Master thesis
Mathematical sciences

An estimator for state occupation probabilities in non-Markov multistate models

May 29, 2017

Author:
Paul Reijbroek
3672883

Supervisor:
Dr. Cristian Spitoni
Second reader:
Dr. Martin Bootsma

Preface

This thesis is the result of the final project for the master studies Mathematical Sciences at Utrecht University. I would first like to thank my supervisor dr. Cristian Spitoni, who introduced me to the topic of multistate models and survival analysis and who guided me very well throughout the project. His suggestions, feedback, ideas, explanations and questions were invaluable. I would like to thank dr. Martin Bootsma as well, for being the second reader for my thesis. I would also like to thank dr. Roberto Fernandez, who, after listening to my mathematical interests, suggested to contact dr. Spitoni to be the supervisor of my thesis.

As a result of this thesis, I am now able to comprehend mathematical literature much more comprehensively. I have learnt to study proofs with a critical mind, how to correct them and how to use the proofs and template of proofs to construct my own proofs. I have gained a much more comprehensive knowledge of multistate and survival models and the tools that are often used in the analysis of these models. As I worked through the material, this thesis became more and more enjoyable as I gained more knowledge on these subjects.

Wageningen, May 29, 2017

Abstract

In medical research, the progress of a disease can be modelled using multistate models. Quantities of interest are the transition hazard and the state occupation probabilities. In this thesis, we consider estimators of the integrated transition hazard and state occupation probabilities, with the possibility of right-censoring, in multistate models that are not necessarily Markov. We focus on deriving the Nelson-Aalen estimator and the Aalen-Johansen estimator, and show that these are consistent, by correcting the proofs in [1]. We work out the variance for the distribution of the latter estimator, and propose an estimator for this variance. The contribution of this manuscript is purely theoretical, without data simulations.

Contents

1 Introduction

In medical research, when a patient is diagnosed with a disease or when someone is requiring some form of treatment, he/she will want to know what to expect, what could happen and the probabilities of good and bad events happening to them. This information is useful for physicians as well, because they have to consider all kinds of treatment and whether the treatment that they are considering would be the most effective treatment.

When a patient is diagnosed with a possibly deadly disease, a question he/she often asks is the chances of survival, and if he/she survives, how it will affect his/her life. There are several other types of questions a patient might ask as well.

For example, when a patient is diagnosed with cancer, he/she will want to know if he/she has a chance of surviving and how good his/her chances are. But that is probably not all he/she wants to know. Suppose there are two treatments (A and B) available, but every individual might experience a relapse after either treatment. From relapse, one might go to the definitive state of death. Of course the risk of each treatment and the amount of discomfort will be of importance in choosing a treatment, but the probability of surviving a possible relapse (and the probability of relapse) will be a factor as well.

In such cases, there exist models that estimate the probability of survival (among probabilities of other events happening) in a Markov chain, and its properties have been studied (e.g. [4]). A Markov assumption is rather strong though, and quite often it is not realistic. When looking at the previous example, there are five states: the patient has been diagnosed with cancer ("diagnosis"), the patient during and after treatment A ("treatment A"), the patient during and after treatment B ("treatment B"), "relapse", and "death". Assuming the patient has received treatment A, the rate of transitioning to "relapse" may depend on the transition time between state "diagnosis" and state "treatment A" and on how long ago that transition was. There might be a higher transition rate to "relapse" when the transition time to "treatment A" happened much earlier. In that case, the transition rate to "relapse" doesn't only depend on the current state "treatment A", but also on how long the patient has been in state "treatment A". Similar things could happen in state "treatment B". Furthermore, when a patient is in the state "relapse", the transition rate to death may again depend on how long the patient has been in state "relapse", as well as the treatment the patient received prior to relapse and how long he/she has been in state "treatment A" or "treatment B". Therefore, the transition rate does not only depend on the current state, but it depends on more of the history of the patient.

As we can see, a Markov model for the progress of this disease is not an accurate model, so a model with weaker assumptions on the data is needed for a better prediction.

This thesis will start with a review on models that model different stages of a disease as states. This is called a multistate model. From there on, we use survival data to gather information about the estimated probability of transitioning into one state from another, since the actual probabilities of survival (or from getting into one state from another) for many diseases are unknown, which is why they have to be estimated. Predictions for future patients will be obtained by looking at the data from patients with the same disease who were treated similarly that have been studied until now. We will state the conditions that we work with throughout this thesis, and we will give an overview.

Multistate models

The progress of a disease can be described by a multistate model for each patient. A state is given by any stage of a disease a patient can be in. An event is any direct transition from one state into another, i.e. reaching one state from another state without being in other states inbetween.

The simplest multistate model is called the *survival model* (see figure 1.1) and consists of just two states (usually alive and death) and one transition (alive → death). It should be noted that these two states could be any states. In the survival setting, this model is used to just find the probability of survival without giving more information using available data. The quantity of interest is the survival time, i.e. the time until the transition from state 'alive' to state 'death' takes place.

The survival model is an example of a multistate Markov model. A model which is not necessarily Markov is the following model that we call the *illness-death model* (see figure 1.1): we have the following three states: healthy, ill, death, and the following possible transitions: healthy → ill, ill → healthy, healthy → death and ill → death.

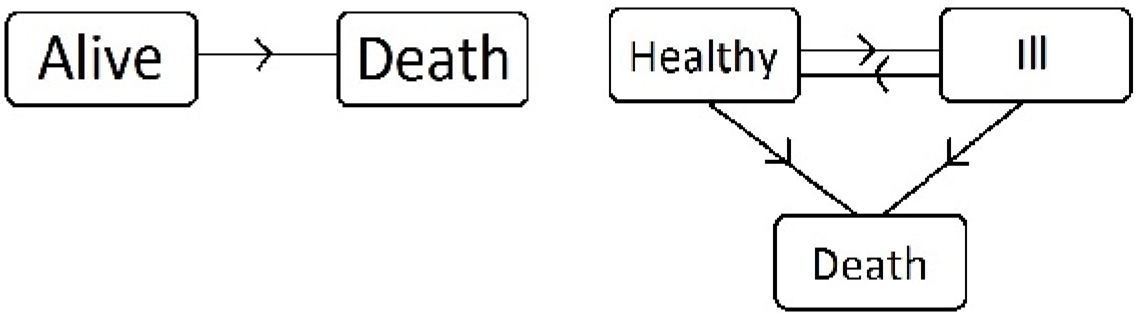

Figure 1.1: A survival model (left) and an illness-death model (right)

It is possible that the probability of death is much higher for a patient that has entered state 'ill' for a second time than one that does so for the first time. Furthermore, it may also depend on how long the patient has been in state ill, i.e. how long it has been since the last transition. Therefore, the survival probability (or the probability of being cured) depends on more than just the state the person is in at the time.

In general, a multistate model may have more than one starting point, more than one absorbing state and multiple intermediate states, and intermediate states can be starting states as well. The illness-death model is an example of a multistate model with one or two starting states, depending on whether or not every individual starts in state 'healthy', two intermediate states ('ill' and 'healthy') and one absorbing state ('death').

In general, the quantities of interest are the *stage occupation probabilities* (the probability of being in a state at a certain time) and the *transition probabilities* (the probability of transitioning into one state from another). In the following example, from Spitoni, Verduijn & Putter (2012) [10], which considers patients with kidney malfunctioning (see figure 1.2), there are five possible states: alive during or after first dialysis (state 1), alive during or after first kidney transplant (state 2), alive during or after second dialysis (state 3), alive during or after second transplant (state 4), and death (state 5).

The interesting quantities are, for each state, the probability of being in that state at a certain time, and the probability of getting into any state from another state before that one. Because getting into state 4 from state 1 is only possible via state 2 and 3, the probability of getting there is not a very mathematically interesting quantity, even though a patient may find it to be relevant (as it may affect their quality of life and because it may have an influence on their expectations as to what to expect). The reason these types of probabilities where there are no direct transitions are not very interesting to study on their own is the fact that the quantities can be found without too much difficulty by finding the transition probabilities of getting into state 4 from state 3, into state 3 from state 2 and into state 2 from state 1. Those quantities are quite interesting, as well as the probability for each state to go into state 5 (either directly or via another (couple of) state(s)).

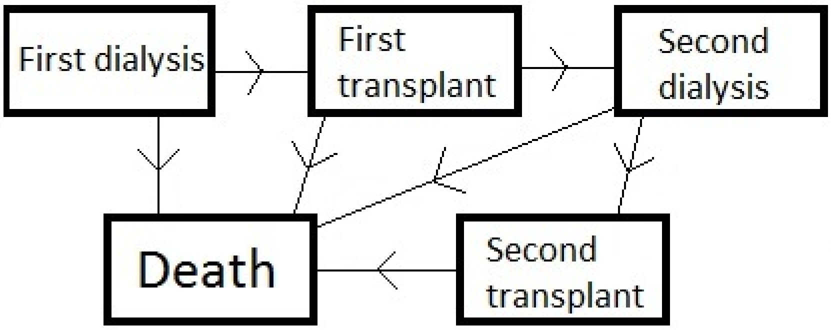

Figure 1.2: Multistate model for kidney malfunctioning

We can see that this model is not necessarily Markov. How long it has been since any procedure has been completed may influence the probability of death or the probability of requiring another procedure. It doesn't only depend on the state a patient currently occupies.

As we can see, even in rather simple models, a Markov assumption may be a bit of a stretch. A violation of this assumption could make calculations for occupation and transition probabilities a lot more difficult, though.

Predictions for state occupation and transition probabilities

Since the state occupation probabilities and transition probabilities are the quantities we are most interested in, these are the ones we will want to estimate. This is done with the help of survival analysis and survival data. Although it would be nice to give an estimation as to how long a patient will be in a state before they make the next transition, it is very difficult to give a good estimator for this. We limit ourselves to the aforementioned probabilities.

Unfortunately, data about patient might be incomplete, for many reasons. We discuss two types of *censoring* of data. The first one is right-censoring, which happens when a study ends while no event (i.e. transition) has happened to a patient, or when a patient leaves the study before it ends and before they enter an absorbing state. Even though we don't have complete data for these patients, we may still use the data, and we want to use as much of the available data as possible, without discarding incomplete observations.

The second one is left-censoring, which happens if we do not have information about a patient at the start of the study, that is, they had been in the initial state before the study started or because (an) event(s) has/have happened to them that we do not know about.

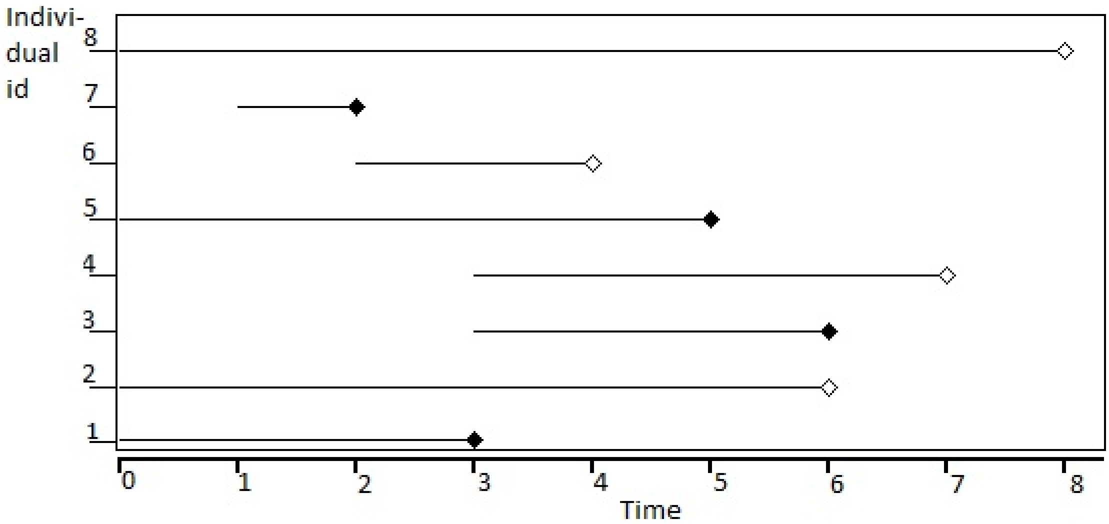

Figure 1.3: A lexis diagram for the possible censoring of individuals, where a closed dot indicates an event and an open dot indicates right-censoring

In figure 1.3, individuals 1, 2, 5 and 8 are possibly left-censored, as they may have been in a starting state before time 0, the beginning of the study, or because events already happened to them before time 0. Individuals 2, 4 and 6 are right-censored because they left the study before an event happened to them and before the study ended. Individual 8 is also right-censored, because nothing had happened to them when the study ended at time 8.

Patients can be censored for a variety of reasons. In many models, *non-informative censoring* is used, which means that the censoring does not depend on the events, transitions and transition times and vice versa. Non-informative censoring takes place if a patient decides to leave the study or is forced to leave the study for, for example, reasons that do not have anything to do with the disease or the treatment thereof. Censoring can also be *informative*, if patients decide to leave the study because, for example, he/she feels the treatment is not helping them or because he/she gets too ill to participate.

Informative censoring, although more realistic in several cases, will complicate analysis of the multistate model very much. For that reason, and because analysis of models with non-informative censoring is still quite useful, we will assume that the censoring in our models is non-informative. On top of that, we will only work with right-censoring, to be able to derive some very neat results.

Estimation of probabilities

As mentioned before, in survival models and multistate models, people are interested in the survival function (which gives the survival probability) in survival models and the transition hazard and the occupation probabilities in multistate models. In this thesis, we are interested in non-parametric estimators of these quantities, so an estimator that does not assume any kind of parametric form on the quantity that we want to estimate.

A non-parametric estimator for the survival function was given by Edward Kaplan & Paul Meier in 1958 [6]. This estimator is now known as the *Kaplan-Meier estimator*. We obtain the estimator by multiplying the estimated probabilities of survival on subintervals of the study period. The Kaplan-Meier estimator may take right-censoring into account. This estimator can be used to obtain different estimators for multistate models.

An estimator to obtain the probabilities of transitioning at any possible moment in time was given by Nelson & Aalen in 1969 [7], and was proved to be consistent for a Markov system. This estimator is now called the *Nelson-Aalen estimator*. The estimator consists of the sum of all transitions from one state to another, where each time such a transition is made, you divide that transition by the number of people at risk of such a transition. As Datta & Satten [1] seemingly proved in 2001, a version of the Nelson-Aalen estimator can be used to estimate the transition probabilities even if the quantities of the model are not necessarily Markov, and this estimator is still consistent. A main part of this thesis will be to prove consistency of this estimator, correcting the proof given in the paper published by Datta & Satten in 2001. This is found in chapter 4

From the Kaplan-Meier estimator and the Nelson-Aalen estimator for non-Markov systems, one may derive the *Aalen-Johansen estimator*, which can be seen as a generalised matrix version of the Kaplan-Meier estimator. This estimator is used to find the probabilities of being in any state at any moment of time. Aalen and Johansen proved consistency of this estimator in a Markov setting in 1978 [4]. Datta & Satten seemingly proved in 2001 that consistency holds in a non-Markov setting as well, but their proof seemed a bit rushed, and did not entirely derive the estimator when right-censoring was included. The derivation of the estimator and an extended proof of consistency for the estimator is given in chapter 5.

Aim of the thesis

In this thesis, we are interested in the theory regarding multistate models. We aim to improve Datta & Satten's proofs of consistency of the Nelson-Aalen estimator for cumulative transition hazards and the Aalen-Johansen estimator for state occupation probabilities in a non-Markov setting with independent right-censoring [1]. Furthermore, we want to derive a formula for the estimated variance of the Aalen-Johansen estimator, using just the quantities this model provides and quantities that can be derived from it.

Outline of the thesis

In chapter 2, we formalise the concept of survival analysis and give a formal definition of independent right-censoring. We introduce the estimators for which we will prove consistency in chapters 4 and 5. We will also recall some notions about Martingales, predictable processes and compensators.

In chapter 3, we discuss the multistate model. We introduce the product integral, the counting process and how they relate to Martingales, we derive the transition hazard, the censoring hazard and for both of these the cumulative distribution function. Furthermore, we give the definition of a Markov process and compare this to the definition of the transition hazard. Finally, we review the estimators we defined in chapter 2 and define them for multistate models.

Chapter 4 focusses on the Nelson-Aalen estimator for the transition hazard. We define the counting processes and at risk processes for our multistate model, we prove a lemma regarding expectations of censored and non-censored data, and we use the proof of this lemma to prove consistency of the Nelson-Aalen estimator for cumulative transition hazards under reasonable conditions.

In chapter 5, we derive the Aalen-Johansen estimator for the occupation probabilities. We use a balance equation for the case where we have uncensored data and use the Law of Large Numbers to prove why this is a consistent estimator. Next, we derive the same estimator in a similar way, although it requires more calculations, for right-censored data. We then use the results of chapter 4 to prove consistency of this estimator in a way that it similar to the proof of consistency for the Nelson-Aalen estimator in chapter 4

Information about the distribution of the derived Aalen-Johansen estimator is given in chapter 6. We prove consistency of the covariance function of the distribution of the estimator, and propose a formula for the variance of this estimator.

In the final chapter, we discuss a different estimator for the occupation probability and possible requirements for this estimator to be consistent. We also discuss the possibility to include more known data to acquire the Nelson-Aalen estimator, and what other requirements might be needed for this estimator to be consistent.

2 Survival analysis

Throughout the thesis, we assume we are working on a (general) probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Unless stated otherwise, random variables are defined on this probability space. Furthermore, we will often use so called *càdlàg processes*, which are processes that are right-continuous and have left-hand limits (continue à droite, limite à gauche, which means 'continuous on the right, limits on the left' in French).

In this chapter, we formalise the concept of survival analysis and non-informative right-censoring. We will show how to find the probability of survival without being censored. Furthermore, we introduce two estimators that will be useful later on for multistate models (see chapter 3). We end this chapter by recalling notions on Martingales, predictable processes and compensators and some interesting properties they have.

As has been stated in the introduction, we will assume, throughout this thesis, that any censoring in our models is non-informative right-censoring, i.e. it does not depend on the state an individual is in and on transition times.

2.1 Survival analysis without censoring

If our study consists of individuals with only complete data, we can introduce a random variable, the survival time T^*, that gives us all the information we need. Assuming i.i.d. individuals, as we do, estimating the probability of surviving up to time t is very easy. A consistent estimator for the survival probability $S(t) = \mathbb{P}(T^* > t)$ would be

$$\hat{S}(t) := \frac{\#\{\text{people who have survived up to time } t\}}{\#\{\text{people in the study at time } 0\}} \tag{2.1.1}$$

and this estimator gives us all the information we could want for this model. In many studies, this is not a realistic situation, because very often there are people who leave the study before it is finished and before they make the transition. In the following section, we will discuss the censoring hazard and censoring probability, as well as the survival hazard.

2.2 Survival hazard and censoring hazard

For simple survival models with right-censoring, we introduce two random variables: the survival time T^* and the censoring time C. Quantities of interest will then be $T := \min\{T^*, C\}$ and $\delta := I(C > T^*)$, the indicator whether or not the individual was ever censored. Assuming no censoring, we may define the survival hazard α by

$$\alpha(t) := \lim_{\Delta t \downarrow 0} \frac{\mathbb{P}(T^* \in [t, t + \Delta t) | T^* \geq t)}{\Delta t} \tag{2.2.1}$$

i.e. the rate of death at time t given that the individual has survived up to time t. Then $A(t) := \int_0^t \alpha(s)\mathbf{d}s$ is the cumulative hazard.

A similar approach can be used to acquire the cumulative censoring hazard, as well as the probability of not having been censored while having survived up to time t. The censoring hazard is given by

$$\lambda(t) := \lim_{\Delta t \downarrow 0} \frac{\mathbb{P}(C \in [t, t + \Delta t), \delta = 0 | T \geq t)}{\Delta t} \tag{2.2.2}$$

With the cumulative censoring hazard $\Lambda(t) = \int_0^t \lambda(s)\mathbf{d}s$.

We now assume censoring may take place anywhere on an interval $[0, \tau)$. For any $t \in (0, \tau)$, we take a partition $0 = t_0 < t_1 < \ldots < t_m = t$ of the interval $[0, t]$. Let $\Delta\Lambda(t_i) := \Lambda(t_i) - \Lambda(T_{i-1})$. We find

$$\begin{aligned}
&\mathbb{P}(C > t, \delta = 0) = \lim_{\max|t_i - t_{i-1}| \to 0} \mathbb{P}(C > t_m, \delta = 0 | C > t_{m-1}) \cdot \\
&\mathbb{P}(C > t_{m-1}, \delta = 0 | C > t_{m-2}) \cdot \ldots \cdot \mathbb{P}(C > t_1, \delta = 0 | C > t_0)\mathbb{P}(C > t_0) \\
&= \lim_{\max|t_i - t_{i-1}| \to 0} (1 - \mathbb{P}(t_{m-1} < C \leq t_m, \delta = 0 | C > t_{m-1})) \cdot \\
&(1 - \mathbb{P}(t_{m-2} < C \leq t_{m-1}, \delta = 0 | C > t_{m-2})) \cdot \ldots \cdot (1 - \mathbb{P}(0 < C \leq t_1, \delta = 0 | C > t_0)) \cdot 1 \\
&= \lim_{\max|t_i - t_{i-1}| \to 0} (1 - \Delta\Lambda(t_m)) \cdot (1 - \Delta\Lambda(t_{m-1})) \cdot \ldots \cdot (1 - \Delta\Lambda(t_1)) \\
&= \lim_{\max|t_i - t_{i-1}| \to 0} \prod_{i=1}^{m} (1 - \Delta\Lambda(t_i))
\end{aligned} \tag{2.2.3}$$

This limit of a product leads to the *product integral*, which is further explained in section 3.3. The representation of the censoring function as a product integral will be useful in later computations.

2.3 Non-parametric estimators

In many cases the survival hazard is unknown and has to be estimated. In case we have a uniform population (which will be assumed from here on out. If they all share a certain quality, it still remains a uniform population), we may use non-parametric estimation to find the cumulative survival hazard.

2.3.1 The Nelson-Aalen estimator

Let $t_1 < \ldots < t_m \leq t$ denote all distinct event times up to time t, let n_j be the number of people experiencing some event at time t_j, and let y_j be the number of people at risk of an event at time t_j-. The *Nelson-Aalen* estimator is then given by

$$\hat{A}(t) := \sum_{j:t_j \leq t} \frac{n_j}{y_j} \tag{2.3.1}$$

In a multistate model (see chapter 3), we will use a similar estimator for the transition hazard for all possible transitions. With a couple of restrictions, we will actually prove convergence in probability of each estimated transition hazard to the actual hazard, uniformly on $[0, t]$ (where $t \in (0, \tau)$, with τ defined as in section 4.2 of the Nelson-Aalen estimator in chapter 4) for these transition hazards in a multistate model. The Nelson-Aalen estimator for a multistate model is defined in section 3.6.

It should be noted that a similar estimator is available for the cumulative censoring hazard. Because we focus on the (cumulative) survival hazard and the occupation (or state or survival) probabilities, the details are omitted.

2.3.2 The Kaplan-Meier estimator

Having found the Nelson-Aalen estimator for the cumulative survival hazard, we may continue to get an estimator for the survival function. Let $\Delta A(t_j) = A(t_j) - A(t_{j-1})$ where $t_0 = 0$ and $0 < t_1 < \ldots < t_m \leq t$ are all the event times up to time t. In the same way we found a formula for the censoring function, we may find a formula with a product for the survival function, i.e.

$$S(t) := \mathbb{P}(T > t) = \lim_{\max|t_j - t_{j-1}| \to 0} \prod_{j=1}^{m} (1 - \Delta A(t_j)) \tag{2.3.2}$$

Since we have distinct event times t_j, it is easy to see that $\Delta \hat{A}(t_j) = \frac{n_j}{y_j}$. Plugging in $\hat{A}$ in (2.3.2) for A, we find the so called *Kaplan-Meier estimator*

$$\hat{S}(t) := \prod_{j:t_j \leq t} (1 - \frac{n_j}{y_j}) \tag{2.3.3}$$

In section 3.6, we will define a similar estimator for multistate models and see what they estimate. Convergence of that estimator is proved in chapter 5.

2.4 Martingales

A very important type of stochastic process that we will use is the *Martingale*. A Martingale is essentially a stochastic process that, given the value of the process at a certain time, remains the same in expectation.

Definition 2.4.1. *Let $M = \{M(t) : t \geq 0\}$ be a càdlàg stochastic process and let $\{\mathcal{F}_t : t \geq 0\}$ be a filtration, defined on a probability space. Then the process M is called a* Martingale *with respect to $\mathcal{F}_t$ (or an $\mathcal{F}_t$-Martingale) if it satisfies the following properties:*

1. *M is adapted to $\{\mathcal{F}_t : t \geq 0\}$*
2. *$\mathbb{E}|M(t)| < \infty$ for all $t < \infty$*
3. *$\mathbb{E}(M(t)|\mathcal{F}_s) = M(s)$ a.s. for all $t \geq s \geq 0$*

Replacing property 3 in definition 2.4.1 by $\mathbb{E}(M(t)|\mathcal{F}_s) \geq M(s)$ a.s. gives us a *submartingale*, and replacing (3) by $\mathbb{E}(M(t)|\mathcal{F}_s) \leq M(s)$ a.s. gives us a *supermartingale.*

It is important to note that, due to property 3, a Martingale M satisfies $\mathbb{E}(M(t + \Delta t) - M(t)|\mathcal{F}_s) = 0$ a.s..

A Martingale M is called *square integrable* if $\mathbb{E}(M(t)^2) < \infty$ for all $t < \infty$. Also, if M is a square integrable Martingale, then $M^2 = \{M(t)^2 : t \geq 0\}$ is a submartingale. This is a result of Jensen's inequality, a result on convex functions and expectations, proved by Johan Jensen in 1906. The result is omitted here.

Another class of stochastic processes of vital importance is the class of *predictable processes.* If a stochastic process is predictable, it has some nice properties with regards to Martingales on the same filtration.

Definition 2.4.2. *A stochastic process $X = \{X(t) : t \geq 0\}$ is called* predictable *with respect to a filtration $\{\mathcal{F}_t : t \geq 0\}$ (or $\mathcal{F}_t$-predictable) if it is measurable on the σ-algebra $\mathcal{F}_{t-} := \sigma(\bigcup_{h>0} \mathcal{F}_{t-h})$.*

A very deep result, that will be the basis of the work we will do with counting processes (see section 3.4), is what is called the Doob-Meyer decomposition. Before stating the theorem, we have to define a submartingale of class D.

Definition 2.4.3. *A submartingale $X = \{X(t) : t \geq 0\}$ is a* submartingale of class D *if the class of random variables $X(T)$ where T is an arbitrary stopping time is uniformly integrable.*

Now we can state the theorem:

Theorem 2.4.1 (Doob-Meyer decomposition)**.** *Let $X = \{X(t) : t \geq 0\}$ be a càdlàg submartingale of class D. Then there exists a unique, nondecreasing, predictable process $\tilde{X} = \{\tilde{X}(t) : t \geq 0\}$ such that $\tilde{X}(0) = 0$ a.s. and $X - \tilde{X}$ is a uniformly integrable Martingale. This process $\tilde{X}$ is called the* compensator *of X.*

Using this theorem, and remembering that M^2 is a submartingale if M is a square integrable Martingale, we can find a process $\tilde{M^2}$ as in theorem 2.4.1 such that $M^2 - \tilde{M^2}$ is a Martingale. This process is usually denoted $\langle M \rangle$ and is called the *predictable variation process* of M.

The predictable variation process can be very useful in deriving convergence in probability for Martingales. This is done using the following version of the *Lenglart inequality* (see section II.5.2 in [9]):

Theorem 2.4.2. *Let M be a Martingale with predictable variation process $\langle M \rangle$. Then the following inequality holds:*

$$\mathbb{P}(\sup_{s \in [0,t]} |M(s)| > \eta) \leq \frac{\delta}{\eta^2} + \mathbb{P}(\langle M \rangle(t) > \delta) \tag{2.4.1}$$

Using this result, one may see that a Martingale doesn't take on large values on $[0, t]$ if its predictable variation process takes on small values. Proving that a Martingale has mean 0 and proving that the predictable variation process of the process converges to 0 in probability is then enough to prove convergence in probability to 0 of the Martingale itself.

Combining these properties of (sub)Martingales and predictable processes, one may derive a truly wonderful result using stochastic integrals when the integrand is a predictable process and when we integrate with respect to a martingale, that is, the resulting process is again a Martingale. The result also gives us a way to calculate the predictable variation process of the Martingale. The following is theorem II.3.1 in [9].

Theorem 2.4.3. *Assume M is a finite variation, square integrable Martingale on a filtration $\{\mathcal{F}_t : t \geq 0\}$, and H is a predictable process on the same filtration and $\int H^2 \mathbf{d}\langle M\rangle$ is locally finite. Then $\int H \mathbf{d}M$ is a square integrable Martingale with predictable variation process $\langle \int H \mathbf{d}M\rangle = \int H^2 \mathbf{d}\langle M\rangle$.*

3 Multistate models

This chapter focusses on formally defining multistate models and its quantities. We introduce the product integral and counting processes and use the notions about Martingales to acquire useful results on counting processes. We define the transition hazard and the censoring hazard for a multistate model. At the end of this chapter, we shortly discuss Markov processes and why the transition hazard as we define it does not imply Markovianess of the model, and review the estimators we defined in chapter 2 for multistate models.

3.1 Quantities in a multistate model

In our multistate models, we assume a model where an individual can be in one state at a time, and we assume there are finitely many states. Transition times will be the times an individual makes a transition from one state into another. In any finite time interval $[0, t]$, we assume an individual makes a finite amount of transitions. In most models in survival analysis, there exists at least one absorbing state (often, this state is death), so a finite number of transitions for each individual is a reasonable assumption.

We may then define the following quantities as in [1]:

T^*_{ik} = time of kth transition for individual i (which we will define to be ∞ if the ith individual enters an absorbing state before transition k. Let $T^*_{i0} \equiv 0$);

C_i = censoring time for individual i;

$s_i(t)$ = state of individual i at time t;

s_{ik} = state of individual i between times $T^*_{i,k-1}$ and T^*_{ik};

$T^*_i = \sup_k \{T^*_{ik} : T^*_{ik} < \infty\}$;

$\delta_i = I(C_i > T^*_i)$ the indicator of whether individual i was (never) censored;

Furthermore, we define $T_{ik} = \min(T^*_{ik}, C_i)$ and $T_i = \min(T^*_i, C_i)$ to be the right-censored transition times for individual i. Denote $\underline{T}^*_i = (T^*_{ik} : k \geq 1)$ and $\underline{s}_i = (s_{ik} : k \geq 1)$ the collection of transition times and states occupied by individual i.

This way, we model every transition, all possible censoring and every state an individual can be in at any time.

3.2 The model without censoring

Similar to what we did at the start of chapter 2, we may investigate the state occupation probabilities, assuming there is no censoring at all. Again, with i.i.d. individuals, it is easy to see that a consistent estimator for the state occupation probability $p_j(t) = \mathbb{P}(s_i(t) = j)$ is the following:

$$\hat{p}^*_j(t) := \frac{\#\text{number of people in state } j \text{ at time } t}{\#\text{number of people in the studies at time } 0} \tag{3.2.1}$$

With right-censoring, acquiring an estimator for this probability is not easy, and proving consistency is rather difficult as well. Without censoring, however, it is very easy to find consistency of this estimator through the Law of Large Numbers. We encounter this fact again in chapter 5. Since consistency holds for any $t \in [0, \tau)$ (see theorem 4.2.1 for the definition of τ), we have an even stronger result:

Theorem 3.2.1. *In a multistate model with no censoring, $\hat{p}_j^*(t)$ is uniformly consistent on $[0, \tau)$, i.e.*

$$\sup_{t \in [0,\tau)} |\hat{p}_j^*(t) - p_j(t)| \xrightarrow{\mathbb{P}} 0$$

as $n \to \infty$.

Since right-censoring is almost always present in studies with multistate models, we have to take it into account, so this estimator, however nice and easy it is to work with, is not an estimator we can often determine. To develop an estimator that does take right-censoring into account, the rest of this chapter is devoted to the necessary tools and background to enable us to acquire an estimator for the state occupation probabilities in chapter 5.

3.3 Product integral

A useful tool in survival analysis and multistate models is the so called *product integral.* Like the way a regular integral can be thought of as a limit of a sum, the product integral can be thought of as the limit of a product. The product integral arises naturally in survival analysis (hence its usefulness), and it gives us nice ways to represent quantities, even if they are discrete products. In this thesis, we will use the definition and theorems regarding the product integral of Andersen et al, 1993 [9] (see section II.6).

Definition 3.3.1. *Let $X(s)$ be a $p \times p$ matrix of càdlàg distribution functions of locally bounded variation, let I be the $p \times p$ identity matrix (where $p \in \mathbb{Z}_{\geq 1}$). We define $Y(t) := \prod_{s \leq t}[I + \mathbf{d}X(s)]$ to be the* product integral *of X over intervals of the form $[0, t]$ as follows:*

$$Y(t) = \prod_{s \leq t}[I + \mathbf{d}X(s)] = \lim_{\max|t_i - t_{i-1}| \to 0} \prod_{i: t_i \leq t} (I + X(t_i) - X(t_{i-1})) \tag{3.3.1}$$

where $0 = t_0 < t_1 < \ldots < t_m$ is a partition of the interval $[0, t]$, and $\mathbf{d}X(s) = X(s) - X(s-)$

Different articles and books use different notations within the product integral. Some writers prefer $\mathbf{d}X(s)$ (e.g. Datta & Satten, 2001 [1]) or even $\mathbf{d}X$ (e.g. Glidden, 2002 [3] and and Andersen et al, 1993 [9]), omitting the time altogether. Another way to write this quantity is $X(\mathbf{d}s)$ (e.g. Glidden, 2002). In this thesis, we will use the representation as used in Datta & Satten, 2001. In practice, it doesn't matter which one you use, though: $\mathbf{d}X(s) = X(s) - X(s-) = X([0, s]) - X([0, s)) = X([0, s] - [0, s)) = X(\mathbf{d}s)$.

The product integral is the (unique) solution to an integral equation, introduced by Volterra in 1887 [11]). This theorem will prove to be essential later on.

Theorem 3.3.1 (Volterra equation)**.** $\prod_{s \leq t} [I + \mathbf{d}X(s)]$ *exists and is a càdlàg function of locally bounded variation. Furthermore, it is the unique solution to the integral equation*

$$Y(t) = I + \int_0^t Y(s-)\mathbf{d}X(s) \tag{3.3.2}$$

The product integral has some other interesting qualities. The first one is multiplicativity, i.e. for $s \leq u \leq t$, we have $\prod_{(s,t]}(I + \mathbf{d}X(w)) = \prod_{(s,u]}(I + \mathbf{d}X(w)) \cdot \prod_{(u,t]}(I + \mathbf{d}X(w))$. This follows easily from definition 3.3.1.

A very important result regarding the product integral is the *Duhamel equation*. It will have a major role in most of the proofs in this thesis. The Duhamel equation is essentially a continuous version of the discrete equality

$$\prod_{i=1}^{n}[I + A_i] - \prod_{i=1}^{n}[I + B_i] = \sum_{i=1}^{n}\left(\prod_{j=1}^{i-1}[I + A_j][A_i - B_i]\prod_{j=i+1}^{n}[I + B_j]\right) \tag{3.3.3}$$

As the Duhamel equation is a continuous version of (3.3.3), the continuous version seems quite natural.

Theorem 3.3.2 (Duhamel equation)**.** *Let* $Y(t) = \prod_{s \leq t}[I + \mathbf{d}X(s)]$ *and* $Y'(t) = \prod_{s \leq t}[I + \mathbf{d}X'(s)]$. *Then*

$$Y(t) - Y'(t) = \int_0^t \prod_{u<s}[1 + \mathbf{d}X(u)]\mathbf{d}[X(s) - X'(s)]\prod_{s<u\leq t}[1 + \mathbf{d}X'(u)] \tag{3.3.4}$$

The version we use more often is the following corollary:

Corollary 3.3.3. *Assume the conditions of theorem 3.3.2. If* $Y'(t)$ *is non-singular, then equation* (3.3.4) *can be rewritten:*

$$Y(t)Y'(t)^{-1} - 1 = \int_0^t \prod_{u<s}[1 + \mathbf{d}X(u)]\mathbf{d}[X(s) - X'(s)]\left(\prod_{u\leq s}[1 + \mathbf{d}X'(u)]\right)^{-1} \tag{3.3.5}$$

A final important property of the product integral worth mentioning is the fact that it is a continuous mapping [5]. This fact will be very useful in chapter 5, when we have a product integral as an estimator for the state occupation probabilities.

3.4 Counting processes

An important class of stochastic processes in survival analysis is the class of *counting processes*. A counting process is basically a stochastic process that counts the number of occurrences of events in a model.

Definition 3.4.1. *A* counting process $N = \{N(t) : t \geq 0\}$ *is a stochastic process adapted to a filtration* $\{\mathcal{F}_t : t \geq 0\}$ *satisfying the following properties:*

1. *N is right-continuous*
2. $N(t) \geq 0$ *and* $N(0) = 0$
3. $N(t) < \infty$ *a.s. for all* $t \geq 0$

4. $N(t)$ is an integer for any $t \geq 0$

5. $N(t) \geq N(s)$ if $t \geq s$

Quite often, an additional constraint is added, that is, the jumps in the counting process all have size +1. We do not use this constraint, although the results would not change if we did use it. If the jumps all have size +1, that means there is at most one occurrence at any time t.

From the definition, we can see that a counting process is a piecewise constant, non-decreasing, càdlàg process with finite expectation at every time t. It is therefore easy to see that a counting process is a submartingale. By theorem 2.4.1, a counting process N has a compensator, which is usually denoted A, such that $M = N - A$ is a Martingale.

A counting process has an intensity process $\alpha = (\alpha(t) : t \geq 0)$. Because the counting process is integrable, we can define the intensity process:

$$\alpha(t) := \lim_{\mathbf{d}t \downarrow 0} \frac{\mathbb{E}(\mathbf{d}N(t)|\mathcal{F}_{t-})}{\mathbf{d}t} \tag{3.4.1}$$

Now $A(t) = \int_0^t \alpha(s)\mathbf{d}s$ is the compensator of $N(t)$. We can now, informally, write

$$\mathbf{d}N(t) = \mathbf{d}M(t) + \mathbf{d}A(t) = \alpha(t)\mathbf{d}t + \mathbf{d}M(t) \tag{3.4.2}$$

This equation will be useful later on.

Another very useful property of the counting process is the form of the predictable variation process of the Martingale M associated with the counting process N. According to equation (2.4.3) in Andersen et al, section II, the predictable variation process of M is given by $\langle M \rangle = A$. This result is also a corollary of theorem 2.5.1 in Fleming & Harrington [2]:

Theorem 3.4.1. *Let N be a counting process with compensator A. If A is continuous, then* $\mathbb{E}(M(t)^2) = \mathbb{E}A(t)$

3.5 Stochastic processes in a multistate model

We have defined $s_i(t)$, the state occupied by individual i at time t. These are random variables, so $(s_i(t) : t \geq 0)$ is a stochastic process, determining the path through the model of individual i, where we can see the state they are in at any time t. This means that, if individual i transitions from state j to state j' at time t, that $s_i(t) = j'$, where $j, j' \in \{1, \ldots, K\} = \mathscr{S}$, the state space of our multistate model. Therefore, the sample paths of s_i are right-continuous.

When starting at time 0, each individual is in a certain state already. We define the starting state the probability that an individual starts in each state. For now, we denote this as the vector $p(0) = (\mathbb{P}(s_i(0) = 1), \ldots, \mathbb{P}(s_i(0) = K))$.

Furthermore, we define our filtration $\{\mathcal{F}_t : t \geq 0\}$, where $\mathcal{F}_t = \sigma(\{\underline{T_i^*}, \underline{s_i}, I(T_i \leq u, \delta_i = 0), 0 \leq u \leq t, i = 1, \ldots, n\})$. This is the σ-algebra that contains the history of the model up to time t. Because all s_i are right-continuous, so is our filtration.

We can now give expressions for a counting process, counting the number of transitions from state j to state j' of individual i, as well as the (uncensored) transition hazard, which is quite similar to the survival hazard we saw in equation (2.2.1).

Let$j \neq j'$. Define

$$N^*_{i;jj'}(t) = \sum_{k\geq 1} I(T^*_{ik} \leq t, s_{ik} = j, s_{i,k+1} = j') \tag{3.5.1}$$

to be the counting process as described above. Furthermore, let

$$\alpha_{jj'}(t) = \lim_{\mathbf{d}t\downarrow 0} \sum_{k\geq 1} \frac{\mathbb{P}[T^*_{ik} \in [t, t+\mathbf{d}t), s_{ik} = j' | s(t-) = j]}{\mathbf{d}t} \tag{3.5.2}$$

be the transition hazard of the uncensored chain from state j to state j'. Like the survival hazard, we can see this equals

$$\alpha_{jj'}(t) = \lim_{\mathbf{d}t\downarrow 0} \frac{\mathbb{E}[\mathbf{d}N^*_{i;jj'}(t) | s_i(t-)]}{\mathbf{d}t} \tag{3.5.3}$$

with the help of Fubini's theorem. Then $\mathbf{A} = \{A_{jj'}\}$ is the cumulative hazard matrix, where $A_{jj'}(t) = \int_0^t \alpha_{jj'}(s)\mathbf{d}s$, where $\alpha_{jj}(t) = -\sum_{j'\neq j} \alpha_{jj'}(t)$.

For the censoring hazard, we assume it is independent of transition times and the states occupied. Furthermore, since we assume all individuals are i.i.d., the censoring hazard is the same for each individual.

We have to redefine the censoring hazard for each individual as well, because of the multistate model. We define the censoring hazard like we did in section 2.2.

$$\lambda_c(t) = \lim_{\mathbf{d}t\downarrow 0} \frac{\mathbb{P}(C_i \in [t, t+\mathbf{d}t), \delta_i = 0 | T_i \geq t)}{\mathbf{d}t} \tag{3.5.4}$$

The cumulative censoring hazard is then given by $\Lambda_c(t) = \int_0^t \lambda_c(s)\mathbf{d}s$. With the definition of the product integral and using (2.2.3), we find

$$\mathbb{P}(C_i > t, \delta_i = 0) = \prod_{(0,t]} (1 - \mathbf{d}\Lambda_c(s)) =: K(t) \tag{3.5.5}$$

Which is essentially the probability that individual i has survived up to time t without being censored.

3.5.1 Markov processes

In survival and multistate models, the cumulative transition hazard (or survival hazard) is often estimated under the assumption that the system is Markov. As we stated in the introduction of this thesis, this is a strong and often not very realistic assumption. We will still state the definition of a Markov process, and show how the process being Markov affects the transition hazard compared to the process not necessarily being Markov.

Definition 3.5.1. *Let $(S, \mathcal{S})$ be a measurable space. A stochastic process $X = \{X_t : t \geq 0\}$ adapted to a filtration $\{\mathcal{F}_t : t \geq 0\}$ is called a* Markov process *if, for all $A \in \mathcal{S}$ and for all $s < t$*

$$\mathbb{P}(X(t) \in A | \mathcal{F}_s) = \mathbb{P}(X(t) \in A | X_s) \tag{3.5.6}$$

Looking at the definition of the transition hazard (see (3.5.2)), we can see that this is a conditional probability which only assumes knowledge of the state of an individual at time $t-$. This is not equal to a Markov assumption, though, because we do not assume that

$$\lim_{\mathbf{d}t\downarrow 0}\sum_{k\geq 1}\frac{\mathbb{P}[T^*_{ik}\in[t,t+\mathbf{d}t),s_{ik}=j'|s(t-)=j]}{\mathbf{d}t}$$

and

$$\lim_{\mathbf{d}t\downarrow 0}\sum_{k\geq 1}\frac{\mathbb{P}[T^*_{ik}\in[t,t+\mathbf{d}t),s_{ik}=j'|s(t-)=j,\mathcal{F}_{t-}]}{\mathbf{d}t}$$

are equal. If we can prove consistency of an estimator for the integrated version of (3.5.2) (which we will for the Nelson-Aalen estimator do in chapter 4), we may estimate the cumulative transition hazard for non-Markov systems.

3.6 Estimators in a multistate model

In case of a multistate model, we require estimators for every possible transition between states. A similar estimator for each transition can be found in a similar way as the Nelson-Aalen estimator from section 2.3.1: let $n_{k;jj'}$ be the number of people experiencing a transition from state j to state j' at time t_k and let $y_{k;j}$ be the number of people at risk in state j at time t_k- (where $j\neq j'$). In that case, the Nelson-Aalen estimator for the transition hazard between j and j' is given by

$$\hat{A}_{jj'}(t):=\sum_{k:t_k\leq t}\frac{n_{k;jj'}}{y_{k;j}} \tag{3.6.1}$$

In a similar way as in section 2.3.2, one may now find the *Aalen-Johansen estimator* for the state occupation probability for each state j. First, define n_k the $K\times K$ matrix containing all values $n_{k;jj'}$ (with $n_{k;jj}:=\sum_{j'\neq j}n_{k;jj'}$) and define y_k the diagonal matrix with elements $y_{k;j}$ on its diagonal. The Aalen-Johansen estimator of the state occupation probability for state j is then:

$$\hat{p}_j(t)=\left(y_0\prod_{k:t_k\leq t}(I+n_ky_k^{-1})\right)_j \tag{3.6.2}$$

where y_0 denotes the initial state occupation distribution at time 0.

We prove consistency of the Nelson-Aalen estimator in chapter 4 and consistency of the Aalen-Johansen estimator in chapter 5.

4 The Nelson-Aalen estimator for non-Markov data

In this chapter, we prove consistency of the Nelson-Aalen estimator for cumulative transition hazards in multistate models that are not necessarily Markov. The theorem that we will prove is a very nice result, first published by Datta & Satten in 2001 [1]. We first define a counting process for each transition, as well as an "at risk of transitions in this state" process, hereafter called the at risk process, for each state. Then we define similar processes, taking the risk of right-censoring into account. We prove a lemma regarding the expectations of these processes to make it seem reasonable that the censored processes can be used in the Nelson-Aalen estimator instead of the uncensored processes, and finally we prove that the estimator that we find is indeed asymptotically consistent.

The results were derived before in [1]. However, while working through the proofs of both the lemma and the theorem stating consistency of the estimator, a couple of small errors were detected. They have been fixed in the proofs in this thesis.

4.1 Uncensored and censored processes

4.1.1 Nelson-Aalen estimator for uncensored processes

Define $N^*_{jj'}(t) = \sum_{i=1}^n N^*_{i;jj'}(t)$ and $Y^*_j(t) = \sum_{i=1}^n \sum_{k\geq 1} I(T^*_{i,k-1} < t \leq T^*_{ik}, s_{ik} = j)$ for $j \neq j'$. An estimator of A is given by the matrix $\hat{A}^* = \{\hat{A}^*_{jj'}\}$ where

$$\hat{A}^*_{jj'}(t) = \begin{cases} \displaystyle\int_0^t J_j(u) Y^*_j(u)^{-1} \mathbf{d}N^*_{jj'}(u) & j \neq j' \\ \displaystyle -\sum_{j'\neq j} \hat{A}^*_{jj'}(t) & j = j' \end{cases} \tag{4.1.1}$$

where $J_j(u) = I(Y^*_j(u) > 0)$. It takes little work to show that this is in fact the *Nelson-Aalen estimator* as defined in section 3.6.

In [8], section 3.1.5, this estimator is almost derived. We will do the same derivation here.

First, notice that the compensator of $N^*_{jj'}(t)$ is equal to $\int_0^t Y^*_j(s)\alpha_{jj'}(s)\mathbf{d}s$, because the transition rate is proportional to the number of people at risk. Therefore, (3.4.2) becomes

$$\mathbf{d}N^*_{jj'}(t) = \mathbf{d}M^*_{jj'}(t) + Y^*_j(t)\alpha_{jj'}(t)\mathbf{d}t \tag{4.1.2}$$

Multiplying both sides by $J_j(t)$ to be able to divide both sides by $Y^*_j(t)$, we find

$$\frac{J_j(t)}{Y^*_j(t)}\mathbf{d}N^*_{jj'}(t) = \frac{J_j(t)}{Y^*_j(t)}\mathbf{d}M^*_{jj'}(t) + J_j(t)\alpha_{jj'}(t)\mathbf{d}t \tag{4.1.3}$$

Integrating both sides of the equation from 0 to t, we find

$$\int_0^t \frac{J_j(s)}{Y^*_j(s)}\mathbf{d}N^*_{jj'}(s) = \int_0^t \frac{J_j(s)}{Y^*_j(s)}\mathbf{d}M^*_{jj'}(s) + \int_0^t J_j(s)\alpha_{jj'}(s)\mathbf{d}s \tag{4.1.4}$$

Taking expectations on both sides and noting that $\int_0^t \frac{J_j(s)}{Y^*_j(s)}\mathbf{d}M^*_{jj'}(s)$ is a mean-zero Martingale by theorem 2.4.3, we find an unbiased estimator for $\int_0^t J_j(s)\alpha_{jj'}(s)\mathbf{d}s$. Since it seems likely for the transition hazard to be low when nobody is at risk, a logical conslusion would be that this estimator is also a good estimator for $\int_0^t \alpha_{jj'}(s)\mathbf{d}s$. For the case of censored data, this is shown in theorem 4.2.1 (see below).

4.1.2 Nelson-Aalen estimator for censored processes

The problem with the estimator in (4.1.1) is the fact that these estimators do not take censoring into account. Whenever the data are censored, these processes cannot be determined, because these processes count the actual number of transitions, rather than the observed transitions. When we do take censoring into account, we immediately notice that the number of observed transitions is less than the number of actual transitions. Even though the censored individuals left the study early, the information that has been gathered before that time is still useful. To be able to still work with the total number of people in the study, we have to account for them in a way. Since we assumed non-informative censoring, it makes sense to assume that the transition rate between two states j and j' for the censored individuals and the uncensored individuals are the same. Stating this differently, the probability of observing a transition, given the fact that a transition takes place, is equal to the probability of having survived without having been censored at that time. We find the following processes:

$$\begin{aligned} N_{jj'}(t) &= \sum_{i=1}^{n}\sum_{k\geq 1} I(T^*_{ik}\leq t, C_i \geq T^*_{ik}, s_{ik}=j, s_{i,k+1}=j')/K(T^*_{ik}-) \\ Y_j(t) &= \sum_{i=1}^{n}\sum_{k\geq 1} I(T^*_{i,k-1} < t \leq T^*_{ik}, C_i \geq T^*_{ik}, s_{ik}=j)/K(t-) \end{aligned} \tag{4.1.5}$$

We expect these quantities to be equal in expectation to N^* and Y^* respectively. In lemma 4.1.1, we show that this is in fact the case. This lemma suggests that we may replace the counting process N^* and the at risk process Y^* by N and Y respectively. We acquire the following estimator:

$$\hat{A}_{jj'}(t) = \begin{cases} \int_0^t J_j(u)Y_j(u)^{-1}\mathbf{d}N_{jj'}(u) & j \neq j' \\ -\sum_{j'\neq j}\hat{A}_{jj'}(t) & j = j' \end{cases} \tag{4.1.6}$$

Note how this estimator is again the Nelson-Aalen estimator from section 3.6, written in a slightly different way.

Remark: Often, the distribution function of the censoring hazard is unknown as well. For the Nelson-Aalen estimator, this is not a problem, because the censoring probability is divided away in the fraction $J_j(u)Y_j(u)^{-1}\mathbf{d}N_{jj'}(u)$.

To give the strong indication that this estimator could be consistent, we prove the following lemma.

Lemma 4.1.1. *Suppose $K(t) > 0$ and j is not an absorbing state. Then*
(a) $\mathbb{E}[N_{jj'}(t)] = \mathbb{E}[N^*_{jj'}(t)]$
(b) $\mathbb{E}[Y_j(t)] = \mathbb{E}[Y^*_j(t)]$

Proof. (a) By Fubini's theorem, we find

$$\mathbb{E}(N_{jj'}(t)) = n \sum_{k\geq 1} \mathbb{E}\left\{ \frac{I(T^*_{ik} \leq t, C_i \geq T^*_{ik}, s_{ik} = j, s_{i,k+1} = j')}{K(T^*_{ik})} \right\} \tag{4.1.7}$$

Dropping the index i to improve readability, we have $I(C \geq T^*_k) = \prod_{s<T^*_k}[1 + \mathbf{d}\bar{X}(s)]$, with $\bar{X}(s) = -I(C \leq s)$. Now noting that $I(C \leq s) = I(C \leq s, \delta = 0)$ for $s \leq T^*_k$ and defining $X(s) := -I(C \leq s, \delta = 0)$, we find

$$I(T^*_k \leq t, C \geq T^*_k, s_k = j, s_{k+1} = j') = I(T^*_k \leq t, s_k = j, s_{k+1} = j') \prod_{s<T^*_k} [1 + \mathbf{d}X(s)]$$

By definition of $K(s)$, we find, on the set $\{T^*_k \leq t, C \geq T^*_k, s_k = j, s_{k+1} = j'\}$, with $X'(s) := -\int_0^s \lambda_c(u) I(T \geq u) \mathbf{d}u$, that $K(T^*_k-) = \prod_{s<T^*_k}[1 + \mathbf{d}X'(s)]$. By theorem 2.4.1, $X'(s)$ is the compensator of $X(s)$, so we can see that $X(s) - X'(s) =: -M^c(s)$ is a martingale with mean 0 w.r.t. $\mathcal{F}_s$.

From the Duhamel equation (3.3.4), we obtain

$$\begin{aligned}
&\frac{I(T^*_k \leq t, C \geq T^*_k, s_k = j, s_{k+1} = j')}{K(T^*_k-)} \\
&= I(T^*_k \leq t, s_k = j, s_{k+1} = j') \prod_{s<T^*_k} [1 + \mathbf{d}X(s)] \left(\prod_{s<T^*_k} [1 + \mathbf{d}X'(s)] \right)^{-1} \\
&= I(T^*_k \leq t, s_k = j, s_{k+1} = j') \\
&\left(1 + \int_0^{T^*_k-} \prod_{u<s} [1 + \mathbf{d}X(u)] \mathbf{d}[X(s) - X'(s)] \prod_{s<u<T^*_k} [1 + \mathbf{d}X'(u)] \left(\prod_{s<T^*_k} [1 + \mathbf{d}X'(s)] \right)^{-1} \right) \\
&= I(T^*_k \leq t, s_k = j, s_{k+1} = j') \left(1 + \int_0^{T^*_k-} \prod_{u<s} [1 + \mathbf{d}X(u)] \mathbf{d}[X(s) - X'(s)] \left(\prod_{u<s} [1 + \mathbf{d}X'(u)] \right)^{-1} \right) \\
&= I(T^*_k \leq t, s_k = j, s_{k+1} = j') \left(1 + \int_0^{T^*_k-} \frac{I(C \geq s)}{K(s)} \mathbf{d}[X(s) - X'(s)] \right)
\end{aligned} \tag{4.1.8}$$

For the integral part on the right-hand side of equation (4.1.8), note that it is equal to 0 for $C < s$. Also note that it is non-negative for $C > s$ and $T^*_k \leq t$, so the integral itself is non-negative, hence it is bounded from above by the same integral from 0 to t. We find

$$\begin{aligned}
&\frac{I(T^*_k \leq t, C \geq T^*_k, s_k = j, s_{k+1} = j')}{K(T^*_k-)} \\
&= I(T^*_k \leq t, s_k = j, s_{k+1} = j') \left(1 + \int_0^{T^*_k-} \frac{I(C \geq s)}{K(s)} \mathbf{d}[X(s) - X'(s)] \right) \\
&\leq I(T^*_k \leq t, s_k = j, s_{k+1} = j') \left(1 + \int_0^{t} \frac{I(C \geq s)}{K(s)} \mathbf{d}[X(s) - X'(s)] \right)
\end{aligned} \tag{4.1.9}$$

Since the integral in the right-hand side of equation (4.1.8) is non-negative, it is bounded from below by 0. Now note that $\frac{I(C\geq s)}{K(s)}$ is predictable w.r.t. $\mathcal{F}_s$. Since $X(s) - X'(s)$ is a mean-zero martingale w.r.t. $\mathcal{F}_s$, we find the integral from 0 to t is a martingale w.r.t. $\mathcal{F}_t$ as well by theorem 2.4.3. Taking expectations on both sides of equation (4.1.8), we get the following:

$$\begin{aligned}
&\mathbb{E}\left(\frac{I(T_k^* \le t, C \ge T_k^*, s_k = j, s_{k+1} = j')}{K(T_k^*-)}\right) \\
&= \mathbb{E}\left(I(T_k^* \le t, s_k = j, s_{k+1} = j')\left(1 + \int_0^{T_k^*-} \frac{I(C \ge s)}{K(s)} \mathbf{d}[X(s) - X'(s)]\right)\right) \\
&\le \mathbb{E}\left(I(T_k^* \le t, s_k = j, s_{k+1} = j')\left(1 + \int_0^{t} \frac{I(C \ge s)}{K(s)} \mathbf{d}[X(s) - X'(s)]\right)\right) \\
&\le \mathbb{E}\left(I(T_k^* \le t, s_k = j, s_{k+1} = j')\right) + \mathbb{E}\left(\int_0^{t} \frac{I(C \ge s)}{K(s)} \mathbf{d}[X(s) - X'(s)]\right) \\
&= \mathbb{E}(I(T_k^* \le t, s_k = j, s_{k+1} = j'))
\end{aligned} \tag{4.1.10}$$

So now the second term is smaller than or equal to the last term. But by non-negativity of the integral, the inequality the other way around is satisfied as well. Therefore, the two are equal, and we find

$$\mathbb{E}\left(\frac{I(T_k^* \le t, C \ge T_k^*, s_k = j, s_{k+1} = j')}{K(T_k^*-)}\right) = \mathbb{E}(I(T_k^* \le t, s_k = j, s_{k+1} = j')) \tag{4.1.11}$$

Hence, by Fubini's theorem, we may conclude $\mathbb{E}(N_{jj'}(t)) = \mathbb{E}(N^*_{jj'}(t))$

(b) By Fubini's theorem, We have

$$\mathbb{E}[Y_j(t)] = n \sum_{k \ge 1} \mathbb{E}\left\{\frac{I(T_{k-1}^* < t \le T_k^*, C \ge t, s_k = j}{K(t-)}\right\} \tag{4.1.12}$$

Define $\bar{X}(s) := -I(C \le s)$, $X(s) = -I(C \le s, \delta := 0)$, like in the proof of part (a). Now on the set $\{T_k^* \ge t, s_k = j\}$ we have $X(s) = \bar{X}(s)$ for $s < t$. Also note $I(C \ge t) = \prod_{s<t}[1 + \mathbf{d}\bar{X}(s)]$. Hence we find $I(T_{k-1}^* < t \le T_k^*, C \ge t, s_k = j) = I(T_{k-1}^* < t \le T_k^*, s_k = j) \prod_{s<t}[1 + \mathbf{d}X(s)]$. Now we also have, on the set $\{T_{k-1}^* < t \le T_k^*, C \ge t, s_k = j\}$, that $K(t) = \prod_{s<t}[1 + \mathbf{d}X'(s)]$ where $X'(s)$ is defined as in the proof of part (a). Following the application of the Duhamel equation as in part (a) and using the second inequality similarly as in equation (4.1.10), we get the desired result. Hence $\mathbb{E}(Y_j(t)) = \mathbb{E}(Y_j^*(t))$.

□

Remark: We have not justified the equalities of the form $I(C \ge T_k^*) = \prod_{s<T_k^*}[1 + \mathbf{d}\bar{X}(s)]$. We will justify the use of the product integral in the proof of lemma 4.1.1 here with help of definition 3.3.1. We will justify the equality for one of these product integrals. The others are done in a similar fashion.

Knowing definition 3.3.1, we find $\prod_{s<T_k^*}[1 + \mathbf{d}\bar{X}(s)] = \lim_{\max|t_i - t_{i-1}| \to 0} \prod_{t_i < T_k^*}[1 + \bar{X}(t_i) - \bar{X}(t_{i-1})]$ where $t_0 < \ldots < t_m$ is a partitition of $[0, T_k^*)$. Now recall the definition of $\bar{X}(t) = -I(C \le s)$. There are two cases to be treated: $C \ge T_k^*$ and $C < T_k^*$.

Case 1: $C \ge T_k^*$
In this case, for any $t_i < T_k^*$, we have $\bar{X}(t_i) = 0$, hence we have

$\lim_{\max|t_i - t_{i-1}|\to 0} \prod_{t_i < T_k^*}[1 + \bar{X}(t_i) - \bar{X}(t_{i-1})] = \lim_{\max|t_i - T_{i-1}|\to 0} \prod_{t_i < T_k^*}[1] = 1 = I(C \geq T_k^*)$. So in this case the equality holds.

Case 2: $C < T_k^*$
In this case, $I(C \geq T_k^*) = 0$. So there exists $u \in (0, T_k^*)$ such that $C = u$, and more importantly, in our partition $t_0 < \ldots < t_m$, there exists $j \leq m$ s.t. $u \in (t_{j-1}, t_j]$. Then $\bar{X}(t_i) - \bar{X}(t_{i-1}) = 0$ for all $i \neq j$ and $\bar{X}(t_j) - \bar{X}(t_{j-1}) = -1$. But in this case we multiply by $1 - 1 = 0$, hence our product integral must be 0 as well.

In both cases, the two coincide, hence the two are equal.

4.2 Consistency of the estimator

Since the expectations of N and N^* are the same (and Y and Y^* as well), we expect the Nelson-Aalen estimator with the censored quantities to be consistent as well. However, a surprising amount of work is needed to prove consistency of this estimator, and a few other assumptions have to be made as well. We mainly require two integrated processes that we encounter to be finite, because the bounds that we produce may otherwise not be good bounds.

To be able to state and prove the theorem that gives us consistency, we need a couple of definitions first. Let $\tau = \sup\{t : \int_0^t \alpha_{jj'}(u)\mathbf{d}u < \infty, j \neq j', K(t) > 0\}$ and denote $\|\cdot\|$ any norm on the space of $K \times K$ matrices. Define $\overline{y}_j^*(t) = \mathbb{E}\{n^{-1}Y_j^*(t)\} = \mathbb{P}\{s_i(t) = j\}$.

We should now be able to prove the following theorem:

Theorem 4.2.1. *Assume all conditions of lemma* 4.1.1*. Let* $t \in [0, \tau)$*. Furthermore, assume*

$$\int_0^t \frac{I(\overline{y}_j^*(u) > 0)\alpha_{jj'}(u)}{\overline{y}_j^*(t)}\mathbf{d}u < \infty \quad \textit{and} \quad \int_0^t \frac{I(\overline{y}_j^*(u) > 0)\lambda_c(u)}{\overline{y}_j^*(t)}\mathbf{d}u < \infty \quad \forall j, j'$$

Then $\sup\limits_{[0,t]}\|\hat{A}(u) - A(u)\| \xrightarrow{\mathbb{P}} 0$

Proof. Fix a non-absorbing state j such that $j' \neq j$ for any other state j', and recall the definitions of N, N^*, Y and Y^*. Let $\tilde{A}_{jj'}^*(u) = \int_0^u J_j(s)\alpha_{jj'}(s)\mathbf{d}s$.

Now first note that $\sup_{[0,t]}|\hat{A}_{jj'}(u) - A_{jj'}(u)| \leq \sup_{[0,t]}|\hat{A}_{jj'}(u) - \tilde{A}_{jj'}^*(u)| + \sup_{[0,t]}|\tilde{A}_{jj'}^*(u) - A_{jj'}(u)|$. We will show that both terms on the right-hand side converge to 0, uniformly, in probability on $[0, t]$.

Now when $\alpha_{jj'}(s) > 0$, we have $\overline{y}_j^*(t) > 0$ and since $\frac{Y_j^*(t)}{n} \xrightarrow{\mathbb{P}} \overline{y}_j^*(t)$ as $n \to \infty$, we find $(1 - J_j(s))\alpha_{jj'}(s) \xrightarrow{\mathbb{P}} 0$ as $n \to \infty$.

Then for all $s \in [0, t]$ we have $\alpha_{jj'}(s) \geq 0$, so $|(1 - J_j(s))\alpha_{jj'}(s)| \leq \alpha_{jj'}(s)$. Also note that $\int_0^t |\alpha_{jj'}(s)|\mathbf{d}s = \int_0^t \alpha_{jj'}(s)\mathbf{d}s < \infty$ by assumption. Then, by the Dominated Convergence Theorem (and non-negativity of the integral below), we find

$$
\begin{aligned}
&\sup_{[0,t]} |\tilde{A}^* jj'(u) - A_{jj'}(u)| \\
&= \sup_{[0,t]} |\int_0^u (1 - J_j(s))\alpha_{jj'}(s)\mathbf{d}s| \\
&\leq \int_0^t |(1 - J_j(s))\alpha_{jj'}(s)|\mathbf{d}s \xrightarrow{\mathbb{P}} \int_0^t 0\mathbf{d}s = 0
\end{aligned} \tag{4.2.1}
$$

So the first term converges to 0 uniformly in probability on $[0, t]$. For the second term in the right-hand side, we split it into more terms:

$$
\begin{aligned}
&\hat{A}_{jj'}(u) - \tilde{A}^*_{jj'}(u) = \int_0^u J_j(s)Y_j(s)^{-1}\mathbf{d}N_{jj'}(s) - \int_0^u J_j(s)\alpha_{jj'}(s)\mathbf{d}s \\
&= \int_0^u \frac{J_j(s)}{Y_j(s)}\mathbf{d}N_{jj'}(s) - \int_0^u \frac{J_j(s)}{Y_j^*(s)}Y_j^*(s)\alpha_{jj'}(s)\mathbf{d}s \\
&= \int_0^u \frac{J_j(s)}{Y_j(s)}\mathbf{d}N_{jj'}(s) - \int_0^u \frac{J_j(s)}{Y_j^*(s)}\mathbf{d}N^*_{jj'}(s) + \int_0^u \frac{J_j(s)}{Y_j^*(s)}\mathbf{d}N^*_{jj'}(s) - \int_0^u \frac{J_j(s)}{Y_j^*(s)}Y_j^*(s)\alpha_{jj'}(s)\mathbf{d}s \\
&= \int_0^u \frac{J_j(s)}{Y_j^*(s)}\mathbf{d}M^*_{jj'}(s) + \int_0^u \frac{J_j(s)}{Y_j(s)}\mathbf{d}N_{jj'}(s) - \int_0^u \frac{J_j(s)}{Y_j(s)}\mathbf{d}N^*_{jj'}(s) + \int_0^u \frac{J_j(s)}{Y_j(s)}\mathbf{d}N^*_{jj'}(s) \\
&- \int_0^u \frac{J_j(s)}{Y_j^*(s)}\mathbf{d}N^*_{jj'}(s) \\
&= \int_0^u \frac{J_j(s)}{Y_j^*(s)}\mathbf{d}M^*_{jj'}(s) + \int_0^u J_j(s)\left(\frac{\mathbf{d}N_{jj'}(s)}{Y_j(s)} - \frac{\mathbf{d}N^*_{jj'}(s)}{Y_j(s)}\right) \\
&+ \int_0^u J_j(s)\left(\frac{1}{Y_j(s)} - \frac{1}{Y_j^*(s)}\right)\mathbf{d}N^*_{jj'}(s)
\end{aligned} \tag{4.2.2}
$$

Now we have divided the second term into three more terms. We will prove convergence in probability to 0 for each of these terms.

First term:

$$\int_0^u \frac{J_j(s)}{Y_j^*(s)}\mathbf{d}M^*_{jj'}(s) =: M^{**}_{jj'}(u) \tag{4.2.3}$$

Here $M^*_{jj'}(s) = N^*_{jj'}(s) - \int_0^s \alpha_{jj'}(s)Y_j^*(u)\mathbf{d}u$ is a mean-zero Martingale, where $\int_0^s \alpha_{jj'}(s)Y_j^*(u)\mathbf{d}u$ is the predictable compensator for the counting process $N^*_{jj'}(s)$ (see [9], section II.4.1). Theorem 3.4.1 now gives us

$$\langle M^* \rangle_{jj'}(u) = \int_0^u \alpha_{jj'}(s)Y_j^*(s)\mathbf{d}s \tag{4.2.4}$$

Given that $\frac{J_j(s)}{Y_j^*(s)}$ is a predictable process, the term we want to have convergence to 0 for is a mean-zero Martingale M^{**}. Now theorem 2.4.3 and (4.2.4) give us

$$\langle M^{**}_{jj'}(u) \rangle = \left\langle \int_0^u \frac{J_j(s)}{Y_j^*(s)}\mathbf{d}M^*_{jj'}(s) \right\rangle = \int_0^u \left(\frac{J_j(s)}{Y_j^*(s)}\right)^2 \alpha_{jj'}(s)Y_j^*(s)\mathbf{d}s \tag{4.2.5}$$

If we can prove convergence to 0 in probability for (4.2.5), we will be able to prove convergence to 0 in probability for (4.2.3).

We get

$$\begin{aligned}&\int_0^u \left(\frac{J_j(s)}{Y_j^*(s)}\right)^2 \alpha_{jj'}(s) Y_j^*(s)\mathbf{d}s = \int_0^u \frac{J_j(s)}{Y_j^*(s)}\alpha_{jj'}(s)\mathbf{d}s \\ &= \frac{1}{n}\int_0^u \frac{J_j(s)}{\frac{Y_j^*(s)}{n}}\alpha_{jj'}(s)\mathbf{d}s \le \frac{1}{n}\int_0^u \frac{I(\overline{y}_j^*(s))}{\overline{Y}_j^*(s)}\alpha_{jj'}(s)\mathbf{d}s\end{aligned} \tag{4.2.6}$$

where $\overline{Y}_j^*(t) = \frac{Y_j^*(s)}{n}$. Since $Y_j^*(s)$ is the sum of i.i.d. indicators, we may consider $\overline{Y}_j^*(s)$ as a Bernoulli random variable with parameter $\overline{y}_j^*(s)$. Its standardised form is given by

$$Z_j(s) := I(\overline{y}_j^*(s) > 0)\frac{\sqrt{n}(\overline{Y}_j^*(s) - \overline{y}_j^*(s))}{\overline{y}_j^*(s)(1-\overline{y}_j^*(s))} \tag{4.2.7}$$

Now $Z_j(s)$ converges weakly on $D_-[0,t]$, the space of càdlàg functions on $[0,t]$, so for any $\epsilon \in (0,1)$, there exists $K > -\infty$ s.t. on sets of probability of at least $1-\epsilon$, there exists $n_0 \in \mathbb{Z}_{>0}$ such that we have $Z_j(s) > K$ for all $n \ge n_0$ and all $s \in [0,t]$. Let $\epsilon \in (0,1), K, n \ge n_0$ as above. Then $K < I(\overline{y}_j^*(s) > 0)\frac{\sqrt{n}(\overline{Y}_j^*(s)-\overline{y}_j^*(s))}{\overline{y}_j^*(s)(1-\overline{y}_j^*(s))}$. The following calculations will assume $\overline{y}_j^*(s) > 0$, because the inequality we wish to get will be trivially true when $\overline{y}_j^*(s) = 0$.

With K as above and some algebra, we find $K(1-\overline{y}_j^*(s)) < I(\overline{y}_j^*(s) > 0)\sqrt{n}\left(\frac{\overline{Y}_j^*(s)}{\overline{y}_j^*(s)} - 1\right)$. Rearranging terms some more now gives $\frac{K(1-\overline{y}_j^*(s))}{\sqrt{n}} + I(\overline{y}_j^*(s) > 0) < I(\overline{y}_j^*(s) > 0)\frac{\overline{Y}_j^*(s)}{\overline{y}_j^*(s)}$, which directly implies $\overline{Y}_j^*(s) > 0$. We find $\frac{I(\overline{y}_j^*(s)>0)}{\overline{y}_j^*(s)} - \frac{K(1-\overline{y}_j^*(s))}{\overline{Y}_j^*(s)\sqrt{n}} > \frac{I(\overline{y}_j^*(s)>0)}{\overline{Y}_j^*(s)}$. All terms in this inequality will be positive, with the possible exception of K. If K is non-negative, then obviously $\frac{2I(\overline{y}_j^*(s)>0)}{\overline{y}_j^*(s)} > \frac{I(\overline{y}_j^*(s)>0)}{\overline{Y}_j^*(s)}$. If $K < 0$, note that the inequality $\frac{I(\overline{y}_j^*(s)>0)}{\overline{y}_j^*(s)} - \frac{K(1-\overline{y}_j^*(s))}{\overline{Y}_j^*(s)\sqrt{n}} > \frac{1}{\overline{Y}_j^*(s)}$ holds for all $n > n_0$, hence we can pick n such that $-\frac{K(1-\overline{y}_j^*(s)}{\overline{Y}_j^*(s)\sqrt{n}} < \frac{I(\overline{y}_j^*(s)>0)}{\overline{y}_j^*(s)}$. In either case, we find $\frac{2I(\overline{y}_j^*(s)>0)}{\overline{y}_j^*(s)} > \frac{I(\overline{y}_j^*(s)>0)}{\overline{Y}_j^*(s)}$. This is the inequality we need. Now if $\overline{y}_j^*(s) = 0$, our strict inequality becomes an equality, so we have $\frac{2I(\overline{y}_j^*(s)>0)}{\overline{y}_j^*(s)} \ge \frac{I(\overline{y}_j^*(s)>0)}{\overline{Y}_j^*(s)}$.

Plugging this inequality into (4.2.6), we find

$$\int_0^t \left(\frac{J_j(s)}{Y_j^*(s)}\right)^2 \alpha_{jj'}(s) Y_j^*(s)\mathbf{d}s \le \frac{2}{n}\int_0^t \frac{I(\overline{y}_j^*(s) > 0)}{\overline{y}_j^*(s)}\alpha_{jj'}(s)\mathbf{d}s \tag{4.2.8}$$

and now notice that the right-hand-side converges to 0 in probability as $n \to \infty$, because the integral is finite by assumption. Therefore, the predictable variation process of (4.2.3) (which we will now denote $\langle M^{**}\rangle$ converges to 0 in probability as $n \to \infty$.

Knowing that $\langle M^{**}\rangle \xrightarrow{\mathbb{P}}$ as $n \to \infty$, theorem 2.4.2 (Lenglart's inequality) should give us convergence in probability to 0 for M^{**}. Let $\eta = \delta^{\frac{1}{4}}$, then we find, again as $n \to \infty$, that

$\mathbb{P}\left(\sup_{[s\in[0,u]}|M^{**}_{jj'}(s)| > \delta^{\frac{1}{4}}\right) \leq \delta^{\frac{1}{2}}$ for any $\delta > 0$. This proves convergence in probability to 0 of the first term, uniformly on $[0,t]$.

Second term:

$$\int_0^u J_j(s)\left(\frac{\mathbf{d}N_{jj'}(s)}{Y_j(s)} - \frac{\mathbf{d}N^*_{jj'}(s)}{Y_j(s)}\right) \tag{4.2.9}$$

From the proof of lemma 4.1.1, we have

$$\begin{aligned}&\tfrac{I(T^*_{ik}\leq t, C_i\geq T^*_{ik}, s_{ik}=j, s_{i,k+1}=j')}{K(T^*_{ik}-)}\\ &= I(T^*_{ik}\leq t, s_{ik}=j, s_{i,k+1}=j')\left(1+\int_0^{T^*_{ik}-}\frac{I(C_i\geq s)}{K(s)}\mathbf{d}[X_i(s)-X'_i(s)]\right)\end{aligned} \tag{4.2.10}$$

Summing first over k and then i and recalling definitions for $N^*_{jj'}$ and $N_{jj'}$ and rearranging terms as we partially did in the proof of lemma 4.1.1, we find

$$N_{jj'}(u) - N^*_{jj'}(u) = -\sum_i\sum_k I(T^*_{ik}\leq u, s_{ik}=j, s_{i,k+1}=j')\int_0^u \frac{I(T^*_{ik}>s)}{K(s)}\mathbf{d}M^c_i(s) \tag{4.2.11}$$

Plugging this in the second term we would like to have convergence for and applying Fubini's theorem, we get

$$\begin{aligned}&\int_0^u J_j(s)\left(\frac{\mathbf{d}N_{jj'}(s)}{Y_j(s)} - \frac{\mathbf{d}N^*_{jj'}(s)}{Y_j(s)}\right)\\ &= -\int_0^u \frac{J_j(s)}{Y_j(s)}\sum_i\sum_k I(T^*_{ik}\leq u, s_{ik}=j, s_{i,k+1}=j')\frac{I(T^*_{ik}>s)}{K(s)}\mathbf{d}M^c_i(s)\\ &= -\sum_i\sum_k I(T^*_{ik}\leq u, s_{ik}=j, s_{i,k+1}=j')\int_0^u \frac{J_j(s)}{Y_j(s)}\frac{I(T^*_{ik}>s)}{K(s)}\mathbf{d}M^c_i(s)\end{aligned} \tag{4.2.12}$$

All terms under the integral in this equation on the right-hand side are predictable. Therefore, that part is a Martingale with predictable variation process at time t

$$\left\langle\int_0^t \frac{J_j(s)}{Y_j(s)}\frac{I(C_i>s)}{K(s)}\mathbf{d}M^c_i(s)\right\rangle = \int_0^t \frac{J_j(s)}{Y_j(s)^2}\frac{I(T^*_{ik}>s)}{K(s)^2}I(T_i\geq s)\lambda_c(s)\mathbf{d}s \tag{4.2.13}$$

So the predictable variation process of the whole process is bounded by

$$\sum_i\sum_k I(T^*_{ik}\leq t, s_{ik}=j, s_{i,k+1}=j')\int_0^t \frac{J_j(s)}{Y_j(s)^2}\frac{I(T^*_{ik}>s)}{K(s)^2}I(T_i\geq s)\lambda_c(s)\mathbf{d}s \tag{4.2.14}$$

Applying Fubini's theorem again, we find

$$
\begin{aligned}
&\sum_i \sum_k I(T^*_{ik} \le t, s_{ik} = j, s_{i,k+1} = j') \int_0^t \frac{J_j(s)}{Y_j(s)^2} \frac{I(T^*_{ik} > s)}{K(s)^2} I(T_i \ge s)\lambda_c(s)\mathbf{d}s \\
&\le \int_0^t \frac{J_j(s)}{Y_j(s)^2 K(s)^2} \sum_i \sum_k I(T^*_{ik} \le t, s_{ik} = j, s_{i,k+1} = j', T^*_{ik} > s, T_i \ge s)\lambda_c(s)\mathbf{d}s \\
&\le \int_0^t \frac{J_j(s)}{Y_j(s)^2 K(s)^2} \sum_i \sum_k I(T^*_{ik} \le t, s_{ik} = j, s_{i,k+1} = j', T^*_{ik} > s)\lambda_c(s)\mathbf{d}s \\
&\le \int_0^t \frac{J_j(s)}{Y_j(s)^2 K(s)^2} Y^*_j(s)\lambda_c(s)\mathbf{d}s \le \frac{1}{K(y)^2} \int_0^t \frac{J_j(s)}{Y_j(s)^2} Y^*_j(s)\lambda_c(s)\mathbf{d}s \\
&\le \frac{1}{K(y)^2} \int_0^t \frac{J_j(s)}{Y_j(s)} \lambda_c(s)\mathbf{d}s \le \frac{1}{2nK(t)^2} \int_0^t \frac{I(\overline{y}^*_j(s) > 0)}{\overline{y}^*_j(s)} \lambda_c(s)\mathbf{d}s \\
&\xrightarrow{\mathbb{P}} 0 \text{ as } n \to \infty
\end{aligned}
$$

The last two inequalities follow from the proof of lemma 4.1.1 and the proof of the first term of this proof. Convergence to 0 now follows by the assumption we made that the integral is finite. Applying Lenglart's inequality like with the first term of this proof, we have convergence to 0 in probability for (4.2.9), uniformly on $[0, t]$.

Third term:

$$
\int_0^u J_j(s) \left(\frac{1}{Y_j(s)} - \frac{1}{Y^*_j(s)} \right) \mathbf{d}N^*_{jj'}(s) \tag{4.2.15}
$$

We may bound this term from above by

$$
\begin{aligned}
&\int_0^u J_j(s) \left| \frac{1}{Y_j(s)} - \frac{1}{Y^*_j(s)} \right| \mathbf{d}N^*_{jj'}(s) \\
&= \int_0^u J_j(s) \left| \frac{1}{Y_j(s)} - \frac{1}{Y^*_j(s)} \right| \mathbf{d}M^*_{jj'}(s) + \int_0^u J_j(s) \left| \frac{1}{Y_j(s)} - \frac{1}{Y^*_j(s)} \right| Y^*_j(s)\alpha_{jj'}(s)\mathbf{d}s
\end{aligned} \tag{4.2.16}
$$

Which, using predictability of the terms under the integral signs, is a submartingale on $[0, t]$ with compensator

$$
\begin{aligned}
L(t) &:= \int_0^t J_j(s) \left| \frac{1}{Y_j(s)} - \frac{1}{Y^*_j(s)} \right| Y^*_j(s)\alpha_{jj'}(s)\mathbf{d}s \\
&= \int_0^t J_j(s) \left| \frac{1}{\overline{Y}_j(s)} - \frac{1}{\overline{Y}^*_j(s)} \right| \overline{Y}^*_j(s)\alpha_{jj'}(s)\mathbf{d}s \\
&= \int_0^t J_j(s) \left| \frac{\overline{Y}_j(s) - \overline{Y}^*_j(s)}{\overline{Y}^*_j(s)\overline{Y}_j(s)} \right| \overline{Y}^*_j(s)\alpha_{jj'}(s)\mathbf{d}s \\
&= \int_0^t J_j(s) \left| \overline{Y}_j(s) - \overline{Y}^*_j(s) \right| \frac{1}{\overline{Y}_j(s)} \alpha_{jj'}(s)\mathbf{d}s \\
&\le \sup_{[0,t]} \left| \overline{Y}_j(s) - \overline{Y}^*_j(s) \right| \int_0^t \frac{J_j(s)}{\overline{Y}_j(s)} \alpha_{jj'}(s)\mathbf{d}s
\end{aligned} \tag{4.2.17}
$$

From lemma 4.1.1b and the Law of Large Numbers, $\sup_{[0,t]} |\overline{Y}_j(s) - \overline{Y}^*_j(s)| \xrightarrow{\mathbb{P}} 0$ as $n \to \infty$.

The second term is bounded in probability by the proof of convergence of (4.2.3) in this proof. Therefore, this compensator converges to 0 in probability as $n \to \infty$.

The first term on the right-hand side of (4.2.16) is a Martingale. Its predictable variation process at time t is given by

$$\left\langle \int_0^t J_j(s) \left| \frac{1}{Y_j(s)} - \frac{1}{Y_j^*(s)} \right| \mathbf{d}M_{jj'}^*(s) \right\rangle = \int_0^t J_j(s)^3 \left| \frac{1}{Y_j(s)} - \frac{1}{Y_j^*(s)} \right|^3 Y_j^*(s)\alpha_{jj'}(s)\mathbf{d}s \quad (4.2.18)$$

Similar algebra as used for the compensator of the Martingale then gives us

$$\begin{aligned} &\left\langle \int_0^t J_j(s) \left| \frac{1}{Y_j(s)} - \frac{1}{Y_j^*(s)} \right| \mathbf{d}M_{jj'}^*(s) \right\rangle \\ &\leq \left(\frac{1}{n} \sup_{[0,t]} \left(J_j(s) \left| \frac{1}{\overline{Y}_j(s)} - \frac{1}{\overline{Y}_j^*(s)} \right| \right) \right)^2 \sup_{[0,t]} \left| \overline{Y}_j(s) - \overline{Y}_j^*(s) \right| \int_0^t \frac{J_j(s)}{\overline{Y}_j(s)} \alpha_{jj'}(s)\mathbf{d}s \end{aligned} \quad (4.2.19)$$

The latter supremum and the integral have already been treated before. Now for the first supremum: first, note that $Y_j(s) \neq 0$ if $Y_j^*(s) \neq 0$ by the proof of lemma 4.1.1b. Second, applying the continuous mapping theorem to this term and using lemma 4.1.1b itself, we find that this term also converges to 0 in probability as $n \to \infty$. In other words, (4.2.19) converges to 0 in probability. Another application of Lenglart's inequality as we did in the proof for convergence of (4.2.3) now gives us the desired result. The proof is complete.

□

We have now proved consistency of the Nelson-Aalen estimator for censored data for processes that don't necessarily have to be Markov processes. In the next chapter, we use this result to prove consistency of the Aalen-Johansen estimator for state occupation probabilities for non-Markov systems.

5 The Aalen-Johansen estimator for non-Markov data

In this chapter, we aim to find a consistent estimator for the state occupation probabilities $p_j(t) = \mathbb{P}(s_i(t) = j)$, where $s_i(t)$ is the stage occupied by individual i at time t, with $j \in \mathscr{S}$. First, we find an estimator for this quantity based on the uncensored processes N^* and Y^* and use a well-known result to prove that this estimator is consistent. We then do the same derivation for the censored processes to arrive at the same estimator, taking into account when censoring might or might not happen, which complicates our calculations a lot. We finally prove that this estimator is consistent as well.

In [1], Datta & Satten oversimplify the way to obtain the estimator for censored processes, and they omit the proof altogether. Although the proof of this particular estimator is similar to the proof of theorem 4.2.1, the proof requires the result, and an additional step is necessary. Therefore, we will give a formal proof of the result.

5.1 State occupation probabilities for uncensored data

We first consider the situation where the data are uncensored. Let us define an estimator for $p_j(t)$:

$$\hat{p}_j^*(t) := \sum_{k=1}^{K} \frac{Y_k^*(0+)}{n} \hat{p}_{kj}^*(0,t) \tag{5.1.1}$$

with $\hat{p}_{kj}^*(0,t)$ the kjth element of the matrix $\hat{p}^*(0,t) = \prod_{(0,t]}(I + \mathbf{d}\hat{A}^*(s))$. Note the similarities of this estimator and the Kaplan-Meier estimator in section 2.3.2 and the version in section 3.6. This estimator is called the *Aalen-Johansen estimator* [4].

From the Law of Large Numbers, we have $\frac{Y_j^*(t+)}{n} \xrightarrow{\mathbb{P}} p_j(t)$ as $n \to \infty$. Therefore, if we have $\hat{p}_j^*(t) = \frac{Y_j^*(t+)}{n}$, we have convergence in probability for $\hat{p}_j^*(t)$ to $p_j(t)$. It is possible to show that this is true. We obtain the results by working out the details of (5.1.2) (see [9], (section IV.4.1.4)). De details are found below.

$$Y_j^*(t+) = Y_j^*(0+) + \sum_{j' \neq j} N_{j'j}^*(t) - \sum_{j' \neq j} N_{jj'}^*(t) \tag{5.1.2}$$

To derive (5.1.1), remember the identity for $\hat{A}^*(t)$ (see (4.1.1)), assume $j' \neq j$, and note that $J_j(s)$ is only necessary when we divide by $Y_j^*(s)$ (because otherwise it doesn't do anything). We find

$$N_{jj'}^*(t) = \int_0^t Y_j^*(s)\mathbf{d}\hat{A}_{jj'}^*(s) \tag{5.1.3}$$

Plugging in in (5.1.2) and using the second identity from (4.1.1), we find

$$
\begin{aligned}
Y_j^*(t+) &= Y_j^*(0+) + \sum_{j' \neq j} \int_0^t Y_{j'}^*(s) \mathbf{d}\hat{A}_{j'j}^*(s) - \sum_{j' \neq j} \int_0^t Y_j^*(s) \mathbf{d}\hat{A}_{jj'}^*(s) \\
&= Y_j^*(0+) + \sum_{j' \neq j} \int_0^t Y_{j'}^*(s) \mathbf{d}\hat{A}_{j'j}^*(s) + \int_0^t Y_j^*(s) \mathbf{d}\hat{A}_{jj}^*(s) \\
&= Y_j^*(0+) + \sum_{j'=1}^{K} \int_0^t Y_{j'}^*(s) \mathbf{d}\hat{A}_{j'j}^*(s) \\
&= Y_j^*(0+) + \left(\int_0^t \mathbf{Y}^*(s) \mathbf{d}\hat{A}^*(s) \right)_j
\end{aligned}
\tag{5.1.4}
$$

where $\mathbf{Y}^*(s) = (Y_1^*(s), \ldots, Y_K^*(s))$ is a row-vector containing all elements $Y_j^*(s)$. Combining all $j \in \mathscr{S}$, we find

$$
\mathbf{Y}^*(t+) = \mathbf{Y}^*(0+) + \int_0^t \mathbf{Y}^*(s) \mathbf{d}\hat{A}^*(s) \tag{5.1.5}
$$

This integral equation is of the same form as the one in Theorem 3.3.1. Applying this result, we find for $\mathbf{Y}^*(t+)$:

$$
\mathbf{Y}^*(t+) = \mathbf{Y}^*(0+) \prod_{(0,t]} (I + \mathbf{d}\hat{A}^*(s)) \tag{5.1.6}
$$

Hence

$$
\frac{Y_j^*(t+)}{n} = \sum_{k=1}^{K} \frac{Y_k^*(0+)}{n} \hat{p}_{kj}^*(0,t) = \hat{p}_j^*(t) \tag{5.1.7}
$$

and the identity $\hat{p}_j^*(t) = \frac{Y_j^*(t+)}{n}$ is justified.

5.2 State occupation probabilities for censored data

Having found a consistent estimator in case of uncensored data, it seems natural to assume that an estimator of the same form for censored data works as well, i.e. replacing the uncensored quantities $\hat{A}^*$ and Y^* with $\hat{A}$ and Y respectively. We find

$$
\hat{p}_j(t) = \sum_{k=1}^{K} \frac{Y_k(0+)}{n} \hat{p}_{kj}(0,t) \tag{5.2.1}
$$

with $\hat{p}(0,t) = \prod_{(0,t]} (I + \mathbf{d}\hat{A}(s))$. Note how this is the Aalen-Johansen estimator for censored data.

Deriving this estimator is not as simple as deriving (5.1.1). We cannot use the balance equation (5.1.2) with the censored qualities. Not only do we use a quantity in those processes that we do not observe, we have to take the people who are censored while in state j into account as well. For these reasons, to get to this estimator, some more work is required.

Define the following quantities:

$$
\begin{aligned}
N_{jj'}^{uc}(t) &:= \sum_{i=1}^{n}\sum_{k\geq 1} I(T_{ik}^* \leq t, C_i \geq T_{ik}^*, s_{ik}=j, s_{i,k+1}=j') \\
Y_{j}^{uc}(t) &:= \sum_{i=1}^{n}\sum_{k\geq 1} I(T_{i,k-1}^* < t \leq T_{ik}^*, C_i \geq T_{ik}^*, s_{ik}=j) \\
N_{j}^{c}(t) &:= \sum_{i=1}^{n}\sum_{k\geq 1} I(T_{i,k-1}^* < C_i \leq T_{ik}^*, C_i \leq t, s_{ik}=j)
\end{aligned}
\tag{5.2.2}
$$

And define the estimator for the hazard

$$
\hat{A}_{jj'}^{uc}(t) = \begin{cases} \int_0^t J_j^{uc}(u) Y_j^{uc}(u)^{-1} \mathbf{d} N_{jj'}^{uc}(u) & j \neq j' \\ -\sum_{j'\neq j} \hat{A}_{jj'}^{uc}(t) & j = j' \end{cases}
\tag{5.2.3}
$$

with $J_j^{uc}(u) = I(Y_j^{uc}(u) > 0)$. Note how these quantities are the the quantities Y, N without division by $K(.)$, and also note that $\hat{A}^{uc}$ equals $\hat{A}$ for this reason.

Now we can write a balance equation like (5.1.2) with these quantities:

$$
Y_j^{uc}(t+) = Y_j^{uc}(0+) + \sum_{j'\neq j} N_{j'j}^{uc}(t) - \sum_{j'\neq j} N_{jj'}^{uc}(t) - N_j^c(t)
\tag{5.2.4}
$$

Like the uncensored data case, we may rewrite all terms $N_{j'j}^{uc}(t)$ and $N_{jj'}^{uc}(t)$ to $\int_0^t Y_{j'}^{uc}(s)\mathbf{d}\hat{A}_{j'j}^{uc}(s)$ and $\int_0^t Y_j^{uc}(s)\mathbf{d}\hat{A}_{jj'}^{uc}(s)$ respectively, and we may use the definition of $\hat{A}_{jj'}^{uc}(s)$ to find the following equation:

$$
Y_j^{uc}(t+) = Y_j^{uc}(0+) + \sum_{j'=1}^{K} \int_0^t Y_{j'}^{uc}(s) \hat{A}_{j'j}^{uc}(s) - N_j^c(t)
\tag{5.2.5}
$$

Writing in vector form like we did for the uncensored data case and noting that $\sum_{j'=1}^{K}\int_0^t Y_{j'}^{uc}(s)\mathbf{d}\hat{A}_{j'j}^{uc}(s) = \left(\int_0^t \mathbf{Y}^{uc}(s)\mathbf{d}\hat{A}^{uc}(s)\right)_j$, we find

$$
\mathbf{Y}^{uc}(t+) = \mathbf{Y}^{uc}(0+) + \int_0^t \mathbf{Y}^{uc}(s)\mathbf{d}\hat{A}^{uc}(s) - \mathbf{N}^c(t)
\tag{5.2.6}
$$

Where $\mathbf{N}^c(t) = (N_1^c(t), \ldots, N_K^c(t))$

Now note that the hazard for the counting process $\mathbf{N}^c$ is just the censoring hazard. According to formula (3.20) in [8], $\displaystyle\int_0^t \frac{J_j^{uc}(s)}{Y_j^{uc}(s)}\mathbf{d}N_j^c(s)$ is an unbiased estimator for $\Lambda_c(t)$. Rewriting and noting that $J_j^{uc}(s) > 0$ whenever $Y_j^{uc}(s) > 0$, we find

$$\mathbf{N}^c = \int_0^t \mathbf{Y}^{uc}(s)\mathbf{d}\mathbf{\Lambda}_c(s) \tag{5.2.7}$$

where $\mathbf{\Lambda}_c(s) = I_K\Lambda_c(s)$, with I_K the $K \times K$ identity matrix.

Now (5.2.6) can be written the following way:

$$\mathbf{Y}^{uc}(t+) = \mathbf{Y}^{uc}(0+) + \int_0^t \mathbf{Y}^{uc}(s)\mathbf{d}(\hat{A}^{uc} - \mathbf{\Lambda}_c)(s) \tag{5.2.8}$$

This takes the form of a Volterra equation like before. Therefore, with theorem 3.3.1, we find

$$\mathbf{Y}^{uc}(t+) = \mathbf{Y}^{uc}(0+) \prod_{(0,t]} (I + \mathbf{d}(\hat{A}^{uc} - \mathbf{\Lambda}_c)(s)) \tag{5.2.9}$$

Now we note that $\mathbf{Y}^{uc}(t) = \mathbf{Y}(t)\mathbf{K}(t-)$ with $\mathbf{K}(t-) = I_K K(t-)$. We also observe that $\mathbf{K}(0) = I_K$. Finally, remember that $\mathbf{K}(t) = \prod_{(0,t]}(I - \mathbf{d}\mathbf{\Lambda}_c(s))$. Plugging this in (5.2.9) and multiplying both sides by $(\mathbf{K}(t))^{-1}$ gives us

$$\mathbf{Y}(t+) = \mathbf{Y}(0+) \prod_{(0,t]} \left(I + \mathbf{d}(\hat{A}^{uc} - \mathbf{\Lambda}_c)(s)\right) \left(\prod_{(0,t]} (I - \mathbf{d}\mathbf{\Lambda}_c(s))\right)^{-1} \tag{5.2.10}$$

To get to (5.2.1) from (5.2.10), we need two more things.

First, we need $\hat{A}^{uc}(s) = \hat{A}(s)$, which we have seen before, and which is clear from the definition of both quantities.

Second, we need

$$\prod_{(0,t]} \left(I + \mathbf{d}(\hat{A} - \mathbf{\Lambda}_c)(s)\right) \left(\prod_{(0,t]} (I - \mathbf{d}\mathbf{\Lambda}_c(s))\right)^{-1} = \prod_{(0,t]} \left(I + \mathbf{d}\hat{A}(s)\right) \tag{5.2.11}$$

To do this, we first apply the Duhamel equation, corollary 3.3.3, to the left-hand-side of the desired equation. We get the following:

$$\begin{aligned} &\prod_{(0,t]} \left(I + \mathbf{d}(\hat{A} - \mathbf{\Lambda}_c)(s)\right) \left(\prod_{(0,t]} (I - \mathbf{d}\mathbf{\Lambda}_c(s))\right)^{-1} \\ &= I + \int_0^t \prod_{(0,s)} (I + \mathbf{d}(\hat{A} - \mathbf{\Lambda}_c)(u))\mathbf{d}\hat{A}(s) \left(\prod_{(0,s]} (I - \mathbf{d}\mathbf{\Lambda}_c)(u))\right)^{-1} \end{aligned} \tag{5.2.12}$$

Noting that $\left(\prod_{(0,s]}(I - \mathbf{d}\mathbf{\Lambda}_c)(s))\right)^{-1}$ is a diagonal matrix with the same entry on every diagonal element and defining
$Z(t) := \prod_{(0,t]} \left(I + \mathbf{d}(\hat{A} - \mathbf{\Lambda}_c)(s)\right) \left(\prod_{(0,t]} (I - \mathbf{d}\mathbf{\Lambda}_c(s))\right)^{-1}$, we may rewrite (5.2.12) to

$$Z(t) = I + \int_0^t Z(s-)\mathbf{d}\hat{A}(s) \tag{5.2.13}$$

which is the Volterra equation we have seen before, with solution

$$Z(t) = \prod_{(0,t]} \left(I + \mathbf{d}\hat{A}(s)\right) \tag{5.2.14}$$

according to theorem 3.3.1, and now (5.2.10) becomes

$$\mathbf{Y}(t+) = \mathbf{Y}(0+) \prod_{(0,t]} (I + \mathbf{d}\hat{A}(s)) \tag{5.2.15}$$

as desired. Dividing all components of $\mathbf{Y}(t+)$ by n will now finally give algebraic equivalence of $\hat{p}_j(t)$ and $Y_j(t+)/n$. Therefore, we have really derived this estimator for $p_j(u)$ using the observed data, instead of only replacing the uncensored processes with the censored processes in (5.1.1).

5.3 Consistency of the estimator with censored processes

Having proved identity (5.2.15), that is, algebraic equivalence of $\frac{Y_j(t)}{n}$ and $\hat{p}_j(t)$, the Law of Large Numbers could be used to prove convergence in probability of $\hat{p}$, like we did for $\hat{p}^*$. Below, we present a different proof, based on and using (the proof of) theorem 4.2.1.

Theorem 5.3.1. *Assume all conditions of theorem* 4.2.1*. Then $\hat{p}_j(u)$ is uniformly consistent in probability for $p_j(u)$ on* $[0, t]$*.*

Proof. Let $\hat{p}(u) := (\hat{p}_1(u), \ldots, \hat{p}_K(u))$. Similarly, let $\hat{p}^*(u) := (\hat{p}_1^*(u), \ldots, \hat{p}_K(u))$. Furthermore, let $\tilde{p}^*(u) := \frac{\mathbf{Y}^*(0+)}{n} \prod_{(0,u]}(I + \mathbf{d}\tilde{A}^*(s))$ where $\tilde{A}^*(s)$ is defined as in the proof of theorem 4.2.1. Also, denote $p(u) = (p_1(u), \ldots, p_K(u))$. We find the following:

$$\sup_{[0,t]} \|\hat{p}(u) - p(u)\| \le \sup_{[0,t]} \|\hat{p}(u) - \tilde{p}^*(u)\| + \sup_{[0,t]} \|\tilde{p}^*(u) - \hat{p}^*(u)\| + \sup_{[0,t]} \|\hat{p}^*(u) - p(u)\| \tag{5.3.1}$$

The last term on the right-hand side of (5.3.1) converges to 0 in probability by the Law of Large Numbers as $n \to \infty$, as we have seen before in section 5.1. If we can prove convergence in probability to 0 for the first two terms of the right-hand side, we will have proved the theorem.

First term:

$$\sup_{[0,t]} \|\hat{p}(u) - \tilde{p}^*(u)\| \tag{5.3.2}$$

From (5.2.15), we find

$$\begin{aligned}
\hat{p}(u) - \tilde{p}^*(u) &= \frac{\mathbf{Y}(0+)}{n} \prod_{(0,u]} (I + \mathbf{d}\hat{A}(s)) - \frac{\mathbf{Y}^*(0+)}{n} \prod_{(0,u]} (I + \mathbf{d}\tilde{A}^*(s)) \\
&= \frac{\mathbf{Y}(0+)}{n} \prod_{(0,u]} (I + \mathbf{d}\hat{A}(s)) - \frac{\mathbf{Y}^*(0+)}{n} \prod_{(0,u]} (I + \mathbf{d}\hat{A}(s)) \\
&+ \frac{\mathbf{Y}^*(0+)}{n} \prod_{(0,u]} (I + \mathbf{d}\hat{A}(s)) - \frac{\mathbf{Y}^*(0+)}{n} \prod_{(0,u]} (I + \mathbf{d}\tilde{A}^*(s)) \\
&= \left(\frac{\mathbf{Y}(0+)}{n} - \frac{\mathbf{Y}^*(0+)}{n}\right) \prod_{(0,u]} \left(I + \mathbf{d}\hat{A}(s)\right) + \frac{\mathbf{Y}(0+)}{n} \left(\prod_{(0,u]} (I + \mathbf{d}\hat{A}(s)) - \prod_{(0,u]} (I + \mathbf{d}\tilde{A}^*(s))\right)
\end{aligned} \tag{5.3.3}$$

Since $\frac{\mathbf{Y}^*(0+)}{n} = \frac{\mathbf{Y}(0+)}{n}$ (because $\mathbb{P}(C_i > 0) = 1$ for all i because $K(0) = 1$), the first term on the right-hand side of (5.3.3) cancels. For the second term, we may use the continuous mapping theorem to prove that (5.3.2) converges to 0 in probability if $\hat{A}(s)$ converges to $\tilde{A}^*(s)$ in probability. But this is a result that we directly proved while proving theorem 4.2.1. Therefore, (5.3.2) converges to 0 in probability as $n \to \infty$.

Second term:

$$\sup_{[0,t]} \|\tilde{p}^*(u) - \hat{p}^*(u)\| \tag{5.3.4}$$

We have

$$\tilde{p}^*(u) - \hat{p}^*(u) = \frac{\mathbf{Y}^*(0+)}{n} \left(\prod_{(0,u]} (I + \mathbf{d}\tilde{A}^*(s)) - \prod_{(0,u]} (I + \mathbf{d}\hat{A}^*(s)) \right) \tag{5.3.5}$$

Like with term 1, we want to use the continuous mapping theorem to prove convergence in probability to 0. To that end, we must prove that $\hat{A}^*(s) \xrightarrow{\mathbb{P}} \tilde{A}^*(s)$.

Let $j' \neq j$, we find

$$\begin{aligned}
&\hat{A}^*_{jj'}(s) - \tilde{A}^*_{jj'}(s) = \int_0^u J_j(s)\alpha_{jj'}(s)\mathbf{d}s - \int_0^u \frac{J_j(s)}{Y_j^*(s)}\mathbf{d}N^*_{jj'}(s) \\
&= \int_0^u \frac{J_j(s)}{Y_j^*(s)}\alpha_{jj'}(s)Y_j^*(s)\mathbf{d}(s) - \int_0^u \frac{J_j(s)}{Y_j^*(s)}\mathbf{d}N^*_{jj'}(s) \\
&= -\int_0^u \frac{J_j(s)}{Y_j^*(s)}\mathbf{d}M^*_{jj'}(s)
\end{aligned} \tag{5.3.6}$$

But this is (4.2.3), for which we have already proved convergence to 0 in probability in theorem 4.2.1. Therefore, by the continuous mapping theorem, (5.3.4) converges to 0 in probability as well.

As now all terms converge to 0 in probability for every $u \in [0,t]$, $\hat{p}(u)$ is uniformly consistent for $p(u)$ on $[0,t]$. This is what we wanted to prove.

□

6 Asymptotics of the Aalen-Johansen estimator

In this chapter, we explore some properties of the Aalen-Johansen estimator for censored processes. Mainly, we are interested in the distribution and the variance of the estimator. Although we proved that the estimator is asymptotically consistent, it would be nice to find a formula on how far off our estimator could reasonably be. This was done by David Glidden in 2002 [3]. We will use the quantities he defined to adapt his proof for the consistency of his estimator for the covariance function for the distribution of the Aalen-Johansen estimator. To do this, we first recall and redefine a couple of processes and add some new ones along the way. After we have proved the consistency of this covariance function, we propose an estimator for the variance.

6.1 Additional quantities of the model

To find the formula for the consistency of the covariance function, Glidden defined counting processes and at risk processes for a multistate model in a slightly different way than Datta & Satten did in [1]. Although the eventual quantities in the Nelson-Aalen estimator and the Aalen-Johansen estimator are the same, he uses these other quantities for the distribution of the Aalen-Johansen estimator and for the (estimator of the) covariance function of this estimator. To be able to work out his proof, we give the same definitions here, unless they are equivalent to the definitions in [1]. In that case, we just recall the definition. Throughout this chapter, we assume $j \neq j'$ unless stated otherwise.

We first define the counting processes and at risk processes for individual i.

$$N^*_{i;jj'}(t) := \sum_{k\geq 1} I(T^*_{ik} \leq t, s_{ik} = j, s_{i,k+1} = j') \tag{6.1.1}$$

$$N_{i;jj'}(t) := \int_0^t I(C_i \geq s)\mathbf{d}N^*_{i;jj'}(s) \tag{6.1.2}$$

$$Y^*_{ij}(t) := \sum_{k\geq 1} I(T^*_{i,k-1} < t \leq T^*_{ik}, s_{ik} = j) \tag{6.1.3}$$

$$Y_{ij}(t) := Y^*_{ij}(t)I(C_i \geq t) \tag{6.1.4}$$

Note that we just recalled the same definitions for $N^*_{i;jj'}$ as the one in chapter 4.

From these definitions, we define n $K \times K$ matrices $N_i(t)$ and $Y_{iD}(t)$ which are i.i.d.. The (j, j')th element of $N_i(t)$ is $N_{i;jj'}(t)$ and its jth diagonal element is $-\sum_{j'\neq j} N_{i;jj'}(t)$ where $j = 1, \ldots, K$. Furthermore, $Y_{iD}(t)$ is a diagonal matrix with jth diagonal element $Y_{ij}(t)$. These matrices are essentially the counting process matrix and the at risk matrix for individual i.

Recalling the definition for the transition hazard α (see (3.5.2)) and how we used Fubini's theorem for this hazard, we get the transition hazard

$$\alpha_{jj'}(t) = \lim_{\mathbf{d}t\downarrow 0} \frac{\mathbb{E}[\mathbf{d}N^*_{i;jj'}(t)|s_i(t-) = j]}{\mathbf{d}t} = \lim_{\mathbf{d}t\downarrow 0} \frac{\mathbb{E}[\mathbf{d}N^*_{i;jj'}(t)|Y^*_{ij}(t) = 1]}{\mathbf{d}t} \tag{6.1.5}$$

which has the same cumulative transition hazard as before:

$$A_{jj'}(t) = \int_0^t \alpha_{jj'}(s)\mathbf{d}(s) \tag{6.1.6}$$

$$A_{jj}(t) = -\sum_{j' \neq j} A_{jj'}(t) \tag{6.1.7}$$

Because of the proof of theorem 5.3.1, we can see that

$$p(t) = \frac{\mathbf{Y}(0+)}{n} \prod_{(0,t]} (I + \mathbf{d}A(s)) \tag{6.1.8}$$

with help from the continuous mapping theorem and theorem 4.2.1.

Because we want to sum over individuals, it seems natural to sum the matrices of all individuals and define the Aalen-Johansen on the result. Let $\overline{Y}_j(t) = n^{-1} \sum_{i=1}^n Y_{ij}(t)$, $\overline{N}_{jj'}(t) = n^{-1} \sum_{i=1}^n N_{i;jj'}(t)$. Now let $\overline{N}(t) := n^{-1} \sum_{i=1}^n N_i(t)$, $\overline{Y}_D(t) := n^{-1} \sum_{i=1}^n Y_{iD}(t)$. We get the following estimator of A:

$$\hat{A}(t) = \int_0^t \overline{Y}_D^{-1}(s) \mathbf{d}\overline{N}(s) \tag{6.1.9}$$

Which leads to the estimator

$$\hat{p}(t) := \frac{\mathbf{Y}(0+)}{n} \prod_{(0,t]} (I + \mathbf{d}\hat{A}(s)) \tag{6.1.10}$$

of $p(t)$. Note that this is equivalent to the definition of the Aalen-Johansen estimator in (5.2.1).

6.2 The distribution function and covariance function

The quantity of interest is now $n^{\frac{1}{2}}(\hat{p} - p)(t)$, because results for the distribution of this quantity will give us information about the distribution of $\hat{p}$.

In appendix A of [3], Glidden showed that $n^{\frac{1}{2}}(\hat{p} - p)(t)$ converges weakly to $\mathcal{G}(t) = \{\mathcal{G}_1(t), \ldots, \mathcal{G}_K(t)\}$ which is a vector of Gaussian processes. Furthermore, he showed $n^{\frac{1}{2}}(\hat{p} - p)(t)$ has the same distribution as the process $W(t) = \{W_1(t), \ldots, W_K(t)\} =: n^{-\frac{1}{2}} \sum \Phi_i(t)$, with

$$\Phi_i(t) := \frac{\mathbf{Y}(0+)}{n} \int_0^t \prod_{(0,s)} (I + \mathbf{d}A(u)) \mathbf{d}\Psi_i(s) \prod_{(s,t]} (I + \mathbf{d}A(u)) \tag{6.2.1}$$

where

$$\Psi_i(t) := \int_0^t \mathcal{Y}_D^{-1}(s)(\mathbf{d}N_i(s) - Y_{iD}(s)\mathbf{d}A(s)) \tag{6.2.2}$$

and

$$\mathcal{Y}_D(s) = \lim_{n \to \infty} \overline{Y}_D(s) \tag{6.2.3}$$

where $0 \leq s \leq t \leq \tau$. He did so by using the compact differentiability of the integral and product integral at these quantities. Because we don't have to make any changes to the proof because every step is quite clear, we omit it here.

The elements of $\mathcal{G}(t)$ have covariance $\xi_{jj'}(s,t) =: \mathrm{Cov}\{\mathcal{G}_{\mathrm{j}}(\mathrm{s}), \mathcal{G}_{\mathrm{j'}}(\mathrm{t})\} = \mathbb{E}\{\Phi_{1\mathrm{j}}(\mathrm{s})\Phi_{1\mathrm{j'}}(\mathrm{t})\}$ where $j = 1, \ldots, K$ and $j' = 1, \ldots, K$. This is not difficult to see:

$$
\begin{aligned}
&\mathrm{Cov}\{\mathcal{G}_{\mathrm{j}}(\mathrm{s}), \mathcal{G}_{\mathrm{j'}}(\mathrm{t})\} \\
&= \mathbb{E}(n^{-1}\sum_{i=1}^{n}\sum_{i'=1}^{n}\Phi_{ij}(s)\Phi_{i'j'}(t)) - \mathbb{E}(n^{-\frac{1}{2}}\sum_{i=1}^{n}\Phi_{ij}(s))\mathbb{E}(n^{-\frac{1}{2}}\sum_{i=1}^{n}\Phi_{ij'}(t)) \\
&= \mathbb{E}(n^{-1}\sum_{i=1}^{n}\Phi_{ij}(s)\Phi_{ij'}(t)) - n^{-1}\sum_{i=1}^{n}\mathbb{E}(\Phi_{ij}(s))\mathbb{E}(\Phi_{ij'}(t)) \\
&= n^{-1}\sum_{i=1}^{n}\mathbb{E}(\Phi_{ij}(s)\Phi_{ij'}(t)) \\
&= n^{-1}n\mathbb{E}(\Phi_{1j}(s)\Phi_{1j'}(t))
\end{aligned}
$$

Because the Gaussian processes have mean 0, because all Φ_i are i.i.d..

The unknown quantities used in (6.2.1) can be estimated. To do so, it seems natural to replace $\mathcal{Y}_D$ by $\overline{Y}_D$ and A by $\hat{A}$. To estimate the covariance function, we now define

$$\hat{\Phi}_i(t) = \frac{\mathbf{Y}(0+)}{n}\int_0^t \prod_{(0,s)}(I + \mathbf{d}\hat{A}(u))\mathbf{d}\hat{\Psi}_i(s)\prod_{(s,t]}(I + \mathbf{d}\hat{A}(u)) \tag{6.2.4}$$

$$\hat{\Psi}_i(t) = \int_0^t \overline{Y}_D^{-1}(s)(\mathbf{d}N_i(s) - Y_{iD}(s)\mathbf{d}\hat{A}(s)) \tag{6.2.5}$$

A consistent estimator for the covariance function is then given by

$$\hat{\xi}_{jj'}(s,t) := n^{-1}\sum_{i=1}^{n}\hat{\Phi}_{ij}(s)\hat{\Phi}_{ij'}(t) \tag{6.2.6}$$

This statement was proved in [3] in appendix B. Because some steps were unclear in his proof, we shall give a slightly modified proof later on in this chapter. First, we define several matrix norms.

Definition 6.2.1 (Matrix norms)**.** *The* norm *of a $K \times K$ matrix A is given by the formula*

$$\|A\| = \max_{1\le j\le K}\sum_{j'=1}^{K}|A_{jj'}| \tag{6.2.7}$$

The supremum norm of a $K \times K$ matrix on $[0,\tau]$ is given by

$$\|B\|_\infty = \sup_{t\in[0,\tau]}\|B(t)\| \tag{6.2.8}$$

Furthermore, the variation norm is given by

$$\|C\|_v = \max_{1\le j\le k}\left(\sum_{j'=1}^{K}|C_{jj'}|^v\right)^{\frac{1}{v}} \tag{6.2.9}$$

With the help of these norms, consistency of $\hat{\xi}$ can be proved.

Theorem 6.2.1. *The covariance estimator $\hat{\xi}_{jj'}$ is consistent for all $j, j' = 1, \ldots, K$*

Proof. Define

$$\xi(s,t) = \mathbb{E}(\Phi_1(s)^T\Phi_1(t)) = n^{-1}\sum_{i=1}^{n}\Phi_i(s)^T\Phi_i(t) \tag{6.2.10}$$

and

$$\hat{\xi}(s,t) = n^{-1}\sum_{i=1}^{n}\hat{\Phi}_i(s)^T\hat{\Phi}_i(t) \tag{6.2.11}$$

where $\xi_{jj'}(s,t)$ and $\hat{\xi}_{jj'}(s,t)$ are the elements of 6.2.10 and 6.2.11. To prove consistency of $\hat{\xi}_{jj'}(s,t)$, we have to prove that

$$\left\| n^{-1}\sum_{n=1}^{n}\hat{\Phi}_i(s)^T\hat{\Phi}_i(t) - \Phi_i(s)^T\Phi_i(t)\right\| \xrightarrow{\mathbb{P}} 0 \tag{6.2.12}$$

as $n \to \infty$. To show this, we have to show the following three properties:

$$\max_{1\le i\le n}\|\hat{\Phi}_i(t)\| = O_q(1) \tag{6.2.13}$$

$$\max_{1\le i\le n}\|\Phi_i(t)\| = O_q(1) \tag{6.2.14}$$

$$\max_{1\le i\le n}\|\hat{\Phi}_i(t) - \Phi_i(t)\| \xrightarrow{\mathbb{P}} 0 \text{ as n } \to \infty \tag{6.2.15}$$

Where q denotes the outer probability. This is easy to see, since there exists n_0 such that for all $n > n_0$:

$$\begin{aligned}
&\left\| n^{-1}\sum_{i=1}^{n}\hat{\Phi}_i(s)^T\hat{\Phi}_i(t) - \Phi_i(s)^T\Phi_i(t)\right\| \\
&\le \left\| n^{-1}\sum_{i=1}^{n}\hat{\Phi}_i(s)^T\hat{\Phi}_i(t) - \hat{\Phi}_i(s)^T\Phi_i(t)\right\| + \left\| n^{-1}\sum_{i=1}^{n}\hat{\Phi}_i(s)^T\Phi_i(t) - \Phi_i(s)^T\Phi_i(t)\right\| \\
&\le M\left\| n^{-1}\sum_{i=1}^{n}\hat{\Phi}_i(t) - \Phi_i(t)\right\| + M\left\| n^{-1}\sum_{i=1}^{n}\hat{\Phi}_i(s)^T - \Phi_i(s)^T\right\| \xrightarrow{\mathbb{P}} 0
\end{aligned}$$

as $n \to \infty$, if (6.2.13), (6.2.14) and (6.2.15) hold. Here, $M = \max\{M_1, M_2\}$, where M_1 and M_2 are the quantities such that $\mathbb{P}(\max_{1\le i\le n}\|\hat{\Phi}_i(t)\| > M_1) \xrightarrow{q} 0$ and $\mathbb{P}(\max_{1\le i\le n}\|\Phi_i(t)\| > M_2) \xrightarrow{q} 0$ as $n \to \infty$.

Since the elements of $\hat{\Lambda}$ are all monotone and bounded in probability (by the assumption that there are finitely many transitions for each individual), we find that the same is true for $\|\hat{A}\|_\infty$ and $\|\hat{A}\|_v$, i.e. they are $O_q(1)$. Since $\overline{Y}_D^{-1}$, $\mathbf{d}N_i$ and Y_{iD} are bounded as well for all $i = 1, \ldots, n$, we find that $\max_{1\le i\le n}\|\hat{\Psi}_i\|_v =: \|\hat{\Psi}\|_v$ is $O_q(1)$ as well.

Let $v > 1$, and let $w > 1$ such that $\frac{1}{v} + \frac{1}{w} = 1$, and assume without loss of generality that $v > w$. Then, for (6.2.13), we have the following:

$$\begin{aligned}
&\max_{1\le i\le n} \|\hat{\Phi}_i(t)\| = \max_{1\le i\le n} \left\| \frac{\mathbf{Y}(0+)}{n} \int_0^t \prod_{(0,s)} (I + \mathbf{d}\hat{A}(u)) \mathbf{d}\hat{\Psi}_i(s) \prod_{(s,t]} (I + \mathbf{d}\hat{A}(u)) \right\| \\
&\le \left\| \prod_{(0,t]} (I + \mathbf{d}\hat{A}(u)) \right\|_\infty \left\|\hat{\Psi}\right\|_v \left\| \prod_{(0,t]} (I + \mathbf{d}\hat{A}(u)) \right\|_w \\
&\le K^{\frac{1}{w}-\frac{1}{v}} \left\| \prod_{(0,t]} (I + \mathbf{d}\hat{A}(u)) \right\|_\infty \left\|\hat{\Psi}\right\|_v \left\| \prod_{(0,t]} (I + \mathbf{d}\hat{A}(u)) \right\|_v \\
&\le K^2 \left\| \prod_{(0,t]} (I + \mathbf{d}\hat{A}(u)) \right\|_\infty \left\|\hat{\Psi}\right\|_v \left\| \prod_{(0,t]} (I + \mathbf{d}\hat{A}(u)) \right\|_v
\end{aligned} \tag{6.2.16}$$

because of the Hölder inequality.

For $\| \prod_{(0,t]}(I + \mathbf{d}\hat{A}(u))\|_v$, we have
$\| \prod_{(0,t]}(I + \mathbf{d}\hat{A}_{jj'}(u))\|_v = \sup_{1\le j\le K} \left(\sum_{j'=1}^K \prod_{(0,t]}(I + \mathbf{d}\hat{A}_{jj'}(u))^v\right)^{\frac{1}{v}} \le K^{\frac{1}{v}} \| \prod_{(0,t]}(I + \mathbf{d}\hat{A}(u)\|_\infty$.

Now $\| \prod_{(0,t)}(I + \mathbf{d}\hat{A})\|_\infty$ is bounded from above by $f(t)\exp(K|\hat{A}'|)$, with $f(t) := (1 + KC^*)^t$ where $C^* = \sup_{t\in[0,\tau]} \max_{i,j\in\{1,\ldots,K\}} \hat{A}_{jj'}(t)$ and where $\hat{A}'(t) = \max_{j,j'\in\{1,\ldots,K\}} \hat{A}_{jj'}(t)$ with $\hat{A}' = \sup_{t\in[0,\tau]} A'(t)$. This can be seen the following way, if we define T to be the set of all transition times $0 < t_1 < t_2 < \ldots < t_k \le \tau$ with $t_0 = 0$:

$$\prod_{(0,t]} (I + \mathbf{d}\hat{A}(u)) = \prod_{t_i\le t, t_i\in T} (I + \Delta\hat{A}(t_i)) \tag{6.2.17}$$

where $\Delta\hat{A}(t_i) = \hat{A}(t_i) - \hat{A}(t_{i-1})$. With the non-negativity and monotonicity of $\hat{A}$, we find the following:

$$\begin{aligned}
&\Delta\hat{A}_{jj'}(t_i) = \hat{A}_{jj'}(t_i) - \hat{A}_{jj'}(t_{i-1}) \\
&= \hat{A}_{jj'}(t_i) - \hat{A}'(t_{i-1}) + \hat{A}'(t_{i-1}) - \hat{A}_{jj'}(t_{i-1}) \\
&\le \hat{A}'(t_i) - \hat{A}'(t_{i-1}) + \hat{A}'(t_{i-1}) - \hat{A}_{jj'}(t_{i-1}) \le \hat{A}'(t_i) - \hat{A}'(t_{i-1}) + C^*
\end{aligned} \tag{6.2.18}$$

which, with again the help of the monotonicity of $\hat{A}$ (and $\hat{A}'$), implies

$$\left\|\Delta\hat{A}(t_i)\right\|_\infty \le K|\hat{A}'(t_i)| - K|\hat{A}'(t_{i-1})| + |C^*| \tag{6.2.19}$$

Plugging this all in in (6.2.17), we get

$$\begin{aligned}
&\left\| \prod_{(0,t]} (I + \mathbf{d}\hat{A}(u)) \right\|_\infty = \left\| \prod_{t_i\le t, t_i\in T} (I + \Delta\hat{A}(t_i)) \right\|_\infty \\
&\le \prod_{t_i\le t, t_i\in T} \left\|(I + \Delta\hat{A}(t_i))\right\|_\infty \le \prod_{t_i\le t, t_i\in T} (1 + \left\|\Delta\hat{A}(t_i)\right\|_\infty) \\
&\le \prod_{t_i\le t, t_i\in T} (1 + K\Delta|\hat{A}'(t_i)| + KC^*) \le \prod_{(0,t]} (1 + K\mathbf{d}|\hat{A}'(s)|) \prod_{(0,t]} (1 + KC^*) \\
&\le f(t)\exp(K|\hat{A}'|)
\end{aligned} \tag{6.2.20}$$

Where for the second to last inequality an equality is used that we proved in theorem 5.3.1 (see (5.2.11)) Since (6.2.20) is bounded in probability, (6.2.17) is now bounded in probability as well.

Now that we have shown that all three terms of (6.2.16) are bounded in probability, this implies that (6.2.13) is $O_q(1)$. (6.2.14) is done in a similar way, noting that A is bounded in probability as well. All we have left to prove is (6.2.15).

To do so, note the following:

$$
\begin{aligned}
&\max_{1\le i\le n} \|\hat{\Phi}_i(t) - \Phi_i(t)\| \\
&\le \max_{1\le i\le n} \left\| \int_0^t \prod_{(0,s)} (I + \mathbf{d}\hat{A}(u))\mathbf{d}\hat{\Psi}_i(s) \left(\prod_{(s,t]} (I + \mathbf{d}\hat{A}(u)) - \prod_{(s,t]} (I + \mathbf{d}A(u)) \right) \right\|_\infty \\
&+ \left\| \int_0^t \prod_{(0,s)} (I + \mathbf{d}\hat{A}(u))\mathbf{d}\left(\hat{\Psi}_i(s) - \Psi_i(s)\right) \prod_{(s,t]} (I + \mathbf{d}A(u)) \right\|_\infty \\
&+ \left\| \int_0^t \left(\prod_{(0,s)} (I + \mathbf{d}\hat{A}(u)) - \prod_{(0,s)} (I + \mathbf{d}A(u)) \right) \mathbf{d}\Psi_i(s) \prod_{(s,t]} (I + \mathbf{d}A(u)) \right\|_\infty
\end{aligned}
$$

We may rewrite each of these terms the same we did for inequality (6.2.16). The first and third term of the above equation then converge to 0 in probability because the the difference in the product integral converges to 0 and because all elements are bounded in probability, as we proved before when we proved uniform consistency for the estimator $\hat{p}$. The second term has $\hat{\Psi} - \Psi$ converging to 0 in probability, because of convergence of $\overline{Y}_D$ and $\hat{A}$ to $\mathcal{Y}_D$ and A, respectively, and with help of the continuous mapping theorem. This means all terms converge to 0 in probability as $n \to \infty$. Therefore, (6.2.15) holds, and our theorem is proved. $\square$

6.3 Variance of the distribution

Although we have found a consistent estimator for the covariance function, the expression is rather difficult. The estimator for the variance of the elements of $n^{\frac{1}{2}}(\hat{p} - p)(t)$ is no different, since the elements of $\mathcal{G}(t)$ have variance $\xi_{jj}(t,t) =: \xi_j(t)$. It would be nice to have a formula in terms of quantities that we know or quantities which we can estimate with consistent estimators.

Notice how $\mathrm{\mathbb{V}ar\{\mathcal{G}_j(t)\} = Cov\{\mathcal{G}_j(t), \mathcal{G}_j(t)\} = \mathbb{E}(n^{-1}\sum_{i=1}^n \Phi_{ij}(t)^2) = n^{-1}\sum_{i=1}^n \mathbb{E}\Phi_{ij}(t)^2}$. Denote $p_{jj'}(s,t)$ the (j,j')th element of $\prod_{(s,t)}(I + \mathbf{d}A(u))$. Define $\hat{p}_{jj'}(s,t)$ similarly. Let $p(s,t) = \prod_{(s,t)}(I + \mathbf{d}A(u))$ and define $\hat{p}(s,t)$ likewise. Then

$$
\begin{aligned}
\Phi_{ij}(t) &= \sum_{k=1}^K \frac{Y_k(0+)}{n} \left[\int_0^t p_{jj'}(0,s)\mathbf{d}\Psi_i(s)p_{jj'}(s,t) \right]_{kj} \\
&= \sum_{k=1}^K \sum_{l=1}^K \sum_{m=1}^K \frac{Y_k(0+)}{n} \int_0^t p_{kl}(0,s)[\mathbf{d}\Psi(s)]_{lm} p_{mj}(s,t)
\end{aligned}
\tag{6.3.1}
$$

which we may now use for the variance. Recalling (6.2.2) for Ψ, We find, with help of Fubini's theorem

$$
\begin{aligned}
&\mathbb{V}\mathrm{ar}(\mathcal{G}_{\mathrm{j}}(\mathrm{t})) = \\
&\frac{1}{n}\sum_{i=1}^{n}\sum_{k,l,m\in\mathscr{S}}\sum_{k',l',m'\in\mathscr{S}}\mathbb{E}\frac{Y_k(0+)}{n}\frac{Y_{k'}(0+)}{n}\int_0^t\int_0^t p_{kl}(0,s)p_{k'l'}(0,s')p_{mj}(s,t)p_{m'j}(s',t)\cdot \\
&(\mathcal{Y}_D^{-1}(s)(\mathbf{d}N_i(s)-Y_{iD}(s)\mathbf{d}A(s)))_{lm}(\mathcal{Y}_D^{-1}(s')(\mathbf{d}N_i(s')-Y_{iD}(s')\mathbf{d}A(s')))_{l'm'} \\
&=\sum_{i=1}^{n}\sum_{k,l,m\in\mathscr{S}}\sum_{k',l',m'\in\mathscr{S}}\frac{Y_k(0+)}{n}\frac{Y_{k'}(0+)}{n}\int_0^t\int_0^t p_{kl}(0,s)p_{k'l'}(0,s')p_{mj}(s,t)p_{m'j}(s',t)\cdot \\
&\frac{1}{n}\mathbb{E}\left((\mathcal{Y}_D^{-1}(s)(\mathbf{d}N_i(s)-Y_{iD}(s)\mathbf{d}A(s)))_{lm}(\mathcal{Y}_D^{-1}(s')(\mathbf{d}N_i(s')-Y_{iD}(s')\mathbf{d}A(s')))_{l'm'}\right)
\end{aligned}
\tag{6.3.2}
$$

Now it is useful to note that $\mathbf{d}N_i(s)-Y_{iD}(s)\mathbf{d}A(s)$ is the increment of a Martingale (and is therefore a Martingale itself with mean zero), because $N_i(t)-\int_0^t Y_iD(s)\mathbf{d}A(s) =: M_i(t)$ is a Martingale. Then, using the tower property for conditional expectations and using the Martingale property, assuming $s\geq s'$, we find:

$$
\begin{aligned}
&\mathbb{E}((\mathbf{d}M_i(s))_{lm}(\mathbf{d}M_i(s'))_{l'm'})=\mathbb{E}(\mathbb{E}((\mathbf{d}M_i(s))_{lm}(\mathbf{d}M_i(s'))_{l'm'}|\mathcal{F}_{s'})) \\
&=\mathbb{E}((\mathbf{d}M_i(s'))_{lm}\mathbb{E}((\mathbf{d}M_i(s))_{l'm'}|\mathcal{F}_{s'}))=\mathbb{E}((\mathbf{d}M_i(s'))_{lm}(\mathbf{d}M_i(s'))_{l'm'})
\end{aligned}
\tag{6.3.3}
$$

Define

$$
\begin{aligned}
&\Sigma_{l,m,l',m'}(s,s') := \\
&\frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{\mathcal{Y}_D(s)}(\mathbf{d}N_i(s)-Y_{iD}(s)\mathbf{d}A(s))\right)_{lm}\left(\frac{1}{\mathcal{Y}_D(s')}(\mathbf{d}N_i(s')-Y_{iD}(s')\mathbf{d}A(s'))\right)_{l'm'}
\end{aligned}
\tag{6.3.4}
$$

for all $l,m,l',m'\in\mathscr{S}$.

Through (6.3.3) and theorem 3.4.1, we find, when $l=l'$ and $m=m'$, that
$\mathbb{E}(\Sigma_{m,l,m,l}(s,s'))=\mathbb{E}(\frac{1}{\mathcal{Y}_D(s)}\frac{1}{\mathcal{Y}_D(s')}Y_{iD}(s)\mathbf{d}A(s))_{lm}$.

Applying this information to (6.3.2) and applying theorem 3.4.1, we find

$$
\begin{aligned}
&\mathbb{V}\mathrm{ar}(\mathcal{G}_{\mathrm{j}}(\mathrm{t}))=\sum_{\mathrm{i=1}}^{\mathrm{n}}\sum_{\mathrm{k,l,m}\in\mathscr{S}}\sum_{\mathrm{k',l',m'}\in\mathscr{S}}\frac{\mathrm{Y_k(0+)}}{\mathrm{n}}\frac{\mathrm{Y_{k'}(0+)}}{\mathrm{n}}\int_0^{\mathrm{t}}\int_0^{\mathrm{t}}\mathrm{p_{kl}(0,s)p_{k'l'}(0,s')p_{mj}(s,t)p_{m'j}(s',t)}\cdot \\
&\frac{1}{n}\mathbb{E}\left((\mathcal{Y}_D^{-1}(s)(\mathbf{d}N_i(s)-Y_{iD}(s)\mathbf{d}A(s)))_{lm}(\mathcal{Y}_D^{-1}(s')(\mathbf{d}N_i(s')-Y_{iD}(s')\mathbf{d}A(s')))_{l'm'}\right) \\
&=\sum_{k,l,m\in\mathscr{S}}\sum_{k',l',m'\in\mathscr{S}}\frac{Y_k(0+)}{n}\frac{Y_{k'}(0+)}{n}\int_0^t\int_0^t p_{kl}(0,s)p_{k'l'}(0,s')p_{mj}(s,t)p_{m'j}(s',t)\cdot \\
&\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left((\mathcal{Y}_D^{-1}(s)(\mathbf{d}N_i(s)-Y_{iD}(s)\mathbf{d}A(s)))_{lm}(\mathcal{Y}_D^{-1}(s')(\mathbf{d}N_i(s')-Y_{iD}(s')\mathbf{d}A(s')))_{l'm'}\right) \\
&=\sum_{k,l,m\in\mathscr{S}}\sum_{k',l',m'\in\mathscr{S}}\frac{Y_k(0+)}{n}\frac{Y_{k'}(0+)}{n}\int_0^t\int_0^t p_{kl}(0,s)p_{k'l'}(0,s')p_{mj}(s,t)p_{m'j}(s',t)\Sigma_{l,m,l',m'}(s,s')
\end{aligned}
$$

Although this is still a rather long and nasty expression, it contains nothing but quantities that we know or can consistently estimate. As we proved in chapter 4, A can be consistently estimated by $\hat{A}$, and by chapter 5, p is consistently estimated by $\hat{p}$. Furthermore, as $\mathcal{Y}_D(s) = \lim_{n\to\infty} \overline{Y}_D(s)$, this quantity can be used to estimate $\mathcal{Y}_D(s)$.

Define the following estimator for Σ:

$$\hat{\Sigma}_{l,m,l',m'}(s,s') := \\ \frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{\overline{Y}_D(s)}(\mathbf{d}N_i(s) - Y_{iD}(s)\mathbf{d}\hat{A}(s))\right)_{lm}\left(\frac{1}{\overline{Y}_D(s')}(\mathbf{d}N_i(s') - Y_{iD}(s')\mathbf{d}\hat{A}(s'))\right)_{l'm'} \tag{6.3.5}$$

Noting again that $\hat{A}$ and $\overline{Y}_d$ are consistent estimators, an application of the continuous mapping theorem gives us consistency of this estimator for Σ. We find the following estimator for $\xi_j(t)$:

$$\hat{\xi}_j(t) = \\ \sum_{k,l,m\in\mathscr{S}}\sum_{k',l',m'\in\mathscr{S}}\frac{Y_k(0+)}{n}\frac{Y_{k'}(0+)}{n}\int_0^t\int_0^t \hat{p}_{kl}(0,s)\hat{p}_{k'l'}(0,s')\hat{p}_{mj}(s,t)\hat{p}_{m'j}(s',t)\hat{\Sigma}_{l,m,l',m'}(s,s') \tag{6.3.6}$$

Investigating whether this estimator is consistent or not could be an interesting topic to research. A suggestion on how to do this would be to prove that this mapping to get this integral is a continuous mapping, and then to use consistency of the estimators involved. Assuming this estimator is in fact consistent, we may then estimate the variance using just the data that has been provided.

7 Discussion

Conclusion

In this thesis, we have corrected Datta & Satten's [1] proof of consistency of the Nelson-Aalen estimator and the Aalen-Johansen estimator for multistate models and made the proofs more rigorous. We have discussed a way to model progress of diseases for individuals. We have provided the necessary conditions and tools to find those estimators. Using these results, these estimators can be used to give patients who have fallen ill a good idea of the possibilities and likelihoods of what might happen to them.

We have succeeded in our aim to correct and give proofs for the consistency of the Nelson-Aalen estimator and the Aalen-Johansen estimator. We have also made the proof of consistency of the estimated covariance function for the distribution of the Aalen-Johansen estimator more accessible. With the help of these tools, we have been able to provide a formula for the variance of this distribution.

Related research

The results do lead to several other questions that need to be answered. The first one is about the Nelson-Aalen estimator and its definition of the transition hazard. In the Nelson-Aalen estimator, not much information about the past is used. Although there is no Markov-assumption, it seems they use a weighted average on the history of what could have happened to patients. This gives an asymptotically correct estimator, but in individual situations, this estimator could be off, because the information on what happened to the patient itself is not used. This can be remedied, by defining a transition hazard as follows (with $j \neq j'$):

$$\alpha_{jj'}(t) = \lim_{\mathbf{d}t \downarrow 0} \frac{\mathbb{E}[\mathbf{d}N^*_{i;jj'}(t)|\mathcal{F}_{t-}]}{\mathbf{d}t}$$

Through this definition, the transition hazard is now a random variable. Because of the nature of the proof of theorem 4.2.1, this theorem cannot be applied for this transition hazard. To use all information up to time $t-$, other conditions are necessary. A way to do so could be to bound the hazard by a function $k(s)$ with $\int_0^t k(s)\mathbf{d}s < \infty$ for all $t \in (0, \tau)$ and imposing this bound in the conditions for theorem 4.2.1 as well. The proof may then be adapted to fit the new model. However, this is a strong requirement, and weaker assumptions to get the same result could be nice.

Another question is the possibility of a much simpler estimator for the state occupation probabilities. In chapter 5, we proved that the number of people at risk with right-censoring, divided by n times the probability of having survived up to that time without being censored is a consistent estimator. Therefore, when the probability of being censored is known, much less work is required to find the state occupation probabilities. A problem might arise when the censoring probability is unknown, which is usually the case. Assuming that there is a probability larger than 0 that an individual is not censored and has survived, the surviving without censoring probability can be estimated from the censoring hazard, which can be estimated the same way the cumulative transition hazard can be estimated. With independent censoring, the steps to prove consistency for the Nelson-Aalen estimator of the censoring hazard are the same for the estimator of the transition hazard. An estimator for $p_j(t)$ would then be

$$\hat{p'}(t) := \frac{Y_j^{uc}(t+)}{n\hat{K}(t-)}$$

where $\hat{K}$ denotes the estimator for K.

Alternatively, one may use another estimator which is rather naive, that is, $\hat{p'}^*(t) := \frac{Y_j^{uc}(t+)}{n'(t)}$, where $n'(t)$ is the number of people in the study that have not been censored. Under the assumption of uninformative right-censoring, this is a consistent estimator. Although it requires not nearly as many calculations as the estimator $\hat{p'}$, it is a bit naive, as you lose a lot of information by not including the data you have about the censored patients. An interesting topic of research could be the influence of the censoring on the variance of this estimator $\hat{p'}^*$, as well as the variance of the estimator $\hat{p'}$.

A third problem could be the proposed estimator in chapter 6 and calculations thereof. David Glidden recognised the problem, and came up with a different way of estimating the distribution of $n^{\frac{1}{2}}(\hat{p}-p)(\cdot)$, namely through simulation [3]. He multiplied the estimators $\hat{\Phi}_i$ by independent standard normal random variables Z_i that are independent of the data as well. It can be shown quite easily that this process has the same covariance function and therefore the same limiting distribution as the original process. He further explored this method and used it to develop confidence bands for the distribution of $n^{\frac{1}{2}}(\hat{p}-p)(\cdot)$. Although he showed that the method works using bootstrapping, he acquires a simulated result, rather than a result obtained purely from the data from the model, which could be preferable. Calculations can be carried out quicker, though, and sampling from a standard normal distribution is quite easy.

References

[1] Somnath Datta and Glen A. Satten. Validity of the aalen-johansen estimators of stage occupation probabilities and nelson-aalen estimators of integrated transition hazards for non-markov models. *Statistics & Probability Letters*, 55(4), 2001. URL `http://EconPapers.repec.org/RePEc:eee:stapro:v:55:y:2001:i:4:p:403-411`.

[2] Thomas R. Fleming and Wiley David P. Harrington. Counting processes and survival analysis. , new york, 1991. no. of pages: xiii + 429. isbn: 0-471-52218-x. 11(13), 1992.

[3] David V. Glidden. Robust inference for event probabilities with non-markov event data. *Biometrics*, 58(2), 2002.

[4] Soeren Johansen, Odd O. Aalen, and Søren Johansen. An empirical transition matrix for non-homogeneous markov chains based on censored observations. *Scandinavian Journal of Statistics*, 5, 1978.

[5] Søren Johansen, Soeren Johansen, Richard D. Gill, and Søren Johansen. A survey of product-integration with a view toward application in survival analysis. *The Annals of Statistics*, 18, 1990.

[6] Edward L. Kaplan and Paul Meier. Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association*, 53, 1958.

[7] Wayne Nelson. Hazard plotting for incomplete failure data. *Journal of Quality Technology*, 1(1), 1969.

[8] Håkon K. Gjessing Odd O. Aalen, Ørnulf Borgan. Survival and event history analysis: A process point of view. 77(3), 2009.

[9] Richard D. Gill Per Kragh Andersen, Ørnulf Borgan and Niels Keiding. Statistical models based on counting processes. , springer-verlag, new york, 1993. no. of pages: xi + 767. isbn: 0-387-97872-0. 13(17), 1993.

[10] Cristian Spitoni, Marion Verduijn, and Hein Putter. Estimation and asymptotic theory for transition probabilities in markov renewal multi-state models. *The International Journal of Biostatistics*, 8(1-A18), 2012.

[11] V Volterra. Sulle equazioni differenziali lineari. *Rend.Accad. Lincei*, 4(3), 1887.