ABSTRACT

Title of dissertation: MARKOV MULTI-STATE MODELS FOR SURVIVAL ANALYSIS WITH RECURRENT EVENTS

Tianhui Zhang
Doctor of Philosophy, 2019

Dissertation directed by: Professor Grace Yang
Statistics Program,
Department of Mathematics

Markov models are a major class within the system of multi-state models for the analysis of lifetime or event-time data. Applications abound, including the estimation of lifetime of ultra-cold neutrons, the bias correction of the apparent magnitude distribution of the stars in a certain area of the sky, and the survival analysis of clinical trials. This thesis addresses some of the problems arising in the analysis of right-censored lifetime data. Clinical trials are used as examples to investigate these problems. A Markov model that takes a patient's disease development into account for the analysis of right-censored data was first constructed by Fix and Neyman (1951). The Fix-Neyman (F-N) model is a homogeneous Markov process with two transient and two absorbing states that describes a patient's status over a period of time during a cancer clinical trial.

This thesis extends the F-N model by assuming the transition rates (hazard rates) to be both state and time dependent. Recurrent transitions between relapse and recovery are allowed in the extended model. By relaxing the condition of time-

independent hazard rates, the extension increases the applicability of the Markov models. The extended models are used to compute the model survival functions, cumulative hazard functions that take into consideration of right censored observations as it has been done in the celebrated Kaplan-Meier estimator. Using the Fix-Neyman procedure and the Kolmogorov forward equations, closed-form solutions are obtained for certain irreversible 4-state extended models while numerical solutions are obtained for the model with recurrent events. The 4-state model is motivated by an Aplastic Anemia data set. The computational method works for general irreversible and reversible models with no restriction on the number of states.

Simulations of right-censored Markov processes are performed by using a sequence of competing risks models. Simulated data are used for checking the performance of nonparametric estimators for various sample sizes. In addition, applying Aalen's (1978) results, estimators are shown have asymptotic normal distributions.

A brief review of some of the literature relevant to this thesis is provided. References are readily available from a vast literature on the survival analysis including many text books. General Markov process models for survival analysis are described, e.g., in Andersen, Borgan, Gill and Keiding (1993).

MARKOV MULTI-STATE MODELS FOR SURVIVAL ANALYSIS WITH RECURRENT EVENTS

by

Tianhui Zhang

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2019

Advisory Committee:
Dr. Grace Yang, Advisor
Dr. Paul J. Smith
Dr. Benjamin Kedem
Dr. Joan Jian-Jian Ren
Dr. Xin He

Dedication

To my family.

Acknowledgments

Firstly, I would like to express my sincere appreciation to my advisor, Dr. Grace Yang for her knowledgeable guidance and encouragement over the past several years. Her advice on both research and career have been invaluable. Dr. Yang spent countless hours proofreading my research and discussing new research ideas with me. I am very fortunate to have Dr. Yang as my advisor and own immeasurable thanks for all she has done for me.

Secondly, I would like to thank my committee members. I benefit a lot from Dr. Paul Smith's linear model class and Dr. Joan Ren's survival analysis class. I'm grateful to Dr. Benjamin Kedem and Dr. Xin He for their willingness to take time out of their busy schedules to serve on my committee. I would also like to thank the faculty in the Statistics Program at the University of Maryland for the substantial guidance and help they have provided throughout my graduate studies.

Last but not least, I would like to express my deepest gratitude to my parents and friends for their support and encouragement. None of my work would have been possible without them.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

AA Aplastic Anemia
IST Immunosuppressants Therapy

Chapter 1: Introduction

Statistical analyses of lifetime data are routinely carried out in a broad spectrum of scientific disciplines. Depending on the area of applications, a lifetime could be the lifetime of a human being or an item produced in a factory, or the doubling time of the size of a malignant tumour, or the survival time of a cancer patient measured fom the time of diagnosis, or the time to disintegration of an ultra-cold neutron, or the time to move to another house, and numerous others. In other words, a lifetime is considered as a waiting time to an event under study which is measured from a well-defined initial condition of a subject under study to the time of occurrence of a specific event of interest. Thus, a lifetime, a survival time and an event time will be used interchangeably in this thesis.

Of particular interest in the statistical analysis of lifetime data are the study of the survival function, life expectation function, failure rate (hazard rate), cumulative hazard rate function. Under some conditions, these functions provide alternative ways of describing the distribution of a lifetime. Widely used life tables are con-

structed that display simultaneously the estimates of these functions based on data sets of lifetimes.

Collection of lifetime data is by no means simple. The method of collection determines how a model for lifetime data should be constructed for statistical analysis. Due to technical difficulties that vary with the field of applications, sampling limitations, and cost considerations, it is rather unusual to obtain a data set that has all completely observed lifetimes. Typically some of the lifetime observations in the data set can only be partially recorded. Statistical theory shows that partially observed data must be included in the statistical analysis. Otherwise the findings would be biased. Inclusion of partially observed lifetimes in the analysis depends on the sampling plan for data collection and the model constructed for the data.

Depending on the particular sampling plan employed, an observed lifetime can be incomplete in the sense that it is censored, randomly truncated, length biased, size biased, interval censored or possibly other forms of incompleteness. There is a vast statistical literature dealing with statistical theory, model constructions and applications of incomplete lifetimes.

This thesis addresses some of the statistical problems arising in right-censored data. One of the most widely used model is the traditional right-censoring models (to be called right-censoring model for short) for survival analysis which refers to

the joint distribution of (V, δ) defined in (2.1) under the stochastic independence assumption of X and C. Some (Meira-Machado *et al.* (2009) [37]) call it a mortality model where the concern is with the time-to-death of a patient regardless a patient's disease development leading to the end point of death. A stochastic model that takes patient's disease development into account for the analysis of right-censored data was constructed by Fix and Neyman (1951) [23]. The Fix-Neyman (F-N) model is a homogeneous Markov process with two transient and two absorbing states that describes a patient's status during a cancer clinical trial where the statuses are in-remission (or recovery), relapse, death and loss to follow up. When casting the right-censoring model into a 3-state Markov process with one transient state and two absorbing states of death and loss to follow up, the right-censoring model becomes a special case of the F-N model. The difference between the models is that all of the transition rates (hazard rates) in the F-N model are assumed to be independent of the time t while those in the right-censoring model can be both state and time dependent.

In this thesis, we extend the F-N model by letting all of the transition rates (risks) to be both state and time dependent. That is, we extend the Fix-Neyman model to a non-homogeneous finite-state Markov process. The extension is necessary as it is well known that in many applications, time-independent hazard rates do not

fit the data. A glaring example is the survival function of human lifetimes shown in Figure 1.1. The survival curve does not have an exponential distribution. It would be too good to be true if the survival curve had an exponential distribution with age-independent failure.

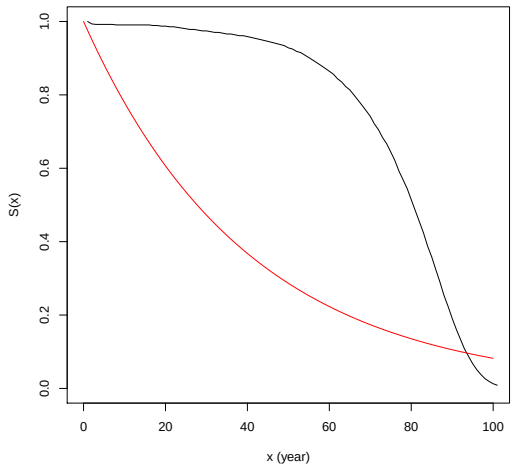


Figure 1.1: Comparison of an exponential survival function and the survival function of lifetime obtained by converting the life expectancy function given in the United States Life Tables 2008 (Arias (2012) [10]). The red line is the exponential survival function. The black line is the actual survival function of lifetime.

Applying the Fix-Neyman (F-N) procedure, in Section 2.2 we derive the survival function of an individual that takes into account right-censored observations (due to the risk of loss to follow-up). This result answers an identifiability problem, namely that from right-censored non-homogeneous Markov process, the survival function of an individual under study is identifiable. Although identifying

the survival function using the F-N procedure is elegantly simple, computational complexities increase with the number of states in the model. For instance, it is straightforward to compute the survival function in the right-censoring model, but for the extended F-N model with time and state-dependent transition rates, an explicit analytical solution for the survival function is only available in some special cases. For more complicated transition patterns, we obtained the solutions for the survival function numerically.

Markov models have been used broadly in the analysis of lifetime data. General Markov process models for survival analysis are succinctly described in Andersen, Borgan, Gill and Keiding (Chapter III.1.2, 1993) [7]. The Fix-Neyman (1951) [23] paper was discussed in early years after its publication, e.g. Chiang (1968) [14], Cox (1972) [18], Prentice and Kalbfleisch (1978) [43]. But it seems to have disappeared in more recent literature. As far as we know, the F-N procedure is hardly used. Although the F-N procedure may not be needed for some Markov models with 2 or 3 states and simple transition paths, for more states with complicated transition paths, the advantage of the F-N procedure becomes clear.

Chapter 2 gives a brief review of some of the literature relevant to this thesis. Section 2.2 presents the F-N procedure and compares it with the usual method of calculating the survival function in the right-censoring model. Details can be found

in Yang (2013) [50]. Section 2.3 shows that technically the Markov models considered in this thesis unify the study of the survival function for some important multi-state models such as the multiple decrement model, the illness-death model, Chiang's staging model, the progressive model and the F-N model. The section also includes a review of an extension to models for bivariate survival lifetimes. Section 2.4 deals with various approaches to estimation in Markov multi-state models. Parametric, nonparametric estimation and regression problems are reviewed separately. Section 2.5 notes some more recent results on non-Markovian models.

Chapter 3 is devoted to the construction of an irreversible Markov model. The model is motivated by an clinical trial of patients with Aplastic Anemia(AA). We use an irreversible 4-state non-homogeneous Markov model $\boldsymbol{\xi}^o$ to describe the progression of a patient with AA. This is covered in Section 3.2. Section 3.3 gives a construction of the model by using competing risks at every transition time. The construction is needed for estimation and simulation. In Section 3.4, we state the estimation problems of the cumulative hazard functions, transition probabilities, and survival function and explain the available data. Section 3.5 carries out the estimation problems. Section 3.5.1 constructs the estimators for cumulative hazard functions. Section 3.5.2 deals with estimation of transition probabilities. The estimation requires solving the Kolmogorov forward equations for explicit analytical

forms of the transition probabilities. Solutions are obtained. Section 3.5.3 derives the model solution for the sought-after survival function $S(t)$ which is a function of transition probabilities obtained in Section 3.5.2. the estimator of the survival function is obtained whose distributional properties are checked by simulation data. Section 3.7 shows the estimator of the cumulative hazard function constructed in this thesis is identical to that given by Aalen (1978b) [2]. Therefore the asymptotic properties established by Aalen apply to our estimators as well.

Chapter 4 contains the simulation of irreversible Markov processes. Simulated data are used to examine the distributional properties of the estimators developed in Chapter 3. Results are presented in Section 4.2. In Section 4.3, we carry out goodness of fit testing of the model by statistical hypothesis testing using both a modified Kolmogorov-Smirnov test and a chi-squared test.

Chapter 5 is concerned with estimation and simulation in reversible Markov models. The treatment is similar but not identical to that for irreversible Markov models. Only differences between the two are discussed.

Chapter 6 concludes this thesis.

Chapter 2: Literature Review

2.1 Introduction

Markov models are a major part of the multi-state models. The review is primarily on Markov multi-state models.

Section 2.2 discusses the F-N procedure and compares it with the usual method of calculating the survival function in the traditional right-censoring model. Stochastic models for lifetime data and statistical estimation are reviewed separately in Sections 2.3 and 2.4 respectively. Section 2.5 notes some recent development on non-Markovian models. Section 2.6 reviews several well-known models for recurrent event analysis.

2.2 Comparison of the Fix-Neyman Model and the Right-Censoring Model

It is convenient to compare the F-N model and the right-censoring model by an example.

A prototype example of right-censored data is the collection of survival times from a clinical trial of cancer patients who have received a certain method of treatment. After the treatment, each patient is followed up over a period of time. If the death occurs in the followup period, his/her survival time can be determined. Otherwise, it is only known that the survival time is larger than the length of the followup period. However, the length of the follow up period varies with each patient. This could be due to a patient's withdraw from the clinical trial (loss to followup), or the termination of the clinical trial, or possible other reasons. Formally, let X denote the survival time of a patient and C the length of his/her followup time. The observable data on X are given by a pair of random variables

$$V = \min(X, C) \quad \text{and} \quad \delta \tag{2.1}$$

where $\delta = I[X \leq C]$ is an indicator which is one if $X \leq C$ and zero otherwise. Only if $\delta = 1$, is X completely observable. When $\delta = 0$, we say X is right-censored by

C. Under the assumption of stochastic independence of X and C, (2.1) is called the (classical) right-censoring model.

Thus instead of a direct measurement of X, we can only observe V and δ. The question arising is whether the survival function $S(t) = P[X > t]$ for all t can be identified from the joint distribution of V and δ. The answer is affirmative if we assume that X and C are stochastically independent random variables. (To be correct, we also require that the range of C is larger than that of X for nonparametric estimation of the survival function). Under the independence assumption, (2.1) is called a right-censoring model of X. The right-censored data from all n patients are

$$(V_j, \delta_j) \qquad j = 1, \ldots, n. \tag{2.2}$$

The right censored data are broadly used in the estimation of the survival function of a patient as well as in other scientific investigations. Asymptotic properties of the celebrated Kaplan-Meier (K-M) nonparametric estimator (Kaplan and Meier (1958) [32]) of S have been established under the right-censoring model (2.1).

The F-N model (1951) [23] was constructed for analyzing survival times of patients who received a particular treatment of breast cancer. In calculating the survival function of a patient, Fix and Neyman took into account of the available data on each patient's status with regard to relapse, recovery, loss to followup and

death. Let $\xi^o(t)$ represent the status of a patient at time t. Fix and Neyman constructed a model $\{\xi^o(t) : 0 \leq t \leq T)\}$ where a patient's changes in status are governed by a 4-state homogeneous Markov process, and T is the termination time of the clinical trial. The four states are:

S_0: a patient has completed treatment and is in remission

S_1: a patient has relapsed

S_2: a patient is dead

S_3: a patient is lost to followup

At any time t, the infinitesimal matrix Q^o and possible transitions of a patient are shown in Figure 2.1 with transition rates q_{01}, q_{02}, q_{03} from S_0 to S_1, S_2, S_3,

$$\mathbf{Q^o} = \begin{pmatrix} q_{00} & q_{01} & q_{02} & q_{03} \\ 0 & q_{11} & q_{12} & q_{13} \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

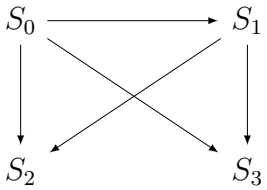


Figure 2.1: F-N model.

respectively and q_{12}, q_{13} from S_1 to S_2, S_3, respectively. Note that Fix and Neyman [23] allow the transition rate $q_{10} > 0$ from S_1 to S_0. For simplicity of discussing the F-N procedure, a simpler version of setting $q_{10} = 0$ is used in Figure 2.1 (see Yang [50]).

Over the course of the clinical trial, a patient can experience one of the following four possible transition paths:

(1) $S_0 \to S_1 \to S_2$

(2) $S_0 \to S_1 \to S_3$

(3) $S_0 \to S_2$

(4) $S_0 \to S_3$

The right-censoring in the F-N model is much more complicated than that of the right-censoring model (2.1). Note that path (2) or (4) results in a right-censored survival time, that is, knowing only the survival time $X^o > t$, while path (1) or (3) gives a complete observation of survival time. For simplicity of illustration, we set $T = \infty$. The survival function of X^o of a patient in the presence of loss to followup can be identified in the Markov process ξ^o by

$$S^o(t) = P[X^o > t] = 1 - P[\xi^o(t) = 2 | \xi^o(0) = 0], \qquad t \geq 0. \tag{2.3}$$

Fix and Neyman showed that the (net) survival function S of a patient can be obtained by by introducing another Markov process $\{\xi(t) : t \geq 0\}$ with states S_0, S_1 and S_2, where $\xi(t)$ represents the status of a patient in the 3-state Markov

process with the risks q_{03} and q_{13} to loss-to-followup eliminated, and the transition rates q_{01}, q_{02}, q_{12} remain the same as in Figure 2.1.

$$\mathbf{Q} = \begin{pmatrix} q_{00} & q_{01} & q_{02} \\ 0 & q_{11} & q_{12} \\ 0 & 0 & 0 \end{pmatrix}$$

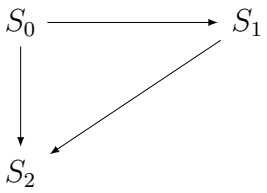


Figure 2.2: F-N model with censoring state eliminated.

The infinitesimal matrix Q and possible transitions in this Markov process are shown in Figure 2.2. The true survival function S is given by

$$S(t) = P[X > t] = 1 - P[\xi(t) = 2 | \xi(0) = 0], \qquad t \geq 0 \tag{2.4}$$

This establishes the identifiability of S from the F-N model.

To compare the F-N model with the right-censoring model, we reformulate the right censoring model (2.1) as a 3-state Markov process, $\{\eta^o(t) : t \geq 0\}$ where $\eta^o(t)$ represents the status of a patient at time t. The three states are S_0, S_2, S_3. At any given time t, the possible transitions of a patient are shown in Figure 2.3 with transition rates q_{02} and q_{03} from S_0 to S_2 and S_3 respectively. Over the time of clinical trial, a patient takes one of the two possible transition paths:

(1) $S_0 \to S_2$

$$\mathbf{Q} = \begin{pmatrix} q_{00} & q_{02} & q_{03} \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

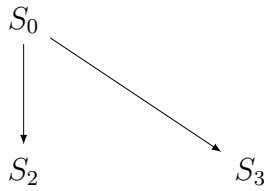


Figure 2.3: Right-censoring model.

(2) $S_0 \to S_3$

It is easy to show that the joint distribution of the right-censored observations $V = \min(X, C)$ and δ in (2.1) correspond to that of the Markov process $\boldsymbol{\eta}^o$ by

$$P[V \le t, \delta = 1] = P[\eta^o(t) = 2 | \eta^o(0) = 0] \tag{2.5}$$

$$P[V \le t, \delta = 0] = P[\eta^o(t) = 3 | \eta^o(0) = 0] \tag{2.6}$$

Here, we take the hazard rates of X and C to be q_{02} and q_{03} respectively. To have exact correspondence, we can assume that q_{02} and q_{03} are time-dependent. To identify the survival function S using the F-N procedure is to derive it from another Markov process,$\{\eta(t) : t \ge 0\}$ which is obtained by eliminating the risk q_{03} of loss-to-followup in the process $\boldsymbol{\eta}^o$ but keeping the same risk q_{02} for S_2. The process $\boldsymbol{\eta}$ has only one transient state S_0 and one absorbing state S_2. It is trivial to compute the survival

function S. It is given by

$$S(t) = P[X \geq t] = P[\eta(t) = 0 | \eta(0) = 0] \tag{2.7}$$

For computing the survival function that is free of right censoring, Neyman (1951) [23] [40] introduced the notion of crude and net survival probabilities in his discussion of an illness and death model. The survival function S^o (see (2.3)) computed from the F-N model for the crude data is called crude survival function. The sought-after survival function S (see (2.4)) that is free from the risk of right-censoring is called the net survival function. Identifying the survival function in the classical right-censoring model is the same problem that Fix and Neyman investigated but for more general right-censoring Markov models.

The F-N model includes the relapse information in the estimation of the survival function, which is a better utilization of the available data than that of the classical right-censoring model in (2.1). However it increases the complexity of the censoring patterns and computations. Fix and Neyman obtained an explicit solution for S for constant transition rates and carried out parametric estimation of S.

With the F-N procedure we can unify the method for identifying and computing survival probabilities for a large class of Markov multi-state models with time and state dependent transition rates. However the computational complexities increase

quickly with the increase of the number of states. We are able to obtain closed-form solutions only in special cases. This is shown in the derivation of closed-form solutions for transition probabilities and survival function for irreversible 4-state Markov models (see Figure 2.4). For reversible 4-state Markov models (see Figure 2.5), we developed numerical methods to obtain the solutions. The numerical methods can be used for general Markov models.

$$\mathbf{Q} = \begin{pmatrix} q_{00}(t) & q_{01}(t) & q_{02}(t) & q_{03}(t) \\ 0 & q_{11}(t) & q_{12}(t) & q_{13}(t) \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

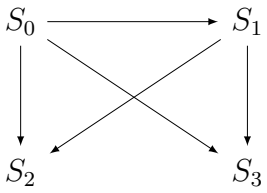


Figure 2.4: Irreversible model.

$$\mathbf{Q} = \begin{pmatrix} q_{00}(t) & q_{01}(t) & q_{02}(t) & q_{03}(t) \\ q_{10}(t) & q_{11}(t) & q_{12}(t) & q_{13}(t) \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

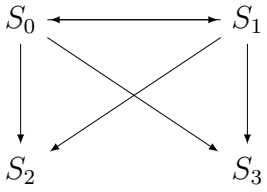


Figure 2.5: Reversible model.

2.3 Absorbing Markov Multi-state Models

A multi-state model is a stochastic process, $\boldsymbol{\xi} = \{\xi(t) : t \geq 0\}$, that describes the transitions of a study subject from one state to another over a period of observation. Multi-state models are used extensively in science and engineering

especially in biomedical research such as observing the status of patients over time in a clinical trial. There are a multiple number of states that a patient may advance to before reaching the end point (the event of interest). When such change data are available in the trial, multi-state models take into account of these changes in the computation and estimation of a patient's survival function, cumulative hazard function and others. Finite state absorbing Markov processes play important roles in multi-state models. One of the advantages of absorbing Markov models is that the survival function of a subject under study is identifiable because the Markovian property presupposes that the competing risks at every transition are stochastically independent. Tsiatis (1975) [49] gave an example of nonindentifiable (net) survival functions in a multiple decrement model with dependent competing risks.

For short, we shall refer to absorbing Markov multi-state models as Markov models. In Section 2.5 some non-Markov models are presented. Many commonly used models for survival analysis are Markov models. These models differ by the number of transient states and absorbing states, the transition paths and the assumptions of the transition rates as deemed appropriate for applications. For example, the right-censoring model (see (2.3)) has one transient state (alive) and two absorbing states (death and loss to followup). Multiple decrement models (see (2.6)) (commonly called competing risks models) are perhaps the earliest extension of right-censoring

models which have one transient state (alive) and a finite number of absorbing states for different causes of death. The simple illness and death model (e.g., Andersen and Keiding (2002) [9], PROVA Study Group (1991) [29]) (see (2.7)) has two transient states (healthy, illness) and one absorbing state (death). Chiang's multiple staging model (1980)[15] (see (2.8)) has k transient states representing different stages of a progressive disease, and one absorbing state (death). A special case of the k-progressive model (see (2.9)), the F-N model (see (2.1)), has two transient states (recovery, relapse) and two absorbing states (death, loss to followup).

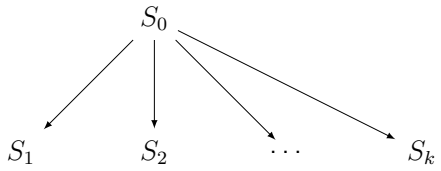


Figure 2.6: Transitions in the multiple decrement model.

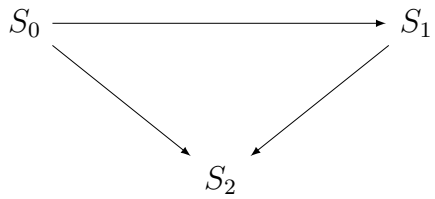


Figure 2.7: Transitions in the simple illness-death model.

Suppose a Markov model $\{\xi(t) : t \geq 0\}$ is a $(k+1)$-state irreversible model. Let q_{ij} denote the transition rate of an individual moving from state i to state j for

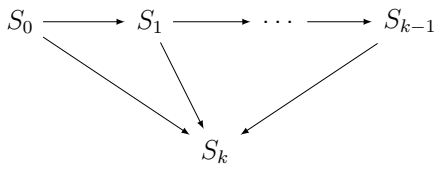


Figure 2.8: Transitions in Chiang's multi-staging model.

Figure 2.9: Transitions in the k-progressive model.

all states i, j. Then its infinitesimal matrix is given by

$$\mathbf{Q} = \begin{bmatrix} q_{00} & q_{01} & \cdots & \cdots & \cdots & q_{0k} \\ 0 & q_{11} & q_{12} & \cdots & \cdots & q_{1k} \\ 0 & 0 & q_{22} & \cdots & \cdots & q_{2k} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \tag{2.8}$$

One sees that the q_{ij} below the diagonal are zero and the transition paths are in one direction and irreversible (e.g Figure 2.4). Then the transition probabilities can be solved recursively by using the Chapman-Kolmogorov equations. The survival function can then be solved explicitly:

$$S(t) = P[X > t] = P[\xi(t) \neq k | \xi(0) = 0]. \tag{2.9}$$

Calculation of the transition probabilities from the given transition rates are of independent interest in survival analysis.

Chiang (1964, 1968) [13][14] extended the F-N model of two transient and two absorbing states by including more transient and absorbing states in the model. Results obtained by Chiang include explicit solutions for the transition probabilities and multiple transition probabilities. In the special case of multiple decrement model, explicit formulas for the crude and net survival probabilities are available (Chiang, Chapter 11, 1968 [14]) where the transition rates are assumed to be a product $c_{ij}\lambda(t)$ of positive constants c_{ij} and a hazard rate function, $\lambda(t)$ where $\lambda(t)$ is a function of the time t but independent of states for all i, j. Note that this model includes the Cox regression model as a special case.

Chiang's multi-staging model was generalized by J.Q. Fang in his unpublished Ph.D thesis (1985) [21]. The generalization allows any finite number of transient states and absorbing states. Fang's model deals with time-dependent covariates under the assumption of proportional hazards. The baseline hazard is assumed to have log-linear form. See Section 2.4.3.

In the models discussed so far, the interest is in the survival time distribution of a single individual. Due to possible dependence of the survival times of the study subjects, there are many practical problems that require modeling of bivariate survival times (X_1, X_2) of two subjects, X_1 and X_2 (Freund 1961 [24], Marshall and Olkin 1967 [35]), such as the survival times of twins, or the times to loss of hearing

of each of the two ears, or the failure times of two related components of a system in engineering reliability. The structure of the model is shown in Figure 2.10. Extensions to the two classical models includes Kvam and Samaniego (1997) [33] which considers a generalization to k dimensions.

$$\mathbf{Q} = \begin{pmatrix} q_{00} & q_{01} & q_{02} & 0 \\ 0 & q_{11} & 0 & q_{13} \\ 0 & 0 & q_{22} & q_{23} \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

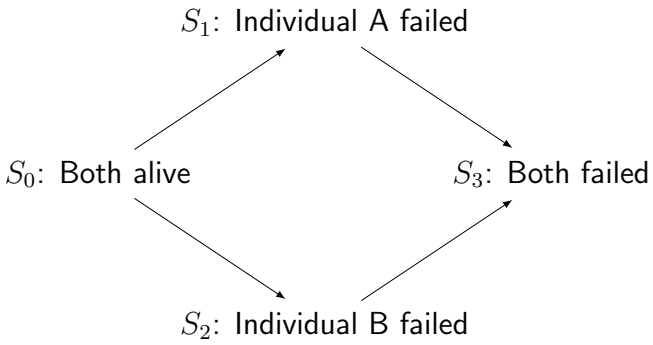


Figure 2.10: Transitions in the bivariate model.

2.4 Approaches to Estimation in Markov Multi-state Models

As noted before, in the context of multi-state models, our Markov models always refer to absorbing Markov processes. Also hazard rate functions or transition rate functions are referred to as hazard functions or transition rates respectively.

Typically Markov models are defined by specifying the transition rates $q_{ij}(t)$, $i, j \in S$. The transition rates are unknown and to be estimated from the data. The literature on the estimation is huge and there are plenty of books on the subject, e.g. Andersen (1997) [6], Andersen *et al.*(1993) [7], Kalbfleisch and Prentice (2002) [31]

and Collett (2003) [16], to name a few. Our review is necessarily narrowly focused on those more closely related to this thesis. We look at some of the very important results in each of the three categories: parametric, nonparametric and regression estimation.

2.4.1 Parametric Estimation

Estimation of constant transition rates in the homogeneous Markov models are fairly well-developed. Maximum likelihood estimates (MLE) for the staging model with constant transition rates are given in Chiang (1980) [15]. MLE under the F-N model with covariates are obtained by Beck (1979) [11] and the model includes the illness and death model as a special case. See also Beck and Chiang (1981) [12].

In many applications, modeling time-dependent transition rates is needed. Effort aiming at developing time-dependent Markov model includes a procedure to partition the observation period $[0, T]$ into small time intervals $[a_{l-1}, a_l]$, $i = 1, 2, \ldots, n$, $0 = a_0 < a_1 < \ldots < a_n = T$. On each time interval the transition rate $q_{ij}(t)$ is assumed to be constant. Then $q_{ij}(t)$ is a piecewise function defined by

$$q_{ij}(t) = q_{ij}^l, \quad a_{l-1} < t \leq a_l, \qquad\qquad l = 1, 2, \ldots, n$$

where the a_l's are known constants. See examples in Andersen *et al.*(1993) [7] and

Pérez-Ocón *et al.*(2001) [42]. The assumption of constant transition rates can be checked by hypothesis testing. A likelihood ratio test can be used to check the fit of the piecewise constant model to the constant transition rates model.

A better model could be obtained by increasing the number of intervals at the cost of increasing number of unknown parameters (transition rates), which not only increases the computation complexity but also requires larger sample sizes. Instead of piece-wise constant functions, for a special Markov model, explicit solutions for age-specific prevalence probabilities for time (age) and state dependent transition rates were derived (Yang and Chang (1990) [51]). Depending on the choice of transition rate functions, these prevalence probabilities can be non-monotonic, a requirement for the model to fit the seroepidemiology surveys that exhibit declining prevalence in older age. In Efromovich and Chu (2018) [20], transition rates are approximated by Fourier series. The coefficients of the Fourier series are estimated by moment estimators.

Generally, the requirement on the data is less stringent for parametric inferences than that for nonparametric inferences. For example, one can carry out estimation of the parameter λ in the exponential distribution with truncated data observed over the time interval $[0, T]$ for a finite T. Once an estimator $\hat{\lambda}$ is obtained, an estimated survival function is $e^{-\hat{\lambda}t}$ for all $t \geq 0$. But one cannot obtain a non-parametric

estimator of the survival function larger than the value of T with the truncated data.

Since the publication of Aalen (1975,1978) [1][2], counting processes formulation of the data and models has become one of the fundamental methods for lifetime analysis. We can convert our observed Markov multi-state processes into counting processes. Suppose there are n individuals in a study. Then the Markov multi-state process $\boldsymbol{\xi}_k$ with state space S for the k^{th} individual is observed over the time interval $[0, \tau_k]$, $k = 1, 2, \ldots, n$. This is equivalent to recording $\xi_k(0)$ and the following counting processes:

$$N_{ij}^k(t) = \text{number of direct transitions from i} \to j \text{ in [0,t]}, \quad i, j \in S, \quad t > 0$$

described by the times of these transitions

$$0 < Y_{ij}^{k1} < \ldots < Y_{ij}^{kN_{ij}(\tau_k)},$$

where Y_{ij}^{km} is the length of time between the m^{th} direct transition from state i and j for the k^{th} individual, $m = 1, 2, \ldots, N_{ij}(\tau_k)$, $k = 1, 2, \ldots, n$. Let $N_{ij}(t) =$

$\sum_{k=1}^{n} N_{ij}^k(t)$, $Z_i^k(t) = I\{\xi_k(t-) = i\}$, $Z_i(t) = \sum_{k=1}^{n} Z_i^k(t)$, then the likelihood is

$$\prod_{k=1}^{n} \prod_{i \neq j} \prod_{l=1}^{N_{ij}^k(\tau_k)} q_{ij}(Y_{ij}^{il}) \exp\Big(- \int_0^{\tau_k} q_{ij}(t) Z_i^k(t) dt \Big).$$

This is the likelihood for the general case. It can be simplified if constant or piecewise constant transition rates are assumed. Parametric inference is readily available in statistics books. In situations where the maximum likelihood estimator is difficult or even impossible to compute, M-estimators can be used. See Andersen *et al.*(1993) [7] for examples.

2.4.2 Nonparametric Estimation

Aalen (1975, 1978) [1][2] introduced a nonparametric estimator $\hat{\Lambda}_{ij}(t)$ of the cumulative hazard function $\Lambda_{ij}(t)$ for right censored data as a stochastic integral with respect to a counting process. Such a formulation permits the use of martingale calculus to obtain statistical properties of the estimator. Suppose there are n individuals under observation. Let $N_{ij}(t)$ be the number of individuals moving directly to state j from state i in $[0, t]$, let Y_{ij}^k be the observed sojourn time in state i before moving to state j for the k^{th} individual, and $H_i(t) = \sum_j \sum_{k=1}^{n} I[Y_{ij}^k \geq t]$,

$J_i(t) = I[H_i(t) > 0]$. Then the estimator $\hat{\Lambda}_{ij}(t)$ is given by

$$\hat{\Lambda}_{ij}(t) = \int_0^t \frac{J_i(u)}{H_i(u)} dN_{ij}(u). \tag{2.10}$$

Aalen and Johansen (1978) [5] proposed a nonparametric estimator in terms of counting processes for the transition probability matrix in a finite nonhomogeneous Markov model:

$$\hat{P}(s,t) = \prod_{u \in (s,t]} (I + \hat{\Lambda}(du))$$

where $\hat{\Lambda}(t) = (\hat{\Lambda}_{ij}(t))$ is the matrix of estimated transition rates and $P(s,t)$ is the transition probability matrix. The basic tool in their derivation is the matrix product integral. The estimator can be thought of as the generalization of the Kaplan-Meier estimator for general Markov multi-state models with a finite number of states. The exact and asymptotic properties of these estimators are studied based on stochastic integrals and martingales.

Frydman (1992) [25] proposed a nonparametric maximum likelihood procedure for the estimation of the cumulative hazard rates in the irreversible illness-death model. A nonparametric estimator for interval-censored data in the ilness-death model was introduced in Frydman (1995) [26]. Both of these results were generalized to incorporate observations with unknown intermediate event status (Frydman and

Szarek (2009) [27]).

2.4.3 Regression models

In all of the aforementioned models, each individual under study is described using a stochastic process $\boldsymbol{\xi}_k$, $k = 1, 2, \ldots, n$ with state space $S = \{1, 2, \ldots, m\}$. The n stochastic processes are assumed to be independent and have identical distribution. In practical situations, however, the personal characteristics vary from one individual to another and may contain valuable information which might impact the survival time distribution. One of the most important extensions of the Kaplan-Meier estimator was given by Cox (1972) [18], who introduced the covariate vector (Z) associated with an individual under study in the form of a regression component of the hazard function of that individual. As such the sample consists of n independent but not identically distributed possibly right-censored survival times. The Cox model facilitates multiple sample comparisons. Specifically, the Cox regression model for the mortality model (right-censoring model) has the following hazard rate function for an individual:

$$\lambda(t) = \lambda_0(t) \exp(-\beta' Z), \tag{2.11}$$

where $\lambda_0(t)$ is a baseline hazard rate function common to all individuals under study (baseline hazard rate function) and the hazard rate function of an individual is proportional to $\lambda_0(t)$ by a factor $\exp(-\beta' Z)$, where β is an unknown vector regression coefficient.

The Cox regression model has been extensively investigated and there are many generalizations. The proportionality assumption may fail sometimes. For example, consider a clinical trial in which patients are randomized into either a treatment group or control group. The event under study is the time of death of each patient. Within the framework of Markov multi-state models, a generalized Cox regression model can be expressed in terms of the following hazard function:

$$q_{ij}(t, Z) = \phi(q_{ij,0}(t), \beta_{ij}^T Z), \quad t > 0$$

where $q_{ij}(t)$ is the hazard function of moving directly from state i to state j at time t for $i \neq j$, ϕ is the link function, $q_{ij,0}(t)$ is the baseline hazard function governing the transition from state i and j, β_{ij} is the vector of regression coefficients, and Z is the covariate vector representing the characteristics of an individual in the study. The vector Z can be either time-independent or time-dependent. In the Cox model which deals with one hazard rate function, $\phi(q_{ij,0}(t), \beta_{ij}^T Z) = \lambda_0(t) \exp(\beta_{ij}^T Z)$. The estimation of the baseline hazard function $\lambda_0(t)$ and the regression parameters β_{ij}

are commonly treated separately.

In a different approach, J.Q. Fang (1985) [21] introduced a fully parametric model with time-dependent covariates and the baseline hazard function $\lambda_0(t)$ of a log-linear form given by

$$\lambda_0(t) = \sum_{s=1}^{r} \gamma_s Y_s(t), \tag{2.12}$$

where the $Y_s(t)$ are specified functions of t, and the γ_s are coefficients. For example, if $r = 1$, $Y_1(t) = 1$, then $\lambda_0(t) = \gamma_1$ which gives an exponential distribution. If $r = 2$, $Y_1(t) = 1$, $Y_2(t) = log(t)$, we have a Weibull distribution. One can use a r^{th} degree polynomial by taking $Y_s(t) = t^{s-1}$. Under Fang's assumptions (Fang 1985 [21], p.20), the Cox regression model can be written as

$$\lambda(t) = \exp(\beta' X(t)) \tag{2.13}$$

where $X(t)$ is a column vector, $X(t) = (x_1(t), \cdots, x_p(t))$ in which $x_1(t) = 1$, $x_2(t)$, $\cdots$, $x_{p_1+1}(t)$ are p_1 time-dependent covariates, $x_{p_1+2}(t) = t$, $\cdots$, $x_s(t) = t^{s-p_1-1}$, $\cdots$, $x_p(t) = t^{p-p_1-1}$ are quasi covariates (specified time-dependent functions) in $\lambda_0(t)$ and it is assumed that $p \geq 1 + p_1$. This is a fully parametric model and the value p can be estimated from the data set by using likehood ratio tests. Using Le Cam's theory [34], the log likehood ratios are proven to have asymptotic normal distributions

for contiguous alternatives. Furthermore, a general Markov multi-state model with transition rates $q_{ij}(t)$ having a parametric form similar to (2.13) is provided in Fang (1985) [21]. With appropriate choices of the functions $Y_s(t)$ for $\lambda_0(t)$, it becomes a semi-Markov model. The model with irreversible transition paths is applied to studying the University of California faculty promotion data.

Another approach that does not assume the proportional hazards is the additive hazards model (Aalen 1989 [3]) given by:

$$q_{ij}(t, Z) = q_{ij,0}(t) + \beta_{ij}^T(t)Z, \tag{2.14}$$

where the regression coefficients β_{ij} are allowed to depend on time.

Therneau and Grambsch (2013) [48] showed that if the form of the covariates are incorrectly specified, it will lead to a diagnosis of non-proportional hazards. To study the effects of covariates, a general proportional hazard model with arbitrary covariate effects is proposed:

$$q_{ij}(t, Z) = q_{ij,0}(t) \exp(f(Z)) \tag{2.15}$$

where $f(Z)$ is a smooth function. In model (2.15), the effect of the covariate vector Z is in the form of an arbitrary smooth function f.

More recent works include Huang and Liu (2006) [30] which studied the parametric estimator of the hazard functions in the following model:

$$q(t, Z) = q_0(t, Z) \exp(\psi(\beta^T Z))$$

where ψ is an unknown smooth link function which can model the possible nonlinearity of the effect of the covariates. The link function is approximated by a polynomial spline.

When the time-dependent covariates are categorical, Cortese and Andersen (2009) [17] suggest to incorporate the information carried by the covariate process into the multi-state model. Each possible value that the covariate can assume over time can be represented as a transient state.

2.5 Non-Markov Models

In numerous practical problems Markov models are not suitable. In particular the assumption of stochastic independence in competing risks is sometimes hard to defend. For example, in the multiple decrement model, elimination of the risk of one type of failure might increase the risk of another type of failure (Moeschberger (1974) [39], Prentice *et al.*(1978) [43]). Another example, in the illness-death model

described in Section 2.1, Figure 2.7, if the transition rate from diseased (state 1) to death (state 2) depends on the time of entry into the diseased state, then the process $\{\xi(t) : t \geq 0\}$ is non-Markovian.

Strauss and Shavelle (1998) [46] extended the K-M estimator without the Markovian assumption. Aalen *et al.*(2001) [4], Datta and Satten (2001) [19] studied the performance of the estimators of state occupation probabilities derived from the Aalen–Johansen estimators (2.10) under a non-Markovian model. More recently, using simulation, Glidden (2002) [28] obtained confidence bands for the Aalen-Johansen estimator for the transition probabilities and showed that under a non-Markovian condition, the Aalen-Johansen estimator may be biased. Also see Meira-Machado *et al.*(2006) [36]. Meira-Machado *et al.*(2006) demonstrated by simulation that a nonparametric estimator of transition probabilities which these authors introduced outperforms the Aalen–Johansen estimator in a non-Markov situation. Putter and Spitoni (2018) [45] proposes a relatively simple and intuitive procedure called landmark Aalen–Johansen (LMAJ) that will provide consistent estimators of transition probabilities for general multi-state models.

2.6 Recurrent Event Analysis

The Cox regression model has also been generalized for analyzing recurrent events. For example, a commonly used one is the Andersen-Gill (AG) model (1982) [8] which assumes that the recurrent events are independent. The intensity process is:

$$\lambda(t) = Y(t)\lambda_0(t)\exp(X(t)\beta),$$

where $Y(t)$ is the at-risk indicator ($Y(t) = 1$ if the individual is still under observation, $Y(t) = 0$ otherwise), $\lambda_0(t)$ is the baseline intensity function, $X(t)$ is the covariate process and β is the coefficient vector. The difference between the AG model and the Cox model lies in the indicator $Y(t)$. The AG model is also applicable to right censored data under the independent censoring assumption. In Miloslavsky *et al.*(2004) [38], an estimator that accounts for dependent censoring is introduced to improve the AG model.

The Prentice-Williams-Peterson (PWP) model [44] is similar to the AG model, but for each recurrent event a separate intensity function is modeled:

$$\lambda_j(t) = Y_j(t)\lambda_{0j}(t)\exp(X(t)\beta_j), \quad j = 1, 2, \ldots, k.$$

It can be seen as a stratified AG model. In the PWP model, the at-risk indicator for the j^{th} event $Y_j(t)$ is zero until the $(j-1)^{st}$ event and only then becomes one. When the j^{th} event occurs it becomes zero again.

More recently, multi-state models have been extended for recurrent events, e.g., Andersen and Keiding (2002) [9]. For example, Figure 2.11 shows a multi-state model for recurrent events. The intensities in the model can be estimated using approaches discussed in Section 2.4.

Figure 2.11: Multi-state model for recurrent events.

Chapter 3: Irreversible Markov Multi-state Model

3.1 Introduction

Our ultimate goal is to estimate the survival function. Section 3.2 gives a motivating example of the progression of patients with Aplastic Anemia (AA). Section 3.3 constructs a 4-state irreversible Markov model $\boldsymbol{\xi}^o$ using competing risks. Section 3.4 states the problems to investigate where the survival function is written in terms of a transition probability of $\boldsymbol{\xi}^o$. Section 3.5 gives nonparametric estimators of the cumulative hazard functions and an estimator of the net survival function is obtained in Section 3.5.3. Section 3.6 deals with the case when the observation period is a finite time interval. Section 3.7 shows the asymptotic properties of the estimators. Section 3.8 applies the model to a real data set collected in the clinical trial on patients with AA.

3.2 A Motivating Example

We use a Markov model to describe the progression of a patient with Aplastic Anemia (AA) with data from a clinical trial. Aplastic Anemia (AA) is a form of bone marrow failure where the bone marrow does not produce new blood cells, leaving the patient susceptible to bleeding and infection. Possible treatments include blood transfusion, bone marrow transplant and medical therapy. In the clinical trial, for patients with severe AA, treatment with immunosuppressants therapy (IST) or a bone marrow transplant is necessary. The IST treatment involves a drug that suppresses the activity of immune cells which damage the bone marrow. The treatment helps the bone marrow to recover and generate new blood cells. Patients younger than 40 years old with a blood-matched sibling who can donate bone marrow are usually treated by a bone marrow transplant, while patients over 40 or without a blood-matched sibling donor are usually treated by drug therapy.

Upon the completion of the first IST, a patient is considered to be in remission. A recovered patient may also relapse. The follow up of each patient begins immediately after the first IST until death or loss to followup. Dr. Wu of NIH, one of the authors of Sloand *et al.*(2008) [47] brought this research problem to our attention and we gratefully acknowledge the discussion with Dr. Wu. For each patient in the

clinical trial, the times at which the events of interest occurred were recorded. The events of interest include: the times of first IST, relapse, death and loss to followup. For example, after the first IST, if a patient relapses and then dies, we record his/her data as follows: date of IST, date of relapse, date of death; if a patient is lost to followup directly after the treatment, then the observations were: date of IST, data of last count. The data structure will be described in details in Section 3.4. In Section 3.3 through Section 3.5, we assume that the censoring time is a random variable with support $[0, \infty]$. However in reality, clinical trials usually terminate at a prespecified non-negative finite time, say T, which makes the support of the censoring time variable on a finite time interval $[0, T]$. This case will be discussed in Section 3.6. The comparison of the model with $T = \infty$ and the model with $T < \infty$ will also be made in Section 3.6 and Section 4.2.4.

3.3 Irreversible Markov Model

Applying the F-N procedure, we construct a four-state Markov process $\boldsymbol{\xi}^o = \{\xi^o(t) : t \geq 0\}$ for the AA example described in Section 3.2, where $\xi^o(t)$ represents the status of a patient at time t, and $t = 0$ corresponds to the time the patient immediately after the first IST. Throughout this chapter we assume in the model that a patient who has relapsed will not receive a second IST. So the Markov model

is irreversible. In the Chapter 5 we study a reversible Markov model in which a relapsed patient may receive a second IST. The four possible states are S_0, S_1, S_2 and S_3 where:

S_0: a patient is in remission immediately after first IST

S_1: a patient relapses after the first remission

S_2: a patient has died

S_3: a patient is lost to followup

In terms of Markov processes, S_2 and S_3 are absorbing states, S_0 and S_1 are transient states. Possible transitions in the process $\{\xi^o(t), t \geq 0\}$ are described schematically in Figure 3.1. Note that this is the same model as illustrated in Figure 3.1.

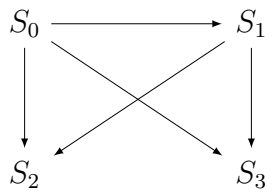


Figure 3.1: Possible transition paths of a patient during the clinical trial.

A patient will take one and only one of the following four mutually exclusive transition paths during the clinical trial:

(1) $S_0 \to S_1 \to S_2$

(2) $S_0 \to S_1 \to S_3$

(3) $S_0 \to S_2$

(4) $S_0 \to S_3$

Patients on transition path (2) and (4) are lost to followup and thus the corresponding data is right-censored. The exact time at which a transition occurs and the state where a transition takes place are recorded. For notational convenience and consistency with symbols used in Markov model, we shall denote S_i by i, for $i = 0, 1, 2, 3$.

The transition probabilities for the Markov process are defined as

$$P_{ij}^o(s,t) = P(\xi^o(t) = j \mid \xi^0(s) = i), 0 \leq s < t, \quad i, j = 0, 1, 2, 3, \tag{3.1}$$

where $P_{ij}^o(s,t)$ is the probability that a patient will be in state j at time t given that the patient is in state i at time s. Throughout this chapter, the initial state $\xi^o(0)$ is assumed to be 0. It is assumed that the transition rates $q_{ij}(t)$ are well-defined as

$$q_{ij}(t) = \lim_{h \to 0} \frac{P(\xi^o(t+h) = j \mid \xi^o(t) = i)}{h}, \quad t > 0, \quad i \neq j, \quad i, j = 0, 1, 2, 3, \tag{3.2}$$

$$q_{ii}(t) = -\sum_{j \neq i} q_{ij}(t), \quad t > 0, \quad i = 0, 1, 2, 3. \tag{3.3}$$

We assume that all $q_{ij}(t)'s$ are continuous in t and

$$\int_0^\infty q_{ij}(u)du = \infty, \quad i \neq j, \quad i, j = 0, 1, 2, 3. \tag{3.4}$$

Constructing the Markov model from the sample paths as follows. This will be used in simulating the Markov processes in Chapter 4.

After the initial IST, over time a patient may either die, relapse or be lost to followup. For $j = 1, 2, 3$, let Y_{0j} be the sojourn time of a patient who starts from state 0 and stays in state 0 for the duration of Y_{0j} before moving directly to state j. In terms of the Markov model, Y_{01}, Y_{02} and Y_{03} are nonnegative independent random variables that represent time to relapse, time to death and time to loss to followup (from the IST). For $j = 1, 2, 3$, let $F_{0j}(t)$, $S_{0j}(t)$ and $q_{0j}(t)$ be the cumulative distribution function (CDF), the survival function and the hazard function of Y_{0j}, respectively. The hazard functions $q_{0j}(t)$ are given in (3.2). Death, relapse and loss to followup are considered as competing risks with hazard rate $q_{0j}(t)$, $j = 1, 2, 3$. Under the competing risks model, when a transition (an event) is observed, it means that the time of the event occurrence is the minimum of Y_{01}, Y_{02} and Y_{03}. The observable

quantities are a pair of random variables (Y_0, δ_0) defined by

$$\begin{aligned} Y_0 &= \min(Y_{01}, Y_{02}, Y_{03}), \\ \{\delta_0 = j\} &= \{Y_{0j} = \min(Y_{01}, Y_{02}, Y_{03})\}, \quad j = 1, 2, 3. \end{aligned} \tag{3.5}$$

The random variable Y_0 is the length of time that the process $\boldsymbol{\xi}^o$ stays in state 0 before leaving 0 and $\delta_0 = j$ indicates $\boldsymbol{\xi}^o$ moves directly from state 0 to state j for $j = 1, 2, 3$. Then the survival function of Y_0 is

$$\begin{aligned} S_0(t) &= P(Y_0 > t) \\ &= P(\min(Y_{01}, Y_{02}, Y_{03}) > t) \\ &= S_{01}(t)S_{02}(t)S_{03}(t), \quad t \geq 0 \end{aligned} \tag{3.6}$$

and

$$\begin{aligned} P(\delta_0 = j) &= P(Y_{0j} = \min(Y_{01}, Y_{02}, Y_{03})) \\ &= P(Y_0 = Y_{0j}) \\ &= \int_0^\infty S_0(u) q_{0j}(u) du, \quad j = 1, 2, 3. \end{aligned} \tag{3.7}$$

The joint distribution of Y_0 and δ_0 is:

$$\begin{aligned} G_0(t,j) &= P(Y_0 \le t, \delta_0 = j) \\ &= \int_0^t S_0(u) q_{0j}(u) du, \quad t \ge 0, \quad j = 1,2,3. \end{aligned} \tag{3.8}$$

For convenience in simulation, we introduce random variables Y_{0j}^c, $j = 1,2,3$. The distribution function of Y_{0j}^c is:

$$\begin{aligned} F_{0j}^c(t) &= P(Y_{0j}^c \le t) \\ &= P(Y_0 \le t \mid \delta_0 = j) \\ &= \frac{G_0(t,j)}{P(\delta_0 = j)} \\ &= \int_0^t S_0(u) q_{0j}(u) du \Big/ \int_0^\infty S_0(u) q_{0j}(u) du, \quad t \ge 0, \quad j = 1,2,3. \end{aligned} \tag{3.9}$$

Equation (3.9) shows that the cdf of Y_{0j}^c is the conditional cdf of Y_0 given $\delta_0 = j$. The differences between the random variables Y_{0j}^c and Y_{0j} are as follows. A patient in state 0 will move directly to either state 1 or 2 or 3, and his/her sojourn time Y_0 in state 0 is $\min(Y_{01}, Y_{02}, Y_{03})$. The observations are Y_0 and δ_0. However, for the purpose of simulation, we find it convenient to use δ_0 and the conditional distribution of Y_0 given δ_0. Recall that Y_{0j}^c is the observed sojourn time in state 0 given that the

patient will move to state j. The CDF of Y_0 and Y_{0j}^c's are related by the following formula:

$$\begin{aligned} F_0(t) &= P(Y_0 \leq t) \\ &= \sum_{j=1}^{3} P(Y_0 \leq t | \delta_0 = j) P(\delta_0 = j) \\ &= \sum_{j=1}^{3} P(Y_{0j}^c \leq t) P(\delta_0 = j) \\ &= \sum_{j=1}^{3} F_{0j}^c(t) P(\delta_0 = j), \quad t \geq 0, \quad j = 1, 2, 3. \end{aligned} \tag{3.10}$$

Once a patient enters state 1, the possible next state to enter is either state 2 or state 3. By the same token, we define the following random variables for $j = 2, 3$:

Y_{1j}: The sojourn time in state 1 before moving directly to state j,

Y_1: The sojourn time in state 1 before moving to the next state,

δ_1: The indicator random variable, $\delta_1 = j$ indicates that the patient enters state j.

Y_{1j}^c: The sojourn time in state 1 conditioned on the patient will enter state j directly from state 1.

Let $F_{1j}(t)$, $S_{1j}(t)$ and $q_{1j}(t)$ be the cdf, survival function and the hazard rate

function of Y_{1j}, respectively, where $q_{1j}(t)$ is defined in (3.2), $j = 2, 3$. We then have

$$\begin{aligned} S_1(t) &= P(Y_1 > t) \\ &= S_{12}(t)S_{13}(t), \quad t \geq 0, \end{aligned} \tag{3.11}$$

$$\begin{aligned} P(\delta_1 = j) &= P(Y_{1j} = \min_{m=2,3} Y_{1m}) \\ &= \int_0^\infty S_1(u)q_{1j}(u)du, \end{aligned} \tag{3.12}$$

the cdf of Y_{1j}^c is

$$\begin{aligned} F_{1j}^c(t) &= P(Y_{1j}^c \leq t) \\ &= \int_0^t S_1(u)q_{1j}(u)du \Big/ \int_0^\infty S_1(u)q_{1j}(u)du, \quad t \geq 0, \quad j = 2, 3, \end{aligned} \tag{3.13}$$

and the joint distribution of (Y_1, δ_1) is:

$$\begin{aligned} G_1(t, j) &= P(Y_1 \leq t, \delta_1 = j) \\ &= \int_0^t S_1(u)q_{1j}(u)du, \quad t \geq 0, \quad j = 2, 3. \end{aligned} \tag{3.14}$$

3.4 Problems to Investigate and Data Structure

Problems: We are interested in estimating the survival function of an AA patient in the clinical trial described in Section 3.2. In our model (Figure 3.1), the survival function computed from the process $\{\xi^o(t) : t \geq 0\}$ is:

$$\begin{aligned} S^o(t) &= P(X^o > t \mid \xi^o(0) = 0) \\ &= 1 - P(\xi^o(t) = 2 | \xi^o(0) = 0) \\ &= 1 - P^o_{02}(0, t), \end{aligned}$$

where X^o denotes the survival time of a patient in the $\boldsymbol{\xi}^o$ model. Recall that the Fix and Neyman procedure of identifying the (net) survival function S involves two Markov processes, $\boldsymbol{\xi}^o$ and $\boldsymbol{\xi}$ (Section 2.2), where $\boldsymbol{\xi} = \{\xi(t) : t \geq 0\}$ is a 3-state Markov process ($\{S_0, S_1, S_2\}$) with the same transition rates q_{01}, q_{02}, q_{12} as in $\boldsymbol{\xi}^o$. To compute S requires the computation of $P_{02}(0, t)$, from which the net survival function of a patient is:

$$\begin{aligned} S(t) &= P(X > t \mid \xi(0) = 0) \\ &= 1 - P_{02}(0, t). \end{aligned}$$

For any Markov model considered in this thesis, given the infinitesimal matrix Q, we use the following Kolmogorov forward equations to solve for its transition probabilities $P_{ij}(0,t)$:

$$\frac{dP_{ij}(s,t)}{dt} = \sum_{l \neq j} P_{il}(s,t) q_{lj}(t) + P_{ij}(s,t) q_{jj}(t), \quad 0 \leq s < t, \quad i,j = 0,1,2, \tag{3.15}$$

with initial conditions

$$\begin{cases} P_{ii}(0,0) = 1, \\ \lim\limits_{s \to t} P_{ij}(s,t) = 1, & \text{if } i = j, \\ \lim\limits_{s \to t} P_{ij}(s,t) = 0, & \text{if } i \neq j. \end{cases} \tag{3.16}$$

If all of the transition rates are constant in t, the solutions for the transition probabilities are known. For time dependent $q_{ij}(t)$, Feller (1940) [22] proved the existence and uniqueness of the solutions. However, except in some special cases such as an irreversible Markov process, there are no closed-form solutions. We rely on numerical methods to obtain the solutions, in particular for the survival function.

Data: Following the AA example (Section 3.2), suppose there are n patients in the clinical trial. The changes over time of the k^{th} patient, for $k = 1, \ldots, n$ are assumed to be n *i.i.d.* Markov processes $\boldsymbol{\xi}_k^o = \{\xi_k^o(t), t \geq 0\}$. For any patient, we observe his/her sojourn time in state 0 or 1 before moving directly to the next state. Our observed data are as follows:

(1) $\{(Y_{0,k}, \delta_{0,k}) \mid k \in 1, \ldots, n, \delta_{0,k} = 1\}$,

(2) $\{(Y_{0,k}, \delta_{0,k}) \mid k \in 1, \ldots, n, \delta_{0,k} = 2\}$,

(3) $\{(Y_{0,k}, \delta_{0,k}) \mid k \in 1, \ldots, n, \delta_{0,k} = 3\}$,

(4) $\{(Y_{1,k}, \delta_{1,k}) \mid k \in 1, \ldots, N_{01}, \delta_{1,k} = 2\}$,

(5) $\{(Y_{1,k}, \delta_{1,k}) \mid k \in 1, \ldots, N_{01}, \delta_{1,k} = 3\}$.

Note that, if for $j = 1, 2, 3$ we let $N_{0j}(t)$ be the number of patients who started from state 0 and moved directly to state j in $[0, t]$, $t > 0$, for $j = 2, 3$ let $N_{1j}(t)$ be the number of patients who landed in state 1 and moved to state j in $[0, t]$, $t > 0$. Then for any $t > 0$ we should have $N_{01}(t)+N_{02}(t)+N_{03}(t) = n$ and $N_{12}(t)+N_{13}(t) = N_{01}(t)$. Let $N_0(t) = (N_{01}(t), N_{02}(t), N_{03}(t))$. Then given $t = y_0$, $N_0(y_0)$ follows a multinomial distribution with parameters n and $p_0(y_0) = (p_{01}(y_0), p_{02}(y_0), p_{03}(y_0))$ where

$$
\begin{aligned}
p_{0j}(y_0) &= \lim_{dt \to 0} P(Y_{0j} = \min(Y_{01}, Y_{02}, Y_{03}) \mid Y_0 \in [y_0, y_0 + dt)) \\
&= \lim_{dt \to 0} \frac{P(Y_{0j} = \min(Y_{01}, Y_{02}, Y_{03}), Y_0 \in [y_0, y_0 + dt))}{P(Y_0 \in [y_0, y_0 + dt))} \\
&= \lim_{dt \to 0} \frac{\Big(\int_{y_0}^{y_0 + dt} S_0(t) q_{0j}(t) dt \Big) / dt}{\Big(F_0(y_0 + dt) - F_0(y_0) \Big) / dt} \\
&= -\frac{q_{0j}(y_0)}{q_{00}(y_0)}, \quad \text{subject to} \sum_{j=1}^{3} p_{0j}(y_0) = 1,
\end{aligned}
$$

$$q_{00}(y_0) = -\sum_{j=1}^{3} q_{0j}(y_0). \tag{3.17}$$

The probabilities p_{0j} are functions of time, evaluated at y_0, which give the probability of event j happening at time y_0, $j = 1, 2, 3$. Similarly given $t = y_1$, $N_1(y_1) = (N_{12}(y_1), N_{13}(y_1))$ follows a binomial distribution with parameters $N_{01}(y_0)$ and $p_1(y_1) = (p_{12}(y_1), p_{13}(y_1))$ where

$$p_{1j}(y_1) = -\frac{q_{1j}(y_1)}{q_{11}(y_1)}, \quad j = 2, 3, \tag{3.18}$$

$$q_{11}(y_1) = -\sum_{j=2}^{3} q_{1j}(y_1). \tag{3.19}$$

3.5 Nonparametric Estimators

Following Section 3.3, under model $\{\xi^o(t) : t \geq 0\}$, we construct nonparametric estimators of the cumulative hazard functions, transition probabilities and the survival function respectively.

3.5.1 Estimators of Cumulative Hazard Functions

The denominator in (3.9) is the probability

$$\begin{aligned}\alpha_{0j} &= P(\delta_0 = j) \\ &= \int_0^\infty S_0(u) q_{0j}(u) du, \quad j = 1, 2, 3. \end{aligned} \tag{3.20}$$

The probability α_{0j} is a constant independent of time t. Then (3.9) becomes

$$F_{0j}^c(t) = \frac{1}{\alpha_{0j}} \int_0^t S_0(u) q_{0j}(u)\ du. \tag{3.21}$$

Differentiating with respect to t on both side of (3.21) yields

$$dF_{0j}^c(t) = \frac{1}{\alpha_{0j}} S_0(t) q_{0j}(t) dt. \tag{3.22}$$

It follows that a formula for the hazard rate $q_{0j}(t)$ is:

$$\begin{aligned} q_{0j}(t) &= \alpha_{0j} \frac{dF_{0j}^0(t)}{S_0(t)dt} \\ &= \alpha_{0j} \frac{dG_0(t,j)}{\alpha_{0j} S_0(t) dt} \\ &= \frac{dG_0(t,j)}{S_0(t)dt}, \quad t \geq 0, \quad j = 1, 2, 3. \end{aligned} \tag{3.23}$$

Note that in (3.23), the unknown constant α_{0j} in the numerator and denominator is cancelled. For $j = 1, 2, 3$, a nonparametric estimation of $q_{0j}(t)$ can be performed by the estimation of $G_0(t, j)$ and $S_0(t)$ with the data $\{Y_{0,k}, \delta_{0,k}, k = 1, 2, \ldots, n\}$ and $\{Y_{1,k}, \delta_{1,k}, k = 1, 2, \ldots, N_{01}\}$ (see Section 3.4).

Let $t_{0j} = \max_{k=1,2,\ldots,n} \{Y_{0,k} \mid \delta_{0,j} = j\}$ be the largest uncensored observation of Y_{0j}, $j = 1, 2, 3$. Let $t_{1j} = \max_{k=1,2,\ldots,n} \{Y_{1,k} \mid \delta_{1,j} = j\}$ be the largest uncensored observation for Y_{1j}, $j = 2, 3$. Let $t_0 = \max(t_{01}, t_{02}, t_{03})$, $t_1 = \max(t_{12}, t_{13})$. Then the nonparametric maximum likelihood estimators of $G_0(t, j)$ and $S_0(t)$ are:

$$\hat{G}_0(t, j) = \frac{1}{n} \sum_{k=1}^{n} I\{Y_{0,k} \leq t, \delta_{0,k} = j\}, \quad t \in [0, t_{0j}], \tag{3.24}$$

$$\hat{S}_0(t) = \frac{1}{n} \sum_{k=1}^{n} I\{Y_{0,k} > t\}, \qquad t \in [0, t_0]. \tag{3.25}$$

Plug (3.24), (3.25) into (3.23) we obtain:

$$\hat{q}_{0j}(t) = \frac{d\hat{G}_0(t, j)}{\hat{S}_0(t)dt}, \quad t \in [0, t_{0j}], \quad j = 1, 2, 3. \tag{3.26}$$

In (3.26), $\hat{G}_0(t, j)$ is a step function. Therefore

$$d\hat{G}_0(t, j) = \hat{G}_0(t, j) - \hat{G}_0(t-, j). \tag{3.27}$$

Integrating both sides of (3.26), we obtain an estimator of the cumulative hazard function $\Lambda_{0j}(0,t) = \int_0^t q_{0j}(u)du$:

$$\begin{aligned}\hat{\Lambda}_{0j}(0,t) &= \int_0^t \frac{1}{\hat{S}_0(u)} d\hat{G}_0(t,j) \\ &= \sum_{k=1}^{n} \frac{I(Y_{0,k} \le t, \delta_{0,k} = j)}{\sum_{m=1}^{n} I(Y_{0,m} > Y_{0,k})}, \quad t \in [0, t_{0j}], \quad j = 1,2,3. \end{aligned} \tag{3.28}$$

Estimators of $q_{1j}(t)$ and $\Lambda_{1j}(0,t)$ can be derived similarly:

$$\hat{q}_{1j}(t) = \frac{d\hat{G}_1(t,j)}{\hat{S}_1(t)dt}, \quad t \in [0, t_{1j}], \quad j = 2,3, \tag{3.29}$$

$$\begin{aligned}\hat{\Lambda}_{1j}(0,t) &= \int_0^t \frac{1}{\hat{S}_1(u)} d\hat{G}_1(u,j) \\ &= \sum_{k=1}^{N_{01}} \frac{I(Y_{1,k} \le t, \delta_{1,k} = j)}{\sum_{m=1}^{N_{01}} I(Y_{1,m} > Y_{1,k})}, \quad t \in [0, t_{1j}], \quad j = 2,3. \end{aligned} \tag{3.30}$$

3.5.2 Estimators of Transition Probabilities

Corresponding to the transition paths depicted in Figure 3.1, the infinitesimal matrix of our Markov model is $\boldsymbol{Q}^o(t)$, $t > 0$:

$$\mathbf{Q}^o(t) = \begin{pmatrix} q_{00}(t) & q_{01}(t) & q_{02}(t) & q_{03}(t) \\ 0 & q_{11}(t) & q_{12}(t) & q_{13}(t) \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \tag{3.31}$$

Given $\boldsymbol{Q}^o(t)$, only seven transition probabilities are nonzero. Applying the Kolmogorov forward equations (3.15) to all nonzero $P^o_{ij}(0,t), i,j = 0,1,2,3$, we have the following differential equations:

$$\begin{aligned}
\frac{dP^o_{00}(0,t)}{dt} &= P^o_{00}(0,t)q_{00}(t),\\
\frac{dP^o_{01}(0,t)}{dt} &= P^o_{00}(0,t)q_{01}(t) + P^o_{01}(0,t)q_{11}(t),\\
\frac{dP^o_{02}(0,t)}{dt} &= P^o_{00}(0,t)q_{02}(t) + P^o_{01}(0,t)q_{12}(t),\\
\frac{dP^o_{03}(0,t)}{dt} &= P^o_{00}(0,t)q_{03}(t) + P^o_{01}(0,t)q_{13}(t),\\
\frac{dP^o_{11}(0,t)}{dt} &= P^o_{11}(0,t)q_{11}(t),\\
\frac{dP^o_{12}(0,t)}{dt} &= P^o_{11}(0,t)q_{12}(t),\\
\frac{dP^o_{13}(0,t)}{dt} &= P^o_{11}(0,t)q_{13}(t),
\end{aligned}$$

with initial conditions (3.16). The solutions are:

$$P^o_{00}(0,t) = \exp\Big(\int_0^t q_{00}(u)\,du\Big), \tag{3.32}$$

$$P^o_{01}(0,t) = \exp\Big(\int_0^t q_{11}(u)\,du\Big)\int_0^t \exp\Big(\int_0^u -q_{11}(v)\,dv\Big)P^o_{00}(u)q_{01}(u)\,du, \tag{3.33}$$

$$P^o_{02}(0,t) = \int_0^t \Big[P^o_{00}(0,u)q_{02}(u) + P^o_{01}(0,u)q_{12}(u)\Big]\,du, \tag{3.34}$$

$$P^o_{03}(0,t) = \int_0^t \Big[P^o_{00}(0,u)q_{03}(u) + P^o_{01}(0,u)q_{13}(u)\Big]\, du, \tag{3.35}$$

$$P^o_{11}(0,t) = \exp\Big(\int_0^t q_{11}(u)\, du\Big), \tag{3.36}$$

$$P^o_{12}(0,t) = \int_0^t P^o_{11}(0,u)q_{12}(u)\, du, \tag{3.37}$$

$$P^o_{13}(0,t) = \int_0^t P^o_{11}(0,u)q_{13}(u)\, du. \tag{3.38}$$

Estimators of the transition probabilities $P^o_{ij}(0,t)$ can be obtained by plugging (3.26) and (3.29) into (3.32), ..., (3.38).

3.5.3 Estimator of Net Survival Function

The infinitesimal matrix of the 3-state Markov process $\boldsymbol{\xi}$ is

$$\boldsymbol{Q}^*(t) = \begin{pmatrix} q^*_{00}(t) & q_{01}(t) & q_{02}(t) \\ 0 & q^*_{11}(t) & q_{12}(t) \\ 0 & 0 & 0 \end{pmatrix}, \tag{3.39}$$

where

$$q^*_{00}(t) = -q_{01}(t) - q_{02}(t),$$

$$q^*_{11}(t) = -q_{12}(t).$$

The transition probabilities $P_{ij}(0,t)$ in $\boldsymbol{\xi}$ can be solved using Kolmogorov forward equations (3.15). The solutions are:

$$P_{00}(0,t) = \exp\Big(\int_0^t q^*_{00}(u)\,du\Big), \tag{3.40}$$

$$P_{01}(0,t) = \exp\Big(\int_0^t q^*_{11}(u)\,du\Big)\int_0^t \exp\Big(\int_0^u -q^*_{11}(v)\,dv\Big)P_{00}(u)q_{01}(u)\,du, \tag{3.41}$$

$$P_{02}(0,t) = \int_0^t \Big[P_{00}(0,u)q_{02}(u) + P_{01}(0,u)q_{12}(u)\Big]\,du. \tag{3.42}$$

Actually, instead of solving the Kolmogorov equations for the 3-state $\boldsymbol{\xi}$, we can easily obtain the solutions $P_{00}(0,t)$ in (3.40), by setting $q_{03}(t) = 0$ in $q_{00}(t)$ in the solution $P^o_{00}(0,t)$ in (3.32). Similarly, $P_{01}(0,t)$ in (3.41) is obtained by setting $q_{13}(t) = 0$ in $q_{11}(t)$ in the solution $P^o_{01}(0,t)$ in (3.33), and $P_{02}(0,t)$ in (3.42) is obtained by setting $q_{03}(t) = q_{13}(t) = 0$ in the solution $P^o_{02}(0,t)$ in (3.34).

For estimation of $P_{02}(0,t)$, we need estimates of $q_{01}(t)$, $q_{02}(t)$ and $q_{12}(t)$. We can use $\hat{q}_{01}(t)$, $\hat{q}_{02}(t)$ and $\hat{q}_{12}(t)$ in (3.26) and (3.29). An estimator of $P_{02}(0,t)$ in $\boldsymbol{\xi}$ can be obtained by plugging (3.26) and (3.29) into (3.40) ... (3.42). Denote the obtained estimator as $\hat{P}_{02}(0,t)$. Then an estimator of the net survival function of a

patient is

$$\begin{aligned}\hat{S}(t) &= 1 - \hat{P}_{02}(0,t) \\ &= 1 - \int_0^t \left[\hat{P}_{00}(0,u)\hat{q}_{02}(u) + \hat{P}_{01}(0,u)\hat{q}_{12}(u)\right] du. \end{aligned} \tag{3.43}$$

3.6 Clinical Trial Terminating at a Finite Time T

In Sections 3.2 through 3.5, the estimation is carried out under the assumption that the clinical trial terminates when all the patients are dead or lost to followup where there is no termination date for the clinical trial. In this section, we will discuss estimation of $\Lambda_{ij}(0,t)$, $P_{ij}(0,t)$ and $S(t)$ for a clinical trial that terminates at a prespecified finite time T.

In (3.5), $Y_0 = \min(Y_{01}, Y_{02}, Y_{03})$, where Y_{03} is the sojourn time in state 0 before being lost to followup and the support of Y_{03} is assumed to be $[0, \infty]$. In this section, the clinical trial terminates at time T, therefore we assume that $0 < Y_{03} \leq T$. Then our observations are limited to $[0, T]$. Similar to Section 3.5.1, for each patient starting from state 0, we observe a pair of random variables (Y_0, δ_0), where:

$$Y_0 = \min(Y_{01}, Y_{02}, Y_{03}), \quad Y_0 \in (0, T],$$

$$\{\delta_0 = j\} = \{Y_0 = Y_{0j}\}, \quad j = 1, 2, 3.$$

The survival function of Y_0 is:

$$S_0(t) = \begin{cases} S_{01}(t)S_{02}(t)S_{03}(t), & 0 < t \leq T \\ 0, & t > T. \end{cases}$$

The distribution function of δ_0 is:

$$P(\delta_0 = j) = \int_0^T S_0(u)q_{0j}(u)du, \quad j = 1, 2, 3,$$

The joint distribution of Y_0 and δ_0 is:

$$\begin{aligned} G_0(t, j) &= P(Y_0 \leq t, \delta_0 = j) \\ &= \int_0^t S_0(u)q_{0j}(u)du, \quad 0 < t \leq T, \quad j = 1, 2, 3. \end{aligned} \tag{3.44}$$

The estimators of the cumulative hazard functions $\Lambda_{0j}(0, t), j = 1, 2, 3$ can be derived in the same way as in Section 3.5.1. Here we display the estimators without derivation:

$$\hat{\Lambda}_{0j}(0, t) = \sum_{k=1}^{n} \frac{I(Y_{0,k} \leq t, \delta_{0,k} = j)}{\sum_{m=1}^{n} I(Y_{0,m} > Y_{0,k})}, \quad t \in [0, t_{0j}], \quad j = 1, 2, 3. \tag{3.45}$$

For patients moved to state 1 before the end of the clinical trial, i.e., given

$Y_{01} < T$, we observe another pair of random variables (Y_1, δ_1), where:

$$\begin{aligned} Y_1 &= \min(Y_{12}, Y_{13}, T - Y_{01}), \\ \{\delta_1 = 2\} &= \{Y_1 = Y_{12}\}, \\ \{\delta_1 = 3\} &= \{Y_1 = \min(Y_{13}, T - Y_{01})\}. \end{aligned}$$

If we define $Y_{13}^* = \min(Y_{13}, T - Y_{01})$, then $Y_1 = \min(Y_{12}, Y_{13}^*)$. The survival function of Y_{13}^* is:

$$\begin{aligned} S_{13}^*(t) &= P(Y_{13}^* > t) \\ &= P(Y_{13} > t, T - Y_{01} > t \mid Y_{01} < T) \\ &= S_{13}(t)F_{01}(T - t)/F_{01}(T), \quad 0 < t \leq T. \end{aligned}$$

The survival function of Y_1 is:

$$\begin{aligned} S_1(t) &= P(Y_1 > t) \\ &= P(Y_{12} > t, Y_{13}^* > t) \\ &= S_{12}(t)S_{13}^*(t), \quad 0 < t \leq T. \end{aligned}$$

The joint distribution of (Y_1, δ_1) is:

$$\begin{aligned} G_1(t,2) &= P(Y_1 \le t, \delta_1 = 2) \\ &= \int_0^t S_1(u) q_{12}(u) du, \quad 0 < t \le T, \end{aligned}$$

$$\begin{aligned} G_1(t,3) &= P(Y_1 \le t, \delta_1 = 3) \\ &= \int_0^t S_1(u) q^*_{13}(u) du, \quad 0 < t \le T, \end{aligned}$$

where $q^*_{13}(t)$ is the hazard rate function of Y^*_{13}:

$$q^*_{13}(t) = \Big[f_{13}(t) F_{01}(T-t) + S_{13}(t) f_{01}(T-t) \Big] \Big/ \Big[S_{13}(t) F_{01}(T-t) \Big], \quad 0 < t \le T.$$

The estimators of $\Lambda_{1j}(0,t), j = 2,3$ are as follows:

$$\hat{\Lambda}_{1j}(0,t) = \sum_{k=1}^{N_{01}} \frac{I(Y_{1,k} \le t, \delta_{1,k} = j)}{\sum_{m=1}^{N_{01}} I(Y_{1,m} > Y_{1,k})}, \quad t \in [0, t_{1j}], \quad j = 2,3. \tag{3.46}$$

We see that the estimators (3.45) and (3.46) are the same as (3.28) and (3.30) on $[0,T]$. Let $(y_{0,k}, \delta_{0,k})$, $k = 1, 2, \ldots, n$, be the observed data when the observation period is $[0, \infty)$. We consider two consecutive time intervals $[0,T] \cup (T, \infty)$. For a

clinical trial that terminates at a finite time T, its observations in $[0, T]$ for patients starting in state 0 are the same as those with observation period $[0, \infty)$ but observed in $[0, T]$. We observe exactly the same data on $[0, T]$ whether the observation period is $[0, T]$ or $[0, \infty)$. Therefore, for $j = 1, 2, 3$ the estimated $\hat{\Lambda}_{0j}(0, t)$ and $\hat{\sigma}^2_{0j}(t)$ are the same on $[0, T]$ when the observation periods are $[0, T]$ and $[0, \infty)$.

However, the observations on patients in state 1 are different when $T < \infty$ instead of $T = \infty$. The observed patients are in state 1 before the termination of the clinical trial. Then for each of them the sojourn time in state 0 is less than T. Suppose we observe $(y_{1,k}, \delta_{1,k}), k = 1, 2, \ldots, n_{01}$ when $T = \infty$. When T becomes finite, for the k^{th} patient, the remaining observation time becomes $T - y_{0,k}$ given $y_{0,k} < T$. Therefore the observation on the k^{th} patient becomes $\min(y_{1,k}, T - y_{0,k})$. This differs from the case for $T = \infty$. The termination of the clinical trial ($T < \infty$) also has an impact on the variance of $\hat{\Lambda}_{1j}(0, t)$ because we observe fewer transitions. When $T < \infty$, the variance of $\hat{\Lambda}_{1j}(0, t)$ is larger than that when $T = \infty$. It can be seen in the Figures from the simulations in Section 4.2.4.

3.7 Asymptotic Distributions of the Estimators

3.7.1 Cumulative Hazard Functions

Estimators of the cumulative hazard functions $\Lambda_{0j}(0,t)$ and $\Lambda_{1j}(0,t)$ are given in (3.28) and (3.30) respectively. The estimators are derived from a Markov multi-state model. We can also formulate the estimators using counting processes. Consider a multivariate counting process $N_0(t) = (N_{01}(t), N_{02}(t), N_{03}(t))$, where $N_{0j}(t) = \sum_{k=1}^{n} I[Y_{0,k} \leq t, \delta_{0,k} = j]$, $j = 1, 2, 3$. Assume that the intensity process $\lambda_0(t) = (\lambda_{01}(t), \lambda_{02}(t), \lambda_{03}(t))$ satisfies the multiplicative model $\lambda_{0j}(t) = q_{0j}(t)H_0(t)$, where $H_0(t) = \sum_{k=1}^{n} I[Y_{0,k} \geq t]$. Then the estimator $\hat{\Lambda}_{0j}(0,t)$ becomes:

$$\hat{\Lambda}_{0j}(0,t) = \int_0^t H_{0j}(u)^{-1} dN_{0j}(u), \quad t \in [0, t_{0j}], \quad j = 1, 2, 3. \tag{3.47}$$

Similarly we can also formulate $\hat{\Lambda}_{1j}(0,t)$ in terms of counting processes. This is known as the Nelson-Aalen estimator.

The properties of the Nelson-Aalen estimator have been studied by Aalen (1975, 1978b) [1][2] and Andersen *et al.*(1993) [7]. In this section, we will derive the bias, estimated variance, uniform consistency and asymptotic normality of $\hat{\Lambda}_{0j}(0,t)$ and $\hat{\Lambda}_{1j}(0,t)$ using results from Aalen (1978b). See also Section 4.1 of Andersen *et*

al.(1993). The results are stated using Markov multi-state model notations.

(1) Bias

We have the following equations:

$$\mathbf{E}\hat{\Lambda}_{0j}(0,t) = \int_0^t q_{0j}(u)P(\hat{S}_0(u) > 0)du, \quad t \in [0, t_{0j}], \quad j = 1, 2, 3,$$
$$\mathbf{E}\hat{\Lambda}_{1j}(0,t) = \int_0^t q_{1j}(u)P(\hat{S}_1(u) > 0)du, \quad t \in [0, t_{1j}], \quad j = 2, 3.$$

If for all $t \in [0, t_{0j}]$, $\hat{S}_0(t) > 0$ and $\hat{S}_1(t) > 0$, then $\hat{\Lambda}_{0j}(0,t)$ and $\hat{\Lambda}_{1j}(0,t)$ are unbiased estimators of $\Lambda_{0j}(0,t)$ and $\Lambda_{1j}(0,t)$ on $[0, t_{0j}]$ and $[0, t_{1j}]$ respectively. Since the nonparametric estimator $\hat{\Lambda}_{0j}(0,t)$ is identifiable up to the last uncensored observation, $\hat{S}_0(t)$ is positive for all $t \in [0, t_{0j}]$. Therefore, $\hat{\Lambda}_{0j}(0,t)$ is an unbiased estimator of $\Lambda_{0j}(0,t)$ on $[0, t_{0j}]$. Similarly, we conclude that $\hat{\Lambda}_{1j}(0,t)$ is an unbiased estimator of $\Lambda_{1j}(0,t)$ on $[0, t_{1j}]$.

(2) Variance

Unbiased estimators of the variance of $\hat{\Lambda}_{0j}(0,t)$ and $\hat{\Lambda}_{1j}(0,t)$ are:

$$\hat{\sigma}^2_{0j}(t) = \int_0^t (\hat{S}_0(u))^{-2} d\hat{G}_0(u,j), \quad t \in [0, t_{0j}], \quad j = 1, 2, 3, \tag{3.48}$$
$$\hat{\sigma}^2_{1j}(t) = \int_0^t (\hat{S}_1(u))^{-2} d\hat{G}_1(u,j), \quad t \in [0, t_{1j}], \quad j = 2, 3. \tag{3.49}$$

(3) Consistency

The following theorem (Theorem 4.1.1 in Andersen *et al.*(1993)) shows that $\hat{\Lambda}_{0j}(0,t)$, $j=1,2,3$, are uniformly consistent on compact intervals.

Theorem 3.7.1. *Let $t \in [0, t_{0j}]$ for $j=1,2,3$, and assume that as $n \to \infty$,*

$$\frac{1}{n}\int_0^t \Big(q_{0j}(u)/\hat{S}_0(u)\Big)du \xrightarrow{P} 0 \tag{3.50}$$

and

$$\int_0^t \Big(1 - I[\hat{S}_0(u) > 0]\Big) q_{0j}(u)du \xrightarrow{P} 0. \tag{3.51}$$

Then, as $n \to \infty$,

$$\sup_{s\in[0,t]} \mid \hat{\Lambda}_{0j}(0,s) - \Lambda_{0j}(0,s) \mid \xrightarrow{P} 0. \tag{3.52}$$

Condition (3.51) is obvious since $\hat{S}_0$ is always positive on $[0, t_{0,j}]$ for $j=1,2,3$. We now check condition (3.50). The empirical survival function $\hat{S}_0$ converges to S_0 uniformly, that is

$$\sup_{s\geq 0} |\hat{S}_0(s) - S_0(s)| \xrightarrow{a.s.} 0.$$

Then for a given $\epsilon_0 > 0$ such that $S_0(s) - \epsilon_0 > 0$ for all $s \in [0,t]$, there exists $N > 0$ such that when $n \geq N$, $0 < S_0(s) - \epsilon_0 \leq \hat{S}_0(s) \leq S_0(s) + \epsilon_0$ for all

$s \in [0, t]$. Then

$$\frac{1}{n}\int_0^t \Big(q_{0j}(u)/\hat{S}_0(u)\Big)du \leq \frac{1}{n}\frac{1}{S_0(t)-\epsilon_0}\int_0^t q_{0j}(u)du. \tag{3.53}$$

Since $S_0(t) - \epsilon_0 > 0$ and q_{0j} is integrable on $[0, t]$, the RHS of (3.53) converges to 0 pointwise. Then condition (3.50) follows directly. The uniform consistency of $\hat{\Lambda}_{1j}(0, t)$, $j = 2, 3$ can be derived in the same way.

(4) Asymptotic Distribution

The asymptotic distribution of $\hat{\Lambda}_{0j}(0, t)$ for $j = 1, 2, 3$ on $[0, t_{0j}]$ is stated in Theorem 4.1.2 in Andersen *et al.*(1993). Define $\hat{\mathbf{\Lambda}}_0(0, t) = (\hat{\Lambda}_{01}(0, t), \hat{\Lambda}_{02}(0, t), \hat{\Lambda}_{03}(0, t))$, and define $\mathbf{\Lambda}_0(0, t)$ in a similar manner.

Theorem 3.7.2. *Let $t \in [0, t_{0j}]$ and assume that q_{0j}/S_0 is integrable over $[0, t_{0j}]$ for $j = 1, 2, 3$. Let*

$$\sigma^2_{0j}(t) = \int_0^t \Big(q_{0j}(u)/S_0(u)\Big)du, \quad j = 1, 2, 3, \tag{3.54}$$

and assume that

(A) For each $t \in [0, t_{0j}]$ and $j = 1, 2, 3$,

$$\int_0^t \Big(q_{0j}(u)/\hat{S}_0(u)\Big)du \xrightarrow{P} \sigma^2_{0j}(t) \quad \text{as } n \to \infty. \tag{3.55}$$

(B) For $j = 1, 2, 3$ and all $\epsilon > 0$,

$$\int_0^{t_{0j}} \frac{q_{0j}(u)}{\hat{S}_0(u)} I\Big\{\frac{1}{\sqrt{n}\hat{S}_0(u)} > \epsilon\Big\} du \xrightarrow{P} 0 \quad as\ n \to \infty. \tag{3.56}$$

(C) For $j = 1, 2, 3$,

$$\sqrt{n} \int_0^{t_{0j}} \Big(1 - I[\hat{S}_0(u) > 0]\Big) q_{0j}(u) du \xrightarrow{P} 0 \quad as\ n \to \infty. \tag{3.57}$$

Then

$$\sqrt{n}(\hat{\mathbf{\Lambda}}_0(0,t) - \mathbf{\Lambda}_0(0,t)) \xrightarrow{\mathcal{D}} \boldsymbol{U} = (U_{01}, U_{02}, U_{03}) \quad as\ n \to \infty \tag{3.58}$$

on $D[0, t_{01}] \times D[0, t_{02}] \times D[0, t_{03}]$, where U_{01}, U_{02}, U_{03} are independent Gaussian martingales with $U_{0j}(0) = 0$ and $cov(U_{0j}(s_1), U_{0j}(s_2)) = \sigma^2_{0j}(s_1 \wedge s_2)$, Also, for $j = 1, 2, 3$,

$$\sup_{t \in [0, t_{0j}]} \big| n\hat{\sigma}^2_{0j}(t) - \sigma^2_{0j}(t) \big| \xrightarrow{P} 0 \quad as\ n \to \infty, \tag{3.59}$$

where $\hat{\sigma}^2_{0j}(s)$ is defined by equation (3.48).

In Theorem 3.7.2, conditions A to C are satisfied if $\inf_{t\in[0,t_{0j}]} S_0(t) > 0$ and

$$\sup_{t\in[0,t_{0j}]} \left|\hat{S}_0(t) - S_0(t)\right| \xrightarrow{P} 0 \quad as\ n \to \infty, \tag{3.60}$$

assuming that $\Lambda_{0j}(0,t) < \infty$ for all $t \in [0, t_{0j}]$, $j = 1, 2, 3$. Since $\hat{S}_0$ is the empirical survival function, condition (3.60) is obviously satisfied. The asymptotic distribution of $\hat{\Lambda}_{1j}(0,t)$ can be derived in the same way.

3.7.2 Net Survival Function

The estimated $\hat{S}(t)$ in (3.43) is a function of $\hat{\Lambda}_{01}(0,t)$, $\hat{\Lambda}_{02}(0,t)$ and $\hat{\Lambda}_{12}(0,t)$:

$$\begin{aligned}
\hat{S}(t) &= 1 - \hat{P}_{02}(0,t) \\
&= 1 - \int_0^t \exp\Big(- \hat{\Lambda}_{01}(0,u) - \hat{\Lambda}_{02}(0,u)\Big) d\hat{\Lambda}_{02}(0,u) \\
&- \int_0^t \exp\Big(- \hat{\Lambda}_{12}(0,u)\Big) \Big[\int_0^u \exp\Big(\hat{\Lambda}_{12}(0,v) - \hat{\Lambda}_{01}(0,v) - \hat{\Lambda}_{02}(0,v)\Big) d\hat{\Lambda}_{01}(0,v)\Big] d\hat{\Lambda}_{12}(0,u).
\end{aligned}$$

We can derive the asymptotic distribution of $\hat{S}(t)$ using the functional delta method. Let

$$\gamma_1(F,G) = \int_0^t \exp(-F-G) dG,$$

$$\gamma_2(F,G,H) = \int_0^t \exp(-H)\Big[\int_0^u \exp(H-F-G)dF\Big]dH.$$

Then

$$\begin{aligned}\hat{S} &= \gamma(\hat{\Lambda}_{01},\hat{\Lambda}_{02},\hat{\Lambda}_{03}) \\ &= 1-\gamma_1(\hat{\Lambda}_{01},\hat{\Lambda}_{02})-\gamma_2(\hat{\Lambda}_{01},\hat{\Lambda}_{02},\hat{\Lambda}_{12}).\end{aligned}$$

We first derive the first-order Hadamard derivatives of γ_1 and γ_2. Let F_0, G_0 and H_0 be functions from $\mathbb{R}^+$ to $\mathbb{R}^+$, $F_{t_n} = F_0 + t_n F_n$, $G_{t_n} = G_0 + t_n G_n$, $H_{t_n} = H_0 + t_n H_n$, where F_n, G_n and H_n are any functions from $\mathbb{R}^+$ to $\mathbb{R}^+$ such that $F_n \to F$, $G_n \to G$, $H_n \to H$ and $t_n \to 0$ as $n \to \infty$, then the derivative of γ_1 at (F_0, G_0) is:

$$\begin{aligned}\gamma'_{1(F_0,G_0)}(F,G) &= \frac{\gamma_1(F_{t_n},G_{t_n})-\gamma_1(F_0,G_0)}{t_n} \\ &= \Big[\int_0^t \exp(-F_{t_n}-G_{t_n})dG_{t_n} - \int_0^t \exp(-F_0-G_0)dG_0\Big]/t_n \\ &= \int_0^t \Big[\frac{\exp(-F_{t_n}-G_{t_n})-exp(-F_0-G_0)}{t_n}\Big]dG_0 + \int_0^t \exp(-F_0-G_0)dG_n \\ &= -\int_0^t \Big[(F_n+G_n)\exp(-F_0-G_0)\Big]dG_0 + \int_0^t \exp(-F_0-G_0)dG_n \\ &= -\int_0^t \Big[(F+G)\exp(-F_0-G_0)\Big]dG_0 + \int_0^t \exp(-F_0-G_0)dG \quad \text{as } n\to\infty.\end{aligned} \tag{3.61}$$

To derive γ_2', notice that if we let $\gamma_3(S,H) = \int_0^t \exp(-H)SdH$, $\gamma_4(F,G,H) = \int_0^t \exp(H-F-G)dF$, $\gamma_5(F,G,H) = (\gamma_4(F,G,H),H)$, then $\gamma_2 = \gamma_3 \circ \gamma_5$. By the chain rule, we have

$$\gamma'_{2(F_0,G_0,H_0)} = \gamma'_{3(\gamma_4(F_0,G_0,H_0),H_0)} \circ \gamma'_{5(F_0,G_0,H_0)}.$$

We can derive $\gamma'_{3(S_0,H_0)}(S,H)$ and $\gamma'_{5(F_0,G_0,H_0)}(F,G,H)$ similarly as follows:

$$\begin{aligned}
\gamma'_{3(S_0,H_0)}(S,H) &= \frac{\gamma_3(S_t,H_{t_n}) - \gamma_3(S_0,H_0)}{t_n} \\
&= \Big[\int_0^t \exp(-H_{t_n})S_t dH_{t_n} - \int_0^t \exp(-H_0)S_0 dH_0\Big]/t_n \\
&= \int_0^t \Big[\frac{\exp(-H_{t_n})S_t - \exp(-H_0)S_0}{t_n}\Big]dH_0 + \int_0^t \exp(-H_0)S_0 dH_n \\
&= \int_0^t \Big[\exp(-H_0)H_n S_0 + \exp(-H_0)S_n\Big]dH_0 + \int_0^t \exp(-H_0)S_0 dH_n \\
&= \int_0^t \Big[\exp(-H_0)HS_0 + \exp(-H_0)S\Big]dH_0 + \int_0^t \exp(-H_0)S_0 dH \quad \text{as } n \to \infty.
\end{aligned} \tag{3.62}$$

$$
\begin{aligned}
\gamma'_{4(F_0,G_0,H_0)}(F,G,H) &= \frac{\gamma_4(F_{t_n},G_{t_n},H_{t_n})-\gamma_4(F_0,G_0,H_0)}{t_n} \\
&= \Big[\int_0^t \exp(H_{t_n}-F_{t_n}-G_{t_n})dF_{t_n} - \int_0^t \exp(H_0-F_0-G_0)dF_0\Big]/t_n \\
&= \int_0^t \Big[\frac{\exp(H_{t_n}-F_{t_n}-G_{t_n})-\exp(H_0-F_0-G_0)}{t_n}\Big]dF_0 \\
&\quad + \int_0^t exp(H_0-F_0-G_0)dF_n \\
&= \int_0^t \Big[(H_n-F_n-G_n)\exp(H_{t_n}-F_{t_n}-G_{t_n})\Big]dF_0 \\
&\quad + \int_0^t \exp(H_{t_n}-F_{t_n}-G_{t_n})dF_n \\
&= \int_0^t \Big[(H-F-G)\exp(H_0-F_0-G_0)\Big]dF_0 \\
&\quad + \int_0^t \exp(H_0-F_0-G_0)dF \quad \text{as } n\to\infty. \qquad (3.63)
\end{aligned}
$$

Then $\gamma'_{5(F_0,G_0,H_0)}(F,G,H) = (\gamma'_{4(F_0,G_0,H_0)}(F,G,H),H)$. Combine (3.61), (3.62) and (3.63) we have

$$
\begin{aligned}
\gamma'_{2(F_0,G_0,H_0)}(F,G,H) &= \gamma'_{3(\gamma_4(F_0,G_0,H_0),H_0)} \circ \gamma'_{5(F_0,G_0,H_0)}(F,G,H) \\
&= \gamma'_{3(\gamma_4(F_0,G_0,H_0),H_0)}\Big(-\int_0^t \Big[(H-F-G)\exp(H_0-F_0-G_0)\Big]dF_0 \\
&\qquad\qquad + \int_0^t \exp(H_0-F_0-G_0)dF, H\Big) \\
&= \int_0^t \Big[-\exp(-H_0)HS_0 + \exp(-H_0)S\Big]dH_0 \\
&\quad + \int_0^t \exp(-H_0)S_0 dH \qquad (3.64)
\end{aligned}
$$

where in (3.64),

$$
\begin{aligned}
S_0 &= \gamma_4(F_0, G_0, H_0) \\
&= \int_0^t \exp(H_0 - F_0 - G_0) dF_0 \\
S &= \int_0^t \Big[(h - F - G) \exp(H_0 - F_0 - G_0) \Big] dF_0 + \int_0^t \exp(H_0 - F_0 - G_0) dF.
\end{aligned}
$$

The first-order Hadamard differentiability of γ_1, γ_3 and γ_4 can be proved as follows.

$$
\begin{aligned}
&\frac{1}{t_n} \Big\{ \gamma_1(F_{t_n}, G_{t_n}) - \gamma_1(F_0, G_0) - \gamma'_{1,(F_0,G_0)}(t_n F_n, t_n G_n) \Big\} \\
&= \frac{1}{t_n} \Big\{ \int_0^t \exp(-F_{t_n} - G_{t_n}) dG_{t_n} - \int_0^t \exp(-F_0 - G_0) dG_0 \\
&\qquad + \int_0^t \Big[(t_n F_n + t_n G_n) exp(-F_0 - G_0) \Big] dG_0 - \int_0^t \exp(-F_0 - G_0) d(t_n G_n) \Big\} \\
&= \int_0^t \frac{\exp(-F_{t_n} - G_{t_n}) - \exp(-F_0 - G_0)}{t_n} dG_0 + \int_0^t (F_n + G_n) exp(-F_0 - G_0) dG_0 \\
&\qquad + \int_0^t \Big[\exp(-F_{t_n} - G_{t_n}) - \exp(-F_0 - G_0) \Big] dG_n \\
&= \int_0^t -(F^* + G^*) \exp(-F_0 - G_0 - t_n(F^* + G^*)) dG_0 + \int_0^t (F_n + G_n) \exp(-F_0 - G_0) dG_0 \\
&\qquad + \int_0^t \Big[\exp(-F_{t_n} - G_{t_n}) - \exp(-F_0 - G_0) \Big] dG_n, \qquad (3.65)
\end{aligned}
$$

where F^* is between F_0 and $F_0 + t_n F_n$, G^* is between G_0 and $G_0 + t_n G_n$. Let $t_n \to 0$

in (3.65). Then we have:

$$\lim_{t_n \to 0} \frac{1}{t_n} \Big\{ \gamma_1(F_{t_n}, G_{t_n}) - \gamma_1(F_0, G_0) - \gamma'_{1,(F_0,G_0)}(t_n F_n, t_n G_n) \} = 0.$$

Similarly we can check that:

$$\lim_{t_n \to 0} \frac{1}{t_n} \Big\{ \gamma_3(S_t, H_{t_n}) - \gamma_3(S_0, H_0) - \gamma'_{3,(S_0,H_0)}(t_n S_n, t_n H_n) \} = 0,$$

and

$$\lim_{t_n \to 0} \frac{1}{t_n} \Big\{ \gamma_4(F_{t_n}, G_{t_n}, H_{t_n}) - \gamma_4(F_0, G_0, H_0) - \gamma'_{4,(F_0,G_0,H_0)}(t_n F_n, t_n G_n, t_n H_n) \} = 0.$$

From Theorem 3.7.2 we have:

$$\sqrt{n} \Big((\hat{\Lambda}_{01}, \hat{\Lambda}_{02}, \hat{\Lambda}_{03}) - (\Lambda_{01}, \Lambda_{02}, \Lambda_{03}) \Big) \xrightarrow{\mathcal{D}} (U_{01}, U_{02}, U_{03}) \quad as \ n \to \infty,$$

$$\sqrt{n_{01}} \Big((\hat{\Lambda}_{12}, \hat{\Lambda}_{13}) - (\Lambda_{12}, \Lambda_{13}) \Big) \xrightarrow{\mathcal{D}} (U_{12}, U_{13}) \quad as \ n \to \infty.$$

Let $N = n + n_{01}$ and assume that $n/N \to \lambda \in (0, 1)$ as $n \to \infty$. We can further

obtain

$$\sqrt{N}\Big((\hat{\Lambda}_{01},\hat{\Lambda}_{02},\hat{\Lambda}_{12})-(\Lambda_{01},\Lambda_{02},\Lambda_{12})\Big)\xrightarrow{\mathcal{D}}(\frac{U_{01}}{\sqrt{\lambda}},\frac{U_{02}}{\sqrt{\lambda}},\frac{U_{12}}{\sqrt{1-\lambda}})\quad as\ n\to\infty.$$

Then the asymptotic distribution of $\hat{S}$ follows from the functional delta method (Van der Vaart (1997), Theorem 20.8):

$$\sqrt{N}\Big(\gamma(\hat{\Lambda}_{01},\hat{\Lambda}_{02},\hat{\Lambda}_{12})-\gamma(\Lambda_{01},\Lambda_{02},\Lambda_{12})\Big)\xrightarrow{\mathcal{D}}\gamma'_{\mathbf{\Lambda}}(\frac{U_{01}}{\sqrt{\lambda}},\frac{U_{02}}{\sqrt{\lambda}},\frac{U_{12}}{\sqrt{1-\lambda}}),\tag{3.66}$$

where $\mathbf{\Lambda}=(\Lambda_{01},\Lambda_{02},\Lambda_{12})$ and $\gamma'_{\mathbf{\Lambda}}=-\gamma'_{1\mathbf{\Lambda}}$ - $\gamma'_{2\mathbf{\Lambda}}$ with $\gamma'_{\mathbf{\Lambda}}$, $\gamma'_{1\mathbf{\Lambda}}$ and $\gamma'_{2\mathbf{\Lambda}}$ being the first-order Hadamard derivatives of γ, γ_1 and γ_2 at $\mathbf{\Lambda}$. By (3.66), the asymptotic distribution of $\hat{S}$ follows.

3.8 Application to Aplastic Anemia Data

In Section 3.2, we discuss the Aplastiic Anemia data as a motivating example. A four-state Markov process is constructed for this problem. We apply the nonparametric estimators of the cumulative hazard functions ((3.28) and (3.30)), transition probabilities and net survival function ((3.43)) to the AA data.

There are 238 patients in the clinical trial. The clinical trial lasts for about 21 years. Each patient was followed after the first IST treatment. Over time, 80 of the

patients relapsed, 151 were lost to followup and 7 died directly after the first IST. These refer to direct transitions from state 0 to state 1, state 2 and state 3. Further, of the 80 relapsed patients, 23 died and 57 were lost to followup. The number of patients on each transition path is illustrated in Figure 3.2. For each patient, the data include the times when transitions happened and hence we can calculate the sojourn times between any two direct transitions. Our purpose is to estimate the survival function of the patients.

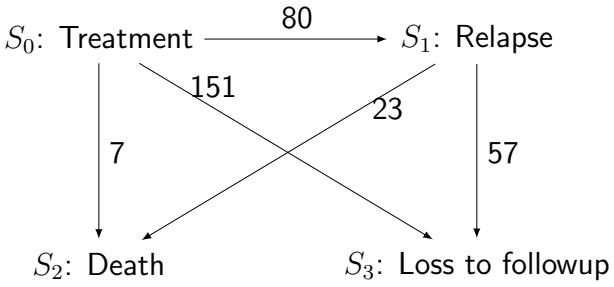
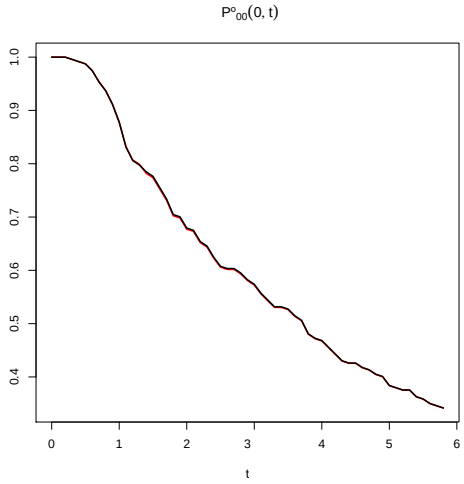


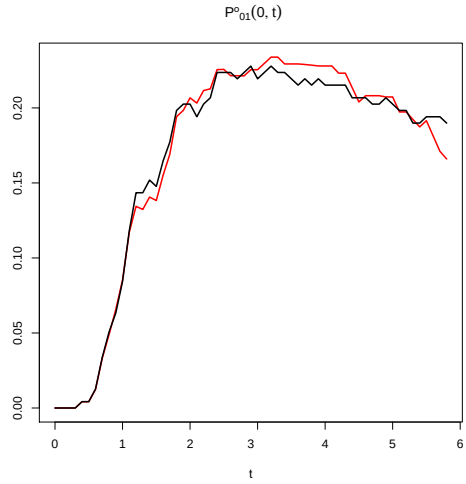
Figure 3.2: Transition paths of AA patients

We plot the estimated crude transition probabilities $\hat{P}^o_{ij}(0,t)$ in Figure 3.3. We compare each $\hat{P}^o_{ij}(0,t)$ with the corresponding empirical transition probability $P^o_{ij,n}(0,t)$ which is defined as

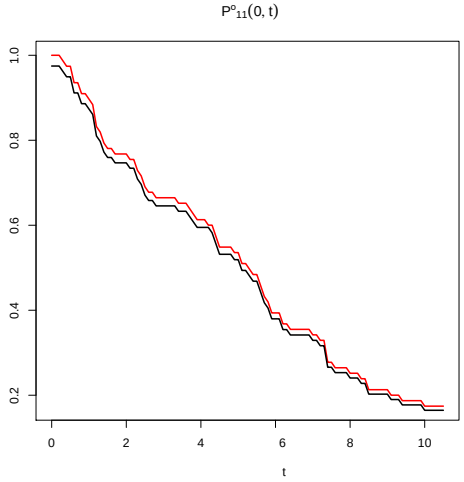
$$P^o_{ij,n}(0,t) = \frac{\#\ \{\text{patients in state i at time 0 and in state j at time t}\}}{\#\ \{\text{patients in state i at time 0}\}}.$$



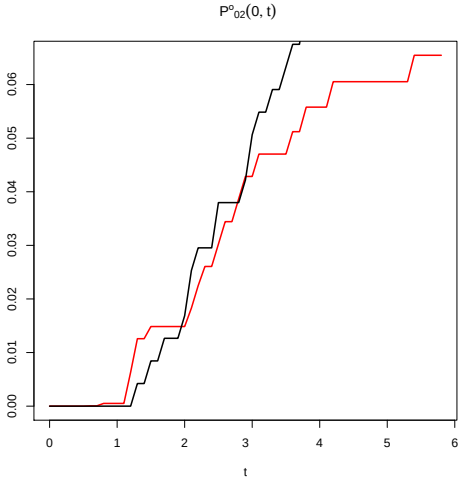
(a) P^o_{00}



(b) P^o_{01}



(c) P^o_{11}



(d) P^o_{02}

Figure 3.3: Comparison of $\hat{P}^o_{ij}(0,t)$ and $P^o_{ij,n}(0,t)$. The black line is $P^o_{ij,n}(0,t)$, the red line is $\hat{P}^o_{ij}(0,t)$.

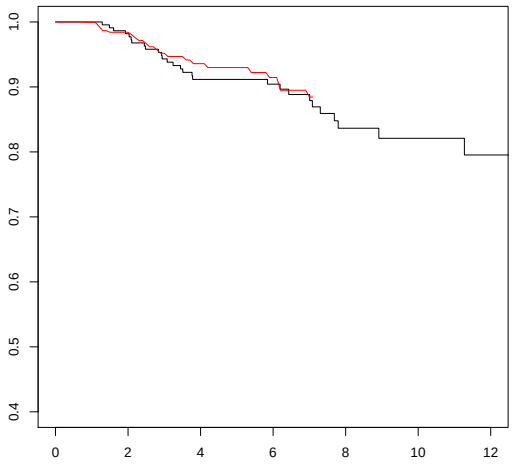


Figure 3.4: Comparison of $\hat{S}(t) = 1 - \hat{P}_{02}(0,t)$ and the Kaplan-Meier estimator. The black line is the Kaplan-Meier estimator, the red line is $\hat{S}(t)$. Censoring rate is $\frac{207}{237} = 0.87$.

The graph of the estimated net survival function $\hat{S}(t) = 1 - \hat{P}_{02}(0,t)$ and the Kaplan-Meier estimator is shown in Figure 3.4.

Chapter 4: Simulation for Irreversible Markov Model

Simulations are often used to study distributional properties of estimators both for large samples and small samples, especially when the finite-sample distribution is hard to derive. Also, accuracy of the estimation results obtained from the simulation can be verified by the theoretical model that are used for the simulation. In this section, simulations are conducted to check the goodness of fit of the estimators of cumulative hazard functions.

Since our clinical trial data are modeled by Markov processes, our basic sample consists of n mutually independent Markov processes $\{\xi_k^o(t), t \geq 0\}$, $k = 1, 2, \ldots, n$. We will simulate $\{\xi^o(t), t \geq 0\}$ for each of the n patients.

4.1 Simulation Procedure

In Section 3.3, we use Y_0 (see (3.5)) to represent the sojourn time of a patient in the initial state 0 before he/she moves to to the next state. It is assumed that Y_0 is a random variable determined by 3 other mutually independent nonnegative random

variables Y_{01}, Y_{02} and Y_{03} as $Y_0 = \min(Y_{01}, Y_{02}, Y_{03})$. After staying in state 0 for Y_0 amount of time, the patient can enter either state 1, 2 or 3 according to the following chance mechanism. Of the three sojourn times $Y_{0j}, j = 1, 2, 3$, we only observe the minimum. As far as simulation goes, there are two different but equivalent ways of simulating the Markov process.

The first method: starting from state 0, we generate three random variables $Y_{01} = y_{01}$, $Y_{02} = y_{02}$ and $Y_{03} = y_{03}$, denote the observed minimum of y_{01}, y_{02} and y_{03} as y_0 and record the index δ_0 of the minimum random variable, $\delta_0 = 1, 2, 3$. If y_{02} or y_{03} is the minimum, i.e., $\delta_0 = 2$ or 3, then the next transition is to state 2 or 3 accordingly. Since both state 2 and state 3 are absorbing states, we are done with this transition path of this patient. If y_{01} is the minimum, i.e., $\delta_0 = 1$, then starting from state 1 (at time y_{01}), we generate two random variables $Y_{12} = y_{12}$ and $Y_{13} = y_{13}$, in which y_{12} is the sojourn time in state 1 before moving to state 2, and y_{13} is the sojourn time in state 1 before moving to state 3. Similar to the direct transition from the initial state 0, we record the minimum Y_1 and the index δ_1 of the minimum random variable for the next transition event. The simulated data are pairs of random variables: (Y_0, δ_0) and (Y_1, δ_1), where Y_j and δ_j are dependent random variables.

The second method is to use the multinomial distributions discussed in (3.17)

$\sim$ (3.19) in Section 3.5.1. To begin with, we generate a sojourn time $Y_0 = y_0$ directly from the distribution $F_0(t) = 1 - S_0(t)$ (see (3.6)) using inverse transformation sampling, where Y_0 is the time spent in state 0 before the patient entering the next state. To be clear, consider an example using the Weibull distribution. Assume that Y_0 follows the Weibull distribution with CDF $F_0(t) = 1 - \exp(-(t/\eta)^\gamma)$. Simulate a $Unif[0, 1]$ random variable x_0, let $y_0 = F_0^{-1}(x_0)$ where F_0^{-1} is the inverse function of F_0. Then y_0 is an observed random sample from $Weibull(\gamma, \eta)$. After we simulate the sojourn time in state 0, we need to decide which state the patient has entered or we say which event has occurred. The index of the next event follows a multinomial distribution. The possible values of a multinomial trial are 1, 2 and 3 with probability p_{01}, p_{02} and p_{03} respectively. The probabilities $p_{0j}, j = 1, 2, 3$ are defined in (3.17). If the next event is to enter state 2 or 3, we stop the simulation. Because they are absorbing states. Otherwise, if the next event is entering state 1, we generate sojourn time $Y_1 = y_1$. Then there are two choices: either state 2 or state 3. Thus instead of a trinomial distribution we have a Bernoulli distribution of choices where the probability of choosing state 2 is $p_{12}(y_1)$ and of choosing state 3 is $p_{13}(y_1) = 1 - p_{12}(y_1)$. The number chosen is the index of the next occurring event.

It is easy to show that the two methods described above are mathematically equivalent. First we note that the distributions of δ_0 and δ_1, are the same for both

methods. The distribution of δ_0 is trinomial. In method one, we have:

$$\begin{aligned} P(\delta_0 = j) &= P(Y_{0j} = \min(y_{01}, y_{02}, y_{03})) \\ &= \int_0^\infty S_0(u) q_{0j}(u) du, \\ &= \alpha_{0j}, \quad j = 1, 2, 3. \end{aligned} \tag{4.1}$$

In method two, the distribution of δ_0 is:

$$\begin{aligned} P(\delta_0 = j) &= \int_0^\infty -\frac{q_{0j}(u)}{q_{00}(u)} dF_0(u) \\ &= \int_0^\infty S_0(u) q_{0j}(u) du, \\ &= \alpha_{0j}, \quad j = 1, 2, 3. \end{aligned} \tag{4.2}$$

Second, the joint distributions of (Y_0, δ_0) in each of the two methods are also the same (see (3.8)). Therefore, the two simulation methods are mathematically equivalent. We can choose either method to simulate the transition paths. The following simulations are carried out using method two as it has a more natural interpretation in terms of the clinical trial.

The simulation procedure is as follows:

(1) Generate $Y_0 = y_0$ from distribution function $F_0(t) = P(Y_0 \leq t) = 1 - S_0(t)$

(see (3.6)) using inverse transform sampling.

(2) Take a number δ_0 from the set $\{1, 2, 3\}$ with trinomial probabilities

$$\Big(- q_{01}(y_0)/q_{00}(y_0), -q_{02}(y_0)/q_{00}(y_0), -q_{03}(y_0)/q_{00}(y_0)\Big).$$

(3) If $\delta_0 = 2$ or $\delta_0 = 3$, stop. The observed transition path is $0 \to 2$ or $0 \to 3$, and the observed time of transition is y_0. Otherwise continue to step (4).

(4) If $\delta_0 = 1$, generate $Y_1 = y_1$ from distribution function $F_1(t) = P(Y_1 \leq t) = 1 - S_1(t)$ (see (3.11)).

(5) Take a number δ_1 from the set $\{2, 3\}$ with binomial probabilities

$$\Big(- q_{12}(y_1)/q_{11}(y_1), -q_{13}(y_1)/q_{11}(y_1)\Big).$$

(6) If $\delta_1 = 2$ or $\delta_1 = 3$, the observed transition path is $0 \to 1 \to 2$ or $0 \to 1 \to 3$, the observed times of transition are y_0 and y_1.

Repeat the procedure (1) through (6) for each of the n patients to obtain a sample of n independent Markov process, where n is the prespecified sample size. In the next section, simulations are conducted under the following assumptions:

(a) $Y_{01} \sim gamma(2.6, 0.8)$, $Y_{02} \sim gamma(4, 1)$ where in $gamma(\alpha, \beta)$, α is the shape parameter and β is the rate parameter.

(b) $Y_{03} \sim Weibull(1.5, 5)$, $Y_{12} \sim Weibull(1.5, 4)$, $Y_{13} \sim Weibull(1.2, 6)$ where in $Weibull(\gamma, \eta)$, γ is the shape parameter and η is the scale parameter.

(c) $Y_{ij}, i, j = 0, 1, 2, 3$ are independent random variables.

The parameters used in the *gamma* and $Weibull$ distributions are selected so that the mean of each Y_{ij} is similar to the corresponding mean calculated from the Aplastic Anemia data set.

4.2 Simulation Results

Estimation of the cumulative hazard functions, the transition probabilities and the net survival function are carried out with the simulated data. Our simulated data are as follows:

(1) $\Big\{(y_{0,k}, \delta_{0,k}) \mid k \in \{1, \ldots, n\}, \delta_{0,k} = 1\Big\}$,

(2) $\Big\{(y_{0,k}, \delta_{0,k}) \mid k \in \{1, \ldots, n\}, \delta_{0,k} = 2\Big\}$,

(3) $\Big\{(y_{0,k}, \delta_{0,k}) \mid k \in \{1, \ldots, n\}, \delta_{0,k} = 3\Big\}$,

(4) $\Big\{(y_{1,k}, \delta_{1,k}) \mid k \in \{1, \ldots, n_{01}\}, \delta_{1,k} = 2\Big\}$,

(5) $\Big\{(y_{1,k}, \delta_{1,k}) \mid k \in \{1, \ldots, n_{01}\}, \delta_{1,k} = 3\Big\}$.

4.2.1 Estimators of Cumulative Hazard Functions with Simulated Data

Estimators of the cumulative hazard functions $\Lambda_{0j}(0,t)$ and $\Lambda_{1j}(0,t)$ can be calculated by plugging the simulated data into (3.28) and (3.30):

$$\hat{\Lambda}_{0j}(0,t) = \sum_{k=1}^{n} \frac{I(y_{0,k} \leq t, \delta_{0,k} = j)}{\sum_{m=1}^{n} I(y_{0,m} > y_{0,k})}, \quad t \in [0, t_{0j}], \quad j = 1,2,3, \tag{4.3}$$

$$\hat{\Lambda}_{1j}(0,t) = \sum_{k=1}^{n_{01}} \frac{I(y_{1,k} \leq t, \delta_{1,k} = j)}{\sum_{m=1}^{n_{01}} I(y_{1,m} > y_{1,k})}, \quad t \in [0, t_{1j}], \quad j = 2,3. \tag{4.4}$$

We only display the graphs of $\Lambda_{01}(0,t)$ and $\Lambda_{03}(0,t)$ and their estimates in Figure 4.1 and Figure 4.2, where $\Lambda_{01}(0,t)$ is the cumulative hazard function of $gamma(2.6, 0.8)$ and $\Lambda_{03}(0,t)$ is the cumulative function of $Weibull(1.5, 5)$. On each figure, we display the graphs corresponding to sample size $n = 200$, $n = 500$ and $n = 1000$. The black line represents $\Lambda_{ij}(0,t)$, the red line represents $\hat{\Lambda}_{ij}(0,t)$, the blue dashed lines represents the 95% confidence intervals of $\hat{\Lambda}_{ij}(0,t)$, the green dot-dashed line indicates the largest uncensored survival time in the sample, that is $\max_{k=1,2,\ldots,n_{ij}} \{y_{i,k} \mid \delta_{i,k} = j\}$. Each $\Lambda_{ij}(0,t)$ is only identifiable up to the largest uncensored survival time. By comparing $\hat{\Lambda}_{ij}(0,t)$ with $\Lambda_{ij}(0,t)$, we see that in general as the sample size increases, the deviation of $\hat{\Lambda}_{ij}(0,t)$ from $\Lambda_{ij}(0,t)$ decreases. The

estimated variance of $\hat{\Lambda}_{0j}(0,t)$ and $\hat{\Lambda}_{1j}(0,t)$ are as follows:

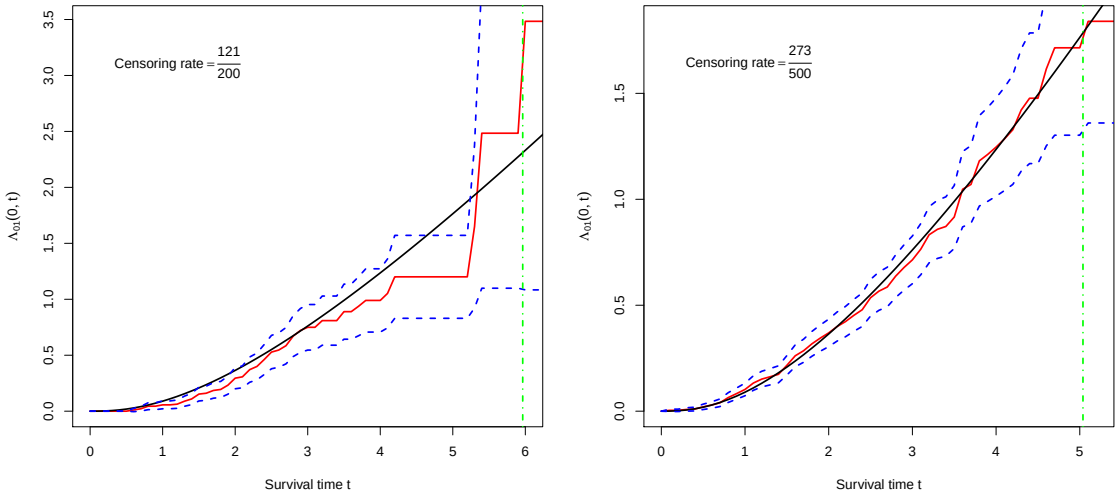
$$\hat{\sigma}^2_{0j}(t) = \sum_{k=1}^{n} \frac{I[Y_{0,k} \leq t, \delta_{0,k} = j]}{\left[\sum_{m=1}^{n} I[Y_{0,m} \geq Y_{0,k}]\right]^2}, \quad t \in [0, t_{0j}], \quad j = 1, 2, 3, \tag{4.5}$$

$$\hat{\sigma}^2_{1j}(t) = \sum_{k=1}^{n_{01}} \frac{I[Y_{1,k} \leq t, \delta_{1,k} = j]}{\left[\sum_{m=1}^{n_{01}} I[Y_{1,m} \geq Y_{1,k}]\right]^2}, \quad t \in [0, t_{1j}], \quad j = 2, 3. \tag{4.6}$$

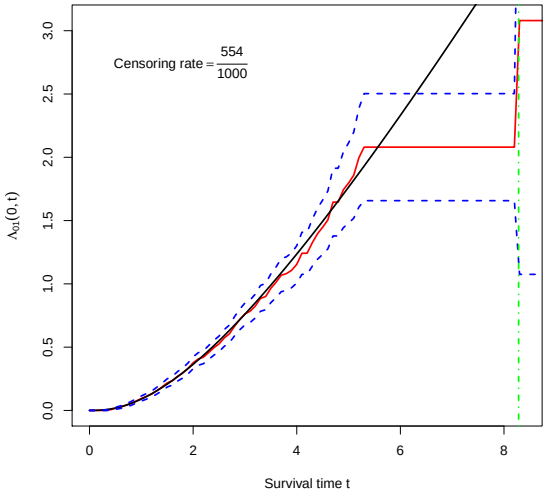
The estimated variances $\hat{\sigma}^2_{0j}(t)$ and $\hat{\sigma}^2_{1j}(t)$ are functions of the sample size n. For any $t \in [0, t_{ij}]$, $\hat{\sigma}^2_{ij}(t) = O(1/n)$. Therefore the variance decreases as the sample size grows. As a result, the width of the confidence interval decreases as the sample size grows.

Figure 4.1 and 4.2 are generated from a single simulation. Now we repeat the simulation for 1000 times and calculate the mean of the estimated cumulative hazard functions at each time point. We plot the mean computed from 1000 simulations. We only display the graph of Λ_{01} for $n = 200$ and $n = 500$ in Figure 4.3 to demonstrate the difference between a single simulation and repeated simulations. Notice that in Figure 4.3, the red line is smooth and almost coincides with the black line. Because when we repeat the simulation for 1000 times, at each time point t, by the Law of Large Numbers the mean of $\hat{\Lambda}_{01}(0,t)$ converges in probability to $\Lambda_{01}(0,t)$ as the sample size approaches infinity.

$$\Lambda_{01}(0,t) = -\log\left(1 - \int_0^t [\beta^\alpha x^{\alpha-1}\exp(-\beta x)]dx/\Gamma(\alpha)\right),\ t>0,\ \alpha = 2.6, \beta = 0.8$$



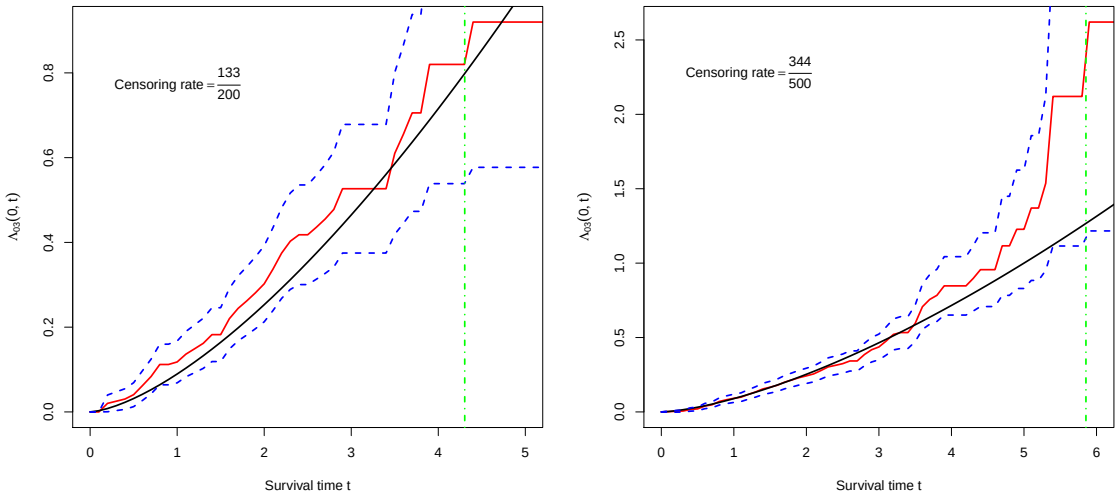
(a) $n = 200$ (b) $n = 500$



(c) $n = 1000$

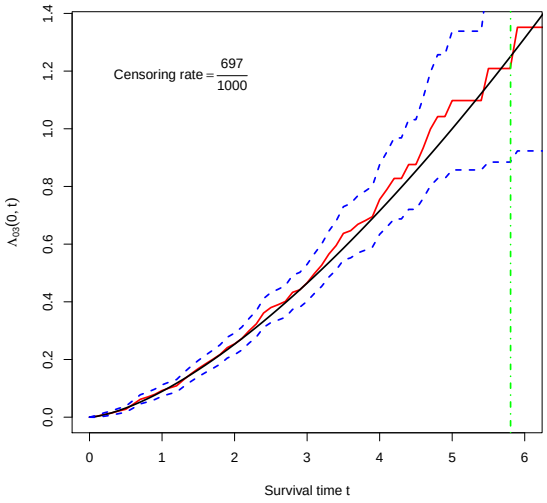
Figure 4.1: Comparison of $\hat{\Lambda}_{01}(0,t)$ and $\Lambda_{01}(0,t)$ for $n = 200, 500, 1000$, where the black line is $\Lambda_{01}(0,t)$, the red line is $\hat{\Lambda}_{01}(0,t)$, the blue lines are the 95% confidence interval, the green line indicates the largest uncensored observation: $\max(y_{0,k}|\delta_{0,k} = 1)$.

$$\Lambda_{03}(0,t) = \left(\frac{t}{\eta}\right)^{\gamma},\ t > 0,\ \gamma = 1.5, \eta = 5$$



(a) $n = 200$

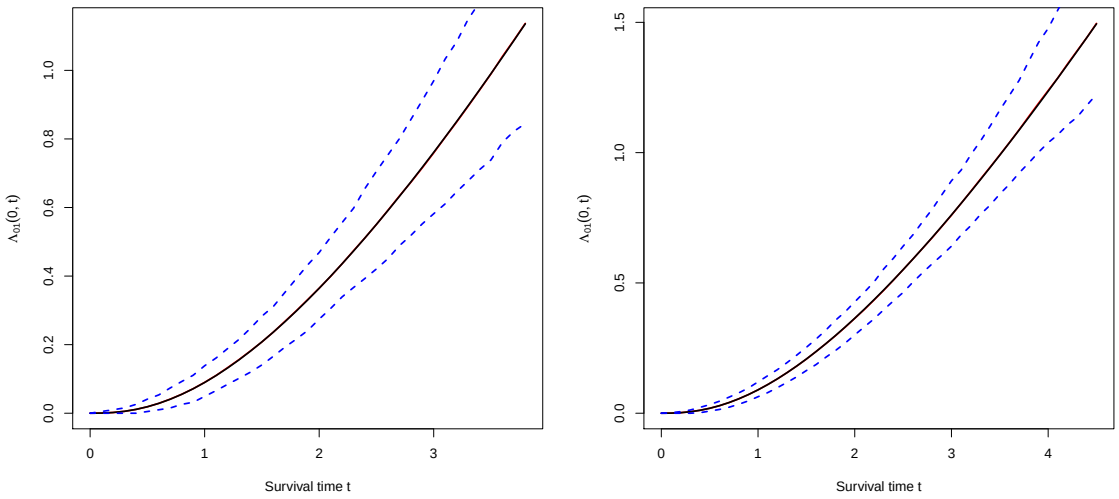
(b) $n = 500$



(c) $n = 1000$

Figure 4.2: Comparison of $\hat{\Lambda}_{03}(0,t)$ and $\Lambda_{03}(0,t)$ for $n = 200$ and $n = 500$, where the black line is $\Lambda_{03}(0,t)$, the red line is $\hat{\Lambda}_{03}(0,t)$, the blue line is the 95% confidence interval, the green line indicates the largest uncensored observation: $\max(y_{0,k}|\delta_{0,k} = 3)$.

$$\Lambda_{01}(0,t) = -\log\left(1 - \int_0^t [\beta^\alpha x^{\alpha-1}\exp(-\beta x)]dx/\Gamma(\alpha)\right),\ t>0,\ \alpha = 2.6, \beta = 0.8$$



(a) $n = 200$ (b) $n = 500$

Figure 4.3: Comparison of $\Lambda_{01}(0,t)$ and the mean of $\hat{\Lambda}_{01}(0,t)$ computed from 1000 simulations for $n = 200, 500$. The black line is $\Lambda_{01}(0,t)$, the red line is the mean of $\hat{\Lambda}_{01}(0,t)$, the blue lines are the 95% confidence interval.

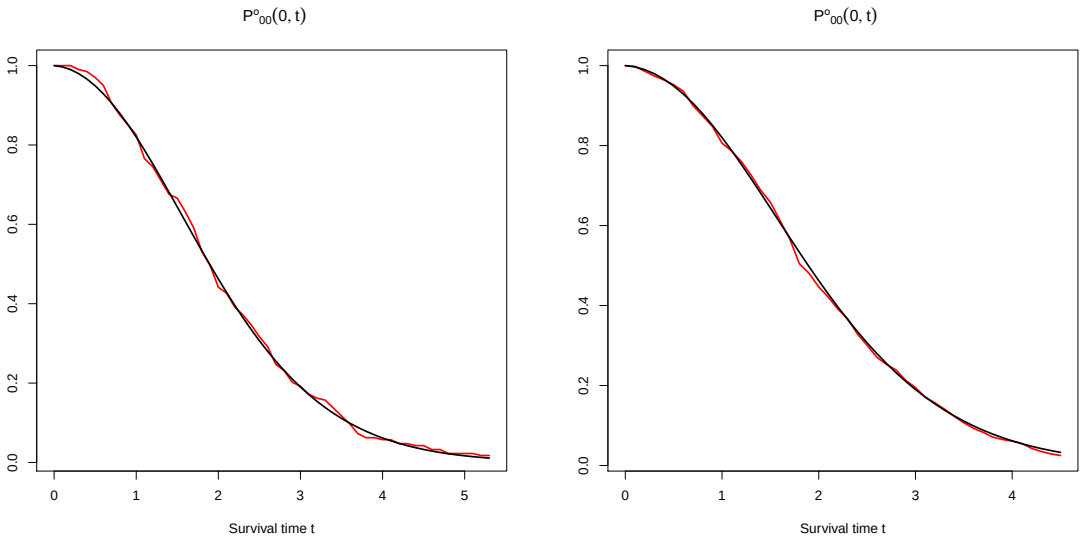
4.2.2 Estimators of Crude Transition Probabilities with Simulated Data

In Section 3.5.2, we obtained estimators of the crude transition probabilities $P_{ij}^{o}(0,t)$ using Kolmogorov froward equations. Of the seven nonzero transition probabilities in $\boldsymbol{\xi}^0$, we are of particular interest in $P_{00}^{o}(0,t)$, $P_{01}^{o}(0,t)$, $P_{11}^{o}(0,t)$ and $P_{02}^{o}(0,t)$ because $1-P_{02}(0,t)$ is the net survival function that is a function of $P_{00}(0,t)$, $P_{01}(0,t)$ and $P_{11}(0,t)$. Because the graph of $P_{11}^{o}(0,t)$ is similar to that of $P_{00}^{o}(0,t)$, we only display the graphs of $P_{00}^{o}(0,t)$, $P_{01}^{o}(0,t)$, $P_{02}^{o}(0,t)$ and their estimates in Figure 4.4 through Figure 4.6. In each figure, we display the graphs corresponding to $n = 200$, $n = 500$ and $n = 1000$.

4.2.3 Estimator of Net Survival Probability with Simulated Data

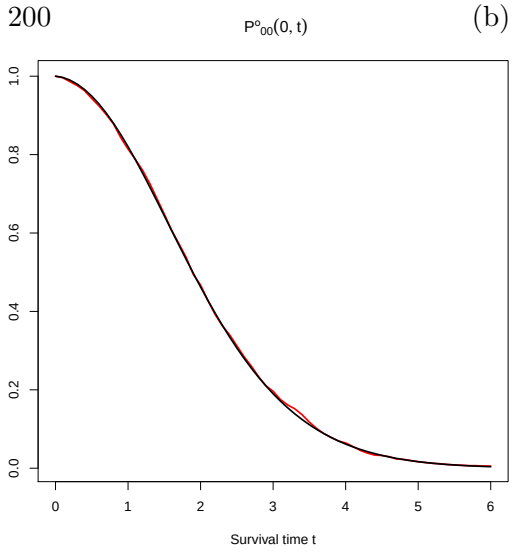
We plot $\hat{S}(t)$ (see (3.43)) under our Markov model and compare it with the Kaplan-Meier estimator for $n = 200$, $n = 500$ and $n = 1000$ in Figure 4.7.

Figure 4.7 shows that the Kaplan-Meier estimator of the survival probability is higher than $\hat{S}(t)$ for all $t > 0$ and all three sample sizes.



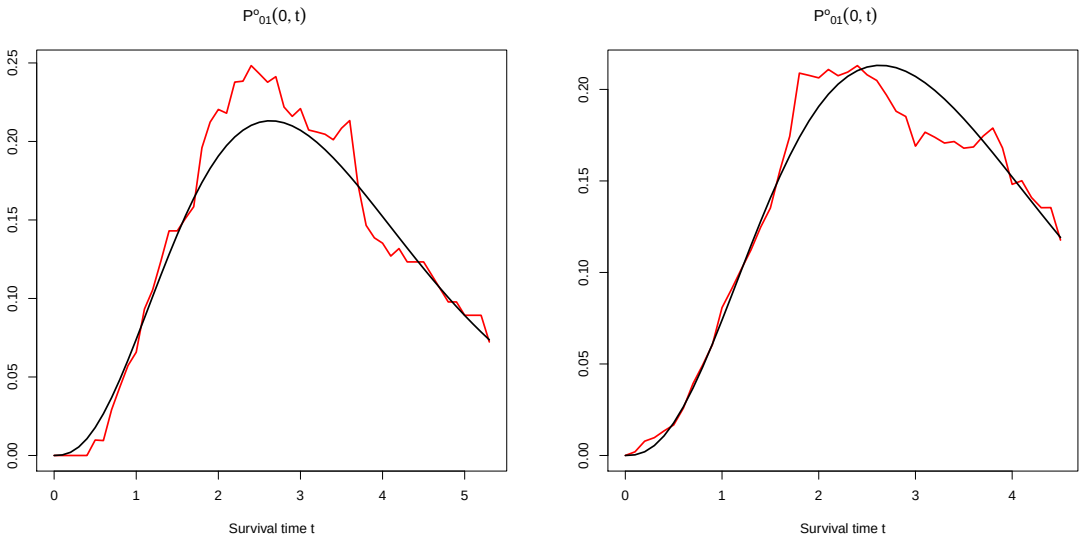
(a) $n = 200$

(b) $n = 500$



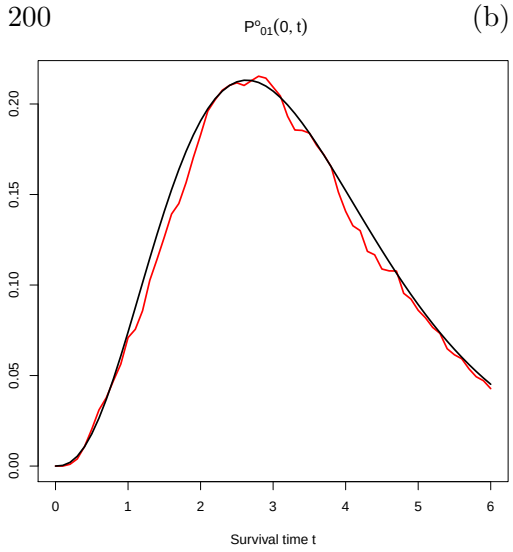
(c) $n = 1000$

Figure 4.4: Comparison of $\hat{P}^o_{00}(0,t)$ and $P^o_{00}(0,t)$ for $n = 200$, $n = 500$ and $n = 1000$. The black line represents $P^o_{00}(0,t)$, the red line represents $\hat{P}^o_{00}(0,t)$.



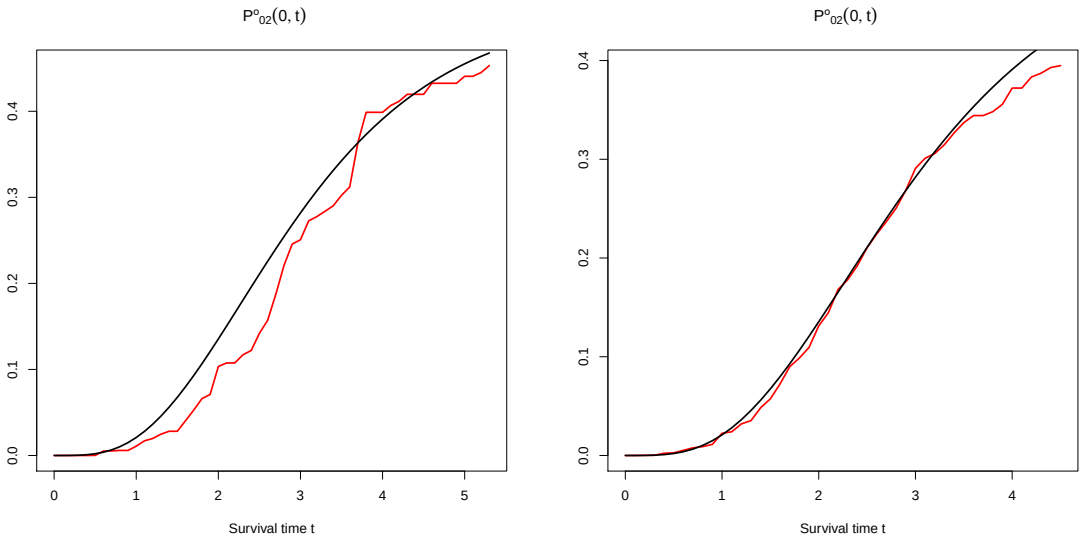
(a) $n = 200$

(b) $n = 500$



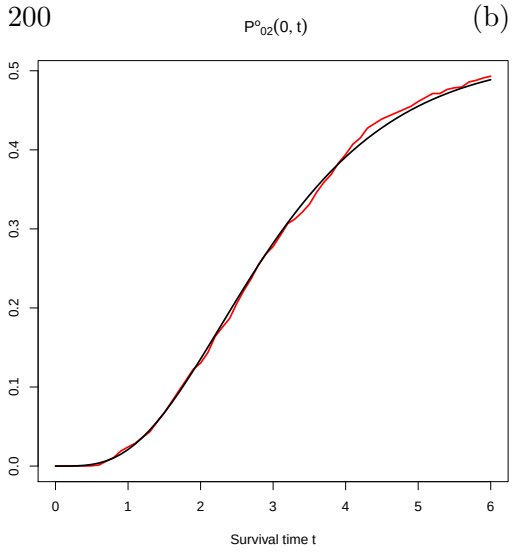
(c) $n = 1000$

Figure 4.5: Comparison of $\hat{P}^o_{01}(0,t)$ and $P^o_{01}(0,t)$ for $n = 200$, $n = 500$ and $n = 1000$. The black line represents $P^o_{01}(0,t)$, the red line represents $\hat{P}^o_{01}(0,t)$.



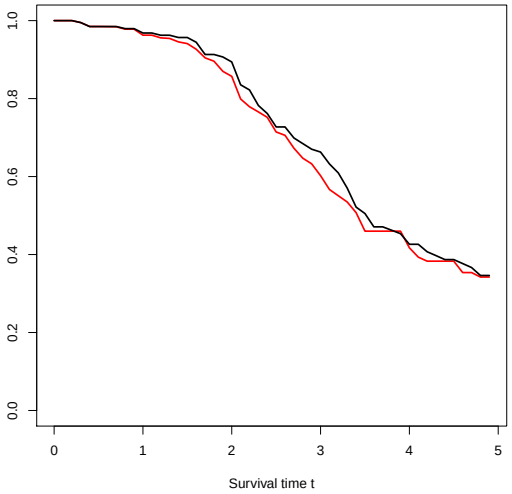
(a) $n = 200$

(b) $n = 500$

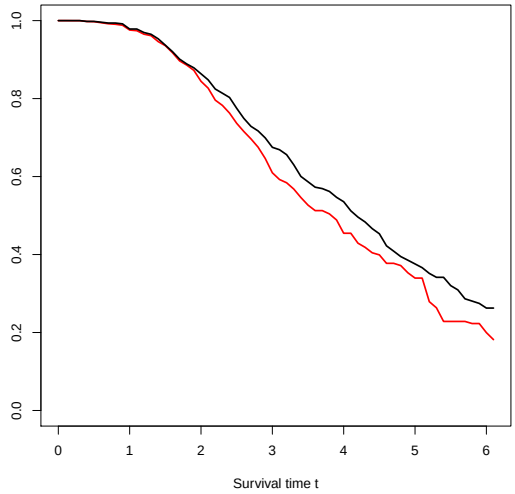


(c) $n = 1000$

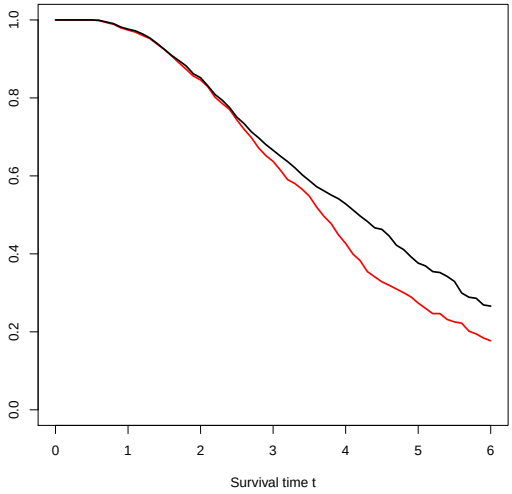
Figure 4.6: Comparison of $\hat{P}^o_{02}(0,t)$ and $P^o_{02}(0,t)$ for $n = 200$, $n = 500$ and $n = 1000$. The black line represents $P^o_{02}(0,t)$, the red line represents $\hat{P}^o_{02}(0,t)$.



(a) $n = 200$, censoring rate $= \frac{103}{200}$



(b) $n = 500$, censoring rate $= \frac{227}{500}$



(c) $n = 1000$, censoring rate $= \frac{489}{1000}$

Figure 4.7: Comparison of $\hat{S}(t) = 1 - \hat{P}_{02}(0,t)$ ((3.43)) and Kaplan-Meier estimator for $n = 200$, $n = 500$ and $n = 1000$. The black line represents the Kaplan-Meier estimator, the red line represents $\hat{S}(t)$.

4.2.4 Clinical Trial Terminating at a Finite Time T

We compare the graphs of $\hat{\Lambda}_{ij}(0,t)$, $\hat{P}^o_{ij}(0,t)$ and $\hat{S}(t)$ for $T=\infty$ and $T<\infty$ in Figure 4.8 through Figure 4.10. We use $T=5$ in the simulation.

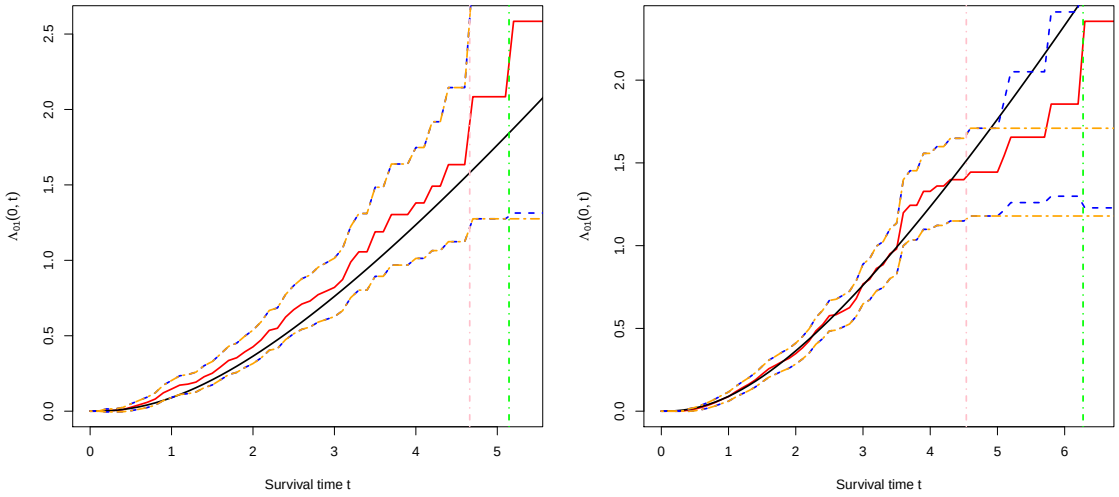
Similar to Section 4.2, we estimated the crude transition probabilities $P^o_{00}(0,t)$, $P^o_{01}(0,t)$, $P^o_{11}(0,t)$, $P^o_{02}(0,t)$ and the net survival function $S(t)$. We only display the graph of $\hat{S}(t)$ and compare it with Kaplan-Meier estimatorfor $n=200$, $n=500$ and $n=1000$.

From the Figure 4.10, we can see that when $T=\infty$ and $T=5$, the Kaplan-Meier estimators are the same on $[0,5]$. And $\hat{S}(t)$ is lower than the Kaplan-Meier estimator for all three sample sizes both when $T=\infty$ and $T=5$. A possible reason for this may be as follows. From Section 2.2 we know that the Kaplan-Meier model can be described by a 3-state Markov process. The three states are: treatment (S_0), death (S_1) and loss to followup (S_3). It does not include a state for relapse, thus it does not utilize all the available patients status data.

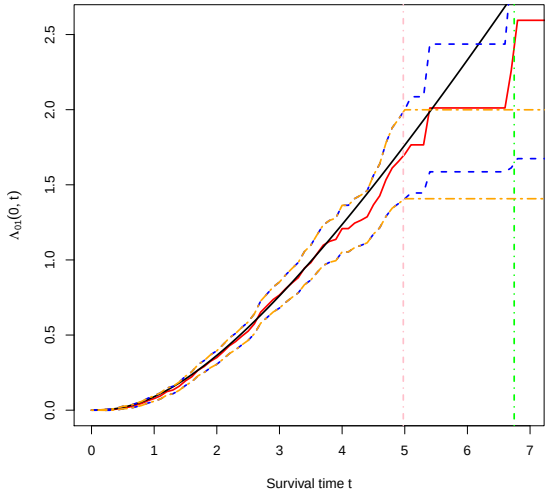
4.3 Hypothesis Testing

It can be seen from Figure 4.1 and Figure 4.2 that $\hat{\Lambda}_{01}(0,t)$ and $\hat{\Lambda}_{03}(0,t)$ are close to the true cumulative hazard functions and the deviation decreases as the

$$\Lambda_{01}(0,t) = -log\Big(1 - \int_0^t [\beta^\alpha x^{\alpha-1} exp(-\beta x)]dx/\Gamma(\alpha)\Big),\ t > 0,\ \alpha = 2.6, \beta = 0.8$$



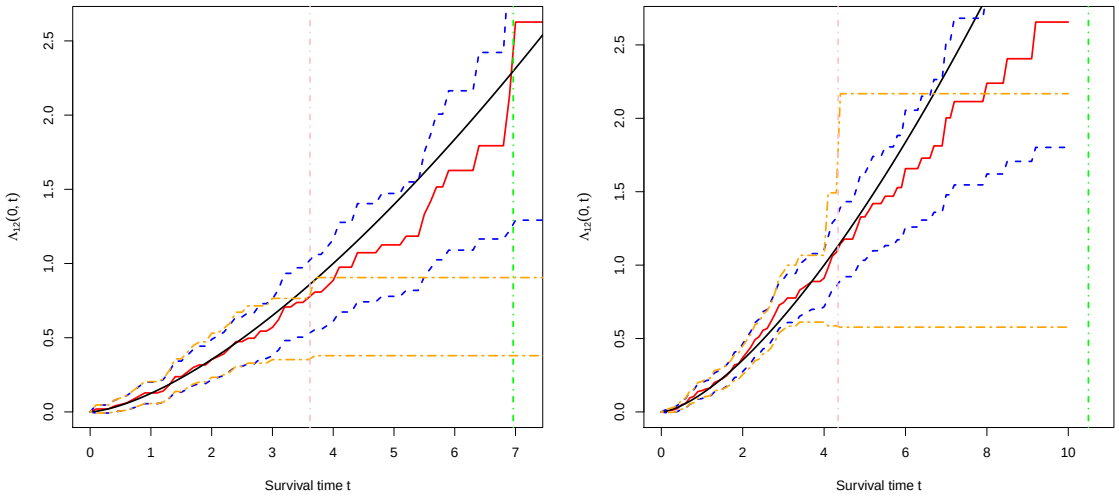
(a) $n = 200$ (b) $n = 500$



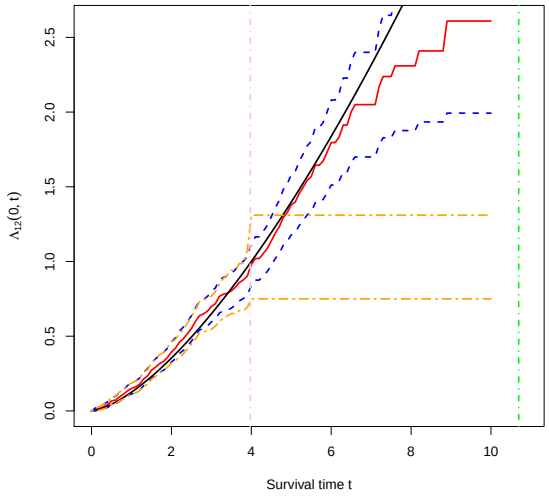
(c) $n = 1000$

Figure 4.8: Comparison of $\hat{\Lambda}_{01}(0,t)$ and $\Lambda_{01}(0,t)$ for $n = 200, 500, 1000$, where the black line is $\Lambda_{01}(0,t)$, the red line is $\hat{\Lambda}_{01}(0,t)$, the blue lines are the 95% confidence interval when $T = \infty$, the orange lines are the 95% confidence interval when $T = 5$, the green line indicates the largest uncensored observation when $T = \infty$, the pink line indicates the largest uncensored observation when $T = 5$.

$$\Lambda_{12}(0,t) = \left(\frac{t}{\eta}\right)^{\gamma},\ t > 0,\ \gamma = 1.5, \eta = 4$$

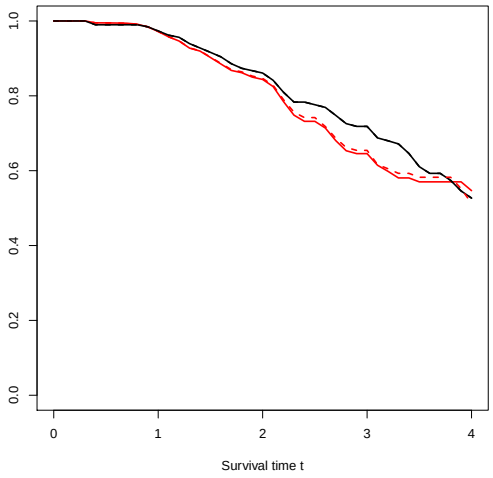


(a) $n = 200$ (b) $n = 500$

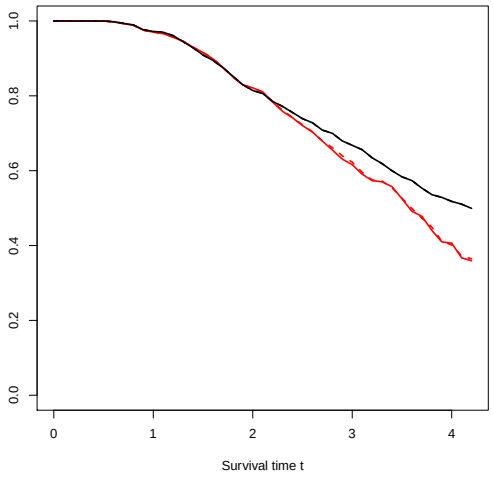


(c) $n = 1000$

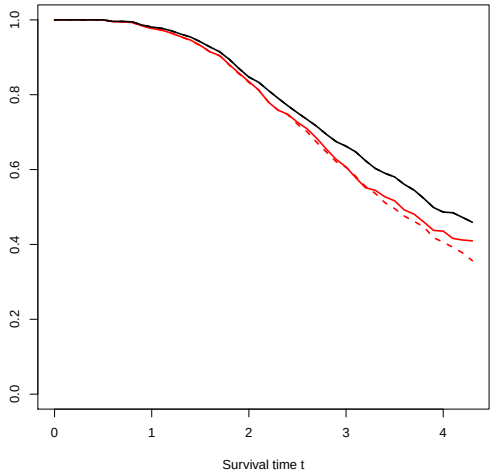
Figure 4.9: Comparison of $\hat{\Lambda}_{12}(0,t)$ and $\Lambda_{12}(0,t)$ for $n = 200, 500, 1000$, where the black line is $\Lambda_{12}(0,t)$, the red line is $\hat{\Lambda}_{12}(0,t)$, the blue lines are the 95% confidence interval when $T = \infty$, the orange lines are the 95% confidence interval when $T = 5$, the green line indicates the largest uncensored observation when $T = \infty$, the pink line indicates the largest uncensored observation when $T = 5$.



(a) $n = 200$, censoring rate $= \frac{103}{200}$ when $T = \infty$, censoring rate $= \frac{123}{200}$ when $T = 5$.



(b) $n = 500$, censoring rate $= \frac{227}{500}$ when $T = \infty$, censoring rate $= \frac{285}{500}$ when $T = 5$.



(c) $n = 1000$, censoring rate $= \frac{489}{1000}$ when $T = \infty$, censoring rate $= \frac{608}{1000}$ when $T = 5$.

Figure 4.10: Comparison of $\hat{S}(t) = 1 - \hat{P}_{02}(0, t)$ and the Kaplan-Meier estimator when $T = \infty$ and $T = 5$ for $n = 200$, $n = 500$ and $n = 1000$. The black solid line is the Kaplan-Meier when $T = 5$, the black dashed line is the Kaplan-Meier estimator when $T = \infty$, the red solid line is $\hat{S}(t)$ when $T = 5$, the red dashed line is $\hat{S}(t)$ when $T = \infty$.

sample size increases. However, formal statistical hypothesis testing is still necessary to theoretically measure the goodness of fit of our Markov model.

4.3.1 Kolmogorov-Smirnov Test

We take Y_{01} as an example to show the hypothesis testing procedure. We derived an estimator $\hat{\Lambda}_{01}(0,t)$ of $\Lambda_{01}(0,t)$. We can further derive an estimator of the distribution function of Y_{01}: $\hat{F}_{01}(t) = 1 - \exp(-\hat{\Lambda}_{01}(0,t))$. Denote the distribution function of Y_{01} in our model by $F_{01,0}$ (in the simulation, we use $gamma(2.6, 0.8)$). Our null and alternative hypotheses are as follows:

$$H_0 : F_{01} = F_{01,0} \quad vs. \quad H_a : F_{01} \neq F_{01,0}$$

The Kolmogorov-Smirnov (K-S) test statistic measures the supreme distance between the empirical distribution function $F_n(t)$ of a complete sample of $i.i.d$ random variables and the hypothetical distribution function $F_0(t)$. The Kolmogorov-Smirnov statistic is D_n defined by

$$D_n = \sup_t |F_n(t) - F_0(t)|. \tag{4.7}$$

Using D_n we would reject H_0 if the computed D_n value, say d_n, is larger than the

critical value c_α. However, the distribution of D_n is only known and independent of F_0 for complete samples. Various modifications of the K-S test statistic are available in the literature for carrying out hypothesis testing of H_0 *vs.* H_a with right censored data. See review in Nikulin, Lemeshko, Chimitova and Tsivinskaya (2011) [41]. To this end, Nikulin, *et al.*. (2011) [41] proposes a test statistic in their equation (1). In terms of our hypothesis test, the modified test statistics D_n^* is given as follows.

$$\begin{aligned} D_n^* &= \sup_{t>0} |\hat{F}_{01}(t) - F_{01,0}(t)|, \\ &= \sup_{k:\ \delta_{01,k}=1} |\hat{F}_{01}(Y_{0,k}) - F_{01,0}(Y_{0,k})|. \end{aligned}$$

The asymptotic distribution of D_n^* involves the amount of censoring in the data. However, the asymptotic distribution of D_n^* under H_0 is unknown and its approximated distribution is obtained by simulation. Based on simulated distribution, the cutoff value c_α and the P-value of the modified test are determined. We adapt the method in Nikulin *et al.*. (2011) [41] to simulated the distribution of D_n^*.

We would reject H_0 if D_n^* is greater than the critical value c_α. We simulate 5000 samples of (Y_0, δ_0) of a given sample size n, for $n = 200, 500$ and 1000. The test statistic D_n^* is calculated for each sample. Then we can use the empirical distribution of the simulated values of D_n^* to approximate the true distribution function of D_n^*

and calculate the P-value of the test.

Apply the modified K-S test to each of the following hypothesis testing problem:

$$H_0 : F_{0j} = F_{0j,0} \quad vs. \quad H_a : F_{0j} \neq F_{0j,0}, \quad j = 1, 2, 3,$$

$$H_0 : F_{1j} = F_{1j,0} \quad vs. \quad H_a : F_{1j} \neq F_{1j,0} \quad j = 2, 3.$$

The test results are summarized in Table 4.1. The censoring rate is the ratio of the number of censored observations to the total number of observations.

		F_{01}	F_{02}	F_{03}	F_{12}	F_{13}
$n = 200$	D_n^*	0.0794	0.1060	0.1091	0.0975	0.1170
	P-value	0.6521	0.5206	0.5157	0.6477	0.6972
	censoring rate	112/200	162/200	126/200	34/88	54/88
$n = 500$	D_n^*	0.0615	0.0683	0.1009	0.0670	0.1203
	P-value	0.7639	0.5598	0.5301	0.9566	0.5933
	censoring rate	262/500	371/500	367/500	103/238	135/238
$n = 1000$	D_n^*	0.0354	0.0434	0.0495	0.0828	0.1020
	P-value	0.9240	0.7164	0.7601	0.7166	0.8004
	censoring rate	545/1000	747/1000	708/1000	290/455	345/455

Table 4.1: Modified Kolmogorv-Smirnov test for $n = 200$, $n = 500$ and $n = 1000$.

Note that all the P-values are greater than 0.1. If we set the significance level $\alpha < 0.1$, the null hypothesis of each F_{ij} is not rejected.

4.3.2 Chi-squared Test

There are problems using a chi-squared statistic, say, $W(\theta)$, for a goodness of fit test. In the complete sample theory, if the unknown parameter θ is estimated by the minimum chi-squared method or some modified minimum chi-squared method, then as the sample size increases to infinity, the asymptotic distribution of $W(\hat{\theta})$ is a chi-squared distribution. According to a Chernoff-Lehmann theorem (1954), if the estimate $\hat{\theta}$ is obtained by the MLE, then $W(\hat{\theta})$ will not have an asymptotic chi-squared distribution. Furthermore, under right censoring the asymptotic distribution of a chi-squared type of test statistic depends on the censoring distribution [41]. We finesse the difficulties by constructing a Pearson's chi-squared statistic for checking the goodness of fit of model F_{ij} as follows.

There are four mutually exclusive transition paths in the irreversible model. We can divide the patients into four categories according to their transition paths. Let n_i be the number of observed patients in category i, and E_i be the expected number of observations in this category (calculated under the model distributions). The chi-squared test statistic is $W = \sum_{i=1}^{4}(n_i - E_i)^2/E_i$. Under the null hypothesis that $F_{0j} = F_{0j,0}$ for $j = 1, 2, 3$ and $F_{1j} = F_{1j,0}$ for $j = 2, 3$, the test statistic W has an asymptotic chi-squared distribution with 3 degrees of freedom.

The probabilities of the four categories p_1, p_2, p_3 and p_4 are calculated as

follows:

(1) $S_0 \to S_1 \to S_2$

This path implies that $Y_{01} = \min(Y_{01}, Y_{02}, Y_{03})$, $Y_{12} = \min(Y_{12}, Y_{13})$. By the independence of each Y_{ij},

$$\begin{aligned} p_1 &= P(Y_{01} = \min(Y_{01}, Y_{02}, Y_{03}))P(Y_{12} = \min(Y_{12}, Y_{13})) \\ &= \int_0^\infty S_{02}(u)S_{03}(u)dF_{01}(u) \int_0^\infty S_{13}(u)dF_{12}(u) \\ &= \int_0^\infty S_0(u)d\Lambda_{01}(u) \int_0^\infty S_1(u)d\Lambda_{12}(u). \end{aligned}$$

(2) $S_0 \to S_1 \to S_3$

This path implies that $Y_{01} = \min(Y_{01}, Y_{02}, Y_{03})$, $Y_{13} = \min(Y_{12}, Y_{13})$.

$$\begin{aligned} p_2 &= P(Y_{01} = \min(Y_{01}, Y_{02}, Y_{03}))P(Y_{13} = \min(Y_{12}, Y_{13})) \\ &= \int_0^\infty S_0(u)d\Lambda_{01}(u) \int_0^\infty S_1(u)d\Lambda_{13}(u). \end{aligned}$$

(3) $S_0 \to S_2$

This path implies that $Y_{02} = \min(Y_{01}, Y_{02}, Y_{03})$.

$$\begin{aligned} p_3 &= P(Y_{02} = \min(Y_{01}, Y_{02}, Y_{03})) \\ &= \int_0^\infty S_0(u) d\Lambda_{02}(u). \end{aligned}$$

(4) $S_0 \to S_3$

This path implies that $Y_{03} = \min(Y_{01}, Y_{02}, Y_{03})$.

$$\begin{aligned} p_4 &= P(Y_{03} = \min(Y_{01}, Y_{02}, Y_{03})) \\ &= \int_0^\infty S_0(u) d\Lambda_{03}(u). \end{aligned}$$

The expected number of observations in each category under the null hypothesis is $E_i = np_i$.

The test results (test statistic, P-values) are summarized in Table 4.2.

Sample size	$n = 200$	$n = 500$	$n = 1000$
W	2.05	1.42	0.14
P-value	0.5604	0.7019	0.9864

Table 4.2: Chi-squared test for $n = 200$, $n = 500$ and $n = 1000$.

All the P-values are greater than 0.1. If we set the significance level $\alpha < 0.1$,

the null hypothesis of each F_{ij} is not rejected. In Table 4.2, the P-values increase as the sample size increases. However, the P-values are random, if we repeat the simulation and perform the hypothesis testing again, we will get different P-values. For each sample size, we repeat the chi-squared test for 100 times and calculate the average of the P-values. Corresponding to $n = 200, 500$ and 1000, the average P-values are 0.4724, 0.4301 and 0.4638 respectively which indicates that the P-values are stable.

Chapter 5: Reversible Markov Multi-state Model

5.1 Introduction

In Chapter 3, we use an irreversible Markov model (Figure 3.1) in which a patient's transitions from relapse to treatment ($S_1 \rightarrow S_0$) are not allowed. In this chapter, we relax this restriction by allowing a relapsed patient to have the possibility of receiving a second IST treatment. This means that a patient's transition from S_0 to S_1 is reversible.

5.2 Reversible Markov Model

Similar to the model described in Chapter 3, consider a four-state stochastic process $\{\xi^o(t), t \geq 0\}$ where $\xi^o(t)$ is the status of a patient at time t. The four states are the same as before:

S_0: a patient is in remission immediately after IST

S_1: a patient relapses after remission

S_2: a patient is dead

S_3: a patient is lost to followup

Transition from S_1 to S_0 is allowed. The transition paths are described in Figure 5.1.

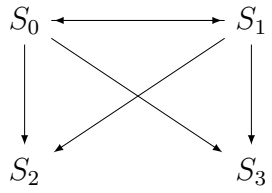


Figure 5.1: Possible transition paths of a patient during the clinical trial.

Assume n individuals are participating in the clinical trial. Each of them will take one and only one of the following possible mutually exclusive transition paths:

(1) $S_0 \to S_2$

(2) $S_0 \to S_3$

(3) $S_0 \to S_1 \to S_0 \to \ldots \to S_2$

(4) $S_0 \to S_1 \to S_0 \to \ldots \to S_3$

(5) $S_0 \to S_1 \to S_0 \to S_1 \to \ldots \to S_2$

(6) $S_0 \to S_1 \to S_0 \to S_1 \to \ldots \to S_3$

Patients on transition path (3) and (4) enter S_2 or S_3 directly from state 0, while patients on transition path (5) and (6) enter S_2 or S_3 directly from S_1. We use the same notations as in Chapter 3 except for the random variable Y_1, which represents the amount of time spent in S_1 before moving to the next state. In the reversible model, transition from S_1 to S_0 is allowed. Therefore, $Y_1 = \min(Y_{10}, Y_{12}, Y_{13})$, where Y_{10} is the random variable representing the sojourn time in S_1 before moving to S_0. The observations are (Y_0, δ_0) and (Y_1, δ_1). Employing the same method in Section 3.5, we can derive the joint distribution of (Y_0, δ_0) and (Y_1, δ_1), and the formulas of the cumulative hazard functions Λ_{1j} for $i = 0, 1, j = 1, 2, 3$. Here we present the results without derivation:

$$\Lambda_{0j}(0,t) = \int_0^t \frac{1}{S_0(u)} dG_0(u,j), \tag{5.1}$$

$$\Lambda_{1j}(0,t) = \int_0^t \frac{1}{S_1(u)} dG_1(u,j), \tag{5.2}$$

where $G_0(t,j)$ (see (3.8)) and $G_1(t,j)$ (see (3.14)) are the joint distribution functions of (Y_0, δ_0) and (Y_1, δ_1) respectively, $j = 1, 2, 3$, S_0 and S_1 are the survival functions of Y_0 and Y_1.

5.3 Nonparametric Estimators

5.3.1 Estimators of Cumulative Hazard Functions

Suppose our observed samples are as follows:

$$(Y_{0,k}, \delta_{0,k}), \quad k = 1, 2, \ldots, n,$$

$$(Y_{1,k}, \delta_{1,k}), \quad k = 1, 2, \ldots, N_{01}.$$

On the basis of (3.28) and (3.30), for $j = 1, 2, 3$, estimators of $q_{0j}(t)$ and $q_{1j}(t)$, $\Lambda_{0j}(0,t)$ and $\Lambda_{1j}(0,t)$ can be constructed as:

$$\hat{q}_{0j}(t) = \frac{d\hat{G}_0(t,j)}{\hat{S}_0(t)dt}, \quad t \in [0, t_{0j}], \tag{5.3}$$

$$\begin{aligned} \hat{\Lambda}_{0j}(0,t) &= \int_0^t \frac{1}{\hat{S}_0(u)} d\hat{G}_0(t,j) \\ &= \sum_{k=1}^{n} \frac{I(Y_{0,k} \le t, \delta_{0,k} = j)}{\sum_{m=1}^{n} I(Y_{0,m} > Y_{0,k})}, \quad t \in [0, t_{0j}], \end{aligned} \tag{5.4}$$

$$\hat{q}_{1j}(t) = \frac{d\hat{G}_1(t,j)}{\hat{S}_1(t)dt}, \quad t \in [0, t_{1j}], \tag{5.5}$$

$$\begin{aligned}\hat{\Lambda}_{1j}(0,t) &= \int_0^t \frac{1}{\hat{S}_1(u)} d\hat{G}_1(t,j) \\ &= \sum_{k=1}^{N_{01}} \frac{I(Y_{1,k} \le t, \delta_{1,k} = j)}{\sum_{m=1}^{N_{01}} I(Y_{1,m} > Y_{1,k})}, \quad t \in [0, t_{1j}].\end{aligned} \tag{5.6}$$

5.3.2 Estimators of Crude Transition Probabilities

When the hazard functions $q_{ij}(t)$ are time-dependent under some conditions, Feller (1940) [22] proved the existence and uniqueness of the solution to the Kolmogorov forward equations. However, it is difficult to find an explicit form for the solution with reversible Markov process that has time-dependent transition rates. Except for some special cases (e.g., Yang and Chang (1990) [51]), we rely on numerical methods to solve the transition probabilities.

Corresponding to the transition paths in Figure 5.1, the infinitesimal matrix $\boldsymbol{Q}^o(t)$, $t > 0$, of our reversible Markov model $\boldsymbol{\xi}^o$ is:

$$\mathbf{Q}^o(t) = \begin{pmatrix} q_{00}(t) & q_{01}(t) & q_{02}(t) & q_{03}(t) \\ q_{10}(t) & q_{11}(t) & q_{12}(t) & q_{13}(t) \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \tag{5.7}$$

Note that $\mathbf{Q}^o(t)$ differs from that in equation (3.31) by a positive risk $q_{10}(t)$. The

associated Kolmogorov forward equations are as follows:

$$\frac{dP^o_{00}(0,t)}{dt} = P^o_{00}(0,t)q_{00}(t) + P^o_{01}(0,t)q_{10}(t), \tag{5.8}$$

$$\frac{dP^o_{01}(0,t)}{dt} = P^o_{00}(0,t)q_{01}(t) + P^o_{01}(0,t)q_{11}(t), \tag{5.9}$$

$$\frac{dP^o_{02}(0,t)}{dt} = P^o_{00}(0,t)q_{02}(t) + P^o_{01}(0,t)q_{12}(t), \tag{5.10}$$

$$\frac{dP^o_{03}(0,t)}{dt} = P^o_{00}(0,t)q_{03}(t) + P^o_{01}(0,t)q_{13}(t), \tag{5.11}$$

$$\frac{dP^o_{10}(0,t)}{dt} = P^o_{10}(0,t)q_{00}(t) + P^o_{11}(0,t)q_{10}(t), \tag{5.12}$$

$$\frac{dP^o_{11}(0,t)}{dt} = P^o_{10}(0,t)q_{01}(t) + P^o_{11}(0,t)q_{11}(t), \tag{5.13}$$

$$\frac{dP^o_{12}(0,t)}{dt} = P^o_{10}(0,t)q_{02}(t) + P^o_{11}(0,t)q_{12}(t), \tag{5.14}$$

$$\frac{dP^o_{13}(0,t)}{dt} = P^o_{10}(0,t)q_{03}(t) + P^o_{11}(0,t)q_{13}(t). \tag{5.15}$$

Equations (5.8) and (5.9) only involve the transition probabilities P^o_{00} and P^o_{01}, therefore form a subsystem. Once we solve (5.8) and (5.9), (5.10) and (5.11) can be solved directly. The subsystem formed by equation (5.8) and (5.9) can be written in a ma-

trix form:

$$\begin{aligned} y'(t) &= A(t)y(t) \\ &= f(t, y(t)), \quad y(0) = (1, 0)^T \end{aligned} \tag{5.16}$$

where

$$y(t) = (P^o_{00}(0, t), P^o_{01}(0, t))^T, \tag{5.17}$$

$$A(t) = \begin{pmatrix} q_{00}(t) & q_{10}(t) \\ q_{01}(t) & q_{11}(t) \end{pmatrix}. \tag{5.18}$$

Note that the elements of the coefficient matrix A are time-dependent. Therefore we can not solve (5.16) by eigenvalue decomposition of matrix A. We proceed with a numerical method. We use Euler's method which is a basic explicit method for numerical integration of ordinary differential equations. Let h be the step size and $t_n = nh$. Then $y'(t_n)$ can be approximated by

$$\frac{y(t_{n+1}) - y(t_n)}{h} = f(t, y(t_n)),$$

which gives an approximation:

$$y(t_{n+1}) = y(t_n) + hf(t, y(t_n)). \tag{5.19}$$

Applying equation (5.19) together with the initial value $y(0) = (1, 0)^T$, we can obtain an approximation of $y(t)$ at each time t_n. We can choose a smaller step size h to get a more precise approximation.

5.3.3 Estimator of Net Survival Function

Similar to Section 3.5.3, the net survival function of a patient is:

$$\begin{aligned} S(t) &= P(X > t \mid \xi(0) = 0) \\ &= 1 - P_{02}(0, t), \end{aligned}$$

where $\boldsymbol{\xi} = \{\xi(t) : t \geq 0\}$ is a 3-state Markov process ($\{S_0, S_1, S_2\}$) with the same transition rates $q_{01}(t)$, $q_{02}(t)$, $q_{10}(t)$ and $q_{12}(t)$ as in $\{\xi^0(t) : t \geq 0\}$. The infinitesimal matrix of $\boldsymbol{\xi}$ is

$$\boldsymbol{Q}^*(t) = \begin{pmatrix} q^*_{00}(t) & q_{01}(t) & q_{02}(t) \\ q_{10}(t) & q^*_{11}(t) & q_{12}(t) \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \tag{5.20}$$

where

$$q_{00}^*(t) = -q_{01}(t) - q_{02}(t),$$

$$q_{11}^*(t) = -q_{10}(t) - q_{12}(t).$$

To derive the transition probability $P_{02}(0,t)$ in $\boldsymbol{\xi}$, we need the following Kolmogorov forward equations:

$$\begin{aligned}
\frac{dP_{00}(0,t)}{dt} &= P_{00}(0,t)q_{00}^*(t) + P_{01}(0,t)q_{10}(t), \\
\frac{dP_{01}(0,t)}{dt} &= P_{00}(0,t)q_{01}(t) + P_{01}(0,t)q_{11}^*(t), \\
\frac{dP_{02}(0,t)}{dt} &= P_{00}(0,t)q_{02}(t) + P_{01}(0,t)q_{12}(t).
\end{aligned}$$

We solve the system of equations for the transition probabilities numerically and obtain an estimator of $P_{02}(0,t)$, denoted as $\hat{P}_{02}(0,t)$. Then an estimator of the net survival function $S(t)$ is:

$$\hat{S}(t) = 1 - \hat{P}_{02}(0,t). \tag{5.21}$$

5.4 Clinical Trial Terminating at a Finite Time T

In Section 5.2 and 5.3, the estimation is carried out under the assumption that the clinical trial terminates when all the patients are dead or lost to followup, where there is no termination date for the clinical trial. In this section, similar to Section 4.3, we will simulate the transition paths under the assumption that the clinical trial terminates at a prespecified finite time T.

For each patient in state 0, we observe a pair of random variables (Y_0, δ_0), for each patient in state 1, we observe a pair of random variables (Y_1, δ_1). However, unlike the irreversible model, in the reversible model, the patients are allowed to visit state 0 and state 1, recurrently. In this case, the distributions of the censoring variable Y_{03} and Y_{13} depend on the number of times the patient has visited state 0 and state 1 respectively. For example, a patient has transition path $0 \to 1 \to 0 \to 1 \to 2$. So initially his/her observation period from state 0 is $[0, T]$. When he/she visits state 0 for the second time, the remaining observation period from state 0 is $[Y_{01} + Y_{10}, T - Y_{01} - Y_{10}]$ given $Y_{01} + Y_{10} < T$. Similarly, when a patient first visits state 1, the remaining observation period from state 1 is $[Y_{01}, T - Y_{01}]$ given $Y_{01} < T$, when the patient visits state 1 for the second time, the remaining observation period from state 1 is $[Y_{01} + Y_{10} + Y_{01}, T - Y_{01} - Y_{10} - Y_{01}]$ given $Y_{01} + Y_{10} + Y_{01} < T$.

Thus the length of observation period is random and the distribution depends on the number of times the patient has visited state 0 or state 1. Therefore when estimating $\Lambda_{ij}(0,t)$, we only use the observations (Y_0, δ_0) in state 0 after the first treatment and the observations (Y_1, δ_1) in state 1 after the first relapse. In terms of observable random variables, we have:

$$
\begin{aligned}
Y_0 &= \min(Y_{01}, Y_{02}, Y_{03}, T), \quad Y_0 \in (0, T], \\
\{\delta_0 = j\} &= \{Y_0 = Y_{0j}\}, \quad j = 1, 2, \\
\{\delta_0 = 3\} &= \{Y_0 = \min(Y_{03}, T)\}. \\
Y_1 &= \min(Y_{10}, Y_{12}, Y_{13}, T - Y_{01} | Y_{01} < T), \quad Y_1 \in (0, T], \\
\{\delta_1 = j\} &= \{Y_1 = Y_{1j}\}, \quad j = 0, 2, \\
\{\delta_1 = 3\} &= \{Y_1 = \min(Y_{13}, T - Y_{01})\}.
\end{aligned}
$$

The joint distribution of (Y_0, δ_0) is the same as that given in the irreversible case (see (3.44)). But the joint distribution of (Y_1, δ_1) is different:

$$
\begin{aligned}
G_1(t, j) &= P(Y_1 \le t, \delta_1 = j) \\
&= \int_0^t S_1(u) q_{1j}(u) du, \quad 0 < t \le T, \ j = 0, 2,
\end{aligned}
$$

$$G_1(t,3) = P(Y_1 \le t, \delta_1 = 3)$$
$$= \int_0^t S_1(u) q_{13}^*(u) du, \quad 0 < t \le T,$$

where

$$S_1(t) = P(Y_1 > t)$$
$$= S_{10}(t) S_{12}(t) S_{13}^*(t), \quad 0 < t \le T,$$
$$S_{13}^*(t) = S_{13}(t) F_{01}(T-t)/F_{01}(T), \quad 0 < t \le T,$$

and $q_{13}^*(t)$ is the hazard rate function corresponding to the survival function $S_{13}^*(t)$. The estimators for $\Lambda_{ij}(t), i, j = 0, 1, 2, 3$ can be derived in the same way as in Section 3.5.1. We display the estimators without derivation:

$$\hat{\Lambda}_{0j}(0,t) = \sum_{k=1}^{n} \frac{I(Y_{0,k} \le t, \delta_{0,k} = j)}{\sum_{m=1}^{n} I(Y_{0,m} > Y_{0,k})}, \quad t \in [0, t_{0j}], \quad j = 1, 2, 3, \tag{5.22}$$

$$\hat{\Lambda}_{1j}(0,t) = \sum_{k=1}^{N_{01}} \frac{I(Y_{1,k} \le t, \delta_{1,k} = j)}{\sum_{m=1}^{N_{01}} I(Y_{1,m} > Y_{1,k})}, \quad t \in [0, t_{1j}], \quad j = 0, 2, 3. \tag{5.23}$$

5.5 Asymptotic Distributions of the Estimators

The estimators of $\Lambda_{ij}(0,t)$ in this chapter are the same as those in Chapter 3. The uniform consistency, unbiasedness, estimated variance and asymptotic distributions of $\hat{\Lambda}_{0j}(0,t)$ for $j=1,2,3$ and $\hat{\Lambda}_{1j}(0,t)$ for $j=0,2,3$ can be proved in the same way as in Section 3.7.

The transition probabilities are calculated numerically. Therefore we do not have a closed form for $\hat{S}(t)=1-\hat{P}_{02}(0,t)$. The asymptotic distribution of $\hat{S}(t)$ can not be derived from the asymptotic distributions of $\hat{\Lambda}_{ij}(0,t)$. However, we can still approximate the asymptotic distribution of $\hat{S}(t)$ using simulation.

5.6 Simulation for Reversible Markov Model

Simulations for the reversible model can be carried out similarly to that of the irreversible model in Chapter 4. The simulation procedures are as follows:

(1) Generate the sojourn time in state 0, $Y_0=y_0$, from the distribution function $F_0(t)=1-S_0(t)$ using inverse transform sampling.

(2) Take a number δ_0 from the set $\{1,2,3\}$ with trinomial probabilities $\Big(-q_{01}(y_0)/q_{00}(y_0), -q_{02}(y_0)/q_{00}(y_0), -q_{03}(y_0)/q_{00}(y_0)\Big)$.

(3) If $\delta_0 = 2$ or $\delta_0 = 3$, stop. The observed transition path is $0 \to 2$ or $0 \to 3$, the observed time of transition is y_0. Otherwise continue to step (4).

(4) If $\delta_0 = 1$

- Generate the sojourn time in state 1, $Y_1 = y_1$, from the distribution function $F_1(t) = 1 - S_1(t)$, where $S_1(t) = S_{10}(t)S_{12}(t)S_{13}(t)$.
- Take a number δ_1 from the set $\{0, 2, 3\}$ with trinomial probabilities $\Big(-q_{10}(y_1)/q_{11}(y_1), -q_{12}(y_1)/q_{11}(y_1), -q_{13}(y_1)/q_{11}(y_1)\Big)$, where $q_{11}(t) = -q_{10}(t) - q_{12}(t) - q_{13}(t)$.

(5) If $\delta_1 = 2$ or $\delta_1 = 3$, stop. Otherwise go to step (1).

Repeat the procedure for each of the n patients to get a sample of n independent Markov processes, where n is the prespecified sample size. As in Section 4.1, gamma and Weibull distributions are used to generate the Y_{ij} random variables. Additionally, $gamma(3.6, 1)$ is used to generate the random variable Y_{10}. The simulations are carried out using sample sizes 200, 500 and 1000.

5.6.1 Estimation of the Cumulative Hazard Functions with Simulated Date

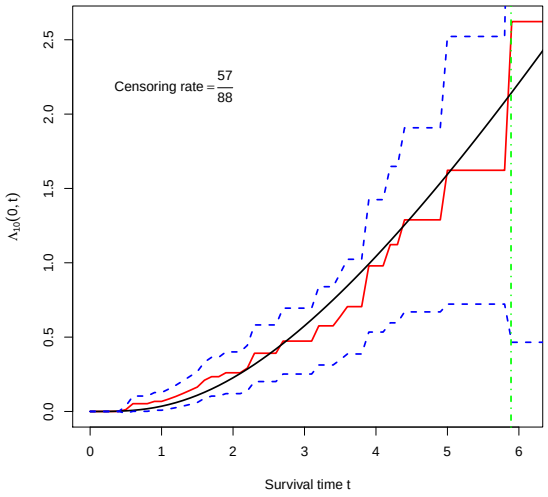
We plot $\hat{\Lambda}_{ij}(0,t)$ against $\Lambda_{ij}(0,t)$. For illustration, we display only the graphs of $\Lambda_{10}(0,t)$ and $\Lambda_{03}(0,t)$ and their estimates in Figures 5.2 and 5.3, where $\Lambda_{10}(0,t)$ is the cumulative hazard function of $gamma(3.6,1.0)$ and $\Lambda_{03}(0,t)$ is the cumulative hazard function of $Weibull(1.5,5)$.

5.6.2 Estimator of the Net Survival Function with Simulated Data

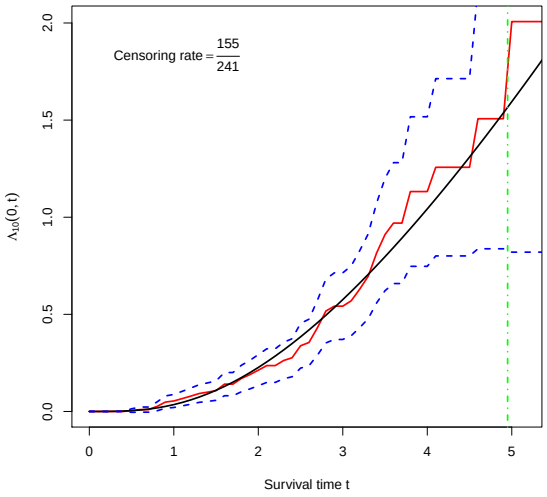
The transition probabilities are solved from the Kolmogorov equations numerically. We do not display the graphs for each $\hat{P}_{ij}(0,t)$ separately. We only show the graph for $\hat{S}(t)$ and compare it with the Kaplan-Meier estimator for $n = 200$, $n = 500$ and $n = 1000$ in Figure 5.4.

Figure 5.4 shows that the Kaplan-Meier estimator is larger than $\hat{S}(t)$ for all $t > 0$ and for all sample sizes. The conclusion is consistent with that in the irreversible model in Chapter 4.

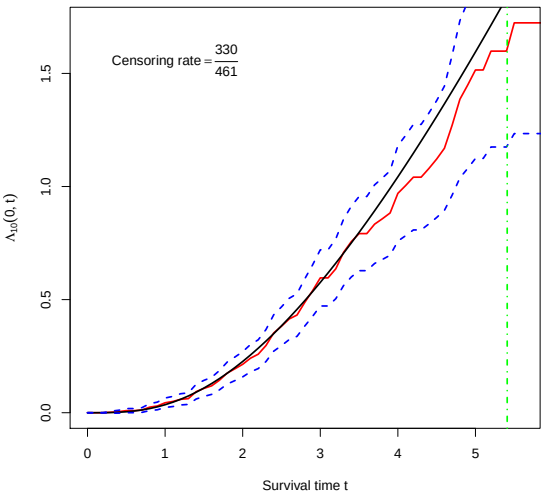
$$\Lambda_{10}(t) = -\log\left(1 - \int_0^t [\beta^\alpha x^{\alpha-1}\exp(-\beta x)]dx/\Gamma(\alpha)\right),\ t > 0,\ \alpha = 3.6, \beta = 1.0$$



(a) $n = 200$



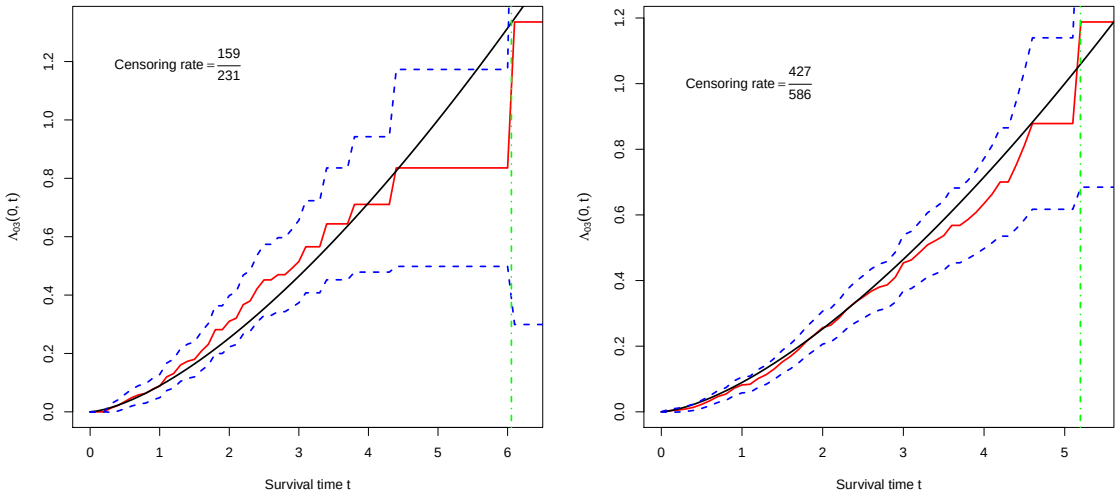
(b) $n = 500$



(c) $n = 1000$

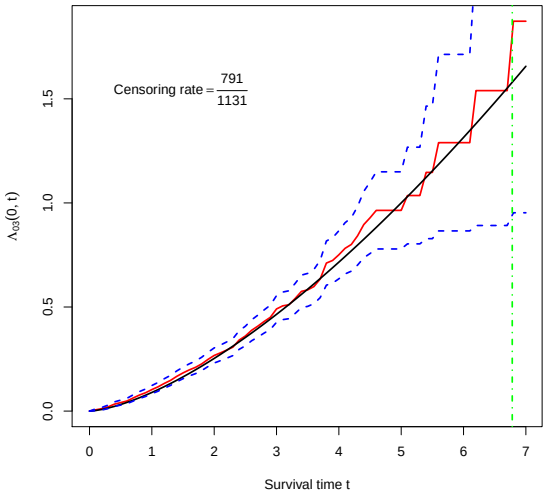
Figure 5.2: Comparison of $\hat{\Lambda}_{10}(0,t)$ and $\Lambda_{10}(0,t)$ for $n = 200, 500, 1000$, where the black line is $\Lambda_{10}(0,t)$, the red line is $\hat{\Lambda}_{10}(0,t)$, the blue line is the 95% confidence interval, and the green line indicates the largest uncensored observation: $\max(y_{1,k}|\delta_{1,k} = 0)$.

$$\Lambda_{03}(t) = \left(\frac{t}{\eta}\right)^{\gamma},\ t > 0,\ \gamma = 1.5, \eta = 5$$



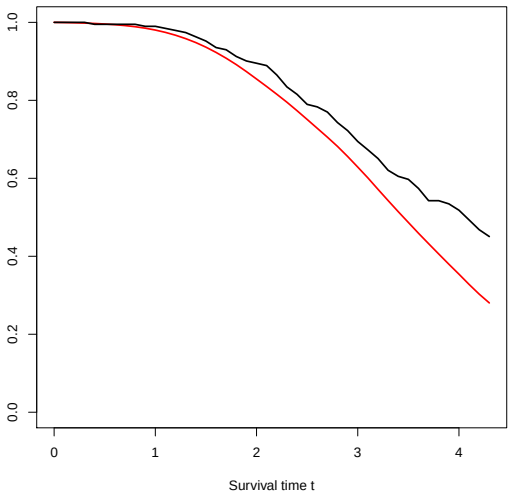
(a) $n = 200$

(b) $n = 500$

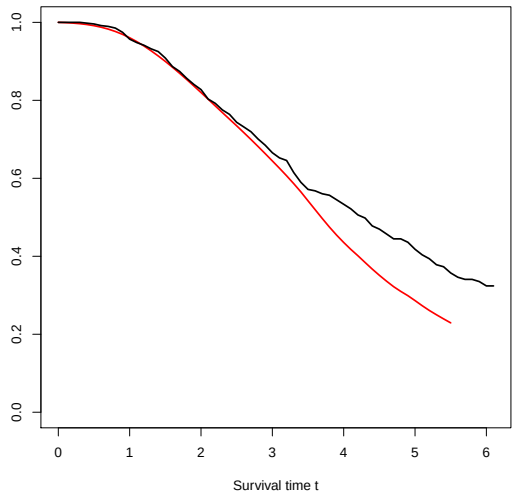


(c) $n = 1000$

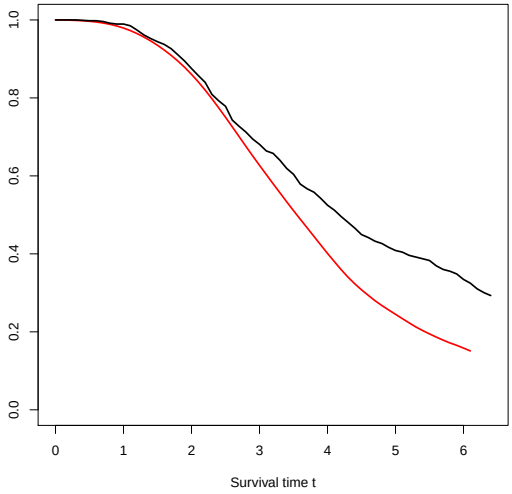
Figure 5.3: Comparison of $\hat{\Lambda}_{03}(0,t)$ and $\Lambda_{03}(0,t)$ for $n = 200, 500, 1000$, where the black line is $\Lambda_{03}(0,t)$, the red line is $\hat{\Lambda}_{03}(0,t)$, the blue line is the 95% confidence interval, the green line indicates the largest uncensored observation: $\max(y_{0,k}|\delta_{0,k} = 3)$.



(a) $n = 200$, censoring rate $= \frac{89}{200}$



(b) $n = 500$, censoring rate $= \frac{240}{500}$



(c) $n = 1000$, censoring rate $= \frac{496}{1000}$

Figure 5.4: Comparison of $\hat{S}(t) = 1 - \hat{P}_{02}(0,t)$ and the Kaplan-Meier estimator for $n = 200$, $n = 500$ and $n = 1000$. The black line represents the Kaplan-Meier estimator, the red line represents $\hat{S}(t)$.

5.6.3 Clinical Trial Terminating at a Finite Time T

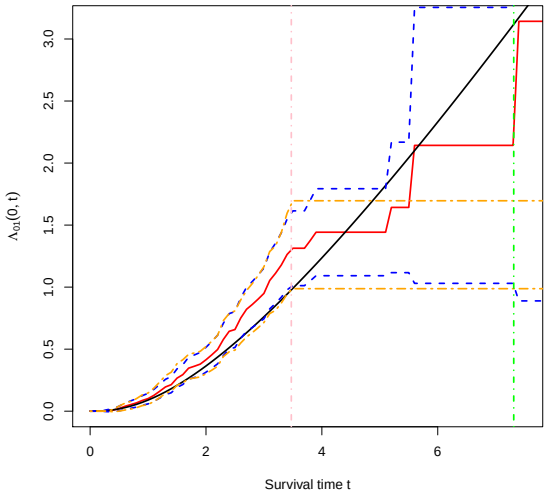
In this section, we carry out the simulation for a clinical trial that terminates at a prespecified finite time T.

In the reversible model with $T < \infty$, both $\hat{\Lambda}_{0j}(0,t), j = 1,2,3$ and $\hat{\Lambda}_{1j}(0,t), j = 0,2,3$ have larger variance compared to that when $T = \infty$ (see Figure 5.2 and 5.3). Because the sample size used to estimate $\Lambda_{ij}(0,t)$ is smaller when $T < \infty$. For illustration, we display the graphs of $\hat{\Lambda}_{01}(0,t)$ and $\hat{\Lambda}_{12}(0,t)$ in Figure 5.5 and 5.6. We use $T = 5$ in the simulation.

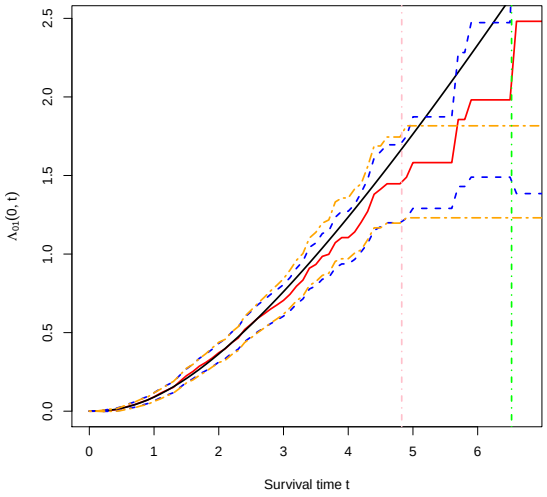
The transition probabilities $P_{00}(0,t)$, $P_{01}(0,t)$, $P_{11}(0,t)$ and $P_{02}(0,t)$ are calculated numerically. Then an estimator of the survival function $S(t)$ follows. We only display the graph of $\hat{S}(t)$ and compare it with the Kaplan-Meier estimator for $n = 200$, $n = 500$ and $n = 1000$ in Figure 5.7.

From the Figure 5.7, we can see that the Kaplan-Meier estimator of the survival probability is higher than $\hat{S}(t)$ both when $T = \infty$ and $T = 5$. It indicates that the Kaplan-Meier estimator may overestimate the survival probability because it does not use the data on relapse. The conclusion is consistent with that in the irreversible model.

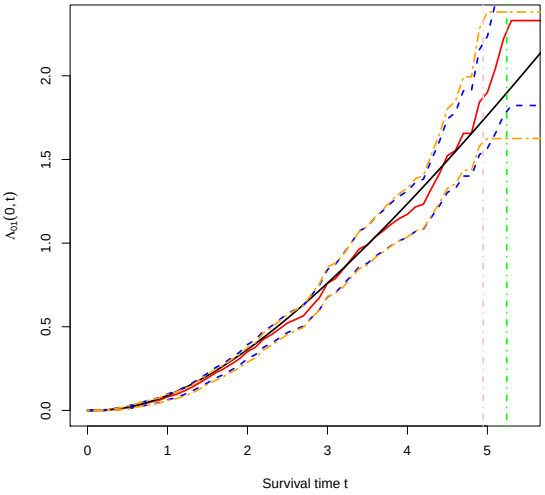
$$\Lambda_{01}(0,t) = -log\Big(1 - \int_0^t [\beta^\alpha x^{\alpha-1} exp(-\beta x)]dx/\Gamma(\alpha)\Big),\ t > 0,\ \alpha = 2.6, \beta = 0.8$$



(a) $n = 200$, censoring rate $= \frac{114}{200}$



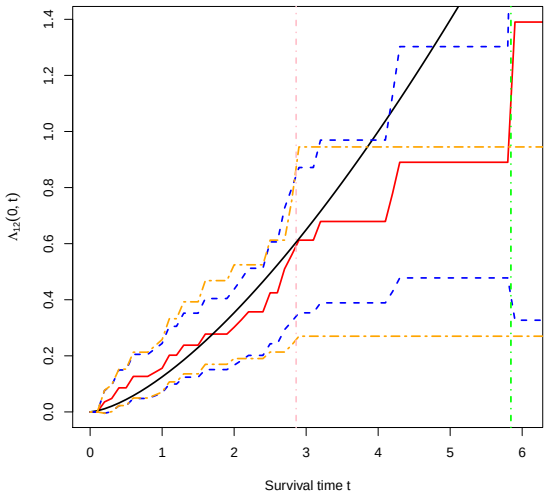
(b) $n = 500$, censoring rate $= \frac{248}{500}$



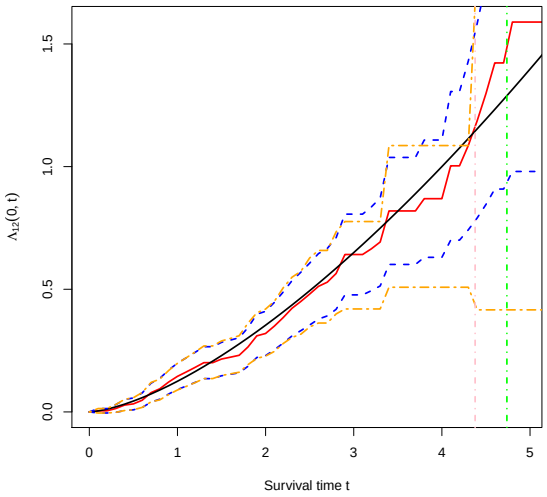
(c) $n = 1000$, censoring rate $= \frac{534}{1000}$

Figure 5.5: Comparison of $\hat{\Lambda}_{01}(0,t)$ and $\Lambda_{01}(0,t)$ for $n = 200, 500, 1000$. The black line is $\Lambda_{01}(0,t)$, the red line is $\hat{\Lambda}_{01}(0,t)$, the blue lines are the 95% confidence interval when $T = \infty$, the orange lines are the 95% confidence interval when $T = 5$, the green line indicates the largest uncensored observation when $T = \infty$, the pink line indicates the largest uncensored observation when $T = 5$.

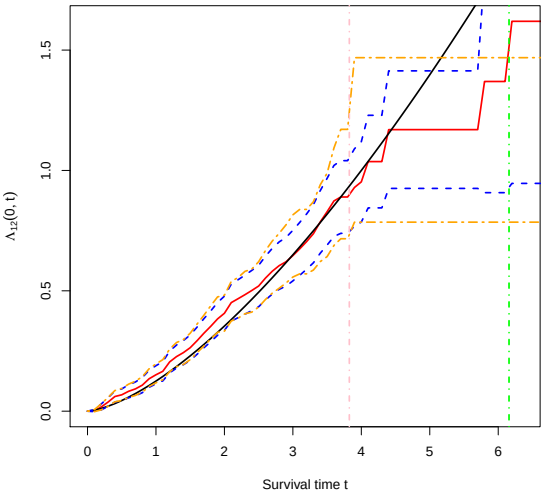
$$\Lambda_{12}(0,t) = \left(\frac{t}{\eta}\right)^{\gamma},\ t > 0,\ \gamma = 1.5, \eta = 4$$



(a) $n = 200$, censoring rate $= \frac{62}{86}$

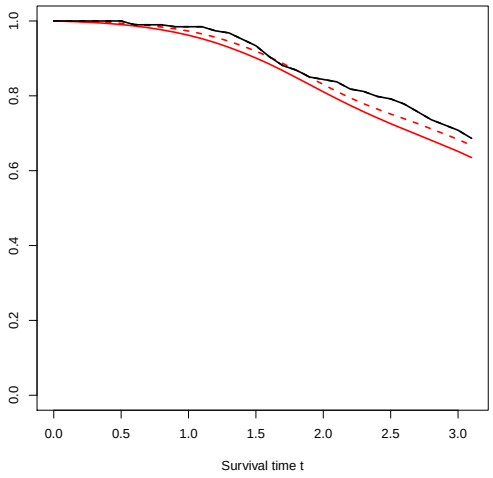


(b) $n = 500$, censoring rate $= \frac{168}{252}$

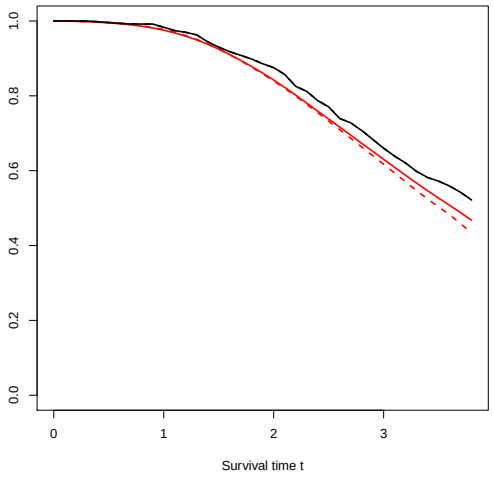


(c) $n = 1000$, censoring rate $= \frac{313}{466}$

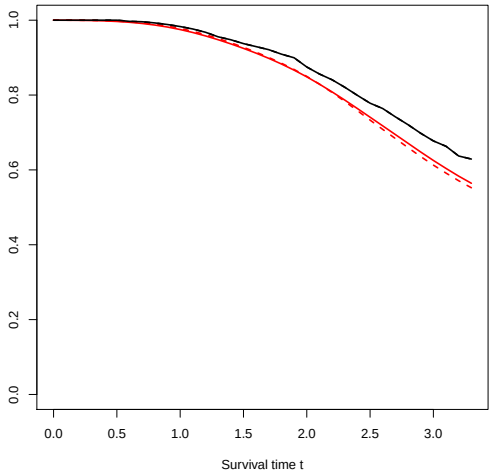
Figure 5.6: Comparison of $\hat{\Lambda}_{12}(0,t)$ and $\Lambda_{12}(0,t)$ for $n = 200, 500, 1000$. The black line is $\Lambda_{12}(0,t)$, the red line is $\hat{\Lambda}_{12}(0,t)$, the blue lines are the 95% confidence interval when $T = \infty$, the orange lines are the 95% confidence interval when $T = 5$, the green line indicates the largest uncensored observation when $T = \infty$, the pink line indicates the largest uncensored observation when $T = 5$.



(a) $n = 200$, censoring rate $= \frac{89}{200}$ when $T = \infty$, censoring rate $= \frac{112}{200}$ when $T = 5$



(b) $n = 500$, censoring rate $= \frac{240}{500}$ when $T = \infty$, censoring rate $= \frac{306}{500}$ when $T = 5$



(c) $n = 1000$, censoring rate $= \frac{496}{1000}$ when $T = \infty$, censoring rate $= \frac{609}{1000}$ when $T = 5$

Figure 5.7: Comparison of $\hat{S}(t) = 1 - \hat{P}_{02}(0t)$ and the Kaplan-Meier estimator when $T = \infty$ and $T = 5$ for $n = 200$, $n = 500$ and $n = 1000$. The black solid line is the Kaplan-Meier estimator when $T = 5$, the black dashed line is the Kaplan-Meier estimator when $T = \infty$, the red solid line is $\hat{S}(t)$ when $T = 5$, the red dashed line is $\hat{S}(t)$ when $T = \infty$.

5.7 Hypothesis Testing

5.7.1 Modified Kolmogorov-Smirnov Test

The hypothesis tests are carried out in a similar way to that in Section 4.3.1 for irreversible model. We only show the testing results as given in Table 5.1.

		F_{01}	F_{02}	F_{03}	F_{10}	F_{12}	F_{13}
$n = 200$	D_n^*	0.0755	0.0805	0.0873	0.1020	0.0797	0.1817
	P-value	0.6722	0.6642	0.6132	0.1828	0.5458	0.6943
	censoring rate	127/228	168/228	161/228	58/86	55/86	59/86
$n = 500$	D_n^*	0.0504	0.1031	0.0754	0.0763	0.0826	0.1083
	P-value	0.6823	0.3610	0.8382	0.4310	0.4489	0.7639
	censoring rate	325/570	429/570	386/570	148/218	129/218	159/218
$n = 1000$	D_n^*	0.0366	0.0566	0.0411	0.0507	0.0449	0.0965
	P-value	0.8932	0.8686	0.8920	0.5904	0.8434	0.8934
	censoring rate	616/1159	871/1159	831/1159	297/456	283/456	332/456

Table 5.1: Modified Kolmogorv-Smirnov test for $n = 200$, $n = 500$ and $n = 1000$.

Note that all the P-values are greater than 0.1. If we set the significance level $\alpha < 0.1$, the null hypothesis of each F_{ij} is not rejected.

5.7.2 Chi-squared Test

The chi-squared test is performed a little different from that in Chapter 4 for irreversible model. We use the same test statistic $W = \sum_{i=1}^{m}(n_i - E_i)^2/E_i$, but the choices of categories are different. In the irreversible model, there are four possible transition paths so we consider each transition path as one category. In the reversible model, the number of possible transition paths is infinite. We can combine different transition paths to form a single category. For example, we can divide the observations into the following four categories: (A) transition paths of length two ending in S_2 (death), (B) transition paths of length two endings in S_3 (loss to followup), (C) transition paths of length three, (D) transition paths of length not less than four. The probabilities of falling into each of the four categories are:

(1) $S_0 \to S_2$

$$\begin{aligned} p_1 &= P(\text{falling into (A)}) \\ &= P(Y_{02} = \min(Y_{01}, Y_{02}, Y_{03})) \\ &= \int_0^\infty S_0(u) d\Lambda_{02}(u). \end{aligned}$$

(2) $S_0 \to S_3$

$$\begin{aligned} p_2 &= P(\text{falling into (B)}) \\ &= P(Y_{03} = \min(Y_{01}, Y_{02}, Y_{03})) \\ &= \int_0^\infty S_0(u) d\Lambda_{03}(u). \end{aligned}$$

(3) $S_0 \to S_1 \to S_3$ or $S_0 \to S_1 \to S_2$

$$\begin{aligned} p_3 &= P(\text{falling into (C)}) \\ &= P(Y_{01} = \min(Y_{01}, Y_{02}, Y_{03}))\Big(1 - P(Y_{10} = \min(Y_{10}, Y_{12}, Y_{13}))\Big) \\ &= \int_0^\infty S_0(u) d\Lambda_{01}(u)\Big(1 - \int_0^\infty S_1(u) d\Lambda_{10}(u)\Big), \end{aligned}$$

where $S_1(t) = S_{10}(t)S_{12}(t)S_{13}(t)$.

(4) The others.

$$\begin{aligned} p_4 &= P(\text{faling into (D)}) \\ &= 1 - p_1 - p_2 - p_3. \end{aligned}$$

The test results (test statistic and P-values) are summarized in Table 5.2. All

Sample size	$n = 200$	$n = 500$	$n = 1000$
W	3.00	1.59	0.62
P-value	0.2690	0.2435	0.8931

Table 5.2: Chi-squared test for $n = 200$, $n = 500$ and $n = 1000$.

the P-values are greater than 0.1. Therefore if we set the significance level $\alpha < 0.1$, the null hypothesis of each F_{ij} is not rejected. In Table 5.2, as the sample size increases, the P-values increase in general. However, the P-values are random. If we repeat the simulation and perform the hypothesis testing again, we will get different P-values. We repeat the simulations for 100 times and for each sample size, we calculate the mean of 100 the P-values. Corresponding to $n = 200, 500$ and 1000, the average P-values are 0.4615, 0.5223 and 0.4964 respectively which indicates that the P-values are stable.

Chapter 6: Conclusion

In a fundamental paper, Fix and Neyman (1951) [23] introduced a method for estimating the survival function with right-censored survival time data collected from clinical trials. The method takes into account of the available data on each patient's status with regard to relapse, recovery, loss to follow-up and death. A patient's survival time is measured from a well-defined starting time such as the time of diagnosis to the time of death or loss to followup and it includes the sojourn times that the patient spends in various states. Fix and Neyman constructed an absorbing Markov process to model the observable data on the disease evolution of a patient during a clinical trial. Taking into consideration of right-censored Markov model, Fix and Neyman introduced a procedure to calculate the model survival function as well as its estimator. The Fix-Neyman model is a homogeneous Markov process, i.e., transition rates are independent of time. In this thesis, we generalize the Fix-Neyman method for nonhomogeneous Markov models, i.e., the transition rates are both time and state dependent. From which one can estimate the survival

function nonparametrically. This problem is raised in Yang (2013) [50] and becomes the topic of this thesis. Although finite nonhomogeneous Markov model for lifetime analyses has been used widely in more recent years, Fix-Neyman's method is hardly used except mostly in Chiang's work (e.g., Chiang (1968) [14]) and by his students.

In this thesis, the model survival function is calculated by solving a system of Kolmogorov forward equations. Dictated by the availability of closed form solutions, the Markov models for statistical analyses are investigated separately for the irreversible and reversible models. Analytical solutions for the survival function can be derived for irreversible models. We use the 4-state Markov model with two transient and two absorbing states to demonstrate the derivations and estimate the cumulative hazard functions and survival function. In the reversible models, except in some special cases, analytical solutions are not available. In this thesis, numerical solutions are developed for statistical analyses.

The generalization of the Fix-Neyman model includes a number of existing multi-state models for right-censored lifetime data as special cases, such as the multiple decrement models, the illness and death models, the staging models, and the classical right censoring model. The optimality of the celebrated Kaplan-Meier nonparametric estimator of the survival function is based on the classical right censoring model. Thus the generalization of the Fix-Neyman model and their procedure unify

the estimation of the survival function with right-censored data. In general, the Fix-Neyman model utilizes more of the available data in the clinical trial than that of the classical right-censoring model which disregards patient's status data in the clinical trial.

Extensive simulations of the Markov models for both irreversible and reversible cases are carried out to examine the accuracy and distributional properties of the estimators. Using the Markovian property, simulated data are obtained by simulating a sequence of competing risks models at each transition time. With simulated data, a comparison is made between the estimated survival function under a particular 4-state model and the Kaplan-Meier estimator under a 3-state model (which ignores the status data). The comparison shows that the Kaplan-Meier estimator has larger values for all times and in all cases of comparisons.

Future work includes studying of power functions of various goodness of fit tests, and determining under what conditions would the Kaplan-Meier estimator be uniformly larger than a nonparametric estimator of the survival function in a 4-state Markov model.

Bibliography

[1] Aalen, O. (1975). *Statistical Inference for a Family of Counting Processes*. PhD thesis, University of California, Berkeley.

[2] Aalen, O. (1978). Nonparametric estimation of partial transition probabilities in multiple decrement models. *Ann. Stat.*, **6**:534–545.

[3] Aalen, O. (1989). A linear regression model for the analysis of life times. *Stat. Med.*, **8**(8):907–925.

[4] Aalen, O. and Gjessing, H. (2001). Understanding the shape of the hazard rate: a process point of view. *Stat. Sci.*, **16**(1):1–22.

[5] Aalen, O. and Johansen, S. (1978). An empirical transition matrix for non-

homogeneous Markov chains based on censored observations. *Scand. J. Stat.*, **5**:141–150.

[6] Andersen, P. K. (1997). *Multi-state Models for Event History Analysis in Clinical Medicine and Epidemiology*. University of Copenhagen.

[7] Andersen, P. K., Borgan, O., Gill, R. D., and Keiding, N. (2012). *Statistical Models Based on Counting Processes*. Springer Science & Business Media.

[8] Andersen, P. K. and Gill, R. (1982). Cox's regression model for counting processes: A large sample study. *Ann. Stat.*, **10**(4):1100–1120.

[9] Andersen, P. K. and Keiding, N. (2002). Multi-state models for event history analysis. *Stat. Methods Med. Res.*, **11**(2):91–115.

[10] Arias, E. (2012). United states life tables 2008. *National Vital Statistics Reports*, **61**(3).

[11] Beck, G. (1979). Stochastic survival models with competing risks and covariates. *Biometrics*, **35**:427–438.

[12] Beck, G. J. and Chiang, C. L. (1981). On maximum likelihood solutions for exponential survival models. *Biom. J.*, **23**(5):451–459.

[13] Chiang, C. L. (1964). A stochastic model of competing risks of illness and competing risks of death. *Stoch. Models Med. Biol.*, pages 323–354.

[14] Chiang, C. L. (1968). *Introduction to Stochastic Processes in Biostatistics*. John Wiley & Sons, New York.

[15] Chiang, C. L. (1980). *An Introduction to Stochastic Processes and Their Applications*. RE Krieger Publishing Company New York.

[16] Collett, D. (2015). *Modelling Survival Data in Medical Research*. Chapman & Hall/CRC.

[17] Cortese, G. and Andersen, P. K. (2010). Competing risks and time-dependent covariates. *Biom. J.*, **52**(1):138–158.

[18] Cox, D. R. (1972). Regression models and life-tables. *J. Royal Stat. Society-Series B.ens*, **34**(2):187–202.

[19] Datta, S. and Satten, G. (2001). Validity of the aalen–johansen estimators of stage occupation probabilities and nelson–aalen estimators of integrated transition hazards for non-Markov models. *Stat. Probab. Lett.*, **55**(4):403–411.

[20] Efromovich, S. and Chu, J. (2018). Small LTRC samples and lower

bounds in hazard rate estimation. *Ann. Inst. Stat. Math.* DOI: 10.1007/s10463-017-0617-x.

[21] Fang, J. Q. (1985). *Multi-state Survival Analysis with Time-dependent Covariates and Censoring*. PhD thesis, University of California, Berkeley.

[22] Feller, W. (1940). On the integro-differential equations of purely discontinuous Markoff processes. *Trans. Amer. Math. Soc.*, **48**:488–515.

[23] Fix, E. and Neyman, J. (1951). A simple stochastic model of recovery, relapse, death and loss of patients. *Hum. Biol.*, **23**(3):205–241.

[24] Freund, J. (1961). A bivariate extension of the exponential distribution. *J. Am. Stat. Assoc.*, **56**(296):971–977.

[25] Frydman, H. (1992). A non-parametric estimation procedure for a periodically observed three-state Markov process, with application to AIDS. *J. Royal Stat. Society-Series B.ens*, **54**:853–866.

[26] Frydman, H. (1995). Nonparametric estimation of a Markov 'illness-death' process from interval-censored observations, with application to diabetes survival data. *Biometrika*, **82**:773–789.

[27] Frydman, H. and Szarek, M. (2009). Nonparametric estimation in a Markov

"illness–death" process from interval censored observations with missing intermediate transition status. *Biometrics*, **65**(1):143–151.

[28] Glidden, D. (2002). Robust inference for event probabilities with non-Markov event data. *Biometrics*, **58**(2):361–368.

[29] Group, P. S. (1991). Prophylaxis of first hemorrhage from esophageal varices by sclerotherapy, propranolol or both in cirrhotic patients: A randomized multicenter trial. *Hepatology*, **14**(6):1016–1024.

[30] Huang, J. and Liu, L. (2006). Polynomial spline estimation and inference of proportional hazards regression models with flexible relative risk form. *Biometrics*, **62**(3):793–802.

[31] Kalbfleisch, J. D. and Prentice, R. L. (2002). *The Statistical Analysis of Failure Time Data*. John Wiley & Sons.

[32] Kaplan, E. and Meier, P. (1958). Nonparametric estimation from incomplete observations. *J. Am. Stat. Assoc.*, **53**(282):457–481.

[33] Kvam, P. H. and Samaniego, F. J. (1997). Multivariate life testing in variably scaled environments. *Lifetime. Data. Anal.*, **3**(4):337–351.

[34] Le Cam, L. and Yang, G. (2000). *Asymptotics in Statistics: Some Basic Concepts*. Springer, 2 edition.

[35] Marshall, A. and Olkin, I. (1967). A generalized bivariate exponential distribution. *J. Appl. Probab.*, **4**(2):291–302.

[36] Meira-Machado, L., de Uña-Álvarez, J., and Cadarso-Suarez, C. (2006). Nonparametric estimation of transition probabilities in a non-Markov illness-death model. *Lifetime Data Anal.*, **12**(3):325–344.

[37] Meira-Machado, L., de Uña-Álvarez, J., Cadarso-Suarez, C., and Andersen, P. K. (2009). Multi-state models for the analysis of time-to-event data. *Stat. Methods Med. Res.*, **18**(2):195–222.

[38] Miloslavsky, M., Keles, S., van der Laan, M., and Butler, S. (2004). Recurrent events analysis in the presence of time-dependent covariates and dependent censoring. *J. Royal Stat. Society-Series B.ens*, **66**(1):239–257.

[39] Moeschberger, M. L. (1974). Life tests under dependent competing causes of failure. *Technometrics*, **16**(1):39–47.

[40] Neyman, J. (1950). *First Course in Probability and Statistics*. New York: Holt, Rinehart and Winston.

[41] Nikulin, M., Lemeshko, B., Chimitova, E., and Tsivinskaya, A. (2011). Non-parametric goodness-of-fit tests for censored data.

[42] Pérez-OcónC, R., Ruiz-Castro, J., and Gámiz-Pérez, M. (2001). A piecewise Markov process for analysing survival from breast cancer in different risk groups. *Stat. Med.*, **20**(1):109–122.

[43] Prentice, R., Kalbfleisch, J., Peterson, J., Arthur, V., Flournoy, N., Farewell, V., and Breslow, N. (1978). The analysis of failure times in the presence of competing risks. *Biometrics*, **34**:541–554.

[44] Prentice, R., Williams, B., and Peterson, A. (1981). On the regression analysis of multivariate failure time data. *Biometrika*, **68**(2):373–379.

[45] Putter, H. and Spitoni, C. (2018). Non-parametric estimation of transition probabilities in non-Markov multi-state models: The landmark aalen–johansen estimator. *Stat. Methods Med. Res.*, **27**(7):2081–2092.

[46] Singer, R., Strauss, D., and Shavelle, R. (1998). Comparative mortality in cerebral palsy patients in california, 1980-1986. *J. Insurance Med., New York*, **30**(4):240–246.

[47] Sloand, E., Wu, C., Greenberg, P., Young, N., and Barrett, J. (2008). Fac-

tors affecting response and survival in patients with myelodysplasia treated with immunosuppressive therapy. *J. Clin. Oncol.*, **26**(15):2505.

[48] Therneau, T. and Grambsch, P. (2013). *Modeling Survival Data: Extending the Cox Model*. Springer Science & Business Media.

[49] Tsiatis, A. (1975). A nonidentifiability aspect of the problem of competing risks. *Proc. Natl. Acad. Sci.*, **72**(1):20–22.

[50] Yang, G. (2013). Neyman, Markov processes and survival analysis. In *Risk Assessment and Evaluation of Predictions*, pages 67–86. Springer.

[51] Yang, G. and Chang, M. (1990). A stochastic model for analyzing prevalence surveys of hepatitis a antibody. *Math. Biosci.*, **98**(2):157–169.