M - 1. \tag{3.15}$$

Some of the I_r's may be empty sets, denoted here by ϕ. Let

$$\mathcal{C} = \{s_r : I_r \neq \phi, 1 \leq r \leq 2^M - 1\}, \tag{3.16}$$

$$\mathcal{A} = \{I_r : I_r \neq \phi, 1 \leq r \leq 2^M - 1\}. \tag{3.17}$$

It can be verified that the elements of $\mathcal{A}$ are distinct and disjoint.

Note that each of the intervals $B_1, \ldots, B_M$ is a union of disjoint sets that are members of $\mathcal{A}$. For any Borel set A, suppose $P(A)$ is the probability assigned to A as per the probability distribution F. Let $p_r = P(I_r)$, for $I_r \in \mathcal{A}$. Then the likelihood (3.12) reduces to

$$\begin{aligned} L = \prod_{i \in \mathcal{I}_1} \left(\sum_{\substack{r: I_r \subseteq A_i \\ s_r \in \mathcal{C}}} p_r \right) &\times \prod_{i \in \mathcal{I}_2} \left(1 - \sum_{l=1}^{k} b_l I_{(T_i \in A_{i(l+1)} \setminus A_{il})} \right) \\ &\cdot \left[\left(\sum_{\substack{r: I_r \subseteq A_i \\ s_r \in \mathcal{C}}} p_r \right) - \left(\sum_{\substack{r: I_r \subseteq A_i' \\ s_r \in \mathcal{C}}} p_r \right) \right] \\ \times \prod_{i \in \mathcal{I}_3} \left[\sum_{l=1}^{k} b_l \left\{ \left(\sum_{\substack{r: I_r \subseteq A_{i(l+1)} \\ s_r \in \mathcal{C}}} p_r \right) - \left(\sum_{\substack{r: I_r \subseteq A_{il} \\ s_r \in \mathcal{C}}} p_r \right) \right\} \right]&, \end{aligned} \tag{3.18}$$

which simplifies to

$$\begin{aligned} L = \prod_{i \in \mathcal{I}_1} \left(\sum_{\substack{r: I_r \subseteq A_i \\ s_r \in \mathcal{C}}} p_r \right) &\times \prod_{i \in \mathcal{I}_3} \left[\sum_{l=1}^{k} b_l \left(\sum_{\substack{r: I_r \subseteq A_{i(l+1)} \setminus A_{il} \\ s_r \in \mathcal{C}}} p_r \right) \right] \\ \times \prod_{i \in \mathcal{I}_2} \left(1 - \sum_{l=1}^{k} b_l I_{(T_i \in A_{i(l+1)} \setminus A_{il})} \right) &\cdot \left(\sum_{\substack{r: I_r \subseteq A_i \setminus A_i' \\ s_r \in \mathcal{C}}} p_r \right). \end{aligned} \tag{3.19}$$

Thus, maximizing the likelihood (3.12) is equivalent to maximizing the likelihood (3.19) with respect to p_r for $s_r \in \mathcal{C}$.

There is a partial order among the members of $\mathcal{C}$ in the sense that some sets are contained in others. We consider the following subsets of $\mathcal{C}$.

$$\begin{aligned} \mathcal{C}_1 &= \{s : s \in \mathcal{C};\ \text{there is another element } s' \in \mathcal{C},\ \text{such that } s \subset s'\}, \\ \mathcal{C}_2 &= \{s : s \in \mathcal{C};\ \text{there is another element } s' \in \mathcal{C},\ \text{such that} \\ &\qquad s' \backslash (s \cap s') \text{ consists of a singleton } j \text{ and } s \backslash (s \cap s') = \{j + n_2\}\}, \\ \mathcal{C}_0 &= \mathcal{C} \backslash (\mathcal{C}_1 \cup \mathcal{C}_2). \end{aligned} \tag{3.20}$$

Our next result shows that the maximization of the likelihood can be restricted to $\mathcal{C}_0$.

Theorem 3.2. *Maximizing the likelihood* (3.19) *with respect to p_r for $s_r \in \mathcal{C}$ is equivalent to maximizing it with respect to p_r for $s_r \in \mathcal{C}_0$, i.e.,*

$$\max_{p_r : p_r \in [0,1], \sum_{s_r \in \mathcal{C}} p_r = 1} L = \max_{p_r : p_r \in [0,1], \sum_{s_r \in \mathcal{C}_0} p_r = 1} L.$$

Proof. By definitions of $\mathcal{C}$ and $\mathcal{C}_0$, we can rewrite the likelihood (3.19) as follows.

$$\begin{aligned} L = &\prod_{i \in \mathcal{I}_3} \left[\sum_{l=1}^{k} b_l \left(\sum_{\substack{r: I_r \subseteq A_{i(l+1)} \backslash A_{il} \\ s_r \in \mathcal{C} \backslash \mathcal{C}_0}} p_r + \sum_{\substack{r: I_r \subseteq A_{i(l+1)} \backslash A_{il} \\ s_r \in \mathcal{C}_0}} p_r \right) \right] \\ &\times \prod_{i \in \mathcal{I}_2} \left(1 - \sum_{l=1}^{k} b_l I_{(T_i \in A_{i(l+1)} \backslash A_{il})} \right) \cdot \left(\sum_{\substack{r: I_r \subseteq A_i \backslash A_i' \\ s_r \in \mathcal{C} \backslash \mathcal{C}_0}} p_r + \sum_{\substack{r: I_r \subseteq A_i \backslash A_i' \\ s_r \in \mathcal{C}_0}} p_r \right) \\ &\times \prod_{i \in \mathcal{I}_1} \left(\sum_{\substack{r: I_r \subseteq A_i \\ s_r \in \mathcal{C} \backslash \mathcal{C}_0}} p_r + \sum_{\substack{r: I_r \subseteq A_i \\ s_r \in \mathcal{C}_0}} p_r \right) \end{aligned} \tag{3.21}$$

For any $s_r \in \mathcal{C} \backslash \mathcal{C}_0$, let $\mathcal{A}_r = \{I_{r'} : s_{r'} \in \mathcal{C}_0, s_r \subset s_{r'}\}$. By the construction of $\mathcal{C}_0$, $\mathcal{A}_r$ is a non-empty set. The elements of $\mathcal{A}_r$ are disjoint sets consisting of unions of intervals, which are subsets of $[t_{min}, t_{max}]$. Let I_{r*} be that member of $\mathcal{A}_r$ which satisfies the condition 'there is $\alpha \in I_{r^*}$ such that $\alpha < \beta$ whenever $\beta \in I_{r^\dagger}$ for any $I_{r^\dagger} \in \mathcal{A}_r$'. We shall show that by shifting mass from any I_r to I_{r^*}, there will be no reduction in the contribution of any individual to the likelihood (3.21).

We now check the effect of shifting mass on the likelihood (3.21). The change in the likelihood would only be through the sets B_j such that $j \in s_{r^*} \backslash s_r$. Further, it is easy to see that two such sets cannot affect the contribution of the same individual to the likelihood.

CASE (i). For any $j \in s_{r^*} \backslash s_r$, let i_j be such that $B_j = A_{i_j}$ and $i_j \in \mathcal{I}_1$. Since $I_{r^*} \subseteq A_{i_j}$ but $I_r \not\subseteq A_{i_j}$, the factor contributed by individual i_j in the likelihood is increased when mass is shifted from I_r to I_{r^*}.

CASE (ii). For any $j \in s_{r^*} \backslash s_r$, let i_j be such that $B_j = A_{i_j} \backslash A'_{i_j}$ and $i_j \in \mathcal{I}_2$. In this case, $I_{r^*} \subseteq A_{i_j}$ but $I_{r^*} \not\subseteq A'_{i_j}$. By construction, $B_{n_2+j} = A'_{i_j}$, which is disjoint with B_j. Since I_{r^*} is not a null set, we must have $n_2 + j \notin s_{r^*}$ and hence $n_2 + j \notin s_r$. It follows that I_r is not contained in B_j or B_{n_2+j}. Thus, $I_r \not\subseteq A_{i_j}$ and $I_r \not\subseteq A'_{i_j}$ and any shift of mass from I_r to I_{r^*} increases the contribution of the i_jth individual to the likelihood.

CASE (iii). For any $j \in s_{r^*} \backslash s_r$, let i_j be such that $B_j = A_{i_j(l+1)} \backslash A_{i_j l}$ for some $l = 1, \ldots, k$ and $i_j \in \mathcal{I}_3$. Any transfer of mass from I_r to I_{r^*} increases the contribution of the i_jth individual to the likelihood, since $I_{r^*} \in A_{i_j(l+1)}$ and $I_{r^*} \notin A_{i_j l}$, whereas I_r is not in either of them.

It may be concluded that maximizing L can be restricted to $\{p_r : s_r \in \mathcal{C}_0\}$. □

It follows from the above theorem that the likelihood has the same maximum value whether s_r is chosen from the class $\mathcal{C}$ or $\mathcal{C}_0$. Therefore, we can replace $\mathcal{C}$ by $\mathcal{C}_0$ in (3.19).

Let us relabel the intervals I_j, $s_j \in \mathcal{C}_0$, by $J_1, J_2, \ldots, J_v$. Further, let $\mathcal{A}_0 = \{J_1, J_2, \ldots, J_v\}$ and $q_j = P(J_j)$ for $j = 1, 2, \ldots, v$. If the likelihood (3.19) is rewritten with the condition $s_r \in \mathcal{C}$ replaced by the equivalent condition $I_r \in \mathcal{A}$, then Theorem 3.2 shows that the latter condition can be replaced by $I_r \in \mathcal{A}_0$. In other words, maximizing the likelihood (3.19) is equivalent to maximizing

$$\begin{aligned} L(p,\eta) &= \prod_{i\in\mathcal{I}_1}\left(\sum_{j:J_j\subseteq A_i} q_j\right) \times \prod_{i\in\mathcal{I}_3}\left[\sum_{l=1}^{k} b_l\left(\sum_{j:J_j\subseteq A_{i(l+1)}\setminus A_{il}} q_j\right)\right] \\ &\quad\times \prod_{i\in\mathcal{I}_2}\left(1-\sum_{l=1}^{k} b_l I_{(T_i\in A_{i(l+1)}\setminus A_{il})}\right)\cdot\left(\sum_{j:J_j\subseteq A_i\setminus A_i'} q_j\right) \\ &= \prod_{i=1}^{n}\left(\sum_{j=1}^{v}\alpha_{ij}q_j\right), \end{aligned} \tag{3.22}$$

with respect to the vector parameters $p = (q_1, q_2, \ldots, q_v)^T$ and $\eta = (b_1, \ldots, b_k)^T$, subject to the restrictions $\sum_{j=1}^{v} q_j = 1, 0 \le q_1, \ldots, q_v \le 1$, and $0 \le b_1 \le \cdots b_k \le 1$, where

$$\alpha_{ij} = \begin{cases} I_{(J_j\subseteq A_i)} & \text{if } i\in\mathcal{I}_1, \\ 1-\sum_{l=1}^{k} b_l . I_{(T_i\in A_{i(l+1)}\setminus A_{il})} . I_{(J_j\subseteq A_i\setminus A_i')} & \text{if } i\in\mathcal{I}_2, \\ \sum_{l=1}^{k} b_l . I_{(J_j\subseteq A_{i(l+1)}\setminus A_{il})} & \text{if } i\in\mathcal{I}_3, \end{cases} \tag{3.23}$$

for $i = 1, \ldots, n$, and $j = 1, \ldots, v$.

Now consider the set $\mathcal{A}_2 = \{\{T_i\},\ i \in \mathcal{I}_2\}$ as defined in (3.14). The cardinality of this set is the same as that of $\mathcal{I}_2$, which is n_2 (see notation defined after equation (3.14)). The task of maximizing the above likelihood can be simplified further through the following result, which is interesting by its own right.

Theorem 3.3. *The set $\mathcal{A}_2$ is contained in the set $\mathcal{A}_0$ almost surely. Further, if G is a discrete distribution with finite support, then the probability of $\mathcal{A}_0$ being equal to $\mathcal{A}_2$ goes to one as $n \to \infty$.*

Proof. Let $i \in \mathcal{I}_2$ and the index j_i be such that $s_{j_i} = \{j : T_i \in B_j\}$. Since each time-to-event has an absolutely continuous distribution, the recalled times T_i, $i \in \mathcal{I}_2$ are distinct with probability 1. Therefore, $\{T_i\} \in \{B_1, B_2, \ldots, B_{n_2}\}$ almost surely. It follows that $T_i \in I_{j_i} \subseteq \{T_i\}$, i.e., $I_{j_i} = \{T_i\}$ with probability 1. It is also easy to see $s_{j_i} \in \mathcal{C}_0$ and hence $\mathcal{A}_2 \subseteq \mathcal{A}_0$ almost surely.

The interview times are discrete valued with finite domain; $x_1, x_2, \ldots, x_k$ are also finite. Therefore, even when n is large, there is at most a finite number (say

N) of distinct sets of the form

$$A_s = \{\bigcap_{i\in s} B_i\} \bigcap \{\bigcap_{i\in \mathcal{I}_1\cup\mathcal{I}_3\setminus s} B_i^c\},$$

where $s \subseteq \mathcal{I}_1 \cup \mathcal{I}_3$. Denote by $s^{(1)}, s^{(1)}, \ldots, s^{(N)}$ the index sets corresponding to the N distinct sets described above.

Consider a member of $\mathcal{A}_0$, say I_s, where s is a subset of $\{1, 2, \ldots, n\}$. If $s \subseteq \mathcal{I}_2$, then it is already a singleton. If not, it can be written as $s^{(j)} \cup (s\setminus s^{(j)})$, with $s^{(j)} \subseteq \mathcal{I}_1 \cup \mathcal{I}_3$ and $s\setminus s^{(j)} \subseteq \mathcal{I}_2$ for some $j \in \{1, 2, \ldots, N\}$. Let us consider three further special cases.

CASE (i). Let $s = s^{(j)} \cup \{r\}$ for $r \in \mathcal{I}_2$. In this case, I_s is either a singleton or a null set. If it is a null set, then it cannot be a member of $\mathcal{A}_0$. Thus, Case (a) contributes only singletons to $\mathcal{A}_0$.

CASE (ii). Let $s = s^{(j)} \cup \{r_1, r_2, \ldots, r_p\}$, for $r_1, r_2, \ldots, r_p \in \mathcal{I}_2$ when $p > 1$. In this case I_s is either a singleton or a null set. Since the absolute continuity of the time-to-event distribution almost surely precludes coincidence of two sample values (say, T_{r_1} and T_{r_2}), I_s is a null set with probability 1. In summary, Case (ii) cannot contribute anything other than a singleton to $\mathcal{A}_0$.

CASE (iii). Let $s = s^{(j)}$. The probability that a specific individual (say, the i-th one) has the landmark event at an age contained in $A_{s^{(j)}}$ is

$$P(T_i \in A_{s^{(j)}}, i \in \mathcal{I}_2).$$

Since this quantity is strictly positive, the probability that none of the n individuals have had the landmark event in $A_{s^{(j)}}$ and recalled the date is

$$\left(1 - P(T_i \in A_{s^{(j)}}, i \in \mathcal{I}_2)\right)^n,$$

which goes to zero as $n \to \infty$. Thus, the probability that there is $i \in \mathcal{I}_2$ such that $T_i \in A_{s^{(j)}}$ goes to one as $n \to \infty$. Therefore, $I_{s^{(j)}\cup\{i\}} = I_{s^{(j)}} \cap \{T_i\}$ is non-null. It follows that $P[I_s \notin \mathcal{A}_0]$ goes to one.

The statement of the theorem follows by combining the three cases. □

We are now ready for the next result regarding the existence and uniqueness of the NPMLE. The uniqueness is established probabilistically under the condition that n_2, the number of cases with exact recall, goes to infinity.

Theorem 3.4. *The likelihood* (3.22) *has a maximum. Further, if G is a discrete distribution with finite support, then the probability that it has a unique maximum goes to one, as $n_2 \to \infty$.*

Proof. From (3.22), the log-likelihood is given by

$$\ell(p,\eta) = \sum_{i=1}^{n} \left(\ln \left(\sum_{j=1}^{v} \alpha_{ij} q_j \right) \right) \tag{3.24}$$

Consider maximization of $\ell(p,\eta)$ periodically with respect to p and η. Given $(p^{(n)}, \eta^{(n)})$, the iterate at the nth stage, define the next iterate $(p^{(n+1)}, \eta^{(n+1)})$ by

$$\eta^{(n+1)} = \begin{cases} \eta^{(n)} & \text{if } n \text{ is even,} \\ \underset{\eta \in S_2}{\operatorname{argmax}}\, \ell(p^{(n)}, \eta) & \text{if } n \text{ is odd,} \end{cases} \quad p^{(n+1)} = \begin{cases} p^{(n)} & \text{if } n \text{ is odd,} \\ \underset{p \in S_1}{\operatorname{argmax}}\, \ell(p, \eta^{(n)}) & \text{if } n \text{ is even,} \end{cases} \tag{3.25}$$

where $S_1 = \{p : \sum_{j=1}^{v} q_j = 1,\ 0 \le q_1, \ldots, q_v \le 1\}$ and $S_2 = \{\eta : 0 \le b_1 \le \cdots \le b_k \le 1\}$. We shall show that the functions $\ell(p, \cdot)$ and $\ell(\cdot, \eta)$ are concave over the convex sets S_1 and S_2, respectively, so that there exists a maximum at each iteration. Thus, in each stage there is an increase in the likelihood (3.22), which is bounded by $(kv)^n$, and the sequence of partially maximized likelihoods converges. Under the conditions stated in the theorem, we shall also show that the objective function is strictly concave, so that the maximum at each stage is unique, with probability tending one as n_2 goes to infinity. Finally, since $S_1 \times S_2$ is a closed set, the sequence of maxima obtained at successive stages converges to a unique limit, with probability tending to one.

Let $\boldsymbol{B}$ be an $n \times v$ matrix with β_{ij} in the ijth position. For fixed η, the partial derivative of (3.24) with respect to p is

$$\frac{\partial \ell}{\partial p} = \sum_{i=1}^{n} \frac{B_i}{B_i^T p}$$

where B_i is the ith row of $\boldsymbol{B}$ matrix. The second derivative or the Hessian is

$$\frac{\partial \ell}{\partial p \partial p^T} = -\sum_{i=1}^{n} \frac{B_i B_i^T}{(B_i{}^T p)^2} \tag{3.26}$$

which is a non-positive definite matrix. Hence ℓ is a concave function over a convex and bounded domain, which ensures the existence of a maximum (de la Fuente, 2000; Simon and Blume, 1994). Now, we need to show that the probability of the Hessian matrix being negative definite goes to one. It is enough to show for any vector $u \neq 0$,

$$P\left(\sum_{i=1}^{n} \frac{(B_i^T u)^2}{(B_i{}^T p)^2} = 0\right) \rightarrow 0.$$

In other words, we need to show that for any arbitrary vector $u \neq 0$,

$$P\left(B_i^T u = 0 \quad \forall i\right) = P(\boldsymbol{B}u = 0) \rightarrow 0. \tag{3.27}$$

It is clear from (3.23) that for an individual (say i) having exactly recalled age at landmark event, B_i has only one non-zero element. In this situation, the equation $B_i^T u = 0$ implies that the corresponding element of u is zero. Further, Theorem 3.3 shows that, with probability tending to one, the columns of $\boldsymbol{B}$ correspond only to singleton members of $\mathcal{A}_0$ associated with individuals recalling age at event exactly. Therefore, with probability tending to one, the event $\boldsymbol{B}u = 0$ coincides with the event $u = 0$.

For fixed p, the first derivative of (3.24) with respect to η is

$$\frac{\partial \ell}{\partial \eta} = \sum_{i=1}^{n} \frac{\mathbf{A}_i p}{B_i{}^T p}$$

where $\mathbf{A}_i$ is the $k \times v$ matrix with the $(l,j)^{\text{th}}$ element given by $\frac{\partial \beta_{ij}}{\partial b_l}$.

The Hessian with respect to η is

$$\frac{\partial \ell}{\partial \eta \partial \eta^T} = -\sum_{i=1}^{n} \left(B_i^T p\right)^{-2} \mathbf{A}_i p p^T \mathbf{A}_i^T \tag{3.28}$$

which is non-positive definite matrix. Hence ℓ is a concave function over a convex domain, it ensures the existence of a maximum (de la Fuente, 2000; Simon and Blume, 1994).

In order to prove the negative definiteness of the Hessian with probability tending to one, we need to show that for any arbitrary vector $\nu \neq 0$,

$$P\left(\nu^T \mathbf{A}_i p = 0 \quad \forall i\right) \to 0. \tag{3.29}$$

From (3.23), it follows that for $i \in \mathcal{I}_2$,

$$\mathbf{A}_i p = -\left(\sum_{j=1}^{v} q_j \cdot I(J_j \subset A_i)\right) \left(I(T_i \in A_{i1}), \ldots, I(T_i \in A_{ik})\right)^T, \tag{3.30}$$

which is a vector with a non-zero element exactly at one place. The condition $\nu^T \mathbf{A}_i p = 0$ is equivalent to the requirement that the element of ν corresponding to the non-zero element of $\mathbf{A}_i p$ is zero. On the other hand, as $n_2 \to \infty$,

$$P\Big(\sum_{i\in\mathcal{I}_2} I\left((S_i - T_i) \in [x_l, x_{l+1}]\right) = 0\Big) = \Big[P\Big((S_i - T_i) \in [x_l, x_{l+1}] | \delta_i \varepsilon_i = 1\Big)\Big]^{n_2} \to 0 \quad \forall l.$$

Thus, for all $l = 1, \ldots, k$, there is at least one $i \in \mathcal{I}_2$ such that $T_i \in A_{il}$, with probability tending to one. Therefore, the condition $\nu^T \mathbf{A}_i p = 0 \quad \forall i \in \mathcal{I}_2$ reduces, with probability tending to one, to the requirement that all the elements of ν are zero. Therefore, for $\nu \neq 0$, we have

$$P\left(\nu^T \mathbf{A}_i p = 0, \quad \forall i\right) \leq P\left(\nu^T \mathbf{A}_i p = 0, \quad \forall i \in \mathcal{I}_2\right) \to 0.$$

Thus, the probability that the Hessian matrix defined in (3.28) is negative definite goes to one. This completes the proof. □

3.5 Self-consistency approach for estimation

Following the work of Efron (1967) on computing the Kaplan-Meier estimator (Kaplan and Meier, 1958) through a self consistency algorithm and similar work by Turnbull (1976) in the case of interval censored data, we seek to obtain an estimator based on the self consistency approach.

For $i = 1, 2, \ldots, n$, let

$$L_{ij} = \begin{cases} 1 & \text{if } T_i \in J_j, \\ 0 & \text{otherwise,} \end{cases}$$

When $i \in \mathcal{I}_2$, the value of L_{ij} is known. Otherwise, its expectation with respect to the probability vector p is given by

$$E(L_{ij}) = \frac{\alpha_{ij} q_j}{\sum_{j=1}^{v} \alpha_{ij} q_j} = \mu_{ij}(p), \quad \text{say.} \tag{3.31}$$

Thus, $\mu_{ij}(p)$ represents the probability that the i-th observation lies in J_j. The average of these probabilities across the n individuals,

$$\frac{1}{n} \sum_{i=1}^{n} \mu_{ij}(p) = \pi_j(p), \quad \text{say,} \tag{3.32}$$

should indicate the probability of the interval J_j. Thus, it is reasonable to expect that the vector p would satisfy the equation

$$q_j = \pi_j(p) \quad \text{for} \quad 1 \leq j \leq v. \tag{3.33}$$

An estimator of p may be called self consistent if it satisfies the simultaneous equations (3.33).

The form of the above equations suggests the following iterative procedure.

STEP I. Obtain a set of initial estimates q_j^0 $(1 \leq j \leq m)$.

STEP II. At the nth stage of iteration, use current estimate, p^n, to evaluate $\mu_{ij}(p^n)$ for $i = 1, 2, \ldots, n$, $j = 1, 2, \ldots, v$ and $\pi_j(p^n)$ for $j = 1, 2, \ldots, v$ from (3.31) and (3.32), respectively.

STEP III. Obtain updated estimates p^{n+1} by setting $q_j^{n+1} = \pi_j(p^n)$.

STEP IV. Return to Step II with p^{n+1} replacing p^n.

STEP V. Iterate; stop when the required accuracy has been achieved.

The following theorem shows that equation (3.33) defining a self consistent estimator must be satisfied by an NPML estimator of p.

Theorem 3.5. *An NPML estimator of p must be self consistent.*

Proof. We can incorporate the constraint $\sum_{j=1}^{v} q_j = 1$, by using the Lagrange multiplier, to maximize

$$\ell = \sum_{i=1}^{n} \left(\log \left(\sum_{j=1}^{v} \alpha_{ij} q_j \right) \right) + \lambda \left(\sum_{j=1}^{v} q_j - 1 \right) \tag{3.34}$$

By setting the derivative of ℓ with respect to λ equal to 0, we have

$$\frac{\partial \ell}{\partial \lambda} = \sum_{j=1}^{v} q_j - 1 = 0. \tag{3.35}$$

On the other hand, by setting the derivative of ℓ with respect to q_j's equal to 0, we obtain

$$\frac{\partial \ell}{\partial q_j} = \sum_{i=1}^{n} \frac{\alpha_{ij}}{\sum_{r=1}^{v} \alpha_{ir} q_r} - \lambda = 0 \qquad \forall j = 1, 2, \ldots, v. \tag{3.36}$$

By multiplying both sides of (3.36) by q_j and adding them over all values of j, we get

$$\sum_{j=1}^{v} \sum_{i=1}^{n} \frac{\alpha_{ij} q_j}{\sum_{r=1}^{v} \alpha_{ir} q_r} = \lambda \sum_{j=1}^{v} q_j, \tag{3.37}$$

which simplifies, after interchange of the summations and utilization of (3.35), to

$$\lambda = n. \tag{3.38}$$

By substituting into (3.36) the optimum value of λ obtained above, we have

$$\sum_{i=1}^{n} \frac{\alpha_{ij}}{\sum_{r=1}^{v} \alpha_{ir} q_r} = n \qquad \text{for } j = 1, \ldots, v,$$

which is equivalent to (3.33), the equation defining the self consistent estimator.

□

Let $\hat{p} = (\hat{q}_1, \ldots, \hat{q}_v)$ and $\hat{\eta} = (\hat{b}_1, \ldots, \hat{b}_k)$ denote values of p and η, respectively, for which $L(p, \eta)$ attains its maximum over the set

$$\Re = \left\{ (p, \eta) | \sum_{j=1}^{v} q_j = 1, \quad 0 \le q_1, \ldots, q_v \le 1, \quad 0 \le b_1 \le \cdots \le b_k \le 1 \right\}.$$

Then a maximum likelihood estimator $\hat{F}_n$ of F is given by

$$\hat{F}_n(t) = \sum_{j: J_j \subseteq [0,t]} \hat{q}_j. \tag{3.39}$$

In the sequel, we refer to this estimator as an NPMLE of F.

3.6 A computationally simpler estimator

The computational complexity of the NPMLE depends on the number of segments (k) used in the piecewise constant formulation of the function π_η. It follows from equations (3.16) and (3.20) that the cardinality of the class $\mathcal{C}$ can increase exponentially with k, though the cardinality of the sub-class $\mathcal{C}_0$ is smaller. One can conceive of a computational simplification on the basis of Theorem 3.3. According to this theorem, the NPMLE has mass only at points of exact recall of the event, when n is large. In such a case, the likelihood (3.22) involves J_j's that are singletons only. Hence, the crucial task of identifying the appropriate J_j's becomes redundant. Therefore, irrespective of the value of n, one can maximize (3.22) with respect to point masses restricted to the time points of exact recall of the event. This method would produce a computationally simpler estimator that is equivalent to the unique NPMLE for large n.

Formally, let $t_1, \dots, t_{n_2}$ be the ordered set of distinct ages at event that have been perfectly recalled, and $q_1^*, \dots, q_{n_2}^*$ be the probability masses allocated to them. The likelihood (3.22), subject to the constraint that $q_j = 0$ whenever $J_j \notin \mathcal{A}_2$, is equivalent to the unconstrained maximization of

$$L(p^*, \eta) = \prod_{i=1}^{n} \left[\sum_{j=1}^{n_2} \alpha_{ij} q_j^* \right], \tag{3.40}$$

with respect to the parameters $p^* = (q_1^*, \dots, q_{n_2}^*)^T$ and η, over the set

$$\Re^* = \left\{ (p^*, \eta) | \sum_{j=1}^{n_2} q_j^* = 1, \quad 0 \le q_1^*, \dots, q_{n_2}^* \le 1, \ 0 \le b_1 \le \dots \le b_k \le 1 \right\}.$$

Let the likelihood (3.40) be maximized at $(\hat{p}^*, \hat{\eta}^*)$, where $\hat{p}^* = (\hat{q}_1^*, \dots, \hat{q}_{n_2}^*)^T$. We define an approximate NPMLE (AMLE) of F as

$$\tilde{F}_n(t) = \sum_{j:t_j \le t} \hat{q}_j^*. \tag{3.41}$$

3.7 Estimation of variance

The variance of the NPMLE and the AMLE may be estimated through bootstrap resampling. Sen, Banerjee and Woodroofe (2010) have argued that some bootstrap methods for constructing non-parametric confidence intervals of distribution function are not guaranteed to be consistent. In view of the argument given by them, we estimate the variances of (3.39) and (3.41) through m out of n bootstrapping of Bickel, Gotze and van Zwet (1997) with selection of m as in Bickel and Sakov (2008), so that consistency is ensured. Bickel, Gotze and van Zwet (1997) discussed a number of resampling schemes in which $m = o(n)$ observation were resampled. They showed how using bootstrap samples of size m, where $m \to \infty$ and $m/n \to 0$, typically resolves the problem of inconsistency. The choice of m is a key point. Bickel and Sakov (2008) considered an adaptive rule to pick m, and gave general sufficient conditions for validity of the rule.

3.8 Consistency of the estimators

Consider the estimator $\tilde{F}_n$. Let Θ be the set of all distribution functions over the support $[t_{min}, t_{max}]$, i.e.,

$$\Theta = \{F : [t_{min}, t_{max}] \to [0, 1];\ F \text{ right continuous, nondecreasing;} \\ F(t_{min}) = 0;\ F(t_{max}) = 1\}. \tag{3.42}$$

and $\overline{\Theta}$ be the set of all sub-distribution functions, i.e.,

$$\overline{\Theta} = \{F : [t_{min}, t_{max}] \to [0, 1];\ F \text{ right continuous, nondecreasing;} \\ F(t_{min}) = 0;\ F(t_{max}) \leq 1\}. \tag{3.43}$$

Note that, with respect to the topology of vague convergence, $\overline{\Theta}$ is compact by Helley's selection theorem. Further, let F_0 denote the true distribution of the time of occurrence of landmark events with density f_0, and $F_0(t_{min}) = 0$.

For any given distribution $F \in \Theta$ having masses restricted to the set $\{t_1, \ldots, t_{n_2}\}$, the log of the likelihood (3.40) can be rewritten as a function of

F (instead $q_1^*, \ldots, q_{n_2}^*$) as

$$\ell(F) = \sum_{i=1}^{n} \log \left[\sum_{j=1}^{n_2} \alpha_{ij} \{F(t_j) - F(t_j-)\} \right]. \tag{3.44}$$

Define the set

$$\mathcal{E} = \{F \; : \; F \in \Theta, \, E[\ell(F) - \ell(F_0)] = 0\}, \tag{3.45}$$

which is an equivalence class of the true distribution F_0.

Strong consistency of the AMLE is established by the following theorem.

Theorem 3.6. *In the above set-up, the AMLE $\{\tilde{F}_n\}$ converges almost surely to the equivalence class $\mathcal{E}$ of the true distribution F_0, in the topology of vague convergence.*

Proof. The proof relies on an application of Theorem 3.1 of Wang (1985), in the manner it was used by Gentleman and Geyer (1994). The said theorem makes use of five assumptions.

The first assumption requires a separable compactification of the parameter space Θ. In the present case, the set $\overline{\Theta}$ serves this purpose. The Lévy distance can be used as metric, and the compactness follows by the Helley selection theorem. Homeomorphic mapping of $[t_{min}, t_{max}]$ to $[0,1]$ can be used to establish separability (Billingsley, 1968, p.239). The equivalence class $\mathcal{E}$ defined by (3.45) is regarded as a single point in Θ. This takes care of the issue of non-identifiability.

Let, for $r = 1, 2, \ldots$, $V_r(F)$ be the Lévy neighborhood of $F \in \Theta$ with radius $1/r$. For such a sequence of decreasing open neighborhoods, Wang (1985)'s second assumption requires that, for any F_0 in Θ, there is a function $F_r : \overline{\Theta} \to V_r(F_0)$ such that (a) $\ell(F) - \ell(F_r(F))$ is locally dominated on $\overline{\Theta}$ and (b) $F_r(F)$ is in Θ if $F \in \Theta$. We define $F_r(F) = \frac{1}{r+1}F + \frac{r}{r+1}F_0$. Since $\|F_r(F) - F_0\| = \frac{1}{r+1}\|F - F_0\|$, and the Lévy distance is dominated by the Kolmogorov-Smirnov distance, it is clear that $F_r(F) \in V_r(F_0)$. Condition (b) is obviously satisfied. As for condition (a), note that

$$\begin{aligned}
&\sup_{F\in\overline{\Theta}} \left[\ell(F) - \ell(F_{F,r})\right] \\
&= \sup_{F\in\overline{\Theta}} \log \frac{\sum_{j=1}^{n_2} \alpha_{ij}\left(F(t_j) - F(t_j-)\right)}{\frac{1}{r+1}\left[\sum_{j=1}^{n_2}\alpha_{ij}\left(F(t_j) - F(t_j-)\right)\right] + \frac{r}{r+1}\left[\sum_{j=1}^{n_2}\alpha_{ij}\left(F_0(t_j) - F_0(t_j-)\right)\right]} \\
&\le \log(r+1),
\end{aligned}$$

which has finite expectation. Thus, $\ell(F) - \ell(F_r(F))$ is globally dominated on $\overline{\Theta}$.

The third assumption requires that $E[\ell(F) - \ell(F_r(F))] < 0$ for $F_0 \in \Theta$, $F \in \overline{\Theta}$, $F \neq F_0$. Here, F_0 needs to be interpreted as $\mathcal{E}$, and the result follows along the lines of the proof of Lemma 4.4 of Wang (1985).

The fourth and fifth assumptions require that $\ell(F) - \ell(F_r(F))$ is lower and upper semicontinuous for $F \in \overline{\Theta}$ except for a null set of points (which may depend on F only in the case of upper semicontinuity). Both the conditions follow from the portmanteau theorem (Billingsley, 1968, p. 11), as argued by Gentleman and Geyer (1994). No null set needs to be invoked.

Since all the assumptions hold, the stated result follows from Theorem 3.1 of Wang (1985). □

The following theorem establishes consistency of the NPMLE.

Theorem 3.7. *In the set-up described before Theorem 3.6, the NPMLE $\{\hat{F}_n\}$ converges in probability to the equivalence class $\mathcal{E}$ of the true distribution F_0, in terms of the Lévy distance.*

Proof. Theorem 3.6 says that the Lévy distance of $\{\tilde{F}_n\}$ from the equivalence class $\mathcal{E}$ goes to zero almost surely as n goes to infinity, that is,

$$\inf_{F\in\mathcal{E}} d_L(\tilde{F}_n, F) \to 0 \qquad \text{as } n \to \infty \qquad \text{with probability 1.}$$

It follows that $P(\inf_{F\in\mathcal{E}} d_L(\tilde{F}_n, F) > \epsilon) \to 0$.

Using the fact that $P(\omega : \tilde{F}_n(\omega) = \hat{F}_n(\omega)) \to 1$, we conclude

$$P(\inf_{F\in\mathcal{E}} d_L(\hat{F}_n, F) > \epsilon) \to 0.$$

which proves the statement. □

The last theorem of this section ensures that under some conditions the equivalence class used in Theorems 3.6 and 3.7 includes only F_0.

Theorem 3.8. *If either the condition given in part (b) of Theorem 3.1 or the pair of conditions given in part (c) holds, then the equivalence class defined in* (3.45) *is the singleton class* $\{F_0\}$.

Proof. Note that the equivalence class defined in (3.45) is the class of all distribution functions that have Kullback-Liebler 'distance' zero from the true unknown distribution. Let H be the probability measure corresponding to the density h, (which is determined by g, π_η and F through (3.1)). Let H_0 be the 'true' value of H. The Kullback-Liebler 'distance' between H and H_0 is defined as $D(H\|H_0) = \mu(h\log(\frac{h}{h_0}))$. By Jensen's inequality, it is easy to see that $D(H\|H_0) \geq 0$. The equality in Jensen's inequality holds if and only if the argument of the log function is a constant, i.e.,

$$D(H\|H_0) = 0 \quad \text{iff} \quad H = H_0. \tag{3.46}$$

Under the conditions given in part (b) or (c) of Theorem 3.1, H completely identifies F. Hence, $H = H_0$ implies $F = F_0$. It follows that the true distribution of the time-to-event, F_0, is the only member of the equivalence class $\mathcal{E}$. □

3.9 Simulation results

For the purpose of simulation, we generate sample times to landmark event from the Weibull distribution with shape and scale parameters $\alpha = 11$ and $\beta = 13$, respectively, and truncate the generated samples to the interval [8,16]. This truncated distribution has median of 11.57. The corresponding 'time of interview' is generated from the discrete uniform distribution over $\{7, 8, \ldots, 21\}$. These choices are in line with the data set described in Section 1.3, where the time to landmark event is the age at menarche in years. As for the forgetting probability, we use (3.9) with $k = 8$, intervals of equal length and three sets of values of the parameters described in Table 3.1.

TABLE 3.1
Values of $b_1, \ldots, b_8$ in three simulation models and resulting proportion of data with different types of incompleteness

Simulation model	Case (a)	Case (b)	Case (c)
Value of b_1 and b_2	0.1	0.05	0.40
Value of b_3	0.40	0.15	0.40
Value of $b_4, \ldots, b_8$	0.95	0.35	0.40
Percentage of cases with $\delta_i = 0$	39%	39%	39%
Percentage of cases with $\delta_i \varepsilon_i = 1$	36%	51%	27%
Percentage of cases with $\delta_i(1 - \varepsilon_i) = 1$	25%	10%	34%

Case (a) corresponds to rapid forgetting with the passage of time, while Case (b) represents progressively better retention. The choice of constant π_η function in Case (c) makes the censoring non-informative. Case (a) should favour the proposed methods, as the chosen function π_η induces informative censoring. Case (c) is ideal for the Turnbull estimator based on censored duration data, as the censoring is non-informative, while the proposed estimators are burdened with unnecessary nuisance parameters. Case (b) may not favour any method decisively, as the forgetting probability, though informative, is relatively small and consequently the informativeness of the censoring is mild. The proposed methods, on the other hand, have the handicap of nuisance parameters.

The NPMLE and AMLE of F are implemented by assuming that $k, x_1, x_2, \ldots, x_k$ in (3.9) are known, while $\eta = (b_1, b_2, \ldots, b_k)^T$ is estimated. The NPMLE and the AMLE are obtained by maximizing the likelihoods (3.22) and (3.40), respectively. Recursive maximization is carried out alternately with respect to the probability parameter p and the nuisance parameter η. Since k is chosen as 8, there are eight different b_l's to be estimated along with NPMLE and AMLE, even though many of the b_l's have equal values.

We compare the performances of the NPMLE (3.39) and the AMLE (3.41) with the two MLEs based on (2.1) and (2.2), described here as the Turnbull estimator (status) and the Turnbull estimator (duration), respectively. As a benchmark, we also evaluate the performance of the empirical distribution function (EDF), a hypothetical estimator computed from the underlying complete data. The results reported here are based on 500 simulation runs for sample sizes $n = 100$, 300 and 1000. The simulations for the three cases are run parallely. For each run, the complete data as well as the observation times for the three cases are the same, while the events of forgetting are simulated subsequently according to the chosen forgetting probability.

The Turnbull estimator (status) is uniquely defined only at integer ages. Therefore, in all the plots, we represent it through a set of unconnected points at integer ages.

Figure 3.3 shows plots of the bias, the variance and the mean square error (MSE) of the five estimators for different ages, for $n = 100$ and parameters of the forgetting function (3.9) chosen as in Case (a). The NPMLE is found to have smaller bias than the Turnbull estimator (duration), smaller variance than the Turnbull estimator (status), and smaller MSE than both the Turnbull estimators. The bias of the AMLE is only marginally worse than that of the NPMLE, and their MSE's are comparable.

The large negative bias of the Turnbull estimator (duration) is noteworthy. The simulation model induces greater chance of forgetting when the gap between the time of event and the interview time is longer. Thus, for a case of non-recall, the actual time of the event is generally earlier than what it would have been if the memory had not been assumed to fade with time. The Turnbull estimator (duration) corresponds to the latter assumption (constant π_η), and therefore it generally produces an under-estimate of the time-to-event distribution, which corresponds to larger time-to-event, i.e., smaller gap between the times of event and interview.

Figure 3.4 shows these plots for $n = 100$ and parameters of the forgetting function (3.9) chosen as in Case (b). Even though the bias of the estimators reduce, the overall pattern of performances remains the same. The Turnbull estimator (duration) appears to have smaller bias in this case, where there is slower fading of memory with passage of time. The performance of the AMLE is almost identical to that of the NPMLE. The similarity of performances of the Turnbull estimator (duration), the NPMLE and the AMLE may be explained by the fact that the

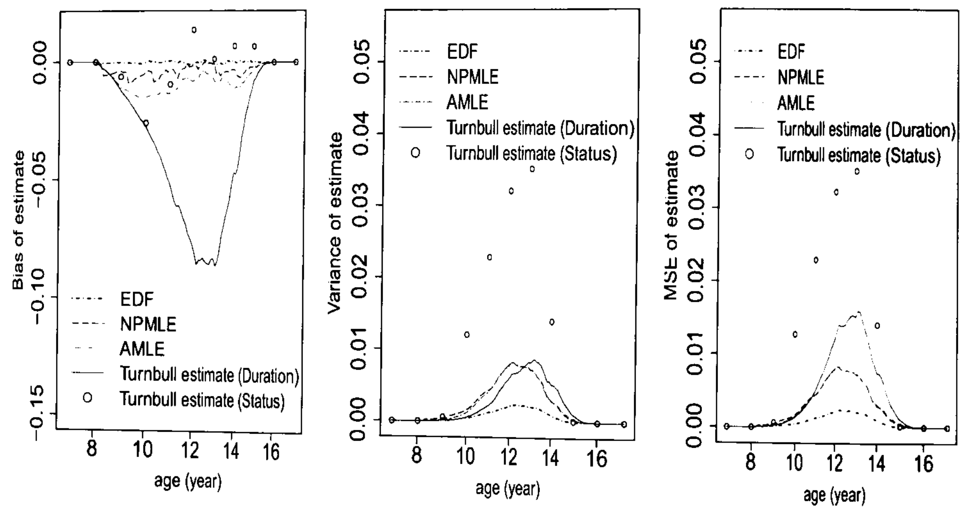

Fig 3.3: Comparison of bias, variance and MSE of the four estimator in case (a) and $n = 100$

nature of treatment of the cases with forgotten dates of events matters less when there is less forgetting. The EDF and the Turnbull estimator (status) have exactly the same performance as depicted in Figure 3.3, since the data required for these estimators remain unchanged.

Figure 3.5 shows these plots for $n = 100$ and parameters of the forgetting function (3.9) chosen as in Case (c). The performances of the EDF and the Turnbull estimator (status) continue to be as seen in Figures 3.3 and 3.4. The Turnbull estimator (duration), the NPMLE and the AMLE have similar patterns of bias, variance and MSE. Note that the constancy of the forgetting probability makes the Turnbull estimator (duration) the appropriate NPMLE in this case. Even though the NPMLE and the AMLE are handicapped with the nuisance parameters $b_1, \ldots, b_8$, their performances are not inferior to that of the Turnbull estimator (duration) in any way.

In all the cases, the performances of the Turnbull estimator (duration), the NPMLE and the AMLE are noticeably worse than that of the EDF. This is

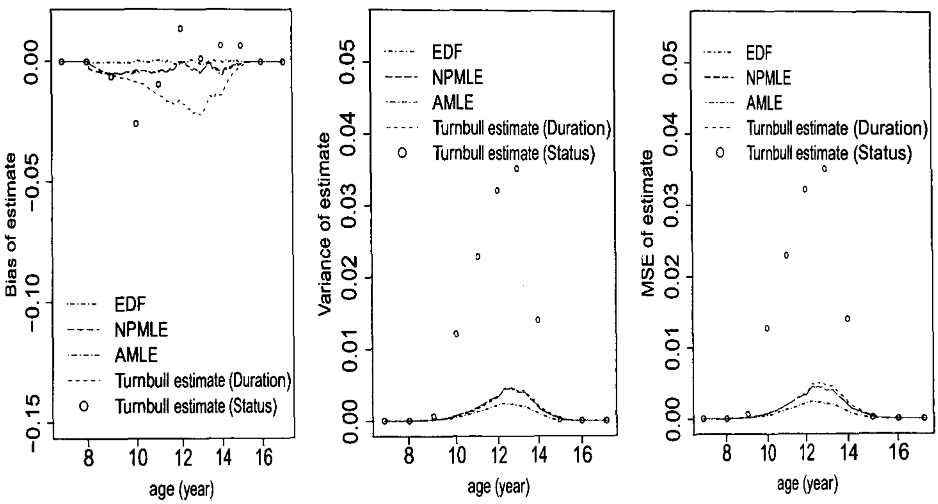

Fig 3.4: Comparison of bias, variance and MSE of the four estimator in case (b) and $n = 100$

because of the substantial number of right censored observations (with $\delta_i = 0$), as seen from Table 3.1. The superior performance of the NPMLE and the AMLE in comparison with the Turnbull (status) shows how gainfully the recall data can be utilized.

Figures 3.6, 3.7 and 3.8 show plots similar to Figures 3.3, 3.4 and 3.5 for $n = 300$. There is a marked reduction in the bias and the variance of the NPNLE, the AMLE and the Turnbull estimator (status). The previously observed pattern of relative performances continues to prevail. The bias of the Turnbull estimator (duration) observed in Figure 3.8 is smaller in comparison with the same case with $n = 100$ (Figure 3.5). This is expected, as the interval censoring associated with forgetting the date of event is chosen to be non-informative in this case. There is no such reduction in Figures 3.6 and 3.7 though. The patterns of bias of the Turnbull (duration) estimator observed in these two figures are of the same order as observed in Figures 3.3 and 3.4 respectively. This occurrence underscores the cost of inadequate handling of the cases of non-recall.

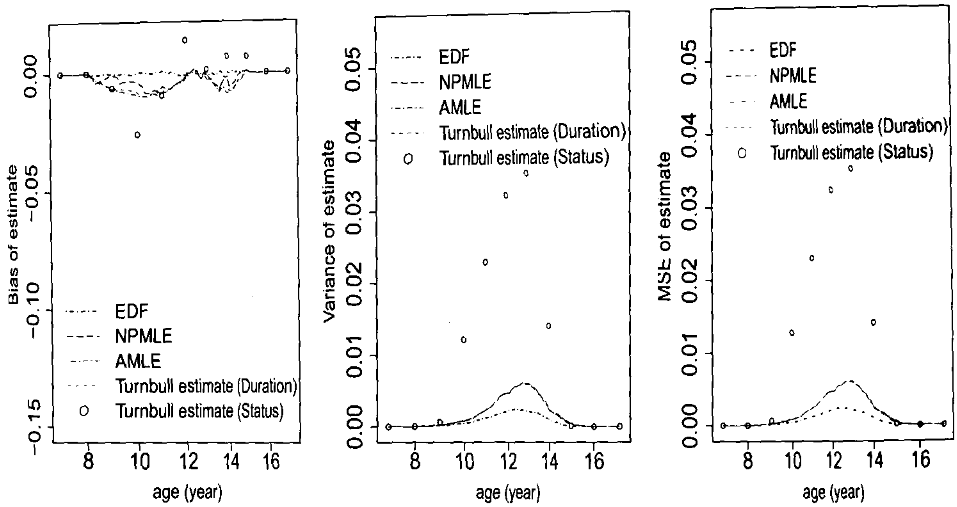

Fig 3.5: Comparison of bias, variance and MSE of the four estimator in case (c) and $n = 100$

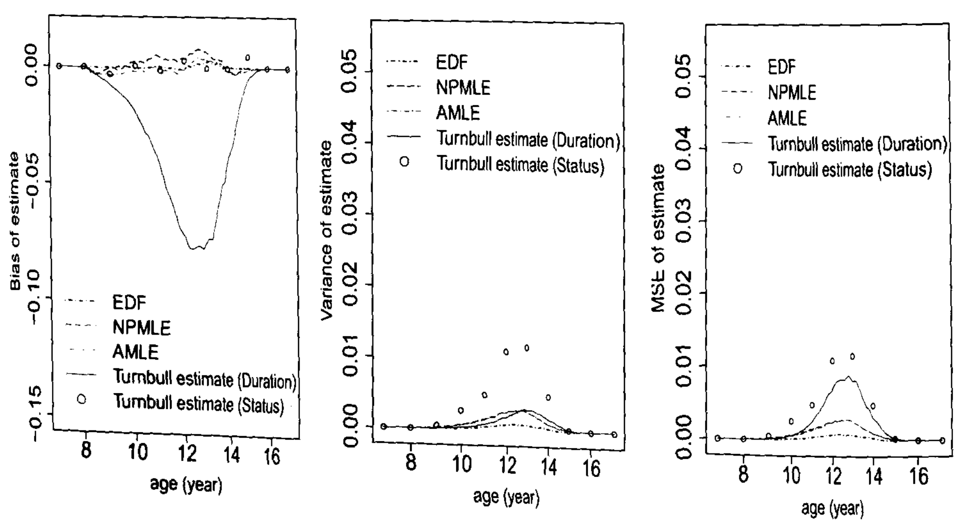

Fig 3.6: Comparison of bias, variance and MSE of the four estimator in case (a) and $n = 300$

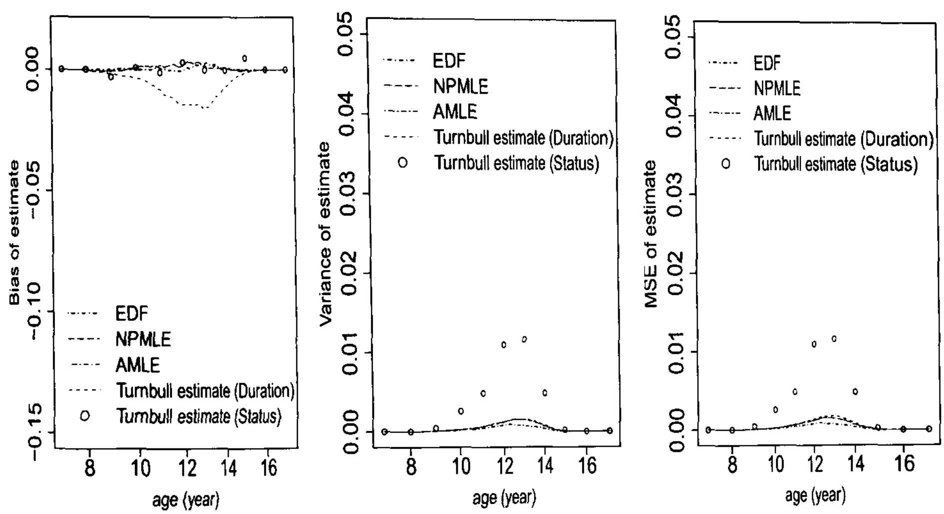

Fig 3.7: Comparison of bias, variance and MSE of the four estimator in case (b) and $n = 300$

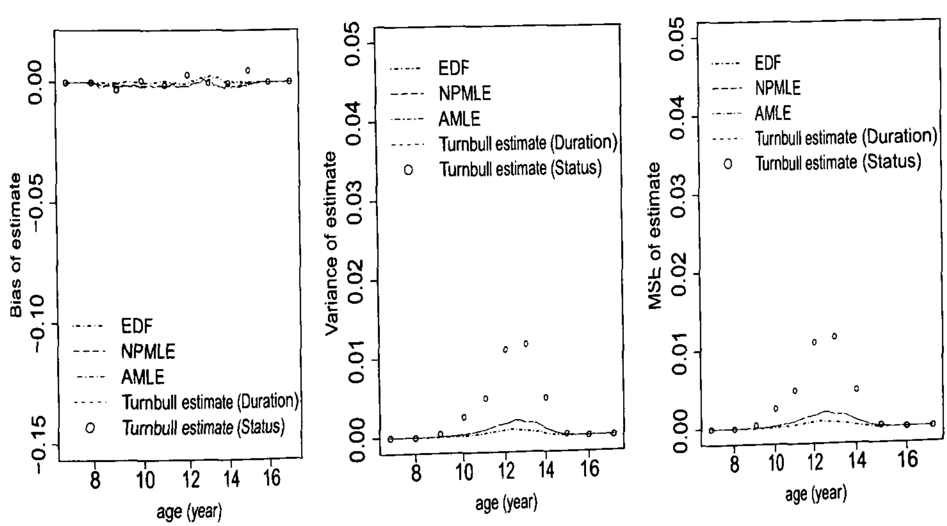

Fig 3.8: Comparison of bias, variance and MSE of the four estimator in case (c) and $n = 300$

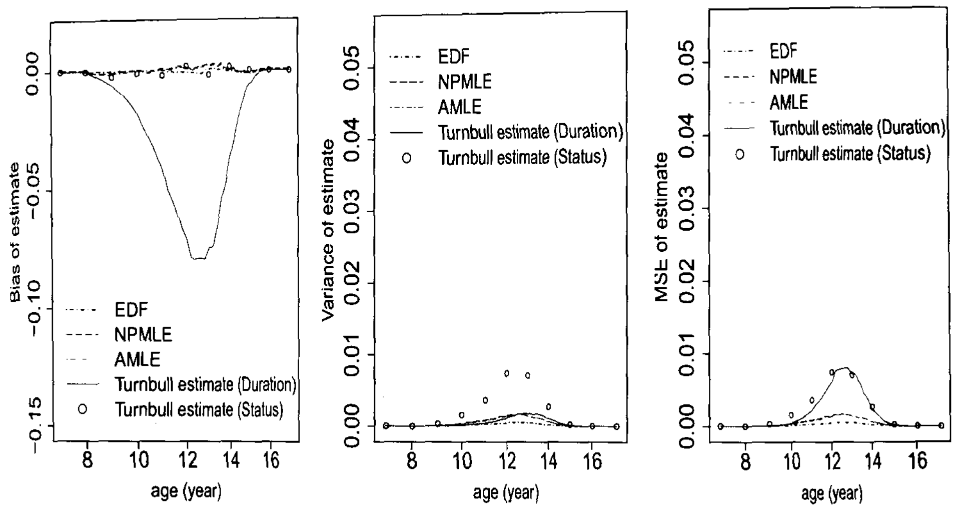

Fig 3.9: Comparison of bias, variance and MSE of the four estimator in case (a) and $n = 1000$

Simulations for $n = 1,000$ in Cases (a), (b) and (c), leading to Figures 3.9, 3.10 and 3.11, show that the bias and the variance of the Turnbull estimator (status), the NPMLE and the AMLE continue to reduce with sample size. The same can be said about the Turnbull estimator (duration) in Case (c), as observed from Figure 3.11. In contrast, the bias of the Turnbull estimator (duration) appears to have stagnated in Cases (a) and (b), as observed in Figures 3.9 and 3.10.

On the basis of the above simulations, the AMLE may be regarded as a reasonable substitute for the NPMLE.

We now turn to the performance of the bootstrap estimator of variance. For this study, we choose $n = 1,000$ and the parameters of the forgetting function as in Case (b). We choose the m out of n bootstrap of Bickel, Gotze and van Zwet (1997), with $m = n^{0.8}$ (see Bickel and Sakov, 2008). Figure 3.12 shows the plots of the average (across 500 runs) of the bootstrap estimate of variance of the NPMLE and the AMLE shown in panel (I) and the sample variance (across 500 runs) of the two estimators in panel (II). The corresponding plots for the other estimators

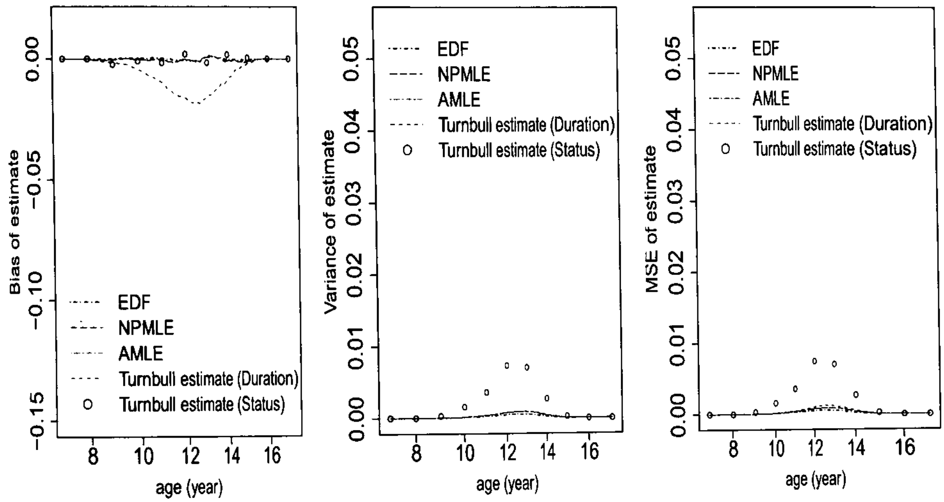

Fig 3.10: Comparison of bias, variance and MSE of the four estimator in case (b) and $n = 1000$

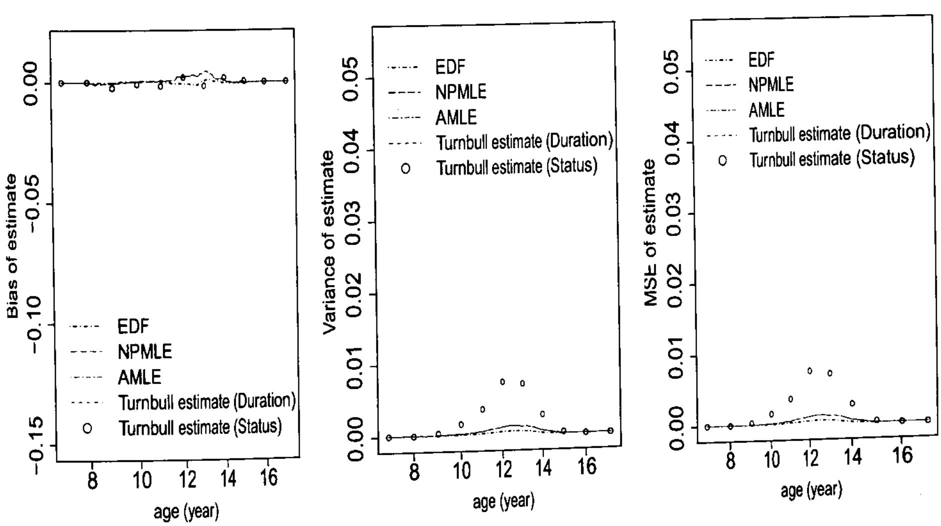

Fig 3.11: Comparison of bias, variance and MSE of the four estimator in case (c) and $n = 1000$

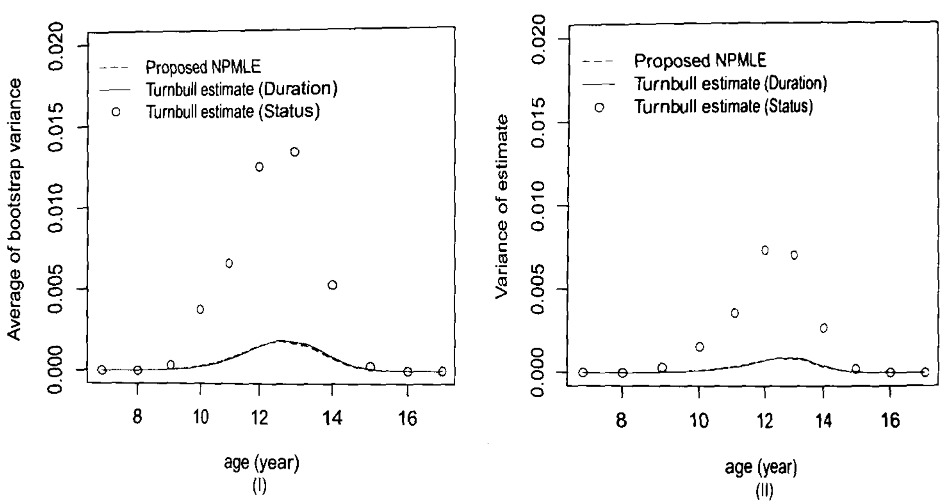

Fig 3.12: (I) Average of bootstrap variance estimator and (II) Sample variance of the four estimators of F.

are also shown. The two sets of the plots show comparable patterns, and mild overestimation of variance on the average. Figure 3.13 shows the standard error (across 500 runs) of the bootstrap estimator of variance, alongside the average (across 500 runs) of the same. It is seen that the standard error is generally much smaller than the average. Thus, the bootstrap estimator of variance appears to be a reasonable one.

Plots for other sample sizes and other values of parameters, which show similar patterns, are omitted for the sake of brevity.

3.10 An example

For the data set explained in Subsection 1.3, the landmark event is the onset of menarche. We modeled the forgetting probability π_η, over the interval 0 to 13 years (maximum possible separation between menarcheal age and age at observation in the sample). We used a piecewise constant model, with $k = 8$ and a uniform grid. Figure 3.14 shows the NPMLE, the AMLE, the Turnbull estimator (duration) and

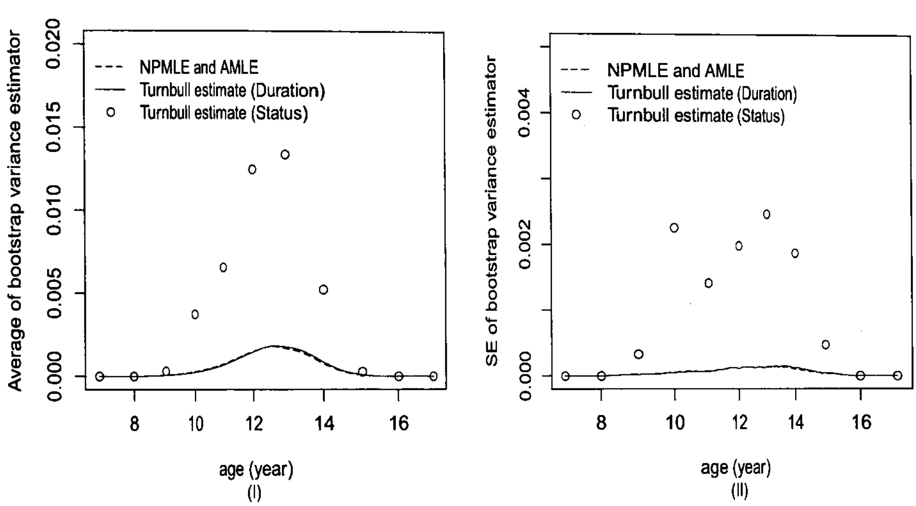

Fig 3.13: (I) Average and (II) Standard error of bootstrap variance estimator using four methods.

the Turbull estimator (Status) of the distribution function of the age at menarche. It can be seen that the NPMLE and the AMLE are indistinguishable. The NPMLE, the AMLE and the Turnbull (status) estimator are closer to one another as compared to the Turnbull (duration) estimator, which is expected to be biased. Since the Turnbull estimator (status) is not uniquely defined at non-integer ages, the NPMLE or the AMLE may be preferred.

In order to get an idea about the estimation error, we estimate the variances of the NPMLE, the AMLE and the two Turnbull estimators through bootstrap resampling. As in the previous section, we use m out of n bootstrap of Bickel, Gotze and van Zwet (1997), with $m = n^{0.8}$, i.e., $m = 472$, and 500 replications. Plots of the bootstrap estimators of variance of the three estimators, shown in Figure 3.15, reveal that the Turnbull (status) estimator has a much larger variance compared to the NPMLE and the AMLE. The proposed methods appear to produce a more accurate and precise estimate than the other two methods.

The chosen value of k for estimation was obtained after considering a coarser and

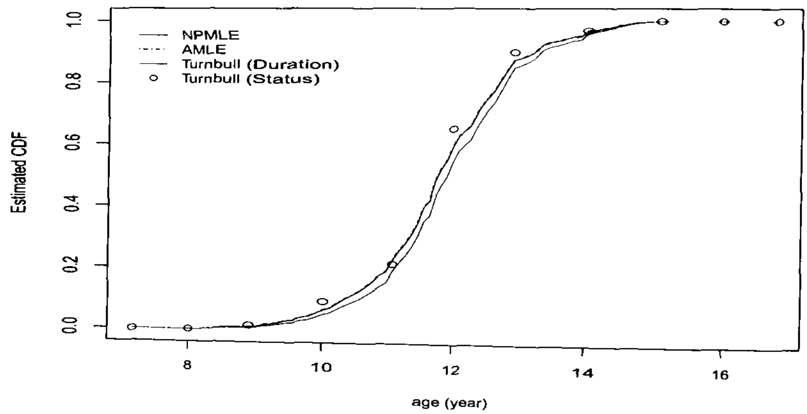

Fig 3.14: Estimated distribution function of data using three methods.

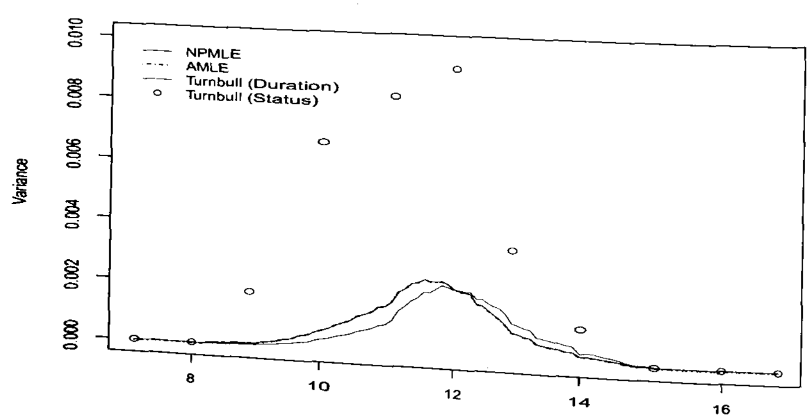

Fig 3.15: Variance of estimated distribution function of data using three methods.

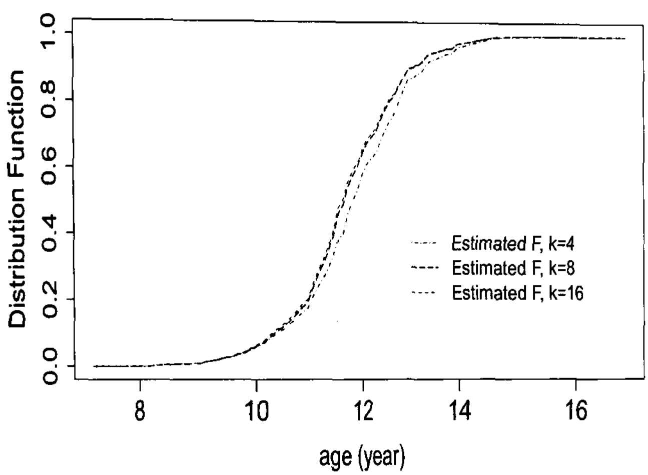

Fig 3.16: Estimated distribution function with different k

a finer partition for the piecewise constant model of π_η. Specifically, the range 0 to 13 years was split experimentally into k equal intervals, with $k = 4, 8$ and 16, and the resulting estimated distribution functions were compared. Figure 3.16 shows plots of the estimated distribution function for different values of k. It is seen that by increasing k from 4 to 8, one observes a substantial change in the estimated distribution function, though the change is much less when k is increased from 8 to 16. The integrated mean square difference between the distribution functions (scaled by the integral of the square of the function for the lower value of k) is 0.85 when one compares $k = 4$ with $k = 8$. The same criterion produces the value 0.019 when the comparison is between the curves for $k = 8$ and $k = 16$. We have chosen $k = 8$, as the alternative choice $k = 16$ does not produce a substantially different estimate of the distribution function. Figure 3.17 shows the estimated function π_η for different values of k. Once again, the estimates of π_η for $k = 8$ and $k = 16$ differ much less than those for $k = 4$ and $k = 8$. This finding justifies the choice $k = 8$.

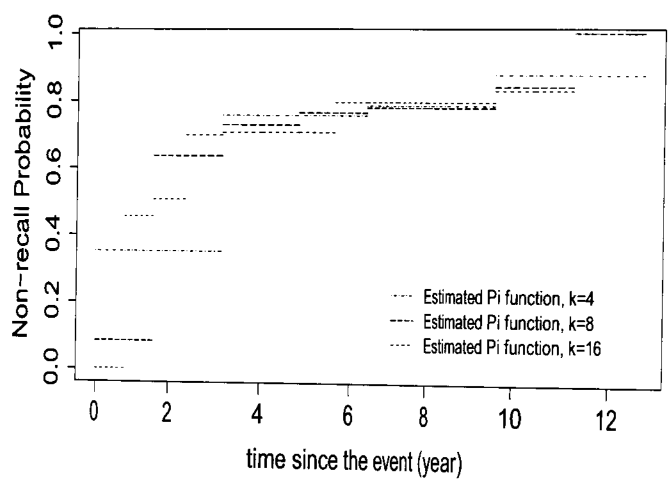

Fig 3.17: Estimated π_η function with different k

3.11 Concluding remarks

In this chapter, we have offered a realistic model and method for estimating the time-to-event distribution based on recall data, in the presence of informative censoring.

As we discussed in Section 2.3, the forgetting function π_η can be regarded as the distribution function of a hypothetical 'time-to-forget'. This formulation requires that π_η should be a non-decreasing function. This constraint can be incorporated in the proposed estimation procedure as follows. Recall that the likelihood (3.22) is maximized alternately with respect to π_η and p. In a particular step, after maximization is done with respect to π_η (with p held fixed), we can use isotonic regression, through the usual algorithm of pooling adjacent violators, on the estimated π_η to obtain a monotonically non-decreasing estimate of π_η. Maximization with respect to p can then proceed after holding π_η fixed at this adjusted estimate. These steps may be repeated until convergence is achieved.

In Section 3.7, the variances of the NPMLE and the AMLE are estimated through bootstrap resampling. A computational formula for the variance is not available, even for large sample size. This may be appreciated in the context of the fact that no result on the asymptotic variance of the NPMLE of the distribution function is available even in the case of general non-informative censoring. Some results for the Turnbull estimator are available in special cases, e.g., for current status data, deterministic censoring times, discrete distribution of censoring time/time-to-event, etc. (Huang, 1999; Yu, Wong and Li, 2001; Chen, Sun and Peace, 2013). However, these results do not apply to the general case of interval censored data of the mixed type, when the underlying time-to-event distribution is continuous.

For the problem of estimating a function, a nonparametric estimator serves as a natural tool for assessing the goodness of fit of a parametric model. We have presented in Section 2.6 a few techniques for checking the adequacy of a parametric model of the time-to-event distribution. The approximate MLE developed in this chapter can be used as an additional check. As an illustration, in Figure 3.18, we compare graphically the closeness of the parametric estimator of time-to-event distribution introduced in Chapter 2 with the AMLE presented here, for the menarcheal data set. The two estimators are very close to one another, indicating appropriateness of the model described in the first paragraph of Section 2.8.

The approach of modeling non-recall through a forgetting function may be adapted to the estimation of the distribution of the time from contracting HIV infection through blood transfusion to the onset of AIDS (Kalbfleisch and Lawless, 1989). Here, the subjects listed in a central registry have a known date of onset of AIDS, but the date of transfusion is sometimes difficult to ascertain retrospectively. However, a range of dates may be available. Since the registry does not include subjects that have had blood transfusion but are yet to develop AIDS, the data are truncated. If one ignores the issue of truncation, as in Kalbfleisch and Lawless (1989), it is possible to incorporate censored data (i.e., cases where a range of transfusion is available), through modeling of recall uncertainty following the ap-

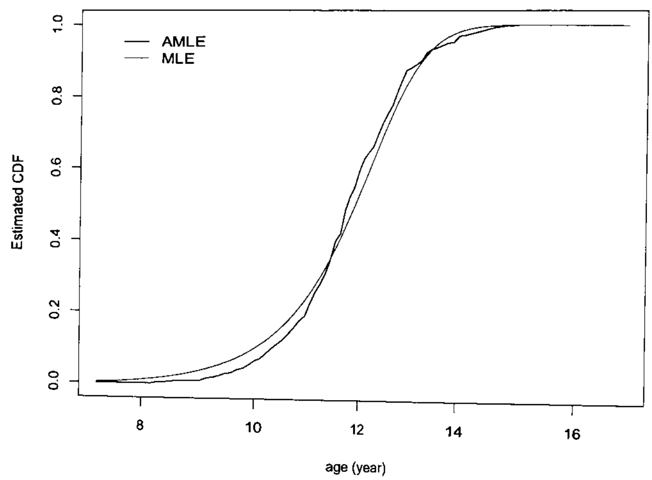

Fig 3.18: Comparison of MLE and AMLE of menarcheal age distribution.

proach used in this chapter. Let Y_i be the date of transfusion leading to infection (if the date is known) and (l_i, r_i) be the range of possible dates of transfusion leading to infection (if exact date is unknown). Let S_i be the date of onset of AIDS and ε_i be the indicator of date of transfusion being known for the ith subject. According to this model, the likelihood is

$$\prod_{i=1}^{n} \left[\int_{l_i}^{r_i} f(S_i - y)\pi_\eta(S_i - y)dy \right]^{1-\varepsilon_i} [f(S_i - Y_i)(1 - \pi_\eta(S_i - Y_i))]^{\varepsilon_i}, \quad (3.47)$$

where f is the density function of duration from infection to the onset of AIDS (also known as incubation period). When (3.9) holds, the likelihood (3.47) simplifies to

$$\prod_{i=1}^{n} \left[\sum_{l=1}^{k} b_l \{F((s_i - r_i) \vee x_l) - F((s_i - l_i) \wedge x_{l+1})\} \right]^{1-\varepsilon_i}$$
$$\left[f(S_i - Y_i) \left(1 - \sum_{l=1}^{k} b_l I_{(x_l < S_i - Y_i \leq x_{l+1})} \right) \right]^{\varepsilon_i}, \quad (3.48)$$

where F is the distribution function of the incubation period.

The likelihood (3.48) involves probabilities assigned to intervals of the type $[t, \infty]$ or $(t, \infty]$, as per the baseline probability distribution. Since these intervals have overlap, we will write them as unions of some disjoint intervals. Let $\mathcal{I}_1$ and $\mathcal{I}_2$ be sets of indices i (between 1 and n) that satisfy the conditions $\varepsilon_i = 0$ and $\varepsilon_i = 1$, respectively. Consider the intervals

$$\begin{aligned} E_l &= (x_l, \infty), && l = 1, \ldots, k; \\ A_i &= [S_i - Y_i, \infty] && \text{for } i \in \mathcal{I}_2; \\ A'_i &= (S_i - Y_i, \infty] && \text{for } i \in \mathcal{I}_2; \\ &\begin{cases} D^l_i = (S_i - l_i, \infty), \\ D^r_i = (S_i - r_i, \infty), \end{cases} && \text{for } i \in \mathcal{I}_1. \end{aligned} \tag{3.49}$$

and the sets

$$\mathcal{B} = \{A_i \backslash A'_i, i \in \mathcal{I}_2\} \cup \{A'_i, i \in \mathcal{I}_2\} \cup \{D^l_i, i \in \mathcal{I}_1\} \cup \{D^r_i, i \in \mathcal{I}_1\} \cup \{E_l, l = 1, \ldots, k\}. \tag{3.50}$$

We collect all the distinct elements of $\mathcal{B}$, and call them as $B_1, B_2, \ldots, B_M$. As in Section 3.4, we can use all non-empty subsets of the index set $\{1, 2, \ldots, M\}$, denoted by $s_1, s_2, \ldots, s_2^{M-1}$, to define the intervals $I_1, \ldots, I_{2^M-1}$ by (3.15), and consider the collection $\mathcal{A}$ of non-empty intervals defined in (3.17). If $\mathcal{J} = \{j : I_j \in \mathcal{A}\}$ and p_j is the probability assigned to I_j, for $j \in \mathcal{J}$, then the likelihood(3.48) simplifies to

$$L = \prod_{i=1}^{n} \left(\sum_{j \in \mathcal{J}} \alpha^*_{ij} p_j \right) \tag{3.51}$$

subject to the restriction $\sum_{j \in \mathcal{J}} p_j = 1$, where,

$$\begin{aligned} \alpha^*_{ij} = \sum_{l=1}^{k} b_l \Big[& I_{\left(I_j \subseteq E_l \backslash E_{l+1}\right)} \cdot I_{\left(l_i < S_i - x_{l+1} < S_i - x_l < r_i\right)} \\ & + I_{\left(I_j \subseteq E_l \backslash E_{l+1}\right)} \cdot I_{\left(l_i < S_i - x_{l+1} < S_i - x_l < r_i\right)} \\ & + I_{\left(I_j \subseteq D^r_i \backslash D^l_i\right)} \cdot I_{\left(S_i - x_{l+1} < l_i < r_i < S_i - x_l\right)} \\ & + I_{\left(I_j \subseteq E_l \backslash D^l_i\right)} \cdot I_{\left(S_i - x_{l+1} < l_i < S_i - x_l < r_i\right)} \Big] \quad \text{for } i \in \mathcal{I}_1,\ j \in \mathcal{J}, \end{aligned}$$

and

$$\alpha_{ij}^* = 1 - \sum_{l=1}^{k} b_l I_{\left(I_j \subseteq A_i \setminus A_i'\right)} \cdot I_{\left(S_i - Y_i \in E_l \setminus E_{l+1}\right)} \quad \text{for} \quad i \in \mathcal{I}_2,\ j \in \mathcal{J}.$$

The likelihood (3.51) is similar to the likelihood (3.22) and it can be maximized similarly. A reduction of the set of related intervals in the spirit of Theorem 3.2 may also be possible.

When the observations come from a heterogeneous population, covariate information may be used to address this issue through a regression model. Such a model can also allow one to examine the relationship between the survival function and the covariates. This problem is taken up in Chapter 4.

Chapter 4

Regression under Cox's model

4.1 Introduction

In the previous two chapters, we focused on the use of either parametric or nonparametric methods/models for the analysis of informatively censored time-to-event data arising from uncertainly recalled landmark event in retrospective study. The methods require the data to be possibly censored samples from a homogeneous population. In practice, there may be heterogenity in the population, and it is important to consider the relation of the time-to-event distribution with other factors. One way of doing this is through a regression model, in which the dependence of time-to-event on concomitant variables is explicitly recognized. As already mentioned in Chapter 1, the relative risk regression model (Cox, 1972) has been very popular for modeling the relationship of covariates with time-to-event. It would be of interest to explore whether the model introduced in Chapter 2 and the nonparametric method developed in Chapter 3 can be adapted to include the effect of covariates through the Cox model.

In this chapter, we consider regression under Cox's model for the special type of informatively censored data arising from uncertainly recalled event times in a retrospective study. In Section 4.2, we introduce the combined model for the time-to-event that incorporates the effect of informative censoring as well as covariates. In Section 4.3, we obtain a simplification of the likelihood that facilitates computation of the MLE of the baseline survival function and the regression coefficients.

In Section 4.4, we make use of a large sample approximation to simplify the computation further. The performance of the resulting approximate MLE is studied through Monte Carlo simulations in Section 4.5. In Section 4.6, we illustrate the proposed method with the data on menarcheal age of adolescent and young adult females, described in Section 1.3. We conclude the chapter with some remarks given in Section 4.7.

4.2 Model, identifiability and likelihood

Consider, for subjects $i = 1, \ldots, n$, the time-to-event T_i which is assumed to be a sample from a distribution F_i with density f_i. Let these subjects be interviewed at times $S_1, \ldots, S_n$, respectively. Let the binary indicator δ_i of the event $T_i < S_i$, as well as the indicator ε_i of the subject i being able to recall the time T_i, follow the model of Section 2.2 with F_θ and f_θ replaced by F_i and f_i, respectively. Let Z_i be the r-dimensional vector of covariates, assumed to be independent of S_i. Note that the distribution of T_i would depend on Z_i. Under Cox's relative risk regression model, the probability of the individual i, with covariate vector Z_i, having the event after time t is

$$\bar{F}_i(t) = [\bar{F}_0(t)]^{\exp(\beta^T Z_i)}, \tag{4.1}$$

where $\bar{F}_0$ is the baseline survival function, assumed to have the density f_0. Thus, the likelihood for the present regression model is

$$\prod_{i=1}^{n} [\bar{F}_i(S_i)]^{1-\delta_i} \left[\{f_i(T_i)(1-\pi_\eta(S_i - T_i))\}^{\varepsilon_i} \left(\int_0^{S_i} f_i(u)\pi_\eta(S_i - u)du \right)^{1-\varepsilon_i} \right]^{\delta_i}, \tag{4.2}$$

with F_i given by (4.1).

Before embarking on the task of estimation, we need to consider the identifiability of β, F_0 and π_η. We drop the subscript i for simplicity. Following Theorem 2.1, it can be seen that a typical factor in the product likelihood is equal to the conditional density of the observable vector (V, δ), given S and Z, where V is as defined in (2.8). The conditional density is written alternatively as

$$h^*(v,\delta|s,z;\beta) = \begin{cases} \bar{F}_0(s)^{\exp(\beta^T z)} & \text{if } v = 0 \text{ and } \delta = 0, \\ \int_0^s -\frac{d}{du}\left(\bar{F}_0(u)^{\exp(\beta^T z)}\right)\pi_\eta(s-u)du & \text{if } v = 0 \text{ and } \delta = 1, \\ -\frac{d}{dv}\left(\bar{F}_0(s-v)^{\exp(\beta^T z)}\right)(1-\pi_\eta(v)) & \text{if } v > 0 \text{ and } \delta = 1, \\ 0 & \text{otherwise.} \end{cases} \tag{4.3}$$

Among the unknown parameters β, F_0 and π_η, the interest lies mainly in β, and possibly in F_0. The following theorem throws light on the issue of identifiability.

Theorem 4.1. *Suppose, for any number τ in the support of F_0, the model (4.3) holds for some $s > \tau$, and for $z = 0$ as well as for at least r linearly independent values of the vector z, where r is the dimension of z. Then the parameters β, F_0 and π are identifiable from h^* under this model.*

Proof. Without loss of generality, we assume that a possible value of z is the 0 vector (this corresponds to a shift of origin, the effect of which can be absorbed through the baseline survival function). For the sake of contradiction, let us assume there are two values of the triplet $(\beta, \bar{F}_0, \pi)$, say $(\beta_1, \bar{F}_{01}, \pi_1)$ and $(\beta_2, \bar{F}_{02}, \pi_2)$, such that their substitutions in the right hand side of (4.3) produce the same function. Then we have, for all z and s and all positive $v < s$,

$$\begin{aligned} -\frac{d}{dv}\left(\bar{F}_{01}(s-v)^{\exp(\beta_1^T z)}\right)(1-\pi_1(v)) &= h(v,1|s,z;\beta) \\ &= -\frac{d}{dv}\left(\bar{F}_{02}(s-v)^{\exp(\beta_2^T z)}\right)(1-\pi_2(v)). \end{aligned}$$

Hence,

$$\frac{\frac{d}{dv}\left(\bar{F}_{01}(s-v)^{\exp(\beta_1^T z)}\right)}{\frac{d}{dv}\left(\bar{F}_{02}(s-v)^{\exp(\beta_2^T z)}\right)} = \frac{1-\pi_2(v)}{1-\pi_1(v)} \quad \forall\ z,\ s,\ v<s, \tag{4.4}$$

i.e.,

$$\frac{\exp(\beta_1^T z)\bar{F}_{01}(s-v)^{\exp(\beta_1^T z)-1} f_{01}(s-v)}{\exp(\beta_2^T z)\bar{F}_{02}(s-v)^{\exp(\beta_2^T z)-1} f_{02}(s-v)} = \frac{1-\pi_2(v)}{1-\pi_1(v)} \quad \forall\ z,\ s,\ v<s. \tag{4.5}$$

In particular, the above identity holds for $z = 0$, i.e.,

$$\frac{f_{01}(s-v)}{f_{02}(s-v)} = \frac{1-\pi_2(v)}{1-\pi_1(v)} \quad \forall\ s,\ v<s, \tag{4.6}$$

After combining the above equation with (4.5), we obtain

$$\frac{\bar{F}_{01}(s-v)^{\exp(\beta_1^T z)-1}}{\bar{F}_{02}(s-v)^{\exp(\beta_2^T z)-1}} = \exp((\beta_2-\beta_1)^T z) \qquad \forall\ z,\ s,\ v<s. \tag{4.7}$$

By taking the limit of the left hand side as v goes to s, we obtain $\exp((\beta_2-\beta_1)^T z) = 1$, i.e.,

$$\beta_2^T z = \beta_1^T z \qquad \forall\ z. \tag{4.8}$$

Since the above equation holds for r linearly independent values of the vector z (as assumed in the statement of the theorem), we have

$$\beta_1 = \beta_2. \tag{4.9}$$

It follows from equations (4.7) and (4.8) that

$$\left[\frac{\bar{F}_{01}(s-v)}{\bar{F}_{02}(s-v)}\right]^{\exp(\beta_1^T z)-1} = 1 \qquad \forall\ z,\ s,\ v<s. \tag{4.10}$$

Therefore,

$$\bar{F}_{01} = \bar{F}_{02}. \tag{4.11}$$

From equations (4.4), (4.9) and (4.11), we have

$$\pi_1(v) = \pi_2(v) \qquad \forall\ v, \tag{4.12}$$

i.e.,

$$(\beta_1, \bar{F}_{01}, \pi_1) = (\beta_2, \bar{F}_{02}, \pi_2),$$

which is a contradiction.

□

We now assume that the distribution of the covariate vector ensures the identifiability of all the unknown parameters, and proceed with the estimation problem using the likelihood (4.2), updated as per (3.9) and (4.1) as

$$L = \prod_{i=1}^{n} [\bar{F}_0(S_i)^{\exp(\beta^T Z_i)}]^{1-\delta_i} \Bigg[\Bigg\{ \left(\bar{F}_0(T_i-)^{\exp(\beta^T Z_i)} - \bar{F}_0(T_i)^{\exp(\beta^T Z_i)}\right) \cdot \left(1 - \sum_{l=1}^{k} b_l I_{(W_{l+1}(S_i)<T_i\le W_l(S_i))}\right) \Bigg\}^{\varepsilon_i} \cdot \left\{ \sum_{l=1}^{k} b_l \left(\bar{F}_0(W_{l+1}(S_i))^{\exp(\beta^T Z_i)} - \bar{F}_0(W_l(S_i))^{\exp(\beta^T Z_i)}\right) \right\}^{1-\varepsilon_i} \Bigg]^{\delta_i}. \tag{4.13}$$

4.3 Maximum likelihood estimation

For simplifying the likelihood, we use the same set-up as described in Section 3.4. The likelihood (4.13) involves probabilities assigned to intervals of the type $[t, t_{max}]$ and $(t, t_{max}]$, as per the baseline probability distribution. Since these intervals have overlap, we express them as unions of some disjoint intervals. Let $\mathcal{I}_1, \mathcal{I}_2$ and $\mathcal{I}_3$ be sets of indices i (between 1 and n) that satisfy the conditions $\delta_i = 0$, $\delta_i\varepsilon_i = 1$ and $\delta_i(1-\varepsilon_i) = 1$, respectively. The set $\mathcal{I}_1$ contains indices of subjects for whom the event is yet to happen till the time of observation, $\mathcal{I}_2$ contains indices of subjects who have experienced the event and remember the date of occurrence, while $\mathcal{I}_3$ is the set of indices of subjects who have experienced the event but forgotten the date. Consider the intervals A_i, A_i' and A_{il} as defined in (3.13), and the sets $\mathcal{A}_1, \mathcal{A}_2, \mathcal{A}_3$ and $\mathcal{A}_4$ as defined in (3.14). Also recall the definition of I_r and the set $\mathcal{C}$ in (3.15) and (3.16), respectively.

As we have seen in Section 3.4, each of the intervals $B_1, \ldots, B_M$ is a union of disjoint sets that are members of $\mathcal{A}$ as defined in (3.17). For any Borel set A, suppose $P_0(A)$ is the probability assigned to A as per the baseline probability distribution corresponding to survival function $\bar{F}_0$. Let $p_r = P_0(I_r)$, for $I_r \in \mathcal{A}$. Then the likelihood (4.13) reduces to

$$L = \prod_{i\in\mathcal{I}_1}\left(\sum_{\substack{r:I_r\subseteq A_i\\ s_r\in\mathcal{C}}} p_r\right)^{\exp(\beta^T Z_i)} \times \prod_{i\in\mathcal{I}_2}\left(1-\sum_{l=1}^{k} b_l I_{(T_i\in A_{i(l+1)}\setminus A_{il})}\right)$$
$$\cdot\left[\left(\sum_{\substack{r:I_r\subseteq A_i\\ s_r\in\mathcal{C}}} p_r\right)^{\exp(\beta^T Z_i)} - \left(\sum_{\substack{r:I_r\subseteq A_i'\\ s_r\in\mathcal{C}}} p_r\right)^{\exp(\beta^T Z_i)}\right]$$
$$\times\prod_{i\in\mathcal{I}_3}\left[\sum_{l=1}^{k} b_l\left\{\left(\sum_{\substack{r:I_r\subseteq A_{i(l+1)}\\ s_r\in\mathcal{C}}} p_r\right)^{\exp(\beta^T Z_i)} - \left(\sum_{\substack{r:I_r\subseteq A_{il}\\ s_r\in\mathcal{C}}} p_r\right)^{\exp(\beta^T Z_i)}\right\}\right]. \quad (4.14)$$

Thus, maximizing the likelihood (4.13) is equivalent to maximizing the likelihood

(4.14) with respect to β, η and p_r for $s_r \in \mathcal{C}$. The p_r's are nuisance parameters when the main objective is to estimate β. The number of these parameters can be very high. This problem is simplified if it can be shown algebraically that some of the estimated p_r's are zero. With this goal, we consider the sub-class $\mathcal{C}_0$ of $\mathcal{C}$, which is defined in (3.20).

Our next result shows that the maximization of the likelihood can be restricted to $\mathcal{C}_0$.

Theorem 4.2. *For fixed values of β and η, maximizing the likelihood (4.14) with respect to p_r for $s_r \in \mathcal{C}$ is almost surely equivalent to maximizing it with respect to p_r for $s_r \in \mathcal{C}_0$, i.e.,*

$$\max_{p_r : p_r \in [0,1], \sum_{s_r \in \mathcal{C}} p_r = 1} L = \max_{p_r : p_r \in [0,1], \sum_{s_r \in \mathcal{C}_0} p_r = 1} L.$$

Proof. Since $\mathcal{C}$ is the union of disjoint sets $\mathcal{C}_0$ and $\mathcal{C}_1 \cup \mathcal{C}_2$, we can rewrite the likelihood (4.14) as

$$L = \prod_{i \in \mathcal{I}_1} \left(\sum_{\substack{r: I_r \subseteq A_i \\ s_r \in \mathcal{C}_0}} p_r + \sum_{\substack{r: I_r \subseteq A_i \\ s_r \in \mathcal{C}_1 \cup \mathcal{C}_2}} p_r \right)^{\exp(\beta^T Z_i)} \times \prod_{i \in \mathcal{I}_2} \left(1 - \sum_{l=1}^{k} b_l I_{(T_i \in A_{il})} \right)$$
$$\cdot \left[\left(\sum_{\substack{r: I_r \subseteq A_i \\ s_r \in \mathcal{C}_0}} p_r + \sum_{\substack{r: I_r \subseteq A_i \\ s_r \in \mathcal{C}_1 \cup \mathcal{C}_2}} p_r \right)^{\exp(\beta^T Z_i)} - \left(\sum_{\substack{r: I_r \subseteq A_i' \\ s_r \in \mathcal{C}_0}} p_r + \sum_{\substack{r: I_r \subseteq A_i' \\ s_r \in \mathcal{C}_1 \cup \mathcal{C}_2}} p_r \right)^{\exp(\beta^T Z_i)} \right]$$
$$\times \prod_{i \in \mathcal{I}_3} \left[\sum_{l=1}^{k} b_l \left\{ \left(\sum_{\substack{r: I_r \subseteq A_{i(l+1)} \\ s_r \in \mathcal{C}_0}} p_r + \sum_{\substack{r: I_r \subseteq A_{i(l+1)} \\ s_r \in \mathcal{C}_1 \cup \mathcal{C}_2}} p_r \right)^{\exp(\beta^T Z_i)} \right.\right.$$
$$\left.\left. - \left(\sum_{\substack{r: I_r \subseteq A_{il} \\ s_r \in \mathcal{C}_0}} p_r + \sum_{\substack{r: I_r \subseteq A_{il} \\ s_r \in \mathcal{C}_1 \cup \mathcal{C}_2}} p_r \right)^{\exp(\beta^T Z_i)} \right\} \right]. \tag{4.15}$$

For every $s_r \in \mathcal{C}_2$, there exists a unique $s_{r^*} \in \mathcal{C}_0$ such that $s_{r^*} \setminus (s_{r^*} \cap s_r) = \{j_r\}$ and $s_r \setminus (s_{r^*} \cap s_r) = \{n_2 + j_r\}$ for some integer j_r in between 1 and n_2, where n_2 is as defined after (3.14). If any probability mass is shifted from I_r to I_{r^*}, the likelihood (4.15) can possibly be affected only through terms that involve the sets B_{j_r} and $B_{n_2+j_r}$, defined as in (3.15). Given the fact that the baseline distribution is absolutely continuous, there is almost surely a unique $i_r \in \mathcal{I}_2$ such that $B_{j_r} = A_{i_r} \setminus A'_{i_r} = \{T_{i_r}\}$ and $B_{n_2+j_r} = A'_{i_r}$. The individual indexed by i_r is the only one whose contribution to the likelihood is affected by the change. For this individual, $I_{r^*} \subset B_{j_r} \subseteq A_{i_r}$, but $I_{r^*} \not\subseteq A'_{i_r}$. On the other hand, $I_r \subseteq B_{n_2+j_r} = A'_{i_r} \subseteq A_{i_r}$. Therefore, the first exponentiated term in the second line of (4.15) remains the same after the shift of mass, while there is a reduction in the subtracted term in that line. The likelihood increases as a result.

We now turn to shifting of probability mass out of I_r, where $s_r \in \mathcal{C}_1$. For any such s_r, define the non-empty set $\mathcal{C}_{s_r} = \{s' : s' \in \mathcal{C}_0, \quad s_r \subset s'\}$. If $\mathcal{C}_{s_r}$ is a singleton, we denote the only member by s_{r^*}. If $\mathcal{C}_{s_r}$ is not a singleton, we denote by s_{r^*} that member which satisfies the condition: 'for all $\beta \in \cup_{j:\, s_j \in \mathcal{C}_{s_r};\, s_j \neq s_{r^*}} I_j$, there is a real number $\alpha \in I_{r^*}$ such that $\alpha < \beta$'. Thus, for every $s_r \in \mathcal{C}_1$, we have a uniquely defined $s_{r^*} \in \mathcal{C}_0$.

If p_r is increased at the expense of p_{r^*}, the likelihood (4.15) can possibly change only through terms that involve sets B_j such that $j \in s_{r^*} \setminus s_r$. We shall show that for an individual i, whose contribution to the likelihood involves such sets, that contribution generally increases due to the said shift of probability mass. In a particular case (Case (iii) below), where this shift cannot be proved to increase the likelihood, there is another way of shifting mass out of p_r that would definitely increase the likelihood.

Case (i). Let $j \in s_{r^*} \setminus s_r$ and $B_j = A_{i_j}$ for some $i_j \in \mathcal{I}_1$. Any shift of probability mass from I_r to I_{r^*} would increase the contribution of the i_jth individual to the likelihood, since $I_{r^*} \subseteq A_{i_j}$ but $I_r \not\subseteq A_{i_j}$.

Case (ii). Let $j \in s_{r^*} \setminus s_r$ and $B_j = A_{i_j} \setminus A'_{i_j}$ for some $i_j \in \mathcal{I}_2$. In this case, $I_{r^*} \subseteq A_{i_j}$ but $I_{r^*} \not\subseteq A'_{i_j}$. By construction, $B_{n_2+j} = A'_{i_j}$, which is disjoint with B_j. In order that I_{r^*} is not a null set, we must have $n_2 + j \notin s_r$. It follows that I_r is not contained in B_j or B_{n_2+j}. Thus, $I_r \not\subseteq A_{i_j}$ and $I_r \not\subseteq A'_{i_j}$.

Clearly, a transfer of probability mass from I_r to I_{r^*} would increase the contribution of the i_jth individual to the likelihood.

CASE (iii). Let $j \in s_{r^*} \setminus s_r$ and $B_j = A'_{i_j}$ for some $i_j \in \mathcal{I}_2$. Since $j \notin s_r$, we have $I_r \subseteq B_j^c = [t_{min}, T_{i_j})$. Therefore, for each of the intervals B_l with $l \in s_r$, $B_l \cap [t_{min}, T_{i_j}) \neq \phi$. On the other hand, since $I_{r^*} \neq \phi$, we have $B_l \cap (T_{i_j}, t_{max}] \neq \phi$ for $l \in s_r$. It follows that each of the intervals B_l, $l \in s_r$, contains a left- and a right-neighborhood of the point T_{i_j}. Consequently, T_{i_j} is contained in these intervals. Hence, the set $s_{r^\dagger} = \{l : T_{i_j} \in B_l\}$ is a superset of s_r contained in $\mathcal{C}_0$, with $I_{r^\dagger} = \{T_{i_j}\} \neq \phi$. As argued in Case (ii), a transfer of probability mass from I_r to $I_{r^\dagger}$ would increase the contribution of the i_jth individual to the likelihood.

CASE (iv). Let $j \in s_{r^*} \setminus s_r$ and $B_j = A_{i_j(l+1)} \setminus A_{i_j l}$ for some $l \in \{1, \ldots, k\}$ and some $i_j \in \mathcal{I}_3$. A transfer of probability mass from I_r to I_{r^*} would increase the contribution of the i_jth individual to the likelihood, since $I_{r^*} \subseteq A_{i_j(l+1)}$ and $I_{r^*} \not\subseteq A_{i_j l}$, whereas I_r is not contained in either of these sets.

It transpires that maximization of L can be achieved even in the presence of the constraint $p_r = 0$ for $s_r \in \mathcal{C}_1 \cup \mathcal{C}_2$. Thus, L can be fully maximized over the restricted set $\{p_r : s_r \in \mathcal{C}_0\}$. □

Let us relabel the intervals I_j, $s_j \in \mathcal{C}_0$, by $J_1, J_2, \ldots, J_v$. Further, let $q_j = P(J_j)$ for $j = 1, 2, \ldots, v$. Theorem 4.2 implies that maximizing the likelihood (4.14) is almost surely equivalent to maximizing

$$
\begin{aligned}
&L(p,\eta,\beta) \\
&= \prod_{i\in\mathcal{I}_1}\Big(\sum_{j:J_j\subseteq A_i} q_j\Big)^{\exp(\beta^T Z_i)} \times \prod_{i\in\mathcal{I}_2}\Big(1-\sum_{l=1}^{k} b_l I_{(T_i\in A_{i(l+1)}\setminus A_{il})}\Big) \\
&\qquad \cdot\left[\Big(\sum_{j:J_j\subseteq A_i} q_j\Big)^{\exp(\beta^T Z_i)} - \Big(\sum_{j:J_j\subseteq A_i'} q_j\Big)^{\exp(\beta^T Z_i)}\right] \\
&\times \prod_{i\in\mathcal{I}_3}\left[\sum_{l=1}^{k} b_l\left\{\Big(\sum_{j:J_j\subseteq A_{i(l+1)}} q_j\Big)^{\exp(\beta^T Z_i)} - \Big(\sum_{j:J_j\subseteq A_{il}} q_j\Big)^{\exp(\beta^T Z_i)}\right\}\right]. \qquad (4.16)
\end{aligned}
$$

with respect to p, η and β, subject to the restriction $\sum_{j=1}^{v} q_j = 1$.

In order to maximize the likelihood (4.16), we need to identify the sets J_j, $j = 1, \ldots, v$, that is, the intervals I_j, $s_j \in \mathcal{C}_0$, defined through (3.15) and (3.20). This identification involves elaborate combinatorial calculations. In fact, simulations reported in Section 3.9 show (in the case of nonparametric estimation in the absence of covariates) that these calculations consume much more computational time than the actual maximization. There it is shown that the set of times of exact recall can serve as a readily available and approximate solution to the combinatorial problem, so that the computational speed can be enhanced several times without sacrificing the quality of the solution substantially. In the next section, we prove a similar result for the regression problem.

4.4 Approximate MLE

Let $\mathcal{A}_0 = \{J_1, J_2, \ldots, J_v\}$, $\mathcal{A}_2 = \{\{T_i\},\ i \in \mathcal{I}_2\}$ as already defined in (3.14), and n_2 is the cardinality of $\mathcal{A}_2$, as before. The task of maximizing the above likelihood can be simplified further through the following result, which is interesting by its own right.

Theorem 4.3. *The set $\mathcal{A}_2$ is contained in the set $\mathcal{A}_0$ almost surely. Further, if the inspection times are samples from a discrete distribution with finite support* *and the range of values of $\beta^T Z_i$ in (4.16) is bounded, then the probability of $\mathcal{A}_0$ being equal to $\mathcal{A}_2$ goes to one as $n_2 \to \infty$.*

Proof. Let $i \in \mathcal{I}_2$ and the index j_i be such that $s_{j_i} = \{j : T_i \in B_j\}$. Since each time-to-event has an absolutely continuous distribution, the recalled times T_i, $i \in \mathcal{I}_2$ are distinct with probability 1. Therefore, $\{T_i\} \in \{B_1, B_2, \ldots, B_{n_2}\}$ almost surely. It follows that $T_i \in I_{j_i} \subseteq \{T_i\}$, i.e., $I_{j_i} = \{T_i\}$ with probability 1. It is also easy to see that s_{j_i} does not belong to $\mathcal{C}_1$ or $\mathcal{C}_2$, with probability 1. Therefore, $s_{j_i} \in \mathcal{C}_0$ and hence $\mathcal{A}_2 \subseteq \mathcal{A}_0$ almost surely.

Let $J_j \in \mathcal{A}_0 \setminus \mathcal{A}_2$. Therefore, there is an index r such that $I_r = J_j \neq \phi$ and $s_r \in \mathcal{C}_0$, even though $I_r \neq \{T_i\}$ for any $i \in \mathcal{I}_2$. We shall show that the existence of I_r implies an event with probability going to zero.

It is easy to see that $i \notin s_r$ for $i = 1, 2, \ldots, n_2$. Thus, I_r can be written as

$$I_r = \left\{\bigcap_{i \in s_r} B_i\right\} \bigcap \left\{\bigcap_{i \notin s_r} B_i^c\right\} = I'_r \setminus \{T_i,\ i \in \mathcal{I}_2\},$$

where

$$I'_r = L_r \bigcap R_r, \quad L_r = \left\{\bigcap_{i \in s_r} B_i\right\}, \quad R_r = \left\{\bigcap_{i \notin s_r, i > n_2} B_i^c\right\}.$$

If there is an $i \in \mathcal{I}_2$ such that, $T_i \in I'_r$, then the index set $s_{r^*} = s_r \cup \{i\}$ corresponds to the non-null interval $I_{r^*} = \{T_i\}$. It follows that $s_r \in \mathcal{C}_1$, which leads to the contradictory conclusion $s_r \notin \mathcal{C}_0$. Therefore $T_i \notin I'_r$ for any $i \in \mathcal{I}_2$.

We now show that an upper bound of the probability of the above event goes to zero as $n_2 \to \infty$. Since the set L_r is obtained as an intersection of sets of the form $(S_i, t_{max}]$, $(T_i, t_{max}]$, $(W_l(S_i), t_{max}]$ or $[t_{min}, t_{max}]$, the intersection itself must be an interval of the form $(l_r, t_{max}]$. On the other hand, since the set R_r is obtained as an intersection of sets that are complements of sets of the above type, the intersection itself must be an interval of the form $[t_{min}, m_r]$. Thus, the set I'_r is the interval $(l_r, m_r]$. By the argument given in the preceding paragraph, neither l_r nor m_r is equal to T_i for any $i \in \mathcal{I}_2$ (otherwise s_r would not be in $\mathcal{C}_0$). Therefore, both l_r and m_r are of the form S_i for some $i \in \mathcal{I}_1$ or of the form $S_i - x_l$ for some $i \in \mathcal{I}_3$ and some $l \in \{1, \ldots, k\}$.

Let $w_1 < w_2 < \cdots < w_K$ be the feasible values of S_i and $S_i - x_l$ (where $1 \leq l \leq k$) that are strictly between t_{min} and t_{max}. Since the baseline distribution is absolutely continuous, we have $1 > \bar{F}_0(w_1) > \bar{F}_0(w_2) > \cdots > \bar{F}_0(w_K) > 0$. The values of l_r and m_r are taken from the set $w_1, w_2, \ldots, w_K$.

The probability of the event "$T_i \notin I_r'$ for any $i \in \mathcal{I}_2$" is

$$\prod_{i\in\mathcal{I}_2}\left[1-\left\{\bar{F}_0^{\exp(\beta^T Z_i)}(l_r) - \bar{F}_0^{\exp(\beta^T Z_i)}(m_r)\right\}\right]$$
$$= \prod_{i\in\mathcal{I}_2}\left[1-\left\{\bar{F}_0^B(l_r)\right\}^{\exp(\beta^T Z_i)/B} + \left\{\bar{F}_0^B(m_r)\right\}^{\exp(\beta^T Z_i)/B}\right],$$

where B is an upper bound on $\exp(\beta^T Z_i)$. Since $u^{\exp(\beta^T Z_i)/B}$ is a strictly concave function of u, we have

$$1-(1-u_2+u_1)^{\exp(\beta^T Z_i)/B} < u_2^{\exp(\beta^T Z_i)/B} - u_1^{\exp(\beta^T Z_i)/B}$$

for $0 < u_1 < u_2 < 1$. Using this inequality for $u_1 = \bar{F}_0^B(m_r)$ and $u_2 = \bar{F}_0^B(l_r)$, we have

$$\prod_{i\in\mathcal{I}_2}\left[1-\left\{\bar{F}_0^{\exp(\beta^T Z_i)}(l_r) - \bar{F}_0^{\exp(\beta^T Z_i)}(m_r)\right\}\right]$$
$$< \prod_{i\in\mathcal{I}_2}\left[1-\bar{F}_0^B(l_r)+\bar{F}_0^B(m_r)\right]^{\exp(\beta^T Z_i)/B}$$
$$< \left[1-\bar{F}_0^B(l_r)+\bar{F}_0^B(m_r)\right]^{n_2 L/B},$$

where L is a lower bound on $\exp(\beta^T Z_i)$. Since $[1 - \bar{F}_0^B(w_{j_1}) + \bar{F}_0^B(w_{j_2})] \in (0,1)$ for any j_1 and j_2 with $1 \leq j_1 < j_2 \leq K$, we have $[1 - \bar{F}_0^B(l_r) + \bar{F}_0^B(m_r)] \in (0,1)$. Therefore, the last expression goes to zero as $n_2 \to \infty$. This completes the proof. □

One can form a computationally simpler estimator on the basis of Theorem 4.3. According to this theorem, the maximum likelihood estimator has mass only at points of exact recall of the event, when n_2 is large. In such a case, the likelihood (4.16) involves J_j's that are singletons only. Therefore, irrespective of the value of n_2, one can maximize (4.16) with respect to point masses restricted to the times of exact recall.

Formally, let $t_1, \ldots, t_{n_2}$ be the ordered set of distinct ages at event that have been exactly recalled, and $q_1^*, \ldots, q_{n_2}^*$ be the probability masses allocated to them, represented by the vector p^*. Maximizing the likelihood (4.16), subject to the constraint that $q_j = 0$ whenever $J_j \notin \mathcal{A}_2$, is equivalent to maximizing the following likelihood subject to $\sum_{j=1}^{n_2} q_j^* = 1$ and $q_j^* \geq 0$:

$$
\begin{aligned}
L_a(p^*, \eta, \beta) = & \prod_{i \in \mathcal{I}_1} \left(\sum_{j: t_j \geqslant t_{m_i}} q_j^* \right)^{\exp(\beta^T Z_i)} \\
& \times \prod_{i \in \mathcal{I}_2} \left(1 - \sum_{l=1}^{k} b_l I_{(T_i \in A_{i(l+1)} \setminus A_{il})} \right) \\
& \qquad \cdot \left[\left(\sum_{j: t_j \geqslant t_{m_i}} q_j^* \right)^{\exp(\beta^T Z_i)} - \left(\sum_{j: t_j > t_{m_i}} q_j^* \right)^{\exp(\beta^T Z_i)} \right] \\
& \times \prod_{i \in \mathcal{I}_3} \left[\sum_{l=1}^{k} b_l \left\{ \left(\sum_{j: t_j \geqslant t_{m_{i(l+1)}}} q_j^* \right)^{\exp(\beta^T Z_i)} - \left(\sum_{j: t_j \geqslant t_{m_{il}}} q_j^* \right)^{\exp(\beta^T Z_i)} \right\} \right].
\end{aligned} \tag{4.17}
$$

where $m_i = \inf\{j : t_j \in A_i\}$ for $i \in \mathcal{I}_1 \cup \mathcal{I}_2$, and, $m_{il} = \inf\{j : t_j \in A_{il}\}$, $l = 1, 2, \ldots, k$ for $i \in \mathcal{I}_3$.

In order to remove the range restriction on the parameters for the underlying baseline survival curve, $\bar{F}_0$, the likelihood is parametrized by

$$
\gamma_d = \log(-\log(\sum_{j: t_j \geq t_d} q_j^*)), \quad d = 1, 2, \ldots, n_2. \tag{4.18}
$$

Thus, the likelihood (4.17) can be expressed as

$$
\begin{aligned}
L_a(\boldsymbol{\gamma}, \eta, \beta) = & \prod_{i \in \mathcal{I}_1} \left(e^{-e^{Z_i\beta + \gamma_{m_i}}} \right) \times \prod_{i \in \mathcal{I}_2} \left(1 - \sum_{l=1}^{k} b_l I_{(T_i \in A_{i(l+1)} \setminus A_{il})} \right) \\
& \qquad \cdot \left[\left(e^{-e^{Z_i\beta + \gamma_{m_i}}} \right) - \left(e^{-e^{Z_i\beta + \gamma_{m_i+1}}} \right) \right] \\
& \times \prod_{i \in \mathcal{I}_3} \left[\sum_{l=1}^{k} b_l \left\{ \left(e^{-e^{Z_i\beta + \gamma_{m_{i(l+1)}}}} \right) - \left(e^{-e^{Z_i\beta + \gamma_{m_{il}}}} \right) \right\} \right].
\end{aligned} \tag{4.19}
$$

where $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_{n_2+1})$ and $\gamma_{n_2+1} = \infty$. A further simplification of (4.19) produces the likelihood

$$\ell_a(\gamma,\eta,\beta) = \sum_{i=1}^{n} \log\left[\sum_{j=1}^{n_2} \alpha_{ij}\left[e^{(-e^{Z_i\beta+\gamma_j})} - e^{(-e^{Z_i\beta+\gamma_{j+1}})}\right]\right]. \tag{4.20}$$

where, for $j = 1, 2, \ldots, n_2$, we have

$$\alpha_{ij} = \begin{cases} I_{(J_j \subseteq A_i)} & \text{if } i \in \mathcal{I}_1, \\ 1 - \sum_{l=1}^{k} b_l . I_{\left(T_i \in A_{i(l+1)} \setminus A_{il}\right)} . I_{(J_j \subseteq A_i \setminus A_i')} & \text{if } i \in \mathcal{I}_2, \\ \sum_{l=1}^{k} b_l . I_{(J_j \subseteq A_{i(l+1)} \setminus A_{il})} & \text{if } i \in \mathcal{I}_3. \end{cases} \tag{4.21}$$

The maximum likelihood estimator of the baseline survival curve $\bar{F}_0$ and the regression parameter β, is obtained by maximizing the above likelihood. The first derivatives for the likelihood (4.20) are

$$\begin{aligned} \frac{\partial \ell_a(\gamma,\eta,\beta)}{\partial b_l} &= \sum_{i=1}^{n} \frac{\sum_{j=1}^{n_2} g_{ij}}{\sum_t \alpha_{it} g_{it}} \xi_{ijl} \\ \frac{\partial \ell_a(\gamma,\eta,\beta)}{\partial \gamma_j} &= \sum_{i=1}^{n} \mu_{ij} h_{ij} \quad \text{for} \quad j = 1, \ldots, n_2, \\ \frac{\partial \ell_a(\gamma,\eta,\beta)}{\partial \beta_k} &= \frac{\sum_{i=1}^{n} \sum_{j=1}^{n_2} \alpha_{ij}\left[h_{ij} - h_{ij+1}\right] x_{ik}}{\sum_l \alpha_{il} g_{il}} \quad \text{for} \quad k = 1, 2, \ldots, r, \end{aligned} \tag{4.22}$$

where $g_{ij} = \left[e^{(-e^{Z_i\beta+\gamma_j})} - e^{(-e^{Z_i\beta+\gamma_{j+1}})}\right]$, $\mu_{ij} = \left(\alpha_{ij-1} - \alpha_{ij}\right) / \sum_t \alpha_{it} g_{it}$,

$\xi_{ijl} = -I\left(T_i \in A_{i(l+1)} \setminus A_{il}\right) . I\left(J_j \subseteq A_i \setminus A_i'\right) I\left(i \in \mathcal{I}_2\right) + I\left(J_j \subseteq A_{i(l+1)} \setminus A_{il}\right) . I\left(i \in \mathcal{I}_3\right)$,

and

$$h_{ij} = \left(e^{-e^{Z_i\beta+\gamma_j}}\right)\left(-e^{Z_i\beta+\gamma_j}\right) \qquad \text{for} \quad j = 1, 2, \ldots, n_2.$$

The MLE's are obtained by setting the partial derivatives equal to zero. A Newton-Raphson iteration can be used to compute the AMLEs $\hat{\gamma}, \hat{\eta}, \hat{\beta}$. The corresponding AMLE of the baseline distribution function $\hat{F}_n$ is

$$\hat{\bar{F}}_0(t) = \sum_{j: t_j \geq t} \hat{q}_j^*, \tag{4.23}$$

where $\hat{q}_1^*, \ldots, \hat{q}_{n_2}^*$ are obtained from (4.18) by substituting $\gamma_d = \hat{\gamma}_d$.

4.5 Simulation results

For the purpose of simulation, we generate samples of time-to-event from a relative risk regression model with survival function $\bar{F}_i(t) = [\bar{F}_0(t)]^{\exp(\beta^T Z_i)}$, where the baseline distribution function $F_0(t)$ is Weibull with shape and scale parameters $\alpha = 11$ and $\beta = 13$, respectively, and discard the samples lying outside the interval [8,16]. This truncated distribution has median 11.57. The vector of covariates, $Z = (Z_1, Z_2)$, consists of a binary variable, taking values 1 and 0 with probabilities 0.25 and 0.75, and a continuous variable having the uniform distribution over the interval [0,5]. We choose the vector of regression coefficients as $\beta = (\beta_1, \beta_2) = (1.5, 1.5)$. The 'time of interview' is generated from the discrete uniform distribution over the set of integers $\{7, 8, \ldots, 21\}$. These choices are in line with the data analytic example of the next section, where the time to landmark event is the age at menarche in years. As for the forgetting probability π_η, we use (3.9) with $k = 8$, $x_1 = 0$, $x_2 = 1.6$, $x_3 = 3.2$, $x_4 = 4.8$, $x_5 = 6.4$, $x_6 = 8$, $x_7 = 9.6$ and $x_8 = 11.2$ and the vector parameter $\eta = (b_1, b_2, \ldots, b_8) = (0.01, 0.15, 0.15, 0.15, 0.15, 0.15, 0.15, 0.15)$.

Note that the approximate log-likelihood (4.20) is maximized alternately with respect to η and (γ, β). For the present simulations, we use an isotonic version of the estimator of π_η in the following way. After each step of maximization with respect to η (with (γ, β) held fixed), we use isotonic regression, through the usual algorithm of pooling adjacent violators, on the estimated π_η to obtain a monotonically non-decreasing estimate of it. Maximization with respect to (γ, β) is then performed after holding π_η fixed. These steps are repeated till convergence.

In the hypothetical situation of all the event times being perfectly recalled, that is, the data are right censored. In this case, one can use the MLE obtained by maximizing Cox's partial likelihood. We refer to this estimator based on 'complete recall' data as the 'complete recall MLE'. On the other hand, if one uses only the 'current status' information, namely whether the event of interest has happened till the time of interview, then the corresponding likelihood is

$$\prod_{i=1}^{n}\left[\bar{F}_0^{\exp(\beta^T Z_i)}(S_i)\right]^{1-\delta_i}\cdot\left[1-\bar{F}_0^{\exp(\beta^T Z_i)}(S_i)\right]^{\delta_i},$$

which can be maximized with respect to β and the values of $\bar{F}_0$ at the possible times of inspection (namely, the integers 7 to 21). We refer to this estimator as the 'current status MLE'. Another option is to use the recalled event time whenever available, but to disregard the informativeness of the censoring. A penalized version of the corresponding likelihood is maximizied in the function `shr` of the `SmoothHazard` package of R, which fits the Cox model by using an approximation of the hazard function by a linear combination of M-splines. We refer to this estimator as the 'SmoothHazard MLE'.

We now compare the performance of the proposed AMLE of the regression coefficients with the three estimators described above. Table 4.1 shows the bias, the standard deviation (Stdev) and the mean squared error (MSE) of the estimated regression coefficients. The results reported here are based on 500 simulation runs for sample sizes $n = 50$, 200 and 1000. It is clear that the standard deviation of the proposed AMLE, as well as its mean square error, is larger than those of the (hypothetical) 'complete recall MLE', but smaller than the 'current status MLE'. The gap between the performances of the first two estimators becomes small as the sample size increases, though the gap between the AMLE and the 'current status MLE' does not reduce as much. The 'SmoothHazard MLE' has a persistent bias even when n is large. This outcome is expected, as the estimator is based on the assumption that the censoring is non-informative. Thus, neither the 'current status MLE' nor the 'SmoothHazard MLE' is able to successfully utilize the information contained in the recalled time-to-event data, while the proposed AMLE is able to do so.

Figures 4.1 and 4.2 show the plots of the empirical bias and the empirical standard deviation of the estimated baseline survival functions, for $n = 50$, 200 and 1000. It is clear that the empirical bias as well as the empirical standard

TABLE 4.1
Bias, Stdev and MSE of estimated regression coefficients

		$n = 50$		$n = 200$		$n = 1000$	
Estimator	Property	β_1	β_2	β_1	β_2	β_1	β_2
Complete recall MLE	Bias	0.2981	−0.0734	0.0089	0.0037	−0.0026	0.0008
	Stdev	0.8321	0.8499	0.1293	0.5048	0.0848	0.2271
	MSE	0.7812	0.7277	0.0168	0.2548	0.0072	0.0515
Proposed AMLE	Bias	0.2593	−0.0547	0.0105	−0.0047	0.0083	-0.0011
	Stdev	1.3145	1.2913	0.1739	0.5057	0.0885	0.2272
	MSE	1.7904	1.6658	0.0303	0.2553	0.0079	0.0516
Current status MLE	Bias	−0.392	−0.3316	−0.026	−0.0367	−0.0032	0.0010
	Stdev	1.4048	1.3405	0.3225	0.9847	0.2353	0.6113
	MSE	2.1271	1.9069	0.1047	0.9709	0.0553	0.3737
Smooth-Hazard MLE	Bias	−0.170	3.551	−0.310	2.841	0.191	1.782
	Stdev	0.740	1.739	0.322	0.850	0.123	0.219
	MSE	0.569	15.648	0.198	8.780	0.0505	3.250

deviation of the estimated baseline survival function become smaller as the sample size increases.

We now turn to the problem of testing for the significance of the estimators of the regression coefficients. The standard theory of parametric estimation generally does not hold in the presence of an infinite dimensional nuisance parameter. However, in the case of the Cox regression model for randomly right censored data, it has been shown that an asymptotic theory based on partial likelihood works in an analogous manner to that based on the asymptotic theory of parametric likelihood (Andersen and Gill, 1982), and that the partial likelihood may be viewed as the full likelihood maximized with respect to the baseline hazard subject to a piecewise linear constraint (Johansen, 1983). We now run some simulations to check whether the likelihood (4.20) with the nuisance parameters $\bar{F}_0$ replaced by the estimator (4.23) can be used similarly to obtain an approximate test of significance of the regression coefficients, even though there is no asymptotic theory as yet to justify such an approximation.

The 'score vector' (borrowing terminology of parametric likelihood theory)

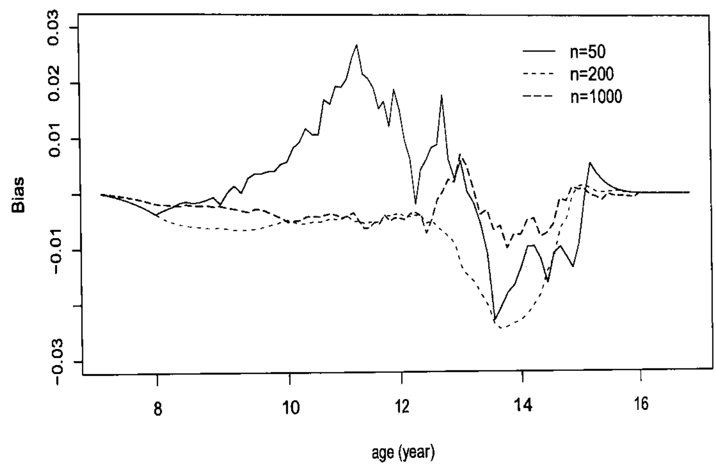

Fig 4.1: Empirical bias of the estimated baseline survival function with $n = 50$, 200 and 1000

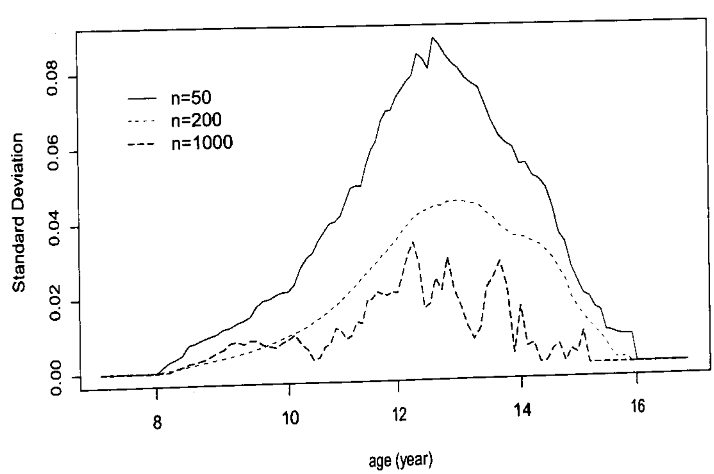

Fig 4.2: Empirical standard deviation of estimated baseline survival function with $n = 50$, 200 and 1000

TABLE 4.2
The empirical type I error probability of test $H_0 : \beta = 0$

Asymtotic	$n = 50$	$n = 200$	$n = 1000$
Type I error	0.041	0.038	0.022

based on $\frac{\partial \ell_a(\gamma,\eta,\beta)}{\partial \beta}$, can be written as

$$U = \sum_{i=1}^{n} \sum_{j=1}^{n_2} \frac{\alpha_{ij}\left(\hat{\hat{F}}(t_j)\log(\hat{\hat{F}}(t_j)) - \hat{\hat{F}}(t_{j+1})\log(\hat{\hat{F}}(t_{j+1}))\right)Z_i}{\sum_l \alpha_{il} g_{il}}. \tag{4.24}$$

The relevant part of the 'information matrix' is $V = A_{22} - A_{21}A_{11}^{-1}A_{12}$, where

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix},$$

$$A_{11} = -\begin{bmatrix} \partial^2\ell_a/\partial\gamma\partial\gamma^T & \partial^2\ell_a/\partial\gamma\partial\eta^T \\ \partial^2\ell_a/\partial\eta\partial\gamma^T & \partial^2\ell_a/\partial\eta\partial\eta^T \end{bmatrix},$$

$$A_{12} = -\left(\partial^2\ell_a/\partial\gamma\partial\beta^T \quad \partial^2\ell_a/\partial\eta\partial\beta^T \right) = A_{21}^T,$$

and

$$A_{22} = -\partial^2\ell_a/\partial\beta\partial\beta^T,$$

the quantities being estimated at $\beta = 0, \gamma = \hat{\gamma}_0$ and $\eta = \hat{\eta}_0$, the restricted AMLE's at $\beta = 0$. The hypothesis $\beta = 0$ may be tested by taking $U^T V^{-1} U$ as an approximate χ^2 statistic with 2 degrees of freedom. In order to check the behavior of this statistic, we generate data of sizes $n = 50, 200$ and 1000 for 1000 runs, under the null hypothesis when the baseline distribution function $F_0(t)$ is the Weibull distribution with shape and scale parameters $\alpha = 11$ and $\beta = 13$, respectively, truncated to the interval [8,16]. Table 4.2 shows the value of empirical type I error probability of this test for different sizes of data. It can be seen that the error probability is less than 0.05. This indicates that the 'score test' is somewhat conservative.

4.6 An example

For the data set explained in Subsection 1.3, the landmark event is the onset of menarche. There are many studies concerning the effects of socioeconomic factors on the measures of body shape (anthropometric indices or ratios) and physical maturation (e.g., biological parameters of the adolescent growth spurt) of children. Some of the important factors which affect age at menarche (maturation in girls) are diet and physical activities which can be directly related to parents' education and monthly family expenditure (Khan et al., 1996; Padez, 2003; Aryeetey, Ashinyo and Adjuik, 2011). We considered three socioeconomic variables: two binary variables indicating whether the father or the mother of the subject had passed high school, and a real variable representing monthly family expenditure in Indian Rupees (indexed with respect to 2008 as base year). We considered a subset of the original data, consisting of 673 respondents who came from a nuclear family and were the only child of their respective parents. Among 673 samples, 241 individuals did not have menarche, 147 individuals had menarche and recalled the date of its onset, while 285 individuals had menarche but could not recall the date. There were 492 individuals with father having passed high school and 420 individuals with mother having passed high school. The median of monthly family expenditure was Rupees 7808. As for the forgetting probability π_η, we modeled it over the interval 0 to 13 years (maximum possible separation between menarcheal age and age at observation in the sample). We used a piecewise constant model, with $k = 8$ and equal length of the intervals over which the probability is constant. There are two binary covariates indicating whether the parents had passed high school, and a continuous covariate representing monthly expenditure of the family. The computational method for AMLE was as described in Section 4.5.

Table 4.3 shows the estimated regression coefficients and the corresponding p-values. The p-values are computed on the basis of the chi-squared distribution with two degrees of freedom, which was seen in the previous section to have led

TABLE 4.3
Estimated regression coefficients and their p–values

Covariates	Estimated value	p–value
Father passed high school	0.091	0.0036
Mother passed high school	0.249	0.0061
Monthly family expenditure	0.0002	0.0047

to conservative decisions. It is found that all the coefficients are significant at the 1% level. The vector of three regression coefficients has p-value 0.00093.

Figure 4.3 shows a plot of the estimated survival functions of four hypothetical subjects with covariate profiles described below.

CASE (a) Neither parent passed high school, monthly family income is equal to the median income of the group (Rs. 7808). We represent this case as $Z = (0, 0, 7808)$.

CASE (b) Only the father passed high school, monthly family income is equal to the median income of the group. We represent this case as $Z = (1, 0, 7808)$.

CASE (c) Both the parents passed high school, monthly family income is equal to the median income of the group. We represent this case as $Z = (1, 1, 7808)$.

CASE (d) Both the parents passed high school, monthly family income is equal to Rupees 10,000. We represent this case as $Z = (1, 1, 10000)$.

It is clear that the fact of any parent having passed high school is associated with earlier maturation. In particular, the mother's educational status is found to account for a greater reduction of the survival function. Also, even a small increase in the monthly family expenditure is found to have a considerable impact on the survival function of the age at menarche.

The chosen value of k was obtained after considering a coarser and a finer partition for the piecewise constant model of π_η. Specifically, the range 0 to 13 years was split experimentally into k equal intervals, with $k = 4$, 8 and 16, and the resulting estimated baseline survival functions were compared. Figure 4.4 shows plots of the estimated baseline survival function for different values of k. There is a substantial change in the estimated baseline survival function when k increases

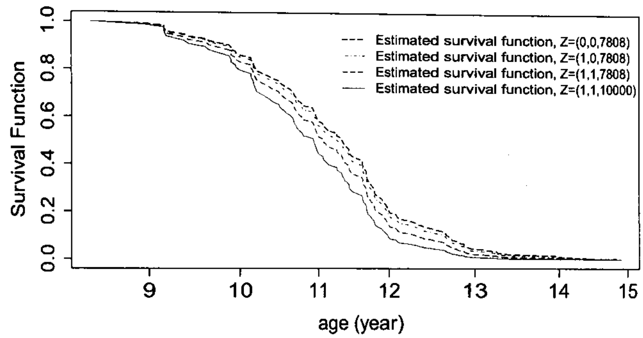

Fig 4.3: Estimated survival function in different cases

from 4 to 8, though the change is much less when k is increased from 8 to 16. When one compares $k = 4$ with $k = 8$ the integrated mean squared difference between baseline survival functions (scaled by the integral of the square of the function for the lower value of k) is 0.92 whereas the same criterion produces the value 0.021 when the comparison is between the curves for $k = 8$ and $k = 16$. This finding justifies the choice of $k = 8$, as the alternative choice $k = 16$ does not produce a substantially different estimate of the baseline survival function. Figure 4.5 shows the estimated function π_η for different values of k. It is seen that, the estimates of π_η for $k = 8$ and $k = 16$ differ less than those for $k = 4$ and $k = 8$.

4.7 Concluding remarks

This chapter presents a method for fitting the Cox regression model to recall-based time-to-event data with covariates, where there is informative censoring. Simulation results indicate that the estimators of the regression coefficients are reasonable. There is no proof of consistency of these estimators as of now. It may be recalled that there is no known proof of consistency of MLEs of the Cox regression parameters even in the case of non-informatively interval-censored data. Some results are available in the special case of status data with fixed and multiple inspection times (and in particular, for the further special case of current status

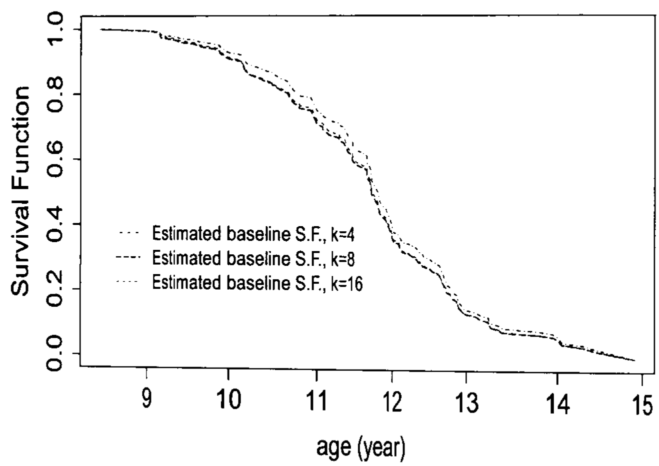

Fig 4.4: Estimated survival function with different k

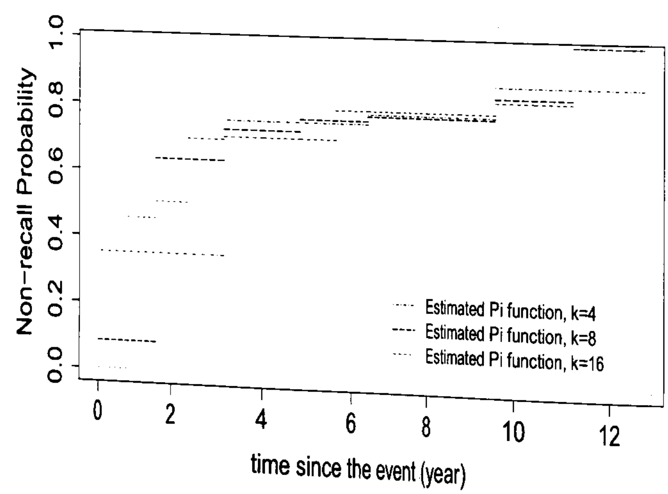

Fig 4.5: Estimated π_η function with different k

data) (Huang, 1996; Yu, Wong and Kong, 2006; Liu and Shen, 2009).

Fitting of a semi-parametric regression model is generally a more complex inferential problem than that of estimating only a distribution. The complexity in the present case is even greater, as the informative interval censoring model leads to a large number of nuisance parameters, including the probability masses allocated, as per the baseline distribution of the Cox model, to intersections of different intervals. The tasks of formation of these intervals and tracking of their probability masses are greatly simplified by the approximation inspired by Theorem 4.3. The Cox regression model appears to be suited to the formulation of the approximate MLE through probability masses assigned to the times of exactly recalled events. It is this matching of the models that makes the AMLE computationally tractable. A different approach may be needed for other regression models.

The proposed approach can be adapted to handle left truncated data. Assuming that there is a time of left truncation associated with each observation, each term in the likelihood would have to be divided by the upper tail probability at the point of truncation. It can be shown that the simplification given through Theorem 4.2 will continue to hold, since the shift of mass envisaged in the proof of that theorem does not alter the factors in the denominator. The objective function (4.20) would then be replaced by

$$\ell_a(\gamma,\eta,\beta) = \sum_{i=1}^{n} \log\left[\sum_{j=1}^{n_2} \alpha_{ij}\left[e^{(-e^{Z_i\beta+\gamma_j})} - e^{(-e^{Z_i\beta+\gamma_{j+1}})}\right]\right] - \log\left[\sum_{j=1}^{n_2} \psi_{ij} e^{(-e^{Z_i\beta+\gamma_j})}\right], \tag{4.25}$$

where ψ_{ij}'s are known constants like α_{ij}'s. The optimization problem is therefore similar.

Chapter 5

Summary of contributions and future Work

5.1 Summary of contributions

In the existing literature of survival analysis, inference from interval censored data is typically drawn after assuming that the censoring is non-informative. If this assumption holds, the likelihood appropriate for this type of censoring can be used to obtain a parametric MLE or a non-parametric MLE. However, in the case of recall based data from retrospective studies, the non-observation window is likely to depend on the time of occurrence of the event. Hence, the censoring is informative. In this thesis, an appropriate model for the underlying censoring mechanism has been proposed and a number of inference procedures have been presented.

After introducing the model connecting the time-to-event distribution with recall based duration data in Chapter 2, we have shown how one can form a likelihood for parametric inference under this model. Since the standard large sample results for the properties of a maximum likelihood estimator based on complete data do not automatically extend to the case of censored data, we have established the consistency and asymptotic normality of the MLE under the chosen model for informatively interval censored data. We have illustrated the utility of the proposed model by showing how oversight of the informativeness in interval censoring can lead to bias and inconsistency of the resulting estimates. Also, we have identified the additional information contained in the recall based duration data and demonstrated theoretically as well as through simulations the advantage of the maximum likelihood estimator based on duration data over that based on status data alone.

In Chapter 3, we have revised the inference problem without any distributional assumption for the time-to-event. Since there is some loss of information through censoring, the time-to-event distribution may not always be identifiable. We have looked into the issue of identifiability and given a mild but sufficient condition for it. We have derived the nonparametric maximum likelihood estimator of the survival function of the time-to-event under the new model of informative censoring. As the non-parametric MLE turns out to be computationally rather complex, we have proposed an approximate MLE and established consistency of both the NPMLE and the approximate MLE under suitable conditions. Monte Carlo simulations indicate that the proposed estimators have smaller bias than the Turnbull estimator based on incomplete duration data, smaller variance than the Turnbull estimator based on current status data, and smaller mean squared error than both of them.

In Chapter 4, we have considered semi-parametric regression under Cox's model for recall based duration data with covariates, arising from retrospective studies. We have proved identifiability of the regression parameters as well as the baseline survival function from the combination of the Cox model and the censoring model proposed in Chapter 2. While a partial likelihood is not available, we have developed a method of semi-parametric inference of the regression parameters and the baseline survival function, under the proposed model. Monte Carlo simulations show reasonable performance of the estimator of the regression parameter, compared to the Cox estimator of the same parameter computed from the complete version of the data. As in the case of non-informatively interval censored survival data, convergence of the estimator under the proposed model, in probability or in distribution, remains to be established. However, simulations indicate that the naive estimator of 'asymptotic variance', obtained from the observed likelihood after replacing the baseline survival function with its estimate, is a conservative one that may be used for testing for the significance of the regression parameters.

In all the chapters, the proposed methods have been illustrated through the analysis of the data set on age at menarche described in Section 1.3.

5.2 Future work

The data set introduced in Section 1.3 also contains 'partial' recall data describing a range of dates for the occurrence of menarche. The intervals specified by the respondents vary from case to case. The specified interval can be a range of days, the calendar month, a range of months, the calendar year or even a range of years. Table 5.1 shows the frequency of these different types of recall, as well as the frequencies of no recall and no menarche till the date of observation, for different age groups.

While the age groups represent cohorts with possibly heterogeneous ages at the onset of menarche, the pattern of heterogeneity is unlikely to differ from one age group to another. Therefore, a comparison of the incidence of partial or no recall among the age groups 16 to 21 (for which all the respondents reported having experienced menarche) is meaningful. The increase in the proportion of cases of 'no recall' with increase in age, seen in the second row of the table from the bottom, corroborates the assumption of memory fading with time, which has been used in Chapter 4. However, no pattern is observed in respect of the incidence of different *types* of partial recall. One might conclude that it would not make sense

TABLE 5.1

Frequency of different recalling type for different ages

Type of recall	Age															Total
	'7'	'8'	'9'	'10'	'11'	'12'	'13'	'14'	'15'	'16'	'17'	'18'	'19'	'20'	'21'	
Menarche not happened	155	155	149	146	103	49	13	4	1							775
Exact recall			2	13	28	52	67	52	58	33	34	31	23	32	18	443
Days recall					3	2	8	6		2	2		1	1		25
Month recall		1		2	3	22	24	32	33	26	22	23	25	18	20	251
Months recall						2	2	4	2	1	4	1	1	2		19
Year recall					4	6	14	21	20	27	16	16	19	26	21	190
Years recall								2	1	1	1					5
No recall					2	12	12	20	33	45	60	65	77	79	82	487
Total	155	156	151	161	143	145	140	141	148	135	139	136	146	158	141	2195

to model the different events of partial recall as functions of time elapsed since the event of interest. There is need for a different model altogether, where calendar time (rather than time since the event) would be relevant. Developing methods of inference under a suitable model would be an interesting research problem.

The possibility of error in recall has been ignored in the models used in Chapters 2-4. Another direction of future research could be inclusion of the possibility of error in recall data, noted by several researchers (for example see Rabe-Hesketh, Yang and Pickles (2001) and Beckett et al. (2001)). The dichotomization of the recall information used here, where all 'partial' recall data have been ignored and regarded as cases of no recall, reduces the impact of recall error. A better handling of the issue through more sophisticated modeling may be attempted.

As mentioned in Section 3.11, modeling non-recall through a forgetting function may be adapted to the estimation of the distribution of the time from contracting HIV infection through blood transfusion to the onset of AIDS. While modeling recall uncertainty in that section, we had ignored the issue of truncation in data. Incorporation of truncation leads to the following likelihood.

$$\prod_{i=1}^{n}\left\{\int_{l_i}^{r_i}\frac{f(S_i-y)}{F(B-y)-F(A-y)}\pi_\eta(S_i-y)dy\right\}^{\varepsilon_i} \cdot\left\{\frac{f(S_i-Y_i)}{F(B-Y_i)-F(A-Y_i)}(1-\pi_\eta(S_i-Y_i))\right\}^{1-\varepsilon_i}, \qquad (5.1)$$

where A and B are the opening and closing dates of registry in respect of the date of onset of AIDS for the data at hand. The above likelihood is more complicated than the likelihood (3.47) and thus a different computational algorithm may be needed for its maximization.

In this thesis, the baseline distribution has been assumed to be absolutely continuous. In the case of coarsely grouped data (say, for recall data with time measured in months) a discrete time model and related methods have to be developed.

As mentioned in the concluding section of Chapter 4, the consistency of the proposed estimator of the regression coefficients has not been established. This problem is yet to be solved even in the case of non-informatively interval censored data, except in some special cases (Huang, 1996; Yu, Wong and Kong, 2006; Liu and Shen, 2009). A study of consistency in the general case would be useful.

The regression model mentioned in Chapter 4 may be extended to include frailty, i.e., shared or individual random effects. In order to get an idea of the nature of the problem at hand, consider individual frailty w_i included in the Cox regression model, so that the hazard rate for the ith subject, given the frailty, is of the form

$$\lambda_i(t) = \lambda_0(t) w_i \exp(\beta^T Z_i), \tag{5.2}$$

where $\lambda_0(t)$ is the unspecified baseline hazard rate, Z_i is the vector of covariates for the ith individual, β is the vector of regression coefficients, and w_i is the unobserved frailty for the ith individual. The w_i's are independent and identically distributed samples from a distribution with mean 1 and some unknown variance. We make the commonly used assumption that the w_i's are independent and identically distributed samples from the gamma distribution with density function

$$g(w;\xi) = \frac{w^{(\frac{1}{\xi}-1)} exp(-w/\xi)}{\Gamma(1/\xi)\xi^{1/\xi}}, \quad w > 0, \quad \xi > 0.$$

It follows that the survival function can be written as (Klein and Moeschberger, 2003)

$$\bar{F}_i(t) = \left(1 + \xi\Lambda_0(t)\exp(\beta^T Z_i)\right)^{-1/\xi}, \tag{5.3}$$

where, $\Lambda_0(t) = \int_0^t \lambda_0(u)du$. Thus, the likelihood (4.2) for the present regression model, with π_η as in (3.9), simplifies to

$$L = \prod_{i=1}^{n} [(1 + \xi\Lambda_0(S_i)\exp(\beta^T Z_i))^{-1/\xi}]^{1-\delta_i}$$
$$\Bigg[\Bigg\{\left((1+\xi\Lambda_0(T_i-)\exp(\beta^T Z_i))^{-1/\xi} - (1+\xi\Lambda_0(T_i)\exp(\beta^T Z_i))^{-1/\xi}\right)$$
$$\cdot \left(1 - \sum_{l=1}^{k} b_l I_{(W_{l+1}(S_i)<T_i\le W_l(S_i))}\right)\Bigg\}^{\varepsilon_i}$$
$$\cdot \Bigg\{\sum_{l=1}^{k} b_l\big((1+\xi\Lambda_0(W_{l+1}(S_i))\exp(\beta^T Z_i))^{-1/\xi}$$
$$- (1+\xi\Lambda_0(W_l(S_i))\exp(\beta^T Z_i))^{-1/\xi}\big)\Bigg\}^{1-\varepsilon_i}\Bigg]^{\delta_i}. \tag{5.4}$$

The form of likelihood (5.4) is more complicated than likelihood (4.16). While $\Lambda_0(\cdot)$ can be written in terms of the masses allocated to different intervals as per the baseline distribution, the presence of the additional nuisance parameter ξ complicates the likelihood to the extent that the computational approach of Chapter 4 does not lead to an elegant iterative procedure. A better solution for (5.3) and other frailty models would require further research.

It would be interesting to explore other regression models for duration data, such as the proportional odds ratio model or the additive hazard model. Under the proportional odds ratio model (Sun, 2006), we have for an individual with covariate vector Z_i

$$\log\left(\frac{\bar{F}_i(t)}{F_i(t)}\right) = \log\left(\frac{\bar{F}_0(t)}{F_0(t)}\right) - \beta^T Z_i, \tag{5.5}$$

where $\bar{F}_0$ is the baseline survival function. Thus, the likelihood (4.2) for the present regression model, with π_η as in (3.9), simplifies to

$$
\begin{aligned}
&L(q_1,\dots,q_v,\beta)\\
&=\prod_{i\in\mathcal{I}_1}\left(\frac{\exp(-\beta^T Z_i)\sum_{j:J_j\subseteq A_i} q_j}{1+(\exp(-\beta^T Z_i)-1)\sum_{j:J_j\subseteq A_i} q_j}\right)\times\prod_{i\in\mathcal{I}_2}\left(1-\sum_{l=1}^{k} b_l I_{\left(T_i\in A_{i(l+1)}\setminus A_{il}\right)}\right)\\
&\cdot\left[\left(\frac{\exp(-\beta^T Z_i)\sum_{j:J_j\subseteq A_i} q_j}{1+(\exp(-\beta^T Z_i)-1)\sum_{j:J_j\subseteq A_i} q_j}\right)-\left(\frac{\exp(-\beta^T Z_i)\sum_{j:J_j\subseteq A_i'} q_j}{1+(\exp(-\beta^T Z_i)-1)\sum_{j:J_j\subseteq A_i'} q_j}\right)\right]\times\\
&\prod_{i\in\mathcal{I}_3}\left[\sum_{l=1}^{k} b_l\left\{\left(\frac{\exp(-\beta^T Z_i)\sum_{j:J_j\subseteq A_{i(l+1)}} q_j}{1+(\exp(-\beta^T Z_i)-1)\sum_{j:J_j\subseteq A_{i(l+1)}} q_j}\right)\right.\right.\\
&\qquad\left.\left.-\left(\frac{\exp(-\beta^T Z_i)\sum_{j:J_j\subseteq A_{il}} q_j}{1+(\exp(-\beta^T Z_i)-1)\sum_{j:J_j\subseteq A_{il}} q_j}\right)\right\}\right], \qquad (5.6)
\end{aligned}
$$

where A_i, A_i' and A_{il} are as defined in (3.13). Likewise, under the additive hazards regression model (Sun, 2006), we have for an individual with covariate vector Z_i

$$\lambda_i(t)=\lambda_0(t)+\beta^T Z_i, \qquad (5.7)$$

where λ_0 is the baseline hazard. Thus, the likelihood (4.2) with π_η as in (3.9) reduces to

$$
\begin{aligned}
&L(q_1,\dots,q_v,\beta)\\
&=\prod_{i\in\mathcal{I}_1}\left(\frac{\sum_{j:J_j\subseteq A_i} q_j}{1+\sum_{j:J_j\subseteq A_i} q_j\exp(\beta^T Z_i)}\right)\times\prod_{i\in\mathcal{I}_2}\left(1-\sum_{l=1}^{k} b_l I_{\left(T_i\in A_{i(l+1)}\setminus A_{il}\right)}\right)\\
&\quad\cdot\left[\left(\frac{\sum_{j:J_j\subseteq A_i} q_j}{1+\sum_{j:J_j\subseteq A_i} q_j\exp(\beta^T Z_i)}\right)-\left(\frac{\sum_{j:J_j\subseteq A_i'} q_j}{1+\sum_{j:J_j\subseteq A_i'} q_j\exp(\beta^T Z_i)}\right)\right]\times\\
&\prod_{i\in\mathcal{I}_3}\left[\sum_{l=1}^{k} b_l\left\{\left(\frac{\sum_{j:J_j\subseteq A_{i(l+1)}} q_j}{1+\sum_{j:J_j\subseteq A_{i(l+1)}} q_j\exp(\beta^T Z_i)}\right)-\left(\frac{\sum_{j:J_j\subseteq A_{il}} q_j}{1+\sum_{j:J_j\subseteq A_{il}} q_j\exp(\beta^T Z_i)}\right)\right\}\right],
\end{aligned}
\qquad (5.8)
$$

where A_i, A_i' and A_{il} are as defined in (3.13). The forms of likelihoods (5.6) and (5.8) are not as simple as (4.16), and therefore new computational techniques have to be devised for analysis under these regression models. For the accelerated failure time model,

$$\bar{F}_i(t) = \bar{F}_0(t\exp(\beta^T Z_i)), \tag{5.9}$$

the likelihood (4.2) with π_η as in (3.9) reduces to

$$L = \prod_{i=1}^{n} \left[\bar{F}_0(S_i \exp(\beta^T Z_i))\right]^{1-\delta_i} \left[\left\{ \left(\bar{F}_0(T_i \exp(\beta^T Z_i)-) - \bar{F}_0(T_i \exp(\beta^T Z_i))\right) \cdot \left(1 - \sum_{l=1}^{k} b_l I_{\left(W_{l+1}(S_i) < T_i \le W_l(S_i)\right)}\right)\right\}^{\varepsilon_i} \cdot \left\{ \sum_{l=1}^{k} b_l \left(\bar{F}_0(W_{l+1}(S_i)\exp(\beta^T Z_i)) - \bar{F}_0(W_l(S_i)\exp(\beta^T Z_i))\right)\right\}^{1-\varepsilon_i} \right]^{\delta_i}. \tag{5.10}$$

This likelihood is even more complicated, as different values of β would produce different arguments of the function $\bar{F}_0$. Therefore, the strategy of optimization over masses attached to exactly recalled times of event would not work. A different approach will be needed for this model.

Bibliography

AALEN, O. O. (1975). *Statistical Inference for a Family of Counting Processes.* Ph.D. dissertation, Department of Statistics, University of California, Berkeley. 4

AKSGLAEDE, L., SORENSEN, K., PETERSEN, J. H., SKAKKEBAK, N. E. and JUUL, A. (2009). Recent decline in age at breast development: The Copenhagen Puberty Study. *Pediatrics* **123,** 932-939. 1, 35

ALIOUM, A. and COMMENGES, D. (1996). A proportional hazards model for arbitrarily censored and truncated data. *Biometrics* **52** 512-524. 5

ALLISON, P. D. (1982). Discrete-time methods for the analysis of event histories. *Sociological Methodology* **13** 61-98. 1

ANDERSEN, P. K. and GILL, R. D. (1982). Cox's regression model for counting processes: A large sample study. *Annals of Statistics* **10** 1100–1120. 90

ARYEETEY, R., ASHINYO, A. and ADJUIK, M. (2011). Age at menarche among basic level school girls in Medina, Accra. *African J. Reprod. Health* **103** 103–110. 93

ASH, R. B. (2000). *Probability and Measure Theory.* Harcourt/Academic Press, Burlington, MA. 16

BAGDONAVICIUS, V. and NIKULIN, M. R. (2003). Statistical modeling in survival analysis and its influence on the duration analysis. In Balakrishnan, N. and Rao, C., editors,. *Advances in Survival Analysis. Handbook of Statistics* **23** 411–429. 5

BECKETT, M., DAVANZO, J., SASTRY, N., PANIS, C. and PETERSON, C. (2001). The quality of retrospective data: An examination of long-term recall in a developing country. *The Journal of Human Resources* **36,** 593-625. 102

BERGSTEN-BRUCEFORS, A. (1976). A note on the accuracy of recalled age at menarche. *Annals of Human Biology* **3** 71-73. 1

BETENSKY, R. A. (2000). On nonidentifiability and noninformative censoring for current status data. *Biometrika* **87** 218–221. 5

BICKEL, P. J., GOTZE, F. and VAN ZWET, W. R. (1997). Resampling fewer than n observations: gains, losses, and remedies for losses. *Statistca Sinica* **7** 1–31. 54, 64, 67

BICKEL, P. J. and SAKOV, A. (2008). On the choice of m in the m out of n bootstrap and confidence bounds for extrema. *Statistica Sinica* **18** 967–985. 54, 64

BILLINGSLEY, P. (1968). *Convergence of Probability Measures.* John Wiley, New York-London-Sydney. 55, 56

BRESLOW, N. E., LUBIN, J. H., MAREK, P. and LANGHOLZ, B. (1983). Multiplicative models and cohort analysis. *Journal of the American Statistical Association,* **78** 1-12. 4

CAMERON, N. (2002). *Human Growth and Development.* Academic Press. 1

CHAUDHURI, P. (2007). On single index regression models for multivariate survival time data. In Nair, V., editor,. *Advances in Statistical Modeling and Inference: Essays in Honor of Kjell A* 223-232. 5

CHEN, D. G., SUN, J. and PEACE, K. E. (2013). *Interval-Censored Time-to-Event Data.* Chapman & Hall, London. 71

CHUMLEA, W. C., SCHUBERT, C. M., ROCHE, A. F., KULIN, H. E., LEE, P. A., HIMES, J. H. and SUN, S. S. (2003). Age at menarche and racial comparisons in US girls. *Pediatrics* **11**, 110-113. 1

CLEMENTS, M. S., MITCHELL, E. A., WRIGHT, S. P., ESMAIL, A., JONES, D. R. and FORD, P. (1997). Influences on breastfeeding in southeast England. *Acta Paediatrica* **86** 51–56. 1

COOK, R. J. and LAWLESS, J. (2007). *The Statistical Analysis of Recurrent Events.* Statistics for Biology and Health. Springer, New York. 6

COX, D. R. (1972). Regression models and life-tables. *Journal of Royal Statistical Society, Series B* **34** 187–220. With discussion by F. Downton, Richard Peto, D. J. Bartholomew, D. V. Lindley, P. W. Glassborow, D. E. Barton, Susannah Howard, B. Benjamin, John J. Gart, L. D. Meshalkin, A. R. Kagan, M. Zelen, R. E. Barlow, Jack Kalbfleisch, R. L. Prentice and Norman Breslow, and a reply by D. R. Cox. 4, 75

DABROWSKA, D. M. and DOKSUM, K. A. (1988). Estimation and testing in a two-sample generalized odds-rate model. *Journal of American Statistical Association* **83** 744–749. 4

DE LA FUENTE, A. (2000). *Mathematical Methods and Models for Economists.* Cambridge University Press, Cambridge. 49

DEMIRJIAN, A., GOLDSTJEN, H. and TANNER, J. M. (1973). A new system of dental age assessment. *Annals of Human Biology* **45** 211–227. 1

DUNSON, D. B. and DINSE, G. E. (2002). Bayesian models for multivariate current status data with informative censoring. *Biometrics* **58**, 79–88. 5

EFRON, B. (1967). The two sample problem with censored data. *Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability* 831-853. 4, 50

EVELETH, P. B. and TANNER, J. M. (1990). *Worldwide Variation in Human Growth*, 2nd ed. Cambridge University Press. 1

FERGUSON, T. S. (1996). *A Course in Large Sample Theory.* Texts in Statistical Science Series. Chapman & Hall, London. 20

FINKELSTEIN, D. M., MOORE, D. F. and SCHOENFELD, D. A. (1993). A proportional hazards model for truncated AIDS data. *Biometrics* **49** 731–740. 5

FINKELSTIEN, D. M., GOGGINES, W. B. and SCHOENFELD, D. A. (2002). Analysis of failure time data with dependent interval censoring. *Biometrics* **58**, 298–304. 7

GENTLEMAN, R. and GEYER, C. J. (1994). Maximum likelihood for interval censored data: consistency and computation. *Biometrika* **81** 618-623. 55, 56

GIBBONS, J. D. and CHAKRABORTI, S. (2003). *Nonparametric Statistical In-*

ference **168**, Fourth ed. Statistics: Textbooks and Monographs. Marcel Dekker Inc., New York. 29

HEDIGER, M. L. and STINE, R. A. (1987). Age at menarche based on recall data. *Annals of Human Biology* **14** 133-142. 5

HOSMER, D. W., LEMESHOW, S. and MAY, S. (2008). *Applied Survival Analysis*, Second ed. Wiley Series in Probability and Statistics. Wiley-Interscience/ John Wiley, Hoboken, NJ. 1, 4

HUANG, J. (1996). Efficient estimation for the proportional hazards model with interval censoring. *The Annals of Statistics* **24** 540–568. 97, 103

HUANG, J. (1999). Asymptotic properties of nonparametric estimation based on partly interval-censored data. *Statistica Sinica* **9** 501–520. 70

ISI (2012). *Annual Report of the Indian Statistical Institute 2011-12.* Available at URL http://library.isical.ac.in/jspui/handle/10263/5345?mode=full. 9

JOHANSEN, S. (1983). An extension of Cox's regression model. *International Statistical Review* **51** 165–174. 90

KACIROTI, N. A., RAGHUNATHAN, T. E. and TAYLOR, J. M. G. (2012). A Bayesian model for time-to-event data with informative censoring. *Biostatistics* **13**, 341–354. 7

KALBFLEISCH, J. D. and LAWLESS, J. F. (1989). Inference based on retrospective ascertainment: an analysis of the data on transfusion-related AIDS. *Journal of American Statistical Association* **84** 360–372. 71, 72

KALBFLEISCH, J. D. and PRENTICE, R. L. (2002). *The Statistical Analysis of Failure Time Data.* John Wiley, New York. 3, 11

KAPLAN, E. L. and MEIER, P. (1958). Nonparametric estimation from incomplete observations. *Journal of American Statistical Association* **53** 457-481. 3, 39, 50

KEIDING, N., BEGTRUP, K., SCHEIKE, T. H. and HASIBEDER, G. (1996). Estimation from current-status data in continuous time. *Lifetime Data Analalysis* **2**, 119-129. 6

KHAN, A. D., SCHROEDER, D. G., MARTORELL, R., HAAS, J. D. and

RIVERA, J. (1996). Early childhood determinants of age at menarche in rural Guatemala. *Amer. J. Hum. Biol.* 8 717–723. 93

KLEIN, J. P. and MOESCHBERGER, M. L. (2003). *Survival Analysis: Techniques for Censored and Truncated Data.* Springer-Verlag, New York. 1, 4, 103

KORN, E. L., GRAUBARD, B. I. and MIDTHUNE, D. (1997). Time-to-event analysis of longitudinal follow-up of a survey: choice of the time-scale. *American Journal of Epidemiology* **145** 72-80. 5

LAWLESS, J. F. (1982). *Statistical Models and Methods for Lifetime Data.* John Wiley, New York. 3

LAWLESS, J. F. (2003). *Statistical Models and Methods for Lifetime Data,* Second ed. John Wiley, New York. 3, 4

LECLERE, M. J. (2005). Modeling time to event: Applications of survival analysis in accounting, economics and finance. *Review of Accounting and Finance* **4** 5–12. 1

LEE, E. T. and WANG, J. W. (2003). *Statistical Methods for Survival Data Analysis.* John Wiley. 3

LEHMAN, E. L. (1999). *Elements of Large-Sample Theory.* Springer-Verlag, New York. 19

LIU, B., LU, W. and ZHANG, J. (2014). Accelerated intensity frailty model for recurrent event data. *Biometrics* **70** 579–587. 5

LIU, H. and SHEN, Y. (2009). A semiparametric regression cure model for interval-censored data. *Journal of American Statistical Association* **104** 1168–1178. 97, 103

MCKAY, H. A., BAILEY, D. B., MIRWALD, R. L., DAVISON, K. S. and FAULKNER, R. A. (1998). Peak bone mineral accrual and age at menarche in adolescent girls: A 6-year longitudinal study. *The Journal of Pediatrics* **13** 682–687. 5

MILLER, R. G. (1981). *Survival Analysis.* John Wiley. 4

MIRZAEI, S. S. and SENGUPTA, D. (2015a). Nonparametric estimation of time-to-event distribution based on recall data in observational studies. *Lifetime Data*

Analysis to be published, DOI:10.1007/s10985-015-9338-8. 9

MIRZAEI, S. S. and SENGUPTA, D. (2015b). Regression under Cox's model for recall-based time-to-event data in observational studies. *Computational Statistics & Data Analysis* to be published, DOI:10.1016/j.csda.2015.07.005. 9

MIRZAEI, S. S., SENGUPTA, D. and DAS, R. (2015). Parametric estimation of menarcheal age distribution based on recall data. *Scandinavian Journal of Statistics.* to be published, DOI:10.1111/sjos.12107. 1, 9

NAKAMURA, T. (1991). Existence of maximum likelihood estimates for interval-censored data from some three-parameter models with a shifted origin. *Journal of Royal Statistical Society, Series B* **53** 211–220. 5

NELSON, E. N., IGLESIAS, C. P., CULLUM, N. and TORGERSON, D. J. (2004). Randomized clinical trial of four-layer and short-stretch compression bandages for venous leg ulcers (VenUS I). *British Journal of Surgery* **91** 1292–1299. 1

NOCEDAL, J. and WRIGHT, S. J. (2006). *Numerical Optimization*, Second ed. Springer Series in Operations Research and Financial Engineering. Springer, New York. 24

PADEZ, C. (2003). Age at menarche of schoolgirls in Maputo, Mozambique. *Annals of Human Biology* **30,** 487–495. 6, 93

RABE-HESKETH, S., YANG, S. and PICKLES, A. (2001). Multilevel models for censored and latent responses. *Statistical Methods in Medical Research* **10,** 409-427. 9, 102

ROBERTS, D. F. (1994). Secular trends in growth and maturation in British girls. *American Journal of Human Biology* **6,** 13–18. 6

SCHARFSTEIN, D. O. and ROBINS, J. M. (2002). Estimation of the failure time distribution in the presence of informative censoring. *Biometrika* **89,** 617–634. 7

SEN, B., BANERJEE, M. and WOODROOFE, M. (2010). Inconsistency of bootstrap: the Grenander estimator. *Annals of Statistics* **38** 1953–1977. 54

SIMON, C. P. and BLUME, L. (1994). *Mathematics for Economists.* W. W. Norton, New York. 49

Skinner, C. J. and Humphreys, K. (1999). Weibull regression for lifetimes measured with error. *Lifetime Data Anal.* **5** 23–37. . MR1750340 34

Stine, R. A. and Small, R. D. (1986). *Estimating the Distribution of Censored Logistic Recall Data* **83**,. Technical Report, Department of Statistic, University of Pennsylvania. 14

Sun, J. (2006). *The Statistical Analysis of Interval-censored Failure Time Data.* Springer, New York. 6, 12, 104, 105

Tanaka, Y. and Rao, P. V. (2005). A proportional hazards model for informatively censored survival times. *Journal of Statistical Planning and Inference* **129** 253–262. 7

Teilmann, G., Petersen, J. H., Gormsen, M., Damgaard, K., Skakkebaek, N. E. and Jensen, T. K. (2009). Early puberty in internationally adopted girls: Hormonal and clinical markers of puberty in 276 girls examined biannually over two years. *Hormone Research Paediatrics* **72**, 236–246. 5

Turnbull, B. W. (1976). The empirical distribution function with arbitrarily grouped, censored and truncated data. *Journal of Royal Statistical Society, Series B* **38** 290–295. 4, 6, 35, 41, 50

Vonta, F. (1996). Efficient estimation in a non-proportional hazards model in survival analysis. *Scandinavian Journal of Statistics* **23** 49–61. 4

Wang, J.-L. (1985). Strong consistency of approximate maximum likelihood estimators with applications in nonparametrics. *Annals of Statistics* **13** 932–946. 55, 56

Wei, L. J. (1992). The accelerated failure time model: A useful alternative to the Cox regression model in survival analysis (with discussion). *Statistics in Medicine* **11** 1871–1879. 4

Wen, C.-C. and Chen, Y.-H. (2014). Functional inference for interval-censored data in proportional odds model with covariate measurement error. *Statistica Sinica* **24**. 9

Wienke, A. (2010). *Frailty Models in Survival Analysis.* Chapman and Hall / CRC Press, Boca Raton, FL. 5

YU, Q., WONG, G. Y. C. and LI, L. (2001). Asymptotic properties of self-consistent estimators with mixed interval-censored data. *Annals of the Institute of Statistical Mathematics* **53** 469–486. 71

YU, Q., WONG, G. Y. C. and KONG, F. (2006). Consistency of the semi-parametric MLE in linear regression models with interval-censored data. *Scandinavian Journal of Statistics* **33** 367–378. 97, 103

ZHIGUO, L., SHIYU, Z., SURESH, C. and CRISPIAN, S. (2007). Failure event prediction using the Cox proportional hazard model driven by frequent failure signatures. *IIE Transactions* **39** 303–315. 1

List of publications by the author

- S. Mirzaei S. and Sengupta, D. 'Human Growth Curve Estimation Through a Combination of Longitudinal and Cross-sectional Data.' *World Academy of Science, Engineering and Technology (WASET)*: **67** pp 865-871. (2012).
- S. Mirzaei Salehabadi. and Sengupta, D. 'A New Technique for Estimating Population Distribution of Growth Curve Parameters with Longitudinal and Cross-sectional Data.' *Springer Proceeding in Mathematics and Statistics*: **46** pp 171-183. (2013).
- S. Mirzaei Salehabadi, Das, R. and Sengupta, D. 'Parametric estimation of menarcheal age distribution based on recall data.' *Scandinavian Journal of Statistics.* [DOI:10.1111/sjos.12107] (2015).
- S. Mirzaei Salehabadi, Sengupta, D. 'Nonparametric estimation of time-to-event distribution based on recall data in observational studies.' *Lifetime Data Analysis* [DOI: 10.1007/s10985-015-9345-9] (2015).
- S. Mirzaei xSalehabadi, Sengupta, D. 'Regression under Cox's Model for Recall-based Time-to-Event Data in Observational Studies.' *Computational Statistics & Data Analysis.* [Doi: 10.1016/j.csda.2015.07.005] (2015).