December 2016

Density Estimation for Lifetime Distributions Under Semi-parametric Random Censorship Models

Carsten Harlass
University of Wisconsin-Milwaukee

Density estimation for lifetime distributions under semi-parametric random censorship models

by

Carsten Harlaß

A Dissertation Submitted in

Partial Fulfillment of the

Requirements for the Degree of

Doctor of Philosophy

in

Mathematics

at

The University of Wisconsin–Milwaukee

December 2016

Abstract

Density estimation for lifetime distributions under semi-parametric random censorship models

by

Carsten Harlaß

The University of Wisconsin–Milwaukee, 2016

Under the Supervision of Professor Gerhard Dikta and Professor Jugal Ghorai

We derive product limit estimators of survival times and failure rates for randomly right censored data as the numerical solution of identifying Volterra integral equations by employing explicit and implicit Euler schemes. While the first approach results in some known estimators, the latter leads to a new general type of product limit estimator. Plugging in established methods to approximate the conditional probability of the censoring indicator given the observation, we introduce new semi-parametric and presmoothed Kaplan-Meier type estimators. In the case of the semi-parametric random censorship model, i.e. the latter probability belonging to some parametric family, we study the strong consistency and asymptotic normality of some linear functionals based on the proposed estimator.

Assuming that the underlying random variable admits a probability density, we define semi-parametric and presmoothed kernel estimators of the density and the hazard rate for randomly right censored data, which rely on the newly derived estimators of the survival function. We determine exact rates of pointwise and uniform convergence as well as the limiting distribution.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

a.s.	Almost surely
CLT	Central limit theorem
d.f.	Probability distribution function
DKW	Dvoretzky-Kiefer-Wolfowitz inequality
e.c.d.f.	Empirical cumulative distribution function
i.i.d.	Stochastically independent and identical distributed
MLE	Maximum likelihood estimator
PHM	Proportional hazards model
PLE	Product limit estimator
p.d.f.	Probability density function
RCM	Random censorship model
SLLN	Strong law of large numbers
SRCM	Semi-parametric random censorship model
w.r.t.	With respect to

List of Symbols

$I_{[A]}$	Indicator function, $I_{[A]} = 1$ if A is true, otherwise 0
$\mathbb{R}_{\geq}$	Positive real line $[0, \infty)$
$R_n(Z_i)$	Rank of Z_i in the sample $(Z_k)_{1 \leq k \leq n}$
$Z_{i:n}$	i-th value of the order statistic of the sample $(Z_k)_{1 \leq k \leq n}$
$\delta_{(i:n)}$	Adjunct censoring indicator of $Z_{i:n}$

Chapter 1

Introduction

In survival analysis it is often not possible to observe the variable of interest. Therefore approximations can only rely on incomplete data. Assuming the framework of the random censorship model (RCM), we deduce a Volterra integral equation for the survival function of the censored random variable. Employing the explicit Euler scheme to numerically solve the equation results in some already known estimators, among them the Kaplan-Meier product limit estimator (PLE), and its semi-parametric and presmoothed equivalents.

Since the corresponding differential equation is stiff, applying an implicit Euler scheme is more suitable from a numerical point of view. This approach leads to a new class of estimators whose members are almost all true distribution functions (d.f.). In comparison, the already established estimators, e.g. the Kaplan-Meier PLE, are in general only subdistribution functions. Plugging in estimators for the conditional probability of the censoring indicator given its actual value, we propose the new semi-parametric and presmoothed PLEs $F_{2,n}^{SE}$ and $F_{2,n}^{PR}$, respectively. For the semi-parametric approach we slightly extend the RCM to the semi-parametric random censorship model (SRCM). In case of $F_{2,n}^{SE}$, we show that the estimator is asymptotically equivalent to the semi-parametric estimators introduced in Dikta (1998, 2000) and therefore inherits its asymptotic properties. In particular, the estimators have the same asymptotic variance. Thus the integral estimator based on $F_{2,n}^{SE}$ is optimal w.r.t. the class of all regular estimators of the integral induced by the true distribution F.

Given that the censored random variable admits a density, kernel based approximations of this underlying probability density function (p.d.f.), evolving from the Kaplan-Meier PLE or its presmoothed version, are considered in the literature. Besides the definition of a new nonparametric density estimator, we focus on the investigation of semi-parametric kernel estimators of the p.d.f. and the hazard rate. We present an asymptotic representation which is used to deduce exact rates of pointwise and uniform convergence as well as the limit distribution. We will show that the semi-parametric density estimator is superior to its Kaplan-Meier counterpart in terms of the asymptotic variance, when assuming the correct parametric model of the plug-in estimator.

Chapter 2 explains the RCM and extends it to the semi-parametric random censorship model (SRCM). In Section 3.1 we present a technique to derive PLEs from identifying Volterra integral equations and propose the new estimators $F_{2,n}^{SE}$ and $F_{2,n}^{PR}$ in Definition 3.7 and Definition 3.8, respectively. Their properties are discussed in Section 3.2. The main results related to $F_{2,n}^{SE}$ are given in Theorem 3.13 and Theorem 3.16. Those are essential for extending the strong law of large numbers (SLLN) and the central limit theorem (CLT) to the semi-parametric setup. Chapter 4 is concerned with kernel based density estimators. New estimators are introduced in Section 4.2 and their asymptotic representations are given in Theorem 4.6 and Theorem 4.7. Based on those we determine exact rates of pointwise and uniform convergence and deduce the pointwise limiting distribution as well as the distribution of the maximal deviation. Due to their complexity, most of the proofs are postponed to the last sections of Chapter 3 or Chapter 4.

Chapter 2

Preliminaries

In this chapter we give a brief introduction to basic concepts in survival analysis and list some preliminary definitions. The ideas are more widely discussed in Klein and Moeschberger (2003) and Kleinbaum and Klein (2012). Comprehensive results can be found in Klein, van Houwelingen, Ibrahim, and Scheike (2013).

2.1 Basics on survival analysis

In classical statistics d.f.s and p.d.f.s are standard tools for data modeling. In survival analysis, where one is interested in survival probabilities and failure rates, usually other instruments are employed. A life time X is the time during which an entity exhibits certain characteristics – for example the time from the entry of a proband into a pharmaceutical study till their death. Other examples are the time a production machine is in working condition before it needs replacement or the time a worker is unemployed until being hired again. Often it is assumed that a life time is concentrated on the positive real line $\mathbb{R}_{\geq} \equiv [0, \infty)$.

Definition 2.1. If not stated otherwise, a life time X is a random variable on some probability space $(\Omega, \mathcal{A}, \mathbb{P})$ which has an absolute continuous distribution w.r.t. the Lebesgue measure on $\mathbb{R}_{\geq}$ and maps into $(\mathbb{R}_{\geq}, \mathcal{B}(\mathbb{R}_{\geq}))$. Let F be its d.f. The induced measure is denoted by dF and the corresponding Radon–Nikodym derivative by f.

Counterparts of the d.f. and the p.d.f. are the survival and the hazard function, respectively.

Definition 2.2. Let X be a random variable as defined in Definition 2.1, then the survival function is defined by $\bar{F}(x) \coloneqq 1 - F(x) = \mathbb{P}(X > x) = \int_x^\infty f(t)dt$. The hazard function is specified by

$$\lambda : \mathbb{R}_\geq \ni x \mapsto \lambda(x) \coloneqq \frac{f(x)}{\bar{F}(x)} \in \mathbb{R}_\geq$$

and

$$\Lambda : \mathbb{R}_\geq \ni x \mapsto \Lambda(x) \coloneqq \int_0^x \lambda(t)dt \in \mathbb{R}_\geq$$

is called the cumulative hazard function.

The survival function evaluated at x is the probability that the random variable X is greater than x. In terms of survival analysis, this can be interpreted as the probability that an individual survives a certain point in time x. By looking at

$$\lambda(x) = \frac{f(x)}{\bar{F}(x)} = \lim_{h \to 0} \left(\frac{\mathbb{P}(X \leq x + h | X \geq x)}{h} \right),$$

the hazard function could be interpreted as the mortality rate at time point x. Both, $\bar{F}$ and Λ, are closely related.

Lemma 2.3. Let X be a random variable as defined in Definition 2.1, then

$$\bar{F}(x) = \exp(-\Lambda(x)).$$

Proof. Applying the exponential function on both sides, the result follows immediately from the definition of the hazard rate and the fundamental theorem of calculus,

$$\Lambda(x) = \int_0^x \lambda(t)dt = \int_0^x \frac{f(t)}{\bar{F}(t)}dt = \int_0^x \frac{F'(t)}{\bar{F}(t)}dt = -\ln(\bar{F}(x)). \qquad \square$$

In many practical applications and scientific fields it is not possible to observe the variable of interest and analysis can only be based on incomplete data, for example see Kalbfleisch and Prentice (2002). When testing for lifetimes or failure rates, incomplete data is primarily caused by censoring. There are different types of censoring; data which is truncated from the right is called right-censored. This kind of data often arises in medical research, cf. Armitage, Berry, and Matthews (2001). For instance in a clinical trial patients start taking a medicine at a certain point in time. A proband could either die during the time frame of the study from the disease which is actually being treated (no censoring), leave the study prior the end (e.g. moving away) or survive the end of the trial. Both last cases are examples for right-censoring. For a more detailed explanation and examples see Klein and Moeschberger (2003, Chapter 3).

A common way to describe randomly right-censored data is the RCM. It is the basis of many publications, among them Kaplan and Meier (1958) and Efron (1967).

Definition 2.4. Let $(X_i)_{1\leq i\leq n}$ be a sequence of independent, identical distributed (i.i.d.), nonnegative random variables defined on the probability space $(\Omega, \mathcal{A}, \mathbb{P})$ and distributed according to the unknown d.f. F; cf. Definition 2.1. Furthermore, let $(Y_i)_{1\leq i\leq n}$ be another sequence of i.i.d., nonnegative random variables defined on the same probability space $(\Omega, \mathcal{A}, \mathbb{P})$ and distributed according to the d.f. G. In addition, assume that the sequences $(X_i)_{1\leq i\leq n}$ and $(Y_i)_{1\leq i\leq n}$ are independent from each other. Under the RCM, data of the form $(Z_i, \delta_i)_{1\leq i\leq n}$ is observed, where $Z_i = \min(X_i, Y_i)$ and $\delta_i = I_{[X_i \leq Y_i]}$. The variable δ_i indicates whether observation Z_i is censored ($\delta_i = 0$) or uncensored ($\delta_i = 1$). Denote the d.f. of the random variable $Z = \min(X, Y)$ by H.

In this scenario, the PLE proposed by Kaplan and Meier (1958) has received great attention both in theory and practice; cf. Section 3.2. Supplementary to the RCM, we here assume the X and Y are absolutely continuous w.r.t. the Lebesgue measure and therefore admit the p.d.f.s f and g, respectively. Furthermore, let $(Z_{i:n})_{1\le i\le n}$ denote the order statistics of the Z-sample and $\left(\delta_{[i:n]}\right)_{1\le i\le n}$ the sequence of indicators adjunct to the ordered Z-sample. For all $1 \le i \le n$, let $R_n(Z_i)$ represent the rank of Z_i in the Z-sample.

Given the RCM, the importance of the conditional probability

$$m(z) \coloneqq \mathbb{P}(\delta = 1|Z = z) = \mathbb{E}\left(\mathbb{1}_{\{X\le Y\}}|Z = z\right), \tag{2.1}$$

the probability of an uncensored observation given its actual value $Z = z$, for the consistency of $\int \varphi dF_n^{KM}$ was pointed out in Stute (1993). Consistently, define $\bar{m}(z) \coloneqq 1 - m(z)$. When looking at the Kaplan-Meier PLE, the probability m is basically estimated by setting it to one or zero, in particular

$$m(Z_i) \approx \delta_i.$$

When we use somehow better estimators for m, which are based on the sample $(Z_i, \delta_i)_{1\le i\le n}$, one can deduce other estimators of F. If for example δ is independent of Z, then $m(x) = \mathbb{E}[\delta]$ and a suitable estimator of m is given by $(1/n)\sum_{i=1}^n \delta_i$. The resulting estimator for F was presented in Abdushukurov (1987) and Cheng and Lin (1987); cf. Remark 3.5. For the semi-parametric PLE introduced by Dikta (1998) it is assumed that m belongs to a parametric family. Hence it can be estimated using common parametric approaches. We now extend the RCM by this assumption.

Definition 2.5. Given the RCM, the semi-parametric random censorship model (SRCM) additionally assumes that the conditional expectation $m(z) = \mathbb{P}(\delta = 1|Z = z)$ belongs to a parametric family where each member is identified by a parameter $\theta \in \Theta$. Hence

$$m(z) = m(z, \theta_0),$$

where $m(\cdot, \theta_0)$ is a parametric function and $\theta_0 = (\theta_{0,1}, ..., \theta_{0,k}) \in \Theta \subset \mathbb{R}^k$ the true parameter.

Parametric models for m can be found in Cox and Snell (1989), Dikta (1998) or Collett (2002). Given the RCM, note the two following basic relationships between F, G, H and m.

Corollary 2.6. Given the definitions of the RCM, it holds that $\bar{H} = \bar{F}\bar{G}$.

Proof. Exploiting the independence of X and Y we have

$$\bar{H}(t) = \mathbb{P}(\min(X, Y) > t) = \mathbb{P}(\{X > t\} \cap \{Y > t\}) = \mathbb{P}(X > t)\mathbb{P}(Y > t) = \bar{F}(t)\bar{G}(t).$$

□

Corollary 2.7. Given the RCM, let $\bar{G}(x^-)$ denote the left-hand limit of $\bar{G}$ at x. Then both subdistribution functions

$$H^1(t) \coloneqq \mathbb{P}(\delta = 1, Z \le t) \qquad \text{and} \qquad H^0(t) \coloneqq \mathbb{P}(\delta = 0, Z \le t)$$

have Radon–Nikodym derivatives w.r.t. dH, and dF or dG, respectively. E.g.

$$H^1(t) = \int_{[0,t]} m(x) H(dx) = \int_{[0,t]} \bar{G}(x^-) F(dx) \tag{2.2}$$

and

$$H^0(t) = \int_{[0,t]} \bar{m}(x) H(dx) = \int_{[0,t]} \bar{F}(x^-) G(dx). \tag{2.3}$$

Moreover, let's denote the Radon–Nikodym derivative of H^1 by h^1 and observe that

$$h^1(x) = m(x)h(x) = \bar{G}(x^-)f(x). \tag{2.4}$$

Proof. Recalling $\delta \coloneqq \mathbb{1}_{\{X \le Y\}}$, the first equality of (2.2) is a shorthand version of

$$\begin{aligned} H^1(t) \coloneqq \mathbb{P}(\delta = 1, Z \le t) =& \mathbb{E}\left(\mathbb{1}_{\{X \le Y\}} \cdot \mathbb{1}_{\{Z \le t\}}\right) = \mathbb{E}\left[\mathbb{E}\left(\mathbb{1}_{\{X \le Y\}} \cdot \mathbb{1}_{\{Z \le t\}} | Z\right)\right] \\ =& \int_{\mathbb{R}_{\ge}} \mathbb{E}\left(\mathbb{1}_{\{Z \le t\}} \mathbb{1}_{\{X \le Y\}} | Z = z\right) H(dz) \\ =& \int_{\mathbb{R}_{\ge}} \mathbb{1}_{\{z \le t\}} \mathbb{E}\left(\mathbb{1}_{\{X \le Y\}} | Z = z\right) H(dz) \\ =& \int_{[0,t]} m(x) H(dx), \end{aligned}$$

where we used the definition of $m(x)$ from (2.1). Similarly, for the second equality in (2.2) consider $Z \coloneqq \min(X, Y)$ and

$$\begin{aligned} H^1(t) =& \mathbb{E}\left(\mathbb{1}_{\{X \le Y\}} \cdot \mathbb{1}_{\{Z \le t\}}\right) = \mathbb{E}\left[\mathbb{E}\left(\mathbb{1}_{\{X \le Y\}} \cdot \mathbb{1}_{\{X \le t\}} | X\right)\right] \\ =& \int_{\mathbb{R}_{\ge}} \mathbb{E}\left(\mathbb{1}_{\{X \le t\}} \mathbb{1}_{\{X \le Y\}} | X = x\right) F(dx) = \int_{\mathbb{R}_{\ge}} \mathbb{1}_{\{x \le t\}} \mathbb{E}\left(\mathbb{1}_{\{x \le Y\}}\right) F(dx) \\ =& \int_{\mathbb{R}_{\ge}} \mathbb{1}_{\{x \le t\}} [1 - \mathbb{E}\left(\mathbb{1}_{\{Y < x\}}\right)] F(dx) = \int_{[0,t]} \bar{G}(x^-) F(dx). \end{aligned}$$

The proof of (2.3) is analogous. □

2.2 Solving differential equations

In Chapter 3, PLEs are derived as the solution of some initial value problem. Already Volterra (1887) was concerned with the numerical solution of Volterra integral equations. He applied Euler schemes to obtain approximate solutions in product form. Gill and Johansen (1990) pointed out that the famous Kaplan-Meier PLE emerges as a solution of an identifying integral equation. However, ordinary differential equations are very well studied and a huge variety of literature is available, for example Hartman (2002), Ascher and Petzold (1998) and Deuflhard and Bornemann (2002). Here we just briefly outline the explicit and implicit Euler scheme as we will use them in the subsequent chapter.

Let $U \subset \mathbb{R}$, $V \subset \mathbb{R}^n$, $u \in C(U, V)$ where $C(U, V)$ denotes the set of continuously differentiable functions mapping $U \mapsto V$. Furthermore let $f \in C(W)$ with W an open subset of $\mathbb{R}^{1+n}$. Then a general first order initial value problem is given by

$$u'(t) = f(t, u(t)) \quad \text{for all } t \in [t_0, t_e], \qquad u(t_0) = u_0,$$

and the corresponding Volterra integral equation is

$$u(t) = u_0 + \int_{t_0}^{t} f(\tau, u(\tau))d\tau \quad \text{for all } t \in [t_0, t_e].$$

Assuming the initial value problem has a unique solution and defining the node points $t_0 \leq t_1 \leq \ldots \leq t_n = t_e$, an intuitive way to numerically approximate $u(t_i) \approx u_i$ is given by the iterative method

$$u_{i+1} = u_i + (t_{i+1} - t_i)f(t_i, u_i) \qquad \text{for } i = 1, \ldots, n, \tag{2.5}$$

which is called the explicit Euler scheme. It is the simplest member of the more general family of Runge-Kutta methods.

An initial value problem is called stiff if $u(t)$ exponentially decreases to zero as t increases but the derivative is significantly larger then $u(t)$ itself. For a more detailed explanation cf. Aiken (1985, pp. 360). In case of the initial value problem being stiff, explicit methods are not applicable any more and A/A(α)- and L-stable methods are recommended, for example see Hairer and Wanner (2010, Chapters IV.3 and IV.5). The simplest L-stable method is the implicit Euler scheme defined by the iteration

$$u_{i+1} = u_i + (t_{i+1} - t_i)f(t_{i+1}, u_{i+1}) \qquad \text{for } i = 1, \ldots, n. \tag{2.6}$$

There are a lot of results available concerning the theory of solving stiff differential equations and there exist much more advanced numerical methods. But here we restrict ourselves to the simplest case and refer to the literature mentioned above.

Chapter 3

Survival time estimators derived from identifying Volterra equations

In literature, the construction of survival time estimators for right censored data is commonly based on the Nelson (1972) and Aalen (1978) estimator. Following the idea of Gill and Johansen (1990) we are going to present a more general technique to derive PLEs of the survival function. Using this method, we will show an alternative way to deduce already known estimators, among those the well-known Kaplan and Meier (1958) PLE. Primarily we are interested in the construction of a new general type of survival time estimators. We will establish a new presmoothed and a new semi-parametric survival time PLE in Section 3.1 and analyze the properties of the latter in Section 3.2. Corollary 3.14 states the asymptotic distribution of the new estimator and Corollary 3.20 represents our strong law result. The proofs of the major theorems are given in Section 3.3.

3.1 Deriving PLEs from Volterra integral equations

Under the RCM and the SRCM, respectively (cf. Section 2.1), data of the form $(Z_i, \delta_i)_{1 \le i \le n}$ is observed but one is usually interested in characteristics of the random variable X, e.g. the d.f. F or the p.d.f. f. In the following we develop Volterra integral equations which identify the d.f. F. Those identifying integral equations will be discretized and approximate solutions are derived by applying the Euler schemes from Section 2.2. Depending on the initial identifying integral equation and the solution method we will end up with different types of PLEs.

Under the assumption that F and G are continuous, we first construct a Volterra type integral equation using G as a starting point. Therefor consider on the one hand

$$G(t) = 1 - \bar{G}(t) = 1 - \frac{\bar{F}(t)\bar{G}(t)}{\bar{F}(t)} = \frac{\bar{F}(t) - \bar{H}(t)}{\bar{F}(t)}, \tag{I}$$

where the last equality follows by Corollary 2.6. On the other hand we have, when applying Corollary 2.7,

$$G(t) = \int_{[0,t]} \frac{\bar{F}(x^-)}{\bar{F}(x^-)} G(dx) = \int_{[0,t]} \frac{1}{\bar{F}(x^-)} H^0(dx). \tag{II}$$

Since we assume continuity of F and G we can omit the left-hand limits. Setting I $=$ II gives the identifying Volterra type integral equation

$$\frac{\bar{F}(t) - \bar{H}(t)}{\bar{F}(t)} = \int_{[0,t]} \frac{1}{\bar{F}(x)} H^0(dx).$$

Replacing H and H^0 by their empirical counterparts,

$$H_n(t) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{\{Z_i \le t\}}, \qquad \bar{H}_n = 1 - H_n, \qquad H_n^0(t) = \frac{1}{n}\sum_{i=1}^{n} \bar{m}_n(Z_i)\mathbb{1}_{\{Z_i \le t\}}, \tag{3.1}$$

with m_n being some estimator of m, leads to the estimating equation

$$\frac{\bar{F}_n^*(t) - \bar{H}_n(t)}{\bar{F}_n^*(t)} = \int_{[0,t]} \frac{1}{\bar{F}_n^*(x)} H_n^0(dx). \tag{3.2}$$

Here F_n^* denotes some estimator of F and consistently $\bar{F}_n^* = 1 - F_n^*$ as well as $\bar{m}_n = 1 - m_n$.

It is easy to construct a Volterra type integral equation based on F using (2.2):

$$F(t) = \int_{[0,t]} \frac{\bar{G}(u^-)}{\bar{G}(u^-)} F(du) = \int_{[0,t]} \frac{1}{\bar{G}(u^-)} H^1(du) = F(0) + \int_{[0,t]} \frac{\bar{F}(u^-)}{\bar{H}(u^-)} H^1(du). \tag{3.3}$$

Due to the continuity of F, G and H it is reasonable to omit the left-hand limits. Again, let F_n^* denote some estimator of F. Then the corresponding estimating equation

$$F_n^*(t) = F_n^*(0) + \int_{[0,t]} \frac{\bar{F}_n^*(u)}{\bar{H}_n(u)} H_n^1(du) = F_n^*(0) + \int_{[0,t]} \frac{\bar{F}_n^*(u) m_n(u)}{\bar{H}_n(u)} H_n(du) \tag{3.4}$$

emerges when replacing H by its empirical distribution function (e.c.d.f.) H_n and, similarly to (3.1), approximating H^1 by

$$H_n^1(t) = \frac{1}{n} \sum_{i=1}^{n} m_n(Z_i) \mathbb{1}_{\{Z_i \le t\}}. \tag{3.5}$$

A common method to numerically solve integral equations like (3.2) and (3.4) is the explicit Euler scheme as outlined in Section 2.2. In our particular case, the grid points are given by the ordered sample $(Z_{k:n})_{1 \le k \le n}$. To actually solve the integral equation define $F_n^*(t) = 0$ for all $t < Z_{1:n}$. Furthermore, it is natural to define $Z_{0:n}$ such that $Z_{0:n} < Z_{1:n}$. Hence $F_n^*(Z_{0:n}) = 0$, which is going to be used as the initial value for the Euler scheme.

Consider the integral equation (3.4). Using $Z_{0:n}, \cdots, Z_{n:n}$ as node points, observe that

$$F_n^*(Z_{k:n}) = F_n^*(Z_{k-1:n}) + \int_{]Z_{k-1:n}, Z_{k:n}]} \frac{\bar{F}_n^*(u) m_n(u)}{\bar{H}_n(u)} H_n(du) \tag{3.6}$$

for $k = 1, \ldots, n$. The application of the explicit Euler scheme and replacing $m_n(Z_{k-1:n})$ by $m_n(Z_{k:n})$ gives

$$F_n^*(Z_{k:n}) = F_n^*(Z_{k-1:n}) + \frac{m_n(Z_{k:n}) \bar{F}_n^*(Z_{k-1:n})}{n \bar{H}_n(Z_{k-1:n})},$$

which, after observing that $n\bar{H}_n(Z_{k-1:n}) = n - k + 1$, is equivalent to

$$\bar{F}_n^*(Z_{k:n}) = \bar{F}_n^*(Z_{k-1:n})\left[1 - \frac{m_n(Z_{k:n})}{n-k+1}\right].$$

This recursive formula can be written explicitly in the typical product form and can be used to define a class of estimators.

Definition 3.1. Let $F_{1,n}^*$ denote a type of product limit estimators defined by

$$1 - F_{1,n}^*(Z_{k:n}) \coloneqq \prod_{i=1}^{k}\left[1 - \frac{m_n(Z_{i:n})}{n-i+1}\right], \tag{3.7}$$

where m_n is some estimator of the conditional probability m defined in (2.1).

Note that this definition is equivalent to

$$1 - F_{1,n}^*(t) = \prod_{i:Z_k \le t}\left[1 - \frac{m_n(Z_i)}{n - R_n(Z_i) + 1}\right],$$

where $R_n(Z_i)$ is the rank of Z_i in the Z-sample and an empty product is considered to be 1. In the remarks below we will see that most of the known estimators, which rely on the RCM, actually belong to this class.

We again turn to the identifying equation (3.3). Assuming that H admits a continuous p.d.f. h with respect to the Lebesgue measure, the corresponding initial value problem is formulated by

$$\frac{\partial \bar{F}(t)}{\partial t} = -\bar{F}(t)\lambda(t), \qquad \lambda(t) = \frac{m(t)h(t)}{\bar{H}(t)}, \qquad \bar{F}(0) = 1, \tag{3.8}$$

where λ is the hazard rate of X. The differential equation (3.8) becomes arbitrarily stiff as λ attains large values, especially if $\lambda(t) \to \infty$ as $t \to \infty$.

In numerical analysis the standard approach to solve such stiff ODEs is the application of A/A(α)- and L-stable methods; cf. Section 2.2. Since all explicit Runge-Kutta methods are not A-stable neither is the explicit Euler scheme. However, the implicit Euler scheme is the simplest L-stable method. Applying the implicit Euler scheme to equation (3.6) gives

$$F_n^*(Z_{k:n}) = F_n^*(Z_{k-1:n}) + \frac{m_n(Z_{k:n})\bar{F}_n^*(Z_{k:n})}{n\bar{H}_n(Z_{k:n})} \qquad \text{for } k = 1, \ldots, n-1.$$

Then solving for $\bar{F}_n^*(Z_{k:n})$ results in the following recursive definition

$$\begin{aligned}
\bar{F}_n^*(Z_{k:n}) &= \bar{F}_n^*(Z_{k-1:n}) - \frac{m_n(Z_{k:n})\bar{F}_n^*(Z_{k:n})}{n\bar{H}_n(Z_{k:n})} \\
\Leftrightarrow \bar{F}_n^*(Z_{k:n}) \left[1 + \frac{m_n(Z_{k:n})}{n\bar{H}_n(Z_{k:n})}\right] &= \bar{F}_n^*(Z_{k-1:n}) \\
\Leftrightarrow \bar{F}_n^*(Z_{k:n}) &= \bar{F}_n^*(Z_{k-1:n}) \frac{n-k}{n-k+m_n(Z_{k:n})} \\
\Leftrightarrow \bar{F}_n^*(Z_{k:n}) &= \bar{F}_n^*(Z_{k-1:n}) \left[1 - \frac{m_n(Z_{k:n})}{n-k+m_n(Z_{k:n})}\right],
\end{aligned}$$

which together with the initial condition $F_n^*(Z_{0:n}) = 0$ is equivalent to the following product form which we use to define a new class of estimators. Even though the calculations hold true only for $k < n$, the last equation is also well-defined for $k = n$ if $m_n(Z_{n:n}) > 0$.

Definition 3.2. Let $F_{2,n}^*$ denote a class of product limit estimators defined by

$$\begin{aligned}
1 - F_{2,n}^*(Z_{k:n}) &:= \prod_{i=1}^{k} \left[1 - \frac{m_n(Z_{i:n})}{n-i+m_n(Z_{i:n})}\right] \qquad (3.9) \\
&= \prod_{i=1}^{k} \left[\frac{n-i}{n-i+m_n(Z_{i:n})}\right] = \prod_{i:Z_i \le t} \left[\frac{n-R_n(Z_i)}{n-R_n(Z_i)+m_n(Z_i)}\right],
\end{aligned}$$

where $R_n(Z_i)$ is the rank of Z_i in the Z-sample and an empty product is interpreted to be 1.

Note that applying the explicit Euler scheme to the estimating equation (3.2) results in $F_{2,n}^*$ defined in (3.9), and similarly, using the implicit version to solve (3.2) eventuates in the definition of $F_{1,n}^*$ given in (3.7).

In the case of no censoring, in particular $\delta \equiv 1$, the conditional probability $m(z) = 1$ for all $z \in \mathbb{R}_{\geq}$. Hence both prototype estimators, $F^*_{1,n}$ and $F^*_{2,n}$, reduce to the well known e.c.d.f.

In case of censoring, it is left to specify the exact form of m_n for both prototype estimators $F^*_{1,n}$ and $F^*_{2,n}$. The choice of m_n strongly depends on the information which is available about m. At first we take a look at the case with the least amount of assumptions.

Remark 3.3. If we assume nothing else but the RCM, then $\delta_{[k:n]}$, the adjunct indicator of $Z_{k:n}$, is a reasonable estimator of $m(Z_{k:n})$. When using

$$m_n(Z_{k:n}) = \delta_{[k:n]}, \qquad \forall k = 1, \dots, n$$

$F^*_{1,n}$ is exactly the estimator introduced by Kaplan and Meier (1958):

$$1 - F_n^{KM}(t) \coloneqq \prod_{i:Z_i \leq t} \left(1 - \frac{\delta_i}{n - R_n(Z_i) + 1}\right), \tag{3.10}$$

where $R_n(Z_i)$ denotes the rank of Z_i for all $1 \leq i \leq n$.

If we assume a SRCM, then $m(z) = m(z, \theta_0)$. Interpreting m to be the link function of the underlying binary regression model of the sample $(Z_i, \delta_i)_{1 \leq i \leq n}$, we can exploit the maximum likelihood estimator (MLE) θ_n of the true value θ_0 to derive an estimator of m.

Remark 3.4. Under the SRCM, we can use a parametric estimator $m_n(\cdot) = m(\cdot, \theta_n)$ and substitute it into (3.7). Setting θ_n to be the MLE, given in Definition 3.7 below, results in the semi-parametric estimator $F^{SE}_{1,n}$ introduced in Dikta (2000):

$$1 - F^{SE}_{1,n}(t) \coloneqq \prod_{i:Z_i \leq t} \left[1 - \frac{m(Z_i, \theta_n)}{n - R_n(Z_i) + 1}\right].$$

Remark 3.5. If, besides the RCM, δ is independent of Z, the model is equivalent to the simple proportional hazards model (PHM) as considered in Koziol and Green (1976). In this case $m(z) = \mathbb{E}[\delta]$ for all $z \in \mathbb{R}_{\geq}$ and a suitable estimator of m is given by

$$m_n(Z_{k:n}, \theta_n) = \theta_n = \frac{1}{n} \sum_{i=1}^{n} \delta_i, \qquad \forall k = 1, \ldots, n.$$

Plugging this m_n into (3.7) gives a slightly modified version of an estimator F_n^{ACL}, which was first considered by Abdushukurov (1987) and Cheng and Lin (1987). To be precise, the estimator $\dot{F}_{1,n}^{SE}$, defined in (3.14) below, is equivalent to F_n^{ACL} under the PHM when setting $m(Z_{k:n}, \theta_n) = 1/n \sum_{i=1}^{n} \delta_i$. It was shown, that this estimator is more efficient than the Kaplan-Meier PLE in terms of asymptotic variance under the PHM.

Remark 3.6. When assuming that m satisfies certain smoothness conditions, we can use a nonparametric regression estimator of m; cf. Definition 3.8. In this case (3.7) becomes the presmoothed Kaplan-Meier estimator $F_{1,n}^{PR}$ as introduced in Ziegler (1995) and Cao, López-de Ullibarri, Janssen, and Veraverbeke (2005).

Definition 3.7. Similar to Remark 3.4, under the SRCM, we have $m_n(\cdot) = m(\cdot, \theta_n)$ for all $k = 1, \ldots, n$. Then $F_{2,n}^{*}$ induces the estimator

$$\begin{aligned} 1 - F_{2,n}^{SE}(t) &:= \prod_{i:Z_i \leq t} \left[1 - \frac{m(Z_i, \theta_n)}{n - R_n(Z_i) + m(Z_i, \theta_n)} \right] \\ &= \prod_{i:Z_i \leq t} \left[\frac{n - R_n(Z_i)}{n - R_n(Z_i) + m(Z_i, \theta_n)} \right], \end{aligned}$$

where θ_n is the MLE of θ_0. In particular θ_n is the maximizer of the (partial) likelihood function

$$L_n(\theta) = \prod_{i=1}^{n} m(Z_i, \theta)^{\delta_i} \cdot (1 - m(Z_i, \theta))^{1-\delta_i}, \qquad \theta_n := \arg\max_{\theta \in \Theta} L_n(\theta).$$

This estimator was first introduced and investigated in Dikta, Reißel, and Harlaß (2016b). Often instead of the likelihood function the (partial) log-likelihood function

$$l_n(\theta) = n^{-1} \sum_{i=1}^{n} [\delta_i w_1(Z_i, \theta) + (1 - \delta_i) w_2(Z_i, \theta)],$$

is used to determine θ_n in the previous definition where

$$w_1(x, \theta) = \ln(m(x, \theta)) \qquad \text{and} \qquad w_2(x, \theta) = \ln(1 - m(x, \theta)). \tag{3.11}$$

Similar to the semi-parametric case, we can rely on the estimators for m from the Remarks 3.3, 3.5 and 3.6 to plug them into the prototype estimator $F_{2,n}^*$. If only assuming the RCM, we use again $m_n(Z_k) = \delta_k$ which together with $F_{2,n}^*$ results in an estimator almost identical to the already mentioned Kaplan-Meier PLE F_n^{KM}. In fact, the definitions only differ in the mass assigned to the largest observation. Since the semi-parametric approach is a generalization of the Cheng-Lin estimator, there is not much value in considering the PHM separately.

Definition 3.8. If we assume, in addition to the RCM, that m is a smooth function we can employ a preliminary nonparametric estimator of m. Then $F_{2,n}^*$ defines a new type of presmoothed PLE of F which we are going to denote by $F_{2,n}^{PR}$:

$$1 - F_{2,n}^{PR}(t) := \prod_{i: Z_i \le t} \left[\frac{n - R_{i,n}}{n - R_{i,n} + p_n(Z_i)} \right],$$

where $m_n = p_n$ is some nonparametric estimator of m, for example the Nadaraya (1964) and Watson (1964) estimator

$$p_n(t) = \frac{n^{-1} \sum_{i=1}^{n} \delta_i b_n^{-1} K\left(\frac{t - Z_i}{b}\right)}{n^{-1} \sum_{i=1}^{n} b_n^{-1} K\left(\frac{t - Z_i}{b}\right)},$$

as used in Ziegler (1995) and Cao and Jácome (2004). Here K is some probability kernel and $(b_n)_{n>1}$ a series of bandwidths.

3.2 Properties of PLEs under the RCM

The estimators derived in Section 3.1 are partially very well studied and a variety of results is available. Often we are interested in quantities of the underlying lifetime X, which can be expressed as an integral w.r.t. F of some Borel-measurable function φ; for example expectation, variance or simply $F(t)$ for some fixed $t \in \mathbb{R}_{\geq}$. Let F_n^* be some estimator of F, then these quantities can be estimated by $\int_0^\infty \varphi dF_n^*$. To ensure the quality of such integral estimates, one is usually interested in some kind of SLLN, i.e.

$$\int_0^\infty \varphi dF_n^* \xrightarrow[n\to\infty]{a.s.} \int_0^{\tau_H} \varphi dF \tag{3.12}$$

for $\tau_H = \inf\{x : H(x) = 1\}$. Moreover, the limiting distribution is an important property of such an integral estimate and is vital for the construction of confidence intervals. Commonly those PLEs are asymptotically normal distributed, e.g.

$$n^{1/2}\left(\int_0^\infty \varphi dF_n^* - \int_0^{\tau_H} \varphi dF\right) \xrightarrow[n\to\infty]{} \mathcal{N}(0, \sigma_*^2) \qquad \text{in distribution.} \tag{3.13}$$

For the already established estimators presented in Section 3.1, results like (3.12) and (3.13) are available. In case of uncensored data, those properties are provided by the ordinary SLLN and CLT, respectively; cf. Cohn (2013, Theorem 10.2.5 and Theorem 10.3.16).

If we make no further assumptions besides the RCM, then the Kaplan-Meier estimator F_n^{KM} from Remark 3.3 is the natural choice. It is by far the most popular estimator for analyzing censored data and is often applied in practice. Under some weak assumptions, (3.12) and (3.13) also hold for F_n^{KM}. In particular, strong consistency of integrals w.r.t. F_n^{KM} was shown in Stute and Wang (1993) and Stute (1995) proved that (3.13) holds under some weak assumptions for F_n^{KM}, where the asymptotic variance $\sigma^2_{F,KM}$ is given in Stute (1995, Corollary 1.2). Thereby Stute extended the work presented in Breslow and Crowley (1974), Gill (1983) and Schick, Susarla, and Koul (1988). Asymptotic optimality of F_n^{KM} was studied in Wellner (1982).

As already seen in Remark 3.5, $F_{1,n}^{SE}$ is almost identical to F_n^{ACL} under the PHM. Hence $F_{1,n}^{SE}$ can be seen as a generalization of F_n^{ACL}, as described in Dikta (2000, Example 1.3). Some of its basic properties are reviewed in Csörgő (1988). Almost sure (a.s.) consistency is given in Stute (1992) and asymptotic normality was proven in Dikta (1995).

As mentioned in Remark 3.6, when assuming that m suffices certain smoothness conditions, it is possible to estimate m by some nonparametric estimator, which, in combination with the prototype estimator $F_{1,n}^*$, gives $F_{1,n}^{PR}$. Along with other results, Cao et al. (2005) provided an a.s. asymptotic representation of $F_{1,n}^{PR}$ and Jácome and Cao (2007) studied its asymptotic distribution. In particular, they proved (3.12) and (3.13) for the special case of $\varphi(t) = \mathbb{1}_{[0,x]}(t)$ for all $x \leq \tau_H$. Dikta, Külheim, Mendonça, and de Uña-Álvarez (2016a) obtained a CLT for presmoothed Kaplan-Meier integrals with covariates.

3.2.1 The semi-parametric PLE $F_{1,n}^{SE}$

In Section 3.2 it is shown that the difference between the semi-parametric estimators $F_{2,n}^{SE}$ and $F_{1,n}^{SE}$ is asymptotically negligible. Hence, $F_{2,n}^{SE}$ inherits some of its properties from $F_{1,n}^{SE}$. For this reason we will give a short overview of $F_{1,n}^{SE}$.

As defined in Remark 3.4 the estimator is given by

$$1 - F_{1,n}^{SE}(t) = \prod_{i:Z_i \leq t} \left(1 - \frac{m(Z_i, \theta_n)}{n - R_n(Z_i) + 1}\right), \tag{3.14}$$

where $m(\cdot, \cdot)$ is the parametric model as described in Definition 2.5 and θ_n the MLE of the true parameter as explained in Definition 3.7. The estimator was proposed in Dikta (2000), where it is also stated that $F_{1,n}^{SE}$ is very close to the semi-parametric estimator introduced in

Dikta (1998):

$$1-\dot{F}_{1,n}^{SE}(t) \coloneqq \prod_{i:Z_i\le t}\left(\frac{n-R_n(Z_i)}{n-R_n(Z_i)+1}\right)^{m(Z_i,\theta_n)}. \tag{3.15}$$

In particular, Lemma 3.11 shows that $\dot{F}_{1,n}^{SE}$ and $F_{1,n}^{SE}$ are asymptotically identical. When making use of the order statistics $(Z_{i:n})_{1\le i\le n}$ of the Z-values, $F_{1,n}^{SE}$ may be written as

$$1-F_{1,n}^{SE}(Z_{i:n}) = \prod_{k=1}^{i}\left(1-\frac{m(Z_{k:n},\theta_n)}{n-k+1}\right).$$

For some Borel integrable function $\varphi : \mathbb{R}_{\ge} \mapsto \mathbb{R}$, it holds that

$$\int \varphi dF_{1,n}^{SE} = \sum_{i=1}^{n}\varphi(Z_{i:n})W_{1,i,n}^{SE}(\theta_n), \tag{3.16}$$

where, for all $1\le i\le n$, $W_{1,i,n}^{SE}(\theta_n)$ is the weight assigned to the observation $Z_{i:n}$, which is

$$W_{1,i,n}^{SE}(\theta_n) \coloneqq F_{1,n}^{SE}(Z_{i:n}) - F_{1,n}^{SE}(Z_{i-1:n}) = \frac{m(Z_{i:n},\theta_n)}{n-i+1}\prod_{k=1}^{i-1}\left(1-\frac{m(Z_{k:n},\theta_n)}{n-k+1}\right). \tag{3.17}$$

Equivalent to $F_{1,n}^{SE}$, it is possible to define a semi-parametric version of the Nelson (1972)-Aalen (1978) estimator, compare Dikta (1998, p. 255),

$$\Lambda_{1,n}^{SE}(t) \coloneqq \int_0^t \frac{m(x,\theta_n)}{1-H_n(x^-)}H_n(dx) = \sum_{i:Z_i\le t}\frac{m(Z_i,\theta_n)}{n-R_n(Z_i)+1}. \tag{3.18}$$

Similar to the other estimators, the results available for $F_{1,n}^{SE}$ rely on some assumptions which we list here.

(A1) There exists a measurable solution $\theta_n \in \Theta$ of the equation $\nabla(\ln(L_n(\theta))) = 0$ converging to the true θ_0 in probability as $n\to\infty$.

(A2) There exists a measurable solution $\theta_n \in \Theta$ of the equation $\nabla(\ln(L_n(\theta))) = 0$ converging to the true θ_0 a.s. as $n\to\infty$.

(A3) For $1 \leq r \leq k$,

$$\mathbb{E}\left(\left[\frac{\nabla_r m(Z,\theta_0)}{m(Z,\theta_0)}\right]^2\right) < \infty \qquad \text{and} \qquad \mathbb{E}\left(\left[\frac{\nabla_r m(Z,\theta_0)}{1-m(Z,\theta_0)}\right]^2\right) < \infty.$$

(A4) For $i = 1, 2$, $w_i(z, \theta)$, as given by (3.11), possesses continuous second order partial derivatives w.r.t. θ for all $\theta \in \Theta$ and $z \geq 0$. Furthermore $\nabla_{r,s} w_i(\cdot, z)$ is measurable for all $\theta \in \Theta$ and there exists a neighborhood $V(\theta_0) \subset \Theta$ of θ_0 and a measurable function M such that for all $\theta \in V(\theta_0)$, $z \geq 0$ and $1 \leq r, s \leq k$

$$|\nabla_{r,s} w_1(z,\theta)| + |\nabla_{r,s} w_2(z,\theta)| \leq M(z) \qquad \text{and} \qquad \mathbb{E}(M(Z)) < \infty.$$

(A5) The matrix $I(\theta_0) = (\sigma_{r,s})_{1 \leq r,s \leq k}$ with $w(\delta, z, \theta) = \delta w_1(z,\theta) + (1-\delta) w_2(z,\theta)$ and

$$\sigma_{r,s} = -\mathbb{E}(\nabla_{r,s} w(\delta, Z, \theta_0)) = \mathbb{E}\left(\frac{\nabla_r(m(Z,\theta_0))\nabla_s(m(Z,\theta_0))}{m(Z,\theta_0)(1-m(Z,\theta_0))}\right)$$

is positive definite.

(A6) There exists a neighborhood $V(\theta_0) \subset \Theta$ of θ_0 such that $m(z,\theta)$ possesses continuous second order derivatives w.r.t. θ for all $\theta \in \Theta$ and $z \geq 0$. Furthermore, for all $\theta \in V(\theta_0)$ and $1 \leq r, s \leq k$, $\nabla_{r,s} m(\cdot, \theta)$ is measurable, and

$$\sup_{0 \leq z < \infty} \|\nabla m(z,\theta_0)\| < \infty \qquad \text{and} \qquad \sup_{\theta \in V(\theta_0)} \sup_{0 \leq z < \infty} \sum_{1 \leq r,s \leq k} |\nabla_{r,s} m(z,\theta)| < \infty.$$

(A7) For $1 \leq r \leq k$, $\nabla_r m(\cdot.\theta_0)$ is Lipschitz continuous on $[0, T]$ for all $T < \tau_H$, i.e.

$$|\nabla_r m(x,\theta_0) - \nabla_r m(y,\theta_0)| \leq c\,|x-y|$$

for an appropriate constant c, possibly depending on T. Here, $\tau_H = \inf\{x : H(x) = 1\}$.

(A8) $m(\cdot, \theta_0)$ is of bounded variation on $[0, \tau_H]$, i.e.

$$\sup \left\{ \sum_{i=1}^{l} |m(z_i, \theta_0) - m(z_{i-1}, \theta_0)| : 0 = z_o \leq z_1 \leq \ldots \leq z_l \leq \tau_H,\, l \geq 1 \right\} < \infty.$$

(A9) For each $\epsilon > 0$ there exists a neighborhood $V(\epsilon, \theta_0) \subset \Theta$ of θ_0 such that for all $\theta \in V(\epsilon, \theta_0)$

$$\sup_{0 \leq z} |m(z, \theta) - m(z, \theta_0)| < \epsilon.$$

For some of the results, φ has to satisfy certain moment conditions:

(M1) $\int_0^{\tau_H} \varphi^2(x)\gamma_0(x)F(dx) < \infty$,

(M2) $\int_0^{\tau_H} \frac{|\varphi(x)|}{(1-H(x))^{1/2}} F(dx) < \infty$,

(M3) $\int_0^{\tau_H} |\varphi(x)|\, \gamma_0(x)H(dx) < \infty$,

(M4) $\int_0^{\tau_H} \frac{|\varphi(x)|}{m(x,\theta_0)(1-H(t))^{\epsilon}} F(dx) < \infty$ for some $\epsilon > 0$,

where γ_0 as defined in Theorem 3.10. Assumptions (A1), (A3) to (A5) are necessary to ensure the asymptotic normality of the MLE θ_n. For the a.s. results we have to strengthen the assumption (A1) to strong consistency in (A2).

Dikta (1998, Theorem 2.4) proves uniform consistency of $\dot{F}_{1,n}^{SE}$ and Dikta (1998, Corollary 2.6) gives a functional central limit theorem of the process $n^{1/2}(\dot{F}_{1,n}^{SE} - F)$. Both results are valid on the compact interval $[0, T]$ with $H(T) < 1$. Moreover, Dikta (1998, Corollary 2.7) shows that $\dot{F}_{1,n}^{SE}$ is more efficient than the Kaplan-Meier estimator F_n^{KM} in terms of asymptotic variance under the SRCM. Since Lemma 3.11 shows that $\dot{F}_{1,n}^{SE}$ and $F_{1,n}^{SE}$ are asymptotically equivalent, those results also hold true for $F_{1,n}^{SE}$; cf. Dikta (2000, p. 3). Furthermore, Dikta (2000, Theorem 1.1) established a SLLN for integrals w.r.t. $F_{1,n}^{SE}$, which we quote in the next theorem.

Theorem 3.9. Assuming that H is continuous and if the assumptions (A2), (A9) and (M4) are satisfied, then with $\tau_H = \inf\{x : H(x) = 1\}$

$$\int_0^\infty \varphi(t) F_{1,n}^{SE}(dt) \xrightarrow[n\to\infty]{a.s.} \int_0^{\tau_H} \varphi(t) F(dt).$$

Therefore, (3.12) holds for $F_{1,n}^{SE}$ under some weak assumptions. Similar to Stute (1995), the proof of Theorem 3.9 relies on martingale theory.

Dikta, Ghorai, and Schmidt (2005) extended the CLT to the SRCM, i.e. they proved the following theorem.

Theorem 3.10. Let Θ be a connected open subset of $\mathbb{R}^k$. Assuming that H is continuous, (A1),(A3)–(A8), and (M1)–(M3) are satisfied, then

$$n^{1/2}\left(\int_0^\infty \varphi dF_{1,n}^{SE} - \int_0^{\tau_H} \varphi dF\right) \xrightarrow[n\to\infty]{} \mathcal{N}(0, \sigma_{F,SE}^2) \qquad \text{in distribution} \tag{3.19}$$

where
$$\begin{aligned}\sigma_{F,SE}^2 = &\operatorname{Var}\Bigg(\varphi(Z)\gamma_0(Z)m(Z,\theta_0) + (1-m(Z,\theta_0))\gamma_1(Z) - \gamma_2(Z) \\ &\qquad - \frac{\delta - m(Z,\theta_0)}{m(Z,\theta_0)(1-m(Z,\theta_0))}(\gamma_3(Z) - \gamma_4(Z))\Bigg),\end{aligned}$$

with

$$\begin{aligned}
\gamma_0(z) &= \exp\left(\int \frac{\mathbb{1}_{\{t<z\}}}{1-H(t)} H^0(dt)\right), \\
\gamma_1(z) &= \frac{1}{1-H(z)} \int \mathbb{1}_{\{z<t\}} \varphi(t)\gamma_0(t) H^1(dt), \\
\gamma_2(z) &= \int\int \frac{\mathbb{1}_{\{z>x,t>x\}}\varphi(t)\gamma_0(t)}{(1-H(x))^2} H^1(dt)H^0(dx), \\
\gamma_3(z) &= \int\int \frac{\mathbb{1}_{\{t>x\}}\alpha(x,z)\varphi(t)\gamma_0(t)}{1-H(x)} H^1(dt)H(dx), \\
\gamma_4(z) &= \int \varphi(t)\gamma_0(t)\alpha(t,z)H(dt) \text{ and}
\end{aligned}$$

$\alpha(x,y) = \langle \nabla m(x,\theta_0) | I^{-1}(\theta_0)\nabla m(y\theta_0)\rangle$, where $\langle\cdot,\cdot\rangle$ denotes the inner product.

Already Dikta (1998) pointed out, that the asymptotic variance of $F_{1,n}^{SE}$ is less or equal to the asymptotic variance of the Kaplan-Meier estimator F_n^{KM}, if the correct model for m is assumed. When looking at Dikta (1998, Corollary 2.7) it is evident, that equality only occurs in exceptional cases. Besides the latter theorem, Dikta et al. (2005), proved that $\sigma^2_{F,SE} \leq \sigma^2_{F,KM}$ under the SRCM, where $\sigma^2_{F,KM}$ is the asymptotic variance of (3.13) in the case of $F_n^* = F_n^{KM}$, cf. Stute (1995, Corollary 1.2). Corollary 2.5 and Remark 2.6 of Dikta et al. (2005) explain why the increase in efficiency is strict, in almost all reasonable cases, that is $\sigma^2_{F,SE} < \sigma^2_{F,KM}$. Dikta (2014) actually showed that $F_{1,n}^{SE}$ is asymptotically efficient w.r.t. the class of all regular estimators of $\int_0^{\tau_H} \varphi dF$ given the SRCM.

Since it has never been stated explicitly in the literature, the next lemma shows that $\dot{F}_{1,n}^{SE}$ from Dikta (1998) and $F_{1,n}^{SE}$ as defined in Dikta (2000) are asymptotically identical.

Lemma 3.11. Assuming (A2) and (A10), it holds for $0 \leq T < \tau_H$ that

$$\sup_{0\leq t\leq T} |\dot{F}_{1,n}^{SE}(t) - F_{1,n}^{SE}(t)| \overset{a.s.}{=} \mathcal{O}(n^{-1}).$$

3.2.2 The semi-parametric PLE $F_{2,n}^{SE}$

In the previous sections it was shown that, under certain assumptions, (3.12) and (3.13) hold for the estimators F_n^{KM}, F_n^{ACL}, $F_{1,n}^{PR}$ and $F_{1,n}^{SE}$. Here we are going to examine the estimator

$$1 - F_{2,n}^{SE}(Z_{i:n}) = \prod_{k=1}^{i} \left[\frac{n-k}{n-k+m(Z_{k:n},\theta_n)} \right],$$

which we proposed in Definition 3.7. Due to their complexity, most of the proofs for the following theorems are postponed to Section 3.3.

First note that

$$1 - F_{2,n}^{SE}(Z_{n:n}) = \left[1 - \frac{m(Z_{n:n},\theta_n)}{m(Z_{n:n},\theta_n)}\right] \prod_{k=1}^{n-1} \left[1 - \frac{m(Z_{k:n},\theta_n)}{n - i + m(Z_{k:n},\theta_n)}\right] = 0,$$

which shows that the entire weight of one gets distributed among the data points. By definition $F_{2,n}^{SE}(t) = 0$ for all $t < Z_{1:n}$, $F_{2,n}^{SE}$ is monotonically increasing, right-continuous and its left-hand limits exist. Hence we immediately have the following consequence.

Corollary 3.12. The estimator $F_{2,n}^{SE}$ is always a proper probability distribution function.

A Nelson-Aalen type estimator based on $F_{2,n}^{SE}$ can be defined by

$$\begin{aligned}\Lambda_{2,n}^{SE}(t) &:= \int_0^t \frac{1}{1 - F_{2,n}^{SE}(x^-)} F_{2,n}^{SE}(dx) = \sum_{i:Z_{i:n} \le t} \frac{1}{1 - F_{2,n}^{SE}(Z_{i:n}{}^-)} W_{2,i,n}^{SE}(\theta_n) \\ &= \sum_{i:Z_i \le t} \frac{m(Z_i,\theta_n)}{n - R_n(Z_i) + m(Z_i,\theta_n)} = \int_0^t \frac{m(x,\theta_n)}{\bar{H}_n(x) + m(x,\theta_n)/n} H_n(dx). \end{aligned} \tag{3.20}$$

Furthermore, for a Borel-measurable function $\varphi : \mathbb{R}_{\ge} \mapsto \mathbb{R}$ it holds that

$$\int \varphi dF_{2,n}^{SE} = \sum_{i=1}^{n} \varphi(Z_{i:n}) W_{2,i,n}^{SE}(\theta_n), \tag{3.21}$$

where, for all $1 \le i \le n$, $W_{2,i,n}^{SE}(\theta_n)$ is the weight assigned to $Z_{i:n}$ by $F_{2,n}^{SE}$. To be precise

$$\begin{aligned} W_{2,i,n}^{SE}(\theta_n) :=& F_{2,n}^{SE}(Z_{i:n}) - F_{2,n}^{SE}(Z_{i-1:n}) \\ =& \prod_{k=1}^{i-1} \left(1 - \frac{m(Z_{k:n},\theta_n)}{n - k + m(Z_{k:n},\theta_n)}\right) \\ & \qquad - \left(1 - \frac{m(Z_{i:n},\theta_n)}{n - i + m(Z_{i:n},\theta_n)}\right) \prod_{k=1}^{i-1} \left(1 - \frac{m(Z_{k:n},\theta_n)}{n - k + m(Z_{k:n},\theta_n)}\right) \\ =& \frac{m(Z_{i:n},\theta_n)}{n - i + m(Z_{i:n},\theta_n)} \prod_{k=1}^{i-1} \left(\frac{n - k}{n - k + m(Z_{k:n},\theta_n)}\right). \end{aligned} \tag{3.22}$$

Just by visual inspection, the difference between $F_{1,n}^{SE}$ and $F_{2,n}^{SE}$ seems to be minor. In fact, from Theorem 3.13 and Theorem 3.16 we can conclude that the difference is asymptotically negligible. Theorem 3.13 shows that $F_{1,n}^{SE}$ and $F_{2,n}^{SE}$ are stochastically equivalent under some assumptions, among them:

(A10) $0 < m(x, \theta) \leq 1$ for all $x > 0$ and for all θ in an open neighborhood of θ_0.

(M5) $\int \frac{|\varphi|}{(1-H)^{1.5+\epsilon}} dH < \infty$ for some $\epsilon > 0$.

Theorem 3.13. If F and G are continuous, assumptions (A1), (A10) and (M5) are satisfied, then

$$n^{1/2} \left(\int \varphi dF_{2,n}^{SE} - \int \varphi dF_{1,n}^{SE} \right) \xrightarrow[n \to \infty]{} 0 \quad \text{in probability.}$$

The last theorem together with Theorem 3.10 immediately yields the following CLT result.

Corollary 3.14. Given the assumptions of Theorem 3.10 and Theorem 3.13, it holds that

$$n^{1/2} \left(\int_0^\infty \varphi dF_{2,n}^{SE} - \int_0^{\tau_H} \varphi dF \right) \xrightarrow[n \to \infty]{} \mathcal{N}(0, \sigma_{F,SE}^2) \quad \text{in distribution}$$

with the asymptotic variance $\sigma_{F,SE}^2$ as given in Theorem 3.10.

Note that Dikta et al. (2005, p. 31) give an estimator of $\sigma_{F,SE}^2$ which can be used to obtain confidence intervals. In addition, Dikta et al. (2005, Theorem 2.1) provided the following asymptotic representation of $F_{1,n}^{SE}$-integrals which, due to the latter corollary, also holds for integrals w.r.t. $F_{2,n}^{SE}$.

Corollary 3.15. Let Θ be a connected open subset of $\mathbb{R}$. Then under the assumptions of Theorem 3.13 and Dikta et al. (2005, Theorem 2.1), that is F and G are continuous, (A1), (A3) to (A8), (A10), (M1) to (M3), and (M5), it holds that

$$\begin{aligned}\int_0^\infty \varphi F_{2,n}^{SE} =& \frac{1}{n}\sum_{i=1}^n \varphi(Z_i)\gamma_0(Z_i)m(Z_i,\theta_0) + \frac{1}{n}\sum_{i=1}^n (1-m(Z_i,\theta_0))\gamma_1(Z_i) - \frac{1}{n}\sum_{i=1}^n \gamma_2(Z_i)\\ &- \frac{1}{n}\sum_{i=1}^n \frac{\delta - m(Z_i,\theta_0)}{m(Z_i,\theta_0)(1-m(Z_i,\theta_0))}\gamma_3(Z_i)\\ &+ \frac{1}{n}\sum_{i=1}^n \frac{\delta - m(Z_i,\theta_0)}{m(Z_i,\theta_0)(1-m(Z_i,\theta_0))}\gamma_4(Z_i) + o_p(n^{-1/2}),\end{aligned}$$

where γ_0 to γ_4 are given in Theorem 3.10.

A similar result as given in Theorem 3.13 holds true almost surely under the assumption

(M6) $\int \frac{|\varphi|}{(1-H)^{1+\epsilon}} dH < \infty$ for some $\epsilon > 0$.

Theorem 3.16. Given that F and G are continuous and assumptions (A2), (A10), and (M6) are satisfied, then

$$\left|\int \varphi dF_{2,n}^{SE} - \int \varphi dF_{1,n}^{SE}\right| \xrightarrow[n\to\infty]{a.s.} 0.$$

Note that with (A2) we require the MLE θ_n to be strongly consistent while we weaken the moment assumption to (M6), in comparison to Theorem 3.13. In the special case of $\varphi(t) = \mathbb{1}_{[0,x]}(t)$ we can give a uniform a.s. convergence order.

Theorem 3.17. Assuming the conditions of Theorem 3.16 hold for $\varphi(t) = \mathbb{1}_{[0,x]}(t)$ for all $x \le T < \tau_H$, we have

$$\sup_{0\le x\le T} \left|F_{2,n}^{SE}(x) - F_{1,n}^{SE}(x)\right| \stackrel{a.s.}{=} \mathcal{O}(n^{-1}).$$

Using the last theorem in combination with Dikta (2000, Corollary 1.4) we easily deduce the next result.

Corollary 3.18. Given the assumptions of Theorem 3.9 and Theorem 3.17 then

$$F_{2,n}^{SE}(x) \xrightarrow[n\to\infty]{a.s.} F(x).$$

Both, Theorem 3.17 together with Dikta (2000, Corollary 1.5) or Corollary 3.18 in combination with Loeve (1977, p. 21) can be used to enhance the last result to uniform convergence.

Corollary 3.19. Assuming the conditions of Theorem 3.9 and Theorem 3.16 hold for $\varphi(t) = \mathbb{1}_{[0,x]}(t)$ for all $x \le T < \tau_H$, we have

$$\sup_{0\le x\le T} \left|F_{2,n}^{SE}(x) - F(x)\right| \xrightarrow[n\to\infty]{a.s.} 0.$$

Theorem 3.9 together with Theorem 3.16 yields the following strong law result.

Corollary 3.20. Given the assumptions of Theorem 3.9 and Theorem 3.16 then

$$\int_0^\infty \varphi dF_{2,n}^{SE} \xrightarrow[n\to\infty]{a.s.} \int_0^{\tau_H} \varphi dF.$$

As seen, Theorem 3.13 and Theorem 3.16 are the key to our analysis of the asymptotic properties of $F_{2,n}^{SE}$. By those arguments $F_{2,n}^{SE}$ also inherits the efficiency properties of $F_{1,n}^{SE}$.

Remark 3.21. Theorem 3.13 shows that the integral estimators w.r.t. $F_{1,n}^{SE}$ and $F_{2,n}^{SE}$ admit the same asymptotic variance $\sigma_{F,SE}^2$. Then due to Dikta (2014, Corollary 3.11), $\int_0^{\tau_H} \varphi dF_{2,n}^{SE}$ is a regular estimator and is asymptotically efficient w.r.t. the class of all regular estimators of $\int_0^{\tau_H} \varphi dF$ given the SRCM.

3.2.3 Discussion of the estimators

The Kaplan-Meier estimator assigns mass only to uncensored observations while the attached weight increases from the smallest to the largest observation; cf. Efron (1967). This is particularly critical if the last observation is censored. In this case the Kaplan-Meier estimator F_n^{KM} lacks the important property of being a proper d.f., that is $\lim_{t\to\infty} F_n^{KM}(t) < 1$ since F_n^{KM} fails to attach the total mass of one to the observations. This deficit becomes even more apparent when noting that the estimator is designed to put the largest amount of weight on last observation. Hence, neglecting this mass could cause significant bias. Because the weight assigned to a data point depends on the number of preceding censored observations, this behavior gets amplified under high censoring rates. Also $F_{1,n}^{PR}$ and $F_{1,n}^{SE}$ suffer from the same shortcoming if $m_n(Z_{n:n}) \neq 1$ and therefore are only subdistribution functions. There are methods to fix this disadvantage, for example rescaling the weights or simply assigning the missing weight to the largest observation while completely ignoring the censoring indicator, but those are not well studied or cause an unreasonable bias. As discussed above, $\dot{F}_{1,n}^{SE}$, as defined in (3.15), is a slightly modified version of $F_{1,n}^{SE}$. Even that $\dot{F}_{1,n}^{SE}(Z_{n:n}) = 1$, there is a similar problem: in case of high censoring rates, especially of the larger observations, $\dot{F}_{1,n}^{SE}$ attaches an unrealistic amount of mass to the largest observation.

However, Corollary 3.12 shows that $F_{2,n}^{SE}$ is a true probability function under every circumstance, while still providing a more realistic distribution of the total mass. This is a big advantage in comparison to F_n^{KM}, $F_{1,n}^{PR}$ and $F_{1,n}^{SE}$ when using those as plug-in estimators in the approximation of linear functionals. Thereby the missing weight could cause some unreasonable bias especially in case of small sample sizes. Moreover, since $F_{2,n}^{SE}$ is a proper d.f., it is possible to use its quantile function to resample according to F. For example, sampling directly from $F_{2,n}^{SE}$ might improve the bootstrap based construction of confidence bands presented in Subramanian and Zhang (2013).

Theorem 3.13 and Theorem 3.16 show that the estimators $F_{1,n}^{SE}$ and $F_{2,n}^{SE}$ are asymptotically equivalent. Hence all asymptotic results available for $F_{1,n}^{SE}$ also hold true for $F_{2,n}^{SE}$. For instance, Corollary 3.14 shows that the integral estimators w.r.t. $F_{1,n}^{SE}$ and $F_{2,n}^{SE}$, respectively, admit the same asymptotic variance $\sigma^2_{F,SE}$. As discussed in Subsection 3.2.1 and Remark 3.21, $\sigma^2_{F,SE}$ is optimal w.r.t. the class of regular estimators of $\int_0^{\tau_H} \varphi dF$ and therefore $F_{2,n}^{SE}$ outperforms the corresponding Kaplan-Meier integral estimator assuming a correctly chosen parametric model for m. In fact, Theorem 3.13 together with Dikta (2014, Corollary 3.11) shows that both, $F_{1,n}^{SE}$ and $F_{2,n}^{SE}$, are more efficient than F_n^{KM}, $F_{1,n}^{PR}$ and $F_{2,n}^{PR}$ and can not be improved in means of asymptotic variance, when the model for m is chosen correctly. Since $F_{1,n}^{SE}$ and $F_{2,n}^{SE}$ incorporate the additional information of the parametric model, the achieved efficiency gain is something one would intuitively expect.

All results in Subsection 3.2.1 and Subsection 3.2.2 were derived under the SRCM, that is, assuming the correct parametric model for m. Simulation studies, conducted in Dikta, Hausmann, and Schmidt (2002), show that $F_{1,n}^{SE}$ still performs well even under wrong assumptions for m. As discussed, m is based on a parametric binary regression model. Hence it is possible to validate the model assumptions via goodness-of-fit tests. Dikta, Kvesic, and Schmidt (2006) presented a general bootstrap based test to verify the model assumptions.

The SLLN and CLT for integral estimators based on F_n^{KM}, obtained in Stute and Wang (1993) and Stute (1995), requires F and G from the RCM only not to have common jumps. This assumption is not well-suited for the SRCM, as explicated in Dikta (2000, Remark 1.6). Hence we expect F and G, and therefore H, to be continuous. However, in his proofs, Stute applied techniques to overcome this restriction. Those might also be applicable in the semi-parametric case.

3.3 Proving the properties of $F_{2,n}^{SE}$

In this section we will give the proofs for the results related to $F_{2,n}^{SE}$, primarily Theorem 3.13 and Theorem 3.16. The technique used for both theorems is very similar. Hence we will start proofing Theorem 3.16 and sketch the second proof more briefly. The following representation of the difference of two products will turn out to be an essential tool.

Remark 3.22. Let $(a_i)_{1\leq i\leq n}$ $(b_i)_{1\leq i\leq n}$ be two complex sequences. Then

$$\prod_{i=1}^{n} a_i - \prod_{i=1}^{n} b_i = \sum_{i=1}^{n}\left(\prod_{k=1}^{i-1} a_k(a_i - b_i)\prod_{k=i+1}^{n} b_k\right). \tag{3.23}$$

Proof. Trivially, this equation holds for $n = 1$ when interpreting empty products to be equal to one. Then using induction, assume that the equality holds for $n \in \mathbb{N}$ and consider

$$\sum_{i=1}^{n+1}\left(\prod_{k=1}^{i-1} a_k(a_i - b_i)\prod_{k=i+1}^{n+1} b_k\right) = \sum_{i=1}^{n}\left(\prod_{k=1}^{i-1} a_k(a_i - b_i)\prod_{k=i+1}^{n} b_k\right) b_{n+1} + (a_{n+1} - b_{n+1})\prod_{k=1}^{n} a_k$$

which, when applying the induction assumption (3.23), is equivalent to

$$= b_{n+1}\prod_{i=1}^{n} a_i - \prod_{i=1}^{n+1} b_i + \prod_{k=1}^{n+1} a_k - b_{n+1}\prod_{k=1}^{n} a_k,$$

and the proof is complete. A similar result can be found in Gill and Johansen (1990, Lemma 1). □

In addition, multiple times we will make use of the quantile representation of Z based on a uniformly distributed random variable; cf. (Shorack and Wellner, 1986, Theorem 1.1.1). Let $(U_i)_{1\leq i\leq n}$ be an i.i.d. sample from the uniform distribution on $[0, 1]$ with d.f. $\tilde{H}$. Hence Z_i and $H^{-1}(U_i)$ are equal in distribution for $i = 1, \ldots, n$. To shorten the notation we will write $Z_i = H^{-1}(U_i)$ for $i = 1, \ldots, n$.

The quantile function H^{-1} of H is defined by

$$H^{-1}(u) = \inf\{z : H(z) \geq u\}, \qquad 0 < u < 1.$$

Similarly, the empirical distribution and quantile function of the U-sample are denoted by $\tilde{H}_n$ and $\tilde{H}_n^{-1}$, respectively. H_n^{-1} is the empirical quantile function of the Z-sample. In the following we list some known results related to H, H_n, $\tilde{H}_n$ and their quantile functions. From Shorack and Wellner (1986, Theorem 1.1.2) we have

$$H_n(t) = \tilde{H}_n(H(t)), \qquad t \geq 0. \tag{3.24}$$

Relying on this equality and Shorack and Wellner (1986, p. 5, Eq. 21), we have for $0 < u < 1$

$$\begin{aligned} H_n^{-1}(u) &= \inf\{t : \tilde{H}_n(H(t)) \geq u\} = \inf\{t : H(t) \geq \tilde{H}_n^{-1}(u)\} \\ &= \inf\{t : t \geq H^{-1}(\tilde{H}_n^{-1}(u))\} = H^{-1}(\tilde{H}_n^{-1}(u)). \end{aligned} \tag{3.25}$$

Furthermore by Shorack and Wellner (1986, Proposition 1.1.1) and since H is considered to be continuous we have $H(H^{-1}(u)) = u$ for $0 < u < 1$. Together with (3.25) this yields

$$H(H_n^{-1}(u)) = \tilde{H}_n^{-1}(u), \qquad 0 < u < 1. \tag{3.26}$$

Moreover, first applying (3.24) and than using (3.26) gives

$$H_n(H_n^{-1}(u)) = \tilde{H}_n(\tilde{H}_n^{-1}(u)), \qquad 0 < u < 1. \tag{3.27}$$

When defining $\tilde{H}_n^{-1}(0) = 0$ and $\tilde{H}_n^{-1}(1) = U_{n:n}$, where $(U_{i:n})_{1\leq i\leq n}$ denotes the order statistics of the U-sample, we can extend the domain of $\tilde{H}_n^{-1}$ to the closed interval $[0, 1]$. Note that the sample points in $(U_i)_{1\leq i\leq n}$ are a.s. distinct. Hence we have a.s. for $\frac{k}{n} < u \leq \frac{k+1}{n}$ and

$k = 1, \dots, n-1$,

$$\tilde{H}_n(\tilde{H}_n^{-1}(u)) = \tilde{H}_n(U_{k+1:n}) = \frac{k+1}{n} < u + \frac{1}{n},$$

which in combination with Shorack and Wellner (1986, Proposition 1.1.1) gives

$$u \le \tilde{H}_n(\tilde{H}_n^{-1}(u)) \le u + \frac{1}{n}, \qquad 0 \le u \le 1. \tag{3.28}$$

For convenience of a brief notation, set $m_i = m(Z_{i:n}, \theta_n)$ and define

$$a_i = \frac{n-i}{n-i+m_i}, \qquad\qquad b_i = \frac{n-i+1-m_i}{n-i+1},$$

$\bar{a}_i = 1 - a_i$, and $\bar{b}_i = 1 - b_i$. Then (3.22) and (3.17) are equal to

$$W^{SE}_{2,i,n} = \bar{a}_i \prod_{k=1}^{i-1} a_k \qquad \text{and} \qquad W^{SE}_{1,i,n} = \bar{b}_i \prod_{k=1}^{i-1} b_k,$$

respectively.

Proof of Theorem 3.16.

Assume that $\varphi \ge 0$. Otherwise decompose φ into its positive and negative part, and proceed as follows. Using the previous abbreviations, (3.16) and (3.21) give

$$\begin{aligned}
&\int \varphi(x) F^{SE}_{2,n}(dx) - \int \varphi(x) F^{SE}_{1,n}(dx) \\
&= \sum_{i=1}^{n} \varphi(Z_{i:n}) \left(W^{SE}_{2,i,n}(\theta_n) - W^{SE}_{1,i,n}(\theta_n) \right) = \sum_{i=1}^{n} \varphi(Z_{i:n}) \left(\bar{a}_i \prod_{k=1}^{i-1} a_k - \bar{b}_i \prod_{k=1}^{i-1} b_k \right).
\end{aligned}$$

Then Remark 3.22 yields

$$
\begin{aligned}
&= \sum_{i=1}^{n} \varphi(Z_{i:n}) \left(\sum_{j=1}^{i-1} \left\{ \prod_{k=1}^{j-1} a_k (a_j - b_j) \bar{b}_i \prod_{k=j+1}^{i-1} b_k \right\} + (\bar{a}_i - \bar{b}_i) \prod_{j=1}^{i-1} a_k \right) \\
&= \sum_{i=1}^{n} \varphi(Z_{i:n}) \left(\sum_{j=1}^{i-1} \left\{ \prod_{k=1}^{j-1} a_k (a_j - b_j) \bar{b}_i \prod_{k=j+1}^{i-1} b_k \right\} \right) + \sum_{i=1}^{n} \varphi(Z_{i:n}) \left((\bar{a}_i - \bar{b}_i) \prod_{j=1}^{i-1} a_k \right) \\
&\equiv I_1(n) + I_2(n). \qquad (3.29)
\end{aligned}
$$

Note that for $j = 1, \ldots, n-1$

$$
\bar{b}_j - \bar{a}_j = a_j - b_j = \frac{m_j^2 - m_j}{(n-j+1)(n-j+m_j)} \leq 0
$$

and $a_n = 0$ since $m_n > 0$, and $b_n = 1 - m_n$. Furthermore, since $0 \leq a_j \leq 1$ and $0 \leq b_j \leq 1$ for all $j = 1, \ldots, n$, we have

$$
0 \leq \prod_{k=1}^{j-1} a_k \leq 1 \qquad \text{and} \qquad 0 \leq \prod_{k=j+1}^{i-1} b_k \leq 1.
$$

Then, because $(a_j - b_j) \leq 0$, for $j = 1, \ldots, n$,

$$
\begin{aligned}
|I_1(n)| &= -\sum_{i=1}^{n} \varphi(Z_{i:n}) \left(\sum_{j=1}^{i-1} \left\{ \prod_{k=1}^{j-1} a_k (a_j - b_j) \bar{b}_i \prod_{k=j+1}^{i-1} b_k \right\} \right) \\
&\leq -\sum_{i=1}^{n} \varphi(Z_{i:n}) \bar{b}_i \left(\sum_{j=1}^{i-1} (a_j - b_j) \right) \\
&= \sum_{i=1}^{n} \varphi(Z_{i:n}) \bar{b}_i \left(\sum_{j=1}^{i-1} \frac{m_j - m_j^2}{(n-j+1)(n-j+m_j)} \right) \\
&\leq \sum_{i=1}^{n} \varphi(Z_{i:n}) \bar{b}_i \left(\sum_{j=1}^{i-1} \frac{1}{(n-j+1)(n-j)} \right).
\end{aligned}
$$

With $(n-j+1)^{-1}(n-j)^{-1} = (n-j)^{-1} - (n-j+1)^{-1}$ and using telescoping sums

$$
\begin{aligned}
&= \sum_{i=1}^{n} \varphi(Z_{i:n}) \frac{m_i}{n-i+1} \left(\frac{1}{n-i+1} - \frac{1}{n} \right) \\
&\leq \sum_{i=1}^{n} \varphi(Z_{i:n}) \frac{m_i}{(n-i+1)^2} \\
&\leq \sum_{i=1}^{n-2} \frac{\varphi(Z_{i:n})}{(n-i)^2} + \varphi(Z_{n-1:n}) + \varphi(Z_{n:n}).
\end{aligned} \tag{3.30}
$$

Furthermore, because $I_2(n)$ is positive,

$$
\begin{aligned}
I_2(n) &\leq \sum_{i=1}^{n} \varphi(Z_{i:n}) \frac{m_j - m_j^2}{(n-j+1)(n-j+m_j)} \\
&\leq \sum_{i=1}^{n-2} \frac{\varphi(Z_{i:n})}{(n-i)^2} + \varphi(Z_{n-1:n}) + \varphi(Z_{n:n}).
\end{aligned} \tag{3.31}
$$

When defining $B(n) \coloneqq \sum_{i=1}^{n-2} \frac{\varphi(Z_{i:n})}{(n-i)^2}$, we have shown that

$$
\left| \int \varphi(x) F_{2,n}^{SE}(dx) - \int \varphi(x) F_{1,n}^{SE}(dx) \right| \leq 2\Big(B(n) + \varphi(Z_{n-1:n}) + \varphi(Z_{n:n}) \Big).
$$

Then together with $nH_n(Z_{i:n}) = i$ for all $1 \leq i \leq n$ we have

$$
\begin{aligned}
B(n) &= \sum_{i=1}^{n-2} \frac{\varphi(Z_{i:n})}{(n-i)^2} = \frac{1}{n^2} \sum_{i=1}^{n-2} \frac{\varphi(Z_{i:n})}{(\bar{H}_n(Z_{i:n}))^2} \\
&= \frac{1}{n} \int_0^{Z_{n-2:n}} \frac{\varphi(t)}{(\bar{H}_n(t))^2} H_n(dt) \\
&= \frac{1}{n} \int_0^{(n-2)/n} \frac{\varphi(H_n^{-1}(u))}{(\bar{H}_n(H_n^{-1}(u)))^2} du.
\end{aligned}
$$

Now exploiting the quantile representation of the Z-sample by using the results (3.25), (3.27) and (3.28), it holds that

$$
\begin{aligned}
B(n) &= \frac{1}{n}\int_0^{(n-2)/n} \frac{\varphi(H^{-1}(\tilde{H}_n^{-1}(u)))}{(1-\tilde{H}_n(\tilde{H}_n^{-1}(u)))^2}du \\
&= \frac{1}{n}\int_0^{(n-2)/n} \frac{\varphi(H^{-1}(\tilde{H}_n^{-1}(u)))}{(1-u)^2}\frac{(1-u)^2}{(1-\tilde{H}_n(\tilde{H}_n^{-1}(u)))^2}du \\
&< \frac{1}{n}\int_0^{(n-2)/n} \frac{\varphi(H^{-1}(\tilde{H}_n^{-1}(u)))}{(1-u)^2}\frac{(1-u)^2}{(1-u-1/n)^2}du.
\end{aligned}
$$

Since $(1-u)^2/(1-u-1/n)^2 \leq 4$ for $u \in [0,(n-2)/n]$

$$
\begin{aligned}
&\leq \frac{4}{n}\int_0^{(n-2)/n} \frac{\varphi(H^{-1}(\tilde{H}_n^{-1}(u)))}{(1-u)^2}du \qquad (3.32)\\
&= \frac{4}{n}\int_0^{(n-2)/n} \frac{\varphi(H^{-1}(\tilde{H}_n^{-1}(u)))}{(1-\tilde{H}_n^{-1}(u))^{1+\epsilon}}\frac{(1-\tilde{H}_n^{-1}(u))^{2+0.5\epsilon}}{(1-u)^2}\frac{(1-\tilde{H}_n^{-1}(u))^{1+\epsilon}}{(1-\tilde{H}_n^{-1}(u))^{2+0.5\epsilon}}du \\
&\leq \frac{4}{n}\sup_{0\leq u\leq (n-2)/n}\left(\frac{1}{(1-\tilde{H}_n^{-1}(u))^{1-0.5\epsilon}}\right)\sup_{0\leq u\leq (n-2)/n}\left(\frac{(1-\tilde{H}_n^{-1}(u))^{2+0.5\epsilon}}{(1-u)^2}\right) \\
&\qquad \times \int_0^{(n-2)/n} \frac{\varphi(H^{-1}(\tilde{H}_n^{-1}(u)))}{(1-\tilde{H}_n^{-1}(u))^{1+\epsilon}}du.
\end{aligned}
$$

Then, for ϵ small enough, $(1-x)^{0.5\epsilon-1}$ is monotone increasing in x, hence

$$
\begin{aligned}
\frac{4}{n}\sup_{0\leq u\leq (n-2)/n}\left(\frac{1}{(1-\tilde{H}_n^{-1}(u))^{1-0.5\epsilon}}\right) &\leq \frac{4}{n\left[1-\tilde{H}_n^{-1}\left(\frac{n-2}{n}\right)\right]^{1-0.5\epsilon}} \\
&= \frac{4}{n\left[1-U_{n-2:n}\right]^{1-0.5\epsilon}} \leq \frac{4}{n\left[1-U_{n:n}\right]^{1-0.5\epsilon}} \overset{a.s.}{=} \mathcal{O}\left(\frac{4}{(\log(n))^{1-0.5\epsilon}n^{0.5\epsilon}}\right).
\end{aligned}
$$

The last a.s. equality follows from $1-U_{n:n} \overset{a.s.}{=} \mathcal{O}\left(\frac{\log(n)}{n}\right)$ as shown in Robbins and Siegmund (1972, Theorem 1 (i)).

Furthermore, from Shorack and Wellner (1986, Theorem 10.6.1) we have for a fixed $0 < \delta < 1$, for n large enough and for all $u \in [0,(n-2)/n]$

$$
1-\tilde{H}_n^{-1}(u) \overset{a.s.}{\leq} 2^{2-\delta}(1-u)^{1-\delta}.
$$

Therefore it yields

$$\frac{(1-\tilde{H}_n^{-1}(u))^{2+0.5\epsilon}}{(1-u)^2} \leq 2^{(2-\delta)(2+0.5\epsilon)}(1-u)^{(1-\delta)(2+0.5\epsilon)-2} = 2^{(2-\delta)(2+0.5\epsilon)}(1-u)^{0.5\epsilon-\delta(2+0.5\epsilon)}$$

and, since $\epsilon > 0$, δ can be chosen in such a way that the exponent of $(1-u)$ is nonnegative for $u \in [0, (n-2)/n]$. For that reason

$$\sup_{0\leq u\leq (n-2)/n} \left(\frac{(1-\tilde{H}_n^{-1}(u))^{2+0.5\epsilon}}{(1-u)^2} \right)$$

is bounded almost surely. Moreover, by the SLLN,

$$\begin{aligned} \int_0^{(n-2)/n} \frac{\varphi(H^{-1}(\tilde{H}_n^{-1}(u)))}{(1-\tilde{H}_n^{-1}(u))^{1+\epsilon}} du &= \int_0^{(n-2)/n} \frac{\varphi(H_n^{-1}(u))}{(1-H(H_n^{-1}(u)))^{1+\epsilon}} du \\ &\leq \int_0^{\infty} \frac{\varphi(u)}{(1-H(u))^{1+\epsilon}} H_n(du) \\ &\xrightarrow[n\to\infty]{a.s.} \int_0^{\infty} \frac{\varphi(u)}{(1-H(u))^{1+\epsilon}} H(du) < \infty, \end{aligned} \tag{3.33}$$

where the last term is bounded due to assumption (M6). Hence $B(n) \xrightarrow[n\to\infty]{a.s.} 0$.

For $\varphi(Z_{n:n})$ we have

$$\varphi(Z_{n:n}) = \varphi(H^{-1}(U_{n:n})) \leq \frac{\log(n)^{1+\epsilon}}{n^\epsilon} \left(\frac{n}{\log(n)}(1-U_{n:n}) \right)^{1+\epsilon} \sup_{1\leq i\leq n} \frac{1}{n} \frac{\varphi(H^{-1}(U_i))}{(1-U_i)^{1+\epsilon}}.$$

Since assumption (M6) holds, Ghosh, Parr, Singh, and Babu (1984, Lemma 3) yields

$$\sup_{1\leq i\leq n} \frac{1}{n} \frac{\varphi(H^{-1}(U_i))}{(1-U_i)^{1+\epsilon}} \xrightarrow[n\to\infty]{a.s.} 0. \tag{3.34}$$

Furthermore from Robbins and Siegmund (1972, Theorem 1 (i))

$$1-U_{n:n} \overset{a.s.}{=} \mathcal{O}\left(\frac{\log(n)}{n} \right).$$

Thus

$$\left(\frac{n}{\log(n)}(1 - U_{n:n})\right)^{1+\epsilon}$$

is bounded almost surely. Because the first factor is also bounded, we have $\varphi(Z_{n:n}) \xrightarrow[n\to\infty]{a.s.} 0$.

In a similar way we can show the same for $\varphi(Z_{n-1:n})$:

$$\varphi(Z_{n-1:n}) \leq \frac{\log(n)^{1+\epsilon}}{n^{\epsilon}} \left(\frac{n}{\log(n)}(1 - U_{n-1:n})\right)^{1+\epsilon} \sup_{1\leq i\leq n} \frac{1}{n} \frac{\varphi(H^{-1}(U_i))}{(1-U_i)^{1+\epsilon}}.$$

Note that $(1 - U_{n-1:n}) = (1 - U_{n:n}) + (U_{n:n} - U_{n-1:n})$. Then again due to Robbins and Siegmund (1972, Theorem 1 (i)), $(1-U_{n:n}) = \mathcal{O}(\log(n)n^{-1})$ a.s. Furthermore, due to Shorack and Wellner (1986, pp. 720-721), $(1 - U_{n:n})$ and $(U_{n:n} - U_{n-1:n})$ are i.i.d. Therefore, from Robbins and Siegmund (1972, Remark 2.1) in conjunction with Robbins and Siegmund (1972, Theorem 1 (i)) we have

$$(U_{n:n} - U_{n-1:n}) \overset{a.s.}{=} \mathcal{O}\left(\frac{\log(n)}{n}\right).$$

Hence

$$\left(\frac{n(1 - U_{n-1:n})}{\log(n)}\right)^{1+\epsilon} = \left(\frac{n(1 - U_{n:n})}{\log(n)} + \frac{n(U_{n:n} - U_{n-1:n})}{\log(n)}\right)^{1+\epsilon}$$

is bounded almost surely. Then again using (3.34) yields $\varphi(Z_{n-1:n}) \xrightarrow[n\to\infty]{a.s.} 0$, which completes the proof. □

The overall concept of the next proof is equivalent to the last one. Since Theorem 3.13 incorporates convergence in probability we have to use slightly different arguments.

Proof of Theorem 3.13.

Again assume that φ is positive. Otherwise employ the decomposition in its positive and negative parts. Then, due to the same argumentation as in the proof of Theorem 3.16, we have

$$\tilde{T}_n \coloneqq n^{-1/2}\left(\int \varphi dF_{2,n}^{SE} - \int \varphi dF_{1,n}^{SE}\right) = n^{-1/2}\Big(I_1(n) + I_2(n)\Big)$$

where I_1 and I_2 are exactly the same as in (3.29). Then from (3.30) and (3.31) we have

$$\tilde{T}_n \leq 2n^{1/2}\Big(B(n) + \varphi(Z_{n-1:n}) + \varphi(Z_{n:n})\Big)$$

where, equivalently to previous proof, $B(n) \coloneqq \sum_{i=1}^{n-2} \frac{\varphi(Z_{i:n})}{(n-i)^2}$. Hence it is left to show that $n^{1/2}B(n)$, $n^{1/2}\varphi(Z_{n-1:n})$ an $n^{1/2}\varphi(Z_{n:n})$ converge to zero in probability.

Inequality (3.32) yields

$$\begin{aligned}
n^{1/2}B(n) \leq& \frac{4}{n^{1/2}} \int_0^{(n-2)/n} \frac{\varphi(H^{-1}(\tilde{H}_n^{-1}(u)))}{(1-u)^2} du \\
=& \frac{4}{n^{1/2}} \int_0^{(n-2)/n} \frac{\varphi(H^{-1}(\tilde{H}_n^{-1}(u)))}{(1-\tilde{H}_n^{-1}(u))^{1.5+\epsilon}} \left(\frac{1-\tilde{H}_n^{-1}(u)}{1-u}\right)^{1.5+\epsilon} \frac{(1-u)^{1.5+\epsilon}}{(1-u)^2} du \\
\leq& \frac{4}{n^{1/2}} \sup_{0\leq u \leq (n-2)/n} \left((1-u)^{-0.5+\epsilon}\right) \sup_{0\leq u\leq (n-2)/n} \left(\frac{1-\tilde{H}_n^{-1}(u)}{1-u}\right)^{1.5+\epsilon} \\
&\times \int_0^{(n-2)/n} \frac{\varphi(H^{-1}(\tilde{H}_n^{-1}(u)))}{(1-\tilde{H}_n^{-1}(u))^{1.5+\epsilon}} du \\
\leq& \frac{4}{2^{1/2-\epsilon} n^{\epsilon}} \sup_{0\leq u\leq (n-2)/n} \left(\frac{1-\tilde{H}_n^{-1}(u)}{1-u}\right)^{1.5+\epsilon} \int_0^{(n-2)/n} \frac{\varphi(H^{-1}(\tilde{H}_n^{-1}(u)))}{(1-\tilde{H}_n^{-1}(u))^{1.5+\epsilon}} du,
\end{aligned}$$

since $(1-u)^{-0.5+\epsilon}$ attains its maximum on $[0,(n-2)/n]$ at the upper bound of the interval. Due to Shorack and Wellner (1986, p. 419, Inequality 1)

$$\sup_{0\leq u\leq (n-2)/n} \left(\frac{1-\tilde{H}_n^{-1}(u)}{1-u}\right)^{1.5+\epsilon}$$

is bounded in probability. Furthermore, similar to (3.33), we have by the ordinary SLLN,

$$\int_0^{(n-2)/n} \frac{\varphi(H^{-1}(\tilde{H}_n^{-1}(u)))}{(1-\tilde{H}_n^{-1}(u))^{1.5+\epsilon}} du \leq \int_0^{\infty} \frac{\varphi(u)}{(1-H(u))^{1.5+\epsilon}} H_n(du)$$
$$\xrightarrow[n\to\infty]{a.s.} \int_0^{\infty} \frac{\varphi(u)}{(1-H(u))^{1.5+\epsilon}} H(du) < \infty.$$

The last term is finite due to assumption (M5). Thus

$$n^{1/2} B(n) \xrightarrow[n\to\infty]{} 0 \qquad \text{in probability.}$$

Moreover it yields

$$\begin{aligned} n^{1/2}\varphi(Z_{n:n}) =& n^{1/2} \frac{\varphi(H^{-1}(U_{n:n}))}{(1-U_{n:n})^{1.5+\epsilon}} (1-U_{n:n})^{1.5+\epsilon} \\ \leq& (1-U_{n:n})^{\epsilon} (n(1-U_{n:n}))^{1.5} \sup_{1\leq i\leq n} \frac{\varphi(H^{-1}(U_i))}{n(1-U_i)^{1.5+\epsilon}}. \end{aligned}$$

Similarly as in the previous proof,

$$\sup_{1\leq i\leq n} \frac{\varphi(H^{-1}(U_i))}{n(1-U_i)^{1.5+\epsilon}} \xrightarrow[n\to\infty]{a.s.} 0$$

due to Ghosh et al. (1984, Lemma 3) because we assume (M5). Note that both, $n(1-U_{n:n})$ and $n(1-U_{n-1:n})$, are bounded in probability. Hence

$$n^{1/2}\varphi(Z_{n:n}) \xrightarrow[n\to\infty]{} 0 \qquad \text{in probability.}$$

Using the same arguments, the equivalent holds true for $n^{1/2}\varphi(Z_{n-1:n})$. □

Proof of Theorem 3.17.

The proof of Theorem 3.17 is a simplified version of the one for Theorem 3.16. To shorten the notation in the following calculation we define $m_i = m(Z_i, \theta_n)$. At first note that

$$\left|F_{2,n}^{SE}(t) - F_{1,n}^{SE}(t)\right| = \left|\prod_{i:Z_i \le t}\left(1 - \frac{m(Z_{i:n}, \theta_n)}{n-i+1}\right) - \prod_{i:Z_i \le t}\left(1 - \frac{m(Z_{i:n}, \theta_n)}{n-i+m(Z_{i:n}, \theta_n)}\right)\right|.$$

Applying Remark 3.22 to rewrite the difference and using 1 as an upper bound of the contained products yields

$$\begin{aligned}
\left|F_{2,n}^{SE}(t) - F_{1,n}^{SE}(t)\right| &\le \sum_{i:Z_i \le t}\left(\frac{m(Z_{i:n}, \theta_n)}{n - R_n(Z_i) + m(Z_{i:n}, \theta_n)} - \frac{m(Z_{i:n}, \theta_n)}{n - R_n(Z_i) + 1}\right) \\
&= \sum_{i:Z_i \le t} \frac{m(Z_{i:n}, \theta_n) - m^2(Z_{i:n}, \theta_n)}{(n - R_n(Z_i) + m(Z_{i:n}, \theta_n))(n - R_n(Z_i) + 1)} \\
&\le \sum_{i:Z_i \le t} \frac{1}{(n - R_n(Z_i))^2} = \frac{1}{n^2} \sum_{i:Z_i \le t} \frac{1}{(1 - R_n(Z_i)/n)^2} = \frac{1}{n} \int_0^t \frac{1}{(\bar{H}_n(x))^2} H_n(dx),
\end{aligned}$$

where we used $R_n(Z_i) = nH_n(Z_i)$. Since $\bar{H}_n(x) \ge \bar{H}_n(t) \ge \bar{H}_n(T)$

$$\le \frac{1}{n} \frac{1}{(\bar{H}_n(T))^2} \int_0^x H_n(dx) \le \frac{1}{n} \frac{1}{(\bar{H}_n(T))^2}.$$

The result follows by the SLLN and $H(T) < 1$

$$(\bar{H}_n(T))^{-2} \xrightarrow[n\to\infty]{a.s.} (\bar{H}(T))^{-2} < \infty,$$

and holds uniformly in $t \in [0, T]$. $\square$

Proof of Lemma 3.11.

We again use the simplified notation $m_i = m(Z_i, \theta_n)$. First consider the basic inequalities

$$-\frac{a}{1-a} \leq \ln(1-a) \leq -a \quad \text{and} \quad 1 + x \leq \exp(x) \leq \frac{1}{1-x}$$

for all $0 \leq a < 1$ and $1 > x \in \mathbb{R}$. Under the given conditions we have for $i = 1, \ldots, n$ on the one hand

$$\begin{aligned}\left(\frac{n - R_n(Z_i)}{n - R_n(Z_i) + 1}\right)^{m_i} &= \left(1 - \frac{1}{n - R_n(Z_i) + 1}\right)^{m_i} = \exp\left(m_i \ln\left(1 - \frac{1}{n - R_n(Z_i) + 1}\right)\right) \\ &\leq \exp\left(-\frac{m_i}{n - R_n(Z_i) + 1}\right) \leq \frac{n - R_n(Z_i) + 1}{n - R_n(Z_i) + 1 + m_i}\end{aligned} \tag{3.35}$$

and on the other hand

$$\left(\frac{n - R_n(Z_i)}{n - R_n(Z_i) + 1}\right)^{m_i} \geq \exp\left(-\frac{m_i}{n - R_n(Z_i)}\right) \geq 1 - \frac{m_i}{n - R_n(Z_i)} = \frac{n - R_n(Z_i) - m_i}{n - R_n(Z_i)}. \tag{3.36}$$

Applying Remark 3.22 to rewrite the difference $\dot{F}^{SE}_{1,n} - F^{SE}_{1,n}$ while using 1 as the upper bound of the occurring products gives

$$\begin{aligned}\left|\dot{F}^{SE}_{1,n}(t) - F^{SE}_{1,n}(t)\right| &= \left|\prod_{i:Z_i \leq t}\left(1 - \frac{m_i}{n - R_n(Z_i) + 1}\right) - \prod_{i:Z_i \leq t}\left(\frac{n - R_n(Z_i)}{n - R_n(Z_i) + 1}\right)^{m_i}\right| \\ &\leq \sum_{i:Z_i \leq t}\left|\left(1 - \frac{m_i}{n - R_n(Z_i) + 1}\right) - \left(\frac{n - R_n(Z_i)}{n - R_n(Z_i) + 1}\right)^{m_i}\right| \equiv \sum_{i:Z_i \leq t}\left|A_i - B_i\right|.\end{aligned}$$

Note that $|A_i - B_i| = \max(A_i - B_i, B_i - A_i)$. Then inequality (3.36) yields

$$\begin{aligned}A_i - B_i &\leq \left(1 - \frac{m_i}{n - R_n(Z_i) + 1}\right) - \left(\frac{n - R_n(Z_i) - m_i}{n - R_n(Z_i)}\right) \\ &= \frac{m_i}{(n - R_n(Z_i))(n - R_n(Z_i) + 1)} \leq \frac{1}{(n - R_n(Z_i))^2}.\end{aligned}$$

Furthermore by inequality (3.35) we have

$$\begin{aligned}
B_i - A_i &\leq \frac{n - R_n(Z_i) + 1}{n - R_n(Z_i) + 1 + m_i} - \left(1 - \frac{m_i}{n - R_n(Z_i) + 1}\right) \\
&= \frac{m_i^2}{(n - R_n(Z_i) + 1 + m_i)(n - R_n(Z_i) + 1)} \leq \frac{1}{(n - R_n(Z_i))^2}.
\end{aligned}$$

Here we used $0 \leq m(\cdot, \cdot) \leq 1$ for n large enough. Combining the last two results leads to

$$\left|\dot{F}_{1,n}^{SE}(t) - F_{1,n}^{SE}(t)\right| \leq \sum_{i:Z_i \leq t} |A_i - B_i| \leq \sum_{i:Z_i \leq t} \frac{1}{(n - R_n(Z_i))^2} \leq \frac{1}{n(\bar{H}_n(T))^2},$$

where the last inequality is derived identically as described in the proof of Theorem 3.17. Since $H(T) < 1$, the assertion follows by the SLLN, $(\bar{H}_n(T))^{-2} \xrightarrow{a.s.} (\bar{H}(T))^{-2} < \infty$ as $n \to \infty$, and holds uniformly in $t \in [0, T]$. □

Chapter 4

Kernel type density estimators for right censored data

We are going to define kernel density estimators applicable under the random censorship model by replacing the e.c.d.f. in the definition of the usual kernel density estimator with the product limit estimators derived in Chapter 3. In Definition 4.3 and Definition 4.4 we propose the new semi-parametric and presmoothed estimators $f_{2,n}^{SE}$ and $f_{2,n}^{PR}$. The main objective is the derivation of the asymptotic representations for $f_{2,n}^{SE}$ in Theorem 4.6 and Theorem 4.7. Relying on those we determine exact rates of pointwise and uniform convergence and deduce the pointwise limiting distribution as well as the distribution of the maximal deviation.

4.1 Kernel density estimation for complete data

As an entry point to density estimation, first consider the case of observing only uncensored data points, e.g. all estimations can rely on a sample $(x_i)_{1\leq i\leq n}$ of X. In many applications one has no information about the existence or structure of a parametric family possibly underlying X. Therefore nonparametric methods have to be applied.

One of the most popular and extensively studied nonparametric estimators is the kernel density estimator introduced by Rosenblatt (1956) and Parzen (1962). A general version of this estimator is given in the following definition.

Definition 4.1. Assume X is a random variable as given in Definition 2.1 and let F_n^* be some consistent estimator of the d.f. F. Then

$$f_n^*(t) := \frac{1}{a_n} \int_{\mathbb{R}} K\left(\frac{t-x}{a_n}\right) F_n^*(dx) \tag{4.1}$$

defines an estimator of the density function f, provided that the kernel K and the sequence of bandwidths $(a_n)_{n\geq 1}$ satisfy certain conditions, which will be concretized during the following discussion.

In the absence of censoring, the e.c.d.f. F_n, based on a sample $(x_i)_{1\leq i\leq n}$, is a natural choice for F_n^*. In particular

$$f_n(t) := \frac{1}{a_n} \int_{\mathbb{R}} K\left(\frac{t-x}{a_n}\right) F_n(dx) \tag{4.2}$$

is the estimator established by Rosenblatt (1956) and Parzen (1962). There are a lot of results available for complete data including Silverman (1986), Härdle (1991) and Wand and Jones (1994) or, more recently for the multivariate case, Scott (2015). Some elementary characteristics of f_n could readily be seen from the definition. If the conditions

(K1) $\int_{-\infty}^{\infty} K(x)dx = 1$ and

(K2) $K(x) \geq 0 \ \forall x \in \mathbb{R}$

hold, e.g. K is a probability density function, then f_n itself is a probability density function. Furthermore, f_n inherits the continuity properties of K. It is shown in Parzen (1962, Corollary 1A) that the estimator $f_n(t)$ is asymptotically unbiased at all continuity points of f given the assumptions

(H1) $\lim_{n\to\infty} a_n = 0$,

(K3) $\sup_{-\infty<x<\infty} |K(x)| < \infty$,

(K4) $\int_{-\infty}^{\infty} |K(x)|dx < \infty$ and

(K5) $\lim_{x\to\infty} |xK(x)| = 0$.

Often an even function satisfying (K1) to (K5) is referred to as *weighting function* or *probability kernel*. Possible kernel functions are for example given in Silverman (1986, p. 43). The Epanechnikov (1969) kernel, $K_{ep}(t) \coloneqq 3/4\,(1-t^2)\,\mathbb{1}_{|t|<1}$, is considered to be optimal among the kernels of second order since K_{ep} minimizes the mean integrated squared error, cf. Wand and Jones (1994, Chapter 2.7).

If we put more restrictive assumptions on the bandwidth a_n which guarantee that the variance tends to zero as $n \to 0$, i.e.

(H2) $\lim\limits_{n\to\infty} na_n = \infty$,

we can ensure weak convergence of $f_n(t)$ to $f(t)$ at all continuity points of f, cf. Parzen (1962, p. 1069). When extending (H2) to $\lim\limits_{n\to\infty} na_n^2 = \infty$ the last result also holds uniformly.

Nadaraya (1965) was the first one who proved a strong consistency result. Silverman (1978) managed to weaken his assumptions. Given the assumptions (K3),

(H3) $\lim\limits_{n\to\infty} na_n \cdot (\log n)^{-1} = \infty$,

(K6) K is of bounded variation, denoted by V_K

(K7) The set of all discontinuities of K has Lebesgue measure zero,

and assuming f is uniformly continuous on $(-\infty, \infty)$, Bertrand-Retali (1978) showed that $f_n(t)$ converges to f a.s. uniformly. Furthermore, given (K1), (K3) to (K5), (H1), (H2) and

(K8) $\int K(x)^{2+\delta} dx < \infty \qquad \forall \delta > 0$,

Parzen (1962) proved that

$$n^{1/2}\left(f_n - \bar{f}_n\right) \xrightarrow[n\to\infty]{} \mathcal{N}(0, \sigma^2) \qquad \text{in distribution,} \tag{4.3}$$

where the asymptotic variance is given by $\sigma^2 = f(x) \int_{-\infty}^{\infty} K^2(x)dx$ and

$$\bar{f}_n(t) \coloneqq \mathbb{E}[f_n] = \frac{1}{a_n} \int_{\mathbb{R}} K\left(\frac{t-x}{a_n}\right) F(dt), \tag{4.4}$$

is the expectation of the kernel density estimator f_n.

4.2 Density estimators for right censored data

In Section 4.1 we have seen that kernel density estimators as given in Definition 4.1 mainly depend on a consistent estimator of F and some generic kernel function K:

$$f_n^*(t) \coloneqq \frac{1}{a_n} \int_{\mathbb{R}} K\left(\frac{t-x}{a_n}\right) F_n^*(dx).$$

In the case of complete data, $F_n^* = F_n$, the e.c.d.f. based on $(x_i)_{1 \le i \le n}$, is used. The resulting estimator f_n is very well-studied.

In Chapter 3 we presented a new technique to derive estimators which extend the usual e.c.d.f. F_n to the RCM. Those can be used to define density estimators applicable in the case of censoring. Possible candidates for F_n^* are F_n^{KM}, F_n^{ACL}, $F_{1,n}^{SE}$ and $F_{1,n}^{PR}$ from remarks 3.3 to 3.6 but also the new estimators $F_{2,n}^{SE}$ and $F_{2,n}^{PR}$ introduced in Definition 3.7 and Definition 3.8, respectively.

In fact, the estimator

$$\begin{aligned} f_n^{KM}(t) \coloneqq& \frac{1}{a_n} \int_{\mathbb{R}} K\left(\frac{t-x}{a_n}\right) F_n^{KM}(dt) \\ =& \frac{1}{a_n} \sum_{i=1}^{n} K\left(\frac{t-Z_{i:n}}{a_n}\right) W_{i,n}^{KM} \end{aligned}$$

has been proposed by Blum and Susarla (1980) and received great attention in practical applications. McNichols and Padgett (1981) introduced the above given representation in terms of the order statistics of the Z-sample and proved that under the assumptions (K1), (K3) to (K5) and (H1) the estimator $f_n^{KM}(t)$ is an asymptotically unbiased estimator of $f(t)$ for all $t \ge 0$. If, in addition to (H2), some further weak assumptions on K hold, then McNichols and Padgett (1981) also showed that $f_n^{KM} \to f$ in mean square as $n \to \infty$.

In Földes, Rejtő, and Winter (1981, Theorem 3.2) it is shown that under the following hypotheses f_n^{KM} is strongly consistent: Let f be bounded, $G(T_F^-) < 1$ with $T_F = \sup\{x|F(x) < 1\}$ and assume in addition to (K6), (H1)

(K9) K is right-continuous,

(H4) $\lim_{n\to\infty} a_n (n/\log n)^{1/8} = \infty$,

then it holds

$$f_n^{KM}(t) \xrightarrow[n\to\infty]{a.s.} f(t),$$

at all continuity points t of f. Földes et al. (1981) also gave conditions for a.s. uniform convergence. Zhang (1998) used strong approximation techniques and counting processes to study strong uniform convergence. His proof uses a similar approach to Silverman (1978) for the uncensored case. A version of f_n^{KM}, where the bandwidths a_n depend on the censored sample $(Z_i, \delta_i)_{1\le i\le n}$, is considered in McNichols and Padgett (1984). The consistency of f_n^{KM} for some other choices of bandwidths based on the distance to the $k(n)$th nearest uncensored observation are investigated in Mielniczuk (1986).

Ramlau-Hansen (1983) and Mielniczuk (1986) were concerned with the asymptotic distribution of f_n^{KM}, but for somewhat suboptimal bandwidths. Also Blum and Susarla (1980) already derived the limit distribution of an estimator similar to f_n^{KM}. More generally, Diehl and Stute (1988) proved that given (K1) to (K5), (H1), (H2) and

(K10) K is continuously differentiable,

(K11) K vanishes outside some finite interval $-\infty < r < 0 < s < \infty$,

and f, g are bounded on $[0, T']$ for some $T < T'$ then for almost all $0 \le t \le T < T' < \tau_H$

$$(na_n)^{1/2}\Big(f_n^{KM}(t) - \bar{f}_n(t)\Big) \xrightarrow[n\to\infty]{} \mathcal{N}(0, \sigma_{KM}^2) \qquad \text{in distribution,}$$

where

$$\sigma_{KM}^2 = \frac{f(t)}{1 - G(t)} \int_{\mathbb{R}} K^2(x)dx, \tag{4.5}$$

and $\bar{f}_n = \mathbb{E}(f_n)$ as given in (4.4).

Note that $\bar{f}_n(t)$ is not the expectation of f_n^{KM} in the presence of censoring. In addition, Zhang (1996) proved several asymptotic results of f_n^{KM}, including asymptotic normality, by using the theory of martingales for counting processes. Optimal bandwidth selection for f_n^{KM} is discussed in Marron and Padgett (1987), among others. Giné and Guillou (2001) extended the results of Diehl and Stute (1988) to hold for adaptive intervals. The approach they used is similar to Einmahl and Mason (2000) for the uncensored case. L_p convergence of f_n^{KM} was considered in Csörgő, Gombay, and Horvath (1991), Ghorai and Pattanaik (1990) and Carbonez and Györfi (1992). The asymptotic normality of the weighted integrated squared error of f_n^{KM} was proven in Ghorai and Pattanaik (1991) by using a version of the martingale central limit theorem. An error bound for the mean integrated absolute error is given in Kulasekera (1995) and an exact asymptotic expression of the L_1-error is determined in Lemdani and Ould-Saïd (2002). Dinwoodie (1993) studied some large deviation properties of F_n^{KM}. More recently Diallo and Louani (2013) stated moderate and large deviation principles for kernel type estimators of the hazard rate in presence of censoring.

If we make, in addition to the RCM, further smoothness assumptions, for example m, f, and h are four times continuously differentiable then

$$f_{1,n}^{PR}(t) := \frac{1}{a_n} \int_{\mathbb{R}} K\left(\frac{t-x}{a_n}\right) F_{1,n}^{PR}(dx)$$

defines an estimator of the p.d.f. f. Among other results, Cao and Jácome (2004) and Jácome and Cao (2007) were concerned with the asymptotic normality of $f_{1,n}^{PR}$. Jácome and Cao (2007) also proved pointwise strong consistency. A comparison of different presmoothing methods is given in Jácome, Gijbels, and Cao (2008). Bandwidth selection is very crucial for those estimators and is discussed in Jácome and Cao (2008).

Proceeding using the same idea of plugging in PLEs of F into Definition 4.1, it is natural to define the following estimators.

Definition 4.2. Given the SRCM, then under certain assumptions on K and $(a_n)_{n\geq 1}$,

$$f_{1,n}^{SE}(t) := \frac{1}{a_n}\int_{\mathbb{R}} K\left(\frac{t-x}{a_n}\right) F_{1,n}^{SE}(dx)$$

defines an estimator of the p.d.f. f where $F_{1,n}^{SE}$ is given in Remark 3.4.

To our knowledge, this estimator has only been considered in simulation studies presented in Jácome et al. (2008) and more elaborately in Harlaß (2011). Further on, making use of the newly defined approximations $F_{2,n}^{SE}$ and $F_{2,n}^{PR}$, we propose the following estimators.

Definition 4.3. Given the SRCM, let K be some probability kernel and $(a_n)_{n\geq 1}$ a series of bandwidths, then

$$f_{2,n}^{SE}(t) := \frac{1}{a_n}\int_{\mathbb{R}} K\left(\frac{t-x}{a_n}\right) F_{2,n}^{SE}(dx)$$

is an estimator of the density function f where $F_{2,n}^{SE}$ is defined in Definition 3.7.

Definition 4.4. Requiring m, f and h to fulfill some smoothness conditions, e.g. see Cao and Jácome (2004) and Jácome and Cao (2007), then

$$f_{2,n}^{PR}(t) := \frac{1}{a_n}\int_{\mathbb{R}} K\left(\frac{t-x}{a_n}\right) F_{2,n}^{PR}(dx)$$

defines an estimator of the density function f, where $F_{2,n}^{PR}$ is given in Definition 3.8, K some probability kernel and $(a_n)_{n\geq 1}$ a series of bandwidths.

In an equivalent way, it is possible to specify kernel based estimators of the hazard rate as the scaled convolution of some probability kernel and the estimators $\Lambda_{1,n}^{SE}$ or $\Lambda_{2,n}^{SE}$, respectively:

$$\lambda_{1,n}^{SE}(t) := \frac{1}{a_n}\sum_{i=1}^{n} K\left(\frac{t-Z_i}{a_n}\right)\frac{m(Z_i,\theta_n)}{n-R_n(Z_i)+1},$$

and

$$\lambda_{2,n}^{SE}(t) := \frac{1}{a_n}\sum_{i=1}^{n} K\left(\frac{t-Z_i}{a_n}\right)\frac{m(Z_i,\theta_n)}{n-R_n(Z_i)+m(Z_i,\theta_n)}.$$

Because of this construction, those estimators behave quite similar to their counterparts for the p.d.f.

Due to the improved properties of the semi-parametric PLEs versus the Kaplan-Meier estimator, in particular the gain in efficiency in terms of the asymptotic variance under the SRCM, see Remark 3.21, it is conceivable that those improvements are carried over to the semi-parametric kernel estimators. The simulation studies in Harlaß (2011) support this conjecture. The simulations indicate a reduction of the asymptotic variance and the mean squared error when comparing $f_{1,n}^{SE}$ to f_n^{KM}. Theorem 4.10 below shows that this hypotheses is indeed true. In particular we show that $f_{1,n}^{SE}$ and $f_{2,n}^{SE}$ are equivalent as $n \to \infty$ and therefore the result holds for both estimators.

Recalling Theorem 3.13 and Theorem 3.17, it is not very surprising that the semi-parametric estimators $f_{1,n}^{SE}$ and $f_{2,n}^{SE}$, given in Definition 4.2 and Definition 4.3, are asymptotically identical. An exact statement is given in the following theorem.

Theorem 4.5. Requiring (K10), (K11) and given the assumptions of Theorem 3.17, that is F and G are continuous (A2), (A10), and (M6) are satisfied, then for $T < T' < \tau_H$

$$(na_n)^{1/2}\sup_{0\le t\le T}\left|f_{2,n}^{SE}(t) - f_{1,n}^{SE}(t)\right| \stackrel{a.s.}{=} \mathcal{O}((na_n)^{-1/2}).$$

The following two theorems represent our major results related to semi-parametric kernel density estimators. The asymptotic representation of $f_{1,n}^{SE}$ heavily depends on h_n^1, the kernel density estimator of h^1, which was defined in (2.4),

$$h_n^1(t) := \frac{1}{a_n} \int_{\mathbb{R}} K\left(\frac{t-x}{a_n}\right) H_n^1(dx).$$

H_n^1 is given in (3.5). Note that this estimator relies on the complete dataset $(Z_i)_{1 \le i \le n}$.

Theorem 4.6. Let Θ be a connected, open subset of $\mathbb{R}^k$ and let K be some probability kernel, that is (K1) to (K5), satisfying (K10) and (K11). Furthermore, assume (H1), (H2) and that f and g are bounded on $[0, T']$ for some $T < T'$. Given the SRCM and that H is continuous, then under the assumptions (A1) to (A7), (A10) it holds for almost all $0 \le t \le T < T' < \tau_H$ that

$$\sup_{0 \le t \le T} (na_n)^{1/2} \left| f_{1,n}^{SE}(t) - \bar{f}_n(t) - \frac{h_n^1(t) - \mathbb{E} h_n^1(t)}{1 - G(t)} \right| = \mathcal{O}((na_n)^{-1/2}) + \mathcal{O}(a_n^{1/2}) \qquad \text{in probability.}$$

Theorem 4.7. Under the assumptions of Theorem 4.6 it holds that

$$\sup_{0 \le t \le T} (na_n)^{1/2} \left| f_{1,n}^{SE}(t) - \bar{f}_n(t) - \frac{h_n^1(t) - \mathbb{E} h_n^1(t)}{1 - G(t)} \right| = \mathcal{O}\left(\frac{\ln \ln n}{(na_n)^{1/2}}\right) + \mathcal{O}\left((a_n \ln \ln n)^{1/2}\right) \qquad \text{a.s.}$$

Corollary 4.8. As a direct consequence of Theorem 4.5, Theorem 4.6 and Theorem 4.7 hold true for $f_{2,n}^{SE}$ under the same assumptions.

Theorems 4.6 and 4.7 are the counterparts to Diehl and Stute (1988, Theorem 1) for the semi-parametric case. To prevent possible misapprehension, Diehl and Stute (1988) defined the kernel density estimator, appearing in their theorem, using $\hat{H}_n^1(x) := n^{-1} \sum_{i=1}^n \delta_i \mathbb{1}_{\{Z_i \le x\}}$. Here we rely on h_n^1 as given above, in particular $H_n^1(x) = n^{-1} \sum_{i=1}^n m(Z_i, \theta_n) \mathbb{1}_{\{Z_i \le x\}}$; cf. (3.5).

Similar asymptotic representations can be derived for $\lambda_{1,n}^{SE}$ and $\lambda_{2,n}^{SE}$, the kernel based estimators of the hazard rate which rely on $\Lambda_{1,n}^{SE}$ and $\Lambda_{2,n}^{SE}$, respectively. We omit the proof since it is analogous to the one given in Section 4.3.

Remark 4.9. If G and $\bar{f}_n$ are replaced by H and

$$\bar{\lambda}_n(t) = \frac{1}{a_n} \int_{\mathbb{R}} K\left(\frac{t-x}{a_n}\right) \Lambda(dx),$$

respectively, then Theorem 4.6 and Theorem 4.7 hold for $\lambda_{1,n}^{SE}$ as well as for $\lambda_{2,n}^{SE}$ under the same assumptions.

In the following we derive the asymptotic distribution of the $f_{1,n}^{SE} - \bar{f}_n$ and the limit distribution of the maximal deviation. Moreover, we give exact rates of pointwise and uniform strong convergence.

Theorem 4.10. Under the assumptions of Theorem 4.6 it holds that

$$(na_n)^{1/2} \left| f_{1,n}^{SE}(t) - \bar{f}_n(t) \right| \xrightarrow[n\to\infty]{} \mathcal{N}(0, \sigma_{SE}^2) \qquad \text{in distribution},$$

with the asymptotic variance

$$\sigma_{SE}^2 = \frac{m(t,\theta_0) f(t)}{1 - G(t)} \int_{\mathbb{R}} K^2(u)\, du.$$

Due to Theorem 4.5, the same holds true for $f_{2,n}^{SE}$.

Corollary 4.11. Under the same assumptions as Theorem 4.6 it holds that

$$(na_n)^{1/2} \sup_{0 \le t \le T} \left| f_{2,n}^{SE}(t) - \bar{f}_n(t) \right| \xrightarrow[n\to\infty]{} \mathcal{N}(0, \sigma_{SE}^2) \qquad \text{in distribution}$$

where σ_{SE}^2 as given in Theorem 4.10.

Proof. The triangle inequality yields

$$(na_n)^{1/2} \left| f_{2,n}^{SE}(t) - \bar{f}_n(t) \right| \leq (na_n)^{1/2} \left| f_{2,n}^{SE}(t) - f_{1,n}^{SE}(t) \right| + (na_n)^{1/2} \left| f_{1,n}^{SE}(t) - \bar{f}_n(t) \right| = \mathrm{I} + \mathrm{II}.$$

Then $\mathrm{I} \to 0$ as $n \to \infty$ almost surely due to Theorem 4.5. The limit distribution follows from Theorem 4.10. □

Remark 4.12. When comparing the asymptotic variance of $f_{2,n}^{SE}$ with the one of the Kaplan-Meier counterpart, we have

$$\sigma_{KM}^2 - \sigma_{SE}^2 = (1 - m(t, \theta_0)) \frac{f(t)}{\bar{G}(t)} \int_{\mathbb{R}} K^2(u)\, du \geq 0,$$

for all $t \in \mathbb{R}_{\geq}$ given the SRCM. Since $0 \leq m(t) \leq 1$, the semi-parametric estimator is more efficient at t, wherever $m(t) < 1$. Both estimators perform equally at all t, if and only if $m(t) = 1$ for all t. But this is only the case if there is no censoring at all. In case of a continuous model m, equality only occurs with probability 0. Moreover, recall that $m \equiv 1$ implies that there is no censoring. In this case both estimators reduce to the e.c.d.f.

Corollary 4.13. Assume that the conditions of Theorem 4.6 are satisfied. Set $a_n = n^{-\delta}$ for some $0 < \delta < 1/2$. Given that h^1 and K satisfy the assumptions of Bickel and Rosenblatt (1973, Theorem 3.1), then

$$\mathbb{P}\left((2\delta \ln n)^{1/2} \left[\sup_{T'' \leq t \leq T} (na_n)^{1/2} \sqrt{\frac{1 - G(t)}{m(t, \theta_n) f(t) \int K^2}} \left| f_{1,n}^{SE}(t) - \bar{f}_n(t) \right| - d_n \right] < x \right) \to e^{-2e^{-x}},$$

for $d_n \to \infty$, depending on K and δ, as $n \to \infty$.

Corollary 4.14. Assume (H1), (H2) and additionally

$$\text{(H5)} \quad \lim_{\epsilon\to 0}\lim_{n\to\infty} \sup_{k:|k-n|\le n\epsilon} \left|\frac{a_k}{a_n}-1\right| = 0$$

and

$$\text{(H6)} \quad \lim_{n\to\infty} \frac{(\ln n)^4}{n a_n \ln\ln n} = 0.$$

Then, under the assumptions of Theorem 4.7,

$$\limsup_{n\to\infty} \sqrt{\frac{n a_n}{2\ln\ln n}} \left(f_{1,n}^{SE}(t) - \bar{f}_n(t)\right) \overset{a.s.}{=} \left(\frac{m(t,\theta_0)f(t)}{1-G(t)}\int K^2(u)du\right)^{1/2}.$$

Proof. This result is a direct consequence of Theorem 4.7 and the application of Hall (1981, Theorem 2). □

Corollary 4.15. Under the assumptions of Theorem 4.7, extend (H1) and (H2) by choosing $(a_n)_{n>1}$ such that the following holds:

$$\text{(H7)} \quad \lim_{n\to\infty} \frac{\ln a_n^{-1}}{n a_n} = 0,$$

and

$$\text{(H8)} \quad \lim_{n\to\infty} \frac{\ln a_n^{-1}}{\ln\ln n} = \infty.$$

Furthermore, assume $0 < k \le f(t)$ for $t \in [T'', T']$ with $0 \le T'' < T < T'$. Then

$$\lim_{n\to\infty} \sqrt{\frac{n a_n}{2\ln a_n^{-1}}} \sup_{T''\le t\le T'} \sqrt{\frac{1-G(t)}{m(t,\theta_n)f(t)}} \left|f_{1,n}^{SE}(t) - \bar{f}_n(t)\right| \overset{a.s.}{=} \left(\int K^2(u)du\right).$$

Proof. The convergence follows directly from Stute (1982, Theorem 1.3) due to Theorem 4.7. □

In practice assumptions (H5) and (H6) are no restriction since usually the bandwidth a_n is taken to behave like Cn^{-b} with $C > 0$ and $0 < b < 1$ for which both assumptions are satisfied. Prerequisites (H7) and (H8) roughly state that the bandwidth ranges between $1/n$ and $1/\ln n$). When choosing for example $a_n = Cn^{-b}$ (H7) and (H8) are satisfied.

The rule-of-thumb given in Silverman (1986, Equation 3.31)

$$a_n = Cn^{-1/5} \tag{4.6}$$

with $C = \min(\text{standard deviation}, \text{inter quartile range}/1.34)$ is a popular choice for a fixed bandwidth and satisfies all of the previously mentioned conditions. Scott (1992) suggests to choose the factor 1.06 instead. For more comprehensive results on bandwidth selection see for example Heidenreich, Schindler, and Sperlich (2013) and Chiu (1996).

Corollary 4.16. Due to Theorem 4.5, the corollaries 4.13, 4.14 and 4.15 also hold for $f_{1,n}^{SE}$ replaced by $f_{2,n}^{SE}$.

4.3 Proving the properties of $f_{1,n}^{SE}$ and $f_{2,n}^{SE}$

In this section we are primarily concerned with the proof of the Theorems 4.6, 4.7 and 4.10. Other results are proved right away in Section 4.1. Note that by assumption (K11), $K((t-x)/a_n)) = 0$ for all $x \notin (t - sa_n, t - ra_n)$. Hence it is sufficient to integrate over $S_n \coloneqq S_n(t) \coloneqq [t - sa_n, t - ra_n]$ instead of the whole real line.

The following two corollaries will be used repeatedly in the course of this section. They can be derived by applying a general variant of integration by parts. Hewitt (1960, p. 423) provides a formula for integration by parts w.r.t. signed measures which fits our needs. The proof of more elementary versions could be based on Hewitt and Stromberg (1965, Theorem 21.67).

Corollary 4.17. Let $\mathcal{H}$ be some arbitrary continuous d.f. and assume (K10) and (K11), then

$$\int_{\mathbb{R}} K\left(\frac{t-x}{a_n}\right)\mathcal{H}(dx) = -\int_{\mathbb{R}} \mathcal{H}(x)K\left(\frac{t-dx}{a_n}\right).$$

Proof. By (K10) and (K11), $K((t-x)/a_n))$ is an absolute continuous function in x which evaluates to zero at the boundaries of S_n. Then Hewitt (1960, p. 423) yields

$$\int_{S_n} K\left(\frac{t-x}{a_n}\right)\mathcal{H}(dx) = -\int_{S_n} \mathcal{H}(x)K\left(\frac{t-dx}{a_n}\right). \qquad \square$$

Corollary 4.18. Let $\psi : \mathbb{R} \mapsto \mathbb{R}$ be some bounded Borel-measurable function and again use $S_n = [t - sa_n, t - ra_n]$. Moreover, assume (K10) and (K11). Then

$$\int_{S_n} \psi(x)K\left(\frac{t-dx}{a_n}\right) = -\int_{[r,s]} \psi(t - ua_n)K'(u)d(u) = -\int_{[r,s]} \psi(t - ua_n)K(du).$$

Proof. Let K' denote the derivative of K and λ the Lebesgue measure. By assumption (K10), $K((t-x)/a_n) = K \circ \mathcal{T}(x)$ with $\mathcal{T}(x) \coloneqq (t-x)/a_n)$ is absolutely continuous.

Note that $\mathcal{T}^{-1}(u) = t - ua_n$. Then by Hewitt (1960, p. 423)

$$\begin{aligned}\int_{S_n} \psi(x)(K \circ \mathcal{T})(dx) &= \int_{S_n} \psi(x)(K \circ \mathcal{T})'(x)\lambda(dx) \\ &= -\int_{S_n} \psi(x)K'(\mathcal{T}(x))a_n^{-1}\lambda(dx) \\ &= -\int_{\mathcal{T}^{-1}(S_n)} \psi(\mathcal{T}^{-1}(u))K'(u)a_n^{-1}\lambda_{\mathcal{T}}(du),\end{aligned}$$

where $\lambda_{\mathcal{T}}$ is the image measure of the Lebesgue measure λ induced by the transformation $\mathcal{T}$. Applying the transformation formula, cf. Cohn (2013, Theorem 6.1.7), gives

$$= -\int_{[r,s]} \psi(t - ua_n)K'(u)\lambda(du) = -\int_{[r,s]} \psi(t - ua_n)K(du),$$

where the last equality again relies on Hewitt (1960, p. 423). $\square$

The guiding idea to eventually obtain an asymptotic representation is to decompose the difference $f_{2,n}^{SE} - \bar{f}_n$ into an asymptotically negligible remainder and a contributing part for which we will give an i.i.d. representation. Having this in mind, Theorem 4.5 examines the difference $f_{2,n}^{SE} - f_{1,n}^{SE}$.

Proof of Theorem 4.5.

From Definition 4.2 and Definition 4.3 it follows

$$\left|f_{2,n}^{SE}(t) - f_{1,n}^{SE}(t)\right| = \left|\frac{1}{a_n}\int_{\mathbb{R}} K\left(\frac{t-x}{a_n}\right) F_{2,n}^{SE}(dx) - \frac{1}{a_n}\int_{\mathbb{R}} K\left(\frac{t-x}{a_n}\right) F_{1,n}^{SE}(dx)\right|.$$

Then first applying Corollary 4.17 and then using Corollary 4.18 gives

$$\begin{aligned}\left|f_{2,n}^{SE}(t) - f_{1,n}^{SE}(t)\right| =& \left|\frac{1}{a_n}\int_{\mathbb{R}} \left[F_{2,n}^{SE}(t - ua_n) - F_{1,n}^{SE}(t - ua_n)\right] K'(u)du\right| \\ \leq& \frac{1}{a_n}\sup_{0\leq x\leq T}\left|F_{2,n}^{SE}(x) - F_{1,n}^{SE}(x)\right| \int_r^s |K'(u)|\, du \\ =& \frac{1}{a_n}\sup_{0\leq x\leq T}\left|F_{2,n}^{SE}(x) - F_{1,n}^{SE}(x)\right| V_K \overset{a.s.}{=} \mathcal{O}((na_n)^{-1}).\end{aligned}$$

Note that $\int_r^s |K'(u)|\, du = V_K$ by Hewitt and Stromberg (1965, Theorem 18.1) where V_K denotes the total variation of the kernel function K which is finite due to (K10). The almost sure result follows then from Theorem 3.17. $\square$

We now determine asymptotic representations of $f_{1,n}^{SE} - \bar{f}_n$ which hold in probability as well as almost surely. Note that $\bar{f}_n$ is not the expectation of $f_{1,n}^{SE}$ if some observations are censored, compare (4.4).

Proof of Theorem 4.6 and Theorem 4.7 .

To avoid $\ln(0)$ when $F_{1,n}^{SE}(x) = 1$ in later calculations, we introduce

$$1 - \tilde{F}_{1,n}^{SE}(t) := \prod_{i:Z_i \le t} \left(1 - \frac{m(Z_i, \theta_n)}{n - R_n(Z_i) + 2}\right), \tag{4.7}$$

which turns out to be close to $1 - F_{1,n}^{SE}$. Note that $F_{1,n}^{SE}(t) \ge \tilde{F}_{1,n}^{SE}(t)$ for all $t \ge 0$. From Definition 4.2 and (4.4) we have

$$(na_n)^{1/2} \left| f_{1,n}^{SE}(t) - \bar{f}_n(t) \right| = (na_n)^{1/2} \left| \frac{1}{a_n} \int_{\mathbb{R}} K\left(\frac{t-x}{a_n}\right) F_{1,n}^{SE}(dx) - \frac{1}{a_n} \int_{\mathbb{R}} K\left(\frac{t-x}{a_n}\right) F(dx) \right|.$$

Then applying Corollary 4.17 for both $f_{1,n}^{SE}$ and $\bar{f}_n$ with $\mathcal{H} = F_{1,n}^{SE}$ and $\mathcal{H} = F$ respectively, gives

$$\begin{aligned}
(na_n)^{1/2} \left| f_{1,n}^{SE}(t) - \bar{f}_n(t) \right| &= (na_n)^{1/2} \left| \frac{1}{a_n} \int_{\mathbb{R}} \left[F_{1,n}^{SE}(x) - F(x)\right] K\left(\frac{t-dx}{a_n}\right) \right| \\
=& (na_n)^{1/2} \left| \frac{1}{a_n} \int_{\mathbb{R}} [F_{1,n}^{SE}(x) - \tilde{F}_{1,n}^{SE}(x)] K\left(\frac{t-dx}{a_n}\right) + \frac{1}{a_n} \int_{\mathbb{R}} [\tilde{F}_{1,n}^{SE}(x) - F(x)] K\left(\frac{t-dx}{a_n}\right) \right| \\
\equiv& (na_n)^{1/2} \left| I_1(n) + I_2(n) \right|.
\end{aligned}$$

As mentioned above, we have to introduce $\tilde{F}_{1,n}^{SE}$ to safeguard against $\ln(0)$. Lemma 4.19 shows that its distance to $F_{1,n}^{SE}$ is asymptotically not significant.

Lemma 4.19. Assuming (A2) and (A10), it holds for $0 \le T < \tau_H$ that

$$\sup_{0 \le t \le T} |F_{1,n}^{SE}(t) - \tilde{F}_{1,n}^{SE}(t)| \overset{a.s.}{=} \mathcal{O}(n^{-1}).$$

Proof. Very similar to the proof of Theorem 3.17, it follows from Remark 3.4 and (4.7)

$$T_n := |F_{1,n}^{SE}(t) - \tilde{F}_{1,n}^{SE}(t)| = \left| \prod_{i:Z_i \le t} \left(1 - \frac{m(Z_i, \theta_n)}{n - R_n(Z_i) + 2}\right) - \prod_{i:Z_i \le t} \left(1 - \frac{m(Z_i, \theta_n)}{n - R_n(Z_i) + 1}\right) \right|.$$

Now first rewriting the difference of the products using the representation given in Remark 3.22 and then using $R_n(Z_i) = nH_n(Z_i)$ yields

$$\begin{aligned}
T_n &\le \sum_{i:Z_i \le t} \left| \frac{m(Z_i, \theta_n)}{n - R_n(Z_i) + 1} - \frac{m(Z_i, \theta_n)}{n - R_n(Z_i) + 2} \right| \\
&= \sum_{i:Z_i \le t} \frac{m(Z_i, \theta_n)}{(n - R_n(Z_i) + 1)(n - R_n(Z_i) + 2)} \\
&\le \sum_{i:Z_i \le t} \frac{m(Z_i, \theta_n)}{(n - R_n(Z_i))^2} = \frac{1}{n^2} \sum_{i:Z_i \le t} \frac{m(Z_i, \theta_n)}{(\bar{H}_n(Z_i))^2} = \frac{1}{n} \int_0^t \frac{m(x, \theta_n)}{(\bar{H}_n(x))^2} H_n(dx).
\end{aligned}$$

Since $0 \le m(\cdot, \cdot) \le 1$ for n large enough due to (A2) and (A10) and $\bar{H}_n(x) \ge \bar{H}_n(t) \ge \bar{H}_n(T)$ we have

$$T_n \le \frac{1}{n} \frac{1}{(\bar{H}_n(T))^2} \int_0^t H_n(dx) \le \frac{1}{n} \frac{1}{(\bar{H}_n(T))^2}.$$

The result follows by the SLLN and since $H(T) < 1$,

$$\bar{H}_n^{-2}(T) \xrightarrow[n \to \infty]{a.s.} \bar{H}^{-2}(T) < \infty,$$

and holds uniformly in $t \in [0, T]$. $\square$

Having obtained the last lemma, we can now examine,

$$I_1(n) = \frac{1}{a_n} \int_{\mathbb{R}} \left[F_{2,n}^{SE}(x) - \tilde{F}_{2,n}^{SE}(x) \right] K\left(\frac{t - dx}{a_n} \right).$$

Using T' such that $0 \leq T < T' < \tau_H$ in Lemma 4.19, in particular $t - ua_n \leq T'$ for all n larger than some N. Then applying Corollary 4.18 yields

$$\begin{aligned} |I_1(n)| &= \left| \frac{1}{a_n} \int_{\mathbb{R}} \left[\tilde{F}_{2,n}^{SE}(t - ua_n) - F_{2,n}^{SE}(t - ua_n) \right] K'(u) du \right| \\ &\leq \frac{1}{a_n} \sup_{0 \leq t \leq T'} |F_{2,n}^{SE}(t) - \tilde{F}_{2,n}^{SE}(t)| \int_{\mathbb{R}} |K'(u)|\, du. \end{aligned}$$

Since (K10) by Hewitt and Stromberg (1965, Theorem 18.1), $\int_{\mathbb{R}} |K'(u)|\, du = V_K$. Hence

$$|I_1(n)| \leq \frac{1}{a_n} \sup_{0 \leq t \leq T'} |F_{2,n}^{SE}(t) - \tilde{F}_{2,n}^{SE}(t)| V_K \stackrel{a.s.}{=} \mathcal{O}((na_n)^{-1}),$$

according to Lemma 4.19. Therefore we have shown that

$$I_1(n) \stackrel{a.s.}{=} \mathcal{O}((na_n)^{-1}). \qquad \square$$

In order to analyze the term $I_2(n)$, the following lemma splits up the difference $F_{1,n}^{SE}(x) - F(x)$ into several parts. Slightly different versions can be found in Diehl and Stute (1988, Lemma 5) and Breslow and Crowley (1974, Formula 7.12).

Lemma 4.20. Let $\tilde{F}_{1,n}^{SE}$ be as given in (4.7). If $\Lambda_{1,n}^{SE}$ is the estimator of the cumulative hazard function Λ defined in (3.18), then for all $x \leq T < \tau_H$ it holds true that

$$\begin{aligned} \tilde{F}_{1,n}^{SE}(x) - F(x) =& [1 - F(x)] \left[\Lambda_{1,n}^{SE}(x) - \Lambda(x) \right] - 2^{-1} e^{-\Lambda_n^*(x)} \left[\Lambda_{1,n}^{SE}(x) - \Lambda(x) \right]^2 \\ &+ e^{-\Lambda_n^{**}(x)} \left[-\ln\left(1 - \tilde{F}_{1,n}^{SE}(x)\right) - \Lambda_{1,n}^{SE}(x) \right], \end{aligned}$$

where $\Lambda_n^*(x)$ is some intermediate point between $\Lambda_{1,n}^{SE}(x)$ and $\Lambda(x)$ and Λ_n^{**} between $\Lambda_{1,n}^{SE}(x)$ and $-\ln(1 - \tilde{F}_{1,n}^{SE}(x))$.

Proof. In order to simplify the notation we leave out the argument x in the following calculation. Use Taylor expansion in combination with the intermediate value theorem to obtain

$$\exp\left(-\Lambda_{1,n}^{SE}\right) = \exp(-\Lambda) - \exp(-\Lambda)\left(\Lambda_{1,n}^{SE} - \Lambda\right) + 2^{-1}\exp(-\Lambda_n^*)\left(\Lambda_{1,n}^{SE} - \Lambda\right)^2,$$

$$\exp\left(\ln\left(1 - \tilde{F}_{1,n}^{SE}\right)\right) = \exp(-\Lambda_{1,n}^{SE}) - \exp(-\Lambda_n^{**})(-\ln(1 - \tilde{F}_{1,n}^{SE}) - \Lambda_{1,n}^{SE}).$$

Applying the basic equality $\Lambda(t) = -\ln(1 - F(t))$ from Lemma 2.3 together with the latter expansions yields

$$\begin{aligned}
F_{1,n}^{SE} - F &= (1 - F) - \left(1 - \tilde{F}_{1,n}^{SE}\right) = \exp(-\Lambda) - \left(1 - \tilde{F}_{1,n}^{SE}\right) \\
&= \left[\exp(-\Lambda) - \exp\left(-\Lambda_{1,n}^{SE}\right)\right] + \left[\exp(-\Lambda_{1,n}^{SE}) - \exp\left(\ln\left(1 - \tilde{F}_{1,n}^{SE}\right)\right)\right] \\
&= \left[\exp(-\Lambda)\left(\Lambda_{1,n}^{SE} - \Lambda\right) - 2^{-1}\exp(-\Lambda_n^*)\left(\Lambda_{1,n}^{SE} - \Lambda\right)^2\right] \\
&\qquad + \left[\exp(-\Lambda_n^{**})\left(-\ln\left(1 - \tilde{F}_{1,n}^{SE}\right) - \Lambda_{1,n}^{SE}\right)\right] \\
&= (1 - F)\left(\Lambda_{1,n}^{SE} - \Lambda\right) - 2^{-1}\exp(-\Lambda_n^*)\left(\Lambda_{1,n}^{SE} - \Lambda\right)^2 \\
&\qquad + \exp(-\Lambda_n^{**})\left(-\ln\left(1 - \tilde{F}_{1,n}^{SE}\right) - \Lambda_{1,n}^{SE}\right). \qquad \square
\end{aligned}$$

Turning to $I_2(n)$, rewrite the integrand by applying Lemma 4.20

$$\begin{aligned}
I_2(n) = &\frac{1}{a_n}\int_{\mathbb{R}}[1 - F(x)]\left[\Lambda_{1,n}^{SE}(x) - \Lambda(x)\right]\tilde{K}_{t,n}(dx) \\
&- \frac{1}{a_n}\int_{\mathbb{R}} 2^{-1}e^{-\Lambda_n^*(x)}\left[\Lambda_{1,n}^{SE}(x) - \Lambda(x)\right]^2\tilde{K}_{t,n}(dx) \\
&+ \frac{1}{a_n}\int_{\mathbb{R}} e^{-\Lambda_n^{**}(x)}\left[-\ln\left(1 - \tilde{F}_{1,n}^{SE}(x)\right) - \Lambda_{1,n}^{SE}(x)\right]\tilde{K}_{t,n}(dx) \\
\equiv\ & I_3(n) - I_4(n) + I_5(n),
\end{aligned}$$

where $\Lambda_n^*(x)$ and Λ_n^{**} are some intermediate points, as explained in Lemma 4.20. Here $\tilde{K}_{t,n} := K((t - x)/a_n)$ denotes the transformed kernel. To examining the parts $I_3(n)$, $I_4(n)$ and $I_5(n)$ separately, we first derive the following lemma.

Lemma 4.21. Assuming (A2) and (A10), it holds for $0 \le T < \tau_H$ that

$$\sup_{0\le t\le T} |-\ln(1-\tilde{F}_{1,n}^{SE}(t)) - \Lambda_{1,n}^{SE}(t)| \overset{a.s.}{=} \mathcal{O}(n^{-1}).$$

Proof. Comparable to Breslow and Crowley (1974, Lemma 7.1), the proof is based on the basic inequalities

$$-\frac{a}{1-a} \le \ln(1-a) \le -a, \qquad \text{for all } 0 \le a < 1. \tag{4.8}$$

To shorten the notation we use $m_i = m(Z_i, \theta_n)$ in the following calculation. Note that $\Lambda_{1,n}^{SE}(t) \ge -\ln(1-\tilde{F}_{1,n}^{SE}(t))$ for all $t > 0$. Due to the left-hand inequality of (4.8) it holds that

$$\begin{aligned}
-\ln(1-\tilde{F}_{1,n}^{SE}(t)) - \Lambda_{1,n}^{SE}(t) &= \sum_{i:Z_i\le t} -\ln\left(1-\frac{m_i}{n-R_n(Z_i)+2}\right) - \frac{m_i}{n-R_n(Z_i)+1} \\
&\le \sum_{i:Z_i\le t} \frac{m_i}{n-R_n(Z_i)+2-m_i} - \frac{m_i}{n-R_n(Z_i)+1} \\
&\le \sum_{i:Z_i\le t} \frac{m_i}{n-R_n(Z_i)+1} - \frac{m_i}{n-R_n(Z_i)+1} = 0.
\end{aligned}$$

Hence the difference is given by

$$\begin{aligned}
\Lambda_{1,n}^{SE}(t) + \ln(1-\tilde{F}_{2,n}^{SE}(t)) &= \Lambda_{1,n}^{SE}(t) + \ln\left(\prod_{i:Z_i\le t}\left(1-\frac{m(Z_i,\theta_n)}{n-R_n(Z_i)+2}\right)\right) \\
&= \Lambda_{1,n}^{SE}(t) + \sum_{i:Z_i\le t} \ln\left(1-\frac{m(Z_i,\theta_n)}{n-R_n(Z_i)+2}\right).
\end{aligned}$$

Then applying the right-hand side inequality of (4.8) yields

$$\le \sum_{i:Z_i\le t}\left[\frac{m(Z_i,\theta_n)}{n-R_n(Z_i)+1} - \frac{m(Z_i,\theta_n)}{n-R_n(Z_i)+2}\right] \le \sum_{i:Z_i\le t} \frac{m(Z_i,\theta_n)}{(n-R_n(Z_i))^2} \le \frac{1}{n(\bar{H}_n(T))^2},$$

where the latter two inequalities are derived equivalently as in the proof of Lemma 4.19.

Since $H(T) < 1$, the assertion follows by the SLLN

$$\bar{H}_n^{-2}(T) \xrightarrow[n\to\infty]{a.s.} \bar{H}^{-2}(T) < \infty$$

and holds uniformly in $t \in [0, T]$. $\square$

Consider $I_5(n)$. Note that $0 < e^{-a} \leq 1 \; \forall \; a \geq 0$. Again using T' such that $0 \leq T < T' < \tau_H$ in Lemma 4.21 and applying Corollary 4.18 yields

$$\begin{aligned}
|I_5(n)| =& \left| \frac{1}{a_n} \int_{\mathbb{R}} e^{-\Lambda_n^{**}(t-ua_n)} \left[\Lambda_{1,n}^{SE}(t-ua_n) + \ln\left(1 - \tilde{F}_{1,n}^{SE}(t-ua_n)\right)\right] K'(u)du \right| \\
\leq & \frac{1}{a_n} \int_{\mathbb{R}} \left| \left[-\ln\left(1 - \tilde{F}_{1,n}^{SE}(t-ua_n)\right) - \Lambda_{1,n}^{SE}(t-ua_n)\right] K'(u) \right| du \\
\leq & \frac{1}{a_n} \sup_{0\leq t\leq T'} \left| -\ln\left(1 - \tilde{F}_{1,n}^{SE}(t)\right) - \Lambda_{1,n}^{SE}(t) \right| V_K \overset{a.s.}{=} \mathcal{O}((na_n)^{-1}).
\end{aligned}$$

Recall that $\int_{\mathbb{R}} |K'(u)| \, du = V_K$, the total variation of K, which is finite by (K10). The almost sure convergence result follows by Lemma 4.21. That is

$$I_5(n) \overset{a.s.}{=} \mathcal{O}((na_n)^{-1}). \qquad \square$$

The remaining terms $I_3(n)$ and $I_4(n)$ are primarily governed by the process $\Lambda_{1,n}^{SE} - \Lambda$. An asymptotic representation of this process was derived in Dikta (1998, Lemma 3.12). This representation could be used to prove our weak convergence results. For the sake of completeness, the result is quoted in the next lemma.

Lemma 4.22. Let Θ be a connected, open subset of $\mathbb{R}^k$. Given that H is continuous, $0 \leq t \leq T < \tau_H$ and assumptions (A1), (A3) to (A6) hold, then

$$\begin{aligned}\Lambda_{1,n}^{SE}(t) - \Lambda(t) =& \frac{1}{n} \sum_{i=1}^{n} \left\{ \frac{m(Z_i, \theta_0) I_{[Z_i \leq t]} - H^1(t)}{\bar{H}(t)} \right. \\ & \left. + \frac{\delta_i - m(Z_i, \theta_0)}{m(Z_i, \theta_0)(1 - m(Z_i, \theta_0))} \int_0^t \frac{\alpha(x, Z_i)}{\bar{H}(x)} H(dx) \right\} \\ & - \int_0^t \frac{H_n^1(x) - H^1(x)}{\bar{H}^2(x)} H(dx) + \int_0^t \frac{H_n(x) - H(x)}{\bar{H}^2(x)} H^1(dx) + o_p(n^{-1/2}).\end{aligned}$$

where H^1 and $\alpha(x, y)$ as given in Corollary 2.7 and Theorem 3.10, respectively. Furthermore $H_n^1(t) \coloneqq \int_0^t m(x, \theta_0) H_n(dx)$.

Since we are also interested in strong convergence results, in particular see Theorem 4.7, this representation is not sufficient. The next lemma provides weak and strong convergence rates for $\Lambda_{1,n}^{SE}$ and therefore extends Dikta (1998, Theorem 2.4).

Lemma 4.23. Let Θ be a connected, open subset of $\mathbb{R}^k$. Given that H is continuous then under the assumptions (A1) to (A7), (A10) it holds for all $0 \leq T < \tau_h$ that

$$\sup_{0 \leq t \leq T} \left| \Lambda_{1,n}^{SE}(t) - \Lambda(t) \right| = \begin{cases} \mathcal{O}\left(\left(\frac{\ln \ln(n)}{n} \right)^{1/2} \right) & \text{a.s.} \\ \mathcal{O}(n^{-1/2}) & \text{in probability.} \end{cases}$$

In order to prove those rates we first derive lemmata 4.24, 4.25 and 4.26.

Lemma 4.24. Given H is continuous, then for all $0 \le t \le T < \tau_H$ and $\theta_0 = (\theta_{0,1}, \ldots, \theta_{0,k})$ we have

$$\begin{aligned}
\Lambda_{1,n}^{SE}(t) - \Lambda(t) =& \frac{H_n^1(t) - H^1(t)}{\bar{H}(t)} - \int_0^t \frac{H_n^1(x) - H^1(x)}{\bar{H}^2(x)} H(dx) \\
&+ \int_0^t \frac{H_n(x) - H(x)}{\bar{H}_n(x)\bar{H}(x)} H_n^1(dx) + \int_0^t \frac{m(x,\theta_n) - m(x,\theta_0)}{\bar{H}_n(x) + 1/n} H_n(dx) \\
&- \frac{1}{n} \int_0^t \frac{m(x,\theta_0)}{\bar{H}(x)(\bar{H}_n(x) + 1/n)} H_n(dx) \qquad (4.9) \\
\equiv& \, Q_{n,1}(t) + Q_{n,2}(t) + Q_{n,3}(t) + Q_{n,4}(t) + Q_{n,5}(t).
\end{aligned}$$

Proof. By simply adding and subtracting

$$\begin{aligned}
\Lambda_{1,n}^{SE}(t) - \Lambda(t) = & \left[\int_0^t \frac{m(x,\theta_n)}{\bar{H}_n(x) + 1/n} H_n(dx) - \int_0^t \frac{m(x,\theta_0)}{\bar{H}_n(x) + 1/n} H_n(dx) \right] \\
&+ \left[\int_0^t \frac{m(x,\theta_0)}{\bar{H}_n(x) + 1/n} H_n(dx) - \int_0^t \frac{m(x,\theta_0)}{\bar{H}_n(x)} H_n(dx) \right] \\
&+ \left[\int_0^t \frac{m(x,\theta_0)}{\bar{H}_n(x)} H_n(dx) - \int_0^t \frac{m(x,\theta_0)}{\bar{H}(x)} H_n(dx) \right] \\
&+ \left[\int_0^t \frac{m(x,\theta_0)}{\bar{H}(x)} H_n(dx) - \int_0^t \frac{m(x,\theta_0)}{\bar{H}(x)} H(dx) \right].
\end{aligned}$$

Note that $(1 - H)^{-2}$ is the Radon–Nikodym derivative of $d([1 - H]^{-1})$ with respect to dH. Therefore, due to integration by parts and equivalently to the proof of Dikta (1998, Lemma 3.12), we have for the difference in the last line above

$$\int_0^t \frac{m(x,\theta_0)}{\bar{H}(x)} d[H_n(x) - H(x)] = \frac{H_n^1(t) - H^1(t)}{\bar{H}(t)} - \int_0^t \frac{H_n^1(x) - H^1(x)}{\bar{H}^2(x)} H(dx).$$

Then simple algebra yields $\Lambda_{1,n}^{SE}(t) - \Lambda(t) = Q_{n,4}(t) + Q_{n,5}(t) + Q_{n,3}(t) + Q_{n,1}(t) + Q_{n,2}(t)$. $\square$

Lemma 4.25. Let θ_n be the MLE of the true parameter θ_0 as introduced in Definition 3.7 and let Θ be a connected, open subset of $\mathbb{R}^k$. Under the assumptions (A1) to (A7) it holds that

$$\|\theta_n - \theta_0\| = \begin{cases} \mathcal{O}\left(\left(\frac{\ln\ln(n)}{n}\right)^{1/2}\right) & \text{a.s.} \\ \mathcal{O}(n^{-1/2}) & \text{in probability.} \end{cases}$$

The weak convergence result follows directly from Dikta (1998, Theorem 2.3). Since the proof of the a.s. convergence rate is rather technical and does not contribute to the topic, it is postponed to the appendix.

Lemma 4.26. Given that H is continuous, $0 \leq t \leq T < \tau_H$ and assumption (A7) holds, then

$$\sup_{0\leq t\leq T'} \left|H_n^1(t) - H^1(t)\right| = \begin{cases} \mathcal{O}\left(\left(\frac{\ln\ln(n)}{n}\right)^{1/2}\right) & \text{a.s.} \\ \mathcal{O}(n^{-1/2}) & \text{in probability.} \end{cases}$$

Proof. By (A7), $m(x, \theta_0)$ is absolutely continuous in x on $[0, T']$ with $0 \leq T' < \tau_H$. By Hewitt and Stromberg (1965, Theorem 18.13) there exist absolute continuous, nondecreasing functions m_a and m_b such that

$$m(x, \theta_0) = m_a(x) - m_b(x).$$

Note that m_a and m_b are bounded on $[0, T']$. Then we have for H^1 and H_n^1 as defined in (2.2) and (3.5), respectively,

$$\begin{aligned} H_n^1(t) - H^1(t) = &\left(\int_{[0,t]} m_a(s)H_n(ds) - \int_{[0,t]} m_a(s)H(ds)\right) \\ &- \left(\int_{[0,t]} m_b(s)H_n(ds) - \int_{[0,t]} m_b(s)H(ds)\right) \equiv A_n(t) - B_n(t). \end{aligned}$$

Exploiting the properties of H and H_n, integration by parts, cf. Hewitt and Stromberg (1965, Theorem 21.67), yields

$$A_n(t) = -\int_{[0,t]} \left[H_n(s^-) - H(s)\right] m_a(ds) + (H_n(t) - H(t))m_a(t).$$

Therefore

$$\begin{aligned}\sup_{0\le t\le T'} |A_n(t)| \le & \Big(\sup_{0\le t\le T'} |H_n(t) - H(t)| + 1/n\Big)\Big(m_a(T') - m_a(0)\Big) \\ & + \sup_{0\le t\le T'} |H_n(t) - H(t)| \sup_{0\le t\le T'} |m_a(t)| \\ = & \begin{cases} \mathcal{O}\left(\left(\frac{\ln\ln(n)}{n}\right)^{1/2}\right) & \text{a.s.} \\ \mathcal{O}(n^{-1/2}) & \text{in probability,}\end{cases}\end{aligned}$$

where the convergence rates are due to the law of iterated logarithm, cf. Serfling (2001, p. 62, Theorem B), and the Dvoretzky, Kiefer, and Wolfowitz (1956) (DKW) inequality. The same holds true for $B_n(t)$. □

Relying on the previous results, we can now prove Lemma 4.23.

Proof of Lemma 4.23. Consider the representation of $\Lambda_{1,n}^{SE} - \Lambda_n$ given in Lemma 4.24 and examine the summands $Q_{n,1}(t)$ to $Q_{n,5}(t)$ separately. Since $x \le t \le T < \tau_H$, $m(\cdot, \theta_0) \le 1$ for n large enough and both, H and H_n are increasing

$$-Q_{n,5}(t) = \frac{1}{n}\int_0^t \frac{m(x,\theta_0)}{\bar{H}(x)(\bar{H}_n(x) + 1/n)} H_n(dx) \le \frac{1}{n}\frac{1}{\bar{H}(T)(\bar{H}_n(T))}.$$

Since $\bar{H}_n(T) \to \bar{H}(T)$ as $n \to \infty$ by SLLN and $H(T) < 1$ we have

$$\sup_{0\le t\le T} Q_{n,5}(t) \overset{a.s.}{=} \mathcal{O}\left(n^{-1}\right),$$

uniformly in $t \in [0, T]$. Moreover, Lemma 4.26 immediately yields

$$\sup_{0 \le t \le T'} (Q_{n,1}(t) + Q_{n,2}(t) + Q_{n,3}(t)) = \begin{cases} \mathcal{O}\left(\left(\frac{\ln\ln(n)}{n}\right)^{1/2}\right) & \text{a.s.} \\ \mathcal{O}(n^{-1/2}) & \text{in probability.} \end{cases}$$

Since (A6), there exists measurable function M such that for all $\theta \in V(\theta_0)$, $x \ge 0$, and $1 \le r, s \le k$, $\nabla_{r,s} m(x\theta) \le M(x)$ and $\mathbb{E}(M(Z)) < \infty$. Then in a similar fashion as Dikta (1998, p. 265), we get by Taylor expansion and strong consistency of θ_n for n large enough

$$|m(x, \theta_n) - m(x, \theta_0)| \le k \|\theta_n - \theta_0\| M(x).$$

Furthermore, for n large and small $\epsilon > 0$ it yields $\forall\, 0 \le x \le t$, $(\bar{H}_n(x) + 1/n) > (\bar{H}(t) - \epsilon) > 0$ almost surely. Therefore

$$\begin{aligned} Q_{n,4}(t) &\le \int_0^t \frac{|m(x, \theta_n) - m(x, \theta_0)|}{\bar{H}_n(x) + 1/n} H_n(dx) \le k \|\theta_n - \theta_0\| \int_0^t \frac{M(x)}{\bar{H}_n(x) + 1/n} H_n(dx) \\ &\le k \|\theta_n - \theta_0\| \frac{1}{\bar{H}_n(T) - \epsilon} \int_0^\infty M(x) H_n(dx). \end{aligned}$$

Note that $\int_0^\infty M(y) H_n(dy) \xrightarrow[n \to \infty]{a.s.} \mathbb{E}(M(Z)) < \infty$ and $\bar{H}_n(T) \to \bar{H}(T) < 1$ as $n \to \infty$ by SLLN. Then Lemma 4.25 yields

$$\sup_{0 \le t \le T} Q_{n,4}(t) = \begin{cases} \mathcal{O}\left(\left(\frac{\ln\ln(n)}{n}\right)^{1/2}\right) & \text{a.s.} \\ \mathcal{O}(n^{-1/2}) & \text{in probability.} \end{cases}$$

The weak consistency result could have been derived as an immediate consequence of Dikta (1998, Theorem 2.5). □

Employing the latter result for $0 \leq T < T' < \tau_H$ and one more time using $0 < e^{-a} \leq 1$ $\forall\, a \geq 0$ as well as Corollary 4.18 gives

$$\begin{aligned}
|I_4(n)| &= \left| \frac{1}{a_n} \int_{\mathbb{R}} 2^{-1} e^{-\Lambda_n^*(t-ua_n)} \left[\Lambda_{1,n}^{SE}(t-ua_n) - \Lambda(x)\right]^2 K'(u) du \right| \\
&\leq \frac{1}{a_n} \int_{\mathbb{R}} \left[\Lambda_{1,n}^{SE}(t-ua_n) - \Lambda(t-ua_n)\right]^2 |K'(u)|\, du \\
&\leq \frac{1}{a_n} \left\{ \sup_{0 \leq t \leq T'} \left|\Lambda_{1,n}^{SE}(t) - \Lambda(t)\right| \right\}^2 V_K = \begin{cases} \mathcal{O}\left(\frac{\ln\ln(n)}{na_n}\right) & \text{a.s.} \\ \mathcal{O}((na_n)^{-1}) & \text{in probability,} \end{cases}
\end{aligned}$$

where V_K denotes the total variation of K. The asymptotic result follows by Lemma 4.23 when replacing T by T'. Recall, we decomposed

$$(na_n)^{1/2} \left| f_{1,n}^{SE}(t) - \bar{f}_n(t) \right| = (na_n)^{1/2} \left| I_1(n) + I_2(n) \right|$$

with $I_2(n) = I_3(n) - I_4(n) + I_5(n)$. So far we have shown that $I_1(n)$, $I_4(n)$ and $I_5(n)$ vanish as $n \to \infty$ at a sufficient rate. Hence it is left to analyze $I_3(n)$. Before applying Lemma 4.24 to handle the remaining term, we further investigate the first summand of the therein given representation. For that reason note

$$\frac{1}{\bar{G}(x)} = \frac{G(x) - G(t)}{\bar{G}(x)\bar{G}(t)} + \frac{1}{\bar{G}(t)}.$$

An application of Corollary 2.6 and subsequently using the previous expansion gives

$$\begin{aligned}
(1 - F(x)) \left[\frac{H_n^1(x) - H^1(x)}{\bar{H}(x)} \right] =& \frac{H_n^1(x) - H^1(x)}{\bar{G}(x)} \\
=& \frac{H_n^1(x) - H^1(x)}{\bar{G}(t)} + \frac{G(x) - G(t)}{\bar{G}(x)\bar{G}(t)} (H_n^1(x) - H^1(x)). \qquad (4.10)
\end{aligned}$$

Now using the representation for $\Lambda_{1,n}^{SE} - \Lambda$ given in Lemma 4.24 in combination with (4.10), $I_3(n)$ can be written as

$$I_3(n) = \frac{1}{a_n} \int_{S_n} [1 - F(x)] \left[\Lambda_{1,n}^{SE}(x) - \Lambda(x)\right] K\left(\frac{t - dx}{a_n}\right) = \mathsf{A}_n + \mathsf{B}_n + \mathsf{C}_n + \mathsf{D}_n + \mathsf{E}_n + \mathsf{F}_n$$

where

$$\begin{aligned}
\mathsf{A}_n(t) =& \frac{1}{a_n} \frac{1}{\bar{G}(t)} \int_{S_n} \left[H_n^1(x) - H^1(x)\right] K\left(\frac{t - dx}{a_n}\right), \\
\mathsf{B}_n(t) =& \frac{1}{a_n} \int_{S_n} \frac{G(x) - G(t)}{\bar{G}(x)\bar{G}(t)} (H_n^1(x) - H^1(x)) K\left(\frac{t - dx}{a_n}\right), \\
\mathsf{C}_n(t) =& \frac{1}{a_n} \int_{S_n} [1 - F(x)] \left[\int_0^x \frac{H_n^1(y) - H^1(y)}{(\bar{H}(y))^2} H(dy)\right] K\left(\frac{t - dx}{a_n}\right), \\
\mathsf{D}_n(t) =& \frac{1}{a_n} \int_{S_n} [1 - F(x)] \left[\int_0^x \frac{H_n(y) - H(y)}{\bar{H}_n(y)\bar{H}(y)} H_n^1(dy)\right] K\left(\frac{t - dx}{a_n}\right), \\
\mathsf{E}_n(t) =& \frac{1}{a_n} \int_{S_n} [1 - F(x)] \left[\int_0^x \frac{m(y, \theta_n) - m(y, \theta_0)}{\bar{H}_n(y) + 1/n} H_n(dy)\right] K\left(\frac{t - dx}{a_n}\right), \\
\mathsf{F}_n(t) =& \frac{1}{n a_n} \int_{S_n} [1 - F(x)] \left[\int_0^x \frac{m(y, \theta_0)}{\bar{H}(y)(\bar{H}_n(y) + 1/n)} H_n(dy)\right] K\left(\frac{t - dx}{a_n}\right).
\end{aligned}$$

The terms C_n to F_n are an immediate result when plugging Lemma 4.24 into $I_3(n)$, whereas A_n and B_n were introduced when applying (4.10).

In the following, we will show that A_n is the only contributing term. The others will turn out to be asymptotically negligible. The technique to actually show this is very similar for the terms B_n to F_n. We present the treatment of F_n in detail and only proof the key parts for the other terms.

First consider F_n and let

$$W_{\mathsf{F},n}(x) := \int_0^x \frac{m(y, \theta_0)}{\bar{H}(y)(\bar{H}_n(y) + 1/n)} H_n(dy)$$

denote its inner integral.

Then for $0 \le t \le T < T' < \tau_H$ by using Corollary 4.18 and $\mathfrak{T}(x) = (t-x)/a_n)$ we have

$$\begin{aligned}
|\mathsf{F}_n(t)| &= \left| \frac{1}{na_n} \int_{S_n} \bar{F}(x) W_{\mathsf{F},n}(x) K\left(\mathfrak{T}(dx)\right) \right| \\
&= \left| \frac{1}{na_n} \int_{[r,s]} \bar{F}(t-ua_n) W_{\mathsf{F},n}(t-ua_n) K'(u) du \right| \\
&\le \left| \frac{1}{na_n} W_{\mathsf{F},n}(t) \int_{[r,s]} \bar{F}(t-ua_n) K'(u) du \right| \\
&\quad + \left| \frac{1}{na_n} \int_{[r,s]} \bar{F}(t-ua_n) \left[W_{\mathsf{F},n}(t-ua_n) - W_{\mathsf{F},n}(t) \right] K'(u) du \right| \\
&\equiv \left| A_{\mathsf{F},n}(t) \right| + \left| B_{\mathsf{F},n}(t) \right|.
\end{aligned}$$

Now examine $W_{\mathsf{F},n}(t)$. Since $y \le t \le T$ and $m(\cdot, \theta_0) \le 1$

$$\sup_{0 \le t \le T} W_{\mathsf{F},n}(t) \le \frac{1}{\bar{H}(T) \bar{H}_n(T)}. \tag{4.11}$$

Now let $0 \le t \le T$ and n large enough such that $|t - ua_n| < T'$ for all $u \in [r,s]$. Then

$$\begin{aligned}
\left| W_{\mathsf{F},n}(t-ua_n) - W_{\mathsf{F},n}(t) \right| &\le \frac{1}{\bar{H}(T') \bar{H}_n(T')} \left| H_n(t-ua_n) - H_n(t) \right| \\
&= \frac{1}{\bar{H}(T') \bar{H}_n(T')} \left(2 \left\| H_n - H \right\| + \left| H(t-ua_n) - H(t) \right| \right) \\
&= \frac{1}{\bar{H}(T') \bar{H}_n(T')} \left(2 \left\| H_n - H \right\| + h(y^*(t, a_n, u)) \left| u \right| a_n \right),
\end{aligned}$$

where $y^*(t, a_n, u)$ between $t - ua_n$ and t. Hence $|y^*(t, a_n, u)| < T'$.

Therefore we have

$$\begin{aligned}
&\sup_{0 \le t \le T} \sup_{\substack{r \le u \le s \\ |t-ua_n| < T'}} \left| W_{\mathsf{F},n}(t-ua_n) - W_{\mathsf{F},n}(t) \right| \\
&\quad \le \sup_{0 \le t \le T} \sup_{\substack{r \le u \le s \\ |t-ua_n| < T'}} \frac{1}{\bar{H}(T') \bar{H}_n(T')} \left(2 \left\| H_n - H \right\| + \sup_{0 \le t \le T'} h(t) \max(|r|, |s|) a_n \right).
\end{aligned}$$

which together with the DKW inequality and the law if iterated logarithm finally leads to

$$\sup_{0\le t\le T}\sup_{r\le u\le s}|W_{\mathsf{F},n}(t-ua_n)-W_{\mathsf{F},n}(t)| = \begin{cases} \mathcal{O}\left(\left(\frac{2\ln\ln(n)}{n}\right)^{1/2}\right)+\mathcal{O}\left(a_n\right) & \text{a.s.} \\ \mathcal{O}((n)^{-1/2})+\mathcal{O}\left(a_n\right) & \text{in probability.} \end{cases} \tag{4.12}$$

Due to (K11) we have $\int_{[r,s]} K'(u)du = 0$ and therefore it holds that

$$\begin{aligned} \int_{[r,s]}[1-F(t-ua_n)]K'(u)du &= \int_{[r,s]}\left[F(t)-F(t-ua_n)\right]K'(u)du \\ &= \int_{[r,s]} f(u^*(t,a_n,u))ua_nK'(u)du, \end{aligned}$$

where we used Taylor expansion in combination with the intermediate value theorem. Here $u^*(t,a_n,u)$ is some value between t and $t-ua_n$. Hence

$$\sup_{0\le t\le T}\left|\int_{[r,s]}[1-F(t-ua_n)]K'(u)du\right| \le a_n \sup_{0\le t\le T'} f(t)\max(|r|,|s|)\int_r^s |K'(u)|\,du. \tag{4.13}$$

Then by (4.11) and (4.13) it follows

$$|A_{\mathsf{F},n}(t)| \le \frac{1}{n\bar{H}(T')\bar{H}_n(T')}\sup_{0\le t\le T'} f(t)\max(|r|,|s|)\int_r^s |K'(u)|\,du.$$

In other words we have

$$\sup_{0\le t\le T}|A_{\mathsf{F},n}(t)| \overset{a.s.}{=} \mathcal{O}(n^{-1}).$$

Similarly for $B_{\mathsf{F},n}(t)$, it follows from (4.12) and (4.13) that

$$\begin{aligned}
\sup_{0\le t\le T}|B_{\mathsf{F},n}(t)| &\le \frac{1}{na_n}\sup_{0\le t\le T}\sup_{r\le u\le s}|W_{\mathsf{F},n}(t-ua_n)-W_{\mathsf{F},n}(t)|\sup_{0\le t\le T}\left|\int_r^s \bar{F}(t-ua_n)K'(u)du\right| \\
&= \begin{cases} \frac{1}{n}\left[\mathcal{O}\left(\left(\frac{2\ln\ln(n)}{n}\right)^{1/2}\right)+\mathcal{O}\left(a_n\right)\right] & \text{a.s.} \\ \frac{1}{n}\left[\mathcal{O}((n)^{-1/2})+\mathcal{O}\left(a_n\right)\right] & \text{in probability.} \end{cases} \\
&= \begin{cases} \mathcal{O}\left(\left(\frac{2\ln\ln(n)}{n^3}\right)^{1/2}\right)+\mathcal{O}\left(a_n n^{-1}\right) & \text{a.s.} \\ \mathcal{O}\left((n)^{-3/2}\right)+\mathcal{O}\left(a_n n^{-1}\right) & \text{in probability.} \end{cases}
\end{aligned}$$

All together we have for F_n

$$\sup_{0\le t\le T}|\mathsf{F}_n(t)| \le \begin{cases} \mathcal{O}\left(\left(\frac{2\ln\ln(n)}{n^3}\right)^{1/2}\right)+\mathcal{O}\left(a_n n^{-1}\right) & \text{a.s.} \\ \mathcal{O}\left((n)^{-3/2}\right)+\mathcal{O}\left(a_n n^{-1}\right) & \text{in probability.} \end{cases}$$

In order to handle E_n we proceed in a similar way. Let

$$W_{\mathsf{E},n}(x) := \int_0^x \frac{m(y,\theta_n)-m(y,\theta_0)}{\bar{H}_n(y)+1/n}H_n(dy)$$

be the inner integral of E_n. Then similarly as above

$$\begin{aligned}
|\mathsf{E}_n(t)| &\le \left|\frac{1}{na_n}W_{\mathsf{E},n}(t)\int_{[r,s]}\bar{F}(t-ua_n)K'(u)du\right| \\
&\quad + \left|\frac{1}{na_n}\int_{[r,s]}\bar{F}(t-ua_n)\left[W_{\mathsf{E},n}(t-ua_n)-W_{\mathsf{E},n}(t)\right]K'(u)du\right| \\
&\equiv |A_{\mathsf{E},n}(t)|+|B_{\mathsf{E},n}(t)|\,.
\end{aligned}$$

Expanding $m(y, \cdot)$ in combination with the intermediate value theorem yields

$$W_{\mathsf{E},n}(x) = \int_0^x \frac{\langle \nabla m(y, \theta^*(y, \theta_n, \theta)), \theta_n - \theta_0 \rangle}{\bar{H}_n(y) + 1/n} H_n(dy)$$

where $\theta^*(y, \theta_n, \theta) \in \Theta$ lies in the interior of the line segment connecting θ_n and θ_0. Hence

$$\|\theta^*(y, \theta_n, \theta) - \theta_0\| \leq \|\theta_n - \theta_0\| .$$

Now let $V(\theta_0)$ be the neighborhood of θ_0 from (A6). Then due to (A2) for n large enough

$$\begin{aligned} \sup_{0 \leq t \leq T} |W_{\mathsf{E},n}(t)| &\overset{a.s.}{=} \frac{1}{H_n(T)} \|\theta_n - \theta_0\| \sup_{0 \leq t \leq T} \sup_{\theta \in V(\theta_0)} \|\nabla m(x, \theta)\| \\ &= \begin{cases} \mathcal{O}\left(\left(\frac{2 \ln \ln(n)}{n} \right)^{1/2} \right) & \text{a.s.} \\ \mathcal{O}\left(n^{-1/2} \right) & \text{in probability.} \end{cases} \end{aligned} \tag{4.14}$$

Furthermore, for n large enough, because of (A2) and (A6) it holds that

$$\begin{aligned} &\sup_{0 \leq t \leq T} \sup_{\substack{r \leq u \leq s \\ |t - ua_n| < T'}} |W_{\mathsf{E},n}(t - ua_n) - W_{\mathsf{E},n}(t)| \\ &\quad \leq \frac{1}{\bar{H}_n(T)} \left(\sup_{0 \leq t \leq T'} \sup_{\theta \in V(\theta_0)} \|\nabla m(x, \theta)\| \right) \|\theta_n - \theta_0\| \sup_{0 \leq t \leq T'} \sup_{\substack{r \leq u \leq s \\ |t - ua_n| < T'}} (H_n(t) - H_n(t - ua_n)) \\ &\quad \leq \frac{1}{\bar{H}_n(T)} \left(\sup_{0 \leq t \leq T'} \sup_{\theta \in V(\theta_0)} \|\nabla m(x, \theta)\| \right) \|\theta_n - \theta_0\| \\ &\qquad\qquad \times \left(2 \|H_n - H\| + \sup_{o \leq t \leq T'} h(t) \max(|r|, |s|) a_n \right) . \end{aligned}$$

Therefore it yields

$$
\begin{aligned}
&\sup_{0\le t\le T}\sup_{|t-ua_n|<T'}|W_{\mathsf{E},n}(t-ua_n)-W_{\mathsf{E},n}(t)|\\
&\qquad=\begin{cases}\mathcal{O}\left(\left(\frac{2\ln\ln(n)}{n}\right)^{1/2}\right)\left\{\mathcal{O}\left(\left(\frac{2\ln\ln(n)}{n}\right)^{1/2}\right)+\mathcal{O}\left(a_n\right)\right\} & \text{a.s.}\\ \mathcal{O}\left(n^{-1/2}\right)\left\{\mathcal{O}\left(n^{-1/2}\right)+\mathcal{O}\left(a_n\right)\right\} & \text{in probability.}\end{cases}\\
&\qquad=\begin{cases}\mathcal{O}\left(a_n\left(\frac{2\ln\ln(n)}{n}\right)^{1/2}\right) & \text{a.s.}\\ \mathcal{O}\left(a_nn^{-1/2}\right) & \text{in probability.}\end{cases}
\end{aligned}
\tag{4.15}
$$

We then have from (4.14) and (4.13)

$$
\sup_{0\le t\le T}|A_{\mathsf{E},n}(t)|=\begin{cases}\mathcal{O}\left(\left(\frac{2\ln\ln(n)}{n}\right)^{1/2}\right) & \text{a.s.}\\ \mathcal{O}\left(n^{-1/2}\right) & \text{in probability,}\end{cases}
$$

and from (4.15) it follows

$$
\sup_{0\le t\le T}|B_{\mathsf{E},n}(t)|=\begin{cases}\mathcal{O}\left(a_n\left(\frac{2\ln\ln(n)}{n}\right)^{1/2}\right) & \text{a.s.}\\ \mathcal{O}\left(a_nn^{-1/2}\right) & \text{in probability.}\end{cases}
$$

Therefore we have shown that

$$
\sup_{0\le t\le T}|\mathsf{E}_n(t)|=\begin{cases}\mathcal{O}\left(\left(\frac{2\ln\ln(n)}{n}\right)^{1/2}\right) & \text{a.s.}\\ \mathcal{O}\left(n^{-1/2}\right) & \text{in probability.}\end{cases}
$$

Consider $\mathsf{C}_n(t)$ and $\mathsf{D}_n(t)$ and denote the inner integrals by

$$W_{\mathsf{C},n}(x) := \int_0^x \frac{H_n^1(y) - H^1(y)}{\bar{H}^2(y)} H(dy), \qquad W_{\mathsf{D},n}(x) := \int_0^x \frac{H_n(y) - H(y)}{\bar{H}_n(y)\bar{H}(y)} H_n^1(dy).$$

Then, similar as above, it follows that

$$\sup_{0\le t\le T} |W_{\mathsf{D},n}(t)| \le \|H_n - H\| \frac{1}{\bar{H}_n(T)\bar{H}(T)} = \begin{cases} \mathcal{O}\left(\left(\frac{2\ln\ln(n)}{n}\right)^{1/2}\right) & \text{a.s.} \\ \mathcal{O}\left(n^{-1/2}\right) & \text{in probability.} \end{cases} \tag{4.16}$$

and

$$\sup_{0\le t\le T} |W_{\mathsf{C},n}(t)| \le \left\|H_n^1 - H^1\right\| \frac{1}{\bar{H}^2(T)} = \begin{cases} \mathcal{O}\left(\left(\frac{2\ln\ln(n)}{n}\right)^{1/2}\right) & \text{a.s.} \\ \mathcal{O}\left(n^{-1/2}\right) & \text{in probability.} \end{cases} \tag{4.17}$$

Furthermore it yields

$$\begin{aligned} &\sup_{0\le t\le T} \sup_{\substack{r\le u\le s \\ |t-ua_n|<T'}} |W_{\mathsf{C},n}(t-ua_n) - W_{\mathsf{C},n}(t)| \\ &\qquad \le \sup_{0\le t\le T'} \left|H_n^1(t) - H^1(t)\right| \frac{1}{\bar{H}^2(T')} \sup_{0\le t\le T'} \sup_{\substack{r\le u\le s \\ |t-ua_n|<T'}} |H(t) - H(t-uan)| \\ &\qquad \le \sup_{0\le t\le T'} \left|H_n^1(t) - H^1(t)\right| \frac{1}{\bar{H}^2(T')} \sup_{0\le t\le T'} h(t)\max(|r|, |s|)a_n, \end{aligned}$$

which leads to

$$\sup_{0\le t\le T} \sup_{r\le u\le s} |W_{\mathsf{C},n}(t-ua_n) - W_{\mathsf{C},n}(t)| = \begin{cases} \mathcal{O}\left(a_n\left(\frac{2\ln\ln(n)}{n}\right)^{1/2}\right) & \text{a.s.} \\ \mathcal{O}\left(a_n n^{-1/2}\right) & \text{in probability.} \end{cases} \tag{4.18}$$

Now turning to the term $\mathsf{D}_n(t)$, we have

$$\begin{aligned}
&\sup_{0\le t\le T} \sup_{\substack{r\le u\le s \\ |t-ua_n|<T'}} |W_{\mathsf{D},n}(t-ua_n) - W_{\mathsf{D},n}(t)| \\
&\qquad \le \sup_{0\le t\le T'} |H_n(t)-H(t)| \frac{1}{\bar{H}_n(T')\bar{H}(T')} \sup_{0\le t\le T'} \sup_{\substack{r\le u\le s \\ |t-ua_n|<T'}} \left|H_n^1(t) - H_n^1(t-ua_n)\right| \\
&\qquad \le \sup_{0\le t\le T'} |H_n(t)-H(t)| \frac{1}{\bar{H}_n(T')\bar{H}(T')} \\
&\qquad\qquad \times \left(2 \sup_{0\le t\le T'} (H_n^1(t) - H^1(t)) \sup_{0\le t\le T'} h(t)\max(|r|,|s|)a_n\right),
\end{aligned}$$

which yields

$$\sup_{0\le t\le T} \sup_{r\le u\le s} |W_{\mathsf{D},n}(t-ua_n) - W_{\mathsf{D},n}(t)| = \begin{cases} \mathcal{O}\left(a_n \left(\frac{2\ln\ln(n)}{n}\right)^{1/2}\right) & \text{a.s.} \\ \mathcal{O}\left(a_n n^{-1/2}\right) & \text{in probability.} \end{cases} \tag{4.19}$$

In order to derive an asymptotic representation of $\mathsf{B}_n(t)$ consider

$$\begin{aligned}
|\mathsf{B}_n(t)| &= \left|\frac{1}{a_n}\int_{S_n} \frac{G(x)-G(t)}{\bar{G}(x)\bar{G}(t)}(H_n^1(x) - H^1(x))K\left(\frac{t-dx}{a_n}\right)\right| \\
&= \frac{1}{a_n}\frac{1}{\bar{G}^2(T')} \sup_{0\le t\le T'} (H_n^1(t) - H^1(t)) \left|\int_r^s g(u^*(t,u,a_n))ua_n K'(u)du\right| \\
&= \frac{1}{\bar{G}^2(T')} \sup_{0\le t\le T'} (H_n^1(t) - H^1(t)) \sup_{0\le t\le T'} g(t)\max(|r|,|s|)\int_r^s K'(u)du.
\end{aligned}$$

Now use (4.17) and (4.18) as well as (4.16) and (4.19) in combination with (4.13) to obtain

$$\sup_{0\le t\le T} |\mathsf{B}_n(t) + \mathsf{C}_n(t) + \mathsf{D}_n(t)| = \begin{cases} \mathcal{O}\left(\left(\frac{\ln\ln(n)}{n}\right)^{1/2}\right) & \text{a.s.} \\ \mathcal{O}(n^{-1/2}) & \text{in probability.} \end{cases}$$

Recapitulating, we have shown so far that

$$(na_n)^{1/2}\left|f_{2,n}^{SE}(t)-\bar{f}_n(t)\right| = (na_n)^{1/2}\mathsf{A}_n + (na_n)^{1/2}\begin{cases}\mathcal{O}\left(\left(\frac{\ln\ln(n)}{na_n}\right)\right)+\mathcal{O}\left(\left(\frac{\ln\ln(n)}{n}\right)^{1/2}\right) & \text{a.s.}\\ \mathcal{O}(n^{-1/2})+\mathcal{O}((na_n)^{-1}) & \text{in probability.}\end{cases} \tag{4.20}$$

with $\mathsf{A}_n = \frac{1}{1-G(t)}\frac{1}{a_n}\int_{S_n}\left[H_n^1(x)-H^1(x)\right]K\left(\frac{t-dx}{a_n}\right)$. Using Corollary 4.17, note that

$$\begin{aligned}\mathsf{A}_n &= -\frac{1}{\bar{G}(t)}\frac{1}{a_n}\int_{S_n}K\left(\frac{t-x}{a_n}\right)\left[H_n^1-H^1\right](dx)\\ &= -\frac{1}{\bar{G}(t)}\left[\frac{1}{a_n}\int_{S_n}K\left(\frac{t-x}{a_n}\right)H_n^1(dx)-\frac{1}{a_n}\int_{S_n}K\left(\frac{t-x}{a_n}\right)H^1(dx)\right]\\ &= -\frac{h_n^1(t)-\mathbb{E}(h_n^1(t))}{\bar{G}(t)}.\end{aligned}$$

Hence we have

$$\begin{aligned}&(na_n)^{1/2}\sup_{0\le t\le T}\left|f_{2,n}^{SE}(t)-\bar{f}_n(t)-\frac{h_n^1(t)-\mathbb{E}(h_n^1(t))}{\bar{G}(t)}\right|\\ &\qquad=\begin{cases}\mathcal{O}\left(\left(\frac{\ln\ln(n)}{(na_n)^{1/2}}\right)\right)+\mathcal{O}\left((a_n\ln\ln(n))^{1/2}\right) & \text{a.s.}\\ \mathcal{O}(a_n^{1/2})+\mathcal{O}((na_n)^{-1/2}) & \text{in probability.}\end{cases}\end{aligned}$$

□

Proof of Theorem 4.10.

Recall (4.20), the definition of A_n from the latter proof, and

$$\frac{H_n^1(x) - H^1(x)}{\bar{G}(t)} = \frac{1}{n}\sum_{i=1}^{n} \frac{m(Z_i,\theta_0)I_{[Z_i \le x]} - H^1(x)}{\bar{G}(t)}.$$

Therefore

$$\begin{aligned}\mathsf{A}_n =& \frac{1}{\bar{G}(t)}\frac{1}{a_n}\int_{S_n}\left[H_n^1(x) - H^1(x)\right] K\left(\frac{t-dx}{a_n}\right)\\ =& \frac{1}{n}\sum_{i=1}^{n}\frac{1}{a_n}\frac{1}{\bar{G}(t)}\int_{S_n}\left[m(Z_i,\theta_0)I_{[Z_i\le x]} - H^1(x)\right] K\left(\frac{t-dx}{a_n}\right) \equiv \frac{1}{n}\sum_{i=1}^{n}\mathsf{A}_{i,n},\end{aligned}$$

where $\mathsf{A}_{i,n}$ only depends on the random variable (Z_i, δ_i). Hence, by CLT, the left hand side of (4.20) is asymptotically normal distributed:

$$(na_n)^{1/2}\left|f_{2,n}^{SE}(t) - \bar{f}_n(t)\right| \xrightarrow[n\to\infty]{} \mathcal{N}(\mu, \sigma_{SE}^2) \qquad \text{in distribution,}$$

where

$$\mu = \lim_{n\to\infty} a_n^{1/2}\mathbb{E}\left[\mathsf{A}_{i,n}\right] \qquad \text{and} \qquad \sigma_{SE}^2 = \lim_{n\to\infty} a_n \mathrm{Var}\left[\mathsf{A}_{i,n}\right].$$

For the sake of a brief notation let's set $m(Z_i) = m(Z_i,\theta_0)$ for $i = 1,\ldots,n$. Then calculating the expected value using Corollary 4.18 and Fubini's theorem, cf. Cohn (2013, Theorem 5.2.2), yields

$$\begin{aligned}\mathbb{E}\left[\mathsf{A}_{i,n}\right] =& \frac{1}{a_n\bar{G}(t)}\mathbb{E}\left[\int_{S_n}\left[m(Z_1)I_{[Z_1\le x]} - H^1(x)\right] K\left(\frac{t-dx}{a_n}\right)\right]\\ =& -\frac{1}{a_n\bar{G}(t)}\int_{[r,s]}\left[\mathbb{E}\left(m(Z_1)I_{[Z_1\le t-ua_n]}\right) - H^1(t-ua_n)\right] K'(u)du = 0,\end{aligned}$$

since $\mathbb{E}\left(m(Z_1)I_{[Z_1\le t-ua_n]}\right) = H^1(t-ua_n)$.

Thus $\mu = 0$ and it is left to calculate the variance of asymptotic variance:

$$\begin{aligned}
a_n \mathrm{Var}\left[\mathsf{A}_{i,n}\right] &= \mathrm{Var}\left[(na_n)^{1/2}\mathsf{A}\right] \\
&= \frac{1}{\bar{G}^2(t)}\frac{n}{a_n}\mathrm{Var}\left[\int_{S_n}\left[H_n^1(x) - H^1(x)\right] K\left(\frac{t-dx}{a_n}\right)\right] \\
&= \frac{1}{\bar{G}^2(t)}\frac{n}{a_n}\mathrm{Var}\left[\int_{S_n} K\left(\frac{t-x}{a_n}\right) H_n^1(dx) - \int_{S_n} K\left(\frac{t-x}{a_n}\right) H^1(dx)\right] \\
&= \frac{1}{\bar{G}^2(t)}\frac{n}{a_n}\mathrm{Var}\left[\frac{1}{n}\sum_{i=1}^{n} K\left(\frac{t-Z_i}{a_n}\right) m(Z_i)\right] \\
&= \frac{1}{\bar{G}^2(t)}\frac{1}{a_n}\left[\int_{S_n} K^2\left(\frac{t-x}{a_n}\right) m^2(x)h(x)dx - \left(\int_{S_n} K\left(\frac{t-x}{a_n}\right) m(x)h(x)dx\right)^2\right] \\
&= \frac{1}{\bar{G}^2(t)}\left[\int_{[r,s]} K^2\left(u\right) m^2(t-ua_n)h(t-ua_n)du - a_n\int_{S_n} K\left(\frac{t-x}{a_n}\right) m(x)h(x)du\right] \\
&\xrightarrow[n\to\infty]{} \frac{m^2(t)h(t)}{\bar{G}^2(t)}\int_{\mathbb{R}} K^2\left(u\right)du = \frac{m(t)h^1(t)}{\bar{G}^2(t)}\int_{\mathbb{R}} K^2\left(u\right)du = \frac{m(t)f(t)}{\bar{G}(t)}\int_{\mathbb{R}} K^2\left(u\right)du.
\end{aligned}$$

In conclusion, we have shown that

$$\sigma_{SE}^2 = \frac{m(t)h^1(t)}{(1-G(t))^2}\int_{\mathbb{R}} K^2\left(u\right)du = \frac{m(t)f(t)}{1-G(t)}\int_{\mathbb{R}} K^2\left(u\right)du. \qquad \square$$

Chapter 5

Simulation study

In Chapter 4 it is shown that under the SRCM $f_{1,n}^{SE}$ and $f_{2,n}^{SE}$ admit a smaller asymptotic variance when compared to f_n^{KM}. In this chapter we are going to perform a simulation study in order to demonstrate the improvement regarding the asymptotic variance and to investigate the bias admitted by these estimators.

Recall the SRCM from Definition 2.5. For the simulations we choose the lifetime X to be Weibull distributed and randomly right censored by an independent Weibull variable Y, that is

$$X \sim F \hat{=} \text{Weibull}(\alpha_1, \beta_1) \qquad \text{and} \quad Y \sim G \hat{=} \text{Weibull}(\alpha_2, \beta_2)$$

where f and g denote the p.d.f.s of X and Y, respectively. The p.d.f. of a Weibull(α,β) distribution is given by $f(t) = \alpha\beta(\alpha t)^{\beta-1}\exp(-(\alpha t)^{\beta})$ for $t \geq 0$. As explained in Dikta (1998, Example 2.9.) and Harlaß (2011, Section 3.1.1.), this results in a two-parameter model for the conditional probability m, in particular

$$m(t, \theta) = \frac{\theta_1}{\theta_1 + t^{\theta_2}}, \qquad \theta_1 > 0,\, \theta_2 \in \mathbb{R} \tag{5.1}$$

with $\theta = (\theta_1, \theta_2)^{\intercal}$, $\theta_1 = (\alpha_1^{\beta^1}\beta_1)/(\alpha_2^{\beta^2}\beta_2)$ and $\theta_2 = \beta_2 - \beta_1$. Note that this setup describes a generalized proportional hazard model and that $\tau_H = \infty$.

As a first step, one single dataset of the form $(Z_i, \delta_i)_{1\leq i\leq n}$ was generated according to the SRCM with a parametric model as given in (5.1). We choose the parameters $\alpha_1 = 1$ and $\beta_1 = 5$ for the lifetime distribution F and $\alpha_2 = 1.7$ and $\beta_2 = 1$ for the censoring distribution G which causes approximately 80% of the observations to be censored. Based on this dataset the estimators f_n^{KM}, $f_{1,n}^{SE}$ and $f_{2,n}^{SE}$ were used to estimate the true p.d.f. f. The bandwidth was determined by the rule of thumb given in (4.6). A plot of the estimating curves is given in Figure 5.1.

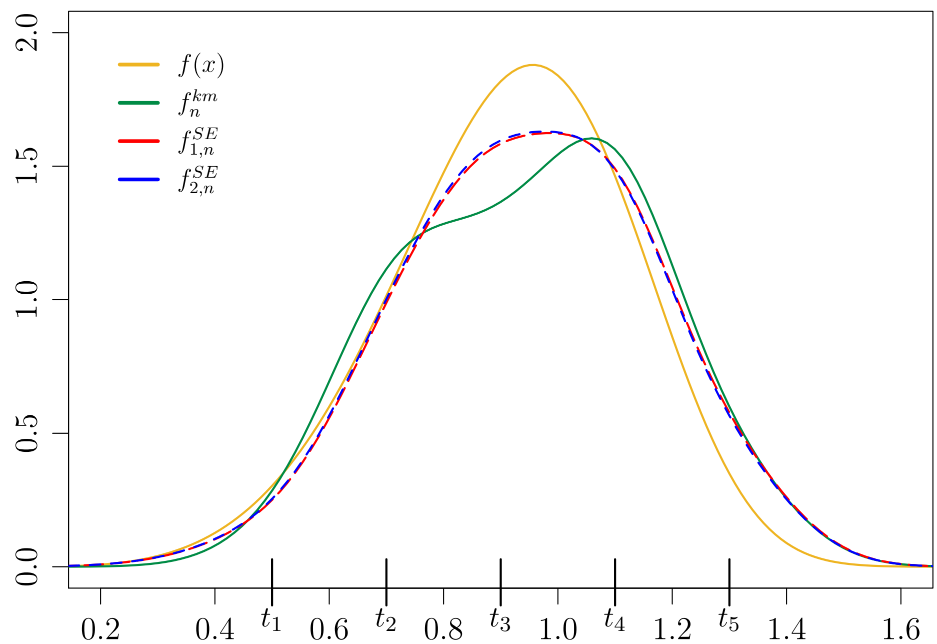

Figure 5.1: Comparison of f_n^{KM}, $f_{1,n}^{SE}$ and $f_{2,n}^{SE}$ based on a single Weibull-Weibull dataset

Table 5.1 lists the bias and the squared bias for each of the three estimators at the node points $(t_i)_{i=1,\ldots,5}$. That is

$$\mathrm{Bias}_{SE2}(t_i) := f(t_i) - f_{2,n}^{SE}(t_i) \qquad \text{and} \qquad \mathrm{Bias}^2_{SE2}(t_i) := \left(f(t_i) - f_{2,n}^{SE}(t_i)\right)^2$$

and similarly for $f_{1,n}^{SE}$ and f_n^{KM}. The additional column labeled with $\overline{\mathrm{Bias}^2}$ shows the averaged squared bias taken over the five node points which in case of $f_{2,n}^{SE}$ is

$$\frac{1}{5}\sum_{i=1}^{5}\left(f(t_i) - f_{2,n}^{SE}(t_i)\right)^2.$$

The data in Table 5.1 shows that the semi-parametric estimations lead to a smaller deviation from the true value when compared to the Kaplan-Meier estimator. Except in one node point, the squared bias of the semi-parametric estimators is smaller than the ones of the Kaplan-Meier approximation. In addition, the results suggest that the new semi-parametric estimator $f_{2,n}^{SE}$ leads to a smaller bias in comparison to $f_{1,n}^{SE}$.

	$t_1 = 0.5$	$t_2 = 0.7$	$t_3 = 0.9$	$t_4 = 1.1$	$t_5 = 1.3$	
$\texttt{Bias}_{KM}$	-0.0186	+0.1000	-0.4528	+0.0994	+0.2498	
$\texttt{Bias}_{SE1}$	-0.0531	-0.0268	-0.2346	+0.0266	+0.2272	
$\texttt{Bias}_{SE2}$	-0.0493	-0.0131	-0.2224	+0.0227	+0.2159	
	$t_1 = 0.5$	$t_2 = 0.7$	$t_3 = 0.9$	$t_4 = 1.1$	$t_5 = 1.3$	$\overline{\text{Bias}^2}$
$\texttt{Bias}^2_{KM}$	0.0004	0.0100	0.2050	0.0099	0.0624	0.011426
$\texttt{Bias}^2_{SE1}$	0.0028	0.0007	0.0550	0.0007	0.0516	0.007793
$\texttt{Bias}^2_{SE2}$	0.0024	0.0001	0.0495	0.0005	0.0466	0.007772

Table 5.1: Bias and squared bias of the estimators at particular node points

To further investigate the behavior of the estimators we generate $k = 100$ datasets

$$D_j = (Z_i, \delta_i)_{1 \le i \le n}, \qquad j = 1, \dots, k$$

of sample size $n = 40$. We again use the same Weibull-Weibull model as described in (5.1) but with the parameters $\alpha_1 = 4$, $\beta_1 = 0.6$, $\alpha_2 = 2$, $\beta_2 = 2$. This leads to roughly 30% censored observations. Similarly as before we apply the estimators f_n^{KM}, $f_{1,n}^{SE}$ and $f_{2,n}^{SE}$ to each of the datasets in order to approximate the true p.d.f. f at 50 equally spaced node points $0.01 < t_1 < \cdots < t_{50} < 0.5$. To indicate the dependence of the estimate $f_{2,n}^{SE}(t)$ upon a particular dataset D we write

$$f_{2,n}^{SE}(t) \equiv f_{2,n}^{SE}(D, t),$$

and similarly for f_n^{KM} and $f_{1,n}^{SE}$.

Based on the $k = 100$ approximations of $f(t_i)$ the average, the MSE and the sample variance are calculated for each of the estimates $f_n^{KM}(t_i)$, $f_{1,n}^{SE}(t_i)$ and $f_{2,n}^{SE}(t_i)$ for $i = 1, \ldots, 50$. That is

$$\begin{aligned}
\mathrm{AVG}_{SE2}(t_i) &:= \frac{1}{k}\sum_{j=1}^{k} f_{2,n}^{SE}(D_j, t_i), \\
\mathrm{MSE}_{SE2}(t_i) &:= \frac{1}{k}\sum_{j=1}^{k} \left(f_{2,n}^{SE}(D_j, t_i) - f(t_i)\right)^2, \\
\mathrm{VAR}_{SE2}(t_i) &:= \frac{1}{k}\sum_{j=1}^{k} \left(f_{2,n}^{SE}(D_j, t_i) - AVE_{SE2}(t_i)\right)^2,
\end{aligned}$$

and similarly for $f_{1,n}^{SE}$ and f_n^{KM}. The results for five particular note points are shown in Table 5.2. In order to obtain some global measures we calculate the average of the MSEs and the variances over all 50 node points, in particular

$$\overline{\mathrm{MSE}}_{SE2} = \frac{1}{50}\sum_{i=1}^{50} \mathrm{MSE}_{SE2}(t_i) \qquad \text{and} \qquad \overline{\mathrm{Var}}_{SE2} = \frac{1}{50}\sum_{i=1}^{50} \mathrm{Var}_{SE2}(t_i),$$

and similarly for $f_{1,n}^{SE}$ and f_n^{KM}. These averages are given in Table 5.3.

	$t_{13} = 0.126$	$t_{21} = 0.202$	$t_{27} = 0.260$	$t_{35} = 0.336$	$t_{43} = 0.412$
$f(t_i)$	1.6264	1.0824	0.8500	0.6462	0.5094
AVG_{KM}	1.9840	1.2693	0.9504	0.6470	0.4498
AVG_{SE1}	1.1656	0.6864	0.7606	0.9621	1.1656
AVG_{SE2}	1.9963	1.2671	0.9487	0.7100	0.5875
MSE_{KM}	2.2073	1.8997	2.0132	2.2701	2.5874
MSE_{SE1}	2.1509	1.8259	1.9495	2.1878	2.4186
MSE_{SE2}	2.1737	1.8229	1.9254	2.1313	2.3191
VAR_{KM}	1.2928	0.7992	0.8479	1.0433	1.3226
VAR_{SE1}	1.1655	0.6864	0.7606	0.9621	1.1655
VAR_{SE2}	1.1071	0.6569	0.7371	0.9294	1.1065

Table 5.2: MSE and variance of the estimators based on $k = 100$ datasets

From the results presented in Chapter 4 one might suggest that the semi-parametric estimators produce smaller variances also for finite sample sizes. The data shown in Table 5.2 confirms this conjecture: At all 50 node points the estimators $f_{1,n}^{SE}$ and $f_{2,n}^{SE}$ attained a smaller sample variance than f_n^{KM}. Moreover $f_{2,n}^{SE}$ produces smaller variances than $f_{1,n}^{SE}$ at all node points t_i. Both facts are also reflected by the averaged variances, in particular Table 5.3 shows that

$$\overline{\text{VAR}}_{SE2} < \overline{\text{VAR}}_{SE1} < \overline{\text{VAR}}_{KM}.$$

This suggest that for a fixed sample size $f_{2,n}^{SE}$ provides approximations with a smaller variance when compared to $f_{1,n}^{SE}$.

$\overline{\text{MSE}}_{KM}$	$\overline{\text{MSE}}_{SE1}$	$\overline{\text{MSE}}_{SE2}$
2.4387	2.3510	2.3180

$\overline{\text{Var}}_{KM}$	$\overline{\text{Var}}_{SE1}$	$\overline{\text{Var}}_{se2}$
1.372486	1.256572	1.200321

Table 5.3: Average of the MSE and the variance taken over all node points

Furthermore Table 5.2 shows that the behavior of the bias is very similar to the one of the variances. The Kaplan-Meier estimator results in larger MSEs at all node points when compared to the semi-parametric counterparts. In addition, the semi-parametric estimator $f_{2,n}^{SE}$ gives even smaller MSEs than the estimator $f_{1,n}^{SE}$, except for some of the smaller node points. The averaged MSEs given in Table 5.3 show that $f_{2,n}^{SE}$ produces overall a smaller bias in comparison to $f_{1,n}^{SE}$ for a fixed sample size:

$$\overline{\text{MSE}}_{SE2} < \overline{\text{MSE}}_{SE1} < \overline{\text{MSE}}_{KM}.$$

This corresponds with the insights drawn from Table 5.1.

Chapter 6

Conclusion

Product limit estimators of the survival time can be derived as the solution of identifying integral equations. The widely used Kaplan Meier PLE F_n^{KM} and also its semi-parametric and presmoothed extensions $F_{1,n}^{SE}$ and $F_{1,n}^{PR}$ are in general only sub-distribution functions. In comparison, the proposed semi-parametric estimator

$$1 - F_{2,n}^{SE}(t) := \prod_{i:Z_i \leq t} \left[1 - \frac{m(Z_i, \theta_n)}{n - R_n(Z_i) + m(Z_i, \theta_n)} \right]$$

is a true distribution function and therefore should perform better w.r.t. the bias especially in the case of small sample sizes. In addition, it is possible to directly sample according to $F_{2,n}^{SE}$ which is particularly useful for the construction of confidence bands of the underlying survival function. Theorem 3.13 and Theorem 3.16 show that $F_{1,n}^{SE}$ and $F_{2,n}^{SE}$ are asymptotically equivalent, i.e., for some Borel-measurable function φ, $\int_0^\infty \varphi dF_{2,n}^{SE}$ is a strong consistent estimator of the linear functional $\int_0^{\tau_H} \varphi dF$ where $\tau_H = \inf\{x : H(x) = 1\}$. Furthermore $\int_0^{\tau_H} \varphi dF$ admits the same asymptotic variance as the corresponding functional w.r.t. $F_{1,n}^{SE}$. This attained variance is optimal w.r.t. to the class of all regular estimators of $\int_0^{\tau_H} \varphi dF$ and therefore $F_{2,n}^{SE}$ outperforms its Kaplan-Meier and presmoothed competitors. A more detailed discussion can be found in Subsection 3.2.3.

Relying on $F_{2,n}^{SE}$, it is possible to extend the usual kernel density estimator to the semi-parametric random censorship model. Key to the analysis of the resulting estimator

$$f_{2,n}^{SE}(t) := \frac{1}{a_n} \int_{\mathbb{R}} K\left(\frac{t-x}{a_n}\right) F_{2,n}^{SE}(dx)$$

are the asymptotic representations derived in Theorem 4.6 and Theorem 4.7 which reveal the enhancement of $f_{2,n}^{SE}$ in comparison to the Kaplan-Meier kernel estimator f_n^{KM}. For example, the asymptotic variance introduced by $f_{2,n}^{SE}$ is in almost all scenarios strictly smaller than the one of f_n^{KM} but is at most equal. Further results drawn from those asymptotic representations are improved pointwise and uniform convergence rates of $f_{2,n}^{SE}$ when compared with f_n^{KM}.

The simulation study has shown that, for a fixed sample size, the semi-parametric estimators $f_{1,n}^{SE}$ and $f_{2,n}^{SE}$ produce smaller variances when compared to the Kaplan-Meier estimator f_n^{KM}. In addition, although $f_{1,n}^{SE}$ and $f_{2,n}^{SE}$ are asymptotically identical, the simulation indicates that, for a fixed sample size, $f_{2,n}^{SE}$ results in smaller variances in comparison to $f_{1,n}^{SE}$. The admitted bias shows a similar behavior: For a fixed sample size, $f_{2,n}^{SE}$ causes a smaller bias than $f_{1,n}^{SE}$ while the bias of both semi-parametric estimators is smaller then the one induced by f_n^{KM}.

For both, $F_{2,n}^{SE}$ and $f_{2,n}^{SE}$, the bias reduction and the gain in efficiency was shown under the assumption of a correctly chosen parametric model for m. There are bootstrap based goodness-of-fit tests available in order to validate the model assumptions. Simulation studies have shown that $F_{1,n}^{SE}$ performs well even under incorrect model assumptions. It is conceivable that $F_{2,n}^{SE}$ and $f_{2,n}^{SE}$ behave similarly.

Bibliography

Aalen, O. (1978), "Nonparametric Inference for a Family of Counting Processes," *The Annals of Statistics*, 6, 701–726.

Abdushukurov, A. A. (1987), "Nonparametric estimation in the proportional hazards model of random censorship," *Akad. Nauk. UzSSR Tashkent*, VINITI No. 3448–V (in Russian).

Aiken, R. C. (1985), *Stiff computation*, New York: Oxford University Press, Inc.

Armitage, P., Berry, G., and Matthews, J. N. S. (2001), *Statistical Methods in Medical Research*, Wiley-Blackwell, 4th ed.

Ascher, U. M. and Petzold, L. R. (1998), *Computer Method for Ordinary Differential Equations and Differential Algebraic Equations*, SIAM.

Bertrand-Retali, M. (1978), "Convergence uniforme d'un estimateur de la densité par la méthode du noyau," *Rev. Roumanine de Math. Pures Appl.*, 23, 361–385.

Bickel, P. J. and Rosenblatt, M. (1973), "On some global measures of the derivations of density function estimates," *The Annals of Statistics*, 1, 1071–1095.

Blum, J. R. and Susarla, V. (1980), "Maximal deviation theory of density and failure rate function estimates based on censored data," in *Proceedings of the 5th International Symposium on Multivariate Analysis*, vol. V, pp. 213–222.

Breslow, N. and Crowley, J. (1974), "A Large Sample Study of the Life Table and Product Limit Estimates Under Random Censorship," *The Annals of Statistics*, 2, 437–453.

Cao, R. and Jácome, M. (2004), "Presmoothed kernel density estimator for censored data," *Journal of Nonparametric Statistics*, 16, 289–309.

Cao, R., López-de Ullibarri, I., Janssen, P., and Veraverbeke, N. (2005), "Presmoothed Kaplan–Meier and Nelson–Aalen estimators," *Journal of Nonparametric Statistics*, 17, 31–56.

Carbonez, A. and Györfi, L. (1992), "L_1-consistency of randomly censored version of histogram estimate," *L_1-Statistical Analysis and Related Methods*, 389–400.

Cheng, P. E. and Lin, G. D. (1987), "Maximum likelihood estimation of a survival function under the Koziol-Green proportional hazards model," *Statistics & Probability Letters*, 5, 75–80.

Chiu, S.-T. (1996), "A comparative review of bandwidth selection for kernel density estimation," *Statistica Sinica*, 6, 129–145.

Cohn, D. L. (2013), *Measure Theory: Second Edition*, Basel: Birkhäuser, 2nd ed.

Collett, D. (2002), *Modelling Binary Data*, Chapman and Hall/CRC, 2nd ed.

Cox, D. R. and Snell, E. J. (1989), *Analysis of Binary Data*, Chapman and Hall/CRC, 2nd ed.

Csörgő, M., Gombay, E., and Horvath, L. (1991), "Central Limit Theorems for L_p Distances of Kernel Estimators of Densities Under Random Censorship," *The Annals of Statistics*, 19, 1813–1831.

Csörgő, S. (1988), "Estimation in the proportional hazards model of random censorship," *Statistics*, 19, 437–463.

Deuflhard, P. and Bornemann, F. (2002), *Scientific Computing with Ordinary Differential Equations*, New York: Springer.

Diallo, A. O. K. and Louani, D. (2013), "Moderate and large deviation principles for the hazard rate function kernel estimator under censoring," *Statistics & Probability Letters*, 83, 735–743.

Diehl, S. and Stute, W. (1988), "Kernel density and hazard function estimation in the presence of censoring," *Journal of Multivariate Analysis*, 25, 299–310.

Dikta, G. (1995), "Asymptotic normality under the Koziol-Green model," *Communications in Statistics - Theory and Methods*, 24, 1537–1549.

— (1998), "On semiparametric random censorship models," *Journal of Statistical Planning and Inference*, 66, 253–279.

— (2000), "The Strong Law under Semiparametric Random Censorship Models," *Journal of Statistical Planning and Inference*, 83, 1–10.

— (2014), "Asymptotically efficient estimation under semi-parametric random censorship models," *Journal of Multivariate Analysis*, Vol. 124, 10–24.

Dikta, G., Ghorai, J., and Schmidt, C. (2005), "The Central Limit Theorem under Semiparametric Random Censorship Models," *Journal of Statistical Planning and Inference*, 127, 23–51.

Dikta, G., Hausmann, R., and Schmidt, C. (2002), "Some Simulation Results under Random Censorship Models," *InterStat., March*, 1–12.

Dikta, G., Külheim, R., Mendonça, J., and de Uña-Álvarez, J. (2016a), "Asymptotic representation of presmoothed Kaplan-Meier integrals with covariates in a semiparametric censorship model," *Journal of Statistical Planning and Inference*, 171, 10–37.

Dikta, G., Kvesic, M., and Schmidt, C. (2006), "Bootstrap Approximations in Model Checks for Binary Data," *Journal of the American Statistical Association*, 101, 521–530.

Dikta, G., Reißel, M., and Harlaß, C. (2016b), "Semi-parametric survival function estimators deduced from an identifying Volterra type integral equation," *Journal of Multivariate Analysis*, 147, 273–284.

Dinwoodie, I. H. (1993), "Large Deviations for Censored Data," *The Annals of Statistics*, 21, 1608–1620.

Dvoretzky, A., Kiefer, J., and Wolfowitz, J. (1956), "Asymptotic Minimax Character of the Sample Distribution Function and of the Classical Multinomial Estimator," *The Annals of Mathematical Statistics*, 27, 642–669.

Efron, B. (1967), "The Two Sample Problem with Censored Data," in *Proceedings of the Fifth Berkley Symposium*, Berkeley: University of California Press, vol. IV, pp. 831–853.

Einmahl, U. and Mason, D. M. (2000), "An empirical process approach to the uniform consistency of kernel-type function estimators," *Journal of Theoretical Probability*, 13, 1–37.

Epanechnikov, V. A. (1969), "Non-Parametric Estimation of a Multivariate Probability Density," *Theory of Probability & Its Applications*, 14, 153–158.

Földes, A., Rejtő, L., and Winter, B. B. (1981), "Strong consistency properties of nonparametric estimators for randomly censored data, II: Estimation of Density and Failure Rate," *Periodica Mathematica Hungarica*, 12, 15–29.

Ghorai, J. K. and Pattanaik, L. M. (1990), "L_1-Consistency of the kernel density estimators based on randomly right censored data," *Communications in Statistics - Theory and Methods*, 19, 2853–2870.

— (1991), "A central limit theorem for the weighted integrated squared error of the kernel type density estimator under the proportional hazard model," *Journal of Nonparametric Statistics*, 1, 111–126.

Ghosh, M., Parr, W. C., Singh, K., and Babu, G. J. (1984), "A Note on Bootstrapping the Sample Median," *The Annals of Statistics*, 12, 1130–1135.

Gill, R. D. (1983), "Large Sample Behaviour of the Product-Limit Estimator on the Whole Line," *The Annals of Statistics*, 11, 49–58.

Gill, R. D. and Johansen, S. (1990), "A Survey of Product-Integration with a View Toward Application in Survival Analysis," *The Annals of Statistics*, 18, 1501–1555.

Giné, E. and Guillou, A. (2001), "On consistency of kernel density estimators for randomly censored data: Rates holding uniformly over adaptive intervals," *Annales de l'institut Henri Poincare (B) Probability and Statistics*, 37, 503–522.

Hairer, E. and Wanner, G. (2010), *Solving Ordinary Differential Equations II: Stiff and Differential-Algebraic Problems*, Berlin, Heidelberg: Springer, 2nd ed.

Hall, P. (1981), "Laws of the iterated logarithm for nonparametric density estimators," *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 56, 47–61.

Härdle, W. (1991), *Smoothing Techniques*, New York: Springer.

Harlaß, C. (2011), "Dichteschätzung der Lebensdauerverteilung unter semiparametrisch zufällig rechtszensierten Daten," Master's thesis, FH Aachen Univerity of Applied Sciences.

Hartman, P. (2002), *Ordinary Differential Equations*, SIAM, 2nd ed.

Heidenreich, N. B., Schindler, A., and Sperlich, S. (2013), "Bandwidth selection for kernel density estimation: A review of fully automatic selectors," *AStA Advances in Statistical Analysis*, 97, 403–433.

Hewitt, E. (1960), "Integration by Parts for Stieltjes Integrals," *The American Mathematical Monthly*, 67, 419–423.

Hewitt, E. and Stromberg, K. (1965), *Real and Abstract Analysis - A Modern Treatment of the Theory of Functions of a Real Variable*, New York: Springer.

Jácome, M. A. and Cao, R. (2007), "Almost sure asymptotic representation for the presmoothed distribution and density estimators for censored data," *Statistics: A Journal of Theoretical and Applied Statistics*, 41, 517–534.

— (2008), "Asymptotic-based bandwidth selection for the presmoothed density estimator with censored data," *Journal of Nonparametric Statistics*, 20, 483–506.

Jácome, M. A., Gijbels, I., and Cao, R. (2008), "Comparison of presmoothing methods in kernel density estimation under censoring," *Computational Statistics*, 23, 381–406.

Kalbfleisch, J. D. and Prentice, R. L. (2002), *The Statistical Analysis of Failure Time Data*, New Jersey: John Wiley & Sons, 2nd ed.

Kaplan, E. L. and Meier, P. (1958), "Nonparametric estimation from incomplete samples," *Journal of the American Statistical Association*, 73, 457–481.

Klein, J. P. and Moeschberger, M. L. (2003), *Survival Analysis - Techniques for Censored and Truncated Data*, New York: Springer, 2nd ed.

Klein, J. P., van Houwelingen, H. C., Ibrahim, J. G., and Scheike, T. H. (2013), *Handbook of Survival Analysis*, Chapman & Hall/CRC.

Kleinbaum, D. G. and Klein, M. (2012), *Survival Analysis: A Self-Learning Text*, New York: Springer, 3rd ed.

Koziol, J. A. and Green, S. B. (1976), "A Cramér-von Mises statistic for randomly censored data," *Biometrika*, 63, 465–474.

Kulasekera, K. B. (1995), "A bound on the L_1-error of a nonparametric density estimator with censored data," *Statistics & Probability Letters*, 23, 233–238.

Lemdani, M. and Ould-Saïd, E. (2002), "Exact asymptotic L_1-error of a kernel density estimator under censored data," *Statistics & Probability Letters*, 60, 59–68.

Loeve, M. (1977), *Probability Theory I*, New York: Springer, 4th ed.

Marron, J. S. and Padgett, W. J. (1987), "Asymptotically Optimal Bandwidth Selection for Kernel Density Estimators from Randomly Right-Censored Samples," *The Annals of Statistics*, 15, 1520–1535.

McNichols, D. T. and Padgett, W. (1981), "Kernel density estimation under random censorship," Tech. rep., Universitiy of South Carolina.

— (1984), "A modified kernel density estimator for randomly right-censored data," *South African Statistical Journal*, 18, 13–27.

Mielniczuk, J. (1986), "Some Asymptotic Properties of Kernel Estimators of a Density Function in Case of Censored Data," *The Annals of Statistics*, 14, 766–773.

Nadaraya, É. A. (1964), "On Estimating Regression," *Theory of Probability & Its Applications*, 9, 141–142.

— (1965), "On Non-Parametric Estimates of Density Functions and Regression Curves," *Theory of Probability & Its Applications*, 10, 186–190.

Nelson, W. (1972), "Theory and Applications of Hazard Plotting for Censored Failure Data," *Technometrics*, 14, 945–966.

Parzen, E. (1962), "On Estimation of a Probability Density Function and Mode," *The Annals of Mathematical Statistics*, 33, 1065–1076.

Ramlau-Hansen, H. (1983), "Smoothing Counting Process Intensities by Means of Kernel Functions," *The Annals of Statistics*, 11, 453–466.

Robbins, H. and Siegmund, D. (1972), "On the law of the iterated logarithm for maxima and minima," in *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability*, Berkeley: University of California Press, vol. III, pp. 51–70.

Rosenblatt, M. (1956), "Remarks on Some Nonparametric Estimates of a Density Function," *The Annals of Mathematical Statistics*, 27, 832–837.

Schick, A., Susarla, V., and Koul, H. (1988), "Efficient estimation of functionals with censored data," *Statistics & Risk Modeling*, 6, 349–360.

Scott, D. W. (1992), *Multivariate Density Estimation: Theory, Practice, and Visualization*, John Wiley & Sons.

— (2015), *Multivariate Density Estimation: Theory, Practice, and Visualization*, New Jersey: John Wiley & Sons, 2nd ed.

Serfling, R. J. (2001), *Approximation Theorems of Mathematical Statistics*, New Jersey: John Wiley & Sons, 1st ed.

Shorack, G. R. and Wellner, J. A. (1986), *Empirical processes with applications to statistics*, New York: John Wiley & Sons.

Silverman, B. W. (1978), "Choosing the Window Width when Estimating a Density," *Biometrika*, 65, 1–11.

— (1986), *Density Estimation for Statistics and Data Analysis*, Chapman and Hall/CRC.

Stute, W. (1982), "A Law of the Logarithm for Kernel Density Estimators," *The Annals of Probability*, 10, 414–422.

— (1992), "Strong consistency under the Koziol-Green model," *Statistics & Probability Letters*, 14, 313–320.

— (1993), "Almost Sure Representations of the Product-Limit Estimator for Truncated Data," *The Annals of Statistics*, 21, 146–156.

— (1995), "The central limit theorem under random censorship," *The Annals of Statistics*, 23, 422–439.

Stute, W. and Wang, J. (1993), "The strong law under random censorship," *The Annals of Statistics*, 21, 1591–1607.

Subramanian, S. and Zhang, P. (2013), "Model-based confidence bands for survival functions," *Journal of Statistical Planning and Inference*, 143, 1166–1185.

Volterra, V. (1887), "Sulle equazioni differenziali lineari," *Rend. Accademica dei Lincei*, 4, 393–396.

Wand, M. P. and Jones, M. C. (1994), *Kernel Smoothing*, Chapman and Hall/CRC.

Watson, G. S. (1964), "Smooth Regression Analysis," *Sankhyā: The Indian Journal of Statistics, Series A*, 26, 359–372.

Wellner, J. A. (1982), "Asymptotic Optimality of the Product Limit Estimator," *The Annals of Statistics*, 10, 595–602.

Witting, H. (1985), *Mathematische Statistik I Parametrische Verfahren bei festem Stichprobenumfang*, Stuttgart: Springer Vieweg + Teubner.

Witting, H. and Müller-Funk, U. (1995), *Mathematische Statistik II Asymptotische Statistik: Parametrische Modelle und nichtparametrische Funktionale*, Stuttgart: Springer Vieweg + Teubner.

Zhang, B. (1996), "Some asymptotic results for kernel density estimation under random censorship," *Bernoulli*, 2, 183–198.

— (1998), "A Note on the Strong Uniform Consistency of Kernel Dinsity Estimatiors under Random Censorship," *Sankhyā: The Indian Journal of Statistics, Series A*, 60, 265–273.

Ziegler, S. (1995), "Ein modifizierter Kaplan–Meier Schätzer," Diploma thesis, University of Giessen.

Appendix A

Convergence rate of the MLE

In the following we will give the proof for Lemma 4.25 and therefore derive an a.s. convergence rate of the MLE defined in Definition 3.7.

Proof of Lemma 4.25.

Recall the log-likelihood function for θ, with w_1, w_2 and w as defined in (A5) and (3.11)

$$l_n(\theta) = \frac{1}{n}\sum_{i=1}^{n} w(\delta_i, Z_i, \theta),$$

in particular $w(\delta, Z, \theta) = \delta w_1(Z, \theta) + (1-\delta) w_2(Z, \theta)$ with $w_1(z, \theta) = \ln(m(z, \theta))$ and $w_2(z, \theta) = \ln(1 - m(z, \theta))$. Furthermore let $\theta = (\theta_1, \ldots, \theta_k)$ and define

$$\nabla_r m(z, \theta_0) \coloneqq D_r m(z, \theta_0) = [\partial/\partial\theta_r m(z, \theta)]|_{\theta=\theta_0}$$

and $\nabla m(z, \theta_0) = \mathrm{Grad}(m(z, \theta_0)) = (D_1 m(z, \theta_0), \ldots, D_k m(z, \theta_0))^\intercal$. Moreover let

$$\mathrm{J}(\nabla m(z, \theta_0)) \coloneqq \begin{bmatrix} \nabla^\intercal \nabla_1 m(z, \theta_0) \\ \vdots \\ \nabla^\intercal \nabla_k m(z, \theta_0) \end{bmatrix} = \begin{bmatrix} D_{1,1} m(z, \theta_0) & \cdots & D_{k,1} m(z, \theta_0) \\ \vdots & \ddots & \vdots \\ D_{1,k} m(z, \theta_0) & \cdots & D_{k,k} m(z, \theta_0) \end{bmatrix}$$

be the Jacobian matrix of $\nabla m(z, \theta_0)$.

Following the reasoning of Witting and Müller-Funk (1995, Theorem 6.35), the expansion of $\mathrm{Grad}(l_n(\theta_n))$ at θ_0 yields

$$\begin{aligned}\mathrm{Grad}(l_n(\theta_n)) &= \mathrm{Grad}(l_n(\theta_0)) + \mathrm{J}(l_n(\tilde{\theta}_n))(\theta_n - \theta_0) \\ &= U_n(\theta_0) + \Big[T_n(\theta_0) + R_n(\theta_0, \theta_n, \tilde{\theta}_n)\Big](\theta_n - \theta_0),\end{aligned} \tag{B.1}$$

where $\tilde{\theta}_n \in \Theta$ lies in the interior of the line segment connecting θ_n and θ_0, and $U_n(\theta_0)$ is a vector with elements

$$U_{n,r}(\theta_0) := \frac{1}{n}\sum_{i=1}^{n} D_r w(\delta_i, Z_i, \theta_0), \quad \text{for all } r = 1, \dots, k.$$

$T_n(\theta_0)$ and $R_n(\theta_0, \theta_n, \tilde{\theta}_n)$ are matrices with elements

$$\begin{aligned} T_{n,r,s}(\theta_0) &:= \frac{1}{n}\sum_{i=1}^{n} D_{r,s} w(\delta_i, Z_i, \theta_0), \\ R_{n,r,s}(\theta_0, \theta_n, \tilde{\theta}_n) &:= \frac{1}{n}\sum_{i=1}^{n} D_{r,s} w(\delta_i, Z_i, \tilde{\theta}_n) - D_{r,s} w(\delta_i, Z_i, \theta_0), \quad \text{for all } r, s = 1, \dots, k. \end{aligned}$$

Now consider U_n. Since $D_r w(\delta_i, Z_i, \theta_0)$ are i.i.d. for all $i = 1, \dots, n$ it follows by SLLN

$$U_{n,r}(\theta_0) \xrightarrow[n\to\infty]{a.s.} \mathbb{E}[D_r w(\delta, Z, \theta_0)] = 0 \qquad \forall\, 1 \le r \le k \tag{B.2}$$

where, since $m(Z, \theta_0) = \mathbb{E}\left[\delta | Z\right]$,

$$\begin{aligned} \mathbb{E}[D_r w(\delta, Z, \theta_0)] &= \mathbb{E}\left[D_r\{\delta \ln(m(Z, \theta_0)) + (1-\delta)\ln(1 - m(Z, \theta_0))\}\right] \\ &= \mathbb{E}\left[\frac{\delta D_r m(Z, \theta_0)}{m(Z, \theta_0)} - \frac{(1-\delta) D_r m(Z, \theta_0)}{1 - m(Z, \theta_0)}\right] \\ &= \mathbb{E}\left[\frac{D_r m(Z, \theta_0)(\delta - m(Z, \theta_0))}{m(Z, \theta_0)(1 - m(Z, \theta_0))}\right] \\ &= \mathbb{E}\left[\mathbb{E}\left[\left.\frac{D_r m(Z, \theta_0)(\delta - m(Z, \theta_0))}{m(Z, \theta_0)(1 - m(Z, \theta_0))}\right| Z\right]\right] \\ &= \mathbb{E}\left[\frac{D_r m(Z, \theta_0)}{m(Z, \theta_0)(1 - m(Z, \theta_0))}\Big(\mathbb{E}\left[\delta | Z\right] - m(Z, \theta_0)\Big)\right] = 0. \end{aligned}$$

Furthermore, note that

$$
\begin{aligned}
&\mathbb{E}\left[D_r w(\delta, Z, \theta_0) D_s w(\delta, Z, \theta_0)\right] \\
&\quad = \mathbb{E}\left[\{\delta D_r \ln(m(Z,\theta_0)) + (1-\delta) D_r \ln(1-m(Z,\theta_0))\}\right. \\
&\qquad\quad \left. \times \{\delta D_s \ln(m(Z,\theta_0)) + (1-\delta) D_s \ln(1-m(Z,\theta_0))\}\right] \\
&\quad = \mathbb{E}\left[\delta^2 D_r \ln(m(Z,\theta_0)) D_s \ln(m(Z,\theta_0))\right. \\
&\qquad\quad \left. + (1-\delta)^2 D_r \ln(1-m(Z,\theta_0)) D_s \ln(1-m(Z,\theta_0))\right] \\
&\quad = \mathbb{E}\left[\delta \frac{D_r m(Z,\theta_0) D_s m(Z,\theta_0)}{m^2(Z,\theta_0)} + (1-\delta) \frac{D_r m(Z,\theta_0) D_s m(Z,\theta_0)}{(1-m(Z,\theta_0))^2}\right] \\
&\quad = \mathbb{E}\left[\frac{D_r m(Z,\theta_0) D_s m(Z,\theta_0)}{m^2(Z,\theta_0)} \mathbb{E}\left[\delta|Z\right] + \frac{D_r m(Z,\theta_0) D_s m(Z,\theta_0)}{(1-m(Z,\theta_0))^2} \mathbb{E}\left[1-\delta|Z\right]\right] \\
&\quad = \mathbb{E}\left[\frac{D_r m(Z,\theta_0) D_s m(Z,\theta_0)}{m(Z,\theta_0)} + \frac{D_r m(Z,\theta_0) D_s m(Z,\theta_0)}{(1-m(Z,\theta_0))}\right] \\
&\quad = \mathbb{E}\left[\frac{D_r m(Z,\theta_0) D_s m(Z,\theta_0)}{m(Z,\theta_0)(1-m(Z,\theta_0))}\right] = \sigma_{r,s}.
\end{aligned}
$$

Now consider T_n. Since $D_{r,s} w(\delta_i, Z_i, \theta_0)$ are i.i.d. for all $i = 1, \dots, n$ it follows by SLLN

$$T_{n,r,s}(\theta_0) \xrightarrow[n\to\infty]{a.s.} \mathbb{E}\left[D_{r,s} w(\delta, Z, \theta_0)\right] = -\sigma_{r,s}.$$

Due to (A4) an (A5), $I(\theta_0)$ is finite and positive definite. Hence we have

$$T_n(\theta_0) \xrightarrow[n\to\infty]{a.s.} -I(\theta_0). \tag{B.3}$$

In the following we will show that $R_n(\theta_0, \theta_n, \tilde{\theta}_n) \xrightarrow{a.s.} 0$ as $n \to \infty$. Therefor define

$$A_{r,s}(\delta, z, \gamma) \coloneqq \sup_{\theta' \in V(\theta_0, \gamma)} \left| D_{r,s} w(\delta, z, \theta') - D_{r,s} w(\delta_i, z_i, \theta_0) \right|.$$

By (A4) and Witting (1985, A3.6) $A_{r,s}$ is measurable. Assumption (A4) also yields that $a_{r,s}(\gamma) := \mathbb{E}\left[A_{r,s}(\delta, z, \gamma)\right] < \infty$ for all $0 < \gamma \leq \gamma_{\theta_0}$ and that $A_{r,s}(\delta, z, \gamma) \to 0$ as $\gamma \to 0$. Hence $a_{r,s}(\gamma) \to 0$ as $\gamma \to 0$ by Lebesgue's Dominated Convergence Theorem. For $\epsilon > 0$ consider

$$\begin{aligned}
&\lim_{N\to\infty} \mathbb{P}\left(\sup_{n\geq N}\left|R_{n,r,s}(\theta_0,\theta_n,\tilde{\theta}_n)\right| > \epsilon\right) \\
&\quad\leq \lim_{N\to\infty} \mathbb{P}\left(\sup_{n\geq N}\left|R_{n,r,s}(\theta_0,\theta_n,\tilde{\theta}_n)\right| > \epsilon, \sup_{n\geq N}\|\theta_n-\theta_0\| < \gamma\right) + \lim_{N\to\infty} \mathbb{P}\left(\sup_{n\geq N}\|\theta_n-\theta_0\| \geq \gamma\right) \\
&\quad\leq \lim_{N\to\infty} \mathbb{P}\left(\sup_{n\geq N}\frac{1}{n}\sum_{i=1}^{n} A_{r,s}(\delta_i, z_i, \gamma) > \epsilon\right) + \lim_{N\to\infty} \mathbb{P}\left(\sup_{n\geq N}\|\theta_n-\theta_0\| \geq \gamma\right) \\
&\quad\leq \lim_{N\to\infty} \mathbb{P}\left(\sup_{n\geq N}\frac{1}{n}\sum_{i=1}^{n} A_{r,s}(\delta_i, z_i, \gamma) - a_{r,s}(\gamma) > \frac{\epsilon}{2}\right) + \lim_{N\to\infty} \mathbb{P}\left(\sup_{n\geq N}\|\theta_n-\theta_0\| \geq \gamma\right) = 0.
\end{aligned}$$

The first term vanishes due to SLLN since $A_{r,s}(\delta_i, z_i, \gamma)$ are i.i.d. for $i = 1, \ldots, n$. The second term is zero because the MLE θ_n is strongly consistent by assumption (A2). So we have $R_{n,r,s}(\theta_0, \theta_n, \tilde{\theta}_n) \xrightarrow{a.s.} 0$ as $n \to \infty$ and hence

$$R_n(\theta_0, \theta_n, \tilde{\theta}_n) \xrightarrow[n\to\infty]{a.s.} 0. \tag{B.4}$$

Using the latter result together with (B.3) gives $|T_n(\theta_0) + R_n(\theta_0, \theta_n, \tilde{\theta}_n)| \xrightarrow{a.s.} |-I(\theta_0)| \neq 0$ as $n \to \infty$ since the determinant is a continuous mapping and $I(\theta_0)$ is positive definite by (A5). Therefore, for n large enough, $[T_n(\theta_0) + R_n(\theta_0, \theta_n, \tilde{\theta}_n)]$ is invertible. Since $\mathrm{Grad}(l_n(\theta_n)) = 0$ it follows from (B.1), (B.3) and (B.4) for n large enough

$$\begin{aligned}
& & 0 &= U_n(\theta_0) + \left[T_n(\theta_0) + R_n(\theta_0,\theta_n,\tilde{\theta}_n)\right](\theta_n - \theta_0) \\
&\Leftrightarrow & (\theta_n - \theta_0) &= -\left[T_n(\theta_0) + R_n(\theta_0,\theta_n,\tilde{\theta}_n)\right]^{-1} U_n(\theta_0) \\
&\Leftrightarrow & (\theta_n - \theta_0) &= \left[I(\theta_0)^{-1} + o(1)\right] U_n(\theta_0) \qquad \text{a.s.} \\
&\Leftrightarrow & \left(\frac{n}{2\ln\ln n}\right)^{1/2}(\theta_n - \theta_0) &= \left[I(\theta_0)^{-1} + o(1)\right]\left(\frac{n}{2\ln\ln n}\right)^{1/2} U_n(\theta_0) \qquad \text{a.s.} \\
& & &\xrightarrow[n\to\infty]{a.s.} I(\theta_0)^{-1} C < \infty
\end{aligned}$$

when applying the continuous mapping theorem and where $C = (C_1, \dots, C_k)$ is the a.s. limit of $\left(\frac{n}{2\ln\ln n}\right)^{1/2} U_n(\theta_0)$. Then by the law of iterated logarithm

$$\limsup_{n\to\infty} \left(\frac{n}{2\ln\ln n}\right)^{1/2} U_{n,r}(\theta_0) \overset{a.s.}{=} \sqrt{\mathrm{Var}(U_{n,r}(\theta_0))} = \sqrt{\sigma_{r,r}}$$

for all $r = 1, \dots, k$. Due to assumption (A5), C is bounded and therefore

$$\|\theta_n - \theta_0\| \overset{a.s.}{=} \mathcal{O}\left(\left(\frac{2\ln\ln(n)}{n}\right)^{1/2}\right). \qquad \square$$

Curriculum Vitae

Carsten Harlaß

Place of birth: Gera, Germany

Education:

Mathematical-technical Assistant (MaTA)/Computer science, August 2009

- IHK/CCI Cologne, Germany

Bachelor of Science in Scientific Programming, September 2009

- University of Applied Sciences Aachen, Germany
- Bachelor's thesis titled «*Java-Interface for the integrated Monte Carlo based gamma-dose rate calculation for nuclear waste containers*»

Master of Science in Technomathematics, July 2011

- University of Applied Sciences Aachen, Germany
- Master's thesis titled «*Dichteschätzung der Lebensdauerverteilung unter semi-parametrisch zufällig rechtszensierten Daten*»

Dissertation Title: Density estimation for lifetime distributions under semi-parametric random censorship models