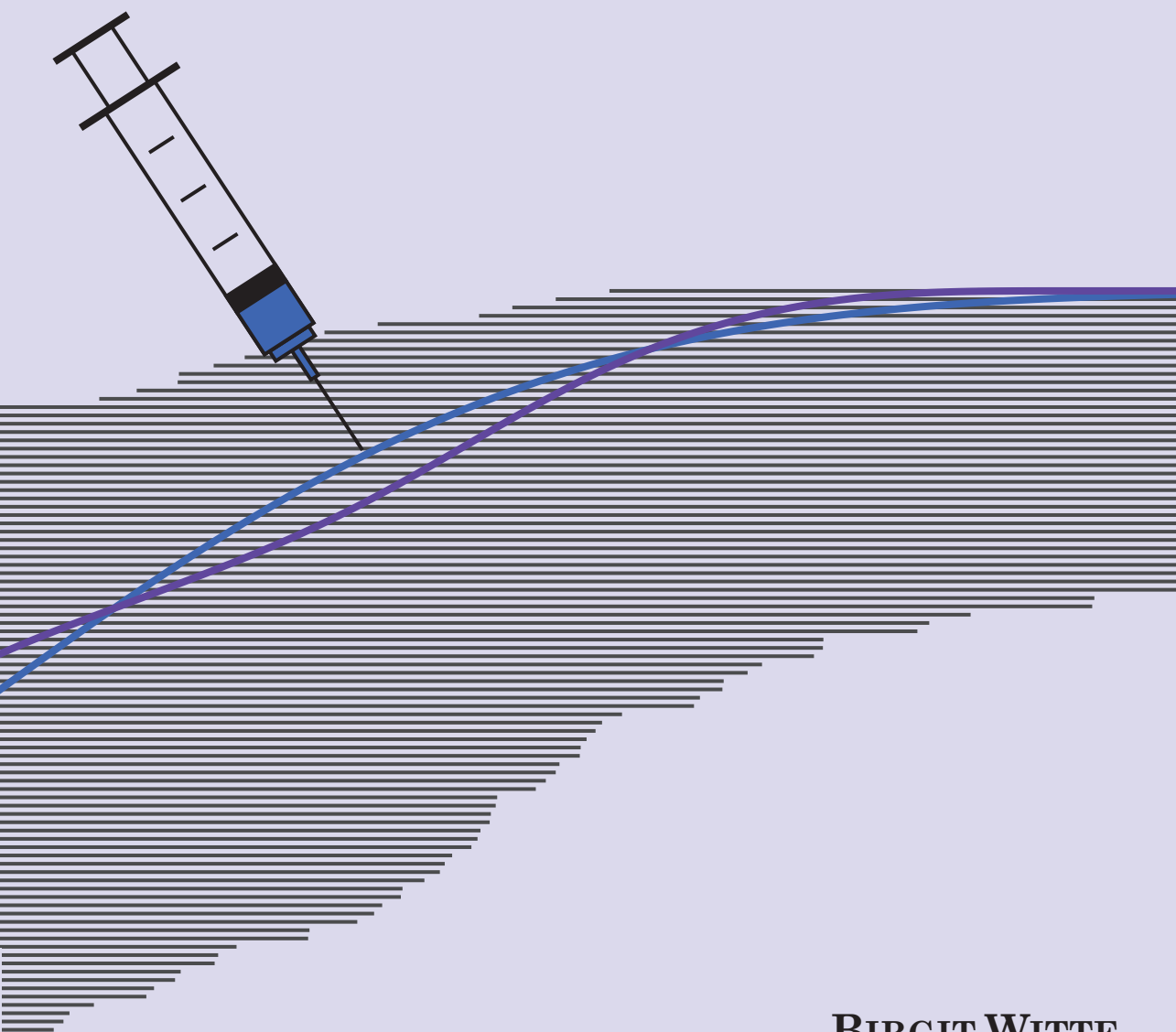


CURRENT STATUS CENSORING MODELS

smooth estimators and their asymptotic properties



BIRGIT WITTE

CURRENT STATUS CENSORING MODELS

smooth estimators and their asymptotic properties

PROEFSCHRIFT

ter verkrijging van de graad doctor
aan de Technische Universiteit Delft,
op gezag van de Rector Magnificus prof. ir. K. C. A. M. Luyben,
voorzitter van het College voor Promoties,
in het openbaar te verdedigen op

dinsdag **15 maart 2011** om *15:00 uur*

door

BIRGIT ILJA WITTE
Master of Science in Mathematics

geboren te Amsterdam

Dit proefschrift is goedgekeurd door de promotoren:

Prof. dr. P. Groeneboom

Prof. dr. ir. G. Jongbloed

Samenstelling promotiecommissie:

Rector Magnificus,	voorzitter
Prof. dr. P. Groeneboom,	Technische Universiteit Delft, promotor
Prof. dr. ir. G. Jongbloed,	Technische Universiteit Delft, promotor
Prof. dr. M. C. M. de Gunst,	Vrije Universiteit Amsterdam
Prof. dr. ir. P. P. B. Eggermont,	University of Delaware
Prof. dr. F. M. Dekking,	Technische Universiteit Delft
Dr. H. P. Lopuhaä,	Technische Universiteit Delft
Dr. ir. M. H. Maathuis,	ETH Zürich
Prof. dr. ir. A. W. Heemink,	Technische Universiteit Delft, reservelid

Dit onderzoek kwam tot stand met steun van NWO.



Nederlandse Organisatie voor Wetenschappelijk Onderzoek

Current Status Censoring Models - smooth estimators and their asymptotic properties

Dissertation at Delft University of Technology

Copyright © 2011 by B. I. Witte

ISBN: 978-90-8891-233-7

Printed by: Proefschriftmaken.nl || Printyourthesis.com

Published by: Uitgeverij BOXPress, Oosterwijk, The Netherlands

ACKNOWLEDGEMENT

Over the past five years I have faced several challenges and learned many things during my PhD research. Fortunately, I did not have to face them all alone and a number of people supported me through the learning process. Therefore, I start this thesis, marking the end point of my PhD research, by thanking all of my friends, family members and colleagues. Listing all of their names would make this thesis twice as thick, but some of them deserve to be mentioned explicitly.

First of all, I wish to express my gratitude to my supervisors Piet Groeneboom and Geurt Jongbloed who made it possible for me to start this PhD project, funded by the Nederlandse Organisatie voor Wetenschappelijk Onderzoek (NWO). Thank you both for guiding me into the academic world. Piet, thank you for sharing and discussing your wonderful ideas with me; they formed the building blocks of this research. Geurt, I still remember your first lecture in Statistische Data Analyse in my third year as a Bachelor student in Mathematics at the Vrije Universiteit. From then on, you introduced the scientific world to me and supervised all my theses (Bachelor, Master and now PhD). Thank you for all of our discussions and your help whenever I had smaller or bigger problems that needed solutions. It was a great pleasure to work with you and hopefully we can continue our collaboration.

I am indebted to all of the members of the reading committee: Michel Dekking, Paul Eggermont, Mathisca de Gunst, Rik Lopuhaä and Marloes Maathuis. I thank all of them for reading the manuscript and especially Paul Eggermont for his valuable remarks.

I also want to thank all of the people from the probability and statistics group, you made my time in Delft quite enjoyable. Special gratitude goes to my (former) office mates Stefanie Michael, Lidka Burzała-Kowalczyk and Tina Nane as well as to our secretary Cindy Bosman. Thank you for the nice time in our office and during our conferences abroad. I really enjoyed our discussions about all sorts of problems (both statistical and non-statistical). Lidka, thank you for being one of my paranimfs. Cindy, thank you for taking care of the non-scientific issues.

I was rather fortunate that there was time for other things besides mathematics during my PhD research. The time on the ice together with my speed skating friends was really pleasant. Although you never quite understood what I have been doing these past five years, you were always supportive and made me forget my problems for several hours. Thank you all for the intensive trainings and races, after which my mind was clear again to tackle my statistical problems.

Last, but absolutely not least, I would like to express my deepest gratitude to my parents, my sister and my dearest Mannes. Mom and Dad, I did not always listen to you; Dad I still remember your words “Wiskunde is niets voor jou”. However, you supported me in whatever I was doing and it is because of that I have managed to get to where I am now. Renske, you were always there to listen whenever I needed someone to talk to. Thank you for all your advice and for being my other paranimf. Mannes, your love and endless support were always encouraging me to continue with my research and finish my PhD. Thank you, with all my heart, for always being there for me.

Birgit Witte
Sint Pancras, January 2011

CONTENTS

1	INTRODUCTION	1
1.1	ESTIMATING A DECREASING DENSITY	2
1.2	CENSORING MODELS	8
1.3	OBJECTIVE AND OUTLINE	12

CURRENT STATUS MODEL

2	SMOOTHED MAXIMUM LIKELIHOOD ESTIMATION	17
2.1	DEFINITION OF THE ESTIMATORS	18
2.2	ASYMPTOTIC DISTRIBUTIONS OF THE ESTIMATORS	24
2.3	OPTIMAL BANDWIDTHS	32
2.4	APPLICATION: HEPATITIS A	35
2.5	TECHNICAL LEMMAS	35
3	MAXIMUM SMOOTHED LIKELIHOOD ESTIMATION	45
3.1	DEFINITION OF THE ESTIMATORS	46
3.2	ASYMPTOTIC DISTRIBUTIONS OF THE ESTIMATORS	51
3.3	APPLICATION: HEPATITIS A	62
3.4	CONCLUDING REMARKS	62
3.5	TECHNICAL LEMMAS	66

CURRENT STATUS CONTINUOUS MARK MODEL

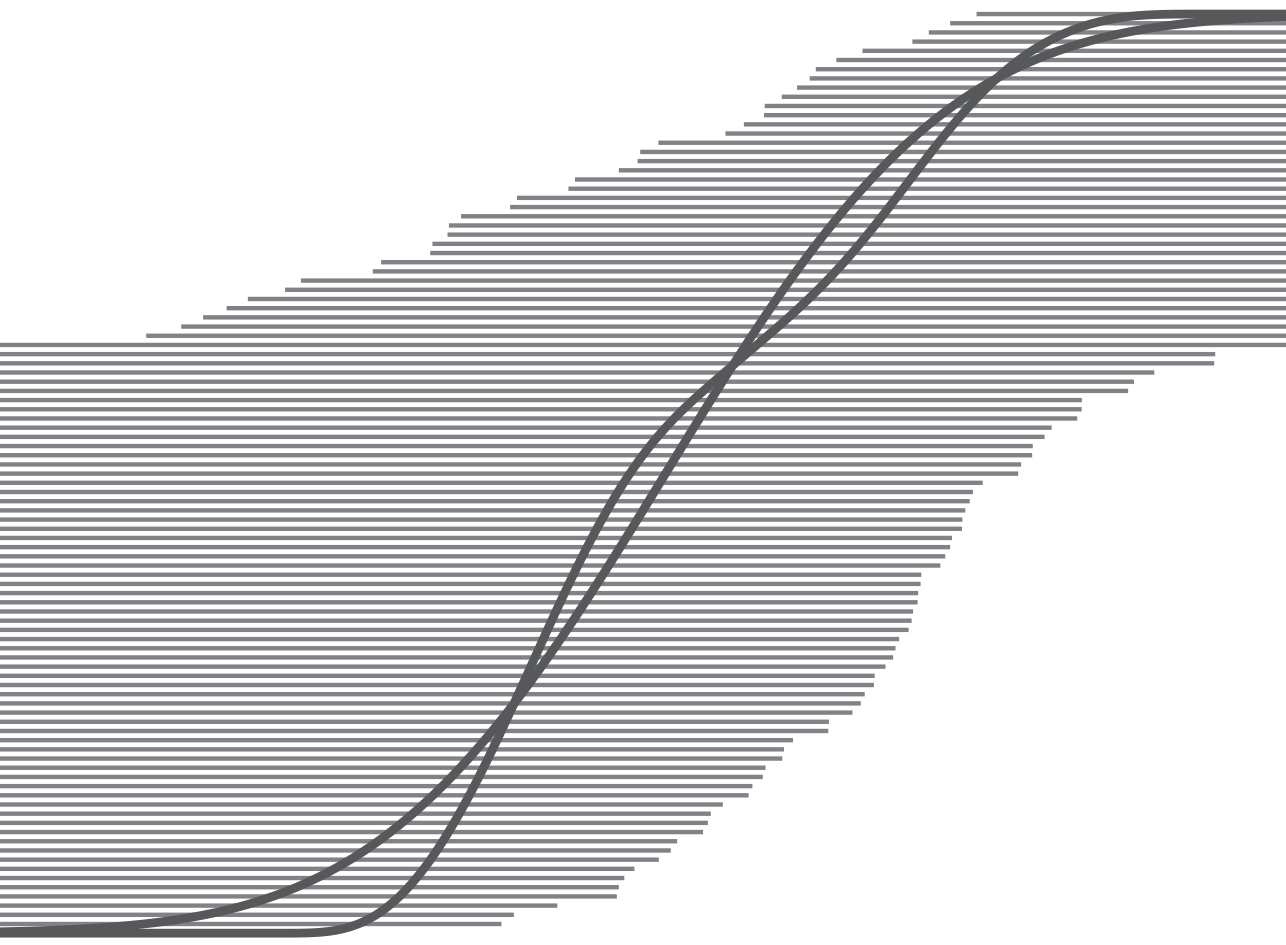
4	MAXIMUM SMOOTHED LIKELIHOOD ESTIMATION	75
4.1	DEFINITION OF THE ESTIMATOR	76
4.2	COMPUTING THE ESTIMATOR	84
4.3	CONSISTENCY OF THE ESTIMATOR	87
4.4	TECHNICAL LEMMAS	90
5	SMOOTH PLUG-IN INVERSE ESTIMATION	97
5.1	DEFINITION OF THE ESTIMATORS	98
5.2	CONSISTENCY AND MONOTONICITY	103
5.3	ASYMPTOTIC DISTRIBUTIONS OF THE ESTIMATORS	106
5.4	SMALL COMPARATIVE SIMULATION STUDY	113
5.5	CONCLUDING REMARKS	115
5.6	TECHNICAL LEMMAS	116

APPENDIX

A	KERNEL ESTIMATION	123
B	NORMS, DISTANCES AND VARIOUS RELATIONS	127
	BIBLIOGRAPHY	131
	SUMMARY	135
	SAMENVATTING	137
	CURRICULUM VITAE	139

CHAPTER 1

INTRODUCTION



Survival analysis is a part of statistics that deals with situations where one is interested in the distribution of a random variable that depicts the time at which an important event occurs. In actuarial applications this could be ‘age of death’, in medical applications ‘time of being declared cured from a disease’ and in reliability theory ‘time of failure of an electrical component’. When estimating such a distribution function, various assumptions can be imposed on it. For example, one could assume that the distribution function F_0 belongs to the class of exponential distributions, i.e. F_0 belongs to the class

$$\mathcal{F}_\vartheta = \{F : F(x) = 1 - e^{-\vartheta x}, \vartheta > 0\}.$$

Sometimes, this parametric assumption is too restrictive, and one only imposes some smoothness condition on F_0 or some shape constraints on its density f_0 , i.e. $F_0 \in \mathcal{F}_s$ or $f_0 \in \mathcal{F}_m$ with

$$\begin{aligned} \mathcal{F}_s &= \{F : F \text{ is a smooth distribution function on } \mathbb{R}\}, \text{ and} \\ \mathcal{F}_m &= \{f : f \text{ is a monotone decreasing density on } [0, \infty)\}. \end{aligned}$$

In many situations the event time is not observed directly or completely due to some sort of censoring. Imagine, for example, one is interested in the distribution of the age of infection with HIV. Data needed to estimate this distribution function is collected in a longitudinal study, where participants are tested for HIV infection during several follow-ups. Therefore, only a time interval is observed which is known to contain the time of HIV infection. This type of censoring is called interval censoring.

In this thesis, we focus on two censoring models, the current status model and the current status continuous mark model. We restrict ourselves to estimators for the distribution function F_0 and its density f_0 that combine smoothness assumptions with monotonicity restrictions. Various estimators are defined and their asymptotic properties are studied.

1.1 ESTIMATING A DECREASING DENSITY

In this section, several estimation methods are described that combine smoothness assumptions and shape restrictions. The methods are illustrated using the problem of estimating a monotone density based on an i.i.d. sample generated by it. That is, in this section we assume that the random variable X is observed directly and its unknown density f_0 belongs to the class $\mathcal{F}_m$.

MAXIMUM LIKELIHOOD ESTIMATION

The Maximum Likelihood Estimator (MLE) $\hat{f}_n$ for f_0 is defined as the maximizer of the log likelihood

$$\ell(f) = \int \log f(u) d\mathbb{F}_n(u), \quad (1.1)$$

over the class $\mathcal{F}_n$ of all decreasing densities on $[0, \infty)$. Here, $\mathbb{F}_n$ is the empirical distribution function of the sample $X_1, X_2, \dots, X_n$. The MLE for the monotone density f_0 is also known as the Grenander estimator, as it was first studied by GRENNANDER (1956). It can be characterized as the left-continuous slope of the least concave majorant of the empirical distribution function $\mathbb{F}_n$ of the data. Figure 1.1(a) illustrates the empirical distribution function of a simulated sample of size $n = 25$ and its least concave majorant. The resulting estimator $\hat{f}_n$ is shown in Figure 1.1(b), as well as the true density f_0 which was taken to be the standard exponential density.

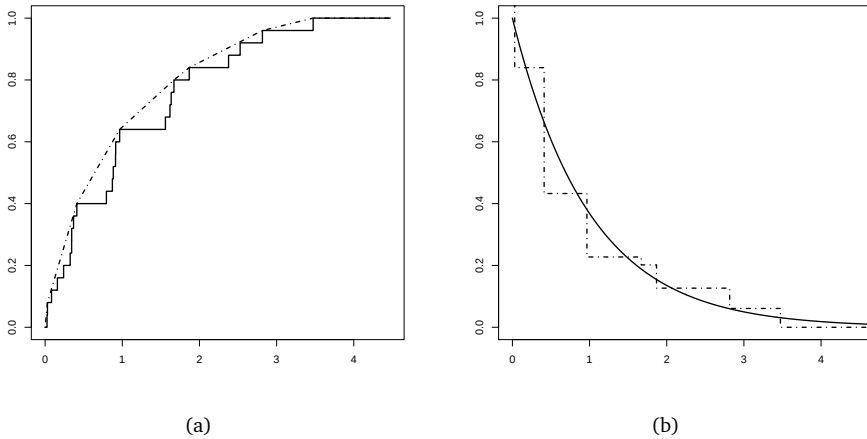


FIGURE 1.1: Characterization of the MLE $\hat{f}_n$ for a simulated sample of size $n = 25$. (a): The empirical distribution function (solid line) and its least concave majorant (dash-dotted line). (b): The estimator $\hat{f}_n$ (dash-dotted line) and the true density f_0 (solid line).

Let F_0 be the distribution function corresponding to the density f_0 and let $\hat{F}_n$ be the least concave majorant of the empirical distribution function $\mathbb{F}_n$. Pointwise consistency of the MLE $\hat{f}_n$ at points $x > 0$ follows by combining the Glivenko-Cantelli

theorem with the following inequality

$$\sup_x |\hat{F}_n(x) - F_0(x)| \leq \sup_x |\mathbb{F}_n(x) - F_0(x)|, \quad (1.2)$$

see e.g., BARLOW *et al.* (1972). PRAKASA RAO (1969) and GROENEBOOM (1985) prove that under mild smoothness assumptions the Grenander estimator $\hat{f}_n$ evaluated at a fixed point x_0 converges in distribution at rate $n^{1/3}$. More specifically, it is shown that

$$n^{1/3} \left| \frac{1}{2} f(x_0) f'(x_0) \right|^{-1/3} (\hat{f}_n(x_0) - f_0(x_0)) \rightsquigarrow 2Z,$$

where Z is the last time where a standard two-sided Brownian motion with parabolic drift reaches its maximum. The distribution of Z is sometimes referred to as Chernoff's distribution, because it occurred for the first time in the literature in CHERNOFF (1964). More information on occurrences and properties of this distribution can be found in GROENEBOOM AND WELLNER (2001).

SMOOTHED MAXIMUM LIKELIHOOD ESTIMATION

As can be seen in Figure 1.1(b), the MLE is a piecewise constant function with its jumps concentrated at a subset of the sample $X_1, \dots, X_n$. This estimator is not smooth. If the density f_0 is assumed to be smooth, it is natural to use an estimator sharing this property. To obtain a smooth estimator, one can use the method of smoothed maximum likelihood estimation. In this procedure, the MLE is computed first and then smoothed using for example kernel smoothing. Let k be a kernel density with support $[-1, 1]$ and denote its scaled version by $k_b(u) = b^{-1}k(u/b)$ for a positive bandwidth (or smoothing parameter) $b = b_n$ depending on the sample size n . Then, the kernel Smoothed Maximum Likelihood Estimator (SMLE) for the density f_0 at the point x_0 is defined as

$$\hat{f}_n^{SM}(x_0) = \int k_b(x_0 - u) d\hat{F}_n(u) = \int k_b(x_0 - u) \hat{f}_n(u) du, \quad (1.3)$$

where $\hat{F}_n$ is the least concave majorant of $\mathbb{F}_n$. Figure 1.2 illustrates the SMLE for the same simulated sample of size $n = 25$ as was used in Figure 1.1. The triweight kernel $k(u) = \frac{35}{32}(1 - u^2)^3$ for $-1 \leq u \leq 1$ was used as kernel density.

The support of a monotone density has to be of the form $[0, \infty)$ or $[0, M]$ for some $M < \infty$. However, the kernel density k has support $[-1, 1]$ so that the SMLE $\hat{f}_n^{SM}$ can also have mass on $[-b, 0]$. To get an estimator with mass concentrated on $[0, \infty)$, we use a boundary correction method. Several correction methods are known, we used the mirror image method of SCHUSTER (1985).

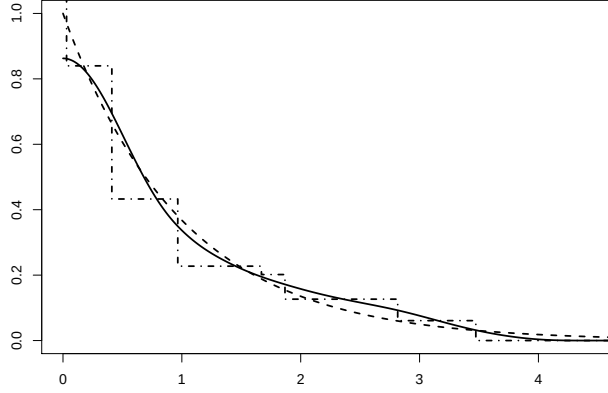


FIGURE 1.2: MLE $\hat{f}_n$ (dash-dotted line) and SMLE $\hat{f}_n^{SM}$ (solid line) for the monotone density f_0 (dashed line) for a simulated sample of size $n = 25$.

As for the MLE $\hat{f}_n$, pointwise consistency of $\hat{f}_n^{SM}$ on $(0, \infty)$ follows from uniform consistency of $\hat{F}_n^{SM}$, where

$$\hat{F}_n^{SM}(x) = \int_0^x \hat{f}_n^{SM}(u) du = \int K_b(x-u) d\hat{F}_n(u)$$

is the SMLE for the distribution function F_0 . Here, K is the distribution function corresponding to the kernel density k , i.e. $K(x) = \int_{-\infty}^x k(u) du$, and $K_b(u) = K(u/b)$ its scaled version. To see this, note that

$$\|\hat{F}_n^{SM} - F_0\|_\infty \leq \|\hat{F}_n^{SM} - \hat{F}_n\|_\infty + \|\hat{F}_n - F_0\|_\infty. \quad (1.4)$$

The second norm on the right-hand side converges in probability to zero by (1.2). To prove the first norm also converges to zero, we first write

$$\begin{aligned} \hat{F}_n^{SM}(x) - \hat{F}_n(x) &= \int \left(K_b(x-u) - \mathbb{1}_{\{u \leq x\}} \right) d\hat{F}_n(u) \\ &= \int \left(K_b(x-u) - \mathbb{1}_{\{u \leq x\}} \right) d(\hat{F}_n - F_0)(u) \\ &\quad + \int \left(K_b(x-u) - \mathbb{1}_{\{u \leq x\}} \right) dF_0(u). \end{aligned} \quad (1.5)$$

Integration by parts yields

$$\begin{aligned}
 & \int_{x-b}^{x+b} (K_b(x-u) - \mathbb{1}_{\{u \leq x\}}) d(\hat{F}_n - F_0)(u) \\
 &= (K_b(x-u) - \mathbb{1}_{\{u \leq x\}}) (\hat{F}_n(u) - F_0(u)) \Big|_{u=x-b}^{x+b} \\
 & \quad + \int_{x-b}^{x+b} k_b(x-u) (\hat{F}_n(u) - F_0(u)) du \\
 &= \int k_b(x-u) (\hat{F}_n(u) - F_0(u)) du.
 \end{aligned}$$

This implies that the absolute value of the first integral in (1.5) is bounded by $\|\hat{F}_n - F_0\|_\infty$, hence converges in probability to zero by (1.2). The absolute value of the second integral in (1.5) is bounded by

$$\int |K_b(x-u) - \mathbb{1}_{\{u \leq x\}}| dF_0(u).$$

The integrand $|K_b(x-u) - \mathbb{1}_{\{u \leq x\}}|$ is bounded by 1 and converges to zero since $b = b_n$ converges to zero as $n \rightarrow \infty$. Hence, this integral converges to zero by the dominated convergence theorem.

MAXIMUM SMOOTHED LIKELIHOOD ESTIMATION

A different, but related, method to obtain a smooth estimator for the monotone density f_0 is given by interchanging the order of smoothing and maximizing the log likelihood in the method of smoothed maximum likelihood estimation. The resulting estimator, the Maximum Smoothed Likelihood Estimator (MSLE), for f_0 is defined as the maximizer of the smoothed log likelihood

$$\hat{f}_n^{MS} = \operatorname{argmax}_{f \in \mathcal{F}} \ell^S(f) = \operatorname{argmax}_{f \in \mathcal{F}} \int \log f(u) d\tilde{F}_n(u), \tag{1.6}$$

where $\tilde{F}_n$ is a smooth estimator for the distribution function F_0 , for example the kernel smoothed empirical distribution function

$$\tilde{F}_n(x) = \int K_b(x-u) dF_n(u).$$

The characterization of the MSLE $\hat{f}_n^{MS}$ can be seen as a continuous version of the characterization of the MLE $\hat{f}_n$, i.e. $\hat{f}_n^{MS}$ is the left-continuous slope of the least concave

majorant of the smooth estimator $\tilde{F}_n$ [see e.g., GROENEBOOM AND JONGBLOED (2010, Example 1)].

Figure 1.3 illustrates the MSLE for the same simulated sample of size $n = 25$ as used in Figures 1.1 and 1.2. The triweight kernel $k(u) = \frac{35}{32}(1 - u^2)^3$ for $-1 \leq u \leq 1$ was used as kernel density. The kernel density estimator $\tilde{f}_n$ on $[0, b)$ was modified similarly as the SMLE.

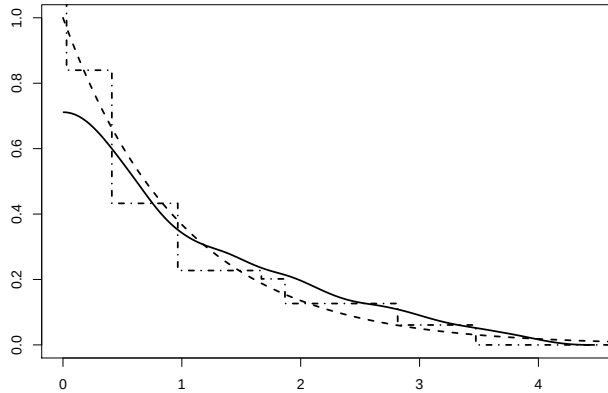


FIGURE 1.3: MLE $\hat{f}_n$ (dash-dotted line) and MSLE $\hat{f}_n^{MS}$ (solid line) for the monotone density f_0 (dashed line) for a simulated sample of size $n = 25$.

EGGERMONT AND LARICCIA (2000) prove that the L_1 -distance between the MSLE $\hat{f}_n^{MS}$ and the true density f_0 is at most equal to the L_1 -distance between the kernel density estimator $\tilde{f}_n$ and f_0 . This implies that $\hat{f}_n^{MS}$ converges to f_0 at least as fast as $\tilde{f}_n$ in L_1 . For bandwidths $b = b_n$ of larger order than $n^{-1/3}$, VAN DER VAART AND VAN DER LAAN (2003) prove that $\hat{f}_n^{MS}$ is asymptotically equivalent with the unrestricted kernel density estimator $\tilde{f}_n$ and converges pointwise to a normal distribution under mild smoothness assumptions. If for example $b = cn^{-1/5}$, then

$$n^{2/5}(\hat{f}_n^{MS}(x_0) - f_0(x_0)) \rightsquigarrow \mathcal{N}(\mu, \sigma^2),$$

with

$$\mu = \frac{1}{2}c^2 f_0''(x_0) \int u^2 k(u) du, \quad \sigma^2 = c^{-1} f_0(x_0) \int k^2(u) du.$$

We conclude this section noting that the MSLE can also be viewed as the Kullback-Leibler projection of a smooth density estimate on the class $\mathcal{F}$. For this, recall that

the Kullback-Leibler projection of a fixed density g on the class $\mathcal{F}$ is defined as

$$\hat{f}_n^{\mathcal{K}} = \operatorname{argmin}_{f \in \mathcal{F}} \mathcal{K}(g, f) = \operatorname{argmin}_{f \in \mathcal{F}} \int g(u) \log \frac{g(u)}{f(u)} du,$$

where $\mathcal{K}(g, f)$ is the Kullback-Leibler divergence of f and g . If g equals the Lebesgue density $\tilde{f}_n$ of the smooth estimator $\tilde{F}_n$ in definition (1.6), then

$$\mathcal{K}(\tilde{f}_n, f) = \int \tilde{f}_n(u) \log \frac{\tilde{f}_n(u)}{f(u)} du = \int \tilde{f}_n(u) \log \tilde{f}_n(u) du - \int \tilde{f}_n(u) \log f(u) du.$$

The first term at the right hand side does not depend on the density f , and the second term is equal to $-\ell^S(f)$. Hence, the minimizer of $\mathcal{K}(\tilde{f}_n, f)$ coincides with the maximizer of $\ell^S(f)$, i.e. the MSLE is the Kullback-Leibler projection of $\tilde{f}_n$ on $\mathcal{F}$.

1.2 CENSORING MODELS

In this section, the two censoring models this thesis focuses on are described in more detail. Furthermore, some examples are given where the data can be described by these censoring models.

THE CURRENT STATUS MODEL

In the current status model, first studied by AYER *et al.* (1955) and VAN EEDEN (1956), the variable of interest is a survival time X with distribution function F_0 . However, instead of observing X , a censoring variable T (with density g) independent of X is observed as well as the indicator variable $\Delta = \mathbb{1}_{\{X \leq T\}}$. This variable indicates whether X is smaller or bigger than T and the observable information on X can be represented by the vector $W = (T, \Delta)$. An example where data can be modeled by the current status model is the Bulgarian hepatitis A prevalence data studied by KEIDING (1991), see also Example 1.1 below.

EXAMPLE 1.1. (HEPATITIS A)

Hepatitis A is a life-long immunity disease and a person can get infected unnoticed. Let X be the age a person gets infected with hepatitis A. To estimate the distribution function F_0 of X , a cross-sectional survey is held where people of various ages have their blood tested for the presence of hepatitis A antibodies. So at time of testing one knows for each person in the survey, the age T and immunization status Δ : infected ($\Delta = 1$) or not ($\Delta = 0$). $\diamond$

The density of the observable vector W with respect to the product of Lebesgue- and counting measure on $[0, \infty) \times \{0, 1\}$ can be written as

$$h_{F_0}(t, \delta) = \delta g(t) F_0(t) + (1 - \delta) g(t) (1 - F_0(t)) = \delta h_1(t) + (1 - \delta) h_0(t). \quad (1.7)$$

Let $\mathbb{H}_n$ be the two-dimensional empirical distribution of an i.i.d. sample $W_1, \dots, W_n$. Then, the log likelihood as a function of F (up to the constant $1/n$) equals

$$\ell(F) = \int \{ \delta \log F(t) + (1 - \delta) \log (1 - F(t)) \} d\mathbb{H}_n(t, \delta). \quad (1.8)$$

Here, a term in the log likelihood that does not depend on the distribution function F is ignored.

GROENEBOOM AND WELLNER (1992) shown that the MLE for F_0 is well defined as a maximizer of (1.8) over all piecewise constant distribution functions with their jumps concentrated at a subset of the sample $T_1, \dots, T_n$. The MLE for F_0 for the Bulgarian hepatitis A prevalence data, studied in KEIDING (1991), is shown in Figure 1.4.

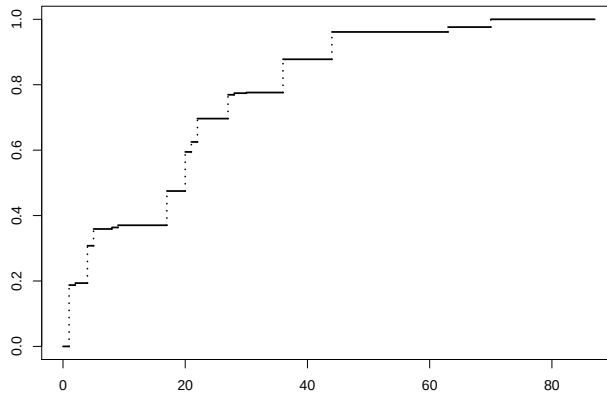


FIGURE 1.4: MLE for hepatitis A prevalence data from Bulgaria.

The MLE can be characterized as the left derivative of the greatest convex minorant of a cumulative sum (cusum) diagram. Let $T_{(i)}$ denote the i -th order statistic of the T_j and $\Delta_{(i)}$ the Δ_j corresponding to it, i.e. $\Delta_{(i)} = \Delta_j$ if $T_{(i)} = T_j$. Then, the cusum diagram consists of the points $P_i = (i, \sum_{j \leq i} \Delta_{(j)})$ for $i = 0, 1, \dots, n$ (with $P_0 = (0, 0)$) and the value of $\hat{F}_n$ in the point $T_{(i)}$ is equal to the left derivative of the convex minorant at the point P_i . The convex minorant is piecewise linear, hence $\hat{F}_n$ is piecewise constant. Denote its jump points by $\tau_1, \dots, \tau_m$ ($m \leq n$), then the set of jump points

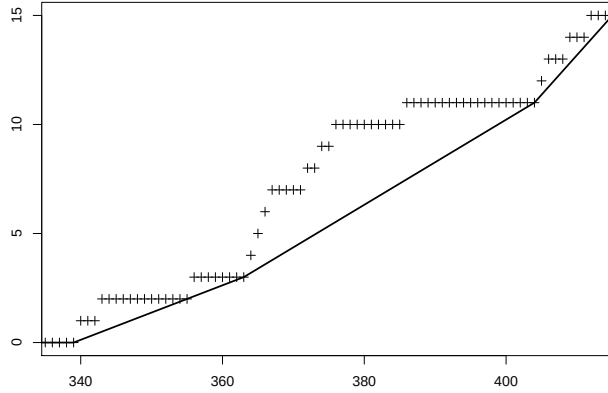


FIGURE 1.5: Part of the cusum diagram and its greatest convex minorant.

$\{\tau_1, \dots, \tau_m\}$ is a subset of the set $\{T_1, \dots, T_n\}$. Figure 1.5 illustrates this characterization. Furthermore, GROENEBOOM AND WELLNER (1992) prove that the estimator is uniformly consistent and that its limiting distribution is the Chernoff distribution.

As noted (and illustrated in Figure 1.4), the MLE $\hat{F}_n$ for the distribution function F_0 of the survival time of interest is piecewise constant, hence does not have a density with respect to Lebesgue measure on $[0, \infty)$. If F_0 is assumed to have a Lebesgue density f_0 , it may be more natural to use estimators other than the MLE $\hat{F}_n$, estimators that are smooth. The smooth estimation methods described in Section 1.1 can be used to obtain smooth estimators for both the distribution function F_0 and its Lebesgue density f_0 (Chapters 2 and 3). In Chapter 3 we prove results for the MSLE $\hat{F}_n^{MS}$ for the distribution function F_0 which are similar to the results for the MSLE $\hat{f}_n^{MS}$ for the monotone density f_0 described in Section 1.1; the characterization of $\hat{F}_n^{MS}$ for F_0 can be seen as continuous version of the characterization of the MLE $\hat{F}_n$ and $\hat{F}_n^{MS}$ is asymptotically equivalent with a naive estimator that does not necessarily satisfy the order constraints of a distribution function.

THE CURRENT STATUS CONTINUOUS MARK MODEL

The second censoring model to be studied in this thesis, the current status continuous mark model, is a special case of the interval censored continuous mark model studied by HUDGENS *et al.* (2007) and MAATHUIS AND WELLNER (2008). The current status continuous mark model is related to the univariate current status model. As before, X is a survival time, but now we are interested in the bivariate vector (X, Y) , where Y is a continuous mark variable which might be correlated to X . The vector (X, Y) has distribution function F_0 with Lebesgue density f_0 . As in the univariate current status model, the variable X is not observed directly. Instead, a censoring variable T

(with density g) independent of (X, Y) is observed as well as the indicator variable $\Delta = \mathbb{1}_{\{X \leq T\}}$. In case X is smaller than or equal to T , i.e. if $\Delta = 1$, the mark variable Y is also observed; in case $\Delta = 0$ the mark variable Y is not observed. Hence, the observable information on (X, Y) can be represented by the vector $W = (T, \Delta \cdot Y) =: (T, Z)$. An example where this model is a good set-up is given in Example 1.2.

EXAMPLE 1.2. (HIV VACCINE TRIALS)

To test the efficacy of a new HIV-vaccine, preventative trials are held where participants are injected with the vaccine and tested for HIV during several follow-ups. One of the questions of interest in these trials is whether the efficacy is related to the genetic sequence of the exposing virus. To answer this question, we estimate the distribution function of the time of infection and the genetic distance between the HIV sequence represented in the vaccine and the HIV sequence the participant is infected with. However, the time of infection is not observed directly but subject to interval censoring and the viral distance is only observed if the participant is already infected at the time of inspection. $\diamond$

In this example, the survival time X is the time of HIV infection and the mark variable Y is the genetic distance, which is treated as a continuous random variable in GILBERT *et al.* (2001). If the participants are only tested once, the censoring variable T is the time of testing and the indicator variable Δ refers to the HIV-infection status at time of inspection: seropositive ($\Delta = 1$) or seronegative ($\Delta = 0$).

Let $B \in \mathcal{B}$, the Borel σ -algebra on $[0, \infty)^2$, let λ_i be Lebesgue measure on $\mathbb{R}^i$ and define the measure λ on $[0, \infty)^2$ by

$$\lambda(B) = \lambda_2(B) + \lambda_1(\{x \in [0, \infty) : (x, 0) \in B\}), \quad B \in \mathcal{B}. \quad (1.9)$$

Then, the density of the observable vector W with respect to this measure λ can be written as

$$\begin{aligned} h_{F_0}(t, z) &= \mathbb{1}_{\{z > 0\}} g(t) \partial_2 F_0(t, z) + \mathbb{1}_{\{z = 0\}} g(t) (1 - F_{0,X}(t)) \\ &= \mathbb{1}_{\{z > 0\}} h_1(t, z) + \mathbb{1}_{\{z = 0\}} h_0(t). \end{aligned} \quad (1.10)$$

Here $F_{0,X}(t) = F_0(t, \infty)$ is the marginal distribution of X and $\partial_2 F_0(t, z) = \frac{\partial}{\partial z} F_0(t, z)$. Let $\mathbb{H}_n$ be the empirical distribution of an i.i.d. sample $W_1, \dots, W_n$. Then, the log likelihood as a function of F (up to the constant $1/n$) equals

$$\ell(F) = \int \{ \mathbb{1}_{\{z > 0\}} \log \partial_2 F(t, z) + \mathbb{1}_{\{z = 0\}} \log (1 - F_X(t)) \} d\mathbb{H}_n(t, z). \quad (1.11)$$

Here, a term in the log likelihood that does not depend on the distribution function F is ignored.

MAATHUIS AND WELLNER (2008) prove that the MLE for the distribution function F_0 is inconsistent. An approach they propose to obtain a consistent estimator is by discretizing the mark variable to K levels. The resulting observations can be viewed as observations from the current status K -competing risk model and asymptotic properties for the MLE in that model are derived in GROENEBOOM *et al.* (2008a, 2008b). An alternative estimator for the distribution function of the undiscretized observations is the MSLE (Chapter 4). Smoothing the maximum likelihood estimator is not sensible since the MLE is inconsistent.

1.3 OBJECTIVE AND OUTLINE

The objective of the research presented in this thesis can be described as follows:

Study the asymptotic properties of smooth likelihood based estimators in two current status censoring models.

Therefore, we first study two estimation methods in the current status model introduced on page 8. If smoothness is assumed, this can be exploited in the estimation process. In Chapters 2 and 3 we study two new smooth estimators: the Smoothed Maximum Likelihood Estimator (SMLE) [see page 4] and the Maximum Smoothed Likelihood Estimator (MSLE) [see page 6].

In Chapter 2, the SMLE is considered for the distribution function F_0 , its Lebesgue density f_0 and the hazard rate λ_0 . The asymptotic distribution of the SMLE for F_0 is derived, using empirical process theory. The asymptotic distributions of the SMLEs for the density f_0 and the hazard rate λ_0 are also derived and the estimators are determined for the data set on hepatitis A from KEIDING (1991).

Interchanging the smoothing and maximization step in the method of smoothed maximum likelihood estimation studied in Chapter 2 yields the MSLE. Chapter 3 is devoted to the MSLEs for F_0 , f_0 and λ_0 . The characterization of the MSLE for F_0 is similar to the characterization of the MLE for F_0 . The asymptotic distribution of the MSLE for F_0 is derived by considering a naive plug-in inverse estimator. This estimator is asymptotically equivalent with the MSLE, and its asymptotic distribution follows by applying the Delta-method. The asymptotic distribution of the MSLEs for the density f_0 and the hazard rate λ_0 follow similarly. The estimators are also determined for the data set on hepatitis A from KEIDING (1991). Most of the results in Chapters 2 and 3 can also be found in GROENEBOOM, JONGBLOED AND WITTE (2010).

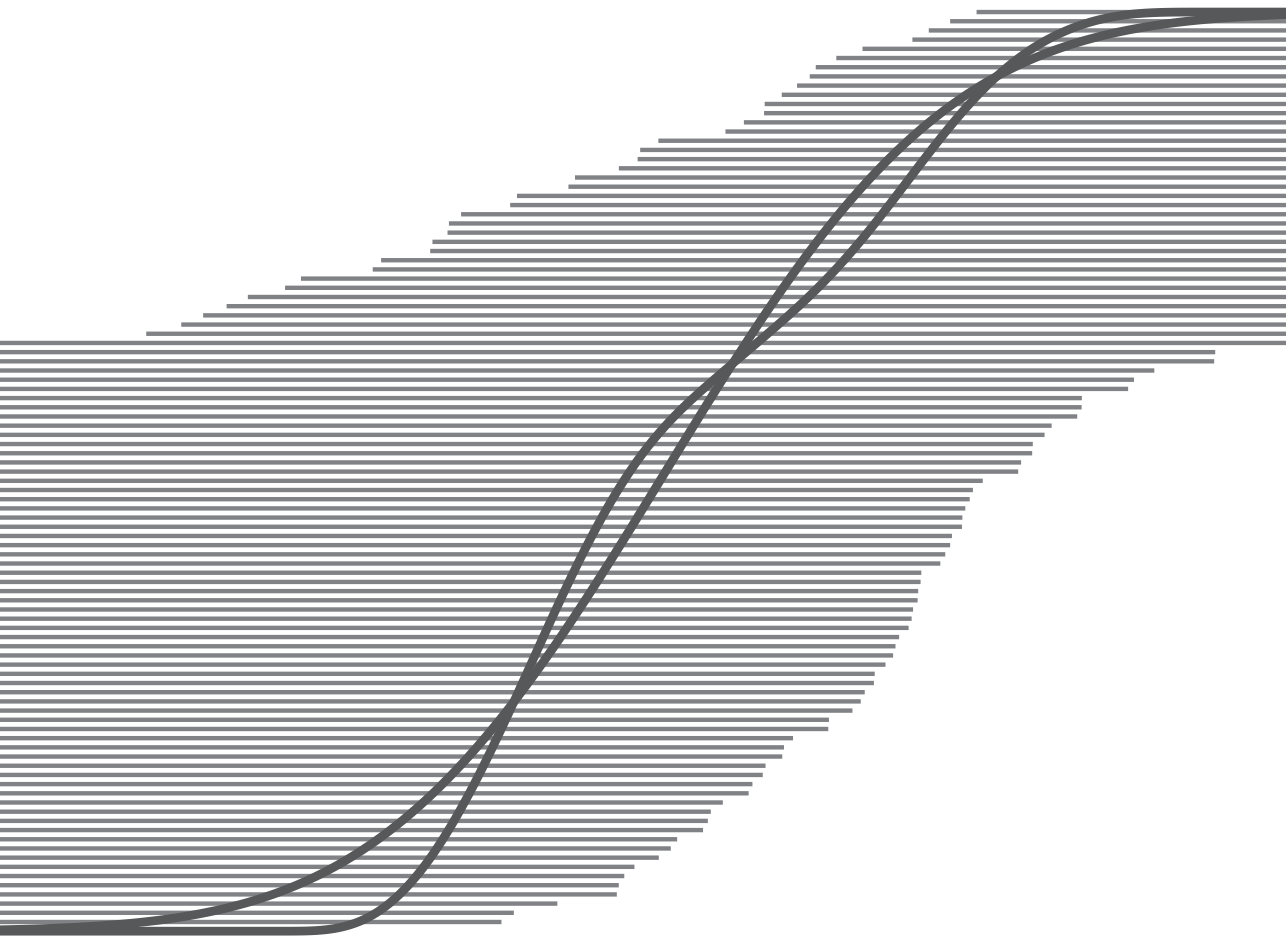
After that, estimators for the bivariate distribution function in the current status continuous mark model, introduced on page 10, are studied. MAATHUIS AND WELLNER (2008) prove that the MLE for the distribution function F_0 in this model is inconsistent. In Chapters 4 and 5 we study two new alternative estimators which turn out to be consistent: the MSLE and the smooth plug-in inverse estimator.

The MSLEs for the bivariate distribution function F_0 , its Lebesgue density f_0 and the marginal distribution function $F_{0,X}$ of X are studied in Chapter 4. The likelihood is smoothed using a bivariate histogram estimator. From Hellinger consistency of the histogram estimator of the observation density, pointwise consistency of the MSLEs for F_0 and $F_{0,X}$ is derived. This chapter is based on GROENEBOOM, JONGBLOED AND WITTE (2011a).

Chapter 5 is devoted to the smooth plug-in inverse estimator for F_0 . This estimator is consistent and asymptotically normally distributed. Although for finite sample size n this estimator has the undesirable property that its Lebesgue density can take negative values, asymptotically it behaves like a bivariate distribution function on large subsets of the support of F_0 . The results in this chapter can also be found in GROENEBOOM, JONGBLOED AND WITTE (2011b).

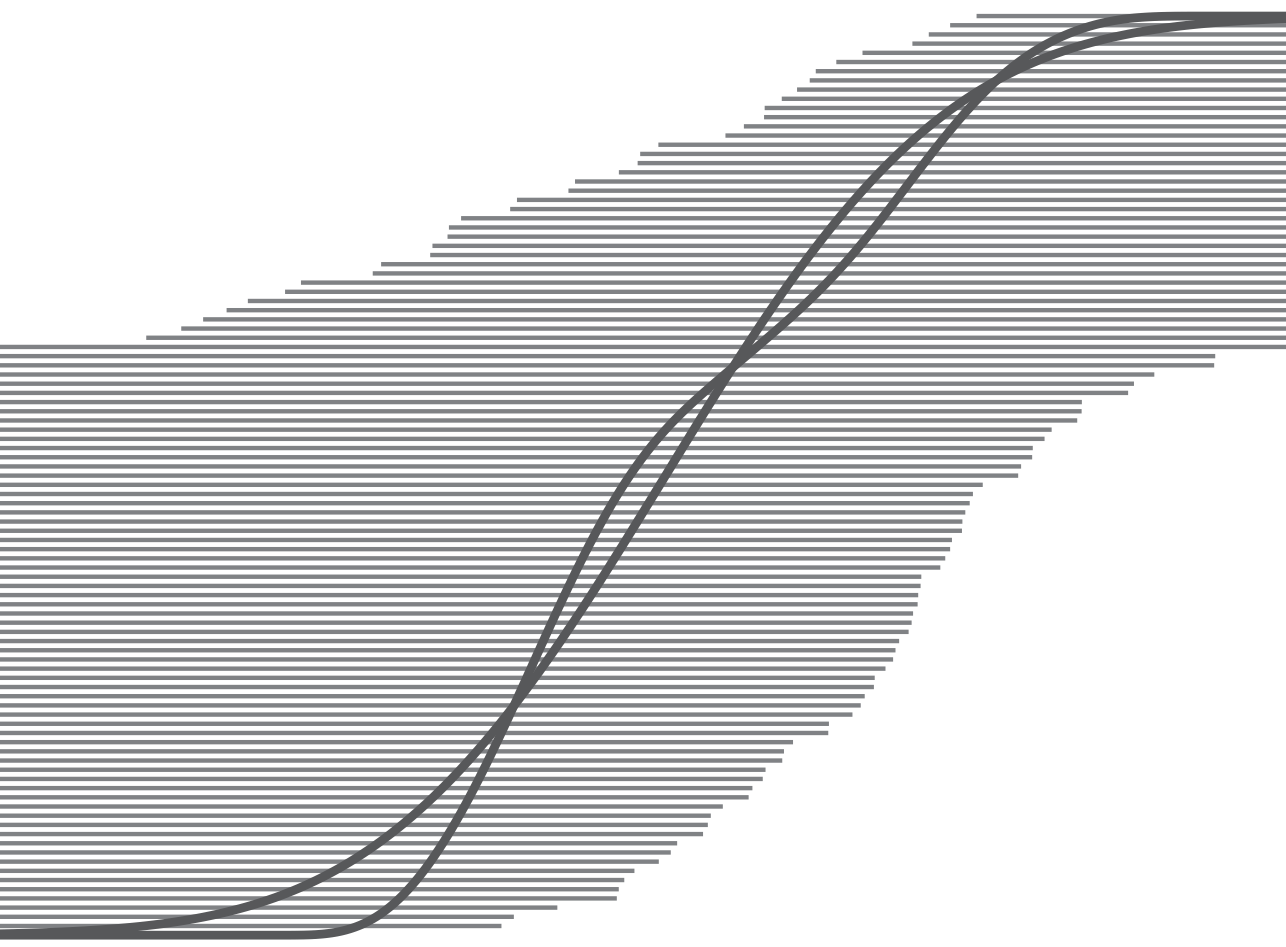
Appendix A contains several asymptotic results on kernel estimators needed in Chapters 2, 3 and 5. Some relations between the Hellinger distance, the Kullback-Leibler divergence and the L_1 -distance needed in Chapter 4 can be found in Appendix B.

CURRENT STATUS MODEL



CHAPTER 2

SMOOTHED MAXIMUM LIKELIHOOD ESTIMATION



The Maximum Likelihood Estimator (MLE) for the distribution function F_0 in the current status model is piecewise constant, consistent and at a fixed point converges at rate $n^{1/3}$ to a Chernoff distribution (GROENEBOOM AND WELLNER, 1992). However, when F_0 is assumed to be smooth, the MLE does not share this property. An estimator for F_0 that is smooth by definition is the Smoothed Maximum Likelihood Estimator (SMLE). In this chapter, the behavior of the SMLE for F_0 is studied in the current status model. Most of the results in this chapter can also be found in GROENEBOOM, JONGBLOED AND WITTE (2010).

This chapter is organized as follows. In Section 2.1 the SMLE for F_0 is defined, as well as the SMLEs for the Lebesgue density f_0 and the hazard rate $\lambda_0 = f_0/(1 - F_0)$. It also contains some results on the MLE, needed in Section 2.2 to derive the asymptotic distributions of the estimators. Kernel smoothing, used to smooth the MLE, requires a smoothing parameter (or bandwidth). Section 2.3 briefly addresses the problem of bandwidth selection in practice. The estimation method is applied to a data set on hepatitis A from KEIDING (1991) in Section 2.4. Technical lemmas needed in some of the proofs can be found in Section 2.5.

2.1 DEFINITION OF THE ESTIMATORS

Recall that in the current status model, as described in Section 1.2, F_0 is the distribution function of a random variable X . However, instead of observing the variable X , the vector $W = (T, \Delta)$ is observed where T is a censoring variable and $\Delta = \mathbb{1}_{\{X \leq T\}}$. The density of the observable vector W with respect to the product of the Lebesgue- and counting measure on $[0, \infty) \times \{0, 1\}$ equals

$$h_{F_0}(t, \delta) = \delta g(t)F_0(t) + (1 - \delta)g(t)(1 - F_0(t)) = \delta h_1(t) + (1 - \delta)h_0(t).$$

In this section, the SMLEs $\hat{F}_n^{SM}$, $\hat{f}_n^{SM}$ and $\hat{\lambda}_n^{SM}$ for F_0 , f_0 and λ_0 are defined, where the MLE is smoothed via kernel smoothing. Let k be a kernel density satisfying condition (K.1) below.

(K.1) The probability density k has support $[-1, 1]$, is symmetric and continuously differentiable on $\mathbb{R}$.

Note that condition (K.1) implies that $m_2(k) = \int u^2 k(u) du < \infty$. Let K be the distribution function corresponding to k , i.e. $K(x) = \int_{-\infty}^x k(u) du$, k' be the derivative of k and $b = b_n > 0$ be a bandwidth (or smoothing parameter) depending on the sample size n . Then we use the following notation for the scaled version of K , k and k'

$$K_b(u) = K(u/b), \quad k_b(u) = \frac{1}{b} k(u/b), \quad \text{and} \quad k'_b(u) = \frac{1}{b^2} k'(u/b). \quad (2.1)$$

Let $\hat{F}_n$ be the MLE for F_0 and denote its jump points by $\tau_1, \dots, \tau_m$ ($m \leq n$), then the SMLE for F_0 at the point t_0 is defined as

$$\hat{F}_n^{SM}(t_0) = \int K_b(t_0 - u) d\hat{F}_n(u) = \sum_{i=1}^m K_b(t_0 - \tau_i) \hat{F}_n(\tau_i). \quad (2.2)$$

The SMLE $\hat{f}_n^{SM}$ for the density f_0 at the point t_0 can be defined similarly,

$$\hat{f}_n^{SM}(t_0) = \int k_b(t_0 - u) d\hat{F}_n(u) = \sum_{i=1}^m k_b(t_0 - \tau_i) \hat{F}_n(\tau_i). \quad (2.3)$$

The hazard rate $\lambda_0 = f_0/(1 - F_0)$ is estimated via the plug-in principle,

$$\hat{\lambda}_n^{SM}(t_0) = \frac{\hat{f}_n^{SM}(t_0)}{1 - \hat{F}_n^{SM}(t_0)}, \text{ for } t_0 \text{ such that } \hat{F}_n^{SM}(t_0) < 1. \quad (2.4)$$

The variable X denotes a positive survival time, i.e. $X \geq 0$, hence we can expect inconsistency problems for the SMLEs at zero. In order to prevent those, we modify the definition of $\hat{f}_n^{SM}$ for $t_0 < b$. To be precise, we define

$$\hat{f}_n^{SM}(t_0) = \int \frac{1}{b} k^\beta \left(\frac{t_0 - u}{b} \right) d\hat{F}_n(u)$$

for $0 \leq t_0 \leq b$ and $\beta = t_0/b$. The so-called boundary kernel k^β is defined by

$$k^\beta(u) = \frac{\nu_{2,\beta}(k) - \nu_{1,\beta}(k)u}{\nu_{0,\beta}(k)\nu_{2,\beta}(k) - \nu_{1,\beta}(k)^2} k(u) \mathbb{1}_{(-1,\beta)}(u),$$

with

$$\nu_{i,\beta}(k) = \int_{-1}^{\beta} u^i k(u) du, \quad i = 0, 1, 2.$$

For $0 \leq t_0 \leq b$, the estimator $\hat{F}_n^{SM}$ is defined as $\hat{F}_n^{SM}(t_0) = \int_{-\infty}^{t_0} \hat{f}_n^{SM}(u) du$. There are other ways to correct the kernel estimator near the boundary, see e.g., SCHUSTER (1985) or JONES (1993). However, simulations show that the results are not much influenced by the boundary correction method used.

For a specific choice of the kernel k , the estimator $\hat{F}_n^{SM}$ can also be identified as the minimizer of the following penalized L_2 -distance

$$\tilde{L}_2(F) := \tilde{L}_2(F; \alpha) = \int (F(x) - \hat{F}_n(x))^2 dx + \alpha \int f(x)^2 dx, \quad \alpha > 0,$$

over the class of all nondecreasing, nonnegative functions. Note that the minimizer, if it exists, will satisfy the “boundary constraints” $\lim_{x \rightarrow -\infty} F(x) = 0$ and $\lim_{x \rightarrow \infty} F(x) = \hat{F}_n(\tau_m)$. Indeed, for any F not satisfying these constraints, $\tilde{L}_2(F) = \infty$. As stated in Theorem 2.1 below, the minimizer $\tilde{F}_n$ of $\tilde{L}_2(F)$ satisfies differential equation (2.5). Theorem 2.2 states that $\hat{F}_n^{SM}$ satisfies this differential equation if $k(u) = \frac{1}{2}e^{-|u|}$ for $u \in \mathbb{R}$ and $b = \sqrt{\alpha}$.

THEOREM 2.1. *Let $\tilde{F}_n$ be a minimizer of $\tilde{L}_2(F)$ over the class of all sub-distribution functions satisfying $\lim_{x \rightarrow -\infty} F(x) = 0$ and $\lim_{x \rightarrow \infty} F(x) = \hat{F}_n(\tau_m)$. If this minimizer $\tilde{F}_n$ is twice continuously differentiable, then it satisfies the following differential equation*

$$F(x) - \alpha F''(x) = \hat{F}_n(x). \quad (2.5)$$

PROOF. The penalized L_2 -distance $\tilde{L}_2(F)$ can be rewritten as follows

$$\begin{aligned} \tilde{L}_2(F) &= \int (F(x) - \hat{F}_n(x))^2 dx + \alpha \int f(x)^2 dx \\ &= \int \{(F(x) - \hat{F}_n(x))^2 + \alpha F'(x)^2\} dx =: \int S(x, F(x), F'(x)) dx. \end{aligned}$$

The minimizer $\tilde{F}_n$ of this functional has to satisfy the Euler-Lagrange equation

$$\begin{aligned} 0 &= \frac{\partial}{\partial F} S(x, F(x), F'(x)) - \frac{d}{dx} \frac{\partial}{\partial F'} S(x, F(x), F'(x)) \\ &= 2(F(x) - \hat{F}_n(x)) - 2\alpha F''(x). \end{aligned}$$

Hence, $\tilde{F}_n$ has to satisfy differential equation (2.5). □

THEOREM 2.2. *Let $\hat{F}_n^{SM}$ be the SMLE for F_0 as defined in (2.2), with kernel density $k(u) = \frac{1}{2}e^{-|u|}$ for $u \in \mathbb{R}$. Then, $\hat{F}_n^{SM}$ is the unique solution to differential equation (2.5) if the smoothing parameter b equals $\sqrt{\alpha}$.*

PROOF. The estimator $\hat{F}_n^{SM}(x)$ is defined as

$$\hat{F}_n^{SM}(x) = \int K_b(x - u) d\hat{F}_n(u),$$

where K_b is defined in (2.1). Then, we have that

$$\frac{d}{dx} \hat{F}_n^{SM}(x) = \int k_b(x - u) d\hat{F}_n(u), \quad \frac{d^2}{dx^2} \hat{F}_n^{SM}(x) = \int k'_b(x - u) d\hat{F}_n(u),$$

for k_b and k'_b as defined in (2.1), hence $\hat{F}_n^{SM}$ is twice continuously differentiable and

$$\hat{F}_n^{SM}(x) - b^2 \frac{d^2}{dx^2} \hat{F}_n^{SM}(x) = \int \{K_b(x-u) - b^2 k'_b(x-u)\} d\hat{F}_n(u).$$

For $k(u) = \frac{1}{2}e^{-|u|}$ we have that

$$k'(u) = -\text{sign}(u) \frac{1}{2} e^{-|u|} \text{ and } K(u) = \begin{cases} \frac{1}{2} e^u & \text{if } u < 0 \\ 1 - \frac{1}{2} e^{-u} & \text{if } u \geq 0 \end{cases},$$

where $\text{sign}(u) = -1$ if $u < 0$ and $\text{sign}(u) = 1$ if $u \geq 0$. This gives that for $x-u < 0$, we have

$$K_b(x-u) - b^2 k'_b(x-u) = \frac{1}{2} e^{(x-u)/b} - b^2 \frac{1}{b^2} \frac{1}{2} e^{(x-u)/b} = 0,$$

and for $x-u \geq 0$ we have

$$K_b(x-u) - b^2 k'_b(x-u) = 1 - \frac{1}{2} e^{-(x-u)/b} + b^2 \frac{1}{b^2} \frac{1}{2} e^{-(x-u)/b} = 1.$$

This implies that

$$\hat{F}_n^{SM}(x) - b^2 \frac{d^2}{dx^2} \hat{F}_n^{SM}(x) = \int \mathbb{1}_{\{u \leq x\}} d\hat{F}_n(u) = \hat{F}_n(x),$$

hence (2.5) is satisfied for $\alpha = b^2$.

To see that $\hat{F}_n^{SM}$ is the unique solution to (2.5), note that by definition of $\hat{F}_n^{SM}$ and the choice of the kernel k

$$\lim_{x \rightarrow -\infty} \hat{F}_n^{SM}(x) = 0, \quad \lim_{x \rightarrow \infty} \hat{F}_n^{SM}(x) = \hat{F}_n(\tau_m).$$

Now, assume $\hat{F}_n^{SM}$ is not unique, i.e. there exists a second solution $\tilde{F}_n$ to (2.5). This implies that the difference $\tilde{F}_n - \hat{F}_n^{SM}$ has to be a solution to the homogeneous equation

$$F(x) - \alpha F''(x) = 0.$$

The general form of this solution is $F(x) = c_1 e^{-x/\sqrt{\alpha}} + c_2 e^{x/\sqrt{\alpha}}$. Since both $\lim_{x \rightarrow -\infty} F(x) = 0$ and $\lim_{x \rightarrow \infty} F(x) = 0$ should hold, both c_1 and c_2 have to be equal to zero. Hence $0 = F(x) = \tilde{F}_n(x) - \hat{F}_n^{SM}(x)$ for all x , implying the uniqueness of $\hat{F}_n^{SM}$. $\square$

Throughout this chapter, the methods are illustrated with simulated data. In these simulations, the true density f_0 is the density of a shifted Gamma(4) distribution, i.e.

$$f_0(x) = \frac{(x-2)^3}{3!} e^{-(x-2)}, \text{ for } x \geq 2.$$

The censoring times are exponentially distributed with mean 3, i.e.

$$g(t) = \frac{1}{3} e^{-t/3}, \text{ for } t \geq 0.$$

The kernel density is taken to be the triweight kernel $k(u) = \frac{35}{32}(1-u^2)^3$ for $-1 \leq u \leq 1$.

For a simulated sample of size $n = 500$, Figure 2.1 shows the estimator $\hat{F}_n^{SM}$ with bandwidth $b = 1.3$ (not obtained by the method described in Section 2.3 but chosen just as illustration). It also shows the maximum likelihood estimator $\hat{F}_n$ and the true distribution function F_0 .

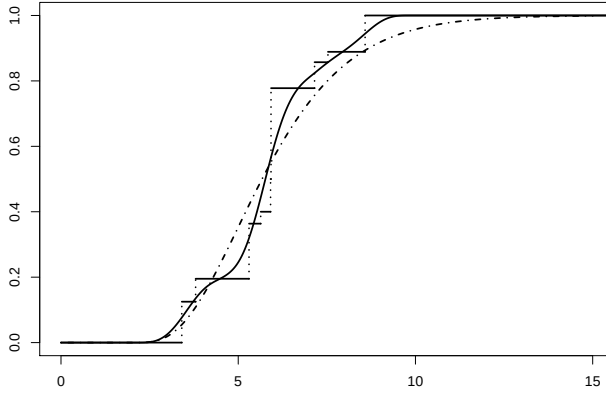


FIGURE 2.1: The true distribution function F_0 (dash-dotted line) and the estimators $\hat{F}_n^{SM}$ with $b = 1.3$ (solid line) and $\hat{F}_n$ (step function) for a simulated sample of size $n = 500$.

Before we derive the asymptotic distributions of the SMLEs some properties of the MLE $\hat{F}_n$ are needed. Lemma 2.3 gives a different characterization of the MLE in terms of an integral with respect to the empirical measure $\mathbb{H}_n$ of the data $W_1, \dots, W_n$.

LEMMA 2.3. For any right continuous piecewise constant function φ with only jumps at the points $\tau_1, \dots, \tau_m$,

$$\int \varphi(u)(\delta - \hat{F}_n(u)) d\mathbb{H}_n(u, \delta) = 0.$$

PROOF Denote the intervals where $\hat{F}_n$ is constant by $J_i = [\tau_i, \tau_{i+1})$ for $0 \leq i \leq m-1$ ($\tau_0 = 0$). Then, for $u \in J_i$

$$\hat{F}_n(u) = \frac{\#\{(T_{(j)}, \Delta_{(j)}) : T_{(j)} \in J_i, \Delta_{(j)} = 1\}}{\#\{(T_{(j)}, \Delta_{(j)}) : T_{(j)} \in J_i\}} = \frac{l}{k},$$

for some k, l such that $l \leq k$. This means that

$$\int_{J_i \times \{0,1\}} \hat{F}_n(u) d\mathbb{H}_n(u, \delta) = \frac{1}{n} \sum_{T_{(j)} \in J_i} \hat{F}_n(T_{(j)}) = \frac{1}{n} \sum_{T_{(j)} \in J_i} \hat{F}_n(\tau_i) = \frac{1}{n} k \frac{l}{k} = \frac{l}{n}.$$

On the other hand, we have

$$\int_{J_i \times \{0,1\}} \delta d\mathbb{H}_n(u, \delta) = \frac{1}{n} \sum_{T_{(j)} \in J_i} \Delta_{(j)} = \frac{l}{n}.$$

This implies

$$\int_{J_i \times \{0,1\}} \varphi(u)(\delta - \hat{F}_n(u)) d\mathbb{H}_n(u, \delta) = \varphi(\tau_i) \int_{J_i \times \{0,1\}} (\delta - \hat{F}_n(u)) d\mathbb{H}_n(u, \delta) = 0,$$

for any right continuous piecewise constant function φ with jump points $\tau_1, \dots, \tau_m$. Hence

$$\int \varphi(u)(\delta - \hat{F}_n(u)) d\mathbb{H}_n(u, \delta) = \sum_{i=1}^{m-1} \int_{J_i \times \{0,1\}} \varphi(u)(\delta - \hat{F}_n(u)) d\mathbb{H}_n(u, \delta) = 0.$$

□

If F_0 and G are assumed to satisfy conditions (F.1) and (G.1) below, GROENEBOOM AND WELLNER (1992) prove [see their Lemma 5.9 and p. 120]

$$\|F_0 - \hat{F}_n\|_\infty = \mathcal{O}_p(n^{-1/3} \log n), \quad (2.6)$$

$$\max_{1 \leq i \leq m} |\tau_{i+1} - \tau_i| = \mathcal{O}_p(n^{-1/3} \log n). \quad (2.7)$$

(F.1) The Lebesgue density f_0 of F_0 has compact support $\mathcal{S}_0 = [0, M_0]$ and stays strictly away from zero.

(G.1) The Lebesgue density g of G stays strictly away from zero on $\mathcal{S}_0$ and its derivative g' is bounded on $\mathcal{S}_0$.

From this, it follows that for fixed $t_0 > 0$, any $v > 0$ and $\mathcal{J}_0 = [t_0 - v, t_0 + v]$

$$\sup_{u \in \mathcal{J}_0} |F_0(u) - \hat{F}_n(u)| = \mathcal{O}_p(n^{-1/3} \log n), \quad (2.8)$$

$$\max_{i: \tau_i \in \mathcal{J}_0} |\tau_{i+1} - \tau_i| = \mathcal{O}_p(n^{-1/3} \log n). \quad (2.9)$$

2.2 ASYMPTOTIC DISTRIBUTIONS OF THE ESTIMATORS

To derive the asymptotic distribution of $\hat{F}_n^{SM}$ we study the behavior of

$$\begin{aligned} \hat{F}_n^{SM}(t_0) - F_0(t_0) &= \int K_b(t_0 - u) d\hat{F}_n(u) - F_0(t_0) \\ &= \int K_b(t_0 - u) d(\hat{F}_n - F_0)(u) + \int K_b(t_0 - u) dF_0(u) - F_0(t_0). \end{aligned}$$

Lemma 2.4 gives a different expression for the first term at the right hand side. With this new expression for the difference $\hat{F}_n^{SM}(t_0) - F_0(t_0)$ we can derive the pointwise asymptotic distribution of $\hat{F}_n^{SM}$.

LEMMA 2.4. *Fix $t_0 > 0$, such that g is strictly positive in a neighborhood of t_0 and define*

$$\psi_{b,t_0}(u) = \frac{k_b(t_0 - u)}{g(u)}.$$

Then

$$\int K_b(t_0 - u) d(\hat{F}_n - F_0)(u) = - \int \psi_{b,t_0}(u) (\delta - \hat{F}_n(u)) dH_{F_0}(u, \delta). \quad (2.10)$$

PROOF. The left-hand side of (2.10) can be rewritten as follows

$$\begin{aligned}
& \int_0^{t_0+b} K_b(t_0-u) d(\hat{F}_n - F_0)(u) \\
&= \int_0^{t_0-b} d(\hat{F}_n - F_0)(u) + \int_{t_0-b}^{t_0+b} K_b(t_0-u) d(\hat{F}_n - F_0)(u) \\
&= \hat{F}_n(t_0-b) - F_0(t_0-b) + K_b(t_0-u)(\hat{F}_n(u) - F_0(u)) \Big|_{u=t_0-b}^{t_0+b} \\
&\quad - \int_{t_0-b}^{t_0+b} -(\hat{F}_n(u) - F_0(u)) k_b(t_0-u) du \\
&= \int_{t_0-b}^{t_0+b} \frac{k_b(t_0-u)}{g(u)} (\hat{F}_n(u) - F_0(u)) dG(u) \\
&= - \int \psi_{b,t_0}(u) (\delta - \hat{F}_n(u)) dH_{F_0}(u, \delta).
\end{aligned}$$

□

By Lemma 2.4 we can consider the integral at the right-hand side of (2.10) to derive the pointwise asymptotic distribution of the estimator $\hat{F}_n^{SM}$. The idea of the proof of this distribution, given in the next theorem, is as follows. By the characterization of the MLE (Lemma 2.3) we could add the term $d\mathbb{H}_n$ for free in the right-hand side of (2.10) if ψ_{b,t_0} were piecewise constant. For most choices of k this function ψ_{b,t_0} is not piecewise constant. Replacing it by an appropriately chosen piecewise constant function results in an additional σ_p -term which does not influence the asymptotic distribution. By some more adding and subtracting, resulting in some more σ_p -terms, we get that

$$\begin{aligned}
& -n^{2/5} \int \psi_{b,t_0}(u) (\delta - \hat{F}_n(u)) dH_{F_0}(u, \delta) \\
&= n^{2/5} \int \psi_{b,t_0}(u) (\delta - F_0(u)) d(\mathbb{H}_n - H_{F_0})(u, \delta) + \sigma_p(1)
\end{aligned}$$

and the pointwise asymptotic distribution follows from the central limit theorem. Based on the asymptotic distribution we can determine the optimal bandwidth which minimizes the asymptotic Mean Squared Error (aMSE).

THEOREM 2.5. *Assume F_0 and G satisfy conditions (F.1) and (G.1). Fix $t_0 > 0$ such that f'_0 is continuous at t_0 and $f'_0(t_0) \neq 0$. Let $b = cn^{-\alpha}$ ($c > 0$) be the bandwidth used in*

the definition of $\hat{F}_n^{SM}$. Then for $\alpha = 1/5$

$$n^{2/5}(\hat{F}_n^{SM}(t_0) - F_0(t_0)) \rightsquigarrow \mathcal{N}(\mu_{F,SM}, \sigma_{F,SM}^2),$$

where

$$\mu_{F,SM} = \frac{1}{2}c^2m_2(k)f_0'(t_0), \quad \sigma_{F,SM}^2 = \frac{F_0(t_0)(1-F_0(t_0))}{cg(t_0)} \int k(u)^2 du. \quad (2.11)$$

For $0 < \alpha < 1/5$,

$$n^{2\alpha}(\hat{F}_n^{SM}(t_0) - F_0(t_0)) \xrightarrow{\mathcal{D}} \mu_{F,SM}.$$

For fixed $t_0 > 0$, the aMSE-optimal bandwidth b for estimating $\hat{F}_n^{SM}(t_0)$ is given by $b_{n,F,SM} = c_{F,SM}n^{-1/5}$, where

$$c_{F,SM} = \left\{ \frac{F_0(t_0)(1-F_0(t_0))}{g(t_0)} \int k(u)^2 du \right\}^{1/5} \left\{ m_2^2(k)f_0'(t_0)^2 \right\}^{-1/5}. \quad (2.12)$$

REMARK. In case t_0 is such that $f_0'(t_0) = 0$, the optimal rate of $b_{n,F,SM}$ is $n^{-1/9}$ resulting in a rate of convergence $n^{4/9}$ and an asymptotic bias $\mu_{F,SM} = \frac{1}{4!}c^4m_4(k)f_0^{(3)}(t_0)$. This is in line with results for other kernel smoothers in case of vanishing first order bias terms.

Figure 2.2 illustrates this pointwise asymptotic result by a simulation study. The value of $n^{2/5}(\hat{F}_n^{SM}(4.75) - F_0(4.75))$ was calculated for a sample of size $n = 100\,000$ with $b = 0.5$ as smoothing parameter. This was repeated $B = 500$ times. Figure 2.2 shows the QQ-plot (with the line $y = \mu_{F,SM} + \sigma_{F,SM}x$) as well as the histogram (with the $\mathcal{N}(\mu_{F,SM}, \sigma_{F,SM}^2)$ -density) of these B values.

Before proving Theorem 2.5, we make the piecewise constant version of the function ψ_{b,t_0} explicit. This function should be constant on the same intervals as the MLE is. Denote the piecewise constant version of ψ_{b,t_0} by $\tilde{\psi}_{b,t_0}$. For $u \in J_i[\tau_i, \tau_{i+1})$, this function is defined as $\tilde{\psi}_{b,t_0}(u) = \psi(\hat{A}_n(u))$ with

$$\hat{A}_n(u) = \begin{cases} \tau_i & \text{if } \forall t \in J_i : F_0(t) > \hat{F}_n(\tau_i), \\ s & \text{if } \exists s \in J_i : \hat{F}_n(s) = F_0(s), \\ \tau_{i+1} & \text{if } \forall t \in J_i : F_0(t) < \hat{F}_n(\tau_i), \end{cases} \quad (2.13)$$

see also Figure 2.3.

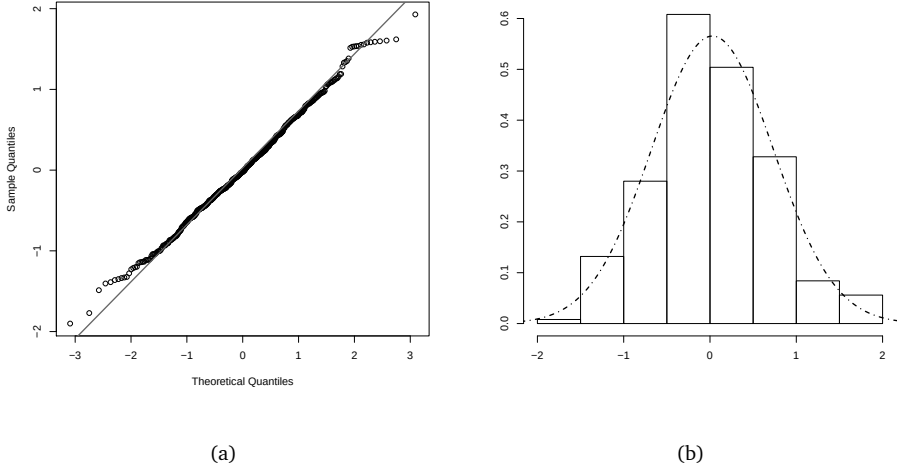


FIGURE 2.2: QQ-plot (a) and histogram (b) of $B = 500$ values of $n^{2/5}(\hat{F}_n^{SM}(4.75) - F_0(4.75))$ for a simulation study with sample size $n = 100\,000$ and $b = 0.5$.

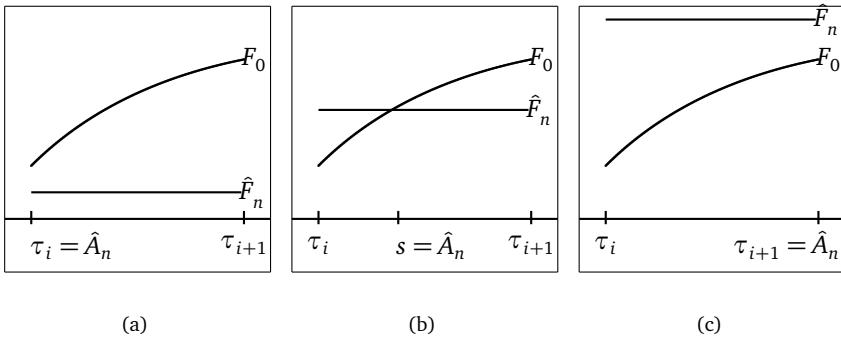


FIGURE 2.3: The 3 different possibilities for the function $\hat{A}_n$. (a) $F_0(t) > \hat{F}_n(\tau_i)$ for all $t \in J_i$; (b) $F_0(s) = \hat{F}_n(\tau_i)$ for some $s \in J_i$; (c) $F_0(t) < \hat{F}_n(\tau_i)$ for all $t \in J_i$.

PROOF OF THEOREM 2.5. Using the piecewise constant version $\bar{\psi}_{b,t_0}$ of ψ_{b,t_0} , we can write

$$\int \psi_{b,t_0}(u)(\delta - \hat{F}_n(u)) dH_{F_0}(u, \delta) = \int \bar{\psi}_{b,t_0}(u)(\delta - \hat{F}_n(u)) dH_{F_0}(u, \delta) + R_n,$$

where for $b = cn^{-\alpha}$ (with $0 < \alpha \leq 1/5$) and n sufficiently large

$$\begin{aligned} |R_n| &= \int (\psi_{b,t_0}(u) - \bar{\psi}_{b,t_0}(u))(\delta - \hat{F}_n(u)) dH_{F_0}(u, \delta) \\ &\leq c_1 b^{-2} \int_{u \in [t_0 - b, t_0 + b]} |F_0(u) - \hat{F}_n(u)|(\delta - \hat{F}_n(u)) dH_{F_0}(u, \delta) \\ &\leq c_1 b^{-2} \int_{u \in [t_0 - b, t_0 + b]} |F_0(u) - \hat{F}_n(u)|^2 dG(u) = \mathcal{O}_p(n^{\alpha-2/3}(\log n)^2) \\ &= \mathcal{O}_p(n^{-2\alpha}) \end{aligned}$$

by (2.8) and Lemma 2.8. So we find

$$\begin{aligned} n^{2\alpha} \int \psi_{b,t_0}(u)(\delta - \hat{F}_n(u)) dH_{F_0}(u, \delta) &= n^{2\alpha} \int \bar{\psi}_{b,t_0}(u)(\delta - \hat{F}_n(u)) d(H_{F_0} - \mathbb{H}_n)(u, \delta) + \mathcal{O}_p(1) \\ &= n^{2\alpha} \int \bar{\psi}_{b,t_0}(u)(\delta - F_0(u)) d(H_{F_0} - \mathbb{H}_n)(u, \delta) \\ &\quad + n^{2\alpha} \int \bar{\psi}_{b,t_0}(u)(F_0(u) - \hat{F}_n(u)) d(G - \mathbb{G}_n)(u) + \mathcal{O}_p(1) \\ &= n^{2\alpha} \int \bar{\psi}_{b,t_0}(u)(\delta - F_0(u)) d(H_{F_0} - \mathbb{H}_n)(u, \delta) + \mathcal{O}_p(1) \end{aligned}$$

using that $n^{2\alpha}R_n = \mathcal{O}_p(1)$, Lemma 2.3 and the first result in Lemma 2.10. By the second result in Lemma 2.10, we get

$$\begin{aligned} n^{2\alpha} \int \bar{\psi}_{b,t_0}(u)(\delta - F_0(u)) d(H_{F_0} - \mathbb{H}_n)(u, \delta) &= n^{2\alpha} \int \psi_{b,t_0}(u)(\delta - F_0(u)) d(H_{F_0} - \mathbb{H}_n)(u, \delta) + \mathcal{O}_p(1). \end{aligned}$$

Applying the central limit theorem with $\alpha = 1/5$, gives

$$n^{2/5} \int \psi_{b,t_0}(u)(\delta - F_0(u)) d(\mathbb{H}_n - H_{F_0})(u, \delta) \rightsquigarrow \mathcal{N}(0, \sigma_{F,SM}^2)$$

for $\sigma_{F,SM}^2$ as in (2.11). For $0 < \alpha < 1/5$, we get

$$n^{2\alpha} \int \psi_{b,t_0}(u)(\delta - F_0(u)) d(\mathbb{H}_n - H_{F_0})(u, \delta) \xrightarrow{\mathcal{P}} 0.$$

Recall that

$$\begin{aligned} n^{2\alpha}(\hat{F}_n^{SM}(t_0) - F_0(t_0)) &= n^{2\alpha} \int \psi_{b,t_0}(u)(\delta - F_0(u)) d(\mathbb{H}_n - H_{F_0})(u, \delta) \\ &\quad + n^{2\alpha} \left(\int K_b(t_0 - u) dF_0(u) - F_0(t_0) \right). \end{aligned}$$

The second term converges to $\mu_{F,SM}$, since

$$\begin{aligned} n^{2\alpha} \left(\int K_b(t_0 - u) dF_0(u) - F_0(t_0) \right) &= n^{2\alpha} \left(\int k_b(t_0 - u) F_0(u) du - F_0(t_0) \right) \\ &= n^{2\alpha} \left(\int k_b(t_0 - u) (F_0(u) - F_0(t_0)) du \right) = \frac{1}{2} c^2 m_2(k) f_0'(t_0) + \mathcal{O}(n^{-\alpha}), \end{aligned}$$

using integration by parts and a Taylor expansion. This now implies that for $\alpha = 1/5$

$$n^{2/5}(\hat{F}_n^{SM}(t_0) - F_0(t_0)) \rightsquigarrow \mathcal{N}(\mu_{F,SM}, \sigma_{F,SM}^2),$$

while for $0 < \alpha < 1/5$

$$n^{2\alpha}(\hat{F}_n^{SM}(t_0) - F_0(t_0)) \xrightarrow{\mathcal{P}} \mu_{F,SM}.$$

To find $c_{F,SM}$ in the optimal bandwidth $b_{n,F,SM}$, we minimize the aMSE with respect to c

$$aMSE(\hat{F}_n^{SM}, c) = \frac{1}{4} c^4 m_2^2(k) f_0'(t_0)^2 + c^{-1} \frac{F_0(t_0)(1 - F_0(t_0))}{g(t_0)} \int k(u)^2 du,$$

which is standard minimization in c , yielding (2.12). $\square$

For the asymptotic distribution of $\hat{f}_n^{SM}$, we study the behavior of

$$\hat{f}_n^{SM}(t_0) - f_0(t_0) = \int k_b(t_0 - u) d(\mathbb{F}_n - F_0)(u) + \int k_b(t_0 - u) dF_0(u) - f_0(t_0).$$

For the first term in the right-hand side, a similar equality as (2.10) holds. The only difference is an extra term b^{-1} coming from the scaled version k_b of k in the definition of $\hat{f}_n^{SM}$. This additional term influences the optimal rate of n in b , which is slightly slower in this case with $n^{-1/7}$. Because of this, the rate of convergence is $n^{2/7}$ in stead of $n^{2/5}$.

Theorem 2.6 states the pointwise asymptotic distribution of $\hat{f}_n^{SM}$ and the aMSE optimal choice of the bandwidth b to estimate $\hat{f}_n^{SM}$. The proof of these results are similar as for the results of $\hat{F}_n^{SM}$, hence they are omitted.

THEOREM 2.6. *Assume F_0 and G satisfy conditions (F.1) and (G.1). Fix $t_0 > 0$ such that f_0'' is continuous at t_0 and $f_0''(t_0) \neq 0$. Let $b = cn^{-1/7}$ ($c > 0$) be the bandwidth used in the definition of $\hat{f}_n^{SM}$. Then*

$$n^{2/7}(\hat{f}_n^{SM}(t_0) - f_0(t_0)) \rightsquigarrow \mathcal{N}(\mu_{f,SM}, \sigma_{f,SM}^2),$$

where

$$\mu_{f,SM} = \frac{1}{2}c^2m_2(k)f_0''(t_0), \quad \sigma_{f,SM}^2 = \frac{F_0(t_0)(1-F_0(t_0))}{c^3g(t_0)} \int k'(u)^2 du.$$

For fixed $t_0 > 0$, the aMSE-optimal value b for estimating $\hat{f}_n^{SM}(t_0)$ is given by $b_{n,f,SM} = c_{f,SM}n^{-1/7}$, where

$$c_{f,SM} = \left\{ 3 \frac{F_0(t_0)(1-F_0(t_0))}{g(t_0)} \int k'(u)^2 du \right\}^{1/7} \left\{ m_2^2(k)f_0''(t_0)^2 \right\}^{-1/7}. \quad (2.14)$$

The hazard rate λ_0 is estimated using the plug-in method. The density f_0 and the distribution function F_0 are estimated by $\hat{f}_n^{SM}$ and $\hat{F}_n^{SM}$, respectively, to obtain the estimator $\hat{\lambda}_n^{SM}(t_0) = \hat{f}_n^{SM}(t_0)/(1 - \hat{F}_n^{SM}(t_0))$. Because $\hat{\lambda}_n^{SM}$ is a quotient, we could estimate the numerator and denominator by choosing one bandwidth $b = cn^{-1/7}$ to compute $\hat{f}_n^{SM}(t_0)$ and a different bandwidth $b_1 = c_1n^{-1/5}$ to compute $\hat{F}_n^{SM}(t_0)$. However, by the relation

$$\lambda_0(t_0) = \frac{d}{dz} - \log(1 - F_0(z)) \Big|_{z=t_0} = \frac{f_0(t_0)}{1 - F_0(t_0)}$$

it is more natural to estimate $f_0(t_0)$ and $F_0(t_0)$ with the same bandwidth. Theorem 2.7 states the asymptotic distribution if both $\hat{f}_n^{SM}$ and $\hat{F}_n^{SM}$ are computed with bandwidth $b = cn^{-1/7}$, since $\hat{f}_n^{SM}$ does not converge in distribution at rate $n^{2/5}$ if $b = cn^{-1/5}$ while $\hat{F}_n^{SM}$ does converge in probability at rate $n^{2/7}$ if $b = cn^{-1/7}$.

COROLLARY 2.7. Assume F_0 and G satisfy conditions (F.1) and (G.1). Fix $t_0 > 0$ such that $F_0(t_0) < 1$, f_0'' is continuous at t_0 and $f_0''(t_0) \neq 0$. Let $b = cn^{-1/7}$ ($c > 0$) be the bandwidth used to compute $\hat{F}_n^{SM}$ and $\hat{f}_n^{SM}$. Then

$$n^{2/7}(\hat{\lambda}_n^{SM}(t_0) - \lambda_0(t_0)) \rightsquigarrow \mathcal{N}(\mu_{\lambda,SM}, \sigma_{\lambda,SM}^2)$$

where

$$\begin{aligned} \mu_{\lambda,SM} &= \frac{\frac{1}{2}c^2 m_2(k)}{1 - F_0(t_0)} \left(f_0''(t_0) + \frac{f_0(t_0)f_0'(t_0)}{1 - F_0(t_0)} \right), \\ \sigma_{\lambda,SM}^2 &= \frac{F_0(t_0)}{c^3 g(t_0)(1 - F_0(t_0))} \int k'(u)^2 du. \end{aligned}$$

for t_0 such that $(1 - F_0(t_0))f_0''(t_0) + f_0(t_0)f_0'(t_0) \neq 0$.

For fixed $t_0 > 0$, the aMSE-optimal bandwidth b for $\hat{\lambda}_n^{SM}(t_0)$ is given by $b_{n,\lambda,SM} = \sigma_{\lambda,SM}^2 n^{-1/7}$, where

$$\begin{aligned} c_{\lambda,SM} &= \left\{ 3 \frac{F_0(t_0)}{g(t_0)(1 - F_0(t_0))} \int k'(u)^2 du \right\}^{1/7} \\ &\quad \times \left\{ \frac{m_2^2(k)}{(1 - F_0(t_0))^2} \left(f_0''(t_0) + \frac{f_0(t_0)f_0'(t_0)}{1 - F_0(t_0)} \right)^2 \right\}^{-1/7}. \end{aligned} \quad (2.15)$$

PROOF. Write $\hat{F}_n^{SM}(t_0) = F_0(t_0) + R_n$, then

$$n^{2/7}(\hat{\lambda}_n^{SM}(t_0) - \lambda_0(t_0)) = \frac{n^{2/7}}{1 - F_0(t_0)}(\hat{f}_n^{SM}(t_0) - f_0(t_0)) + T_n \quad (2.16)$$

with

$$\begin{aligned} T_n &= n^{2/7} \hat{f}_n^{SM}(t_0) \left(\frac{1}{1 - F_0(t_0) - R_n} - \frac{1}{1 - F_0(t_0)} \right) \\ &= n^{2/7} \hat{f}_n^{SM}(t_0) \frac{R_n}{(1 - F_0(t_0))(1 - F_0(t_0) - R_n)}. \end{aligned}$$

If $b = cn^{-1/7}$ is the bandwidth for $\hat{F}_n^{SM}(t_0)$, then by Theorem 2.5 $n^{2/7} R_n \xrightarrow{\mathcal{P}} \mu_{F,SM}$ and

$$T_n \xrightarrow{\mathcal{P}} \mu_{F,SM} \frac{f_0(t_0)}{(1 - F_0(t_0))^2}.$$

Since we also have that

$$\frac{n^{2/7}}{1-F_0(t_0)}(\hat{f}_n^{SM}(t_0) - f_0(t_0)) \rightsquigarrow \mathcal{N}\left(\frac{\mu_{f,SM}}{1-F_0(t_0)}, \frac{\sigma_{f,SM}^2}{(1-F_0(t_0))^2}\right)$$

we get that

$$n^{2/7}(\hat{\lambda}_n^{SM}(t_0) - \lambda_0(t_0)) \rightsquigarrow \mathcal{N}\left(\mu_{\lambda,SM}, \frac{\sigma_{f,SM}^2}{(1-F_0(t_0))^2}\right),$$

for $\mu_{\lambda,SM} = \mu_{f,SM}/(1-F_0(t_0)) + \mu_{F,SM}f_0(t_0)/(1-F_0(t_0))^2$.

The optimal c given in (2.15) is obtained by minimizing

$$\begin{aligned} aMSE(\hat{\lambda}_n^{SM}, c) = & \frac{1}{4}c^4 \frac{m_2^2(k)}{(1-F_0(t_0))^2} \left(f_0''(t_0) + \frac{f_0(t_0)f_0'(t_0)}{1-F_0(t_0)} \right)^2 \\ & + c^{-3} \frac{F_0(t_0)}{g(t_0)(1-F_0(t_0))} \int k'(u)^2 du. \end{aligned}$$

□

REMARK. If we use two different bandwidths to compute $\hat{\lambda}_n^{SM}(t_0)$, i.e. compute $\hat{F}_n^{SM}(t_0)$ with bandwidth $c_1 n^{-1/5}$ and $\hat{f}_n^{SM}(t_0)$ with bandwidth $cn^{-1/7}$, we have

$$n^{2/7}(\hat{\lambda}_n^{SM}(t_0) - \lambda_0(t_0)) \rightsquigarrow \mathcal{N}(\tilde{\mu}_{\lambda,SM}, \sigma_{\lambda,SM}^2)$$

where $\tilde{\mu}_{\lambda,SM} = \mu_{f,SM}/(1-F_0(t_0))$. This follows since now we have $n^{2/7}R_n = o_p(1)$ by Theorem 2.5 hence $T_n = o_p(1)$ in the previous proof. This means that the additional term $\mu_{F,SM}f_0(t_0)/(1-F_0(t_0))^2$ does not appear in the bias of $n^{2/7}(\hat{\lambda}_n^{SM}(t_0) - \lambda_0(t_0))$.

2.3 OPTIMAL BANDWIDTHS

Theorems 2.5 and 2.6 and Corollary 2.7 give the aMSE optimal smoothing parameters to estimate $\hat{F}_n^{SM}(t_0)$, $\hat{f}_n^{SM}(t_0)$ and $\hat{\lambda}_n^{SM}(t_0)$, respectively. Let $\vartheta_0(F)$ be the function to estimate (i.e. $F_0(t_0)$, $f_0(t_0)$ or $\lambda_0(t_0)$) and $\hat{\vartheta}(F)$ be its estimate (i.e. $\hat{F}_n^{SM}(t_0)$, $\hat{f}_n^{SM}(t_0)$ or $\hat{\lambda}_n^{SM}(t_0)$). Then, these optimal smoothing parameters can be written as $b_{n,\hat{\vartheta}(F)} = c_{\hat{\vartheta}(F)} n^{-\alpha}$ for some $\alpha > 0$ (either $1/5$ or $1/7$), where $c_{\hat{\vartheta}(F)}$ is defined as the minimizer of $aMSE(c)$ over all positive c . However, the aMSE depends on the unknown distribution F_0 , so $c_{\hat{\vartheta}(F)}$ and $b_{n,\hat{\vartheta}(F)}$ are unknown.

Several data dependent methods are known to overcome this problem by estimating the aMSE, e.g., the bootstrap method of EFRON (1979) or plug-in methods where the unknown quantities, like f_0 or f_0'' , in the aMSE are replaced by estimates [see e.g., SHEATHER (1983)]. The method used in this chapter is the smoothed bootstrap method, which is commonly used to estimate the optimal bandwidth in density-type problems, see e.g., HAZELTON (1996) and GONZÁLEZ-MANTEIGA *et al.* (1996).

For $\hat{\vartheta}(F) = \hat{F}_n^{SM}(t_0)$, the smoothed bootstrap works as follows. Let n be the sample size and $b_0 = c_0 n^{-1/5}$ an initial choice of the bandwidth. Instead of sampling from the empirical distribution (as is done in the usual empirical bootstrap) we sample $X_1^{*,1}, X_2^{*,1}, \dots, X_m^{*,1}$ ($m \leq n$) from the distribution $\hat{F}_{n,b_0}^{SM}$ (note that the initial choice b_0 is explicitly stated now in the notation of the estimator). Furthermore, we sample $T_1^{*,1}, \dots, T_m^{*,1}$ from $\hat{G}_{n,b_0}$ (a kernel estimator for the distribution function G of the censoring variable T) independent of $X_1^{*,1}, \dots, X_m^{*,1}$ and define $\Delta_i^{*,1} = \mathbb{1}_{\{X_i^{*,1} \leq T_i^{*,1}\}}$. The estimator $\hat{F}_{m,cm^{-1/5}}^{SM}(t_0)$ with bandwidth $b = cm^{-1/5}$ is determined based on the sample $(T_1^{*,1}, \Delta_1^{*,1}), \dots, (T_m^{*,1}, \Delta_m^{*,1})$. This is repeated B times, resulting in B estimators $\hat{F}_{m,cm^{-1/5}}^{SM,1}(t_0), \hat{F}_{m,cm^{-1/5}}^{SM,2}(t_0), \dots, \hat{F}_{m,cm^{-1/5}}^{SM,B}(t_0)$. The $aMSE(c)$ is then estimated by

$$\widehat{MSE}_B(c) = B^{-1} \sum_{i=1}^B (\hat{F}_{m,cm^{-1/5}}^{SM,i}(t_0) - \hat{F}_{n,b_0}(t_0))^2.$$

Let $\hat{c}_{F,SM}$ be the minimizer of $\widehat{MSE}_B(c)$ over all positive c , then we estimate the optimal bandwidth $b_{n,F,SM}$ by $\hat{b}_{n,F,SM} = \hat{c}_{F,SM} n^{-1/5}$. For the other estimators, the smoothed bootstrap works similarly.

Table 2.1 contains the values of $\hat{c}_{F,SM}$ and $\hat{b}_{n,F,SM}$ for the different choices of c_0 and two different points t_0 based on a simulation study with $n = 10000$, $m = 2000$ and $B = 500$. The table also contains the theoretical aMSE optimal value $c_{F,SM}$, given in (2.12), and the values of $\tilde{c}_{F,SM}$ using Monte Carlo simulations of size $n = 10000$ and $m = 2000$ with their corresponding values of $b_{n,F,SM}$ and $\tilde{b}_{n,F,SM}$. In the Monte Carlo simulation, we sample $X_i^{*,j}$ and $T_i^{*,j}$ ($i = 1, \dots, n$ (or m), $j = 1, \dots, B$) from the true underlying distributions F_0 and G . The aMSE is estimated by

$$\widetilde{MSE}_B(c) = B^{-1} \sum_{i=1}^B (\hat{F}_{n,cm^{-1/5}}^{SM,i}(t_0) - F_0(t_0))^2.$$

Then $\tilde{b}_{F,SM} = \tilde{c}_{F,SM} n^{-1/5}$, where $\tilde{c}_{F,SM}$ is defined as the minimizer of $\widetilde{MSE}_B(c)$ over all positive c . Figure 2.4 shows the $aMSE(c)$ for $t_0 = 4$ and its estimates $\widehat{MSE}_B(c)$ with $c_0 = 15$ and $\widetilde{MSE}_B(c)$ as a function of c .

	$t_0 = 4.0$		$t_0 = 6.5$	
	$\hat{c}_{F,SM}$	$\hat{b}_{n,F,SM}$	$\hat{c}_{F,SM}$	$\hat{b}_{n,F,SM}$
$c_0 = 5$	6.050	0.959	9.150	1.450
$c_0 = 10$	7.350	1.165	10.100	1.601
$c_0 = 15$	7.700	1.220	12.050	1.910
$c_0 = 20$	7.850	1.244	14.150	2.243
$c_0 = 25$	9.850	1.561	15.500	2.457
MC-simulation ($n = 10\,000$)	6.700	1.062	10.700	1.696
MC-simulation ($m = 2\,000$)	6.750	1.070	11.600	1.838
Theoretical value	6.467	1.025	10.426	1.652

TABLE 2.1: Minimizing values $\hat{c}_{F,SM}$ for estimating $\hat{F}_n^{SM}(t_0)$ and corresponding values of the bandwidth $\hat{b}_{n,F,SM}$ based on the smoothed bootstrap method for different values of c_0 , based on Monte Carlo simulations and the theoretical values.

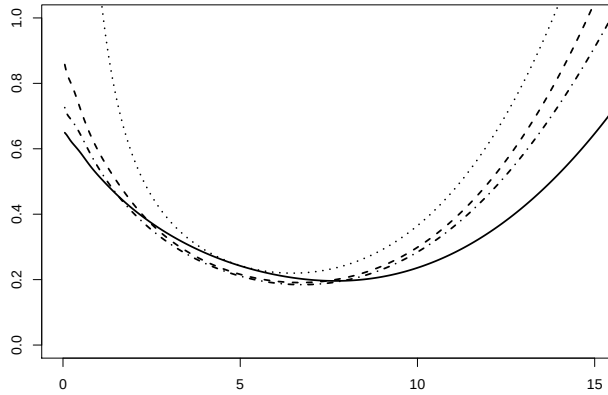


FIGURE 2.4: The theoretical aMSE of $\hat{F}_n^{SM}(4)$ (dotted line) as a function of c and its estimates based on the smoothed bootstrap (solid line) with $c_0 = 15$ and the Monte Carlo simulations with sample size $n = 10\,000$ (dashed line) and $m = 2\,000$ (dash-dotted line).

2.4 APPLICATION: HEPATITIS A

In this section, the SMLEs for F_0 , f_0 and λ_0 are computed for the hepatitis A prevalence data ($n = 850$) described by KEIDING (1991). The smoothing parameter is chosen by the smoothed bootstrap, described in Section 2.3, with $m = 425$ and $B = 500$.

Table 2.2 contains the values of $\hat{c}_{F,SM}$ and $\hat{b}_{n,F,SM}$ for three different time points, $t_0 = 20$, $t_0 = 45$ and $t_0 = 70$ and for different values of c_0 . Table 2.3 contains the values of $\hat{c}_{f,SM}$ and $\hat{b}_{n,f,SM}$, and Table 2.4 contains the values of $\hat{c}_{\lambda,SM}$ and $\hat{b}_{n,\lambda,SM}$.

Figure 2.5 shows the estimator $\hat{F}_n^{SM}$ with smoothing parameter $b = 25$ (which is somewhere in the middle of the results in Table 2.2). The maximum likelihood estimator $\hat{F}_n$ is also plotted in Figure 2.5. The estimators $\hat{f}_n^{SM}$ and $\hat{\lambda}_n^{SM}$ are shown in Figure 2.6 and 2.7, respectively. The smoothing parameter b is equal to 35 for $\hat{f}_n^{SM}$ and 40 for $\hat{\lambda}_n^{SM}$ (somewhere in the middle of the results in Tables 2.3 and 2.4, respectively)

2.5 TECHNICAL LEMMAS

LEMMA 2.8. *Let $t_0 > 0$ be such that f_0 is strictly positive and continuous in a neighborhood of t_0 . Then there exists a constant $c > 0$ such that for n sufficiently large*

$$|\bar{\psi}_{b,t_0}(u) - \psi_{b,t_0}(u)| \leq \frac{c}{b^2} |\hat{F}_n(u) - F_0(u)| \mathbb{1}_{\{|t_0 - u| \leq b\}} \quad (2.17)$$

PROOF. For n sufficiently large, we have for all $s \in \mathcal{J}_0 = [t_0 - b, t_0 + b]$ that $f_0(s) \geq \frac{1}{2}f_0(t_0)$. Fix $u \in \mathcal{J}_0$, then the interval $J_i = [\tau_i, \tau_{i+1})$ it belongs to is one of the following three types:

- (i) $F_0(x) > \hat{F}_n(\tau_i)$ for all $x \in J_i$,
- (ii) $F_0(x) = \hat{F}_n(x)$ for some $x \in J_i$,
- (iii) $F_0(x) < \hat{F}_n(\tau_i)$ for all $x \in J_i$.

First, we consider the situation where $\hat{F}_n(u) = F_0(u)$. Recall that $\bar{\psi}_{b,t_0}(u) = \psi(\hat{A}_n(u))$ with $\hat{A}_n$ defined in (2.13). Then

$$\bar{\psi}_{b,t_0}(u) = \psi_{b,t_0}(u),$$

and both the left- and right-hand side of (2.17) are equal to zero, and the upper bound holds. Note that for each $\hat{F}_n(u) = F_0(u)$ implies $\hat{A}_n(u) = u$, because F_0 is strictly increasing near t_0 .

	$t_0 = 20$		$t_0 = 45$		$t_0 = 70$	
	$\hat{c}_{F,SM}$	$\hat{b}_{n,F,SM}$	$\hat{c}_{F,SM}$	$\hat{b}_{n,F,SM}$	$\hat{c}_{F,SM}$	$\hat{b}_{n,F,SM}$
$c_0 = 50$	107.7	27.947	60.3	15.647	128.9	33.448
$c_0 = 60$	105.6	27.402	67.6	17.541	128.7	33.396
$c_0 = 70$	106.7	27.687	67.8	17.593	127.4	33.059
$c_0 = 80$	101.8	26.416	71.6	18.579	130.4	33.837
$c_0 = 90$	92.5	24.003	70.4	18.268	131.0	33.993
$c_0 = 100$	91.9	23.847	76.5	19.851	127.5	33.085
$c_0 = 110$	90.5	23.484	75.9	19.695	126.2	32.747
$c_0 = 120$	89.8	23.302	80.8	20.967	124.3	32.254
$c_0 = 130$	89.4	23.198	81.0	21.018	124.5	32.306
$c_0 = 140$	84.2	21.849	81.9	21.252	120.2	31.190
$c_0 = 150$	87.3	22.653	88.7	23.017	117.4	30.464

TABLE 2.2: Minimizing values $\hat{c}_{F,SM}$ for estimating $\hat{F}_n^{SM}(t_0)$ and corresponding values of the bandwidth $\hat{b}_{n,F,SM}$ based on the smoothed bootstrap method for different values of c_0 and for three different values of t_0 .

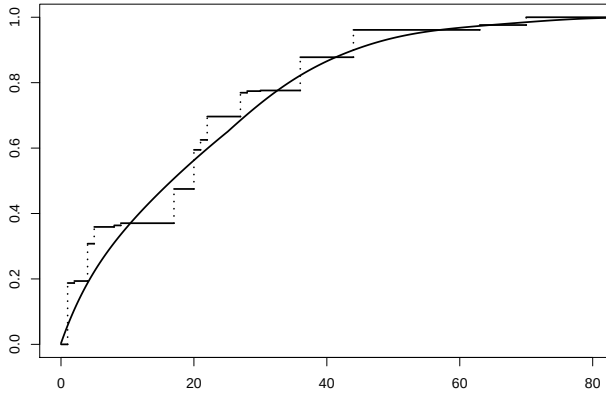


FIGURE 2.5: The estimators $\hat{F}_n^{SM}$ with $b = 25$ (solid line) and $\hat{F}_n$ (step function) for the hepatitis A prevalence data.

	$t_0 = 20$		$t_0 = 45$		$t_0 = 70$	
	$\hat{c}_{f,SM}$	$\hat{b}_{n,f,SM}$	$\hat{c}_{f,SM}$	$\hat{b}_{n,f,SM}$	$\hat{c}_{f,SM}$	$\hat{b}_{n,f,SM}$
$c_0 = 50$	150.0	57.227	91.0	34.718	119.6	45.629
$c_0 = 60$	96.1	36.664	87.6	33.421	104.5	39.868
$c_0 = 70$	98.5	37.579	80.2	30.598	91.0	34.718
$c_0 = 80$	103.3	39.410	87.0	33.192	84.3	32.162
$c_0 = 90$	113.9	43.455	91.8	35.023	82.0	31.284
$c_0 = 100$	150.0	57.227	94.3	35.977	79.9	30.483
$c_0 = 110$	150.0	57.227	101.5	38.724	81.7	31.170
$c_0 = 120$	150.0	57.227	104.3	39.792	84.0	32.047
$c_0 = 130$	150.0	57.227	106.3	40.555	87.1	33.230
$c_0 = 140$	150.0	57.227	109.7	41.852	93.3	35.595
$c_0 = 150$	150.0	57.227	114.5	43.683	93.5	35.672

TABLE 2.3: Minimizing values $\hat{c}_{f,SM}$ for estimating $\hat{f}_n^{SM}(t_0)$ and corresponding values of the bandwidth $\hat{b}_{n,f,SM}$ based on the smoothed bootstrap method for different values of c_0 and for three different values of t_0 .

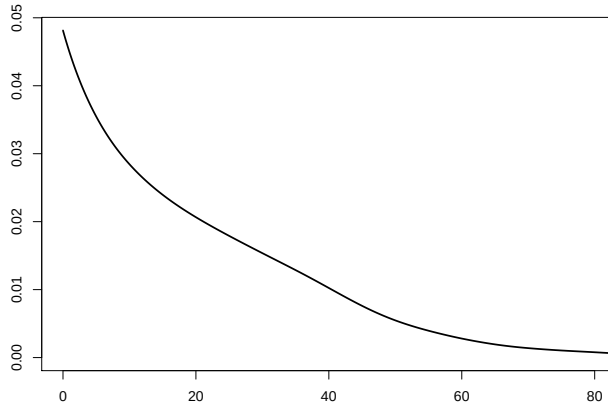


FIGURE 2.6: The estimator $\hat{f}_n^{SM}$ with $b = 35$ for the hepatitis A prevalence data.

	$t_0 = 20$		$t_0 = 45$		$t_0 = 70$	
	$\hat{c}_{\lambda,SM}$	$\hat{b}_{n,\lambda,SM}$	$\hat{c}_{\lambda,SM}$	$\hat{b}_{n,\lambda,SM}$	$\hat{c}_{\lambda,SM}$	$\hat{b}_{n,\lambda,SM}$
$c_0 = 50$	135,6	51.733	78.0	29.758	103.7	39.563
$c_0 = 60$	150.0	57.227	82.7	31.551	119.7	45.667
$c_0 = 70$	150.0	57.227	92.7	35.366	132.7	50.627
$c_0 = 80$	150.0	57.227	95.0	36.244	143.5	54.747
$c_0 = 90$	150.0	57.227	93.3	35.595	145.8	55.625
$c_0 = 100$	112.8	43.035	92.1	35.138	144.1	54.976
$c_0 = 110$	99.6	37.999	94.4	36.015	137.9	52.611
$c_0 = 120$	87.3	33.306	100.2	38.228	135.2	51.581
$c_0 = 130$	82.6	31.513	104.1	39.716	136.8	52.191
$c_0 = 140$	77.9	29.720	107.5	41.013	136.0	51.886
$c_0 = 150$	78.0	29.758	112.4	42.882	133.9	51.085

TABLE 2.4: Minimizing values $\hat{c}_{\lambda,SM}$ for estimating $\hat{\lambda}_n^{SM}(t_0)$ and corresponding values of the bandwidth $\hat{b}_{n,\lambda,SM}$ based on the smoothed bootstrap method for different values of c_0 and for three different values of t_0 .

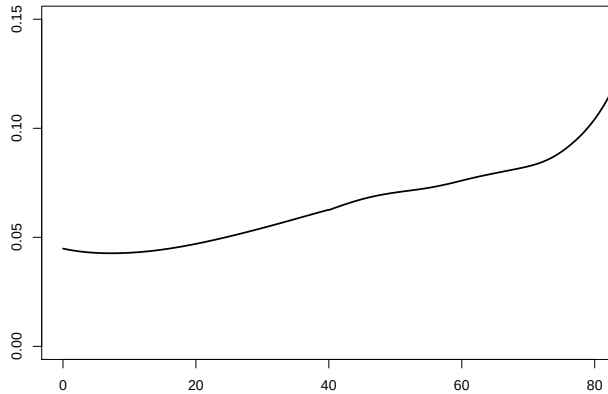


FIGURE 2.7: The estimator $\hat{\lambda}_n^{SM}$ with $b = 40$ for the hepatitis A prevalence data.

Now, we consider the situation where $\hat{F}_n(u) \neq F_0(u)$. For $v, \xi \in J_i$, we get by using a Taylor expansion

$$|\hat{F}_n(u) - F_0(u)| = |\hat{F}_n(v) - F_0(u)| = |\hat{F}_n(v) - F_0(v) - (u - v)f_0(\xi)|.$$

Now, we have three possibilities. If $\hat{A}_n(u) = \tau_i$, then we have that $F_0(\tau_i) - \hat{F}_n(\tau_i) \geq 0$ giving that

$$\begin{aligned} |\hat{F}_n(u) - F_0(u)| &= |\hat{F}_n(\tau_i) - F_0(\tau_i) - (u - \tau_i)f_0(\xi)| \\ &= |(u - \tau_i)f_0(\xi) + F_0(\tau_i) - \hat{F}_n(\tau_i)| \geq |u - \tau_i|f_0(\xi). \end{aligned}$$

If $\hat{A}_n(u) = v$ for some $v \neq u \in J_i$, then we have that $\hat{F}_n(v) = F_0(v)$, so that

$$|\hat{F}_n(u) - F_0(u)| = |\hat{F}_n(v) - F_0(u)| = |\hat{F}_n(v) - F_0(v) - (u - v)f_0(\xi)| = |u - v|f_0(\xi).$$

If $\hat{A}_n(u) = \tau_{i+1}$, then we have $\hat{F}_n(\tau_{i+1}-) - F_0(\tau_{i+1}) \geq 0$ giving that

$$\begin{aligned} |\hat{F}_n(u) - F_0(u)| &= |\hat{F}_n(\tau_{i+1}-) - F_0(\tau_{i+1}) - (u - \tau_{i+1})f_0(\xi)| \\ &= |(\tau_{i+1} - u)f_0(\xi) + \hat{F}_n(\tau_{i+1}-) - F_0(\tau_{i+1})| \\ &\geq |\tau_{i+1} - u|f_0(\xi). \end{aligned}$$

For $v \in [\tau_i, \tau_{i+1}]$, this gives

$$|\hat{F}_n(u) - F_0(u)| \geq |u - v|f_0(\xi) \geq \frac{1}{2}f_0(t_0)|u - v| \geq 0.$$

Since it also holds that

$$|\bar{\psi}_{b,t_0}(u) - \psi_{b,t_0}(u)| = |\psi_{b,t_0}(v) - \psi_{b,t_0}(u)| \leq \tilde{c}b^{-2}|v - u|$$

the upper bound in (2.17) holds if $c = 2\tilde{c}/f_0(t_0)$. $\square$

In Lemma 2.10 below, empirical process theory is used. To prove this lemma, first some definitions are introduced and a result from VAN DER VAART AND WELLNER (1996) is reformulated for the current status model.

Let $\mathcal{F}$ be a class of functions on $\mathbb{R}_+$ and $L_2(Q)$ the L_2 -norm defined by a probability measure Q on $\mathbb{R}_+$, i.e. for $g \in \mathcal{F}$

$$L_2(Q)[g] = \|g\|_{Q,2} = \left(\int_{\mathbb{R}_+} |g|^2 dQ \right)^{1/2}.$$

For any probability measure Q , let $N(\varepsilon, \mathcal{F}, L_2(Q))$ be the minimal number of balls $\{g \in \mathcal{F} : \|g - f\|_{Q,2} < \varepsilon\}$ of radius ε needed to cover the class $\mathcal{F}$. The entropy

$H(\varepsilon, \mathcal{F}, L_2(Q))$ of $\mathcal{F}$ is then defined as

$$H(\varepsilon, \mathcal{F}, L_2(Q)) = \log N(\varepsilon, \mathcal{F}, L_2(Q)),$$

and $J(\delta, \mathcal{F})$ is defined as

$$J(\delta, \mathcal{F}) = \sup_Q \int_0^\delta \sqrt{1 + H(\varepsilon, \mathcal{F}, L_2(Q))} d\varepsilon.$$

An envelope function of a function class $\mathcal{F}$ on $\mathbb{R}_+$ is any function F such that $|f(x)| \leq F(x)$ for all $x \in \mathbb{R}_+$ and $f \in \mathcal{F}$.

THEOREM 2.9. [VAN DER VAART AND WELLNER (1996, THEOREM 2.14.1)] *Let H_{F_0} be the distribution of the observable vector W and $\mathcal{F}$ be a H_{F_0} -measurable class of measurable functions with measurable envelope function F . Then*

$$\mathbb{E} \sup_{f \in \mathcal{F}} \left| \int f d\sqrt{n}(\mathbb{H}_n - H_{F_0}) \right| \lesssim J(1, \mathcal{F}) \|F\|_{H_{F_0}, 2},$$

where $\lesssim$ means $\leq$ up to a multiplicative constant.

LEMMA 2.10. *Assume F_0 and G satisfy conditions (F1) and (G.1) and let $d : [0, \infty) \times \{0, 1\} \rightarrow [-1, 1]$ be defined as $d(u, \delta) = F_0(u) - \delta$. Then for $\alpha \leq 1/5$ and $n \rightarrow \infty$*

$$R_{1,n} = n^{2\alpha} \int \bar{\psi}_{b,t_0}(u) (\hat{F}_n(u) - F_0(u)) d(\mathbb{G}_n - G)(u) = o_p(1), \quad (2.18)$$

$$R_{2,n} = n^{2\alpha} \int \{\bar{\psi}_{b,t_0}(u) - \psi_{b,t_0}(u)\} d(u, \delta) d(\mathbb{H}_n - H_{F_0})(u, \delta) = o_p(1). \quad (2.19)$$

PROOF. Define $\mathcal{J}_0 = [t_0 - \nu, t_0 + \nu]$ for some $\nu > 0$ and note that by (2.8) and (2.9) for any $\eta > 0$ we can find $M_1, M_2 > 0$ such that for all n sufficiently large

$$\mathbb{P}(\mathcal{E}_{1,n,M_1}) := \mathbb{P} \left(\sup_{u \in \mathcal{J}_0} |\hat{F}_n(u) - F_0(u)| \leq M_1 n^{-1/3} \log n \right) \geq 1 - \eta/2, \quad (2.20)$$

$$\mathbb{P}(\mathcal{E}_{2,n,M_2}) := \mathbb{P} \left(\sup_{u \in \mathcal{J}_0} |\hat{A}_n(u) - u| \leq M_2 n^{-1/3} \log n \right) \geq 1 - \eta/2. \quad (2.21)$$

Also note that $\|d\|_\infty \leq 1$. Moreover, denote by $\mathcal{A}$ the class of monotone functions on $\mathcal{J}_0$, with values in $[0, 2t_0]$. Then we know, see e.g., VAN DE GEER (2000, (2.5)), that for all $\delta > 0$

$$H(\delta, \mathcal{A}, L_2(Q)) \lesssim \delta^{-1},$$

for any probability measure Q . For the same reason, the class $\mathcal{B}_M$ of functions of bounded variation on $[0, 2t_0]$, absolutely bounded by M , has entropy function of the same order:

$$H(\delta, \mathcal{B}_M, L_2(Q)) \lesssim \delta^{-1}$$

for all $\delta > 0$.

Let us now start the main argument. Choose $\eta > 0$ and $M_1, M_2 > 0$ related to (2.20) and (2.21), correspondingly. Let $\nu_{1,n}, \nu_{2,n}$ be vanishing sequences of positive numbers and write for $i = 1, 2$

$$\begin{aligned} P(|R_{i,n}| > \nu_{i,n}) &= P(\{|R_{i,n}| > \nu_{i,n}\} \cap \mathcal{E}_{i,n,M_i}) + P(\{|R_{i,n}| > \nu_{i,n}\} \cap \mathcal{E}_{i,n,M_i}^c) \\ &\leq P(\{|R_{i,n}| > \nu_{i,n}\} \cap \mathcal{E}_{i,n,M_i}) + \eta/2 \\ &\leq \nu_{i,n}^{-1} E |R_{i,n}| \mathbb{1}_{\mathcal{E}_{i,n,M_i}} + \eta/2. \end{aligned}$$

Here, we use the Markov inequality, (2.20) and (2.21). We now concentrate on the terms $\nu_{i,n}^{-1} E |R_{i,n}| \mathbb{1}_{\mathcal{E}_{i,n,M_i}}$ for $i = 1, 2$. It will be shown that if we take e.g., $\nu_{i,n} = \varepsilon n^{-\beta_i} (\log n)^2$ for $\beta_1 = 5/6 - 7\alpha/2$ and $\beta_2 = 5/6 - 4\alpha$ and any $\varepsilon > 0$ these terms will be smaller than $\eta/2$ for all n sufficiently large, showing that $R_{i,n} = o_p(n^{-\beta_i} (\log n)^2) = o_p(1)$ for $\alpha \leq 1/5$ and $i = 1, 2$.

We start with some definitions. Define for

$$C_n(u) = \frac{k(n^\alpha(t_0 - u)/c)}{cg(u)} \mathbb{1}_{\mathcal{J}_0}(u),$$

the functions $\xi_{A,B,n}$ and $\zeta_{B,n}$ by

$$\begin{aligned} \xi_{A,B,n}(u) &= C_n(A(u))B(u), \\ \zeta_{B,n}(u, \delta) &= n^{1/3-\alpha} (\log n)^{-1} d(u, \delta) \{C_n(n^{-1/3}B(u) \log n + u) - C_n(u)\} \end{aligned}$$

and let

$$\mathcal{G}_{1,n} = \{\xi_{A,B,n} : A \in \mathcal{A}, B \in \mathcal{B}_{M_1}\}, \quad \mathcal{G}_{2,n} = \{\zeta_{B,n} : B \in \mathcal{B}_{M_2}\}.$$

Note that by condition (K.1) $|C_n(u) - C_n(v)| \leq n^\alpha \rho |u - v|$ for all $u, v \in \mathcal{J}_0$ and some constant $\rho > 0$ depending only on the kernel k , the point t_0 and the constant c in the smoothing parameter b . Also note that both classes $\mathcal{G}_{1,n}$ and $\mathcal{G}_{2,n}$ have a constant ρ_i times $\mathbb{1}_{\mathcal{J}_0}$ as envelope function, where the constant ρ_i only depend on k , t_0 , c and

M_i , $i = 1, 2$. For $\kappa_{1,n} = n^{3\alpha-5/6} \log n$ and $\kappa_{2,n} = n^{4\alpha-5/6} \log n$, we now have that

$$\begin{aligned} \mathbb{E} |R_{1,n}| \mathbb{1}_{\mathcal{E}_{1,n,M_1}} &\leq \mathbb{E} \sup_{A \in \mathcal{A}, B \in \mathcal{B}_{M_1}} \left| n^{2\alpha-1/3} \log n \int \psi(A(u)) B(u) d(\mathbb{G}_n - G)(u) \right| \\ &\quad \times \mathbb{1}_{\mathcal{E}_{1,n,M_1}} \\ &\leq \kappa_{1,n} \mathbb{E} \sup_{\xi \in \mathcal{G}_{1,n}} \left| \int \xi(u) d\sqrt{n}(\mathbb{G}_n - G)(u) \right|, \end{aligned}$$

and

$$\begin{aligned} \mathbb{E} |R_{2,n}| \mathbb{1}_{\mathcal{E}_{2,n,M_2}} &\leq \mathbb{E} \sup_{B \in \mathcal{B}_{M_2}} \left| n^{2\alpha-1/2} \int d(u, \delta) \{ \psi(n^{-1/3} B(u) \log n + u) - \psi(u) \} \right. \\ &\quad \left. d\sqrt{n}(\mathbb{H}_n - H_{F_0})(u, \delta) \right| \mathbb{1}_{\mathcal{E}_{2,n,M_2}} \\ &\leq \mathbb{E} \sup_{\zeta \in \mathcal{G}_{2,n}} \kappa_{2,n} \left| \int \zeta(u, \delta) d\sqrt{n}(\mathbb{H}_n - H_{F_0})(u, \delta) \right|. \end{aligned}$$

To bound these expectations, we use Theorem 2.9. Using the entropy results for $\mathcal{A}$ and $\mathcal{B}_M$ together with smoothness properties, we bound the entropies of the classes $\mathcal{G}_{1,n}$ and $\mathcal{G}_{2,n}$. Therefore, we fix an arbitrary probability measure Q and $\delta > 0$.

We start with the entropy of $\mathcal{G}_{1,n}$. Select a minimal $n^{-\alpha} \delta / (2\rho M_1)$ -net $A_1, \dots, A_{N_A}$ in $\mathcal{A}$ and a minimal $\delta / (2\|C_n\|_\infty)$ -net $B_1, B_2, \dots, B_{N_B}$ in $\mathcal{B}_{M_1}$ and construct the subset of $\mathcal{G}_{1,n}$ consisting of the functions $\xi_{A_i, B_j, n}$ corresponding to these nets. The number of functions in this net is then given by

$$\begin{aligned} N_A N_B &= \exp \left(H(n^{-\alpha} \delta / (2\rho M_1), \mathcal{A}, L_2(Q)) + H(\delta / (2\|C_n\|_\infty), \mathcal{B}_{M_1}, L_2(Q)) \right) \\ &\leq \exp(Cn^\alpha / \delta), \end{aligned}$$

where $C > 0$ is a constant. This set is a δ -net in $\mathcal{G}_{1,n}$. Indeed, choose a $\xi = \xi_{A_i, B_j, n} \in \mathcal{G}_{1,n}$ and denote the closest function to A in the $\mathcal{A}$ -net by A_i and similarly the function in the $\mathcal{B}_{M_1}$ -net closest to B by B_j . Then

$$\begin{aligned} \|\xi_{A,B,n} - \xi_{A_i, B_j, n}\|_{Q,2} &\leq \|C_n\|_\infty \|B(\cdot) - B_j(\cdot)\|_{Q,2} + M_1 \|C_n(A_i(\cdot)) - C_n(A(\cdot))\|_{Q,2} \\ &\leq \delta/2 + M_1 \rho n^\alpha \|A_i - A\|_{Q,2} \leq \delta \end{aligned}$$

This implies that

$$H(\delta, \mathcal{G}_{1,n}, L_2(Q)) \lesssim n^\alpha / \delta, \quad J(\delta, \mathcal{G}_{1,n}) \leq \int_0^\delta \sqrt{1 + H(\varepsilon, \mathcal{G}_{1,n}, L_2(Q))} d\varepsilon \lesssim n^{\alpha/2} \sqrt{\delta}.$$

To bound the entropy of $\mathcal{G}_{2,n}$, we select a minimal δ/ρ -net $B_1, B_2, \dots, B_N$ in $\mathcal{B}_{M_2}$ and construct the subset of $\mathcal{G}_{2,n}$ consisting of the functions $\zeta_{B_i,n}$ corresponding to this net. The number of functions in this net is then given by

$$N = \exp\left(H(\delta/\rho, \mathcal{B}_{M_2}, L_2(Q))\right) \leq \exp(C/\delta),$$

where $C > 0$ is a constant. This set is a δ -net in $\mathcal{G}_{2,n}$. Indeed, choose a $\zeta = \zeta_{B,n} \in \mathcal{G}_{2,n}$ and denote the closest function to B in the $\mathcal{B}_{M_2}$ -net by B_i , then

$$\begin{aligned} \|\zeta_{B,n} - \zeta_{B_i,n}\|_{L_2(Q)} &\leq n^{1/3-\alpha}(\log n)^{-1}\|d\|_\infty \\ &\quad \times \|C_n(n^{-1/3}B(\cdot)\log n + \cdot) - C_n(n^{-1/3}B_i(\cdot)\log n + \cdot)\|_{L_2(Q)} \\ &\leq n^{1/3-\alpha}(\log n)^{-1}n^\alpha\rho n^{-1/3}\log n\|B_i - B\|_{L_2(Q)} \leq \delta. \end{aligned}$$

This implies that

$$H(\delta, \mathcal{G}_{2,n}, L_2(Q)) \lesssim 1/\delta, \quad J(\delta, \mathcal{G}_{2,n}) \lesssim \sqrt{\delta}.$$

We now obtain via Theorem 2.9 that

$$\begin{aligned} \mathbb{E}|R_{1,n}|\mathbb{1}_{\mathcal{E}_{1,n,M_1}} &\leq \kappa_{1,n} \mathbb{E} \sup_{\xi \in \mathcal{G}_{1,n}} \left| \int \xi(u) d\sqrt{n}(\mathbb{G}_n - G)(u) \right| \lesssim \kappa_{1,n} J(1, \mathcal{G}_{1,n}) \\ &\lesssim n^{7\alpha/2-5/6} \log n \\ \mathbb{E}|R_{2,n}|\mathbb{1}_{\mathcal{E}_{2,n,M_2}} &\leq \kappa_{2,n} \mathbb{E} \sup_{\zeta \in \mathcal{G}_{2,n}} \left| \int \zeta(u, \delta) d\sqrt{n}(\mathbb{H}_n - H_{F_0})(u, \delta) \right| \lesssim \kappa_{2,n} J(1, \mathcal{G}_{2,n}) \\ &\lesssim n^{4\alpha-5/6} \log n. \end{aligned}$$

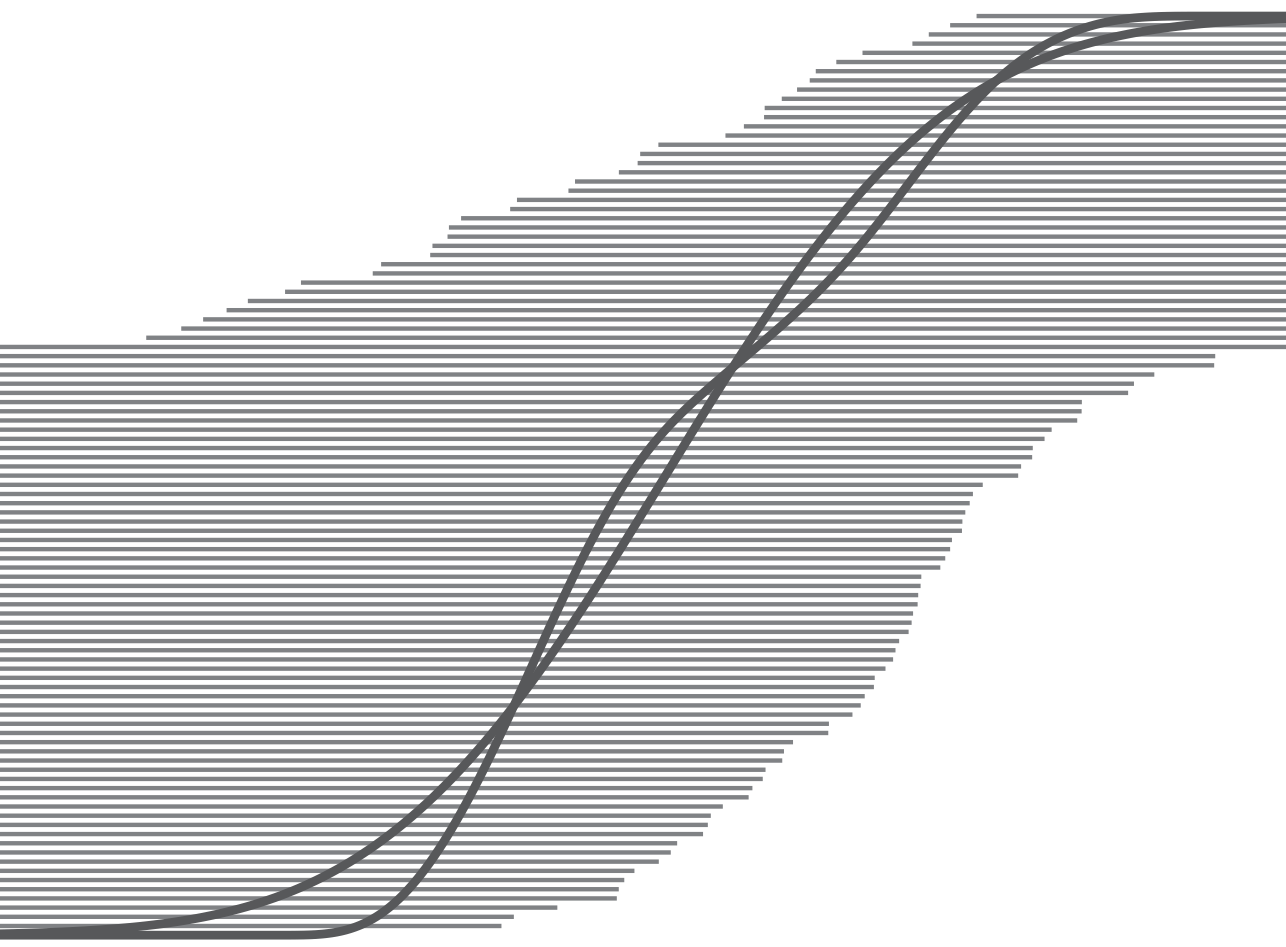
Hence, we can take $\nu_{i,n} = \varepsilon n^{-\beta_i}(\log n)^2$ for $\beta_1 = 5/6 - 7\alpha/2$, $\beta_2 = 5/6 - 4\alpha$ and any $\varepsilon > 0$ to conclude that

$$\mathbb{P}\left(\frac{n^{\beta_i}}{(\log n)^2}|R_{i,n}| > \varepsilon\right) \leq \frac{n^{\beta_i}}{\varepsilon(\log n)^2} \mathbb{E}|R_{i,n}|\mathbb{1}_{\mathcal{E}_{i,n,M_i}} + \eta/2 \lesssim \frac{1}{\varepsilon \log n} + \eta/2 < \eta,$$

for n sufficiently large. $\square$

CHAPTER 3

MAXIMUM SMOOTHED LIKELIHOOD ESTIMATION



The Smoothed Maximum Likelihood Estimator (SMLE) for the distribution function F_0 in the current status model, studied in Chapter 2, has a faster rate of convergence than the Maximum Likelihood Estimator (MLE) under some smoothness assumptions on F_0 . In this chapter, the behavior of a similar estimator, the Maximum Smoothed Likelihood Estimator (MSLE), for F_0 is studied in the current status model. The MSLE and SMLE differ in the order of smoothing and maximizing the likelihood function. The rates of convergence of both estimators are the same, while their asymptotic distributions are not equal but similar. This chapter is based on results of GROENEBOOM, JONGBLOED AND WITTE (2010).

The MSLEs for F_0 , its Lebesgue density f_0 and the hazard rate $\lambda_0 = f_0/(1 - F_0)$ are defined in Section 3.1. The characterization of the MSLE for F_0 , also given in Section 3.1, can be seen as the continuous analogue of the characterization of the MLE for F_0 given in Chapter 1. The asymptotic distribution of the estimators, derived in Section 3.2, are based on the asymptotic distributions of a naive plug-in inverse estimator. The estimation method is applied to a data set on hepatitis A from KEIDING (1991) in Section 3.3. Concluding remarks on the estimators studied in this chapter and in Chapter 2 can be found in Section 3.4. Section 3.5 contains some technical lemmas needed in the proofs in this chapter.

3.1 DEFINITION OF THE ESTIMATORS

In the current status model [see also Section 1.2] the variable of interest X , with distribution function F_0 , is not observed directly. Instead, the vector $W = (T, \Delta)$ is observed where T is a censoring variable and Δ indicates whether X is smaller than or equal to T or not. The observable vector W has density

$$h_{F_0}(t, \delta) = \delta g(t)F_0(t) + (1 - \delta)g(t)(1 - F_0(t)) = \delta h_1(t) + (1 - \delta)h_0(t),$$

with respect to the product of Lebesgue- and counting measure on $[0, \infty) \times \{0, 1\}$. Let $\mathbb{H}_n$ be the empirical distribution of an i.i.d. sample $W_1, \dots, W_n$. Then, the log likelihood as a function of F (up to the multiplicative constant $1/n$) equals

$$\ell(F) = \int \{ \delta \log F(t) + (1 - \delta) \log (1 - F(t)) \} d\mathbb{H}_n(t, \delta) \quad (3.1)$$

Here, we ignore a term in the log likelihood that does not depend on the distribution function F .

In this section, the MSLE $\hat{F}_n^{MS}$ for F_0 in the current status model is defined as well as the MSLEs $\hat{f}_n^{MS}$ and $\hat{\lambda}_n^{MS}$ for f_0 and λ_0 respectively. First we need some definitions. Define the empirical subdistribution functions of the W_j 's with $\Delta_j = 0$ and 1,

respectively by

$$\mathbb{H}_{n,i}(t) = \frac{1}{n} \sum_{j=1}^n \mathbb{1}_{[0,t] \times \{i\}}(T_j, \Delta_j), \text{ for } i = 0, 1$$

and note that the empirical distribution of the data $W_1, \dots, W_n$ can be expressed as $d\mathbb{H}_n(t, \delta) = \delta d\mathbb{H}_{n,1}(t) + (1 - \delta)d\mathbb{H}_{n,0}(t)$. Let $\hat{H}_{n,1}$ and $\hat{H}_{n,0}$ be smoothed versions of $\mathbb{H}_{n,1}$ and $\mathbb{H}_{n,0}$ respectively (e.g., via kernel smoothing), let $\hat{h}_{n,1}$ and $\hat{h}_{n,0}$ be their densities with respect to Lebesgue measure on $[0, \infty)$ and define $d\hat{H}_n(t, \delta) = \delta d\hat{H}_{n,1}(t) + (1 - \delta)d\hat{H}_{n,0}(t)$. This is a smoothed version of the empirical measure $\mathbb{H}_n$, where smoothing is only performed ‘in the t -direction’. Furthermore, note that the empirical distribution function $\mathbb{G}_n$ of all T_j ’s can be written as $\mathbb{G}_n(t) = \mathbb{H}_{n,1}(t) + \mathbb{H}_{n,0}(t)$, since

$$\mathbb{G}_n(t) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{[0,t]}(T_j) = \frac{1}{n} \sum_{j=1}^n \mathbb{1}_{[0,t] \times \{1\}}(T_j, \Delta_j) + \frac{1}{n} \sum_{j=1}^n \mathbb{1}_{[0,t] \times \{0\}}(T_j, \Delta_j),$$

and define $\hat{G}_n(t) = \hat{H}_{n,1}(t) + \hat{H}_{n,0}(t)$ as the smoothed version of $\mathbb{G}_n$. The Lebesgue density $\hat{g}_n$ satisfies $\hat{g}_n(t) = \hat{h}_{n,0}(t) + \hat{h}_{n,1}(t)$.

Following the general approach of maximum smoothed likelihood estimation as described in EGGERMONT AND LARICCIA (2001), we replace the empirical distribution $\mathbb{H}_n$ in the definition of the log likelihood (3.1) by this smoothed version $\hat{H}_n$, and define the smoothed log likelihood on the class of all distribution functions by

$$\begin{aligned} \ell^S(F) &= \int \{ \delta \log F(t) + (1 - \delta) \log (1 - F(t)) \} d\hat{H}_n(t, \delta) \\ &= \int \log (1 - F(t)) d\hat{H}_{n,0}(t) + \int \log F(t) d\hat{H}_{n,1}(t). \end{aligned} \quad (3.2)$$

The maximizer of the smoothed log likelihood is characterized similarly as the maximizer of the log likelihood [see Section 1.2]. The next theorem makes this precise.

THEOREM 3.1. *Consider the following parameterized curve in $\mathbb{R}_+^2$, a continuous cusum diagram (CCSD):*

$$t \mapsto (\hat{G}_n(t), \hat{H}_{n,1}(t)), \quad (3.3)$$

for $t \in [0, \tau]$, with $\tau = \sup\{t \geq 0 : \hat{g}_n(t) > 0\}$. Let $\hat{F}_n^{MS}(t)$ be the right-continuous slope of the lower convex hull of the CCSD (3.3), evaluated at the point with x -coordinate $\hat{G}_n(t)$. Then $\hat{F}_n^{MS}$ is the unique maximizer of (3.2) over the class of all sub-distribution functions. We call $\hat{F}_n^{MS}$ the maximum smoothed likelihood estimator of F_0 .

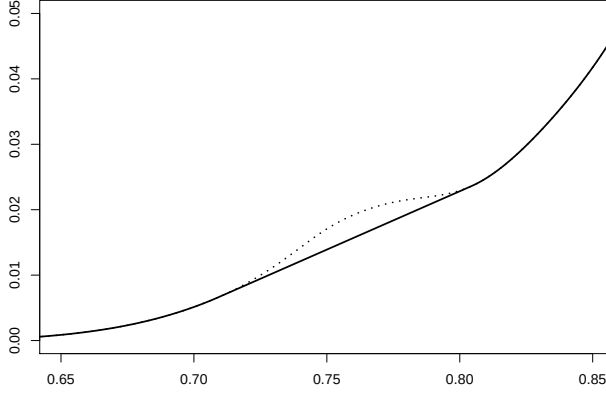


FIGURE 3.1: Part of a CCSD (dotted line) and its lower convex hull (solid line).

Figure 3.1 shows part of a CCSD and its lower convex hull.

REMARK. The MLE $\hat{F}_n$ for F_0 is defined as the right-continuous slope of the lower convex hull of the following parameterized curve in $\mathbb{R}_+^2$

$$t \mapsto (\mathbb{G}_n(t), \mathbb{H}_{n,1}(t)).$$

This curve is limit of the CCSD defined in (3.3) as $b \downarrow 0$.

PROOF OF THEOREM 3.1. Use the equality part of Lemma 3.7 to rewrite (3.2) as

$$\ell^S(\hat{F}_n^{MS}) = \int \left(\hat{F}_n^{MS}(t) \log \hat{F}_n^{MS}(t) + (1 - \hat{F}_n^{MS}(t)) \log (1 - \hat{F}_n^{MS}(t)) \right) d\hat{G}_n(t).$$

By the inequality part of Lemma 3.7, we get for each distribution function F that

$$\ell^S(F) \leq \int \hat{F}_n^{MS}(t) \log F(t) d\hat{G}_n(t) + \int (1 - \hat{F}_n^{MS}(t)) \log (1 - F(t)) d\hat{G}_n(t).$$

Now note, using the convention $0 \cdot \infty = 0$, that for all $p, p' \in [0, 1]$

$$p \log p' + (1 - p) \log(1 - p') \leq p \log p + (1 - p) \log(1 - p). \quad (3.4)$$

This implies that $\ell^S(F) \leq \ell^S(\hat{F}_n^{MS})$, i.e. ℓ^S is maximal for $\hat{F}_n^{MS}$.

Uniqueness of $\hat{F}_n^{MS}$ follows from strict concavity of $\ell^S(F)$. □

We assume the estimators $\hat{H}_{n,i}$ and $\hat{G}_n$ are continuously differentiable, hence, $\hat{F}_n^{MS}$ is continuous and its derivative exists. So we can define the MSLEs for f_0 and λ_0 by

$$\hat{f}_n^{MS}(t_0) = \frac{d}{du} \hat{F}_n^{MS}(u) \Big|_{u=t_0}, \quad \hat{\lambda}_n^{MS}(t_0) = \frac{\hat{f}_n^{MS}(t_0)}{1 - \hat{F}_n^{MS}(t_0)}, \quad (3.5)$$

for $t_0 > 0$ such that $\hat{F}_n^{MS}(t_0) < 1$.

In Theorem 3.1 no particular choice for $\hat{H}_{n,0}$ and $\hat{H}_{n,1}$ was made. For what follows, we define these estimators explicitly as kernel smoothed versions of $\mathbb{H}_{n,0}$ and $\mathbb{H}_{n,1}$. Let k be a probability density satisfying condition (K.1).

(K.1) The probability density k has support $[-1, 1]$, is symmetric and continuously differentiable.

Note that condition (K.1) implies that $m_2(k) = \int u^2 k(u) du < \infty$. Let K be the distribution function corresponding to k , i.e. $K(x) = \int_{-\infty}^x k(u) du$, k' be the derivative of k and $b = b_n > 0$ be a bandwidth (or smoothing parameter) depending on the sample size n . Then we use the following notation for the scaled version of K , k and k'

$$K_b(u) = K(u/b), \quad k_b(u) = \frac{1}{b} k(u/b), \quad \text{and} \quad k'_b(u) = \frac{1}{b^2} k'(u/b). \quad (3.6)$$

For $i = 0, 1$, let

$$\hat{h}_{n,i}(t) = \int k_b(t-u) d\mathbb{H}_{n,i}(u) = \frac{1}{n} \sum_j \mathbb{1}_{\{\Delta_j=i\}} k_b(t-T_j)$$

be kernel (sub-density) estimates based on the observations T_j for which $\Delta_j = i$, and note that $\hat{g}_n(t) = \hat{h}_{n,1}(t) + \hat{h}_{n,0}(t)$ is the usual kernel density estimator for the censoring density g . Also define the associated (sub-)distribution functions and (sub-)density derivatives by

$$\begin{aligned} \hat{H}_{n,i}(t) &= \int_0^t \hat{h}_{n,i}(u) du, \quad \text{for } i = 0, 1 \quad \text{and} \quad \hat{G}_n(t) = \int_0^t \hat{g}_n(u) du, \\ \hat{h}'_{n,i}(t) &= \int k'_b(t-u) d\mathbb{H}_{n,i}(u), \quad \text{for } i = 0, 1 \quad \text{and} \quad \hat{g}'_n(t) = \hat{h}'_{n,0}(t) + \hat{h}'_{n,1}(t). \end{aligned} \quad (3.7)$$

Because $X \geq 0$, we can expect inconsistency problems for the kernel densities near zero. In order to prevent those, we modify the definition of $\hat{h}_{n,i}$ for $t_0 < b$. To be precise, we define

$$\hat{h}_{n,i}(t_0) = \int \frac{1}{b} k^\beta\left(\frac{t-u}{b}\right) d\mathbb{H}_{n,i}(u)$$

for $0 \leq t_0 \leq b$ and $\beta = t_0/b$. The so-called boundary kernel k^β is defined by

$$k^\beta(u) = \frac{v_{2,\beta}(k) - v_{1,\beta}(k)u}{v_{0,\beta}(k)v_{2,\beta}(k) - v_{1,\beta}(k)^2} k(u) \mathbb{1}_{(-1,\beta)}(u),$$

with

$$v_{i,\beta}(k) = \int_{-1}^{\beta} u^i k(u) du, \quad i = 0, 1, 2.$$

For $0 \leq t \leq b$, the estimators $\hat{H}_{n,i}$ are defined as $\hat{H}_{n,i}(t) = \int_{-\infty}^t \hat{h}_{n,i}(u) du$ and $\hat{h}'_{n,i}$ are the derivatives of $\hat{h}_{n,i}$. There are other ways to correct the kernel estimator near the boundary, see e.g., SCHUSTER (1985) or JONES (1993). However, simulations show that the results are not much influenced by the boundary correction method used.

Having made these choices for the smoothed empirical distribution $\hat{H}_n$, let us return to the MSLE. As mentioned in Section 1.1, the MSLE can also be seen as a minimal Kullback-Leibler projection. Let $\hat{h}_n$ be the density of $\hat{H}_n$, i.e. $\hat{h}_n(t, \delta) = \delta \hat{h}_{n,1}(t) + (1 - \delta) \hat{h}_{n,0}(t)$, and define the smooth plug-in inverse estimator for F_0 by

$$\hat{F}_n^{PI}(t) = \frac{\hat{h}_{n,1}(t)}{\hat{g}_n(t)}.$$

This estimator satisfies both inverse relations

$$F_0(t) = \frac{h_1(t)}{g(t)}, \quad F_0(t) = 1 - \frac{h_0(t)}{g(t)}.$$

Furthermore, it can be seen that

$$h_{\hat{F}_n^{PI}}(t, \delta) = \delta \hat{g}_n(t) \hat{F}_n^{PI}(t) + (1 - \delta) \hat{g}_n(t) (1 - \hat{F}_n^{PI}(t)) = \hat{h}_n(t, z).$$

This implies that

$$\mathcal{K}(\hat{h}_n, h_{\hat{F}_n^{PI}}) = \mathcal{K}(\hat{h}_n, \hat{h}_n) = 0,$$

hence the smooth plug-in inverse estimator is a minimal Kullback-Leibler projection (hence a MSLE). However, $\hat{F}_n^{PI}$ might decrease locally, and therefore is a naive MSLE. In Section 3.2, we prove that asymptotically $\hat{F}_n^{PI}$ is monotonically increasing and equivalent with $\hat{F}_n^{MS}$.

Throughout this chapter, the methods are illustrated with simulated data. In these simulations, the true density f_0 is the density of a shifted Gamma(4) distribution, i.e.

$$f_0(x) = \frac{(x-2)^3}{3!} e^{-(x-2)}, \quad \text{for } x \geq 2.$$

The censoring times are exponentially distributed with mean 3, i.e.

$$g(t) = \frac{1}{3}e^{-t/3}, \text{ for } t \geq 0.$$

The kernel density is taken to be the triweight kernel $k(u) = \frac{35}{32}(1-u^2)^3$ for $-1 \leq u \leq 1$.

For a simulated sample of size $n = 500$ (the same as used in Figure 3.1), Figure 3.2 shows the estimator $\hat{F}_n^{MS}$ with bandwidth $b = 0.7$ (not obtained by the method described in Section 2.3 but chosen just as illustration). It also shows the naive plug-in inverse estimator $\hat{F}_n^{PI}$ and the true distribution function F_0 .

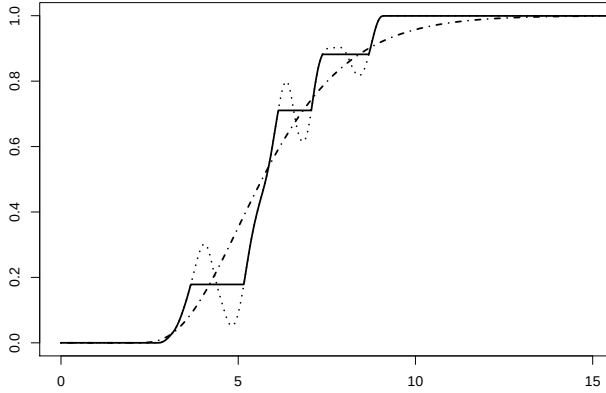


FIGURE 3.2: The true distribution function F_0 (dash-dotted line) and the estimators $\hat{F}_n^{MS}$ (solid line) and $\hat{F}_n^{PI}$ (dotted line) for a simulated sample of size $n = 500$.

3.2 ASYMPTOTIC DISTRIBUTIONS OF THE ESTIMATORS

In this section, the asymptotic distributions of $\hat{F}_n^{MS}$, $\hat{f}_n^{MS}$ and $\hat{\lambda}_n^{MS}$ are derived. To derive the asymptotic distribution of $\hat{F}_n^{MS}$, we first prove that the estimators $\hat{F}_n^{MS}$ and $\hat{F}_n^{PI}$ are pointwise asymptotically equivalent, i.e. they have the same pointwise asymptotic distribution. For these results we need the following conditions on the Lebesgue densities f_0 and g of F_0 and G .

- (F1) The Lebesgue density f_0 of F_0 has compact support $\mathcal{S}_0 = [0, M_0]$ and stays strictly away from zero.
- (G1) The Lebesgue density g of G stays strictly away from zero on $\mathcal{S}_0$ and its derivative g' is bounded on $\mathcal{S}_0$.

THEOREM 3.2. Assume F_0 and G satisfy conditions (F.1) and (G.1). Let $\hat{g}_n$ and $\hat{h}_{n,1}$ be kernel estimators for g and h_1 with kernel density k satisfying condition (K.1). Let $b = b_n$ be the bandwidth used in the definition of $\hat{g}_n$ and $\hat{h}_{n,1}$. Then for all $0 < m < M < M_0$ and b_n such that $b_n \rightarrow 0$ and $nb_n^3 \rightarrow \infty$

$$P(\hat{F}_n^{PI} \text{ is monotonically increasing on } [m, M]) \rightarrow 1.$$

PROOF. Take $0 < m < M < M_0$. By assumption (G.1) and Lemma 3.9, with probability arbitrarily close to one, we have for n sufficiently large that $\hat{g}_n(t) > 0$ for all $t \in [m, M]$. We then have that $\hat{F}_n^{PI}(t) = \hat{h}_{n,1}(t)/\hat{g}_n(t)$ is well defined on $[m, M]$ and to prove that $\hat{F}_n^{PI}(t)$ is monotonically increasing on $[m, M]$ with probability tending to one, it suffices to show that there exists a $\delta > 0$ such that for all $\eta > 0$

$$P\left(\forall t \in [m, M] : \frac{d}{dt} \hat{F}_n^{PI}(t) \geq \delta\right) \geq 1 - \eta, \quad (3.8)$$

for n sufficiently large. We have that

$$\frac{d}{dt} \hat{F}_n^{PI}(t) = \frac{\hat{g}_n(t)\hat{h}'_{n,1}(t) - \hat{h}_{n,1}(t)\hat{g}'_n(t)}{\hat{g}_n(t)^2},$$

which is also well defined.

To prove (3.8), it suffices to prove that there exists a $\delta > 0$ such that for all $\eta > 0$

$$P\left(\forall t \in [m, M] : \hat{g}_n(t)\hat{h}'_{n,1}(t) - \hat{h}_{n,1}(t)\hat{g}'_n(t) \geq \delta\right) \geq 1 - \eta, \quad (3.9)$$

for n sufficiently large. For this, we write

$$\begin{aligned} \hat{g}_n(t)\hat{h}'_{n,1}(t) - \hat{h}_{n,1}(t)\hat{g}'_n(t) &= \hat{g}_n(t)(\hat{h}'_{n,1}(t) - h'_1(t)) + \hat{h}_{n,1}(t)(g'(t) - \hat{h}'_n(t)) \\ &\quad + h'_1(t)(\hat{g}_n(t) - g(t)) + g'(t)(h_1(t) - \hat{h}_{n,1}(t)) + g(t)h'_1(t) - g'(t)h_1(t) \\ &\geq - \sup_{t \in [m, M]} |\hat{h}'_{n,1}(t) - h'_1(t)| \sup_{t \in [m, M]} \hat{g}_n(t) - \sup_{t \in [m, M]} |\hat{g}'_n(t) - g'(t)| \sup_{t \in [m, M]} \hat{h}_{n,1}(t) \\ &\quad - \sup_{t \in [m, M]} |\hat{g}_n(t) - g(t)| \sup_{t \in [m, M]} h'_1(t) - \sup_{t \in [m, M]} |\hat{h}_{n,1}(t) - h_1(t)| \sup_{t \in [m, M]} g'(t) \\ &\quad + g(t)h'_1(t) - g'(t)h_1(t). \end{aligned}$$

By Lemma 3.9 all random terms converge to zero in probability. Since $g(t)h'_1(t) - g'(t)h_1(t) = g(t)^2 f_0(t)$ we have that the last term is bounded below by $\inf_{t \in [m, M]} g(t)^2 f_0(t)$, hence (3.8) follows with $\delta < \inf_{t \in [m, M]} g(t)^2 f_0(t)$. $\square$

Figure 3.3 illustrates the naive plug-in inverse estimator $\hat{F}_n^{PI}$ for three different bandwidth $b = 5n^{-1/3}$, $b = 5n^{-1/5}$ and $b = 5n^{-1/7}$ for a simulated sample with sample size $n = 500$, illustrating Theorem 3.2.

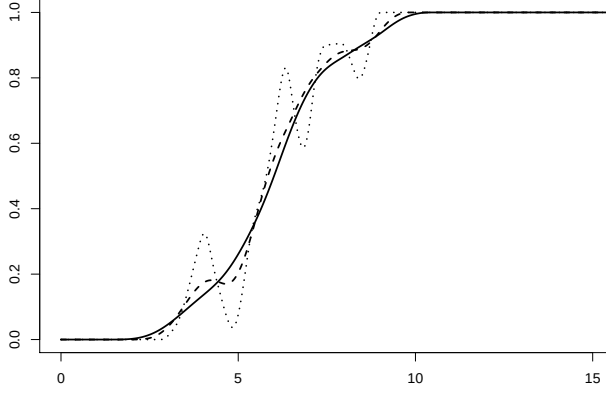


FIGURE 3.3: Illustration of the naive plug-in inverse estimator $\hat{F}_n^{PI}$ for three different values of α in the bandwidth $b = 5n^{-\alpha}$ for a simulated sample of size $n = 500$: $\alpha = 1/3$ (dotted line), $\alpha = 1/5$ (dashed line) and $\alpha = 1/7$ (solid line).

Note that this theorem as it stands does not imply that $\hat{F}_n^{MS}(t) = \hat{F}_n^{PI}(t)$ on $[m, M]$ with probability tending to one. Some additional control on the behavior of $\hat{F}_n^{PI}$ on $[0, m)$ and $(M, M_0]$ is needed. The proof of the corollary below makes this precise.

COROLLARY 3.3. *Under the assumptions of Theorem 3.2, it holds that for all $0 < m < M < M_0$ and $b = b_n$ such that $b_n \rightarrow 0$ and $nb_n^3 \rightarrow \infty$,*

$$P(\hat{F}_n^{PI}(t) = \hat{F}_n^{MS}(t) \text{ for all } t \in [m, M]) \rightarrow 1.$$

Consequently, for all $t > 0$ the asymptotic distributions of $\hat{F}_n^{MS}(t)$ and $\hat{F}_n^{PI}(t)$ are the same.

PROOF. Fix $\delta > 0$ arbitrarily. We will prove that for n sufficiently large

$$P(\hat{F}_n^{PI}(t) = \hat{F}_n^{MS}(t) \text{ for all } t \in [m, M]) \geq 1 - \delta.$$

Define for $\eta_1 \in (0, m)$, $\eta_2 \in (0, M_0 - M)$ and $n \geq 1$ the event A_n by

$$A_n = \left\{ \hat{F}_n^{PI}(t) \text{ is monotonically increasing and } \hat{g}_n(t) > 0 \text{ for } t \in [m - \eta_1, M + \eta_2] \right\}.$$

By Lemma 3.9 and Theorem 3.2, we have for all n sufficiently large $P(A_n) \geq 1 - \delta/10$.

Define the 'linearly extended $\hat{H}_{n,1}$ ' by

$$C_n^*(t) = \begin{cases} \hat{H}_{n,1}(m) + (\hat{G}_n(t) - \hat{G}_n(m))\hat{F}_n^{PI}(m), & \text{for } t \in [0, m) \\ \hat{H}_{n,1}(t), & \text{for } t \in [m, M] \\ \hat{H}_{n,1}(M) + (\hat{G}_n(t) - \hat{G}_n(M))\hat{F}_n^{PI}(M), & \text{for } t \in (M, M_0]. \end{cases}$$

It now suffices to prove that for all n sufficiently large

- (i) $P(\{(\hat{G}_n(t), C_n^*(t)) : t \geq 0\} \text{ convex}) \geq 1 - \delta/2,$
- (ii) $P(\forall t \in [0, M_0] : C_n^*(t) \leq \hat{H}_{n,1}(t)) \geq 1 - \delta/2.$

Indeed, then with probability $\geq 1 - \delta$ the curve $\{(\hat{G}_n(t), C_n^*(t)) : t \geq 0\}$ is a lower convex hull of the CCSD $\{(\hat{G}_n(t), \hat{H}_{n,1}(t)) : t \geq 0\}$ with $C_n^*(t) = \hat{H}_{n,1}(t)$ for all $t \in [m, M]$. From this, it follows that $C_n^*(t) = C_n(t)$ for all $t \in [m, M]$, hence also $C_n(t) = \hat{H}_{n,1}(t)$ for all $t \in [m, M]$. This implies that for n sufficiently large

$$P\left(\forall t \in [m, M] : \hat{F}_n^{PI}(t) = \frac{d\hat{H}_{n,1}(t)}{d\hat{G}_n(t)} = \frac{dC_n(t)}{d\hat{G}_n(t)} = \hat{F}_n^{MS}(t)\right) \geq 1 - \delta.$$

We now prove (i). For the intervals $[0, m)$ and $(M, M_0]$, the curve $\{(\hat{G}_n(t), C_n^*(t)) : t \geq 0\}$ is the tangent line of the CCSD at the points $(\hat{G}_n(m), \hat{H}_{n,1}(m))$ and $(\hat{G}_n(M), \hat{H}_{n,1}(M))$, respectively, so on the event A_n the curve is convex. This gives for n sufficiently large

$$P(\{(\hat{G}_n(t), C_n^*(t)) : t \geq 0\} \text{ convex}) \geq P(A_n) \geq 1 - \delta/10 \geq 1 - \delta/2.$$

To prove (ii), we split up the interval $[0, M_0]$ in five different intervals $\mathcal{J}_1 = [0, m - \eta_1]$, $\mathcal{J}_2 = [m - \eta_1, m]$, $\mathcal{J}_3 = [m, M]$, $\mathcal{J}_4 = (M, M + \eta_2]$ and $\mathcal{J}_5 = (M + \eta_2, M_0]$ and prove that for $1 \leq i \leq 5$

$$P(C_i) = P(\forall t \in \mathcal{J}_i : C_n^*(t) \leq \hat{H}_{n,1}(t)) \geq 1 - \delta/10. \quad (3.10)$$

For $t \in \mathcal{J}_3$, $C_n^*(t) = \hat{H}_{n,1}(t)$, hence (3.10) holds trivially. For the interval $\mathcal{J}_2$, we use that

$$\hat{H}_{n,1}(u) - \hat{H}_{n,1}(v) = (\hat{G}_n(u) - \hat{G}_n(v))\hat{F}_n^{PI}(\xi),$$

for some $\xi \in [u, v]$ (depending on u and v). This gives

$$\begin{aligned} P(\forall t \in \mathcal{J}_2 : \hat{H}_{n,1}(t) - C_n^*(t) \geq 0) \\ &= P(\forall t \in \mathcal{J}_2 : (\hat{G}_n(t) - \hat{G}_n(m))(\hat{F}_n^{PI}(\xi) - \hat{F}_n^{PI}(m)) \geq 0) \\ &= P(\forall t \in \mathcal{J}_2 : \hat{F}_n^{PI}(\xi) - \hat{F}_n^{PI}(m) \leq 0) \geq P(A_n) \geq 1 - \delta/10. \end{aligned}$$

For $\mathcal{J}_4$, we can reason similarly.

Now consider (3.10) for $i = 1$. For every $t \in \mathcal{J}_1$, we have

$$\begin{aligned} H_1(t) - H_1(m) - F_0(m)(G(t) - G(m)) &= \int_t^m (F_0(m) - F_0(u)) dG(u) \\ &\geq \int_{m-\eta_1}^m (F_0(m) - F_0(u)) dG(u). \end{aligned}$$

This means we have

$$\begin{aligned} \hat{H}_{n,1}(t) - C_n^*(t) &\geq \hat{H}_{n,1}(t) - H_1(t) + H_1(m) - \hat{H}_{n,1}(m) + F_0(m)(G(t) - \hat{G}_n(t)) \\ &\quad + F_0(m)(\hat{G}_n(m) - G(m)) + (\hat{F}_n^{PI}(m) - F_0(m))(\hat{G}_n(m) - \hat{G}_n(t)) \\ &\quad + \int_{m-\eta_1}^m (F_0(m) - F_0(u)) dG(u) \\ &\geq -2\|\hat{H}_{n,1} - H_1\|_\infty - 2\|\hat{G}_n - G\|_\infty - 2|\hat{F}_n^{PI}(m) - F_0(m)| \\ &\quad + \int_{m-\eta_1}^m (F_0(m) - F_0(u)) dG(u). \end{aligned}$$

By assumption (F1), we have $\int_{m-\eta_1}^m (F_0(m) - F_0(u)) dG(u) > 0$ so (3.10) follows for $i = 1$ by the uniform consistency of $\hat{G}_n$ and $\hat{H}_{n,1}$ [see Lemma 3.9] and the pointwise consistency of $\hat{F}_n^{PI}$.

For $i = 5$, the proof of (3.10) is similar as for $i = 1$. □

By Corollary 3.3, the estimators $\hat{F}_n^{MS}(t_0)$ and $\hat{F}_n^{PI}(t_0)$ have the same asymptotic distribution for fixed $t_0 > 0$. The pointwise asymptotic distribution of $\hat{F}_n^{PI}(t_0)$ can be derived from the Lindeberg–Feller central limit theorem and the Delta-method. The resulting pointwise asymptotic normality of both $\hat{F}_n^{MS}(t_0)$ and $\hat{F}_n^{PI}(t_0)$ is stated in the next theorem. Based on the asymptotic distribution we can determine the optimal bandwidth which minimizes the asymptotic Mean Squared Error (aMSE).

THEOREM 3.4. *Assume F_0 and G satisfy conditions (F1) and (G.1). Fix $t_0 > 0$ such that f_0'' and g'' exist and are continuous at t_0 and $g(t_0)f_0'(t_0) + 2f_0(t_0)g'(t_0) \neq 0$. Let $b = cn^{-1/5}$ ($c > 0$) be the bandwidth used in the definition of $\hat{g}_n$ and $\hat{h}_{n,1}$. Then*

$$n^{2/5}(\hat{F}_n^{MS}(t_0) - F_0(t_0)) \rightsquigarrow \mathcal{N}(\mu_{F,MS}, \sigma_{F,MS}^2)$$

where

$$\mu_{F,MS} = \frac{1}{2}c^2m_2(k) \left\{ f_0'(t_0) + 2 \frac{f_0(t_0)g'(t_0)}{g(t_0)} \right\},$$

$$\sigma_{F,MS}^2 = c^{-1} \frac{F_0(t_0)(1-F_0(t_0))}{g(t_0)} \int k(u)^2 du.$$

This also holds if we replace $\hat{F}_n^{MS}$ by $\hat{F}_n^{PI}$.

For fixed $t_0 > 0$, the aMSE-optimal bandwidth b for estimating $\hat{F}_n^{MS}(t_0)$ is given by $b_{n,F,MS} = c_{F,MS}n^{-1/5}$, where

$$c_{F,MS} = \left\{ \frac{F_0(t_0)(1-F_0(t_0))}{g(t_0)} \int k(u)^2 du \right\}^{1/5} \times \quad (3.11)$$

$$\times \left\{ m_2^2(k) \left\{ f_0'(t_0) + 2 \frac{f_0(t_0)g'(t_0)}{g(t_0)} \right\}^2 \right\}^{-1/5}.$$

PROOF. For fixed $c > 0$, the asymptotic distribution of $\hat{F}_n^{PI}$ follows immediately by applying the Delta-method with $\varphi(u, v) = v/u$ to the first result in Lemma 3.10. By Corollary 3.3, this also gives the asymptotic distribution of $\hat{F}_n^{MS}$.

To find the optimal bandwidth $b_{n,F,MS}$, we minimize the aMSE with respect to c

$$aMSE(\hat{F}_n^{MS}, c) = \frac{1}{4}c^4m_2^2(k) \left\{ f_0'(t_0) + 2 \frac{f_0(t_0)g'(t_0)}{g(t_0)} \right\}^2$$

$$+ c^{-1} \frac{F_0(t_0)(1-F_0(t_0))}{g(t_0)} \int k(u)^2 du,$$

which is standard minimization in c , yielding (3.11). $\square$

REMARK. In case t_0 is such that $g(t_0)f_0'(t_0) + 2f_0(t_0)g'(t_0) = 0$, the optimal rate of $b_{n,F,MS}$ is $n^{-1/9}$ resulting in a rate of convergence $n^{4/9}$ for $\hat{F}_n^{MS}$ as was also noted for $\hat{F}_n^{SM}$ in the remark following Theorem 2.5.

Figure 3.4 illustrates this pointwise asymptotic result by a simulation study. The value of $n^{2/5}(\hat{F}_n^{MS}(4.75) - F_0(4.75))$ was calculated for a sample of size $n = 100\,000$ with $b = 0.5$ as smoothing parameter. This was repeated $B = 500$ times. Figure 3.4 shows the QQ-plot (with the line $y = \mu_{F,MS} + \sigma_{F,MS}x$) as well as the histogram (with the $\mathcal{N}(\mu_{F,MS}, \sigma_{F,MS}^2)$ -density) of these B values.

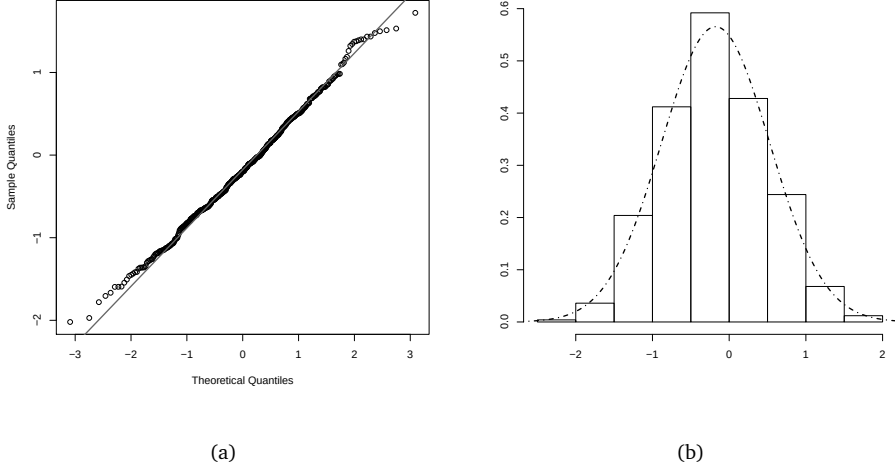


FIGURE 3.4: QQ-plot (a) and histogram (b) of $B = 500$ values of $n^{2/5}(\hat{F}_n^{MS}(4.75) - F_0(4.75))$ for a simulation study with sample size $n = 100\,000$ and $b = 0.5$.

To derive the asymptotic distribution of $\hat{f}_n^{MS}$ we use the Lebesgue density $\hat{f}_n^{PI}$ of the naive plug-in inverse estimator, which can be written as

$$\hat{f}_n^{PI}(t) = \frac{\hat{g}_n(t)\hat{h}'_{n,1}(t) - \hat{g}'_n(t)\hat{h}_{n,1}(t)}{\hat{g}_n(t)^2}.$$

THEOREM 3.5. Consider $\hat{f}_n^{MS}$ as defined in (3.5) and assume F_0 and G satisfy conditions (F.1) and (G.1). Fix $t_0 > 0$ such that $f_0^{(3)}$ and $g^{(3)}$ exist and are continuous at t_0 . Let $b = cn^{-1/7}$ ($c > 0$) be the bandwidth used to define $\hat{F}_n^{MS}$. Then

$$n^{2/7}(\hat{f}_n^{MS}(t_0) - f_0(t_0)) \rightsquigarrow \mathcal{N}(\mu_{f,MS}, \sigma_{f,MS}^2)$$

where

$$\begin{aligned} \mu_{f,MS} &= \frac{1}{2}c^2m_2(k) \left(f_0''(t_0) + 2\frac{g''(t_0)f_0(t_0) + g'(t_0)f_0'(t_0)}{g(t_0)} - 2\frac{g'(t_0)^2f_0(t_0)}{g(t_0)^2} \right) \\ &=: \frac{1}{2}c^2m_2(k)q(t_0) \\ \sigma_{f,MS}^2 &= \frac{F_0(t_0)(1 - F_0(t_0))}{c^3g(t_0)} \int k'(u)^2 du, \end{aligned}$$

for t_0 such that $q(t_0) \neq 0$. This also holds if we replace $\hat{f}_n^{MS}$ by $\hat{f}_n^{PI}$.

For fixed $t_0 > 0$, the aMSE-optimal bandwidth b for estimating $\hat{f}_n^{MS}(t_0)$ is given by $b_{n,f,MS} = c_{f,MS} n^{-1/7}$, where

$$c_{f,MS} = \left\{ 3 \frac{F_0(t_0)(1 - F_0(t_0))}{g(t_0)} \int k'(u)^2 du \right\}^{1/7} \left\{ m_2^2(k) q(t_0)^2 \right\}^{-1/7}. \quad (3.12)$$

PROOF. Write $\hat{g}_n(t_0) = g(t_0) + R_n$ and $\hat{h}_{n,1}(t_0) = h_1(t_0) + R_{n,1}$, so

$$\begin{aligned} & n^{2/7} (\hat{f}_n^{PI}(t_0) - f_0(t_0)) \\ &= n^{2/7} \left(\frac{g(t_0)\hat{h}_{n,1}'(t_0) - \hat{g}_n'(t_0)h_1(t_0)}{g(t_0)^2} - \frac{g(t_0)h_1'(t_0) - g'(t_0)h_1(t_0)}{g(t_0)^2} \right) + T_n, \end{aligned}$$

for

$$\begin{aligned} T_n &= n^{2/7} \frac{(g(t_0) + R_n)\hat{h}_{n,1}'(t_0) - \hat{g}_n'(t_0)(h_1(t_0) + R_{n,1})}{(g(t_0) + R_n)^2} \\ &\quad - n^{2/7} \frac{g(t_0)\hat{h}_{n,1}'(t_0) - \hat{g}_n'(t_0)h_1(t_0)}{g(t_0)^2} \\ &= n^{2/7} \frac{R_n\hat{h}_{n,1}'(t_0) - \hat{g}_n'(t_0)R_{n,1}}{(g(t_0) + R_n)^2} \\ &\quad - n^{2/7} (g(t_0)\hat{h}_{n,1}'(t_0) - \hat{g}_n'(t_0)h_1(t_0)) \frac{R_n(2g(t_0) + R_n)}{g(t_0)^2(g(t_0) + R_n)^2}. \end{aligned}$$

Applying the Delta-method with $\varphi(u, v) = (g(t_0)v - h_1(t_0)u)/g(t_0)^2$ to the last result in Lemma 3.10 gives that

$$n^{2/7} \left(\frac{g(t_0)\hat{h}_{n,1}'(t_0) - \hat{g}_n'(t_0)h_1(t_0)}{g(t_0)^2} - \frac{g(t_0)h_1'(t_0) - g'(t_0)h_1(t_0)}{g(t_0)^2} \right) \rightsquigarrow \mathcal{N}(\mu_1, \sigma_{f,MS}^2),$$

for

$$\mu_1 = \frac{1}{2} c^2 m_2(k) \left(f_0''(t_0) + 3 \frac{g''(t_0)f_0(t_0) + g'(t_0)f_0'(t_0)}{g(t_0)} \right). \quad (3.13)$$

By Lemma 3.10, $n^{2/7}R_n \xrightarrow{\mathcal{P}} \frac{1}{2}c^2m_2(k)g''(t_0)$ and $n^{2/7}R_{n,1} \xrightarrow{\mathcal{P}} \frac{1}{2}c^2m_2(k)h_1''(t_0)$, so by the consistency of $\hat{g}_n'$ and $\hat{h}_{n,1}'$ [see Lemma 3.9] and the continuous mapping

theorem we have

$$\begin{aligned} T_n &\xrightarrow{\mathcal{D}} \frac{1}{2}c^2m_2(k) \frac{g''(t_0)h_1'(t_0) - g'(t_0)h_1''(t_0)}{g(t_0)^2} \\ &\quad - \frac{1}{2}c^2m_2(k)(g(t_0)h_1'(t_0) - g'(t_0)h_1(t_0)) \frac{2g''(t_0)g(t_0)}{g(t_0)^4} \\ &= -\frac{1}{2}c^2m_2(k) \left(2 \frac{g'(t_0)^2 f_0(t_0)}{g(t_0)^2} + \frac{g''(t_0)f_0(t_0) + g'(t_0)f_0'(t_0)}{g(t_0)} \right) = \mu_2. \end{aligned}$$

Hence, we get

$$n^{2/7}(\hat{f}_n^{PI}(t_0) - f_0(t_0)) \rightsquigarrow \mathcal{N}(\mu_{f,MS}, \sigma_{f,MS}^2),$$

for $\mu_{f,MS} = \mu_1 + \mu_2$. By Corollary 3.3, this also gives the asymptotic distribution of $\hat{f}_n^{MS}$.

The optimal c given in (3.12) is obtained by minimizing

$$aMSE(\hat{\lambda}_n^{MS}, c) = \frac{1}{4}c^4m_2^2(k)q(t_0)^2 + c^{-3} \frac{F_0(t_0)(1 - F_0(t_0))}{g(t_0)} \int k'(u)^2 du.$$

□

REMARK. If we use two different bandwidths to compute $\hat{f}_n^{PI}$, i.e. compute $\hat{g}_n$ and $\hat{h}_{n,1}$ with bandwidth $c_1n^{-1/5}$ and $\hat{g}'_n$ and $\hat{h}'_{n,1}$ with bandwidth $cn^{-1/7}$, we have

$$n^{2/7}(\hat{f}_n^{PI}(t_0) - f_0(t_0)) \rightsquigarrow \mathcal{N}(\mu_1, \sigma_{f,MS}^2)$$

where μ_1 is given in (3.13). This follows since now we have $n^{2/7}R_n = o_p(1)$ and $n^{2/7}R_{n,1} = o_p(1)$ hence $T_n = o_p(1)$ in the previous proof. This means that the additional term μ_2 does not appear in the bias of $n^{2/7}(\hat{f}_n^{PI}(t_0) - f_0(t_0))$.

The hazard rate λ_0 is estimated using the plug-in method. The density f_0 and the distribution function F_0 are estimated by $\hat{f}_n^{MS}$ and $\hat{F}_n^{MS}$, respectively, to obtain the estimator $\hat{\lambda}_n^{MS}(t_0) = \hat{f}_n^{MS}(t_0)/(1 - \hat{F}_n^{MS}(t_0))$. Because $\hat{\lambda}_n^{MS}$ is a quotient, we could estimate the numerator and denominator by choosing one bandwidth $b = cn^{-1/7}$ to compute $\hat{f}_n^{MS}(t_0)$ and a different bandwidth $b_1 = c_1n^{-1/5}$ to compute $\hat{F}_n^{MS}(t_0)$. However, by the relation

$$\lambda_0(t_0) = \frac{d}{dz} - \log(1 - F_0(z)) \Big|_{z=t_0} = \frac{f_0(t_0)}{1 - F_0(t_0)}$$

it is more natural to estimate $f_0(t_0)$ and $F_0(t_0)$ with the same bandwidth. Theorem 3.6 states the asymptotic distribution if both $\hat{f}_n^{MS}$ and $\hat{F}_n^{MS}$ are computed with bandwidth $b = cn^{-1/7}$, since $\hat{f}_n^{MS}$ does not converge in distribution at rate $n^{2/5}$ if $b = cn^{-1/5}$ while $\hat{F}_n^{MS}$ does converge in probability at rate $n^{2/7}$ if $b = cn^{-1/7}$.

COROLLARY 3.6. *Consider $\hat{\lambda}_n^{MS}$ of λ_0 as defined in (3.5) and let $b = cn^{-1/7}$ ($c > 0$) be the bandwidth used to define $\hat{F}_n^{MS}$. Assume F_0 and G satisfy conditions (F1) and (G.1). Fix $t_0 > 0$ such that $F_0(t_0) < 1$ and $f_0^{(3)}$ and $g^{(3)}$ exist and are continuous at t_0 . Then*

$$n^{2/7}(\hat{\lambda}_n^{MS}(t_0) - \lambda_0(t_0)) \rightsquigarrow \mathcal{N}(\mu_{\lambda,MS}, \sigma_{\lambda,MS}^2)$$

where

$$\begin{aligned} \mu_{\lambda,MS} &= \frac{1}{2}c^2m_2(k)\frac{1}{1-F_0(t_0)}\left(f_0''(t_0) + 2\frac{g''(t_0)f_0(t_0) + g'(t_0)f_0'(t_0)}{g(t_0)}\right. \\ &\quad \left.- 2\frac{g'(t_0)^2f_0(t_0)}{g(t_0)^2}\right) \\ &\quad + \frac{1}{2}c^2m_2(k)\frac{f_0(t_0)}{(1-F_0(t_0))^2}\left(f_0'(t_0) + \frac{2g'(t_0)f_0(t_0)}{g(t_0)}\right) \\ &= \frac{1}{2}c^2m_2(k)r(t_0) \\ \sigma_{\lambda,MS}^2 &= \frac{F_0(t_0)}{c^3g(t_0)(1-F_0(t_0))}\int k'(u)^2du. \end{aligned}$$

for t_0 such that $r(t_0) \neq 0$. This also holds if we replace $\hat{\lambda}_n^{MS}$ by $\hat{\lambda}_n^{PI}$.

For fixed $t_0 > 0$, the aMSE-optimal bandwidth b for estimating $\hat{\lambda}_n^{MS}(t_0)$ is given by $b_{n,\lambda,MS} = c_{\lambda,MS}n^{-1/7}$, where

$$c_{\lambda,MS} = \left\{ \frac{F_0(t_0)}{g(t_0)(1-F_0(t_0))} \int k'(u)^2 du \right\}^{1/7} \left\{ m_2^2(k)r(t_0)^2 \right\}^{-1/7}. \quad (3.14)$$

PROOF. Write $\hat{F}_n^{MS}(t_0) = F_0(t_0) + R_n$, then

$$n^{2/7}(\hat{\lambda}_n^{MS}(t_0) - \lambda_0(t_0)) = \frac{n^{2/7}}{1-F_0(t_0)}(\hat{f}_n^{MS}(t_0) - f_0(t_0)) + T_n$$

with

$$\begin{aligned} T_n &= n^{2/7} \hat{f}_n^{MS}(t_0) \left(\frac{1}{1 - F_0(t_0) - R_n(t_0)} - \frac{1}{1 - F_0(t_0)} \right) \\ &= n^{2/7} \hat{f}_n^{MS}(t_0) \frac{R_n}{(1 - F_0(t_0))(1 - F_0(t_0) - R_n)}. \end{aligned}$$

If $b = cn^{-1/7}$ is the bandwidth for $\hat{F}_n^{MS}(t_0)$, then

$$n^{2/7} \left(\frac{\hat{h}_{n,1}(t_0)}{\hat{g}_n(t_0)} - \frac{h_1(t_0)}{g(t_0)} \right) \xrightarrow{\mathcal{P}} \frac{1}{2} c_2 m_2(k) \left\{ f_0'(t_0) + 2 \frac{f_0(t_0) g'(t_0)}{g(t_0)} \right\} = \mu_{F,MS},$$

by Lemma 3.10 and the Delta-method. This implies that $n^{2/7} R_n \xrightarrow{\mathcal{P}} \mu_{F,MS}$ and

$$T_n \xrightarrow{\mathcal{P}} \frac{f_0(t_0)}{(1 - F_0(t_0))^2} \mu_{F,MS}.$$

Since we also have that

$$\frac{n^{2/7}}{1 - F_0(t_0)} (\hat{f}_n^{MS}(t_0) - f_0(t_0)) \rightsquigarrow \mathcal{N} \left(\frac{\mu_{f,MS}}{1 - F_0(t_0)}, \frac{\sigma_{f,MS}^2}{(1 - F_0(t_0))^2} \right)$$

we get that

$$n^{2/7} (\hat{\lambda}_n^{MS}(t_0) - \lambda_0(t_0)) \rightsquigarrow \mathcal{N} \left(\mu_{\lambda,MS}, \frac{\sigma_{f,MS}^2}{(1 - F_0(t_0))^2} \right),$$

for $\mu_{\lambda,MS} = \mu_{f,MS}/(1 - F_0(t_0)) + \mu_{F,MS} f_0(t_0)/(1 - F_0(t_0))^2$.

The optimal c given in (3.14) is obtained by minimizing

$$aMSE(\hat{\lambda}_n^{MS}, c) = \frac{1}{4} c^4 m_2^2(k) r(t_0)^2 + c^{-3} \frac{F_0(t_0)}{g(t_0)(1 - F_0(t_0))} \int k'(u)^2 du.$$

□

REMARK. If we use two different bandwidths to compute $\hat{\lambda}_n^{MS}$, i.e. compute $\hat{F}_n^{MS}$ with bandwidth $c_1 n^{-1/5}$ and $\hat{f}_n^{MS}$ with bandwidth $cn^{-1/7}$, we have

$$n^{2/7} (\hat{\lambda}_n^{MS}(t_0) - \lambda_0(t_0)) \rightsquigarrow \mathcal{N}(\tilde{\mu}_{\lambda,MS}, \sigma_{\lambda,MS}^2)$$

where $\tilde{\mu}_{\lambda,MS} = \mu_{f,MS}/(1 - F_0(t_0))$. This follows since now we have $n^{2/7}R_n = o_p(1)$ by Lemma 3.10 and the Delta-method hence $T_n = o_p(1)$ in the previous proof. This means that the additional term $\mu_{F,MS}f_0(t_0)/(1 - F_0(t_0))^2$ does not appear in the bias of $n^{2/7}(\hat{\lambda}_n^{MS}(t_0) - \lambda_0(t_0))$.

3.3 APPLICATION: HEPATITIS A

The Bulgarian hepatitis A prevalence data described by KEIDING (1991) can be modeled by the current status model studied in this chapter. In this section, the distribution function F_0 of the time of infection with hepatitis A is estimated by the method of maximum smoothed likelihood estimation. Also the density f_0 and the hazard rate λ_0 are estimated using this method. The smoothing parameters are chosen by the smoothed bootstrap, described in Section 2.3, with $m = 425$ and $B = 500$.

For three different time points, $t_0 = 20$, $t_0 = 45$ and $t_0 = 70$ and for different values of c_0 , Table 3.1 contains the values of $\hat{c}_{F,MS}$ and $\hat{b}_{n,F,MS}$. The values of $\hat{c}_{f,MS}$ and $\hat{b}_{n,f,MS}$ for these different values of t_0 and c_0 are given in Table 3.2, while Table 3.3 contains these values of $\hat{c}_{\lambda,MS}$ and $\hat{b}_{n,\lambda,MS}$.

The resulting estimators $\hat{F}_n^{MS}$ with smoothing parameter $b = 25$, $\hat{f}_n^{MS}$ and $\hat{\lambda}_n^{MS}$ both with $b = 30$ are shown in Figure 3.5, Figure 3.6 and 3.7, respectively. The smoothing parameters b were chosen to be in the middle of the results in Tables 3.1, 3.2 and 3.3, respectively.

3.4 CONCLUDING REMARKS

In this chapter and the previous chapter, the MSLEs and the SMLEs for the distribution function F_0 , its Lebesgue density f_0 and the hazard rate $\lambda_0 = f_0/(1 - F_0)$ were studied in the current status model. For each function, the resulting estimators converges pointwise to a normal distribution. Both $\hat{F}_n^{MS}$ and $\hat{F}_n^{SM}$ converge faster to F_0 than the MLE in this model, since we assumed some smoothness of F_0 . However, the method of smoothed maximum and maximum smoothed likelihood estimation depends on a smoothing parameter b . The smoothed bootstrap was used to determine the optimal smoothing parameter, for hepatitis A prevalence data from KEIDING (1991).

Unfortunately, we cannot say the MSLE is uniformly better (i.e. always has a smaller aMSE) than the SMLE or vice versa. For example, the asymptotic bias of $\hat{F}_n^{SM}(t_0)$ is equal to

$$\mu_{F,SM}(t_0) = \frac{1}{2}c^2m_2(k)f_0'(t_0),$$

	$t_0 = 20$		$t_0 = 45$		$t_0 = 70$	
	$\hat{c}_{F,MS}$	$\hat{b}_{n,F,MS}$	$\hat{c}_{F,MS}$	$\hat{b}_{n,F,MS}$	$\hat{c}_{F,MS}$	$\hat{b}_{n,F,MS}$
$c_0 = 50$	150.0	38.923	48.6	12.611	109.3	28.362
$c_0 = 60$	150.0	38.923	53.0	13.753	115.5	29.971
$c_0 = 70$	150.0	38.923	54.7	14.194	117.1	30.386
$c_0 = 80$	109.6	28.440	53.6	13.909	118.2	30.671
$c_0 = 90$	126.8	32.903	58.9	15.284	116.2	30.152
$c_0 = 100$	119.5	31.009	53.3	13.831	109.7	28.466
$c_0 = 110$	119.5	31.009	54.6	14.168	104.6	27.142
$c_0 = 120$	121.5	31.528	56.7	14.713	103.9	26.961
$c_0 = 130$	124.7	32.358	59.3	15.388	99.8	25.897
$c_0 = 140$	111.8	29.011	62.2	16.140	95.7	24.833
$c_0 = 150$	120.5	31.268	67.8	17.593	83.4	21.641

TABLE 3.1: Minimizing values $\hat{c}_{F,MS}$ for estimating $\hat{F}_n^{MS}(t_0)$ and corresponding values of the bandwidth $\hat{b}_{n,F,MS}$ based on the smoothed bootstrap method for different values of c_0 and for three different values of t_0 .

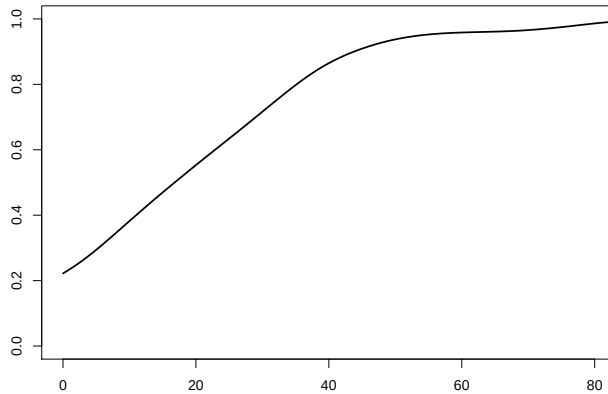


FIGURE 3.5: The estimator $\hat{F}_n^{MS}$ with $b = 25$ for the hepatitis A prevalence data.

	$t_0 = 20$		$t_0 = 45$		$t_0 = 70$	
	$\hat{c}_{f,MS}$	$\hat{b}_{n,f,MS}$	$\hat{c}_{f,MS}$	$\hat{b}_{n,f,MS}$	$\hat{c}_{f,MS}$	$\hat{b}_{n,f,MS}$
$c_0 = 50$	33.8	12.895	100.1	38.190	150.0	57.227
$c_0 = 60$	50.7	19.343	105.8	40.364	150.0	57.227
$c_0 = 70$	53.7	20.487	119.7	45.667	143.3	54.671
$c_0 = 80$	54.0	20.602	130.7	49.864	118.8	45.324
$c_0 = 90$	56.8	21.670	137.8	52.573	101.1	38.571
$c_0 = 100$	58.1	22.166	141.1	53.832	89.7	34.222
$c_0 = 110$	59.1	22.548	136.1	51.924	82.2	31.361
$c_0 = 120$	59.9	22.853	127.0	48.452	76.9	29.339
$c_0 = 130$	58.8	22.433	117.7	44.904	78.4	29.911
$c_0 = 140$	60.6	23.120	112.2	42.806	75.6	28.843
$c_0 = 150$	60.7	23.158	110.0	41.967	76.4	29.148

TABLE 3.2: Minimizing values $\hat{c}_{f,MS}$ for estimating $\hat{f}_n^{MS}(t_0)$ and corresponding values of the bandwidth $\hat{b}_{n,f,MS}$ based on the smoothed bootstrap method for different values of c_0 and for three different values of t_0 .

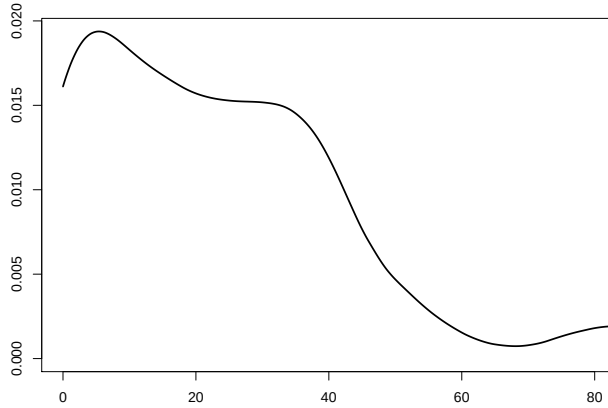


FIGURE 3.6: The estimator $\hat{f}_n^{MS}$ with $b = 30$ for the hepatitis A prevalence data.

	$t_0 = 20$		$t_0 = 45$		$t_0 = 70$	
	$\hat{c}_{\lambda,MS}$	$\hat{b}_{n,\lambda,MS}$	$\hat{c}_{\lambda,MS}$	$\hat{b}_{n,\lambda,MS}$	$\hat{c}_{\lambda,MS}$	$\hat{b}_{n,\lambda,MS}$
$c_0 = 50$	41.9	15.985	44.3	16.901	138.2	52.725
$c_0 = 60$	52.2	19.915	54.8	20.907	137.0	52.268
$c_0 = 70$	54.3	20.716	61.7	23.539	136.3	52.000
$c_0 = 80$	53.6	20.449	63.8	24.341	138.7	52.916
$c_0 = 90$	57.8	22.052	67.6	25.790	150.0	57.227
$c_0 = 100$	59.3	22.624	71.4	27.240	150.0	57.227
$c_0 = 110$	58.9	22.471	69.3	26.439	150.0	57.227
$c_0 = 120$	60.0	22.891	67.8	25.867	150.0	57.227
$c_0 = 130$	59.7	22.776	68.1	25.981	150.0	57.227
$c_0 = 140$	60.1	22.929	67.7	25.829	150.0	57.227
$c_0 = 150$	60.7	23.158	71.0	27.088	150.0	57.227

TABLE 3.3: Minimizing values $\hat{c}_{\lambda,MS}$ for estimating $\hat{\lambda}_n^{MS}(t_0)$ and corresponding values of the bandwidth $\hat{b}_{n,\lambda,MS}$ based on the smoothed bootstrap method for different values of c_0 and for three different values of t_0 .

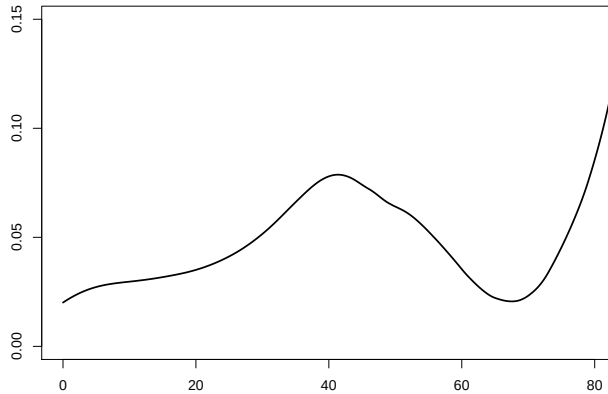


FIGURE 3.7: The estimator $\hat{\lambda}_n^{MS}$ with $b = 30$ for the hepatitis A prevalence data.

and the asymptotic bias of $\hat{F}_n^{MS}(t_0)$ is equal to

$$\mu_{F,MS}(t_0) = \frac{1}{2}c^2m_2(k) \left\{ f_0'(t_0) + 2\frac{f_0(t_0)g'(t_0)}{g(t_0)} \right\} = \mu_{F,SM} + c^2m_2(k)\frac{f_0(t_0)g'(t_0)}{g(t_0)}.$$

The additional term $c^2m_2(k)f_0(t_0)g'(t_0)/g(t_0)$ in the bias of $\hat{F}_n^{MS}$ depends on the unknown densities f_0 and g and the point t_0 we estimate at. For some choices of f_0 , g and t_0 this additional term is positive, for other choices it is negative. This implies that for some t_0 , the aMSE of $\hat{F}_n^{SM}(t_0)$ is smaller than the aMSE of $\hat{F}_n^{MS}(t_0)$ and for other t_0 it is bigger. This phenomenon is illustrated in Figure 3.8 for

$$f_0(x) = \frac{(x-2)^3}{3!}e^{-(x-2)}, \text{ for } x \geq 2, \text{ and } g(t) = \frac{1}{3}e^{-t/3}, \text{ for } t \geq 0.$$

MARRON AND PADGETT (1987) and PATIL *et al.* (1994) obtained similar results for two density estimators in the right-censoring model.

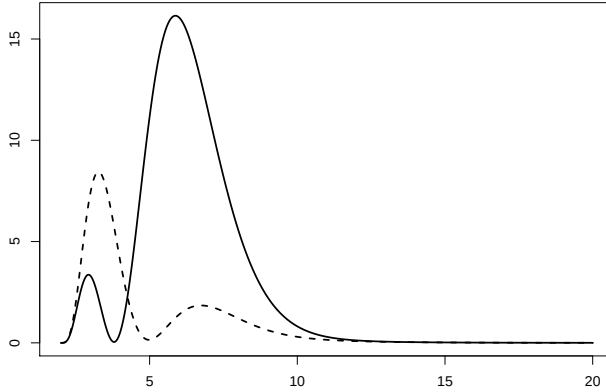


FIGURE 3.8: The aMSE of $\hat{F}_n^{MS}(t)$ (solid line) and $\hat{F}_n^{SM}(t)$ (dashed line) as function of t .

3.5 TECHNICAL LEMMAS

LEMMA 3.7. Let $\hat{F}_n^{MS}$ be defined as in Theorem 3.1. Then for any distribution function F ,

$$\begin{aligned} \int \log F(t) d\hat{H}_{n,1}(t) &\leq \int \hat{F}_n^{MS}(t) \log F(t) d\hat{G}_n(t), \text{ and} \\ \int \log(1 - F(t)) d\hat{H}_{n,0}(t) &\leq \int (1 - \hat{F}_n^{MS}(t)) \log(1 - F(t)) d\hat{G}_n(t), \end{aligned}$$

with equality in case $F = \hat{F}_n^{MS}$.

PROOF. The proof of the inequalities is based on the monotone convergence theorem. Denote the lower convex hull of the CCSD defined in (3.3) by $t \mapsto (\hat{G}_n(t), C_n(t))$ for $t \in [0, \tau]$, where $\tau = \sup\{t \geq 0 : \hat{g}_n(t) > 0\}$. By definition of this convex hull, we have for all $t > 0$

$$\hat{H}_{n,1}(t) = \int \mathbb{1}_{[0,t]}(u) d\hat{H}_{n,1}(u) \geq \int \mathbb{1}_{[0,t]}(u) dC_n(u) = \int \hat{F}_n^{MS}(u) \mathbb{1}_{[0,t]}(u) d\hat{G}_n(u).$$

For fixed t , the function $\mathbb{1}_{[0,t]}(u)$ is decreasing on $[0, \infty)$. Consider an arbitrary distribution function F on $[0, \infty)$ and write $p(t) = -\log F(t)$. Then, on $[0, \tau]$, the function p can be approximated by decreasing step functions

$$p_m(t) = \sum_{i=1}^m a_i \mathbb{1}_{[0, x_i]}(t), \text{ with } a_i \geq 0 \ \forall i \text{ and } 0 < x_1 < \dots < x_m < \tau.$$

The functions p_m can be taken such that $p_m \uparrow p$, on $[0, \tau]$. For each m , we have

$$\begin{aligned} \int p_m(t) d\hat{H}_{n,1}(t) &= \sum_{i=1}^m \int a_i \mathbb{1}_{[0, x_i]}(t) d\hat{H}_{n,1}(t) \geq \sum_{i=1}^m \int a_i \mathbb{1}_{[0, x_i]}(t) dC_n(t) \\ &= \int p_m(t) \hat{F}_n^{MS}(t) d\hat{G}_n(t). \end{aligned} \quad (3.15)$$

The monotone convergence theorem now gives that for each n

$$\begin{aligned} \lim_{m \rightarrow \infty} \int p_m(t) d\hat{H}_{n,1}(t) &= \int p(t) d\hat{H}_{n,1}(t) = - \int \log F(t) d\hat{H}_{n,1}(t), \\ \lim_{m \rightarrow \infty} \int p_m(t) dC_n(t) &= \int p(t) dC_n(t) = - \int \hat{F}_n^{MS}(t) \log F(t) d\hat{G}_n(t). \end{aligned}$$

Combined with (3.15), this implies the first inequality in Lemma 3.7.

To prove the second inequality, it suffices to prove

$$\int \log(1 - F(t)) d\hat{H}_{n,1}(t) \geq \int \hat{F}_n^{MS}(t) \log(1 - F(t)) d\hat{G}_n(t), \quad (3.16)$$

since

$$\int \log(1 - F(t)) d\hat{H}_{n,0}(t) = \int \log(1 - F(t)) d(\hat{G}_n - \hat{H}_{n,1})(t).$$

The proof of (3.16) follows by a similar argument. Then we use approximations $q_m(t)$ of the decreasing function $q(t) = \log(1 - F(t))$ such that $q_m \uparrow q$ to prove (3.16).

For the equality statements for $F = \hat{F}_n^{MS}$, we can also use the monotone approximation by step functions, restricting the jumps to the points of increase of $\hat{F}_n^{MS}$ (i.e. points x for which $\hat{F}_n^{MS}(x + \varepsilon) - \hat{F}_n^{MS}(x - \varepsilon) > 0$ for all $\varepsilon > 0$), implying equality in (3.15). $\square$

LEMMA 3.8. *Let $\hat{g}_n$ be the boundary kernel estimator for g , with smoothing parameter $b = n^{-\alpha}$ ($\alpha < 1/3$). Then with probability converging to one $\hat{g}_n$ is uniformly bounded, i.e. there exists a $C > 0$ such that*

$$\mathbb{P} \left(\sup_{x \in [0,1]} |\hat{g}_n(x)| \leq C \right) \longrightarrow 1. \quad (3.17)$$

PROOF. First note that without loss of generality we can assume $0 \leq k(u) \leq k(0)$. Recall that $v_{i,\beta}(k) = \int_{-1}^{\beta} u^i k(u) du$ for $\beta \in [0, 1]$, for which we have the following bounds

$$v_{0,\beta} \geq \frac{1}{2}, \quad |v_{1,\beta}| \leq \frac{1}{2} \mathbb{E}_k |U|, \quad \frac{1}{2} \text{Var}_k U \leq v_{2,\beta} \leq \text{Var}_k U,$$

where U has density k . Combining this, we get that $v_{0,\beta} v_{2,\beta} - v_{1,\beta}^2 \geq \frac{1}{4} \text{Var}_k |U| > 0$, so that we can uniformly bound the kernel k^β by

$$|k^\beta(u)| = \left| \frac{v_{2,\beta} - v_{1,\beta}u}{v_{0,\beta}v_{2,\beta} - v_{1,\beta}^2} k(u) 1_{[-1,\beta]}(u) \right| \leq \frac{|v_{2,\beta}| + |v_{1,\beta}|}{\frac{1}{4} \text{Var}_k |U|} k(0) = ck(0).$$

For the boundary kernel estimate $\hat{g}_n$, we then have

$$\begin{aligned} |\hat{g}_n(x)| &= \left| b^{-1} \int k^\beta((x-y)/b) d\mathbb{G}_n(y) \right| \leq b^{-1} ck(0) |\mathbb{G}_n(x+b) - \mathbb{G}_n(x-b)| \\ &\leq b^{-1} ck(0) \left\{ |\mathbb{G}_n(x+b) - G(x+b) - \mathbb{G}_n(x-b) + G(x-b)| \right. \\ &\quad \left. + (G(x+b) - G(x-b)) \right\} \\ &\leq ck(0) n^{\alpha-1/2} 2 \sup_{y \geq 0} \sqrt{n} |\mathbb{G}_n(y) - G(y)| + 2 \|g\|_\infty ck(0) \\ &= \mathcal{O}_p(n^{\alpha-1/2}) + 2 \|g\|_\infty ck(0). \end{aligned}$$

Since this bound is uniform in x , (3.17) follows for $C = 3 \|g\|_\infty ck(0)$. $\square$

LEMMA 3.9. Assume F_0 and g satisfy conditions (F.1) and (G.1) and let $\hat{G}_n$, $\hat{H}_{n,1}$, $\hat{g}_n$, $\hat{h}_{n,1}$, $\hat{g}'_n$ and $\hat{h}'_{n,1}$ be kernel estimators for G , H_1 , g , h_1 , g' and h'_1 with kernel density k satisfying condition (K.1) and bandwidth $b = b_n$. For b_n such that $nb_n^3 \rightarrow \infty$,

$$\sup_{t \in [m, \infty)} |\hat{g}_n(t) - g(t)| \xrightarrow{\mathcal{P}} 0, \quad \sup_{t \in [m, \infty)} |\hat{g}'_n(t) - g'(t)| \xrightarrow{\mathcal{P}} 0, \quad (3.18)$$

$$\sup_{t \in [0, 2M_0]} |\hat{G}_n(t) - G(t)| \xrightarrow{\mathcal{P}} 0$$

$$\sup_{t \in [m, \infty)} |\hat{h}_{n,1}(t) - h_1(t)| \xrightarrow{\mathcal{P}} 0, \quad \sup_{t \in [m, \infty)} |\hat{h}'_{n,1}(t) - h'_1(t)| \xrightarrow{\mathcal{P}} 0, \quad (3.19)$$

$$\sup_{t \in [0, 2M_0]} |\hat{H}_{n,1}(t) - H_1(t)| \xrightarrow{\mathcal{P}} 0.$$

PROOF. Let $\hat{g}_n^u$ be the uncorrected kernel estimate for g and note that by properties of the boundary kernel estimator we have for all $x \geq h$

$$\hat{g}_n^u(x) = \hat{g}_n(x).$$

Hence, the first two results in (3.18) follow immediately from Theorems A.2 and A.3. To prove the third result in (3.18), fix $M > M_0$, $\varepsilon > 0$ and choose $0 < \delta < \varepsilon/(2C)$ such that $G(\delta) < \varepsilon/4$, where C is such that (3.17) holds. For all $x \geq 0$ and n sufficiently large (such that $b < \delta$), we then have

$$|\hat{G}_n(x) - G(x)| \leq \delta \sup_{y \in [0, \delta]} |\hat{g}_n(y)| + G(\delta) + \sup_{y \geq \delta} |\hat{G}_n(y) - G(y)|$$

The right-hand side does not depend on x so that

$$\begin{aligned}
 P(\|\hat{G}_n - G\|_\infty > \varepsilon) &\leq P\left(\delta \sup_{y \in [0, \delta]} |\hat{g}_n(y)| + G(\delta) + \sup_{y \geq \delta} |\hat{G}_n^u(y) - G(y)| > \varepsilon\right) \\
 &\leq P\left(\delta \sup_{y \in [0, 1]} |\hat{g}_n(y)| + G(\delta) + \sup_{y \geq \delta} |\hat{G}_n^u(y) - G(y)| > \varepsilon\right) \\
 &= P\left(\left\{\delta \sup_{y \in [0, 1]} |\hat{g}_n(y)| + G(\delta) + \sup_{y \geq \delta} |\hat{G}_n^u(y) - G(y)| > \varepsilon\right\} \right. \\
 &\quad \left. \cap \left\{\sup_{y \in [0, 1]} |\hat{g}_n(y)| \leq C\right\}\right) \\
 &\quad + P\left(\left\{\delta \sup_{y \in [0, 1]} |\hat{g}_n(y)| + G(\delta) + \sup_{y \geq \delta} |\hat{G}_n^u(y) - G(y)| > \varepsilon\right\} \right. \\
 &\quad \left. \cap \left\{\sup_{y \in [0, 1]} |\hat{g}_n(y)| > C\right\}\right) \\
 &\leq P\left(\sup_{y \geq \delta} |\hat{G}_n^u(y) - G(y)| > \varepsilon/4\right).
 \end{aligned}$$

The last probability converges to zero as a consequence of Theorem A.2, hence the last result in (3.18) follows.

For the first result in (3.19), define a binomially distributed random variable $N_1 = \sum_{i=1}^n \Delta_i$ with parameters n and $p = P(\Delta_1 = 1) = \int F_0(u)g(u)du$, and the probability density $\tilde{g}(t) = h_1(t)/p$. Let $V_1, \dots, V_{N_1}$ be the T_i such that $\Delta_i = 1$, and rewrite $\hat{h}_{n,1}(t)$ as $\frac{1}{nb} \sum_{i=1}^{N_1} k_b(t - V_i) = \frac{N_1}{n} \hat{g}_{N_1}(t)$. Then we have by the triangle inequality

$$\|\hat{h}_{n,1} - h_1\|_\infty = \|\hat{g}_{N_1} \frac{N_1}{n} - \tilde{g}p\|_\infty \leq p\|\hat{g}_{N_1} - \tilde{g}\|_\infty + \|\hat{g}_{N_1}\|_\infty \left| \frac{N_1}{n} - p \right|.$$

The first term on the right-hand side converges to zero in probability by Theorem A.2, since $N_1 \xrightarrow{\mathcal{P}} \infty$ as $n \rightarrow \infty$. For the second term on the right-hand side, note that

$$\|\hat{g}_{N_1}\|_\infty = \|\tilde{g} + \hat{g}_{N_1} - \tilde{g}\|_\infty \leq \|\tilde{g}\|_\infty + \|\hat{g}_{N_1} - \tilde{g}\|_\infty,$$

where the last term again converges to zero in probability by Theorem A.2. Combining this with the Law of Large Numbers applied to $\left| \frac{N_1}{n} - p \right|$ gives that $\|\hat{g}_{N_1}\|_\infty \left| \frac{N_1}{n} - p \right| \xrightarrow{\mathcal{P}} 0$ as $n \rightarrow \infty$, hence $\|\hat{h}_{n,1} - h_1\|_\infty \xrightarrow{\mathcal{P}} 0$.

The proofs of the other results in (3.19) are similar. □

LEMMA 3.10. *Under the assumptions of Theorem 3.4 it holds that*

$$n^{2/5} \left(\begin{pmatrix} \hat{g}_n(t_0) \\ \hat{h}_{n,1}(t_0) \end{pmatrix} - \begin{pmatrix} g(t_0) \\ h_1(t_0) \end{pmatrix} \right) \rightsquigarrow \mathcal{N} \left(\begin{pmatrix} \frac{1}{2}c^2m_2(k)g''(t_0) \\ \frac{1}{2}c^2m_2(k)h_1''(t_0) \end{pmatrix}, \Sigma_1 \right), \quad (3.20)$$

with

$$\Sigma_1 = c^{-1} \int k(u)^2 du \begin{pmatrix} g(t_0) & h_1(t_0) \\ h_1(t_0) & h_1(t_0) \end{pmatrix}. \quad (3.21)$$

For $0 < \alpha < 1/5$,

$$\begin{aligned} n^{2\alpha}(\hat{g}_n(t_0) - g(t_0)) &\xrightarrow{\mathcal{P}} \frac{1}{2}c^2m_2(k)g''(t_0), \\ n^{2\alpha}(\hat{h}_{n,1}(t_0) - h_1(t_0)) &\xrightarrow{\mathcal{P}} \frac{1}{2}c^2m_2(k)h_1''(t_0). \end{aligned}$$

Under the assumptions of Theorem 3.5 it holds that

$$n^{2/7} \left(\begin{pmatrix} \hat{g}'_n(t_0) \\ \hat{h}'_{n,1}(t_0) \end{pmatrix} - \begin{pmatrix} g'(t_0) \\ h_1'(t_0) \end{pmatrix} \right) \rightsquigarrow \mathcal{N} \left(\begin{pmatrix} \frac{1}{2}c^2m_2(k)g^{(3)}(t_0) \\ \frac{1}{2}c^2m_2(k)h_1^{(3)}(t_0) \end{pmatrix}, \Sigma_2 \right), \quad (3.22)$$

with

$$\Sigma_2 = c^{-3} \int k'(u)^2 du \begin{pmatrix} g(t_0) & h_1(t_0) \\ h_1(t_0) & h_1(t_0) \end{pmatrix}.$$

PROOF. We start with the proof of (3.20). Define

$$Y_i = \begin{pmatrix} Y_{i;1} \\ Y_{i;2} \end{pmatrix} = n^{-3/5} \begin{pmatrix} k_b(t_0 - T_i) \\ k_b(t_0 - T_i)\Delta_i \end{pmatrix}.$$

By the assumptions on F_0 and g and condition (K.1), we have

$$\begin{aligned} \mathbb{E} Y_{i;1} &= n^{-3/5}g(t_0) + \frac{1}{2}n^{-1}c^2m_2(k)g''(t_0) + o(n^{-1}), \\ \text{Var } Y_{i;1} &= n^{-1}c^{-1}g(t_0) \int k(u)^2 du + o(n^{-6/5}), \\ \mathbb{E} Y_{i;2} &= n^{-3/5}h_1(t_0) + \frac{1}{2}n^{-1}c^2m_2(k)h_1''(t_0) + o(n^{-1}), \\ \text{Var } Y_{i;2} &= n^{-1}c^{-1}h_1(t_0) \int k(u)^2 du + o(n^{-6/5}). \end{aligned}$$

Furthermore we have

$$\text{Cov}(Y_{i;1}, Y_{i;2}) = n^{-1}c^{-1}h_1(t_0) \int k(u)^2 du + \mathcal{O}(n^{-6/5}),$$

so that

$$\begin{aligned} \sum_{i=1}^n \mathbb{E} Y_i &= n^{2/5} \begin{pmatrix} g(t_0) \\ h_1(t_0) \end{pmatrix} + \frac{1}{2}c^2m_2(k) \begin{pmatrix} g''(t_0) \\ h_1''(t_0) \end{pmatrix} + \mathcal{O}(1), \\ \sum_{i=1}^n \text{Var } Y_i &= c^{-1} \int k(u)^2 du \begin{pmatrix} g(t_0) & h_1(t_0) \\ h_1(t_0) & h_1(t_0) \end{pmatrix} + \mathcal{O}(n^{-1/5}). \end{aligned}$$

Here we denote by $\text{Var } Y_i$ the covariance matrix of the vector Y_i . By the Lindeberg–Feller central limit theorem, we get

$$n^{2/5} \left(\begin{pmatrix} \hat{g}_n(t_0) \\ \hat{h}_{n,1}(t_0) \end{pmatrix} - \begin{pmatrix} g(t_0) \\ h_1(t_0) \end{pmatrix} \right) - \frac{1}{2}c^2m_2(k) \begin{pmatrix} g''(t_0) \\ h_1''(t_0) \end{pmatrix} \rightsquigarrow \mathcal{N}(0, \Sigma_1)$$

where Σ_1 is defined in (3.21).

To prove that $n^{2\alpha}(\hat{g}_n(t_0) - g(t_0)) \xrightarrow{\mathcal{P}} \frac{1}{2}c^2m_2(k)g''(t_0)$ for $0 < \alpha < 1/5$, define $W_i = n^{2\alpha-1}k_b(t_0 - T_i)$. Since we have

$$\begin{aligned} \mathbb{E} W_i &= n^{2\alpha-1} \left(g(t_0) + \frac{1}{2}b^2g''(t_0) + \mathcal{O}(b^2) \right) \\ n \text{Var } W_i &= n^{5\alpha-1}c^{-1}g(t_0) \int k(u)^2 du + \mathcal{O}(n^{4\alpha-1}) = \mathcal{O}(n^{5\alpha-1}), \end{aligned}$$

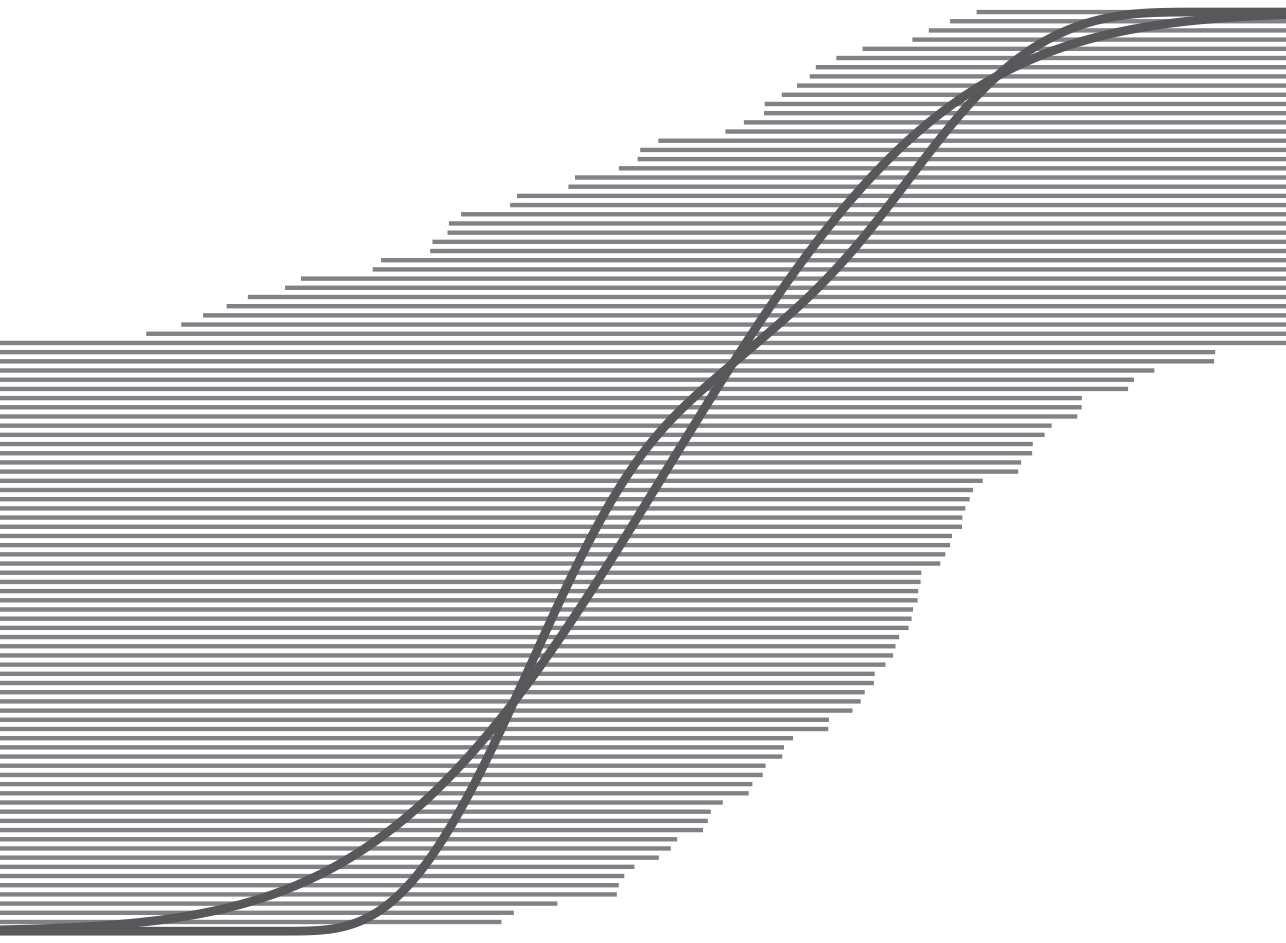
we get that $\sum \text{Var } W_i \rightarrow 0$ for $0 < \alpha < 1/5$, hence

$$n \left(\frac{1}{n} \sum_{i=1}^n W_i - \mathbb{E} W_1 \right) = n^{2\alpha}(\hat{g}_n(t_0) - g(t_0)) - \frac{1}{2}c^2m_2(k)g''(t_0) + \mathcal{O}(1) \xrightarrow{\mathcal{P}} 0.$$

Similarly we can prove that $n^{2\alpha}(\hat{h}_{n,1}(t_0) - h_1(t_0)) \xrightarrow{\mathcal{P}} \frac{1}{2}c^2m_2(k)h_1''(t_0)$.

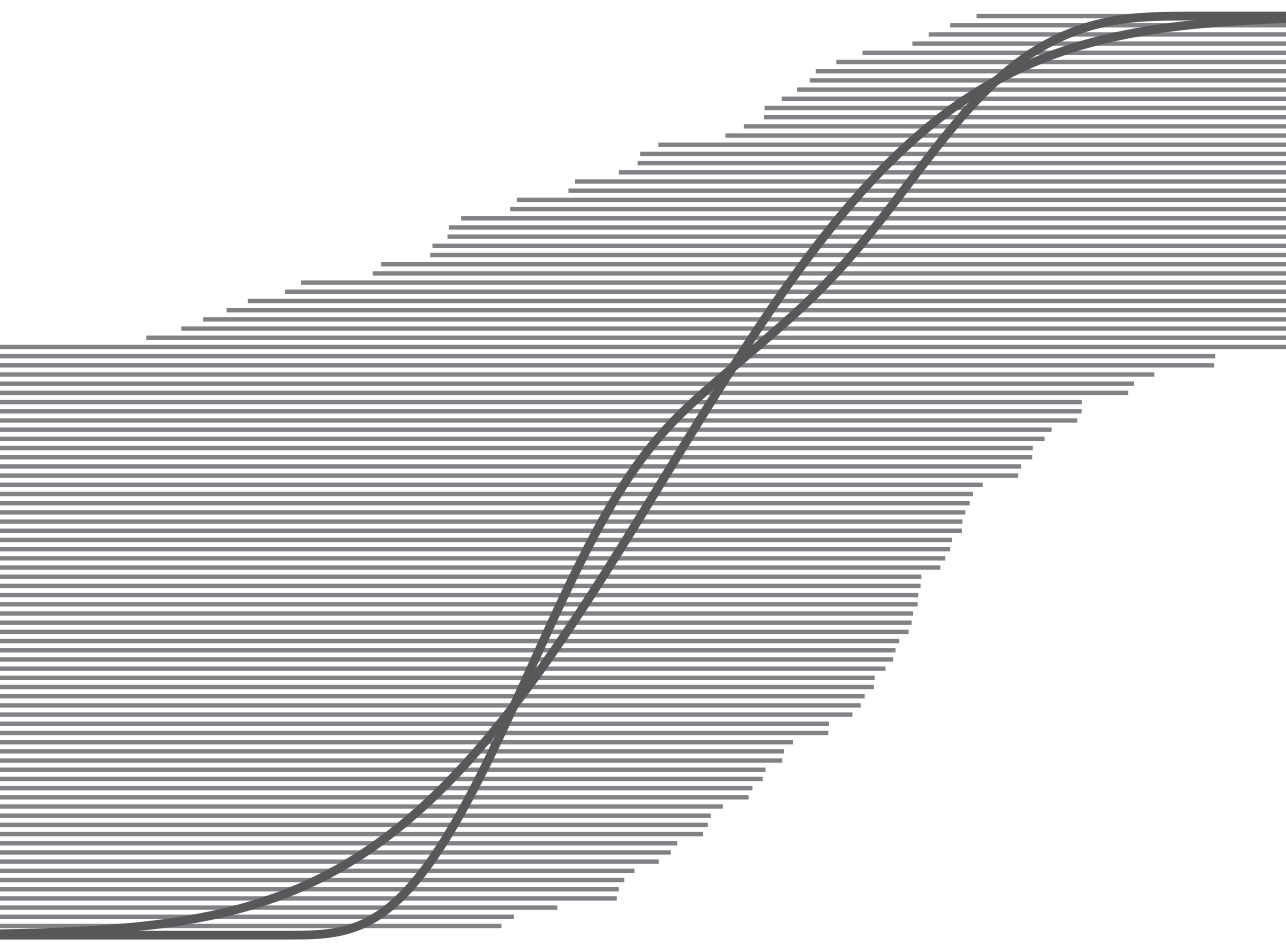
The proof of (3.22) is similar as the proof of (3.20). $\square$

CURRENT STATUS CONTINUOUS MARK MODEL



CHAPTER 4

MAXIMUM SMOOTHED LIKELIHOOD ESTIMATION



Chapters 2 and 3 describe estimators for the distribution function F_0 in the current status model. These estimators attain a faster rate of convergence than the Maximum Likelihood Estimator (MLE), under some smoothness assumptions. In this chapter, the Maximum Smoothed Likelihood Estimator (MSLE) for F_0 in the current status continuous mark model is studied. This estimator is better than the MLE as well; in contrast to the MLE it is consistent. This chapter is based on results of GROENEBOOM, JONGBLOED AND WITTE (2011a).

The outline of this chapter is as follows. Section 4.1 defines the MSLE for F_0 , as well as the MSLEs for the Lebesgue density f_0 and the marginal distribution function $F_{0,X}$ of X . The estimators used for smoothing are bivariate histogram estimators, and the resulting MSLEs for F_0 and $F_{0,X}$ are proved to be pointwise consistent in Section 4.3. Section 4.2 describes an algorithm to maximize the smoothed log likelihood. Technical lemmas can be found in Section 4.4.

4.1 DEFINITION OF THE ESTIMATOR

The current status continuous mark model, as described in Section 1.2, is a bivariate censoring model, where the vector (X, Y) has distribution function F_0 . However, instead of observing the vector (X, Y) , the vector $W = (T, Z)$ is observed where T is a censoring variable and $Z = \Delta \cdot Y$ for $\Delta = \mathbb{1}_{\{X \leq T\}}$. Let $F_{0,X}$ be the marginal distribution of X and $\partial_2 F_0(t, z) = \frac{\partial}{\partial z} F_0(t, z)$, then W has density

$$\begin{aligned} h_{f_0}(t, z) &= \mathbb{1}_{\{z > 0\}} g(t) \partial_2 F_0(t, z) + \mathbb{1}_{\{z = 0\}} g(t) (1 - F_{0,X}(t)) \\ &=: \mathbb{1}_{\{z > 0\}} h_1(t, z) + \mathbb{1}_{\{z = 0\}} h_0(t), \end{aligned} \quad (4.1)$$

with respect to the measure λ defined in (1.9).

Let $\mathbb{H}_n$ be the empirical distribution function of $W_1, \dots, W_n$. HUDGENS *et al.* (2007) define and characterize the nonparametric maximum likelihood estimator (MLE), which maximizes

$$\begin{aligned} \ell(f) &= \int \log h_f(t, z) d\mathbb{H}_n(t, z) \\ &= \int \log \left\{ \mathbb{1}_{\{z > 0\}} \partial_2 F(t, z) + \mathbb{1}_{\{z = 0\}} (1 - F_X(t)) \right\} d\mathbb{H}_n(t, z) \end{aligned} \quad (4.2)$$

over the class of distribution functions with density f with respect to $\lambda_1 \times$ counting measure on the observed marks. However, MAATHUIS AND WELLNER (2008) prove that the MLE is inconsistent. The heart of the consistency difficulties with the MLE seems to reside in the first part of the log likelihood

$$\int \mathbb{1}_{\{z > 0\}} \log \partial_2 F(t, z) d\mathbb{H}_n(t, z),$$

where we really have to deal with a density type expression in z (i.e. $\partial_2 F(t, z) = \int_0^t f(s, z) ds$) instead of a bivariate distribution function.

We now define an alternative likelihood-based method, the method of maximum smoothed log likelihood introduced in Section 1.1, where the resulting estimator for F_0 will turn out to be consistent. Let $\hat{H}_n$ be a smoothed version of the empirical distribution function $\mathbb{H}_n$, then the smoothed log likelihood ℓ^S is defined by replacing $\mathbb{H}_n$ in (4.2) by its smoothed version $\hat{H}_n$, i.e.

$$\ell^S(f) = \int \log \left\{ \mathbb{1}_{\{z>0\}} \partial_2 F(t, z) + \mathbb{1}_{\{z=0\}} (1 - F_X(t)) \right\} d\hat{H}_n(t, z).$$

It is very interesting to see that the factorizing property of the MLE, by which the part involving g drops out, also holds (to a certain extent) in the present case: we do not have to maximize over the unknown g , because it does not play a role in the maximization problem (although it is implicitly contained in the estimator $\hat{H}_n$; we do not have to estimate g separately). The maximum smoothed likelihood estimator (MSLE) $\hat{f}_n^{MS}$ for the density f_0 is then defined as

$$\hat{f}_n^{MS} = \operatorname{argmax}_{f \in \mathcal{F}} \ell^S(f), \quad (4.3)$$

where $\mathcal{F}$ is the class of all distribution functions with density f with respect to Lebesgue measure on $[0, \infty)^2$. The MSLE for the bivariate distribution function F_0 we define as

$$\hat{F}_n^{MS}(t, z) = \int_0^t \int_0^z \hat{f}_n^{MS}(u, v) dv du,$$

and the estimators for $F_{0,X}$ and $\partial_2 F_0$ are defined similarly,

$$\begin{aligned} \hat{F}_{n,X}^{MS}(t) &= \int_0^t \int_0^\infty \hat{f}_n^{MS}(u, v) dv du, \\ \partial_2 \hat{F}_n^{MS}(t, z) &= \frac{\partial}{\partial z} \hat{F}_n^{MS}(t, z) = \int_0^t \hat{f}_n^{MS}(u, z) du. \end{aligned}$$

As mentioned in Section 1.1, the MSLE $\hat{f}_n^{MS}$ can also be seen as a Kullback-Leibler projection, minimizing

$$\begin{aligned} \mathcal{K}(\hat{h}_n, h_f) &= \int \hat{h}_n(t, z) \log \frac{\hat{h}_n(t, z)}{h_f(t, z)} d\lambda(t, z) \\ &= \int \log \hat{h}_n(t, z) d\hat{H}_n(t, z) - \int \log h_f(t, z) d\hat{H}_n(t, z), \end{aligned}$$

over densities $f \in \mathcal{F}$ and g . The first term on the right-hand side does not depend on f and the second term on the right-hand side equals $-\ell^S(f)$, hence

$$\hat{f}_n^{MS} = \operatorname{argmax}_{f \in \mathcal{F}} \ell^S(f) = \operatorname{argmin}_{f \in \mathcal{F}} \mathcal{K}(\hat{h}_n, h_{f_0}).$$

In this chapter, we take a histogram-type estimator for the density $\hat{h}_n$ of $\hat{H}_n$, resulting in a piecewise linear estimator $\hat{H}_n$. There are other possibilities as well to choose $\hat{H}_n$, for example via kernel smoothing. For the ease of computations and proving consistency, we choose a histogram-type estimator, and assume without loss of generality that all observations are in $[0, 1]^2$. To define our estimator $\hat{h}_n$ we take two integers k_n and l_n (the number of bins) and define $A_i = A_{n,i} = ((i-1)/k_n, i/k_n] = (a_{i-1}, a_i]$ and $B_j = B_{n,j} = ((j-1)/l_n, j/l_n] = (b_{j-1}, b_j]$ for $i = 1, \dots, k_n$, $j = 1, \dots, l_n$. Then slightly abusing notation, the estimator $\hat{h}_n$ is defined and denoted by

$$\begin{aligned} \hat{h}_n(t, 0) &= \hat{h}_i = k_n \mathbb{H}_n(A_i \times \{0\}), \text{ if } t \in A_i \\ \hat{h}_n(t, z) &= \hat{h}_{i,j} = k_n l_n \mathbb{H}_n(A_i \times B_j), \text{ if } (t, z) \in A_i \times B_j. \end{aligned}$$

Throughout this chapter, the estimators and results are illustrated with simulated data. In these simulations, the true density f_0 is equal to

$$f_0(x, y) = x + y, \text{ for } (x, y) \in [0, 1]^2.$$

The censoring times have the following density

$$g(t) = 2t, \text{ for } t \in [0, 1].$$

An illustration of the estimator $\hat{h}_n$ is given in Figure 4.1 for a simulated sample of size $n = 5000$ where we took $k_n = l_n = 5$.

We consider the estimator $\hat{f}_n^{MS}$ that is obtained by maximizing $\ell^S(f)$ over the class $\mathcal{F}_n$ of piecewise constant densities with cells $A_i \times B_j$. For the resulting histogram-type estimator $\hat{f}_n^{MS}$ we can prove that it is well defined and unique if all cells contain at least one observation. This holds with high probability under certain conditions on the observation density h_{f_0} and the total number $k_n \cdot l_n$ of cells $A_i \times B_j$.

Before proving the existence and uniqueness of $\hat{f}_n^{MS}$, stated in Theorem 4.2 below, we introduce some notation to relate the class of densities we consider to appropriate subsets of Euclidean space.

$$\begin{aligned} \mathcal{B}_n &= \{f \in \mathbb{R}^{k_n \times l_n} : 0 \leq f_{i,j} \leq k_n l_n \forall i, j\} \\ \alpha_i(f) &= (k_n l_n)^{-1} \sum_{l=i}^{k_n} \sum_{j=1}^{l_n} f_{l,j}, \quad \beta_{i,j}(f) = k_n^{-1} \sum_{l=1}^i f_{l,j}, \quad f \in \mathcal{B}_n \end{aligned} \tag{4.4}$$

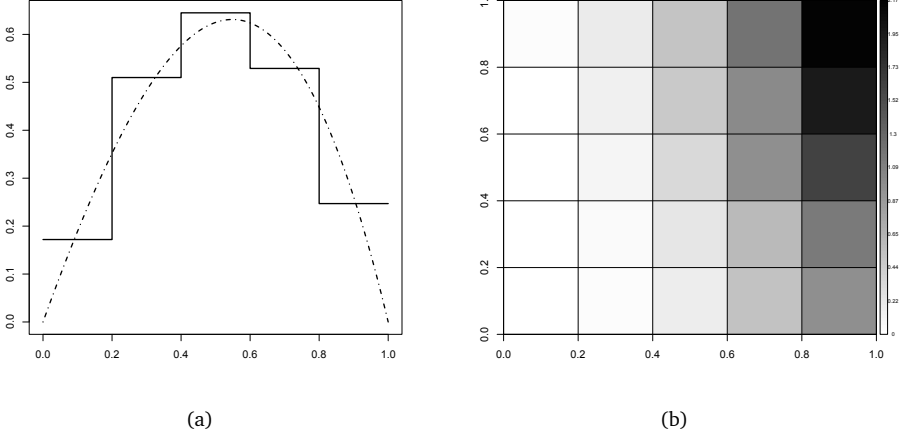


FIGURE 4.1: The estimators $\hat{h}_n(t, 0)$ (a) and $\hat{h}_n(t, z)$ for $z > 0$ (b) for a simulated sample of size $n = 5000$ with $k_n = l_n = 5$. The dash-dotted line in (a) is the true density $h_{f_0}(t, 0)$ for this example.

with $\alpha_{k_n+1}(f) = 0$ and $\beta_{0,j}(f) = 0$ for all j .

LEMMA 4.1. Define

$$\begin{aligned}
 \psi_{-1}(f) &= k_n^{-1} \sum_{i=1}^{k_n} \hat{h}_i \varphi(\alpha_{i+1}(f), \alpha_i(f)) + (k_n l_n)^{-1} \sum_{i=1}^{k_n} \sum_{j=1}^{l_n} \hat{h}_{i,j} \varphi(\beta_{i,j}(f), \beta_{i-1,j}(f)) \\
 &\quad - (k_n l_n)^{-1} \sum_{i=1}^{k_n} \sum_{j=1}^{l_n} f_{i,j} + 1, \\
 \varphi(x, y) &= \begin{cases} (x \log x - y \log y)/(x - y), & x, y \in [0, 1], x \neq y \\ 1 + \log x, & x, y \in [0, 1], x = y, \end{cases} \quad (4.5)
 \end{aligned}$$

with $0 \log 0 := 0$, then

$$\operatorname{argmax}_{f \in \mathcal{F}_n} \ell^S(f) = \operatorname{argmax}_{f \in \mathcal{B}_n} \psi_{-1}(f).$$

PROOF. Let $f \in \mathcal{F}_n$, then for $t \in A_i = (a_{i-1}, a_i]$, $z \in B_j = (b_{j-1}, b_j]$ we can write

$$1 - F_X(t) = \int_{u=t}^{M_1} \int_{z=0}^{M_2} f(u, z) dz du = (a_i - t) \sum_{j=1}^{l_n} l_n^{-1} f_{i,j} + \sum_{l=i+1}^{k_n} \sum_{j=1}^{l_n} (k_n l_n)^{-1} f_{l,j}$$

$$\partial_2 F(t, z) = \int_{u=0}^t f(u, z) du = \sum_{l=1}^{i-1} k_n^{-1} f_{l,j} + (t - a_{i-1}) f_{i,j},$$

so that

$$\int_{A_i} \log(1 - F_X(t)) dt = \int_{A_i} \log \left\{ \sum_{j=1}^{l_n} \sum_{l=i+1}^{k_n} (k_n l_n)^{-1} f_{l,j} + a_i \sum_{j=1}^{l_n} l_n^{-1} f_{i,j} - t \sum_{j=1}^{l_n} l_n^{-1} f_{i,j} \right\},$$

$$\int_{A_i} \int_{B_j} \log \partial_2 F(t, z) dz dt = l_n^{-1} \int_{A_i} \log \left\{ \sum_{l=1}^{i-1} k_n^{-1} f_{l,j} - a_{i-1} f_{i,j} + f_{i,j} t \right\} dt.$$

Since we have for $0 \leq a < b < \infty$, $\sigma \neq 0$ and $\tau \geq -\sigma a$

$$\int_a^b \log(\tau + \sigma t) dt = \frac{1}{\sigma} [u \log u - u]_{u=\tau+\sigma a}^{\tau+\sigma b}$$

$$= \frac{1}{\sigma} (\tau + \sigma b) \log(\tau + \sigma b) - \frac{1}{\sigma} (\tau + \sigma a) \log(\tau + \sigma a) - (b - a),$$

we get

$$\int_{A_i} \log \left\{ \sum_{j=1}^{l_n} \sum_{l=i+1}^{k_n} (k_n l_n)^{-1} f_{l,j} + a_i \sum_{j=1}^{l_n} l_n^{-1} f_{i,j} - t \sum_{j=1}^{l_n} l_n^{-1} f_{i,j} \right\} dt$$

$$= -\frac{1}{l_n^{-1} \sum_{j=1}^{l_n} f_{i,j}} \left\{ \left((k_n l_n)^{-1} \sum_{l=i+1}^{k_n} \sum_{j=1}^{l_n} f_{l,j} \right) \log \left((k_n l_n)^{-1} \sum_{l=i+1}^{k_n} \sum_{j=1}^{l_n} f_{l,j} \right) \right.$$

$$\left. - \left((k_n l_n)^{-1} \sum_{l=i}^{k_n} \sum_{j=1}^{l_n} f_{l,j} \right) \log \left((k_n l_n)^{-1} \sum_{l=i}^{k_n} \sum_{j=1}^{l_n} f_{l,j} \right) \right\} - k_n^{-1}$$

and

$$\begin{aligned} l_n^{-1} \int_{A_i} \log \left\{ \sum_{l=1}^{i-1} k_n^{-1} f_{l,j} - a_{i-1} f_{i,j} + f_{i,j} t \right\} dt \\ = l_n^{-1} \frac{1}{f_{i,j}} \left\{ \left(\sum_{l=1}^i k_n^{-1} f_{l,j} \right) \log \left(\sum_{l=1}^i k_n^{-1} f_{l,j} \right) - \left(\sum_{l=1}^{i-1} k_n^{-1} f_{l,j} \right) \log \left(\sum_{l=1}^{i-1} k_n^{-1} f_{l,j} \right) \right\} \\ - l_n^{-1} k_n^{-1}, \end{aligned}$$

so that

$$\begin{aligned} \ell^S(f) = & - \sum_{i=1}^{k_n} \frac{\hat{h}_i}{l_n^{-1} \sum_{j=1}^{l_n} f_{i,j}} \left\{ \left((k_n l_n)^{-1} \sum_{l=i+1}^{k_n} \sum_{j=1}^{l_n} f_{l,j} \right) \log \left((k_n l_n)^{-1} \sum_{l=i+1}^{k_n} \sum_{j=1}^{l_n} f_{l,j} \right) \right. \\ & \left. - \left((k_n l_n)^{-1} \sum_{l=i}^{k_n} \sum_{j=1}^{l_n} f_{l,j} \right) \log \left((k_n l_n)^{-1} \sum_{l=i}^{k_n} \sum_{j=1}^{l_n} f_{l,j} \right) \right\} \\ & + \sum_{i=1}^{k_n} \sum_{j=1}^{l_n} \frac{l_n^{-1} \hat{h}_{i,j}}{f_{i,j}} \left\{ \left(k_n^{-1} \sum_{l=1}^i f_{l,j} \right) \log \left(k_n^{-1} \sum_{l=1}^i f_{l,j} \right) \right. \\ & \left. - \left(k_n^{-1} \sum_{l=1}^{i-1} f_{l,j} \right) \log \left(k_n^{-1} \sum_{l=1}^{i-1} f_{l,j} \right) \right\} \\ & - \sum_{i=1}^{k_n} k_n^{-1} \hat{h}_i - \sum_{i=1}^{k_n} \sum_{j=1}^{l_n} (k_n l_n)^{-1} \hat{h}_{i,j}. \end{aligned}$$

The last two terms can be left out in the maximization, since they do not depend on f . Now, taking $\alpha_i(f)$ and $\beta_{i,j}(f)$ as in (4.4) we have that

$$l_n^{-1} \sum_{j=1}^{l_n} f_{i,j} = -k_n (\alpha_{i+1}(f) - \alpha_i(f)), \quad f_{i,j} = k_n (\beta_{i,j}(f) - \beta_{i-1,j}(f)),$$

so that $\ell^S(f) = \psi(f)$ for

$$\psi(f) = k_n^{-1} \sum_{i=1}^{k_n} \hat{h}_i \varphi(\alpha_{i+1}(f), \alpha_i(f)) + (k_n l_n)^{-1} \sum_{i=1}^{k_n} \sum_{j=1}^{l_n} \hat{h}_{i,j} \varphi(\beta_{i,j}(f), \beta_{i-1,j}(f))$$

with φ as defined in (4.5), hence the maximizer of $\ell^S(f)$ is the maximizer of $\psi(f)$.

The estimator $\hat{f}_n^{MS}$ has to satisfy the following conditions

$$(S.1) \quad (k_n l_n)^{-1} \sum_{i=1}^{k_n} \sum_{j=1}^{l_n} f_{i,j} = 1,$$

$$(S.2) \quad \forall i, j : f_{i,j} \geq 0.$$

To get condition (S.1) in the objective function $\psi(f)$, we include a Lagrange multiplier $\lambda \in \mathbb{R}$ and maximize

$$\psi_\lambda(f) = \psi(f) + \lambda \left((k_n l_n)^{-1} \sum_{i=1}^{k_n} \sum_{j=1}^{l_n} f_{i,j} - 1 \right)$$

over $\mathcal{P}_n = \{f \in \mathbb{R}^{k_n l_n} : f_{i,j} \geq 0 \ \forall i, j\}$. For any function f satisfying condition (S.1) and each λ , $\psi_\lambda(f)$ equals $\psi(f)$ and for $\hat{f}_\lambda = \operatorname{argmax}_{f \in \mathcal{P}_n} \psi_\lambda(f)$ we have

$$0 = \lim_{\gamma \rightarrow 0} \gamma^{-1} \left(\psi_\lambda((1 + \gamma)\hat{f}_\lambda) - \psi_\lambda(\hat{f}_\lambda) \right) = 1 + \lambda (k_n l_n)^{-1} \sum_{i=1}^{k_n} \sum_{j=1}^{l_n} \hat{f}_{\lambda,i,j}.$$

This means that if we take $\lambda = -1$, the global maximizer of $\psi_\lambda(f)$ over $\mathcal{P}_n$ is contained in $\mathcal{F}_n$, so that

$$\operatorname{argmax}_{f \in \mathcal{P}_n} \psi_{-1}(f) = \operatorname{argmax}_{f \in \mathcal{F}_n} \psi(f).$$

Since we also have that $\mathcal{F}_n \subset \mathcal{B}_n$, it follows that $\operatorname{argmax}_{f \in \mathcal{P}_n} \psi_{-1}(f) \in \mathcal{B}_n$, hence

$$\operatorname{argmax}_{f \in \mathcal{F}_n} \ell^S(f) = \operatorname{argmax}_{f \in \mathcal{B}_n} \psi_{-1}(f).$$

□

The existence and uniqueness of $\hat{f}_n^{MS}$ follows from this lemma.

THEOREM 4.2. *If $\hat{h}_i > 0$ and $\hat{h}_{i,j} > 0$ for all i, j , then the maximizer $\hat{f}_n^{MS}$ of ℓ^S over $\mathcal{F}_n$ exists and is unique.*

REMARK. If $\hat{h}_i = 0$ for some i or $\hat{h}_{i,j} = 0$ for some i and j , we can construct examples where $\hat{f}_n^{MS}$ is not unique. However, we can prove under various conditions on h_{f_0} , k_n and l_n that

$$P(\exists i : \hat{h}_i = 0) \longrightarrow 0, \quad P(\exists i, j : \hat{h}_{i,j} = 0) \longrightarrow 0, \quad (4.6)$$

so that with probability converging to one $\hat{f}_n^{MS}$ is well defined for n sufficiently large. For example, if $k_n = l_n$ and for some $c > 0$

$$\int_{A_i} h_0(t) dt \geq \frac{c}{k_n^2} \text{ and } \int_{A_i \times B_j} h_1(t, z) dt dz \geq \frac{c}{k_n^3}, \quad (4.7)$$

for all i and j , then

$$\begin{aligned} P(\exists i, j : \hat{h}_{i,j} = 0) &\leq \sum_{i,j} P(\hat{h}_{i,j} = 0) \leq k_n^2 \left(1 - \frac{c}{k_n^3}\right)^n \\ &= k_n^2 \left(\left(1 - \frac{c}{k_n^3}\right)^{k_n^3} \right)^{n/k_n^3} \approx k_n^2 e^{-cn/k_n^3} \longrightarrow 0, \end{aligned}$$

if $n^{-1}k_n^3 \log k_n \rightarrow 0$.

A similar argument shows that $P(\exists i : \hat{h}_i = 0) \rightarrow 0$.

Assume for example that f_0 has compact support $\mathcal{W}_M = [0, M_1] \times [0, M_2]$ for some constants $0 < M_1, M_2 < \infty$. If $g \geq \kappa_1 > 0$ on $[0, M_1]$ and $f_0 \geq \kappa_2 > 0$ on $[0, c_1] \times [0, M_2]$ and on $[c_2, M_1] \times [\eta_1, \eta_2]$ for some constants $c_1, c_2, \eta_1, \eta_2 > 0$, then $h_{f_0} \geq \tilde{\kappa} > 0$ on $\mathcal{W}_M$ and

$$\int_{A_i} h_0(t) dt \geq \frac{\tilde{\kappa}}{k_n^2}, \quad \int_{A_i \times B_j} h_1(t, z) dt dz \geq \frac{\tilde{\kappa}}{k_n^3}.$$

Hence, this is a sufficient condition for (4.7).

PROOF OF THEOREM 4.2. The function φ is continuous on $(0, 1]^2$ so ψ_{-1} is continuous on the compact set $\mathcal{B}_n$. Hence ψ_{-1} attains its maximum over $\mathcal{B}_n$ and $\hat{f}_n^{MS}$ exists.

Recall the definitions of $\alpha_i(f)$ and $\beta_{i,j}(f)$

$$\alpha_i(f) = (k_n l_n)^{-1} \sum_{l=i}^{k_n} \sum_{j=1}^{l_n} f_{l,j}, \quad \beta_{i,j}(f) = k_n^{-1} \sum_{l=1}^i f_{l,j},$$

to note that for every $f, g \in \mathcal{B}_n$ and $\gamma \in [0, 1]$

$$\begin{aligned} \alpha_i(\gamma f + (1 - \gamma)g) &= \gamma \alpha_i(f) + (1 - \gamma) \alpha_i(g), \\ \beta_{i,j}(\gamma f + (1 - \gamma)g) &= \gamma \beta_{i,j}(f) + (1 - \gamma) \beta_{i,j}(g). \end{aligned}$$

If $f \neq g$, i.e. $f_{i,j} \neq g_{i,j}$ for some i and j , then this implies that $\alpha_m(f) \neq \alpha_m(g)$ for all $m \leq i$ and $\beta_{m,j}(f) \neq \beta_{m,j}(g)$ for all $m \geq i$. By Lemma 4.5 we then have for each i and

j

$$\begin{aligned} \varphi(\alpha_{i+1}(\gamma f + (1-\gamma)g), \alpha_i(\gamma f + (1-\gamma)g)) &> \gamma \varphi(\alpha_{i+1}(f), \alpha_i(f)) \\ &\quad + (1-\gamma) \varphi(\alpha_{i+1}(g), \alpha_i(g)), \\ \varphi(\beta_{i,j}(\gamma f + (1-\gamma)g), \beta_{i-1,j}(\gamma f + (1-\gamma)g)) &> \gamma \varphi(\beta_{i,j}(f), \beta_{i-1,j}(f)) \\ &\quad + (1-\gamma) \varphi(\beta_{i,j}(g), \beta_{i-1,j}(g)). \end{aligned}$$

This implies that

$$\begin{aligned} \psi_{-1}(\gamma f + (1-\gamma)g) &= k_n^{-1} \sum_{i=1}^{k_n} \hat{h}_i \varphi(\alpha_{i+1}(\gamma f + (1-\gamma)g), \alpha_i(\gamma f + (1-\gamma)g)) \\ &\quad + (k_n l_n)^{-1} \sum_{i=1}^{k_n} \sum_{j=1}^{l_n} \hat{h}_{i,j} \varphi(\beta_{i,j}(\gamma f + (1-\gamma)g), \\ &\quad \quad \quad \beta_{i-1,j}(\gamma f + (1-\gamma)g)) \\ &\quad - (k_n l_n)^{-1} \sum_{i=1}^{k_n} \sum_{j=1}^{l_n} \{\gamma f_{i,j} + (1-\gamma)g_{i,j}\} + 1 \\ &> k_n^{-1} \sum_{i=1}^{k_n} \hat{h}_i \{\gamma \varphi(\alpha_{i+1}(f), \alpha_i(f)) \\ &\quad \quad \quad + (1-\gamma) \varphi(\alpha_{i+1}(g), \alpha_i(g))\} \\ &\quad + (k_n l_n)^{-1} \sum_{i=1}^{k_n} \sum_{j=1}^{l_n} \hat{h}_{i,j} \{\gamma \varphi(\beta_{i,j}(f), \beta_{i-1,j}(f)) \\ &\quad \quad \quad + (1-\gamma) \varphi(\beta_{i,j}(g), \beta_{i-1,j}(g))\} \\ &\quad - \gamma (k_n l_n)^{-1} \sum_{i=1}^{k_n} \sum_{j=1}^{l_n} f_{i,j} - (1-\gamma) (k_n l_n)^{-1} \sum_{i=1}^{k_n} \sum_{j=1}^{l_n} g_{i,j} \\ &\quad + \gamma + (1-\gamma) \\ &= \gamma \psi_{-1}(f) + (1-\gamma) \psi_{-1}(g), \end{aligned}$$

hence $\psi_{-1}(f)$ is strictly concave and its maximizer $\hat{f}_n^{MS}$ is unique. $\square$

4.2 COMPUTING THE ESTIMATOR

By Theorem 4.2 we know that the estimator $\hat{f}_n^{MS}$ defined in (4.3) exists and is unique. In this section we describe an algorithm to compute

$$\hat{f}_n^{MS} = \operatorname{argmax}_{f \in \mathcal{B}_n} \psi_{-1}(f) = \operatorname{argmin}_{f \in \mathcal{B}_n} \tilde{\psi}(f),$$

where $\tilde{\psi} = -\psi_{-1}$.

Write $f \in \mathcal{F}_n$ as a vector, i.e. $f = (f_1 \ f_2 \ \dots \ f_{k_n l_n})^T$. We use Newton's method [see e.g., BAZARAA *et al.* (2006, Section 8.6)] to minimize $\tilde{\psi}$ without the constraint $f \in \mathcal{B}_n$. This method iteratively minimizes the quadratic approximation

$$Q(f; g) = \tilde{\psi}(g) + (f - g)^T \nabla \tilde{\psi}(g) + \frac{1}{2} (f - g)^T H_g (f - g)$$

of $\psi_{-1}(f)$ for fixed $g \in \mathcal{F}_n$, where H_g is the Hessian matrix of $\tilde{\psi}$ evaluated at g . Minimizing $Q(f; g)$ over f is equivalent with minimizing

$$Q^*(f; g) = \frac{1}{2} (f - g + H_g^{-1} \nabla \tilde{\psi}(g))^T H_g (f - g + H_g^{-1} \nabla \tilde{\psi}(g)),$$

and its minimizer is given by

$$\hat{f} = g - H_g^{-1} \nabla \tilde{\psi}(g).$$

If $\hat{f}_i < 0$, we set $\hat{f}_i$ equal to 0 to ensure that each iteration does satisfy the constraint $f \in \mathcal{B}_n$. In our simulations, this modification of Newton's algorithm returned a solution in the interior of $\mathcal{B}_n$. Hence, this algorithm found the unique minimizer of $\tilde{\psi}(f)$, i.e. the unique MSLE $\hat{f}_n^{MS}$.

The dimension of the Hessian H_g is $k_n l_n \times k_n l_n$, so when k_n and l_n are large, inverting H_g can cause instability of the algorithm. A solution would be to replace H_g with a diagonal approximation D_g , where

$$(D_g)_{i,j} = \begin{cases} \frac{\partial^2}{\partial f_i \partial f_j} \psi_{-1}(f) \Big|_{f=g} & \text{if } i = j \\ 0 & \text{if } i \neq j. \end{cases}$$

The minimizer of $Q(f; g)$ is then given by $\hat{f} = g - D_g^{-1} \nabla \tilde{\psi}(g)$. The resulting algorithm can be seen as a Sequential approximating Quadratic Minimization Algorithm. One could end this algorithm with one or more steps of Newton's method.

This function $\hat{f}$ only gives the direction to find the next iterate, it could be that $\tilde{\psi}(\hat{f}) > \tilde{\psi}(g)$. The next iterate is therefore given by $g + \hat{\lambda}(\hat{f} - g)$ where $\hat{\lambda}$ approximately minimizes $\tau(\lambda) = \tilde{\psi}(g + \lambda(\hat{f} - g))$ over $\lambda \in [0, 1]$. To find the approximate minimizer $\hat{\lambda}$, we use a bisection line search method.

Figure 4.2 shows the values of $\tilde{\psi}$ per iteration for the same simulation as in Figure 4.1 and for three different starting points of the algorithm. For any starting point, the algorithm converged to the same minimizing value. The resulting estimators $\hat{f}_n^{MS}$ and $\hat{F}_n^{MS}$ are given in Figure 4.3 for one of the starting points.

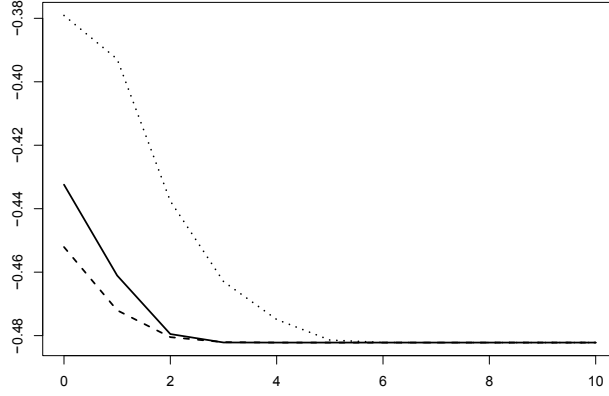


FIGURE 4.2: The values of ψ_{-1} per iteration for three different starting points for a simulated sample of size $n = 5000$ with $k_n = l_n = 5$.

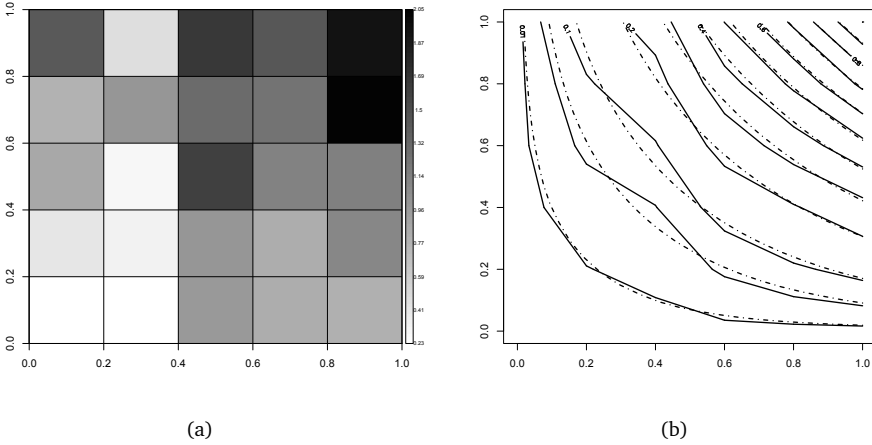


FIGURE 4.3: The estimator $\hat{f}_n^{MS}$ (a) and contour plot of $\hat{F}_n^{MS}$ (b) for a simulated sample of size $n = 5000$ with $k_n = l_n = 5$. The dash-dotted lines is the contour plot of F_0 . The levels of the contour plot are 0.01, 0.05, 0.1, 0.2, $\dots$, 1.0.

4.3 CONSISTENCY OF THE ESTIMATOR

In this section we prove that $\hat{h}_n^{MS} = h_{\hat{f}_n^{MS}}$ and $\hat{F}_n^{MS}$ are consistent estimators for the density h_{f_0} of the observable vector W and the bivariate distribution function of interest F_0 , respectively. To prove the consistency of $\hat{h}_n^{MS}$ and $\hat{F}_n^{MS}$ we assume the densities f_0 and g to satisfy conditions (E.1) and (G.1) below. We also assume f_0 and g are such that h_{f_0} satisfies the conditions needed for (4.6) so that existence and uniqueness of $\hat{f}_n^{MS}$ are guaranteed with probability converging to one. Furthermore, we assume the number of bins k_n and l_n to satisfy condition (B.1) below.

- (E.1) The density f_0 has compact support $\mathcal{W}_M = [0, M_1] \times [0, M_2]$ and is continuous on $\mathcal{W}_M$.
- (G.1) The censoring density g is uniformly continuous and bounded away from zero and infinity on $(0, M_1)$, i.e. $0 < c_g \leq g(t) \leq C_g < \infty$ for all $t \in (0, M_1)$.
- (B.1) The number of bins k_n and l_n converge to ∞ such that $n^{-1}k_n l_n \rightarrow 0$ as $n \rightarrow \infty$.

Note that if $k_n \sim n^{1/4}$ and $l_n \sim n^{1/4}$, condition (B.1) is satisfied. This choice is the L_2 -optimal order for estimating the bivariate histogram estimator $\hat{h}_n$ [see e.g., LECOUTRE (1985)]. Furthermore, note that for this choice of k_n and l_n condition (4.7) holds if F_0 satisfies condition (E.2) below.

- (E.2) $\partial_2 F_0(t, z)$ is either linear in t or in z (but not in both arguments) and $F_{0,X}(t)$ is linear in t .

LEMMA 4.3. *Let f_0 and g satisfy conditions (E.1) and (G.1) and k_n and l_n condition (B.1). Furthermore, assume that $\int_0^{M_1} \log h_0(t) dt < \infty$, $\int_0^{M_1} \int_0^{M_1} \log h_1(t, z) dz dt < \infty$. Then $\hat{h}_n^{MS}$ is $(\mathcal{H})$ -consistent for h_{f_0} , i.e.*

$$\mathcal{H}(\hat{h}_n^{MS}, h_{f_0}) \xrightarrow{\mathcal{P}} 0. \quad (4.8)$$

PROOF. We establish (4.8) using relation (B.3) and the property that $\hat{f}_n^{MS}$ minimizes $\mathcal{K}(\hat{h}_n, h_{\hat{f}_n})$. Since $f_0 \notin \mathcal{F}_n$ in general, the inequality

$$\mathcal{K}(\hat{h}_n, h_{\hat{f}_n^{MS}}) \leq \mathcal{K}(\hat{h}_n, h_{f_0})$$

need not hold. In order to exploit the minimizing property of $\hat{f}_n^{MS}$, we define a piecewise constant representative $\tilde{f}_{n,0}$ of f_0 which belongs to $\mathcal{F}_n$ and approximates f_0

$$\tilde{f}_{n,0}(t, z) = k_n l_n \int_{A_i} \int_{B_j} f_0(u, v) dv du \text{ if } (t, z) \in A_i \times B_j.$$

For this representative it holds that

$$0 \leq 2\mathcal{H}(\hat{h}_n, \hat{h}_n^{MS})^2 \leq \mathcal{H}(\hat{h}_n, \hat{h}_n^{MS}) \leq \mathcal{H}(\hat{h}_n, h_{\hat{f}_{n,0}}), \quad (4.9)$$

also using relation (B.3). The Hellinger distance is a metric, hence applying the triangle inequality twice gives

$$0 \leq \mathcal{H}(\hat{h}_n^{MS}, h_{f_0}) \leq \mathcal{H}(\hat{h}_n^{MS}, \hat{h}_n) + \mathcal{H}(\hat{h}_n, h_{\hat{f}_{n,0}}) + \mathcal{H}(h_{\hat{f}_{n,0}}, h_{f_0}). \quad (4.10)$$

The first term on the right hand side of (4.10) converges in probability to zero by combining (4.9) and Lemma 4.8. The second term converges in probability to zero by combining Lemma 4.8 and relation (B.3). The third term converges to zero by combining relation (B.5) and the second result in Lemma 4.6, hence (4.8) follows. $\square$

From Lemma 4.3, it follows that $\hat{F}_n^{MS}$ converges pointwise and in L_1 -norm to F_0 .

THEOREM 4.4. *Under the conditions of Lemma 4.3,*

$$\|\hat{F}_n^{MS} - F_0\|_1 \xrightarrow{\mathcal{P}} 0. \quad (4.11)$$

Consequently, for all $(t, z) \in \mathcal{W}_M$

$$\hat{F}_n^{MS}(t, z) \xrightarrow{\mathcal{P}} F_0(t, z). \quad (4.12)$$

REMARK. Because $\hat{F}_{n,X}^{MS}(t) = \hat{F}_n^{MS}(t, M_2)$, this lemma implies that for all $t \in [0, M_1]$

$$\hat{F}_{n,X}^{MS}(t) \xrightarrow{\mathcal{P}} F_{0,X}(t).$$

PROOF OF THEOREM 4.4. By combining Lemma 4.3 and relation (B.4), we have that

$$\begin{aligned} \|\hat{h}_n^{MS} - h_{f_0}\|_1 &= \int |\hat{h}_n^{MS}(t, z) - h_{f_0}(t, z)| \, d\lambda(t, z) \\ &= \int_0^{M_1} \int_0^{M_2} g(t) |\partial_2 \hat{F}_n^{MS}(t, z) - \partial_2 F_0(t, z)| \, dz \, dt \\ &\quad + \int_0^{M_1} g(t) |(1 - \hat{F}_{n,X}^{MS}(t)) - (1 - F_{0,X}(t))| \, dt \\ &\leq \sqrt{2} \mathcal{H}(\hat{h}_n^{MS}, h_{f_0}) \xrightarrow{\mathcal{P}} 0. \end{aligned}$$

This implies that

$$\|\partial_2 \hat{F}_n^{MS} - \partial_2 F_0\|_1 = \int |\partial_2 \hat{F}_n^{MS}(t, z) - \partial_2 F_0(t, z)| d\lambda(t, z) \xrightarrow{\mathcal{P}} 0,$$

since $g > 0$ on $(0, M_1)$. To prove (4.11) note that the L_1 -distance between $\hat{F}_n^{MS}$ and F_0 can be bounded by

$$\begin{aligned} \|\hat{F}_n^{MS} - F_0\|_1 &= \int |\hat{F}_n^{MS}(t, z) - F_0(t, z)| d\lambda(t, z) \\ &= \int \left| \int_0^z (\partial_2 \hat{F}_n^{MS}(t, v) - \partial_2 F_0(t, v)) dv \right| d\lambda(t, z) \\ &\leq \int_{t=0}^{M_1} \int_{v=0}^{M_2} \int_{z=v}^{M_2} |\partial_2 \hat{F}_n^{MS}(t, v) - \partial_2 F_0(t, v)| dz dv dt \\ &\leq M_2 \|\partial_2 \hat{F}_n^{MS} - \partial_2 F_0\|_1 \xrightarrow{\mathcal{P}} 0. \end{aligned}$$

To prove (4.12), assume it does not hold for a certain $(t, z) \in \mathcal{W}_M$, i.e. there exists $\varepsilon > 0$, $\delta > 0$ and a subsequence n_k of n such that for all $k \in \mathbb{N}$

$$\mathbb{P} \left(|\hat{F}_{n_k}^{MS}(t, z) - F_0(t, z)| \geq \delta \right) \geq \varepsilon.$$

Assume $\hat{F}_{n_k}^{MS}(t, z) \leq F_0(t, z) - \delta$, then there exists a small $c > 0$ such that

$$\forall (u, v) \in [t - c\delta, t] \times [z - c\delta, z] =: \mathcal{A}_\delta : \hat{F}_{n_k}^{MS}(u, v) \leq F_0(u, v) - \frac{1}{2}\delta,$$

by continuity of F_0 and monotonicity of $\hat{F}_n^{MS}$. This implies that for all k

$$\begin{aligned} \mathbb{P} \left(\int |\hat{F}_{n_k}^{MS}(t, z) - F_0(t, z)| dz dt \geq \frac{c^2}{2} \delta \right) \\ \geq \mathbb{P} \left(\int_{\mathcal{A}_\delta} |\hat{F}_{n_k}^{MS}(u, z) - F_0(u, z)| dz du \geq \frac{c^2}{2} \delta \right) \\ \geq \mathbb{P} \left(\forall (u, v) \in \mathcal{A}_\delta : \hat{F}_{n_k}^{MS}(u, v) \leq F_0(u, v) - \frac{1}{2}\delta \right) \geq \varepsilon. \end{aligned}$$

If $\hat{F}_{n_k}^{MS}(t, z) \geq F_0(t, z) + \delta$, we have by a similar argument that

$$\forall (u, v) \in [t, t + c\delta] \times [z, z + c\delta] =: \mathcal{A}_\delta : \hat{F}_{n_k}^{MS}(u, v) \geq F_0(u, v) + \frac{1}{2}\delta,$$

giving that

$$\begin{aligned} P \left(\int |\hat{F}_{n_k}^{MS}(t, z) - F_0(t, z)| dz dt \geq \frac{c^2}{2} \delta \right) \\ \geq P \left(\forall (u, v) \in \mathcal{A}_\delta : \hat{F}_{n_k}^{MS}(u, v) \geq F_0(u, v) + \frac{1}{2} \delta \right) \geq \varepsilon. \end{aligned}$$

This contradicts (4.11). $\square$

4.4 TECHNICAL LEMMAS

LEMMA 4.5. *The function $\varphi(x, y)$ defined in (4.5) is strictly concave.*

PROOF. Without loss of generality $0 < x < y$. Because φ is twice continuously differentiable, proving that φ is strictly concave is equivalent with proving that $-H$ is positive definite, where

$$H(x, y) = \begin{bmatrix} \frac{\partial^2}{\partial x^2} \varphi(x, y) & \frac{\partial^2}{\partial x \partial y} \varphi(x, y) \\ \frac{\partial^2}{\partial y \partial x} \varphi(x, y) & \frac{\partial^2}{\partial y^2} \varphi(x, y) \end{bmatrix}.$$

To prove $-H$ is positive definite, we have to prove that both eigenvalues λ_1, λ_2 of $-H$ are strictly positive. Since it holds that

$$\det(-H) = \lambda_1 \lambda_2, \quad \text{trace}(-H) = \lambda_1 + \lambda_2,$$

both λ_1 and λ_2 are strictly positive if and only if $\det(-H) > 0$ and $\text{trace}(-H) > 0$.

Write $\varphi(x, y) = (\tau(x) - \tau(y))/(x - y)$ for $\tau(x) = x \log x$, then

$$\begin{aligned} \frac{\partial^2}{\partial x^2} \varphi(x, y) &= 2 \frac{\tau(x) - \tau(y)}{(x - y)^3} - 2 \frac{\tau'(x)}{(x - y)^2} + \frac{\tau''(x)}{x - y} \\ \frac{\partial^2}{\partial x \partial y} \varphi(x, y) &= -2 \frac{\tau(x) - \tau(y)}{(x - y)^3} + \frac{\tau'(x) + \tau'(y)}{(x - y)^2} \\ \frac{\partial^2}{\partial y^2} \varphi(x, y) &= 2 \frac{\tau(x) - \tau(y)}{(x - y)^3} - 2 \frac{\tau'(y)}{(x - y)^2} - \frac{\tau''(y)}{x - y}. \end{aligned}$$

We now prove that $\det(H) = \det(-H) > 0$. We have

$$\begin{aligned}
 (x-y)^4 \det(H(x, y)) &= 4\tau'(x)\tau'(y) + 2(\tau(x) - \tau(y))(\tau''(x) - \tau''(y)) \\
 &\quad - (\tau'(x) + \tau'(y))^2 + 2(x-y)(\tau'(x)\tau''(y) - \tau''(x)\tau'(y)) \\
 &\quad - (x-y)^2\tau''(x)\tau''(y) \\
 &= -(1 + \log x)^2 - (1 + \log y)^2 + 2(1 + \log x)(1 + \log y) \\
 &\quad + \frac{x^2 + y^2 - 2xy}{xy},
 \end{aligned}$$

which gives that

$$\det(H(x, y)) = \frac{(x-y)^2 - xy(\log x - \log y)^2}{xy(x-y)^4}.$$

So to prove that $\det(H) > 0$ it suffices to prove that the denominator is strictly positive, or that

$$(\log x - \log y)^2 < \frac{(x-y)^2}{xy},$$

or that (by putting $u = \sqrt{x}$ and $v = \sqrt{y}$)

$$|\log u - \log v| < \frac{1}{2} \frac{(u+v)|u-v|}{uv}.$$

The function $f(z) = 1/z$ is strictly convex on $(0, \infty)$, so for $z \in (u, v)$, the graph of f lies strictly below the line segment connecting the points $(u, 1/u)$ and $(v, 1/v)$ on this graph. Hence

$$|\log u - \log v| = \left| \int_u^v \frac{1}{z} dz \right| < \frac{1}{2} \left(\frac{1}{u} + \frac{1}{v} \right) |u-v| = \frac{1}{2} \frac{(u+v)|u-v|}{uv}.$$

Proving that $\text{trace}(-H) > 0$ is equivalent with proving that $\text{trace}(H) < 0$. Recall that

$$\begin{aligned}
 \frac{\partial^2}{\partial x^2} \varphi(x, y) &= 2 \frac{\tau(x) - \tau(y)}{(x-y)^3} - 2 \frac{\tau'(x)}{(x-y)^2} + \frac{\tau''(x)}{x-y} \\
 &= \frac{2}{(x-y)^3} (\tau(x) - \tau(y) + (y-x)\tau'(x) + \frac{1}{2}(y-x)^2\tau''(x)).
 \end{aligned}$$

Using a Taylor expansion, we have

$$\tau(y) = \tau(x) + (y-x)\tau'(x) + \frac{1}{2}(y-x)^2\tau''(\zeta_1),$$

for some $\zeta_1 \in (x, y)$. Hence we have

$$\frac{\partial^2}{\partial x^2} \varphi(x, y) = \frac{\tau''(x) - \tau''(\zeta_1)}{x - y}.$$

In a similar way we get

$$\frac{\partial^2}{\partial y^2} \varphi(x, y) = \frac{\tau''(\zeta_2) - \tau''(y)}{x - y},$$

for some $\zeta_2 \in (x, y)$, so that

$$\text{trace}(H) = \frac{\partial^2}{\partial x^2} \varphi(x, y) + \frac{\partial^2}{\partial y^2} \varphi(x, y) = \frac{\tau''(x) - \tau''(\zeta_1) + \tau''(\zeta_2) - \tau''(y)}{x - y}.$$

We assumed $x < y$, so this is strictly negative if $\tau''(x) - \tau''(\zeta_1) + \tau''(\zeta_2) - \tau''(y) > 0$, which is true since $\tau''(z) = \frac{1}{z}$ and $x < \zeta_1, \zeta_2 < y$. $\square$

The piecewise constant representative $\bar{f}_{n,0}$ and the corresponding $h_{\bar{f}_{n,0}}$ converge to the true f_0 and h_{f_0} under condition (B.1). This is stated in Lemma 4.6 below.

LEMMA 4.6. *Under the conditions of Lemma 4.3,*

$$\|\bar{f}_{n,0} - f_0\|_\infty \longrightarrow 0, \quad \|h_{\bar{f}_{n,0}} - h_{f_0}\|_\infty \longrightarrow 0.$$

PROOF. For the proof of the first result, note that for $(t, z) \in A_i \times B_j$

$$\begin{aligned} |\bar{f}_{n,0}(t, z) - f_0(t, z)| &= \left| k_n l_n \int_{A_i} \int_{B_j} f_0(u, v) du dv - f_0(t, z) \right| \\ &\leq k_n l_n \int_{A_i} \int_{B_j} |f_0(u, v) - f_0(t, z)| du dv \\ &\leq \max_{(u,v) \in A_i \times B_j} |f_0(u, v) - f_0(t, z)|. \end{aligned}$$

Using (B.1) and the uniform continuity of f_0 on $\mathcal{W}_M$ we get for $(t, z) \in \mathcal{W}_M$

$$|\bar{f}_{n,0}(t, z) - f_0(t, z)| \leq \max_{i,j} \max_{(u,v) \in A_i \times B_j} |f_0(u, v) - f_0(t, z)| \longrightarrow 0,$$

uniformly in (t, z) as $n \rightarrow \infty$. This implies the first result.

For the proof of the second result, note that for $(t, z) \in \mathcal{W}_M$

$$\begin{aligned} |h_{\bar{f}_{n,0}}(t, z) - h_{f_0}(t, z)| &= \left| g(t) \left\{ \mathbb{1}_{\{z>0\}} \int_{u=0}^t (\bar{f}_{n,0}(u, z) - f_0(u, z)) \, du \right. \right. \\ &\quad \left. \left. + \mathbb{1}_{\{z=0\}} \int_{u=t}^{M_1} \int_{z=0}^{M_2} (\bar{f}_{n,0}(u, z) - f_0(u, z)) \, dz \, du \right\} \right| \\ &\leq \|g\|_\infty \{M_1 \|\bar{f}_{n,0} - f_0\|_\infty + M_1 M_2 \|\bar{f}_{n,0} - f_0\|_\infty\}. \end{aligned}$$

Since this upper bound does not depend on t and z , the second result now follows from the first. $\square$

LEMMA 4.7. Define for $(t, z) \in A_i \times B_j$

$$\bar{h}_n(t, z) = \mathbb{E} \hat{h}_n(t, z) = k_n l_n \int_{A_i} \int_{B_j} h_{f_0}(u, v) \, d\lambda(u, v),$$

then, under the conditions of Lemma 4.3,

$$\|h_{\bar{f}_{n,0}} - \bar{h}_n\|_\infty \longrightarrow 0. \quad (4.13)$$

PROOF. First, note that

$$\|h_{\bar{f}_{n,0}} - \bar{h}_n\|_\infty \leq \|h_{\bar{f}_{n,0}} - h_{f_0}\|_\infty + \|h_{f_0} - \bar{h}_n\|_\infty.$$

The first term converges to zero by Lemma 4.6. We now prove that the second term also converges to zero. To see this, note that

$$\|h_{f_0} - \bar{h}_n\|_\infty \leq \|h_1 - \bar{h}_{1,n}\|_\infty + \|h_0 - \bar{h}_{0,n}\|_\infty.$$

Similarly as the first result in Lemma 4.6, we have for $(t, z) \in A_i \times B_j$

$$\begin{aligned} |h_1(t, z) - \bar{h}_{1,n}(t, z)| &= \left| k_n l_n \int_{A_i} \int_{B_j} (h_1(t, z) - h_1(u, v)) \, dv \, du \right| \\ &\leq k_n l_n \int_{A_i} \int_{B_j} |h_1(t, z) - h_1(u, v)| \, dv \, du \\ &\leq \max_{(u,v) \in A_i \times B_j} |h_1(t, z) - h_1(u, v)|. \end{aligned}$$

Both g and $\partial_2 F_0$ are uniformly continuous, hence h_1 is as well and with condition (B.1) we get for $(t, z) \in \mathcal{W}_M$

$$|h_1(t, z) - \bar{h}_{1,n}(t, z)| \leq \max_{i,j} \max_{(u,v) \in A_i \times B_j} |h_1(t, z) - h_1(u, v)| \longrightarrow 0,$$

uniformly in (t, z) as $n \rightarrow \infty$. Via a similar argument we get that

$$|h_0(t) - \bar{h}_{0,n}(t)| \leq \max_i \max_{u \in A_i} |h_0(t) - h_0(u)| \longrightarrow 0,$$

uniformly in t as $n \rightarrow \infty$, hence

$$\|h_{f_0} - \bar{h}_n\|_\infty \longrightarrow 0.$$

□

LEMMA 4.8. *Under the conditions of Lemma 4.3 such that h_{f_0} satisfies (4.7),*

$$\mathcal{K}(\hat{h}_n, h_{\bar{f}_{n,0}}) \xrightarrow{\mathcal{P}} 0. \quad (4.14)$$

PROOF. We can write

$$\begin{aligned} \mathcal{K}(\hat{h}_n, h_{\bar{f}_{n,0}}) &= \int \hat{h}_n(t, z) \log \frac{\hat{h}_n(t, z)}{h_{\bar{f}_{n,0}}(t, z)} d\lambda(t, z) \\ &= \int \hat{h}_n(t, z) \log \frac{\hat{h}_n(t, z)}{h_{f_0}(t, z)} d\lambda(t, z) - \int \hat{h}_n(t, z) \log \frac{\bar{h}_n(t, z)}{h_{f_0}(t, z)} d\lambda(t, z) \\ &\quad - \int \hat{h}_n(t, z) \log \frac{h_{\bar{f}_{n,0}}(t, z)}{\bar{h}_n(t, z)} d\lambda(t, z). \end{aligned} \quad (4.15)$$

The expectation of the first term converges to zero by BARRON *et al.* (1992, Theorem 5) with $\mu = H_{f_0}$ (the distribution function of the observable vector W), $\nu = \lambda$ as defined in (1.9) and $\mathcal{P} = \{[0, M_1] \times \{0\}, [0, M_1] \times (0, M_2]\}$.

By Fubini's theorem, the expectation of the second term equals

$$\int E \hat{h}_n(t, z) \log \frac{\bar{h}_n(t, z)}{h_{f_0}(t, z)} d\lambda(t, z) = \int \bar{h}_n(t, z) \log \frac{\bar{h}_n(t, z)}{h_{f_0}(t, z)} d\lambda(t, z) = \mathcal{K}(\bar{h}_n, h_{f_0}).$$

This converges to zero by BARRON *et al.* (1992, Theorem 4), so also the expectation of the second term in (4.15) converges to zero.

By (4.7), $\bar{h}_n(t, z) \geq c_{\bar{h}} > 0$ for all $(t, z) \in \mathcal{W}_M$, so that (4.13) implies that for any $\varepsilon > 0$ and n sufficiently large

$$1 - \frac{\varepsilon}{c_{\bar{h}}} \leq 1 + \frac{h_{\bar{f}_{n,0}}(t, z) - \bar{h}_n(t, z)}{\bar{h}_n(t, z)} \leq 1 + \frac{\varepsilon}{c_{\bar{h}}}.$$

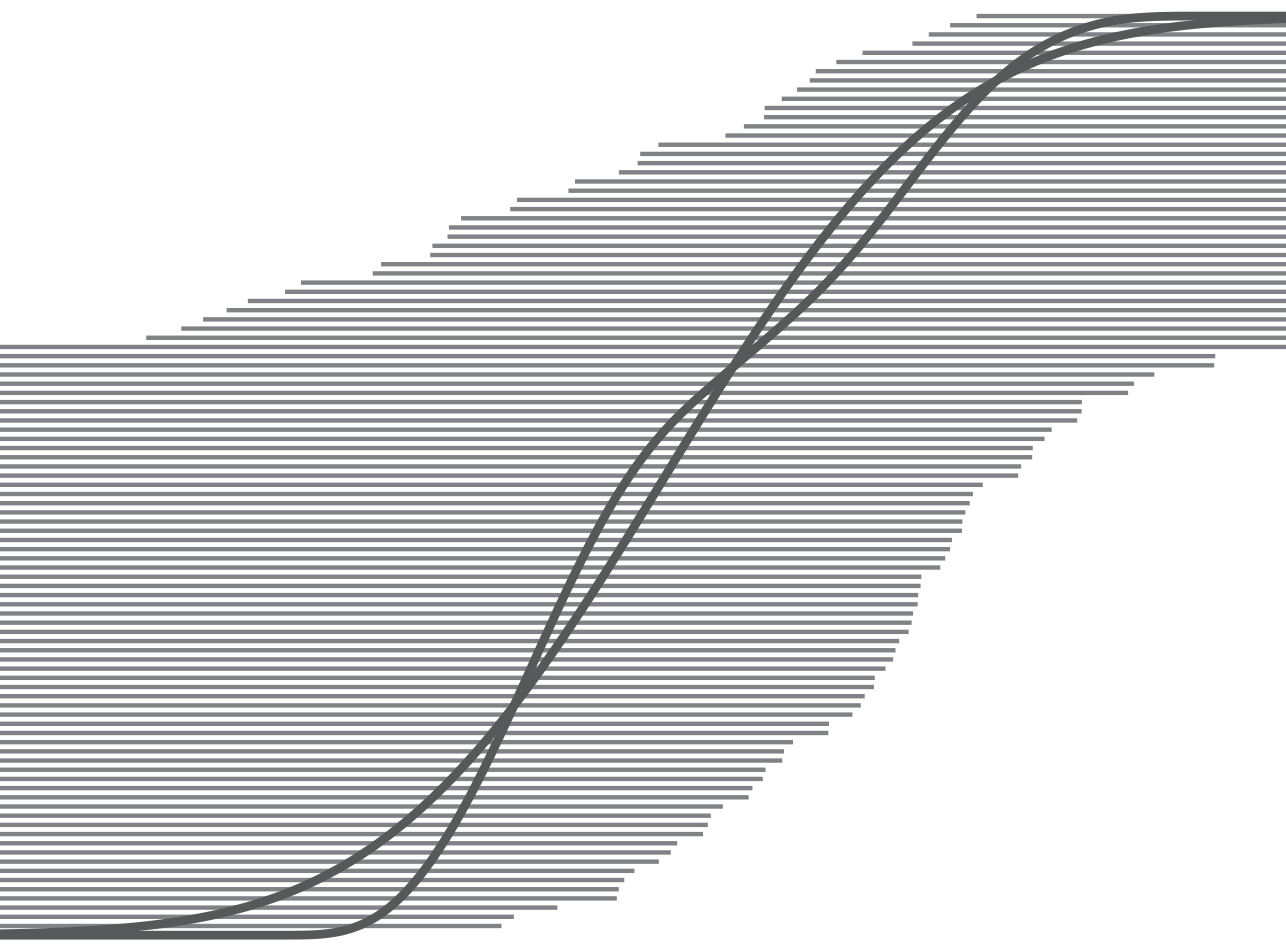
Then

$$\begin{aligned} \log \left(1 - \frac{\varepsilon}{c_{\bar{h}}} \right) &= \int \bar{h}_n(t, z) \log \left(1 - \frac{\varepsilon}{c_{\bar{h}}} \right) d\lambda(t, z) \\ &\leq \int \bar{h}_n(t, z) \log \left(1 + \frac{h_{\bar{f}_{n,0}}(t, z) - \bar{h}_n(t, z)}{\bar{h}_n(t, z)} \right) d\lambda(t, z) \\ &\leq \int \bar{h}_n(t, z) \log \left(1 + \frac{\varepsilon}{c_{\bar{h}}} \right) d\lambda(t, z) = \log \left(1 + \frac{\varepsilon}{c_{\bar{h}}} \right), \end{aligned}$$

so that also the expectation of the third term in (4.15) converges to zero. Therefore, the expectation of $\mathcal{K}(\hat{h}_n, h_{\bar{f}_{n,0}})$ converges to zero, and because $\mathcal{K}(\hat{h}_n, h_{\bar{f}_{n,0}}) \geq 0$ a.s. the convergence in (4.14) now follows. $\square$

CHAPTER 5

SMOOTH PLUG-IN INVERSE ESTIMATION



Chapter 4 studies a Maximum Smoothed Likelihood Estimator (MSLE) for the distribution function F_0 in the current status continuous mark model, based on a histogram-type estimator of the observation density. This estimator is proved to be consistent. Unfortunately, the asymptotic distribution of the MSLE is unknown. In this chapter, a different estimator for F_0 in the current status continuous mark is studied. For this estimator, the smooth plug-in inverse estimator, the asymptotic distribution is derived. Results in this chapter can also be found in GROENEBOOM, JONGBLOED AND WITTE (2011b).

In Section 5.1 two inverse estimators for F_0 are defined, one is based on univariate kernel smoothing and the other on bivariate kernel smoothing. Pointwise consistency of both estimators is proved in Section 5.2. Furthermore, it is proved that the estimator based on bivariate kernel smoothing behaves like a bivariate distribution function for n sufficiently large. The asymptotic distributions of both estimators is derived in Section 5.3. In Section 5.4 the smooth plug-in inverse estimators are compared with the MSLE studied in Chapter 4 and the binned Maximum Likelihood Estimator (MLE) of MAATHUIS AND WELLNER (2008). Concluding remarks on the smooth plug-in inverse estimators and the MSLE can be found in Section 5.5, and technical lemmas are given in Section 5.6.

5.1 DEFINITION OF THE ESTIMATORS

Let (X, Y) be a random vector with bivariate distribution function F_0 . In the current status continuous mark model, see also Section 1.2, this vector (X, Y) is not observed directly. Instead, a censoring variable T is observed and the indicator variable $\Delta = \mathbb{1}_{\{X \leq T\}}$. When X is smaller than T (i.e. $\Delta = 1$), the variable Y is observed, otherwise Y remains unobservable. The vector $W = (T, Z)$ with $Z = \Delta \cdot Y$ is called the observable vector and its density with respect to the measure λ defined in (1.9) equals

$$\begin{aligned} h_{f_0}(t, z) &= \mathbb{1}_{\{z > 0\}} g(t) \partial_2 F_0(t, z) + \mathbb{1}_{\{z = 0\}} g(t) (1 - F_{0,X}(t)) \\ &=: \mathbb{1}_{\{z > 0\}} h_1(t, z) + \mathbb{1}_{\{z = 0\}} h_0(t), \end{aligned} \quad (5.1)$$

where $F_{0,X}$ is the marginal distribution of X and $\partial_2 F_0(t, z) = \frac{\partial}{\partial z} F_0(t, z)$. More general, for convenience of notation, we denote the j th partial derivative with respect to x_i of a function F by $\partial_i^j F$, i.e.

$$\partial_i^j F(x_1, x_2) = \frac{\partial^j}{\partial y_i^j} F(y_1, y_2) \Big|_{(y_1, y_2) = (x_1, x_2)},$$

and omit j when $j = 1$.

Based on the definition $h_1(t, z) = g(t) \partial_2 F_0(t, z)$, we can express the bivariate dis-

tribution function F_0 of (X, Y) in terms of the (sub-)densities g and h_1

$$F_0(t, z) = \frac{1}{g(t)} \int_{-\infty}^z h_1(t, v) dv.$$

Then, we define the plug-in inverse estimator for F_0 in the current status continuous mark model as

$$\hat{F}(t, z) = \frac{1}{\hat{g}(t)} \int_{-\infty}^z \hat{h}_1(t, v) dv,$$

where $\hat{g}$ and $\hat{h}_1$ are estimators for g and h_1 , respectively.

Before explicitly choosing the estimators $\hat{g}$ and $\hat{h}_1$, we introduce some notation. Throughout the paper k denotes a univariate kernel density, $\tilde{k}$ a bivariate kernel density and $b_1 = b_{1,n}$ and $b_2 = b_{2,n}$ positive bandwidths (or smoothing parameters) depending on the sample size n . Denote by k_{b_1} and $\tilde{k}_{b_1, b_2}$ the rescaled versions of k and $\tilde{k}$, i.e.

$$k_{b_1}(u) = \frac{1}{b_1} k(u/b_1), \text{ and } \tilde{k}_{b_1, b_2}(u, v) = \frac{1}{b_1 b_2} \tilde{k}(u/b_1, v/b_2).$$

Furthermore, we define

$$m_2(k) = \int u^2 k(u) du, \quad m_2(\tilde{k}) = \int w_1^2 \tilde{k}(w_1, w_2) dw_1.$$

We estimate g and h_1 by their respective univariate and bivariate kernel (sub-)density estimators

$$\hat{g}_n(t_0) = \frac{1}{n} \sum_{i=1}^n k_{b_1}(t_0 - T_i), \quad \hat{h}_{n,1}^{(2)}(t_0, z_0) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{\{Z_i > 0\}} \tilde{k}_{b_1, b_2}(t_0 - T_i, z_0 - Z_i).$$

The plug-in inverse estimator then becomes

$$\begin{aligned} \hat{F}_n^{(2)}(t_0, z_0) &= \frac{\int_{-\infty}^{z_0} \hat{h}_{n,1}^{(2)}(t_0, z) dz}{\hat{g}_n(t_0)} \\ &= \frac{\int_{-\infty}^{z_0} \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{\{Z_i > 0\}} \tilde{k}_{b_1, b_2}(t_0 - T_i, z - Z_i) dz}{\frac{1}{n} \sum_{i=1}^n k_{b_1}(t_0 - T_i)}. \end{aligned} \quad (5.2)$$

Here, superscript 2 in the notation for the plug-in estimator refers to the fact that there is smoothing in two directions.

In Section 5.3 we also consider a less natural, but computationally and asymptotically more tractable estimator using an estimate for the numerator $\int_{-\infty}^{z_0} h_1(t_0, z) dz$ based on smoothing only in the t -direction, i.e. when we estimate it by

$$\frac{1}{n} \sum_{i=1}^n \mathbb{1}_{(0, z_0]}(Z_i) k_{b_1}(t_0 - T_i).$$

The plug-in inverse estimator then becomes

$$\hat{F}_n^{(1)}(t_0, z_0) = \frac{\frac{1}{n} \sum_{i=1}^n \mathbb{1}_{(0, z_0]}(Z_i) k_{b_1}(t_0 - T_i)}{\frac{1}{n} \sum_{i=1}^n k_{b_1}(t_0 - T_i)}. \quad (5.3)$$

Superscript 1 in the notation for this estimator refers to the fact that there is only smoothing in one direction. Note that if we take $k(y) = \frac{1}{2} \mathbb{1}_{[-1, 1]}(y)$, (5.3) results in

$$\hat{F}_n^{(1)}(t_0, z_0) = \frac{\int_{u \in A_n} \int_{z \leq z_0} \mathbb{1}_{\{z > 0\}} d\mathbb{H}_n(u, z)}{\int_{u \in A_n} \int_{z \geq 0} d\mathbb{H}_n(u, z)},$$

where $\mathbb{H}_n$ is the empirical distribution of the observations $(T_1, Z_1), \dots, (T_n, Z_n)$ and $A_n = A_n(t_0) = [t_0 - b_1, t_0 + b_1]$. This estimator is the total number of observations T_i in A_n with Z -value in $(0, z_0]$ divided by the total number of observations (T_i, Z_i) in the strip $A_n \times [0, \infty)$. Figure 5.1 illustrates the estimator $\hat{F}_n^{(1)}$ for this choice of kernel for two simulated samples, one of sizes $n = 10$ and one of size $n = 100$. As smoothing parameter we took $b_1 = 0.65$ and $b_1 = 0.40$, respectively.

Throughout this chapter, the estimators and the results are illustrated with simulated data. In these simulations, the true density f_0 is equal to

$$f_0(x, y) = x + y, \text{ for } (x, y) \in [0, 1]^2.$$

The censoring times have the following density

$$g(t) = 2t, \text{ for } t \in [0, 1].$$

Note that these estimators are not true bivariate distribution functions, as they decrease locally in the t -direction. Monotonicity of a bivariate function is a necessary (but not sufficient) condition in order to be a bivariate distribution function, hence these estimators can be seen as naive estimators. The estimator $\hat{F}_n^{(2)}$ can also have this undesirable naive behavior.

It is very natural to define the kernel density k in terms of the kernel density $\tilde{k}$ as stated in assumption (K.1):

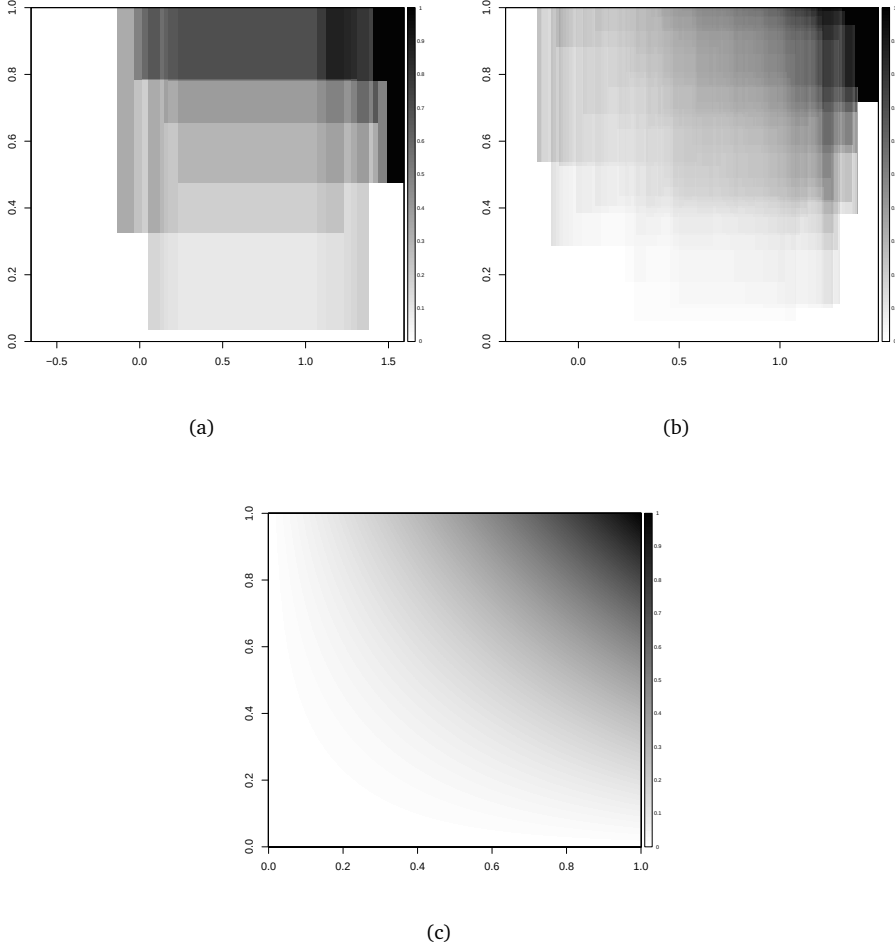


FIGURE 5.1: Two illustrations of the estimator $\hat{F}_n^{(1)}$ for simulated samples of size $n = 10$ with $b_1 = 0.65$ (a) and of size $n = 100$ with $b_1 = 0.40$ (b). The true distribution function F_0 is given in (c).

(K.1) Let $\tilde{k}$ be a bivariate kernel density, then the kernel density k is defined as

$$k(u_1) = \int \tilde{k}(u_1, u_2) du_2.$$

Indeed, if (K.1) holds, the estimator $\hat{F}_n^{(2)}$ also satisfies the inverse relation that follows from substituting $z = 0$ in (5.1)

$$h_0(t) = g(t)(1 - F_{0,X}(t)).$$

To see this, note that we have that

$$\begin{aligned} \hat{g}_n(t_0) &= \frac{1}{n} \sum_{i=1}^n k_{b_1}(t_0 - T_i) \\ &= \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{\{Z_i=0\}} k_{b_1}(t_0 - T_i) + \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{\{Z_i>0\}} k_{b_1}(t_0 - T_i) \\ &= \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{\{Z_i=0\}} k_{b_1}(t_0 - T_i) + \int_{-\infty}^{\infty} \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{\{Z_i>0\}} \tilde{k}_{b_1, b_2}(t_0 - T_i, z - Z_i) dz. \end{aligned}$$

If we define $\hat{h}_{n,0}^{(2)}(t_0) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{\{Z_i=0\}} k_{b_1}(t_0 - T_i)$ as an estimator for the sub-density h_0 in (5.1), then

$$\hat{F}_{n,X}^{(2)}(t_0) = \hat{F}_n^{(2)}(t_0, \infty) = \frac{\int_{-\infty}^{\infty} \hat{h}_{n,1}^{(2)}(t_0, z) dz}{\hat{g}_n(t_0)} = \frac{\hat{g}_n(t_0) - \hat{h}_{n,0}^{(2)}(t_0)}{\hat{g}_n(t_0)} = 1 - \frac{\hat{h}_{n,0}^{(2)}(t_0)}{\hat{g}_n(t_0)}.$$

Note that the product kernel $\tilde{k}(u_1, u_2) = k_1(u_1)k_2(u_2)$ for univariate kernels k_1 and k_2 satisfies this condition for $k = k_1$. In the illustrations of the estimators in this chapter, the kernel density k is taken to be the Epanechnikov kernel, i.e. $k(u) = \frac{3}{4}(1 - u^2)$ for $-1 \leq u \leq 1$, except for Figure 5.1. The kernel density $\tilde{k}$ is taken to be the product kernel of two Epanechnikov kernels, i.e. $\tilde{k}(u_1, u_2) = k(u_1)k(u_2)$.

Define

$$\hat{h}_n^{(2)}(t, z) = \mathbb{1}_{\{z>0\}} \hat{h}_{n,1}^{(2)}(t, z) + \mathbb{1}_{\{z=0\}} \hat{h}_{n,0}^{(2)}(t),$$

then it can be seen that

$$h_{\hat{F}_n^{(2)}}(t, z) = \mathbb{1}_{\{z>0\}} \hat{g}_n(t) \partial_2 \hat{F}_n^{(2)}(t, z) + \mathbb{1}_{\{z=0\}} \hat{g}_n(t) (1 - \hat{F}_n^{(2)}(t, \infty)) = \hat{h}_n^{(2)}(t, z).$$

This implies that

$$\mathcal{K}(\hat{h}_n^{(2)}, h_{\hat{F}_n^{(2)}}) = \mathcal{K}(\hat{h}_n^{(2)}, \hat{h}_n^{(2)}) = 0,$$

hence the smooth plug-in inverse estimator $\hat{F}_n^{(2)}$ can also be seen as a minimal Kullback-Leibler projection (hence as a MSLE). However, $\hat{F}_n^{(2)}$ might decrease locally in t -direction, and therefore is a naive MSLE. By a similar argument, it can be seen that also $\hat{F}_n^{(1)}$ is a naive MSLE where the log likelihood is smoothed only in t direction.

5.2 CONSISTENCY AND MONOTONICITY

In this section we prove that the estimator $\hat{F}_n^{(1)}$ is uniformly consistent and that the estimator $\hat{F}_n^{(2)}$ is pointwise consistent. Furthermore, we prove that for appropriate choices of the bandwidths and n sufficiently large, $\hat{F}_n^{(2)}$ will have all properties of a bivariate distribution function on a large subset of $[0, \infty)^2$, with arbitrarily high probability. To derive these results for $\hat{F}_n^{(2)}$, we assume the distribution function of interest F_0 and the censoring density g satisfy the following conditions.

- (F.1) The Lebesgue density f_0 of F_0 exists for all $(t, z) \in [0, \infty)^2$.
- (G.1) Let $\mathcal{S}_{0,X}^\circ$ denote the interior of the support of the marginal density $f_{0,X}$ of X . On $\mathcal{S}_{0,X}^\circ$, the density g satisfies $0 < g < \infty$ and its derivative g' is uniformly continuous and bounded.

We also impose some conditions on the kernel densities k and $\tilde{k}$, as well as a condition on the smoothing parameters b_1 and b_2 .

- (K.2) The kernel density k has compact support $[-1, 1]$, is continuous and symmetric around 0.
- (K.3) The kernel density $\tilde{k}$ has compact support $[-1, 1]^2$, is continuous and satisfies

$$\int w_i \tilde{k}(w_1, w_2) dw_i = 0 \quad (i = 1, 2), \quad \int w_2^2 \tilde{k}(w_1, w_2) dw_2 = \int w_1^2 \tilde{k}(w_1, w_2) dw_1.$$

- (B.1) The smoothing parameters $b_1 = b_{1,n}$ and $b_2 = b_{2,n}$ converge to zero as $n \rightarrow \infty$

$$\lim_{n \rightarrow \infty} b_{1,n} = \lim_{n \rightarrow \infty} b_{2,n} = 0, \quad \lim_{n \rightarrow \infty} n b_{1,n} = \lim_{n \rightarrow \infty} n b_{2,n} = \infty.$$

Note that the product kernel $\tilde{k}(u, v) = k_1(u)k_2(v)$ for univariate kernels k_1 and k_2 with compact support $[-1, 1]$ that are continuous and symmetric around 0 satisfies condition (K.3) if $m_2(k_1) = m_2(k_2)$. The product of two Epanechnikov kernels, used in the illustrations, certainly satisfies this latter condition.

THEOREM 5.1. *Assume F_0 and g satisfy conditions (F.1) and (G.1). Also assume k is defined via relation (K.1) and satisfies condition (K.2). Furthermore, let b_1 and b_2 satisfy condition (B.1). Then $\hat{F}_n^{(1)}$ is uniformly consistent and $\hat{F}_n^{(2)}(t_0, z_0)$ is pointwise consistent for fixed $t_0 \in \mathcal{S}_{0,X}^\circ$ and $z_0 > 0$.*

PROOF. The uniform consistency of $\hat{F}_n^{(1)}$ follows from HÄRDLE *et al.* (1988, Theorem 3.2), where an estimator is considered for the conditional distribution $F(y|x)$ of Y given X which is similarly defined as our estimator $\hat{F}_n^{(1)}$.

To prove the pointwise consistency of $\hat{F}_n^{(2)}$, fix $t_0 \in \mathcal{S}_{0,X}^\circ$ and $z_0 > 0$ and note that for this t_0 , we have $g(t_0) > 0$ by (G.1). Combining the pointwise consistency of $\int_{-\infty}^{z_0} \hat{h}_{n,1}^{(2)}(t_0, z) dz$, see Lemma 5.6, the pointwise consistency of $\hat{g}_n$ [see e.g., PARZEN (1962, p. 1070)] and Slutsky's lemma we have

$$\hat{F}_n^{(2)}(t_0, z_0) = \frac{\int_{-\infty}^{z_0} \hat{h}_{n,1}^{(2)}(t_0, z) dz}{\hat{g}_n(t_0)} \xrightarrow{\mathcal{P}} \frac{\int_{-\infty}^{z_0} h_1(t_0, z) dz}{g(t_0)} = F_0(t_0, z_0).$$

□

Any function $F : \mathbb{R}^2 \rightarrow [0, 1]$ that satisfies $F(x, y) \rightarrow 0$ for $x, y \rightarrow -\infty$, $F(x, y) \rightarrow 1$ for $x, y \rightarrow \infty$ and

$$\forall x_1 < x_2, y_1 < y_2 : F(x_2, y_2) - F(x_2, y_1) - F(x_1, y_2) + F(x_1, y_1) \geq 0 \quad (5.4)$$

is a bivariate distribution function. Condition (5.4) requires that each rectangle $[x_1, x_2] \times [y_1, y_2]$ has a nonnegative mass. In Theorem 5.2 below, we prove the estimator $\hat{F}_n^{(2)}$ satisfies condition (5.4) asymptotically. This is done by proving that the Lebesgue density $\hat{f}_n^{(2)}$ is positive, with probability converging to one. The estimator $\hat{F}_n^{(1)}$ does not have a density with respect to Lebesgue measure λ_2 , hence we can not prove that $\hat{F}_n^{(1)}$ satisfies condition (5.4) asymptotically by considering its Lebesgue density. To prove Theorem 5.2, we need stronger conditions on b_1 and b_2 than condition (B.1).

(B.2) The smoothing parameters $b_1 = b_{1,n}$ and $b_2 = b_{2,n}$ converge to zero as $n \rightarrow \infty$ and satisfy

$$\lim_{n \rightarrow \infty} n b_{1,n}^2 b_{2,n}^2 = \infty, \quad \lim_{n \rightarrow \infty} n b_{1,n}^3 b_{2,n} = \infty.$$

Note that sequences b_1 and b_2 satisfying condition (B.2) also satisfy condition (B.1) and $n b_{1,n}^3 \rightarrow \infty$.

THEOREM 5.2. Assume F_0 and g satisfy conditions (F1) and (G.1). Also assume k and $\tilde{k}$ satisfy conditions (K.2) and (K.3). In addition, assume k' and $\partial_1 \tilde{k}$ are uniformly continuous. Furthermore, let b_1 and b_2 satisfy condition (B.2). Let $\mathcal{S} \subset [0, \infty)^2$ be compact and such that f_0 is uniformly continuous on an open subset that contains $\mathcal{S}$ and for all $\delta > 0$, $\mathcal{M}_\delta = \{(t, z) \in [0, \infty)^2 : f_0(t, z) \geq 2\delta\} \cap \mathcal{S}$. Then for $\delta > 0$,

$$\mathbb{P} \left(\forall (t, z) \in \mathcal{M}_\delta : \hat{f}_n^{(2)}(t, z) > \frac{l_g^2}{2u_g^2} \delta \right) \rightarrow 1, \quad (5.5)$$

where $\hat{f}_n^{(2)}$ is the Lebesgue density of $\hat{F}_n^{(2)}$ and l_g and u_g are as defined in Lemma 5.8.

PROOF. Fix $\delta > 0$. First note that since

$$\frac{\partial^2}{\partial t \partial z} \int_{-\infty}^z \hat{h}_{n,1}^{(2)}(t, v) dv = \partial_1 \hat{h}_{n,1}^{(2)}(t, z)$$

we have the following expression for $\hat{f}_n^{(2)}$

$$\hat{f}_n^{(2)}(t, z) = \frac{\partial^2}{\partial u \partial v} \hat{F}_n^{(2)}(u, v) \Big|_{(u,v)=(t,z)} = \frac{\hat{g}_n(t) \partial_1 \hat{h}_{n,1}^{(2)}(t, z) - \hat{g}'_n(t) \hat{h}_{n,1}^{(2)}(t, z)}{\hat{g}_n(t)^2}.$$

We first consider the numerator and prove that

$$P(\forall (t, z) \in \mathcal{M}_\delta : \hat{g}_n(t) \partial_1 \hat{h}_{n,1}^{(2)}(t, z) - \hat{g}'_n(t) \hat{h}_{n,1}^{(2)}(t, z) > 2l_g^2 \delta) \longrightarrow 1.$$

For this, note that for all $(t, z) \in \mathcal{M}_\delta$

$$\begin{aligned} \hat{g}_n(t) \partial_1 \hat{h}_{n,1}^{(2)}(t, z) - \hat{g}'_n(t) \hat{h}_{n,1}^{(2)}(t, z) &= \hat{g}_n(t) (\partial_1 \hat{h}_{n,1}^{(2)}(t, z) - \partial_1 h_1(t, z)) \\ &\quad + \hat{h}_{n,1}^{(2)}(t, z) (g'(t) - \hat{g}'_n(t)) + \partial_1 h_1(t, z) (\hat{g}_n(t) - g(t)) \\ &\quad + g'(t) (h_1(t, z) - \hat{h}_{n,1}^{(2)}(t, z)) + g(t) \partial_1 h_1(t, z) - g'(t) h_1(t, z) \\ &\geq - \sup_{t \in \text{proj}_X \mathcal{M}_\delta} \hat{g}_n(t) \sup_{(t,z) \in \mathcal{M}_\delta} |\partial_1 \hat{h}_{n,1}^{(2)}(t, z) - \partial_1 h_1(t, z)| \\ &\quad - \sup_{(t,z) \in \mathcal{M}_\delta} \hat{h}_{n,1}^{(2)}(t, z) \sup_{t \in \text{proj}_X \mathcal{M}_\delta} |g'(t) - \hat{g}'_n(t)| \\ &\quad - \sup_{(t,z) \in \mathcal{M}_\delta} \partial_1 h_1(t, z) \sup_{t \in \text{proj}_X \mathcal{M}_\delta} |\hat{g}_n(t) - g(t)| \\ &\quad - \sup_{t \in \text{proj}_X \mathcal{M}_\delta} g'(t) \sup_{(t,z) \in \mathcal{M}_\delta} |h_1(t, z) - \hat{h}_{n,1}^{(2)}(t, z)| + g(t) \partial_1 h_1(t, z) - g'(t) h_1(t, z), \end{aligned}$$

with $\text{proj}_X \mathcal{M}_\delta = \{t : (t, z) \in \mathcal{M}_\delta \text{ for some } z\}$. By Lemma 5.9 all random terms converge to zero in probability. Since $g(t) \partial_1 h_1(t, z) - g'(t) h_1(t, z) = g(t)^2 f_0(t, z)$ we have that the last term is bounded below by $\inf_{(t,z) \in \mathcal{M}_\delta} g(t)^2 f_0(t, z) \geq 2l_g^2 \delta$ by Lemma 5.8.

By Lemma 5.8 and the uniform consistency of $\hat{g}_n$ [see Lemma 5.9], we have that $0 < \frac{1}{2} l_g < \hat{g}_n(t) < 2u_g < \infty$ for all $t \in \text{proj}_X \mathcal{M}_\delta$ with probability converging to one. This implies that for all $(t, z) \in \mathcal{M}_\delta$

$$\hat{f}_n^{(2)}(t, z) \geq \frac{\hat{g}_n(t) \partial_1 \hat{h}_{n,1}^{(2)}(t, z) - \hat{g}'_n(t) \hat{h}_{n,1}^{(2)}(t, z)}{4u_g^2} > \frac{2l_g^2}{4u_g^2} \delta,$$

with probability converging to one. Hence (5.5) follows. $\square$

REMARK. If, in addition to condition (F1), we assume that f_0 is uniformly continuous on $[0, \infty)^2$, this theorem implies that for each $\delta > 0$ and $M > 0$, the restriction of $\hat{F}_n^{(2)}$ to the set $\{(t, z) \in [0, M]^2 : f_0(t, z) \geq \delta\}$ will asymptotically be the restriction to this set of a bivariate distribution function $\tilde{F}_n$ on $[0, \infty)^2$.

5.3 ASYMPTOTIC DISTRIBUTIONS OF THE ESTIMATORS

In this section we derive the asymptotic distribution of both plug-in inverse estimators. Although the estimator $\hat{F}_n^{(2)}$ is more natural, we start with the estimator $\hat{F}_n^{(1)}$ since deriving its asymptotic distribution is more tractable. Subsequently, we prove that for certain choices of the smoothing parameter b_2 the estimators $\hat{F}_n^{(1)}$ and $\hat{F}_n^{(2)}$ are asymptotically equivalent, yielding the asymptotic distribution of $\hat{F}_n^{(2)}$.

THEOREM 5.3. Assume k satisfies condition (K.2). Fix $t_0, z_0 > 0$ such that $\partial_1^2 F_0(t, z)$ and $g''(t)$ exist and are continuous in a neighborhood of (t_0, z_0) and t_0 , respectively, and $\partial_1^2 F_0(t_0, z_0) + 2g'(t_0)\partial_1 F_0(t_0, z_0)/g(t_0) \neq 0$ and $g(t_0) > 0$. Then for $b_1 = cn^{-1/5}$ ($c > 0$),

$$n^{2/5}(\hat{F}_n^{(1)}(t_0, z_0) - F_0(t_0, z_0)) \rightsquigarrow \mathcal{N}(\mu_1, \sigma^2)$$

where

$$\mu_1 = \frac{1}{2}c^2 m_2(k) \left\{ \partial_1^2 F_0(t_0, z_0) + 2 \frac{g'(t_0)\partial_1 F_0(t_0, z_0)}{g(t_0)} \right\}, \quad (5.6)$$

$$\sigma^2 = c^{-1} \frac{F_0(t_0, z_0)(1 - F_0(t_0, z_0))}{g(t_0)} \int k(u)^2 du. \quad (5.7)$$

PROOF. Note that

$$\hat{F}_n^{(1)}(t_0, z_0) = \varphi \left(\frac{1}{n} \sum_{i=1}^n k_{b_1}(t_0 - T_i), \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{(0, z_0]}(Z_i) k_{b_1}(t_0 - T_i) \right),$$

for $\varphi(u, v) = v/u$, hence the pointwise asymptotic distribution of $\hat{F}_n^{(1)}$ follows immediately by applying the Delta-method to (5.13) in Lemma 5.10. $\square$

REMARK. In case $\partial_1^2 F_0(t_0, z_0) + 2g'(t_0)\partial_1 F_0(t_0, z_0)/g(t_0) = 0$, the rate of convergence changes because the bias is of a different asymptotic order as was also noted for $\hat{F}_n^{SM}$ in the remark following Theorem 2.5.

Figure 5.2 illustrates this pointwise asymptotic result by a simulation study. The value of $n^{2/5}(\hat{F}_n^{(1)}(0.5, 0.5) - F_0(0.5, 0.5))$ was calculated for a sample of size $n = 5000$ with $b_1 = 0.09$ as smoothing parameter. This was repeated $B = 1000$ times. Figure 5.2 shows the QQ-plot (with the line $y = \mu_1 + x\sigma$) as well as the histogram (with the $\mathcal{N}(\mu_1, \sigma^2)$ -density) of these B values.

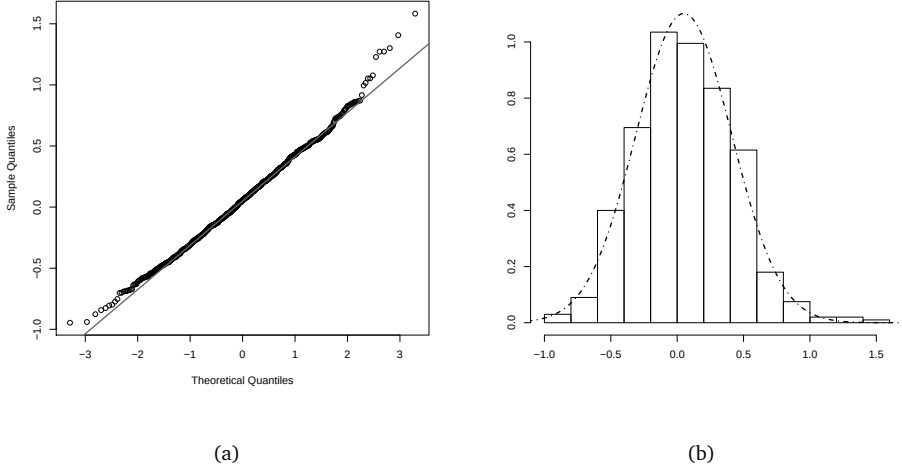


FIGURE 5.2: QQ-plot (a) and histogram (b) of $B = 1000$ values $n^{2/5}(\hat{F}_n^{(1)}(0.5, 0.5) - F_0(0.5, 0.5))$ for a simulation study with sample size $n = 5000$ and $b_1 = 0.09$.

Under definition (K.1) and assumptions (K.2) and (K.3) on the kernel densities k and $\tilde{k}$, we can prove that for $t_0, z_0 > 0$ fixed $n^{2/5}(\hat{F}_n^{(2)}(t_0, z_0) - \hat{F}_n^{(1)}(t_0, z_0))$ converges to zero in probability whenever b_2 converges faster to zero than $n^{-1/5}$. As a consequence, these estimators are (first order) asymptotically equivalent. For b_2 tending to zero slower than $n^{-1/5}$, $n^{2/5}|\hat{F}_n^{(2)}(t_0, z_0) - \hat{F}_n^{(1)}(t_0, z_0)| \rightarrow \infty$ in probability. These results are more precisely stated in Theorem 5.4 and Corollary 5.5.

THEOREM 5.4. Assume k and $\tilde{k}$ satisfy conditions (K.2) and (K.3). Fix $t_0, z_0 > 0$ such that $\partial_2^2 F_0(t, z)$ and $g(t)$ exist and are continuous in a neighborhood of (t_0, z_0) and t_0 , respectively, and $\partial_2^2 F_0(t_0, z_0) \neq 0$ and $g(t_0) > 0$. Let $b_1 = c_1 n^{-1/5}$ and $b_2 = c_2 n^{-\alpha}$ ($c_1, c_2 > 0$), then for $\alpha > 1/5$

$$n^{2/5}(\hat{F}_n^{(2)}(t_0, z_0) - \hat{F}_n^{(1)}(t_0, z_0)) \xrightarrow{\mathcal{P}} 0,$$

for $\alpha = 1/5$

$$n^{2/5}(\hat{F}_n^{(2)}(t_0, z_0) - \hat{F}_n^{(1)}(t_0, z_0)) \xrightarrow{\mathcal{P}} \frac{1}{2}c_2^2 m_2(\tilde{k}) \partial_2^2 F_0(t_0, z_0)$$

while for $\alpha < 1/5$

$$n^{2/5}|\hat{F}_n^{(2)}(t_0, z_0) - \hat{F}_n^{(1)}(t_0, z_0)| \xrightarrow{\mathcal{P}} \infty.$$

PROOF OF THEOREM 5.4. For $i = 1, 2$, let $N_n^{(i)}(t_0, z_0)$ be the numerator in the definitions (5.3) and (5.2) of $\hat{F}_n^{(i)}(t_0, z_0)$ at a fixed point (t_0, z_0) , and note that we can write

$$\begin{aligned} N_n^{(2)}(t_0, z_0) &= \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{(0, z_0 - b_2]}(Z_i) \int_{-\infty}^{z_0} \tilde{k}_{b_1, b_2}(t_0 - T_i, z - Z_i) dz \\ &\quad + \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{(z_0 - b_2, z_0]}(Z_i) \int_{-\infty}^{z_0} \tilde{k}_{b_1, b_2}(t_0 - T_i, z - Z_i) dz \\ &\quad + \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{(z_0, z_0 + b_2]}(Z_i) \int_{-\infty}^{z_0} \tilde{k}_{b_1, b_2}(t_0 - T_i, z - Z_i) dz \\ &= \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{(0, z_0 - b_2]}(Z_i) \int_{-\infty}^{z_0} \tilde{k}_{b_1, b_2}(t_0 - T_i, z - Z_i) dz \\ &\quad + \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{(z_0 - b_2, z_0]}(Z_i) \int_{-\infty}^{Z_i + b_2} \tilde{k}_{b_1, b_2}(t_0 - T_i, z - Z_i) dz \\ &\quad - \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{(z_0 - b_2, z_0]}(Z_i) \int_{z_0}^{Z_i + b_2} \tilde{k}_{b_1, b_2}(t_0 - T_i, z - Z_i) dz \\ &\quad + \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{(z_0, z_0 + b_2]}(Z_i) \int_{-\infty}^{z_0} \tilde{k}_{b_1, b_2}(t_0 - T_i, z - Z_i) dz \\ &= \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{(0, z_0]}(Z_i) k_{b_1}(t_0 - T_i, z_0) \\ &\quad - \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{(z_0 - b_2, z_0]}(Z_i) \int_{z_0}^{Z_i + b_2} \tilde{k}_{b_1, b_2}(t_0 - T_i, z - Z_i) dz \\ &\quad + \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{(z_0, z_0 + b_2]}(Z_i) \int_{Z_i - b_2}^{z_0} \tilde{k}_{b_1, b_2}(t_0 - T_i, z - Z_i) dz. \end{aligned}$$

In the last equality we use (K.1), so that

$$\begin{aligned}
 R_n &= N_n^{(2)}(t_0, z_0) - N_n^{(1)}(t_0, z_0) \\
 &= \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{(z_0, z_0+b_2]}(Z_i) \int_{Z_i-b_2}^{z_0} \tilde{k}_{b_1, b_2}(t_0 - T_i, z - Z_i) dz \\
 &\quad - \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{(z_0-b_2, z_0]}(Z_i) \int_{z_0}^{Z_i+b_2} \tilde{k}_{b_1, b_2}(t_0 - T_i, z - Z_i) dz \\
 &=: \frac{1}{n} \sum_{i=1}^n U_i.
 \end{aligned}$$

First we consider the variance of $n^{2/5}R_n$. Observe that

$$\begin{aligned}
 |U_i| &\leq \mathbb{1}_{(z_0, z_0+b_2]}(Z_i) \int_{Z_i-b_2}^{z_0} \tilde{k}_{b_1, b_2}(t_0 - T_i, z - Z_i) dz \\
 &\quad + \mathbb{1}_{(z_0-b_2, z_0]}(Z_i) \int_{z_0}^{Z_i+b_2} \tilde{k}_{b_1, b_2}(t_0 - T_i, z - Z_i) dz \\
 &\leq \mathbb{1}_{(z_0, z_0+b_2]}(Z_i) \int_{Z_i-b_2}^{Z_i+b_2} \tilde{k}_{b_1, b_2}(t_0 - T_i, z - Z_i) dz \\
 &\quad + \mathbb{1}_{(z_0-b_2, z_0]}(Z_i) \int_{Z_i-b_2}^{Z_i+b_2} \tilde{k}_{b_1, b_2}(t_0 - T_i, z - Z_i) dz \\
 &= \mathbb{1}_{(z_0-b_2, z_0+b_2]}(Z_i) k_{b_1}(t_0 - T_i) := S_i,
 \end{aligned}$$

with

$$\begin{aligned}
 \mathbb{E} S_i^2 &= \int_u \int_v \mathbb{1}_{(z_0-b_2, z_0+b_2]}(v) k_{b_1}(t_0 - u)^2 h_1(u, v) dv du \\
 &= b_1^{-1} b_2 2h_1(t_0, z_0) \int k(x)^2 dx + \mathcal{O}(b_2).
 \end{aligned}$$

Since

$$\text{Var } R_n = \frac{1}{n} \text{Var } U_1 = \frac{1}{n} (\mathbb{E} U_1^2 - (\mathbb{E} U_1)^2) \leq \frac{1}{n} \mathbb{E} S_1^2 = \mathcal{O}(n^{-1} b_1^{-1} b_2),$$

$\text{Var } n^{2/5}R_n \rightarrow 0$ for $b_1 = n^{-1/5}$ and $b_2 = n^{-\alpha}$ with $\alpha > 0$.

Now we consider the expectation of U_i .

$$\begin{aligned}
\mathbb{E} U_i &= \int_v \int_u \mathbb{1}_{(z_0, z_0 + b_2]}(v) \int_{v-b_2}^{z_0} \tilde{k}_{b_1, b_2}(t_0 - u, z - v) dz h_1(u, v) du dv \\
&\quad - \int_v \int_u \mathbb{1}_{(z_0 - b_2, z_0]}(v) \int_{z_0}^{v+b_2} \tilde{k}_{b_1, b_2}(t_0 - u, z - v) dz h_1(u, v) du dv \\
&= b_2 \int_{y=-1}^0 \int_{x=-1}^1 \int_{w=-1}^y \tilde{k}(x, w) dw h_1(t_0 - b_1 x, z_0 - b_2 y) dx dy \\
&\quad - b_2 \int_{y=0}^1 \int_{x=-1}^1 \int_{w=y}^1 \tilde{k}(x, w) dw h_1(t_0 - b_1 x, z_0 - b_2 y) dx dy \\
&= b_2 h_1(t_0, z_0) \left\{ \int_{x=-1}^1 \int_{y=-1}^0 \int_{w=-1}^y \tilde{k}(x, w) dw dy dx - \right. \\
&\quad \left. - \int_{x=-1}^1 \int_{y=0}^1 \int_{w=y}^1 \tilde{k}(x, w) dw dy dx \right\} \\
&\quad - b_2 b_1 \partial_1 h_1(t_0, z_0) \left\{ \int_{x=-1}^1 \int_{y=-1}^0 \int_{w=-1}^y x \tilde{k}(x, w) dw dy dx - \right. \\
&\quad \left. - \int_{x=-1}^1 \int_{y=0}^1 \int_{w=y}^1 x \tilde{k}(x, w) dw dy dx \right\} \\
&\quad - b_2^2 \partial_2 h_1(t_0, z_0) \left\{ \int_{x=-1}^1 \int_{y=-1}^0 \int_{w=-1}^y y \tilde{k}(x, w) dw dy dx - \right. \\
&\quad \left. - \int_{x=-1}^1 \int_{y=0}^1 \int_{w=y}^1 y \tilde{k}(x, w) dw dy dx \right\} \\
&\quad + \mathcal{O}(b_2 b_1^2) + \mathcal{O}(b_2^2 b_1) + \mathcal{O}(b_2^3) \\
&= -b_2 h_1(t_0, z_0) \int_{x=-1}^1 \int_{w=-1}^1 w \tilde{k}(x, w) dw dx \\
&\quad + b_2 b_1 \partial_1 h_1(t_0, z_0) \int_{x=-1}^1 \int_{w=-1}^1 x w \tilde{k}(x, w) dw dx \\
&\quad + \frac{1}{2} b_2^2 \partial_2 h_1(t_0, z_0) \int_{x=-1}^1 \int_{w=-1}^1 w^2 \tilde{k}(x, w) dw dx \\
&\quad + \mathcal{O}(b_2 b_1^2) + \mathcal{O}(b_2^2 b_1) + \mathcal{O}(b_2^3)
\end{aligned}$$

where the last equality follows from changing the order of integration. By condition (K.3), the first two integrals are zero and the last integral equals $m_2(\tilde{k})$, so that

$$En^{2/5}R_n \longrightarrow \begin{cases} \pm\infty & \text{for } \alpha < 1/5, \\ \frac{1}{2}c_2^2m_2(\tilde{k})g(t_0)\partial_2^2F_0(t_0, z_0) & \text{for } \alpha = 1/5, \\ 0 & \text{for } \alpha > 1/5. \end{cases}$$

Applying Slutsky's Lemma to

$$n^{2/5}(\hat{F}_n^{(2)}(t_0, z_0) - \hat{F}_n^{(1)}(t_0, z_0)) = \frac{n^{2/5}R_n}{\hat{g}_n(t_0)}$$

gives the result. $\square$

Figure 5.3 illustrates this result by a simulation study. The value of $n^{2/5}(\hat{F}_n^{(2)}(0.5, 0.5) - \hat{F}_n^{(1)}(0.5, 0.5))$ was calculated for several values of the sample size n with $b_1 = \frac{1}{2}n^{-1/5}$ and $b_2 = \frac{1}{2}n^{-1/3}$ as smoothing parameters. The lines are the lines $\pm \frac{1}{2}n^{-1/6}$, the order of the standard deviation of $n^{2/5}(\hat{F}_n^{(2)}(0.5, 0.5) - \hat{F}_n^{(1)}(0.5, 0.5))$ [see the proof of Theorem 5.4 in Section 5.6].

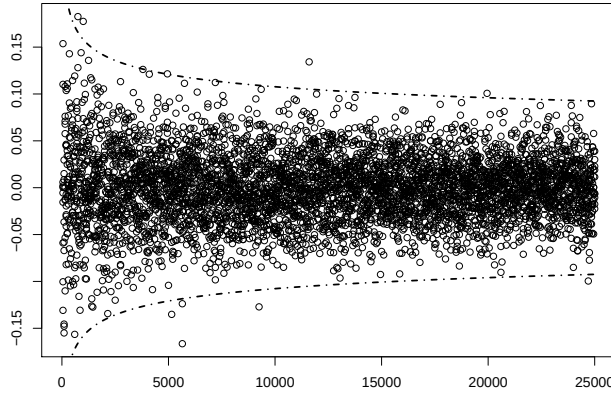


FIGURE 5.3: Illustration of the difference $n^{2/5}(\hat{F}_n^{(2)}(0.5, 0.5) - \hat{F}_n^{(1)}(0.5, 0.5))$ as a function of the sample size n with $b_1 = \frac{1}{2}n^{-1/5}$ and $b_2 = \frac{1}{2}n^{-1/3}$ for a simulation study.

As a consequence of Theorem 5.4, the estimators $\hat{F}_n^{(1)}$ and $\hat{F}_n^{(2)}$ are pointwise asymptotically equivalent for $\alpha > 1/5$, while for $\alpha = 1/5$, $\hat{F}_n^{(2)}(t_0, z_0)$ has an additional (possibly negative) asymptotic bias term.

COROLLARY 5.5. Let $b_1 = c_1 n^{-1/5}$ and $b_2 = c_2 n^{-\alpha}$ ($c_1, c_2 > 0$). Then, under the conditions of Theorem 5.4, for $\alpha > 1/5$

$$n^{2/5}(\hat{F}_n^{(2)}(t_0, z_0) - F_0(t_0, z_0)) \rightsquigarrow \mathcal{N}(\mu_1, \sigma^2)$$

where μ_1 and σ^2 are defined in (5.6) and (5.7) (with $c = c_1$). For $\alpha = 1/5$

$$n^{2/5}(\hat{F}_n^{(2)}(t_0, z_0) - F_0(t_0, z_0)) \rightsquigarrow \mathcal{N}(\mu_2, \sigma^2),$$

where

$$\mu_2 = \mu_1 + \frac{1}{2}c_2^2 m_2(\tilde{k}) \partial_2^2 F_0(t_0, z_0).$$

PROOF. This immediately follows from Theorem 5.4. $\square$

Figure 5.4 illustrates this pointwise asymptotic result by a simulation study. The value of $n^{2/5}(\hat{F}_n^{(2)}(0.5, 0.5) - F_0(0.5, 0.5))$ was calculated for a sample of size $n = 5000$ with $b_1 = 0.09$ and $b_2 = 0.03$ as smoothing parameters. This was repeated $B = 1000$ times. Figure 5.4 shows the QQ-plot (with the line $y = \mu_1 + x\sigma$) as well as the histogram (with the $\mathcal{N}(\mu_1, \sigma^2)$ -density) of these B values.

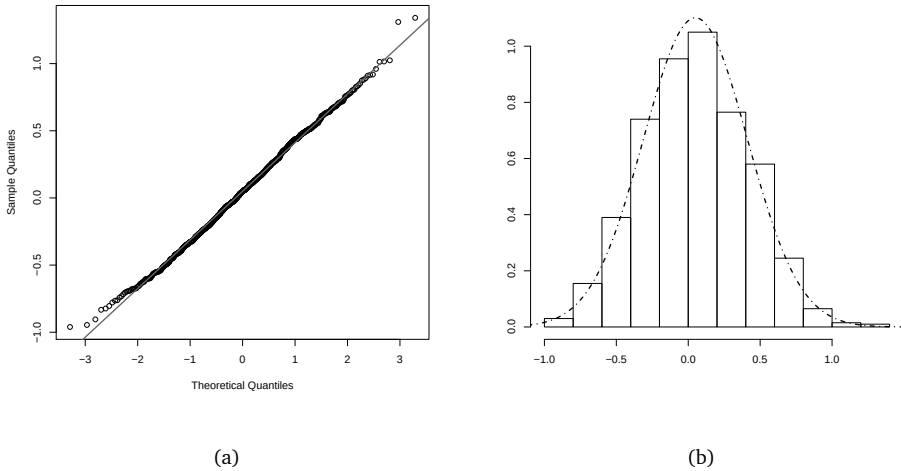


FIGURE 5.4: QQ-plot (a) and histogram (b) of $B = 1000$ values $n^{2/5}(\hat{F}_n^{(2)}(0.5, 0.5) - F_0(0.5, 0.5))$ for a simulation study with sample size $n = 5000$ and $b_1 = 0.09$, $b_2 = 0.03$.

5.4 SMALL COMPARATIVE SIMULATION STUDY

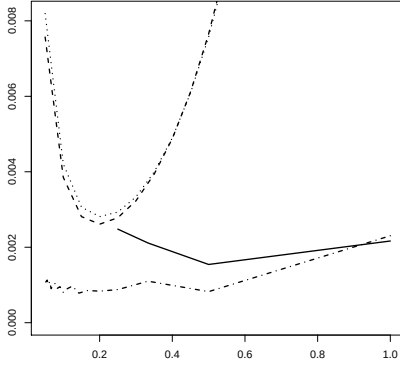
The estimators $\hat{F}_n^{(1)}$ and $\hat{F}_n^{(2)}$ are asymptotically equivalent for certain choices of the smoothing parameter b_2 , but which estimator should one use in practice? To get some insight in the differences between the estimator, we run a simulation study. We simulated data according to $F_0(x, y) = \frac{1}{2}xy(x + y)$ for $x, y \in [0, 1]$ and $g(t) = 2t$ for $t \in [0, 1]$ for different sample sizes $n = 500$, $n = 1000$, $n = 5000$ and $n = 10000$. For each simulation we computed the estimators $\hat{F}_n^{(1)}(t_0, z_0)$ and $\hat{F}_n^{(2)}(t_0, z_0)$ for different values of (t_0, z_0) and different values of the smoothing parameters b_1 and b_2 . We repeated this $B = 250$ times, resulting in 250 estimates $\hat{F}_{n,b_1,b_2}^{(i),1}(t_0, z_0), \hat{F}_{n,b_1,b_2}^{(i),2}(t_0, z_0), \dots, \hat{F}_{n,b_1,b_2}^{(i),250}(t_0, z_0)$ ($i = 1, 2$) for each value of the smoothing parameters b_1 and b_2 . Then, we estimated the Mean Squared Error (MSE) of the estimator $\hat{F}_n^{(i)}(t_0, z_0)$ by

$$\frac{1}{B} \sum_{j=1}^B \left(\hat{F}_{n,b_1,b_2}^{(i),j}(t_0, z_0) - F_0(t_0, z_0) \right)^2.$$

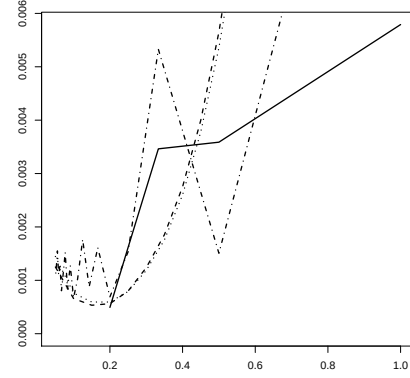
Table 5.1 shows the minimum value of the estimated MSE for each estimator, for each n and in different points (t_0, z_0) . It also highlights the smallest of these values per n and (t_0, z_0) . The binned MLE $\tilde{F}_n$ suggested by MAATHUIS AND WELLNER (2008) and the MSLE $\tilde{F}_n^{MS}$ studied in Chapter 4 are included in this simulation study.

Figure 5.5 shows the resulting values of estimated MSE as function of b_1 . For $\tilde{F}_n$, the smoothing parameter b_1 is the binwidth in z direction, for $\tilde{F}_n^{MS}$ we have that b_1 and b_2 are the binwidths in t - and z -direction, respectively. Both $\hat{F}_n^{(2)}$ and $\hat{F}_n^{MS}$ depend on two smoothing parameters, Figure 5.5 only shows the estimated MSE as function of the first smoothing parameter b_1 , for that value of the second parameter b_2 that has the smallest value of the estimated MSE.

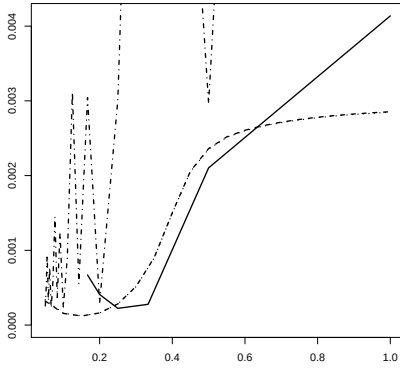
This simulation study, of which some results are also illustrated in Figure 5.5, shows that the binned MLE $\tilde{F}_n$ has the smallest (estimated) MSE for small values of t_0 and all values of z_0 , whenever the sample size is not too big ($n < 5000$). For bigger sample sizes, the minimal values of the MSE of $\tilde{F}_n$ is comparable to the MSE of the smooth plug-in inverse estimators $\hat{F}_n^{(1)}$ and $\hat{F}_n^{(2)}$. The MSEs of $\hat{F}_n^{(1)}$ and $\hat{F}_n^{(2)}$ are comparable except for big t_0 or big z_0 , when the smoothing parameter b_2 is small. If the smoothing parameter b_2 is bigger, $\hat{F}_n^{(1)}$ has a smaller MSE. When t_0 is big, the shapes of the MSE as function of the smoothing parameter b_1 are quit different (except when also z_0 is big) but the minimal MSE values are similar. When z_0 is big, the shapes of the MSE are similar, but $\hat{F}_n^{(2)}$ has a much smaller MSE, even for bigger values of b_2 . The minimal value of the MSE of the MSLE $\tilde{F}_n^{MS}$ is comparable to the minimal values of the smooth plug-in inverse estimators, in some cases a bit smaller and in other cases a bit bigger.



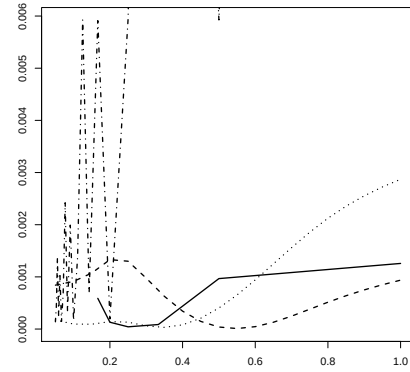
(a) $t_0 = 0.2$, $n = 500$, $b_2 = 0.2$ for $\hat{F}_n^{(2)}$ and $b_2 = 0.25$ for $\hat{F}_n^{MS}$



(b) $t_0 = 0.4$, $n = 1000$, $b_2 = 0.2$ for $\hat{F}_n^{(2)}$ and $b_2 = 0.5$ for $\hat{F}_n^{MS}$



(c) $t_0 = 0.6$, $n = 5000$, $b_2 = 0.05$ for $\hat{F}_n^{(2)}$ and $b_2 = 0.17$ for $\hat{F}_n^{MS}$



(d) $t_0 = 0.8$, $n = 10000$, $b_2 = 1.0$ for $\hat{F}_n^{(2)}$ and $b_2 = 0.2$ for $\hat{F}_n^{MS}$

FIGURE 5.5: The estimated MSE as function of the smoothing parameter b_1 for the smooth plug-in inverse estimators $\hat{F}_n^{(1)}$ (dotted line) and $\hat{F}_n^{(2)}$ (dashed line), the MSLE $\hat{F}_n^{MS}$ (solid line) and the binned MLE $\tilde{F}_n$ (dash-dotted line) for different values of n and in different points $(t_0, 0.6)$ for a simulation study.

		$\hat{F}_n^{(1)}(t_0, 0.6)$	$\hat{F}_n^{(2)}(t_0, 0.6)$	$\hat{F}_n^{MS}(t_0, 0.6)$	$\tilde{F}_n(t_0, 0.6)$
$t_0 = 0.2$	$n = 500$	2.81×10^{-3}	2.61×10^{-3}	1.54×10^{-3}	7.84×10^{-4}
	$n = 1000$	1.53×10^{-3}	1.51×10^{-3}	1.40×10^{-3}	2.01×10^{-4}
	$n = 5000$	2.04×10^{-4}	1.98×10^{-4}	6.11×10^{-4}	1.49×10^{-4}
	$n = 10000$	9.59×10^{-5}	9.56×10^{-5}	3.57×10^{-4}	1.13×10^{-4}
$t_0 = 0.4$	$n = 500$	9.07×10^{-4}	8.24×10^{-4}	1.45×10^{-3}	1.21×10^{-3}
	$n = 1000$	5.94×10^{-4}	5.33×10^{-4}	4.92×10^{-4}	6.74×10^{-4}
	$n = 5000$	1.32×10^{-4}	1.29×10^{-4}	1.24×10^{-4}	2.37×10^{-4}
	$n = 10000$	8.95×10^{-5}	8.91×10^{-5}	1.23×10^{-4}	1.35×10^{-4}
$t_0 = 0.6$	$n = 500$	8.21×10^{-4}	7.82×10^{-4}	5.51×10^{-4}	1.38×10^{-3}
	$n = 1000$	5.31×10^{-4}	5.31×10^{-4}	4.13×10^{-4}	7.79×10^{-4}
	$n = 5000$	1.21×10^{-4}	1.21×10^{-4}	2.24×10^{-4}	2.11×10^{-4}
	$n = 10000$	9.21×10^{-5}	9.14×10^{-5}	1.32×10^{-4}	1.31×10^{-4}
$t_0 = 0.8$	$n = 500$	5.91×10^{-4}	1.80×10^{-4}	6.53×10^{-4}	1.39×10^{-3}
	$n = 1000$	3.14×10^{-4}	1.02×10^{-4}	3.47×10^{-4}	8.59×10^{-4}
	$n = 5000$	5.61×10^{-5}	1.74×10^{-5}	1.09×10^{-4}	2.27×10^{-4}
	$n = 10000$	3.25×10^{-5}	9.86×10^{-6}	4.27×10^{-5}	1.39×10^{-4}

TABLE 5.1: Minimum values of the estimated MSE of the smooth plug-in inverse estimators $\hat{F}_n^{(1)}$ and $\hat{F}_n^{(2)}$, the MSLE $\hat{F}_n^{MS}$ and the binned MLE $\tilde{F}_n$ for different values of n and in different points $(t_0, 0.6)$ for a simulation study. The boldfaced values in each row are the minimal value of the MSEs in that row.

5.5 CONCLUDING REMARKS

All the estimators $\hat{F}_n^{(1)}$, $\hat{F}_n^{(2)}$, $\hat{F}_n^{MS}$ and $\tilde{F}_n$ compared in Section 5.4 are pointwise consistent estimators for the underlying true bivariate distribution function F_0 . But they all have advantages as well as disadvantages. The plug-in inverse estimator $\hat{F}_n^{(2)}$ asymptotically has a positive Lebesgue density on large subsets of $[0, \infty)^2$. This means that “isotonisation” of the estimator $\hat{F}_n^{(2)}$ is not necessary asymptotically, although for finite sample size n it need not be a bivariate distribution function. Put differently, any common shape regularized version of our estimator is asymptotically equivalent with our estimator. However, this only holds asymptotically, and for finite sample size n it might be desirable to have an estimator which is a true bivariate distribution function, satisfying condition (5.4). For example when one wants to sample from an estimator in smoothed bootstrap methods. To satisfy the properties of a bivariate distribution function, the estimator $\hat{F}_n^{(2)}$ can be regularized for example by minimizing $L_2(F) = \|F - \hat{F}_n^{(2)}\|_2$ over the class of bivariate distribution functions. However,

it is unclear to us whether there exists a PAVA-like algorithm [see, e.g. BARLOW *et al.* (1972)] to obtain the minimizer and whether the asymptotic distribution of $\hat{F}_n^{(2)}$ carries over to it. Hence the MSLE $\hat{F}_n^{MS}$ and the binned MLE are good alternatives; they are bivariate distribution functions by definition. These estimators can only be computed using iterative algorithms, whereas the plug-in inverse estimators $\hat{F}_n^{(1)}$ and $\hat{F}_n^{(2)}$ can be computed directly using the data.

For certain choices of the bandwidths b_1 and b_2 , the estimators $\hat{F}_n^{(1)}$ and $\hat{F}_n^{(2)}$ are pointwise asymptotically equivalent, i.e. they have the same pointwise asymptotic distribution. For other choices of the bandwidths, both estimators converge pointwise to a normal distribution with the same variance but a different bias. For example, when $b_1 \sim n^{-1/5}$ and b_2 converges faster to zero than b_1 , the asymptotic bias of $\hat{F}_n^{(1)}(t_0, z_0)$ equals

$$\mu_1(t_0, z_0) = \frac{1}{2}c^2m_2(k) \left\{ \partial_1^2 F_0(t_0, z_0) + 2 \frac{g'(t_0)\partial_1 F_0(t_0, z_0)}{g(t_0)} \right\},$$

and the asymptotic bias of $\hat{F}_n^{(2)}(t_0, z_0)$ equals

$$\mu_2(t_0, z_0) = \frac{1}{2}c^2m_2(k) \left\{ \partial_1^2 F_0(t_0, z_0) + 2 \frac{g'(t_0)\partial_1 F_0(t_0, z_0)}{g(t_0)} \right\} + \frac{1}{2}c_2^2m_2(\tilde{k})\partial_2^2 F_0(t_0, z_0).$$

Thus, the asymptotic bias of $\hat{F}_n^{(2)}(t_0, z_0)$ equals the asymptotic bias of $\hat{F}_n^{(1)}(t_0, z_0)$ plus the term $\frac{1}{2}c_2^2m_2(\tilde{k})\partial_2^2 F_0(t_0, z_0)$. This additional term, depending on the unknown F_0 and the point (t_0, z_0) we estimate at, can be positive for some choices of (t_0, z_0) and negative for other choices of (t_0, z_0) . Therefore, smoothing in one direction does not always yield an estimator with smaller asymptotic Mean Squared Error (aMSE) than smoothing in two directions. More specifically, one estimator is not uniformly better than the other estimator. In Chapters 2 and 3 we found similar results for estimators for the distribution function, its Lebesgue density and the hazard rate in the current status model.

5.6 TECHNICAL LEMMAS

LEMMA 5.6. For $i = 1, 2$, let $N_n^{(i)}(t, z_0)$ be the numerator in the definitions (5.2) and (5.3) of $\hat{F}_n^{(i)}(t, z_0)$. Under the conditions of Theorem 5.1, we have for h_1 defined in (5.1)

$$N_n^{(i)}(t_0, z_0) \xrightarrow{\mathcal{P}} \int_{-\infty}^{z_0} h_1(t_0, z) dz, \quad i = 1, 2. \quad (5.8)$$

PROOF Recall that

$$N_n^{(2)}(t_0, z_0) = \int_{-\infty}^{z_0} \hat{h}_{n,1}^{(2)}(t_0, z) dz = \int_{-\infty}^{z_0} \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{(0, z_0 + b_2]}(Z_i) \tilde{k}_{b_1, b_2}(t_0 - T_i, z - Z_i) dz.$$

Using (K.1), this is bounded from below by

$$\int_{-\infty}^{z_0} \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{(0, z_0 - b_2]}(Z_i) \tilde{k}_{b_1, b_2}(t_0 - T_i, z - Z_i) dz = N_n^{(1)}(t_0, z_0 - b_2),$$

and from above by

$$\int_{-\infty}^{\infty} \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{(0, z_0 + b_2]}(Z_i) \tilde{k}_{b_1, b_2}(t_0 - T_i, z - Z_i) dz = N_n^{(1)}(t_0, z_0 + b_2),$$

so that we have

$$N_n^{(1)}(t_0, z_0 - b_2) \leq N_n^{(2)}(t_0, z_0) \leq N_n^{(1)}(t_0, z_0 + b_2). \quad (5.9)$$

By conditions (F1) and (G.1), we have

$$\begin{aligned} \mathbb{E} N_n^{(1)}(t_0, z_0 \pm b_2) &= \mathbb{E} \mathbb{1}_{(0, z_0 \pm b_2]}(Z_1) k_{b_1}(t_0 - T_1) = \int_{-\infty}^{z_0 \pm b_2} h_1(t_0, z) dz + \mathcal{O}(b_1), \\ \text{Var} N_n^{(1)}(t_0, z_0 \pm b_2) &= n^{-1} \text{Var} \mathbb{1}_{(0, z_0 \pm b_2]}(Z_1) k_{b_1}(t_0 - T_1) = \mathcal{O}(n^{-1} b_1^{-1}). \end{aligned}$$

Since $b_{1,n}, b_{2,n} \rightarrow 0$ and $n b_{1,n} \rightarrow \infty$ as $n \rightarrow \infty$, we now have that

$$\lim_{n \rightarrow \infty} \mathbb{E} N_n^{(1)}(t_0, z_0 \pm b_2) = \int_{-\infty}^{z_0} h_1(t_0, z) dz, \quad \lim_{n \rightarrow \infty} \text{Var} N_n^{(1)}(t_0, z_0 \pm b_2) = 0,$$

implying that $N_n^{(1)}(t_0, z_0 \pm b_2) \xrightarrow{\mathcal{P}} \int_{-\infty}^{z_0} h_1(t_0, z) dz$ and (5.8) follows for $i = 2$ by (5.9). Since $N_n^{(1)}(t_0, z)$ is increasing in z , (5.9) also holds with $N_n^{(2)}$ replaced by $N_n^{(1)}$ and (5.8) follows for $i = 1$. $\square$

LEMMA 5.7. Assume that F_0 and g satisfy conditions (F1) and (G.1). Let $\mathcal{S}$ and $\mathcal{M}_\delta$ be as defined in Theorem 5.2. Then $\text{proj}_X \mathcal{M}_\delta = \{t : (t, z) \in \mathcal{M}_\delta \text{ for some } z\}$ is a closed subset of $\mathcal{S}_{0,X}^\circ$.

PROOF Fix $\delta > 0$. To prove that $\text{proj}_X \mathcal{M}_\delta$ is a closed subset of $\mathcal{S}_{0,X}^\circ$ we prove that

- (i) $\mathcal{M}_\delta$ is closed in $[0, \infty)^2$,
- (ii) $\text{proj}_X \mathcal{M}_\delta$ is closed in $[0, \infty)$,
- (iii) $\text{proj}_X \mathcal{M}_\delta$ is a subset of $\mathcal{S}_{0,X}^\circ$.

We now start with proving (i). By definition of $\mathcal{S}$, there exists an open set $\mathcal{U} \supset \mathcal{S}$ on which f_0 is uniformly continuous. Define

$$\mathcal{U}_\delta = \{(t, z) \in \mathcal{U} : f_0(t, z) \geq 2\delta\} = f_0^{-1}[[2\delta, \infty)].$$

The function f_0 is continuous on $\mathcal{U}$, hence $\mathcal{U}_\delta$ is closed. Since we also have that

$$\mathcal{M}_\delta = \{(t, z) \in [0, \infty)^2 : f_0(t, z) \geq 2\delta\} \cap \mathcal{S} = \mathcal{U}_\delta \cap \mathcal{S},$$

$\mathcal{M}_\delta$ is the intersection of two closed sets, hence closed itself.

For proving (ii), assume $\text{proj}_X \mathcal{M}_\delta$ is not closed. Then, there exists a sequence $(x_n)_n \in \text{proj}_X \mathcal{M}_\delta$ with $x_n \rightarrow x \notin \text{proj}_X \mathcal{M}_\delta$. By (i), the set $\mathcal{M}_\delta$ is closed, hence by definition of $\text{proj}_X \mathcal{M}_\delta$ there exists a sequence $(x_n, y_n)_n \in \mathcal{M}_\delta$. By compactness of $\mathcal{M}_\delta$ (this follows from (i)), there exists a subsequence $(n_k)_k$ and $(x, y) \in \mathcal{M}_\delta$ such that $(x_{n_k}, y_{n_k})_k \rightarrow (x, y)$. From this it follows that $x \in \text{proj}_X \mathcal{M}_\delta$. This yields a contradiction, hence $\text{proj}_X \mathcal{M}_\delta$ is closed.

To prove (iii), first note that there exists $\eta > 0$ such that for all $(t, z), (s, y) \in \mathcal{M}_\delta$

$$\|(t, z) - (s, y)\| < \eta \implies |f_0(t, z) - f_0(s, y)| < \delta,$$

by uniform continuity of f_0 on $\mathcal{M}_\delta$. Now take $t \in \text{proj}_X \mathcal{M}_\delta$. Then for all s in a small neighborhood of t and $z_s > 0$ such that $(s, z_s) \in \mathcal{M}_\delta$

$$f_{0,X}(s) = \int_{-\infty}^{\infty} f_0(s, z) dz \geq \int_{[z_s, z_s + \eta/2]} f_0(s, z) dz \geq \delta\eta/2,$$

hence $t \in \mathcal{S}_{0,X}^\circ$. □

LEMMA 5.8. Assume that F_0 and g satisfy conditions (F1) and (G.1). Let $\mathcal{M}_\delta$ be as defined in Theorem 5.2. Then

$$\exists l_g > 0, u_g < \infty : l_g \leq g(t) \leq u_g \text{ for all } t \in \text{proj}_X \mathcal{M}_\delta. \quad (5.10)$$

PROOF. The set $\text{proj}_X \mathcal{M}_\delta$ is a closed subset of $\mathcal{S}_{0,X}^\circ$ by Lemma 5.7. On $\mathcal{S}_{0,X}^\circ$, we have that $0 < g < \infty$, hence also on $\text{proj}_X \mathcal{M}_\delta$. Now assume (5.10) does not hold. Then,

there exists a sequence $(t_n)_n \rightarrow t \in \text{proj}_X \mathcal{M}_\delta$ such that $g(t_n) \rightarrow 0$. By the uniform continuity of g , this implies that $g(t) = 0$, yielding a contradiction.

The existence of u_g follows immediately from (G.1). $\square$

LEMMA 5.9. *Assume F_0 and g satisfy conditions (F.1) and (G.1). Also assume k and $\tilde{k}$ satisfy conditions (K.2) and (K.3). In addition, assume k' and $\partial_1 \tilde{k}$ are uniformly continuous. Furthermore, let b_1 and b_2 satisfy condition (B.2). Let $\mathcal{S}$ be as defined in Theorem 5.2 and define $\|f\|_{\mathcal{S},\infty} = \sup_{(x,y) \in \mathcal{S}} |f(x,y)|$ and $\mathcal{S}_X = \text{proj}_X \mathcal{S}$. Then*

$$\|\hat{g}_n - g\|_{\mathcal{S}_X,\infty} \xrightarrow{\mathcal{P}} 0, \quad \|\hat{g}'_n - g'\|_{\mathcal{S}_X,\infty} \xrightarrow{\mathcal{P}} 0 \quad (5.11)$$

$$\|\hat{h}_{n,1}^{(2)} - h_1\|_{\mathcal{S},\infty} \xrightarrow{\mathcal{P}} 0, \quad \|\partial_1 \hat{h}_{n,1}^{(2)} - \partial_1 h_1\|_{\mathcal{S},\infty} \xrightarrow{\mathcal{P}} 0 \quad (5.12)$$

PROOF. The results in (5.11) follow directly from Theorems A.2 and A.3. The results in (5.12) follow Theorems A.5 and A.6

The results in Theorems A.5 and A.6 hold for density estimators, whereas the estimator $\hat{h}_{n,1}^{(2)}$ is a sub-density. However, the results in (5.12) follow from those after defining a binomially distributed sample size N_1 and reason similarly as the proof of (3.19). $\square$

LEMMA 5.10. *Under the conditions of Theorem 5.3 it holds that*

$$n^{2/5} \left(\left(\frac{1}{n} \sum_{i=1}^n k_{b_1}(t_0 - T_i) \right) - \left(\frac{g(t_0)}{h_1(t_0, z_0)} \right) \right) \rightsquigarrow \mathcal{N}(\mu, \Sigma_1), \quad (5.13)$$

with

$$\mu = \frac{1}{2} c^2 m_2(k) \left(\frac{g''(t_0)}{\partial_1^2 h_1(t_0, z_0)} \right) \quad (5.14)$$

$$\Sigma_1 = c^{-1} \int k(u)^2 du \begin{pmatrix} g(t_0) & h_1(t_0, z_0) \\ h_1(t_0, z_0) & h_1(t_0, z_0) \end{pmatrix} \quad (5.15)$$

PROOF. Define

$$Y_i = \begin{pmatrix} Y_{i;1} \\ Y_{i;2} \end{pmatrix} = n^{-3/5} \begin{pmatrix} k_{b_1}(t_0 - T_i) \\ \mathbb{1}_{(0, z_0]}(Z_i) k_{b_1}(t_0 - T_i) \end{pmatrix}.$$

By the assumptions on F_0 and g and condition (K.2), we have

$$\begin{aligned} \mathbb{E} Y_{i;1} &= n^{-3/5} g(t_0) + \frac{1}{2} n^{-1} c^2 m_2(k) g''(t_0) + o(n^{-1}), \\ \text{Var } Y_{i;1} &= n^{-1} c^{-1} g(t_0) \int k(u)^2 du + o(n^{-6/5}), \\ \mathbb{E} Y_{i;2} &= n^{-3/5} h_1(t_0, z_0) + \frac{1}{2} n^{-1} c^2 m_2(k) \partial_1^2 h_1(t_0, z_0) + o(n^{-1}), \\ \text{Var } Y_{i;2} &= n^{-1} c^{-1} h_1(t_0, z_0) \int k(u)^2 du + o(n^{-6/5}). \end{aligned}$$

Furthermore we have

$$\text{Cov}(Y_{i;1}, Y_{i;2}) = n^{-1} c^{-1} h_1(t_0, z_0) \int k(u)^2 du + o(n^{-6/5}),$$

so that

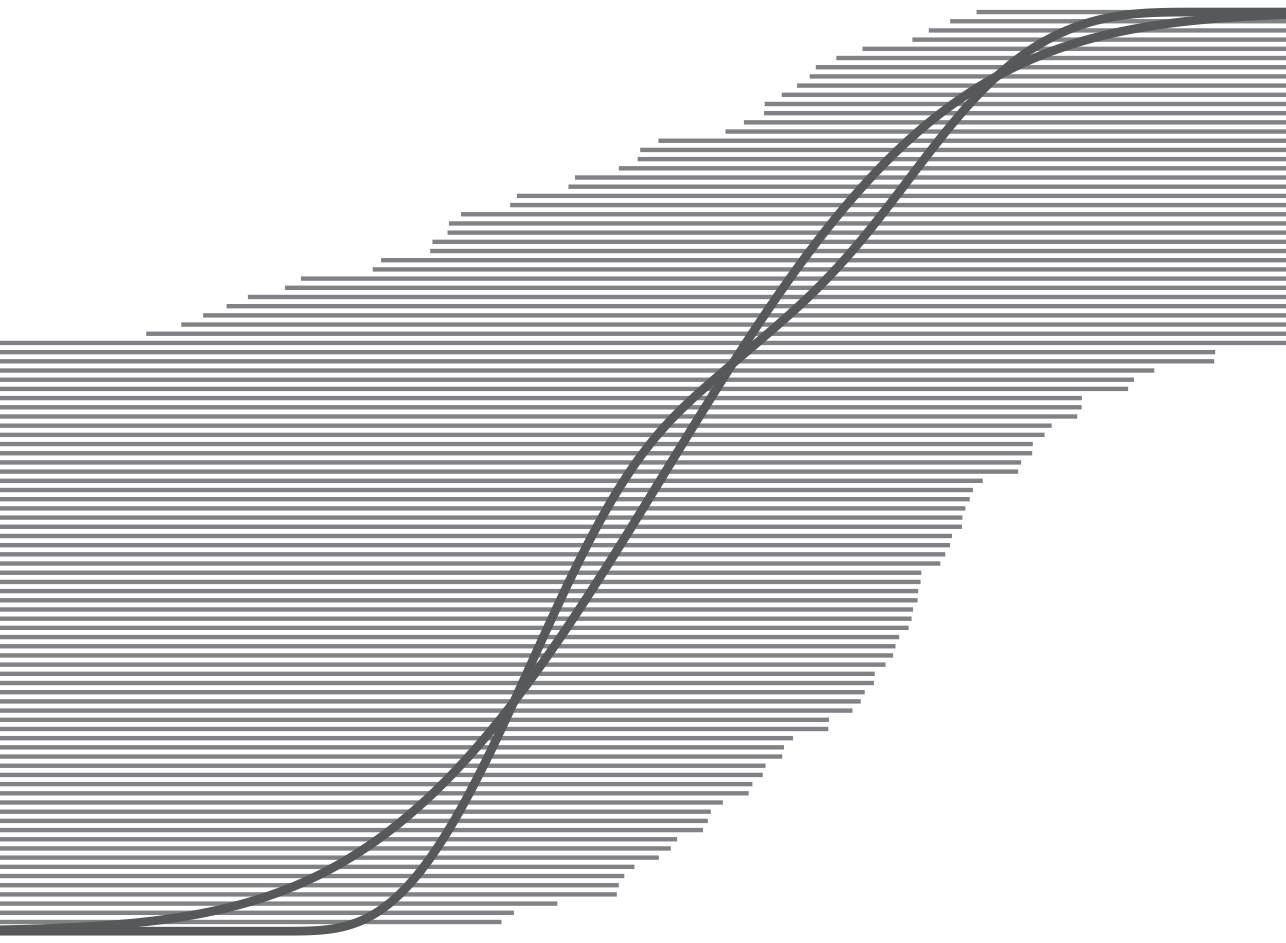
$$\begin{aligned} \sum_{i=1}^n \mathbb{E} Y_i &= n^{2/5} \begin{pmatrix} g(t_0) \\ h_1(t_0, z_0) \end{pmatrix} + \frac{1}{2} c^2 m_2(k) \begin{pmatrix} g''(t_0) \\ \partial_1^2 h_1(t_0, z_0) \end{pmatrix} + o(1) \\ \sum_{i=1}^n \text{Var } Y_i &= c^{-1} \int k(u)^2 du \begin{pmatrix} g(t_0) & h_1(t_0, z_0) \\ h_1(t_0, z_0) & h_1(t_0, z_0) \end{pmatrix} + o(n^{-1/5}). \end{aligned}$$

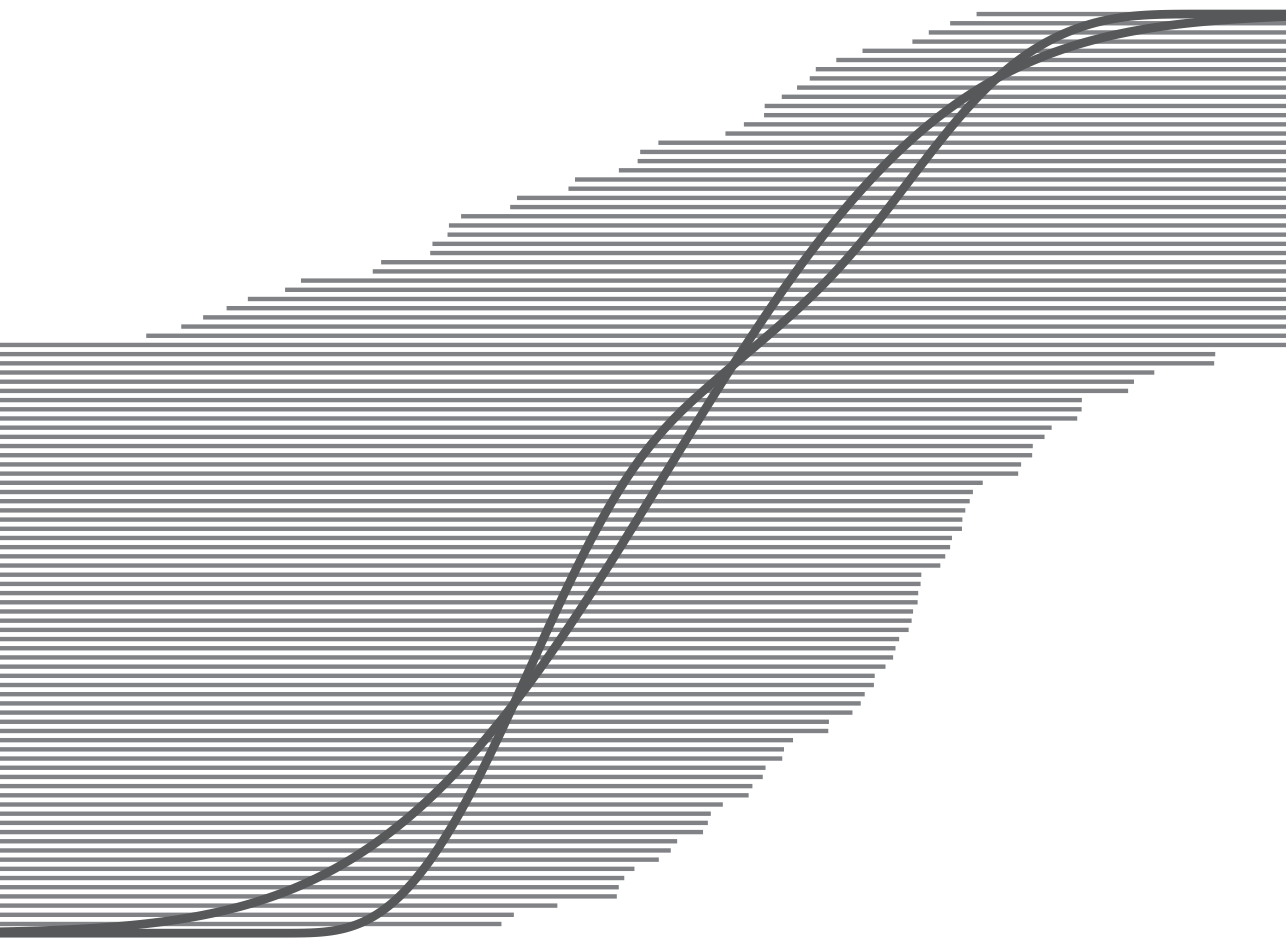
Here we denote by $\text{Var } Y_i$ the covariance matrix of the vector Y_i . By the Lindeberg–Feller central limit theorem we then get

$$\begin{aligned} n^{2/5} \left(\begin{pmatrix} \frac{1}{n} \sum_{i=1}^n k_{b_1}(t_0 - T_i) \\ \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{(0, z_0]}(Z_i) k_{b_1}(t_0 - T_i) \end{pmatrix} - \begin{pmatrix} g(t_0) \\ h_1(t_0, z_0) \end{pmatrix} \right) \\ - \frac{1}{2} c^2 m_2(k) \begin{pmatrix} g''(t_0) \\ \partial_1^2 h_1(t_0, z_0) \end{pmatrix} = \sum_{i=1}^n (Y_i - \mathbb{E} Y_i) + o(1) \rightsquigarrow \mathcal{N}(0, \Sigma_1). \end{aligned}$$

□

APPENDIX





In Chapters 2, 3 and 5 kernel estimators are used as smoothing method. This appendix contains several results on kernel estimators needed in these chapters. We start with recalling the definition of both the univariate and the multivariate kernel estimator.

Let k be a univariate kernel density satisfying condition (K.1) below.

(K.1) The probability density k has support $[-1, 1]$, is symmetric around 0 and continuously differentiable on $\mathbb{R}$.

Note that condition (K.1) implies that $m_2(k) = \int u^2 k(u) du < \infty$. Furthermore, let $b = b_n > 0$ be a bandwidth (or smoothing parameter) depending on the sample size n and satisfying $\lim_{n \rightarrow \infty} b_n = 0$. Denote by $k_b(u) = b^{-1}k(u/b)$ the scaled version of k . Now, we can define the univariate kernel estimator for the density f_0 of a sample $X_1, \dots, X_n$.

DEFINITION A.1. For fixed $x \in \mathbb{R}$, the univariate kernel density estimator for $f_0(x)$ is defined as

$$\hat{f}_n(x) = \frac{1}{n} \sum_{i=1}^n k_b(x - X_i).$$

Alternative ways to write the kernel estimator $\hat{f}_n$ are

$$\hat{f}_n(x) = \int k_b(x - u) d\mathbb{F}_n(u), \text{ or } \hat{f}_n(x) = (k_b * d\mathbb{F}_n)(x),$$

where $\mathbb{F}_n$ is the empirical distribution function of the sample $X_1, \dots, X_n$.

Under some assumptions on the density f_0 and the bandwidth b , the univariate kernel estimator $\hat{f}_n$ is consistent (pointwise and uniform), and converges pointwise to a normal distribution with rate $n^{2/5}$ [see e.g., PARZEN (1962) or SILVERMAN (1978)]. In Chapters 3 and 5 we mostly use the uniform consistency result, which is restated in Theorem A.2.

THEOREM A.2. Assume that f_0 is uniformly continuous on $\mathbb{R}$ and that $b_n \rightarrow 0$ and $nb_n \rightarrow \infty$, then

$$\|\hat{f}_n - f_0\|_\infty := \sup_{x \in \mathbb{R}} |\hat{f}_n(x) - f_0(x)| \xrightarrow{\mathcal{P}} 0.$$

This theorem was proven for example in PARZEN (1962, Theorem 3A) and SILVERMAN (1978, Theorem A).

In Chapters 3 and 5 we also need the uniform consistency of the kernel estimator for the derivative f'_0 of the density f_0 . The kernel estimator for f'_0 is defined as

$$\hat{f}'_n(x) = \frac{1}{n} \sum_{i=1}^n k'_b(x - X_i),$$

where $k'_b(u) = b^{-2}k'(u/b)$ denotes the scaled version of the derivative k' of the kernel density k . Theorem A.3 restates the uniform consistency of this estimator.

THEOREM A.3. *Assume that f'_0 is uniformly continuous on $\mathbb{R}$ and that $b_n \rightarrow 0$ and $nb_n^3 \rightarrow \infty$, then*

$$\|\hat{f}'_n - f'_0\|_\infty \xrightarrow{\mathcal{P}} 0.$$

This theorem follows from SILVERMAN (1978, Theorem C) with $r = 1$.

Similar results also hold for the multivariate kernel estimator, defined in Definition A.4 below. Let k be a multivariate kernel density satisfying condition (K.2) below.

(K.2) The multivariate probability density k has support $[-1, 1]^d$ and is continuous.

Let $b = (b_1, \dots, b_d) = (b_{1,n}, \dots, b_{d,n}) = b_n$ be a vector with bandwidths (or smoothing parameters) depending on the sample size n and satisfying $\lim_{n \rightarrow \infty} b_{i,n} = 0$ for $1 \leq i \leq d$. Denote by $k_b(u) = b_1^{-1} \cdots b_d^{-1} k(u_1/b_1, \dots, u_d/b_d)$ the scaled version of k . Now, we can define the multivariate kernel estimator for the density $f_0 : \mathbb{R}^d \rightarrow \mathbb{R}$ of a sample $X_1, \dots, X_n$.

DEFINITION A.4. *For fixed $x \in \mathbb{R}$, the multivariate kernel density estimator for $f_0(x)$ is defined as*

$$\hat{f}_n(x) = \frac{1}{n} \sum_{i=1}^n k_b(x - X_i),$$

where $(x - X_i) = (x_1 - X_{1,i}, \dots, x_d - X_{d,i})$.

In Chapter 5 we need the uniform consistency of the bivariate [i.e. $d = 2$] kernel estimator $\hat{f}_n$ as well as of $\partial_1 \hat{f}_n$, the partial derivative of $\hat{f}_n$ with respect to the first argument of $\hat{f}_n$. The latter estimator is defined as

$$\partial_1 \hat{f}_n(x) = \frac{1}{n} \sum_{i=1}^n \partial_1 k_b(x - X_i),$$

where $(x - X_i) = (x_1 - X_{1,i}, x_2 - X_{2,i})$ and $\partial_1 k_b(u) = b_1^{-2} b_2^{-1} k(u_1/b_1, u_2/b_2)$ denotes the scaled version of $\partial_1 k$. The uniform consistency of these estimators is stated in Theorems A.5 and A.6.

THEOREM A.5. *Assume that f_0 is uniformly continuous on $\mathbb{R}^2$ and that $nb_{1,n}^2 b_{2,n}^2 \rightarrow \infty$, then*

$$\|\hat{f}_n - f_0\|_\infty \xrightarrow{\mathcal{P}} 0.$$

This theorem follows from GACOULOS (1964, Theorem 3.3) with $p = 2$.

THEOREM A.6. *Assume that $\partial_1 f_0$ is uniformly continuous on $\mathbb{R}^2$ and that $nb_{1,n}^3 b_{2,n} \rightarrow \infty$, then*

$$\|\partial_1 \hat{f}_n - \partial_1 f_0\|_\infty \xrightarrow{\mathcal{P}} 0.$$

PROOF. By MOKKADEM *et al.* (2005, Theorem 3),

$$\lim_{n \rightarrow \infty} (nb_1^3 b_2)^{-1} \log P(\|\partial_1 \hat{f}_n - \partial_1 f_0\|_\infty \geq \delta) = -c,$$

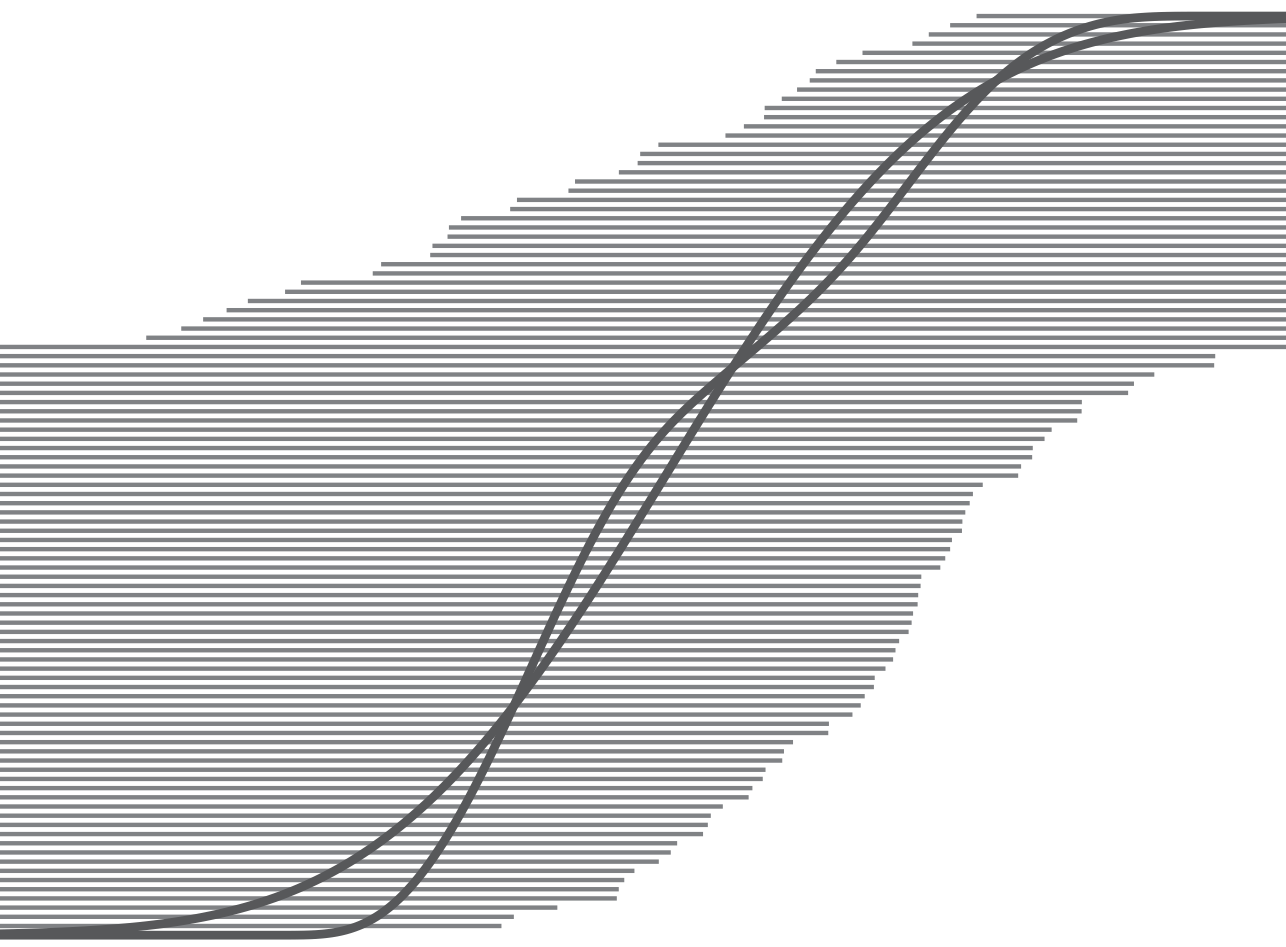
for some constant $c > 0$ only depending on δ and $\partial_1 f_0$. Hence, for n sufficiently large

$$P(\|\partial_1 \hat{f}_n - \partial_1 f_0\|_\infty \geq \delta) \leq 2e^{-nb_1^3 b_2^c} \rightarrow 0,$$

since $nb_{1,n}^3 b_{2,n} \rightarrow \infty$. □

APPENDIX B

NORMS, DISTANCES AND VARIOUS RELATIONS



Let f and g be two probability densities with respect to a dominating measure μ . The L_p -norm of f with respect to the measure μ is defined as

$$\|f\|_p = L_p(f) = \left(\int |f(x)|^p d\mu(x) \right)^{1/p},$$

for $0 < p < \infty$. For $p = \infty$, the L_∞ -norm (or sup-norm) is defined as

$$\|f\|_\infty = \sup_x |f(x)|.$$

The L_p - and sup-distance between f and g are defined as the norms of the difference $f - g$, i.e.

$$\|f - g\|_p = \left(\int |f(x) - g(x)|^p d\mu(x) \right)^{1/p}, \quad \|f - g\|_\infty = \sup_x |f(x) - g(x)|,$$

respectively. The Hellinger distance $\mathcal{H}$ between f and g is defined as

$$\mathcal{H}(f, g) = \sqrt{\frac{1}{2} \int \left(\sqrt{f(x)} - \sqrt{g(x)} \right)^2 d\mu(x)}, \quad (\text{B.1})$$

and the Kullback-Leibler divergence $\mathcal{K}$ of f with respect to g is defined as

$$\mathcal{K}(f, g) = \int f(x) \log \frac{f(x)}{g(x)} d\mu(x). \quad (\text{B.2})$$

LEMMA B.1. *The Hellinger distance $\mathcal{H}$ and Kullback-Leibler divergence $\mathcal{K}$ satisfy the following relation [see also e.g., VAN DE GEER (2000, Lemma 1.3)]*

$$2\mathcal{H}(f, g)^2 \leq \mathcal{K}(f, g). \quad (\text{B.3})$$

PROOF. Since $\frac{1}{2} \log(x) \leq \sqrt{x} - 1$, we have

$$\begin{aligned} \mathcal{K}(f, g) &= - \int f(x) \log \frac{g(x)}{f(x)} d\mu(x) \geq -2 \int f(x) \left(\frac{\sqrt{g(x)}}{\sqrt{f(x)}} - 1 \right) d\mu(x) \\ &= -2 \int \left(\sqrt{f(x)} \sqrt{g(x)} - f(x) \right) d\mu(x) = -2 \int \sqrt{f(x)} \sqrt{g(x)} d\mu(x) + 2 \int f(x) d\mu(x) \\ &= \int \left(f(x) - 2\sqrt{f(x)} \sqrt{g(x)} + g(x) \right) d\mu(x) \\ &= \int \left(\sqrt{f(x)} - \sqrt{g(x)} \right)^2 d\mu(x) = 2\mathcal{H}(f, g)^2. \end{aligned}$$

□

LEMMA B.2. *The Hellinger distance $\mathcal{H}$ and the L_1 -distance satisfy the following inequalities [see also e.g., LE CAM (1986, p. 47)]*

$$\mathcal{H}(f, g)^2 \leq \frac{1}{2} \|f - g\|_1 \leq \sqrt{2} \mathcal{H}(f, g). \quad (\text{B.4})$$

PROOF. To prove the first inequality, we use that $(\sqrt{x} - \sqrt{y})^2 \leq (\sqrt{x} + \sqrt{y})|\sqrt{x} - \sqrt{y}|$. Then

$$\begin{aligned} \mathcal{H}(f, g)^2 &= \int \left(\sqrt{f(x)} - \sqrt{g(x)} \right)^2 d\mu(x) \\ &\leq \frac{1}{2} \int \left(\sqrt{f(x)} + \sqrt{g(x)} \right) \left| \sqrt{f(x)} - \sqrt{g(x)} \right| d\mu(x) \\ &= \frac{1}{2} \int |f(x) - g(x)| d\mu(x) = \frac{1}{2} \|f - g\|_1, \end{aligned}$$

To prove the second inequality, we use Jensen's inequality and the Cauchy-Schwarz inequality. Then

$$\begin{aligned} \left(\frac{1}{2} \|f - g\|_1 \right)^2 &= \left(\frac{1}{2} \int |f(x) - g(x)| d\mu(x) \right)^2 \\ &\leq \frac{1}{2} \int \left(\sqrt{f(x)} - \sqrt{g(x)} \right)^2 d\mu(x) \cdot \frac{1}{2} \int \left(\sqrt{f(x)} + \sqrt{g(x)} \right)^2 d\mu(x) \\ &\leq \mathcal{H}(f, g)^2 \left(1 + \int \sqrt{f(x)g(x)} d\mu(x) \right) \leq 2\mathcal{H}(f, g)^2. \end{aligned}$$

□

If f and g have compact support $\mathcal{C}$ with finite measure $\mu(\mathcal{C}) = C < \infty$, then this lemma implies that

$$\mathcal{H}(f, g)^2 \leq \|f - g\|_1 = \int_{\mathcal{C}} |f(x) - g(x)| d\mu(x) \leq C \|f - g\|_{\infty}. \quad (\text{B.5})$$

BIBLIOGRAPHY

- AYER, M., H. D. BRUNK, G. M. EWING, W. T. REID AND E. SILVERMAN (1955). An empirical distribution function for sampling with incomplete information. *Annals of Mathematical Statistics*, **26**: 641–647.
- BARLOW, R., D. BARTHOLEMUEW, J. BREMMER AND H. BRUNK (1972). *Statistical Inference under Order Restrictions*. John Wiley & Sons, New York.
- BARRON, A. R., L. GYÖRFI AND E. C. VAN DER MEULEN (1992). Distribution estimation consistent in total variation and in two types of information divergence. *IEEE Transactions on Information Theory*, **38**: 1437–1454.
- BAZARAA, M. S., H. D. SHERALI AND C. M. SHETTY (2006). *Nonlinear programming: theory and algorithms*. 3rd edition, Wiley-Interscience, New Jersey.
- CACOULOS, T. (1964). Estimation of a multivariate density. *Annals of the Institute of Statistical Mathematics*, **18**: 179–189.
- CHERNOFF, H. (1964). Estimation of the mode. *Annals of the Institute of Statistical Mathematics*, **16**: 31–41.
- VAN EEDEN, C. (1956). Maximum likelihood estimation of ordered probabilities. *Proceedings Koninklijke Nederlandse Akademie voor Wetenschap A*, **59**: 444–455.
- EFRON, B. (1979). Bootstrap methods: another look at the jackknife. *Annals of Statistics*, **7**: 1–26.
- EGGERMONT, P. P. B. AND V. N. LARICCIA (2000). Maximum likelihood estimation of smooth monotone and unimodal densities. *Annals of Statistics*, **28**: 922–947.
- EGGERMONT, P. P. B. AND V. N. LARICCIA (2001). *Maximum Penalized Likelihood Estimation*. Springer-Verlag, New York.

- EL BARMÍ, H. AND H. MUKERJEE (2005). Inference under a stochastic ordering constraint: The k -sample problem. *Journal of the American Statistical Association*, **100**: 252–261.
- FLYNN, N. M., D. N. FORTHALL AND THE RGP120 HIV VACCINE STUDY GROUP (2005). Placebo-controlled phase 3 trial of a recombinant glycoprotein 120 vaccine to prevent HIV-1 infection. *Journal of Infectious Diseases*, **191**: 654–665.
- VAN DE GEER, S. A. (2000). *Empirical Processes in M-estimation*. Cambridge University Press, New York.
- GESKUS, R. B. AND P. GROENEBOOM (1997). Asymptotically optimal estimation of smooth functionals for interval censoring, part 2. *Statistica Neerlandica*, **51**: 201–219.
- GILBERT, P., S. SELE, M. RAO, A. NAFICY AND J. CLEMENS (2001). Sieve analysis: methods for assessing from vaccine trial data how vaccine efficacy varies with genotypic and phenotypic pathogen variation. *Journal of Clinical Epidemiology*, **54**: 68–85.
- GONZÁLEZ-MANTEIGA, W., R. CAO AND J. S. MARRON (1996). Bootstrap selection of the smoothing parameter in nonparametric hazard rate estimation. *Journal of the American Statistical Association*, **91**: 1130–1140.
- GRENANDER, U. (1956). On the theory of mortality measurement, part II. *Skandinavisk Aktuarietidskrift*, **39**: 125–153.
- GROENEBOOM, P. (1985). Estimating a monotone density. In: L. LeCam and R. A. Olshen (Eds.), *Berkeley Conf. in Honor of Jerzy Neyman and Jack Kiefer II*, 539–555, Wadsworth, Belmont, California.
- GROENEBOOM, P. AND G. JONGBLOED (2010). Generalized continuous isotonic regression. *Statistics & Probability Letters*, **80**: 248 – 253.
- GROENEBOOM, P., G. JONGBLOED AND B. I. WITTE (2010). Maximum smoothed likelihood estimation and smoothed maximum likelihood estimation in the current status model. *Annals of Statistics*, **38**: 352–387.
- GROENEBOOM, P., G. JONGBLOED AND B. I. WITTE (2011a). A maximum smoothed likelihood estimator in the current status continuous mark model, Submitted.
- GROENEBOOM, P., G. JONGBLOED AND B. I. WITTE (2011b). Smooth plug-in inverse estimators in the current status continuous mark model, Submitted.
- GROENEBOOM, P., M. H. MAATHUIS AND J. A. WELLNER (2008a). Current status data with competing risks: consistency and rates of convergence of the MLE. *Annals of Statistics*, **36**: 1031–1063.

- GROENEBOOM, P., M. H. MAATHUIS AND J. A. WELLNER (2008b). Current status data with competing risks: limiting distribution of the MLE. *Annals of Statistics*, **36**: 1064–1089.
- GROENEBOOM, P. AND J. A. WELLNER (1992). *Information Bounds and Nonparametric Maximum Likelihood Estimation*. Birkhäuser Verlag, Basel.
- GROENEBOOM, P. AND J. A. WELLNER (2001). Computing Chernoff's distribution. *Journal of Computational and Graphical Statistics*, **10**: 388–400.
- HÄRDLE, W., P. JANSSEN AND R. SERFLING (1988). Strong uniform consistency rates for estimators of conditional functionals. *Annals of Statistics*, **16**: 1428–1449.
- HAZELTON, M. L. (1996). Bandwidth selection for local density estimators. *Scandinavian Journal of Statistics*, **23**: 221–232.
- HUDGENS, M. G., M. H. MAATHUIS AND P. B. GILBERT (2007). Nonparametric estimation of the joint distribution of a survival time subject to interval censoring and a continuous mark variable. *Biometrics*, **63**: 372–380.
- JONES, M. C. (1993). Simple boundary correction for kernel density estimation. *Statistics and Computing*, **3**: 135–146.
- JONGBLOED, G. (1998). The iterative convex minorant algorithm for nonparametric estimation. *Journal of Computational and Graphical Statistics*, **7**: 310–321.
- KEIDING, N. (1991). Age-specific incidence and prevalence: a statistical perspective. *Journal of the Royal Statistical Society, Series A*, **154**: 371–412.
- LE CAM, L. M. (1986). *Asymptotic methods in statistical decision theory*. Springer-Verlag, New York.
- LECOUTRE, J.-P. (1985). The L_2 -optimal cell width for the histogram. *Statistics & Probability Letters*, **3**: 303 – 306.
- MAATHUIS, M. H. AND J. A. WELLNER (2008). Inconsistency of the MLE for the joint distribution of interval censored survival times and continuous marks. *Scandinavian Journal of Statistics*, **35**: 83–103.
- MAMMEN, E. (1991). Estimating a smooth monotone regression function. *Annals of Statistics*, **19**: 724–740.
- MARRON, J. S. AND W. J. PADGETT (1987). Asymptotically optimal bandwidth selection for kernel density estimators from randomly right-censored samples. *Annals of Statistics*, **15**: 1520–1535.
- MOKKADEM, A., M. PELLETIER AND J. WORMS (2005). Large and moderate deviations principles for kernel estimation of a multivariate density and its partial derivatives. *Australian & New Zealand Journal of Statistics*, **47**: 489–502.

- PARZEN, E. (1962). On estimation of a probability density function and mode. *Annals of Mathematical Statistics*, **33**: 1065–1076.
- PATIL, P. N., M. T. WELLS AND J. S. MARRON (1994). Some heuristics of kernel based estimators of ratio functions. *Journal of Nonparametric Statistics*, **4**: 203–209.
- POLITIS, N. P., J. P. ROMANO AND M. WOLF (1999). *Subsampling*. Springer-Verlag, New York.
- PRAKASA RAO, B. L. S. (1969). Estimation of a unimodal density. *Sankhyā: Indian Journal of Statistics, Series A*, **31**: 23–36.
- ROBERTSON, T., F. T. WRIGHT AND R. L. DYKSTRA (1988). *Order Restricted Statistical Inference*. John Wiley & Sons, New York.
- SCHUSTER, E. F. (1985). Incorporating support constraints into nonparametric estimators of densities. *Communications in Statistics - Theory and Methods*, **14**: 1123–1136.
- SHEATHER, S. J. (1983). A data-based algorithm for choosing the window width when estimating the density at a point. *Computational Statistics & Data Analysis*, **1**: 229–238.
- SILVERMAN, B. W. (1978). Weak and strong uniform consistency of the kernel estimate of a density and its derivative. *Annals of Statistics*, **6**: 177–184.
- SILVERMAN, B. W. (1986). *Density estimation for statistics and data analysis*. Chapman & Hall, London.
- SINGH, R. S. (1976). Nonparametric estimation of mixed partial derivatives of a multivariate density. *Journal of Multivariate Analysis*, **6**: 111–122.
- TANTIYASWADIKUL, C. AND M. B. WOODROOFE (1994). Isotonic smoothing splines under sequential designs. *Journal of Statistical Planning and Inference*, **38**: 75–88.
- VAN DER VAART, A. W. (1998). *Asymptotic Statistics*. Cambridge University Press, New York.
- VAN DER VAART, A. W. AND M. J. VAN DER LAAN (2003). Smooth estimation of a monotone density. *Statistics*, **37**: 189–203.
- VAN DER VAART, A. W. AND J. A. WELLNER (1996). *Weak Convergence and Empirical Processes*. Springer-Verlag, New York.
- WAND, M. P. AND M. C. JONES (1995). *Kernel Smoothing*. Chapman & Hall, London.

SUMMARY

CURRENT STATUS CENSORING MODELS

smooth estimators and their asymptotic properties

In survival analysis, the event time of interest can be censored in various ways. Subjects might drop out of study before the event occurs, which leads to right-censored observations. Or, the study might consist of several follow-ups and the event occurs unnoticed between two successive follow-ups, yielding interval censored observations.

In this thesis, we focus on one specific type of interval censoring: interval censoring case 1. In this case, there are no follow-up times, instead it is checked only once whether the event occurred already or not. In medical or engineering applications, the event usually denotes the onset of a certain disease or the failure of a system. With this type of interval censoring, only the current status of the patient (infected or not) or of the system (broken or not) is checked. Hence, this type of censoring is also referred to as *current status censoring*. We study several *smooth estimators* for the distribution function and the density of the event time. We derive their *asymptotic properties* and address some of their computational aspects.

The univariate *current status model* is treated in Part I. In this model, we are interested in the distribution function F_0 of an event time X . The Maximum Likelihood Estimator (MLE) for F_0 is consistent and can be characterized as the derivative of the greatest convex minorant of a specific cusum diagram. This estimator is piecewise constant between jump points, and, therefore, fails to have a density with respect to Lebesgue measure. Hence, other estimators are more natural if the true underlying distribution function is assumed to have a Lebesgue density.

In Chapter 2 we study the Smoothed Maximum Likelihood Estimator (SMLE) for the smooth distribution function F_0 and its Lebesgue density f_0 . This estimator is obtained by smoothing the MLE using kernel smoothing and is therefore smooth by definition. We prove that the SMLEs for F_0 and f_0 converge pointwise to a normal distribution at rate $n^{2/5}$ and $n^{2/7}$, respectively.

In Chapter 3 we consider a different but related estimator for F_0 : the Maximum Smoothed Likelihood Estimators (MSLEs). This estimator is obtained by interchanging the smoothing and maximization step in the previous estimators, i.e. maximize the smoothed log likelihood. This maximizer turns out to be smooth and its characterization can be seen as a continuous version of the characterization of the MLE. We prove that the MSLE for F_0 is asymptotically equivalent with a naive plug-in inverse estimator. Based on that result we derive the pointwise asymptotic distributions of the MSLEs for F_0 and f_0 , which are similar to the asymptotic distributions of the SMLEs for F_0 and f_0 . More specifically, the rate of convergence and the asymptotic variance of the MSLE and the SMLE for F_0 are equal, while the asymptotic bias of the MSLE for F_0 is equal to the asymptotic bias of the SMLE plus some additional (possibly negative) term depending on the unknown distribution F_0 . The same also holds for the MSLE and the SMLE for f_0 .

Part II covers the bivariate *current status continuous mark model*. In this model, F_0 denotes the joint distribution function of a univariate event time X and a continuous mark variable Y which might be correlated to this event time. The MLE for the bivariate distribution function is inconsistent.

In Chapter 4 we prove pointwise consistency of histogram based MSLEs for the bivariate distribution function F_0 and the marginal distribution $F_{0,X}$ of the event time X . Iterative algorithms are needed for computing the estimator, we describe an iterative quadratic minimization algorithm. Due to the implicit characterization it is hard to study the asymptotic distribution of the MSLEs for F_0 and $F_{0,X}$.

A more tractable estimator for F_0 is studied in Chapter 5, the smooth plug-in inverse estimator using again kernel smoothing as smoothing technique. This estimator is also pointwise consistent and converges at rate $n^{2/5}$ to a normal distribution.

CURRENT STATUS CENSURERINGSMODELLEN

gladde schatters en hun asymptotische eigenschappen

In de overlevingsduuranalyse kan het tijdstip waarop een gebeurtenis plaatsvindt op verschillende manieren gecensureerd zijn. Zo kan een deelnemer de studie verlaten voordat de gebeurtenis heeft plaatsgevonden, wat leidt tot een rechts-gecensureerde observatie. Het kan ook zijn dat de deelnemer tijdens de studie verschillende malen wordt gecontroleerd en dat de gebeurtenis ongemerkt plaatsvindt tussen twee van zulke controletijdstippen, resulterend in een interval gecensureerde observatie.

In dit proefschrift beschouwen we één specifiek type interval censurering: interval censurering type 1. In dit geval zijn er geen meerdere controlemomenten, maar wordt slechts éénmaal gecontroleerd of de gebeurtenis al heeft plaatsgevonden of nog niet. Deze gebeurtenis is vaak het begin van een ziekte of het falen van een systeem in medische of industriële toepassingen. Bij dit type censurering wordt dan alleen de huidige toestand (Engels: current status) van de patiënt (ziek of niet) of het systeem (werkend of niet) geobserveerd. Daarom wordt dit type censurering ook wel *current status censurering* genoemd. Wij bestuderen verschillende *gladde schatters* voor de verdelingsfunctie en de kansdichtheid van het tijdstip waarop de gebeurtenis plaatsvindt. We leiden *asymptotische eigenschappen* van deze schatters af en beschrijven enkele van hun computationele aspecten.

Het eendimensionale *current status model* wordt behandeld in Deel I. In dit model zijn we geïnteresseerd in de verdelingsfunctie F_0 van het tijdstip X waarop de gebeurtenis plaatsvindt. De Meest Aannemelijke Schatter (MLE) voor F_0 is consistent en kan gekarakteriseerd worden als de afgeleide van de grootste convexe minorant van een specifiek puntendiagram. Deze schatter is stuksgewijs constant tussen sprongpunten, en heeft daarom geen dichtheid ten opzichte van de Lebesgue maat. Andere schatters liggen meer voor de hand, als verondersteld wordt dat de werkelijke onderliggende verdelingsfunctie een Lebesgue dichtheid heeft.

In Hoofdstuk 2 bestuderen we de Smoothed Maximum Likelihood Estimator (SMLE) voor de gladde verdelingsfunctie F_0 en zijn Lebesgue dichtheid f_0 . Deze

gladde schatter wordt verkregen door de MLE glad te maken via kernel smoothing. We bewijzen dat de SMLEs voor F_0 en f_0 puntsgewijs in verdeling convergeren naar een normale verdeling met snelheid $n^{2/5}$ en $n^{2/7}$, respectievelijk.

In Hoofdstuk 3 analyseren we een andere schatter voor de gladde verdelingsfunctie F_0 en zijn Lebesgue dichtheid f_0 , de Maximum Smoothed Likelihood Estimator (MSLE). Deze schatter wordt verkregen door een gladgemaakte log likelihood te maximaliseren. De aldus verkregen schatter blijkt glad te zijn en zijn karakterisering kan gezien worden als een continue versie van de karakterisering van de MLE. We bewijzen dat de MSLE voor F_0 asymptotisch equivalent is met een naïeve plug-in schatter. We leiden de puntsgewijze asymptotische verdeling af van de MSLEs voor F_0 en f_0 , gebruikmakend van dat resultaat. Deze verdelingen zijn vergelijkbaar met die van de SMLEs. In het bijzonder zijn de convergentiesnelheid en de asymptotische variantie van de MSLE en de SMLE voor F_0 gelijk, terwijl de asymptotische onzuiverheid van de MSLE voor F_0 gelijk is aan die van de SMLE voor F_0 plus een extra (mogelijk negatieve) term die onder andere afhangt van de onbekende verdelingsfunctie F_0 . Hetzelfde geldt voor de MSLE en de SMLE voor f_0 .

Deel II behandelt het tweedimensionale *current status continuous mark model*. In dit model is F_0 de simultane verdelingsfunctie van het tijdstip X waarop de gebeurtenis optreedt en een continue mark variabele Y . De MLE voor de tweedimensionale verdelingsfunctie F_0 is inconsistent.

In Hoofdstuk 4 bewijzen we puntsgewijze consistentie van op histogramsschatters gebaseerde MSLEs voor de tweedimensionale verdelingsfunctie F_0 en de marginale verdelingsfunctie $F_{0,X}$ van X . Iteratieve algoritmen zijn nodig om de schatters uit te rekenen. Wij beschrijven een iteratief kwadratisch minimaliseringsalgoritme. De impliciete karakterisering van de schatters voor F_0 en $F_{0,X}$ maakt het moeilijk om hun asymptotische verdelingen te bestuderen.

Een meer expliciete schatter voor F_0 wordt in Hoofdstuk 5 bestudeerd. Het betreft een gladde plug-in inverse schatter waarbij wederom kernel smoothing wordt gebruikt om glad te maken. Deze schatter is puntsgewijs consistent en convergeert met snelheid $n^{2/5}$ naar een normale verdeling.

CURRICULUM VITAE



Birgit Witte was born on July 17, 1982 in Amsterdam. From 1994 to 2000 she attended high school at the Spinoza Lyceum Amsterdam, where she obtained her Gymnasium diploma in 2000. After that, she studied Mathematics at the Vrije Universiteit in Amsterdam from 2000 to 2005. During this time she was a student member of the Mathematics curriculum committee (2001-2005). Also, she worked as a student assistant in Mathematics, facilitating various exercise classes for Mathematics, Econometrics and Computer Science courses (2002-2005). In 2003 she wrote her Bachelor's thesis "Isotone regressie & monotone dichtheden" and in 2005 her Master's thesis "Estimating the incubation time distribution function of AIDS under stochastic ordering", both supervised by prof. dr. ir. G. Jongbloed. She obtained her M.Sc. degree in 2005 and immediately there after, she started her PhD research under the supervision of prof. dr. P. Groeneboom and prof. dr. ir. G. Jongbloed. The research was funded by the Dutch Organization for Scientific Research (NWO) and led to the present thesis. The first year and a half she was also working as a communication officer within the Faculty of Science at the Vrije Universiteit Amsterdam. She now works as a scientific researcher at the department of Epidemiology and Biostatistics of the VU University Medical Center in Amsterdam.