Smooth nonparametric estimation under monotonicity constraints

Musta, Eni

DOI
10.4233/uuid:35430f5f-daa8-49df-999f-bf97addd51ab

Publication date
2019

Document Version
Final published version

Citation (APA)
Musta, E. (2019). *Smooth nonparametric estimation under monotonicity constraints*. [Dissertation (TU Delft), Delft University of Technology]. https://doi.org/10.4233/uuid:35430f5f-daa8-49df-999f-bf97addd51ab

SMOOTH NONPARAMETRIC ESTIMATION UNDER MONOTONICITY CONSTRAINTS

ENI MUSTA

Eni Musta

Smooth nonparametric estimation under monotonicity constraints, Delft, 2019

Printed by: GVO drukkers & vormgevers B.V.

ISBN 978-94-6384-012-5

An electronic version of this dissertation is available at
http://repository.tudelft.nl/.

SMOOTH NONPARAMETRIC ESTIMATION UNDER MONOTONICITY CONSTRAINTS

Dissertation

for the purpose of obtaining the degree of doctor
at Delft University of Technology
by the authority of the Rector Magnificus,
prof. dr. ir. T.H.J.J. van der Hagen,
chair of the Board for Doctorates
to be defended publicly on
Wednesday 3, April 2019 at 10:00 o'clock

by

Eni MUSTA

Master of Science in Mathematics
Università di Pisa, Pisa, Italy,

born in Elbasan, Albania.

This dissertation has been approved by the promotors.

Composition of the doctoral committee:

Rector Magnificus,	chairperson
Prof. dr. ir. G. Jongbloed,	Delft University of Technology, promotor
Dr. H.P. Lopuhaä,	Delft University of Technology, promotor

Independent members:

Prof. C. Durot	University of Paris Nanterre, France
Prof. dr. I. Van Keilegom	Catholic University of Leuven, Belgium
Prof. dr. E.A. Cator	Radboud University
Prof. dr. B. de Pagter	Delft University of Technology
Dr. J.J. Cai	Delft University of Technology
Prof. dr. ir. K. I. Aardal	Delft University of Technology, reserve member

TU Delft

CONTENTS

Part I

INTRODUCTION

1 INTRODUCTION

This thesis focuses on asymptotic inference for smooth nonparametric estimators under monotonicity constraints. We start with an introductory chapter where we present fundamental facts about smooth and monotone estimation and essential notations that will be used throughout the thesis. Two aspects of asymptotic inference, namely, the pointwise and the global limit behavior of the estimators, will be treated in Part II and III respectively.

1.1 ESTIMATION UNDER MONOTONICITY CONSTRAINTS

1.1.1 *Motivation*

There are many situations in statistics in which the estimation of an unknown function is of interest. The most basic problem is recovering the underlying probability density that models the randomness of a given set of independent data points. Other examples include estimation of the regression relationship between a response variable and an explanatory variable or estimation of the failure rate in survival analysis. All these problems have been widely studied in the literature and a variety of methods have been proposed. Parametric methods are the first and the most commonly used due to their simplicity in computation, interpretation and prediction. However, assuming that the function of interest has a certain parametric form is usually a very strong and unrealistic assumption which, in case of model misspecification, can lead to incorrect inference. On the other hand, nonparametric methods require fewer or no assumptions on the functional form and as a result are more flexible and robust.

Yet, in many real life problems, one has some prior knowledge on the shape of the curve of interest and it is desirable to have estimators that are consistent with these practical expectations. Shape constrained methods impose certain qualitative assumptions on functional forms without needing to specify parametric models. Monotonicity, in particular, is a shape restriction that arises naturally in various applications. In survival analysis monotonicity constraints reflect the property of aging or becoming more reliable as the

survival time increases. For example, decreasing hazards are used to model survival times of patients after a successful medical treatment. The property of monotonicity plays an important role also when dealing with regression relationships. Indeed, it is often very reasonable to assume that increasing a factor X has a positive (negative) effect on a response Y. Among other examples, the econometric demand functions are monotonic in price while the biometric age-height charts should be monotonic in age over an appropriate range. In situations like these, incorporating monotonicity constraints in the estimation procedure leads to more accurate results and avoids obtaining implausible estimates. Often, monotonicity is also used because it allows for more straightforward inference since the estimators can be constructed without using tuning parameters.

1.1.2 *Isotonic estimation*

Nonparametric inference under shape constraints is currently a very active research area in statistics which initiated with estimation of a monotone real valued function. Two well-known criteria from parametric methods, maximum likelihood and least squares, are usually used also for nonparametric estimation of monotone curves. The first example can be found in Grenander, 1956 in the context of estimating a nonincreasing density f on $[0, \infty)$ on the basis of an i.i.d. sample $X_1, \ldots, X_n$ from f. We start illustrating the main ideas behind isotonic estimation through this simple density model.

The nonparametric maximum likelihood estimator of f is the maximizer $\hat{f}_n$ of the log-likelihood function

$$f \mapsto \sum_{i=1}^{n} \log f(X_i)$$

over all nonincreasing functions $f : [0, \infty) \to [0, \infty)$. In Grenander, 1956 it is shown that $\hat{f}_n$ can be characterized as the left-hand slope of the least concave majorant (LCM) $\widehat{F}_n$ of the empirical distribution function F_n. Hence, it is a piecewise constant function which can jump only at the points $X_1, \ldots, X_n$.

The concave majorant characterization leads to a fast computational algorithm and also plays a role in the asymptotic analysis. Besides, it initiated a more general strategy to produce monotone estimators. The idea is as following. The primitive of a nonincreasing function is concave. Hence, an estimator of the function of interest is the derivative of the least concave majorant of an estimator for its primitive function. Similarly, in case of estimating nondecreasing curves, we start with a naive estimator for the primitive of the curve of interest and then take the left-derivative of the greatest

convex minorant (GCM) of the naive estimator. Such estimators are called Grenander-type estimators. In particular, in the density model the empirical distribution function F_n is taken as a naive estimator for the cumulative distribution function F. The resulting Grenander-type estimator satisfies the maximum likelihood principle.

The Grenander-type procedure has been developed in a variety of other statistical models, e.g., regression (see Brunk, 1958), random censoring (see Huang and Wellner, 1995), or the Cox model (see Lopuhaä and Nane, 2013). In the next section we discuss some of them in more details.

1.1.3 *Examples*

The first two models we are going to consider come from survival analysis, known also as reliability theory, which studies the time until the occurrence of a certain event of interest. In survival analysis, subjects are followed during a period of time (duration of the study) and event times are registered for each of them. However, at the end of the follow-up some of the subjects will not have experienced the event. Hence, these observations will be censored, i.e., the exact event time is not known, only partial information is available. There are different type of censoring schemes but here we focus on right censoring which means that the subject might leave the study before the event occurs or the event occurs after the end of the study. Moreover, we assume that censoring times are independent of survival times.

While random variables are typically characterized by their probability density or distribution function, in survival analysis it is more natural to focus on the hazard function (failure rate). It is defined as the probability that an individual will experience an event within a small time interval given that the subject has survived until the beginning of this interval. Hence, a frequently encountered problem in this field is the estimation of the hazard rate. In this context, monotonicity constraints arise naturally, reflecting the property of aging or becoming more reliable as the survival time increases.

Example 1.1.1. RIGHT CENSORING MODEL Suppose we have an i.i.d. sample $X_1, \ldots, X_n$ with distribution function F and density f, representing the survival times. Let $C_1, \ldots, C_n$ be the i.i.d. censoring variables with a distribution function G and density g. Under the right random censorship model, we assume that the survival time X and the censoring time C are independent and the observed data consists of i.i.d. pairs of random variables $(T_1, \Delta_1), \ldots, (T_n, \Delta_n)$, where T denotes the follow-up time $T = \min(X, C)$ and $\Delta = \mathbb{1}_{\{X \leqslant C\}}$ is the censoring indicator.

The hazard rate λ is characterized by the following relation

$$\lambda(t) = \frac{f(t)}{1 - F(t)} \tag{1.1.1}$$

and we refer to the quantity

$$\Lambda(t) = \int_0^t \lambda(u)\, du = -\log\left[1 - F(t)\right], \tag{1.1.2}$$

as the cumulative hazard function. We aim at estimating λ, subject to the constraint that it is decreasing (the case of an increasing hazard is analogous), on the basis of n observations $(T_1, \Delta_1), \ldots, (T_n, \Delta_n)$.

The likelihood function is given by

$$\prod_{i=1}^{n} \left[f(T_i)\,(1 - G(T_i))\right]^{\Delta_i} \left[g(T_i)\,(1 - F(T_i))\right]^{1-\Delta_i}.$$

Hence, using (1.1.1) and (1.1.2), the nonparametric maximum likelihood estimator $\hat{\lambda}_n$ of λ is the maximizer of the pseudo-loglikelihood

$$l(\lambda) = \sum_{i=1}^{n} \left[\Delta_i \log \lambda(T_i) - \Lambda(T_i)\right]$$

over all decreasing functions $\lambda : [0, \tau_F) \to [0, \infty)$, where τ_F is the end point of the support of F. It is shown in Huang and Wellner, 1995 (see their Theorem 3.3) that $\hat{\lambda}_n$ is the left derivative of the least concave majorant of the cumulative sum diagram consisting of points

$$P_j = \left(W_n(T_{(j)}), V_n(T_{(j)})\right), \qquad j = 0, 1, \ldots, n,$$

where $P_0 = (0, 0)$, $\{T_{(j)}\}_{j=1,\ldots,n}$ is the ordered statistics of $\{T_j\}_{j=1,\ldots,n}$ and

$$W_n(T_{(j)}) = \frac{1}{n} \sum_{i=1}^{j} (n - i + 1) \left(T_{(i)} - T_{(i-1)}\right), \qquad V_n(T_{(j)}) = \frac{1}{n} \sum_{i=1}^{j} \Delta_{(i)}.$$

On the other hand, the Grenander-type estimator $\tilde{\lambda}_n$ of λ is defined as the left-hand slope of the least concave majorant $\widehat{\Lambda}_n$ of the Nelson-Aalen estimator Λ_n for the cumulative hazard function Λ, where

$$\Lambda_n(t) = \sum_{i=1}^{n} \frac{\mathbb{1}_{\{T_i \leqslant t\}} \Delta_i}{\sum_{j=1}^{n} \mathbb{1}_{\{T_j \geqslant T_i\}}}. \tag{1.1.3}$$

It turns out that, in this case the maximum likelihood estimator and the Grenander-type estimator are different. However, as noticed in Huang and

Wellner, 1995, the two estimators are very close for moderately large sample size and, almost always, they have the same jump points.

Now, we consider estimating the density function f assuming that it is nonincreasing. The nonparametric maximum likelihood estimator $\hat{f}_n$ is again characterized as the derivative of the least concave majorant of some cumulative sum diagram (see Corollary 3.1 in Huang and Wellner, 1995). On the other hand, to construct the Grenander-type estimator $\tilde{f}_n$ of f, we take as a naive estimator of the cumulative distribution function F the Kaplan-Meier estimator

$$F_n(t) = 1 - \prod_{i: X_{(i)} \leqslant t} \left(1 - \frac{d_i}{n_i}\right), \tag{1.1.4}$$

where $X_{(i)}$, $i = 1, \ldots, m$ are the ordered observed event times, d_i is the number of events at time $X_{(i)}$ and n_i is the number of individuals at risk prior to time $X_{(i)}$. Then, $\tilde{f}_n$ is defined as the left-hand slope of the least concave majorant $\widehat{F}_n$ of F_n. Once more, the two estimators are different but very close to each other.

Example 1.1.2. THE COX REGRESSION MODEL The semi-parametric Cox regression model is a very popular method in survival analysis that allows incorporation of covariates when studying lifetime distributions in the presence of right censored data. It was initially proposed in biostatistics (Cox, 1972) and quickly became broadly used to study, for example, the time to device failure in engineering, the effectiveness of a treatment in medicine, mortality in insurance problems, duration of unemployment in social sciences etc. The ease of interpretation, resulting from the formulation in terms of the hazard rate as well as the proportional effect of the covariates favor the wide use of this framework.

Let $X_1, \ldots, X_n$ be an i.i.d. sample representing the survival times of n individuals, which can be observed only on time intervals $[0, C_i]$ for some i.i.d. censoring times $C_1, \ldots, C_n$. One observes i.i.d. triplets $(T_1, \Delta_1, Z_1), \ldots, (T_n, \Delta_n, Z_n)$, where $T_i = \min(X_i, C_i)$ denotes the follow up time, $\Delta_i = \mathbb{1}_{\{X_i \leqslant C_i\}}$ is the censoring indicator and $Z_i \in \mathbb{R}^p$ is a time independent covariate vector. Given the covariate vector Z, the event time X and the censoring time C are assumed to be independent. Furthermore, conditionally on $Z = z$, the event time is assumed to be a nonnegative random variable with an absolutely continuous distribution function $F(x|z)$ and density $f(x|z)$. Similarly the censoring time is assumed to be a nonnegative random variable with an absolutely continuous distribution function $G(x|z)$ and density $g(x|z)$. The censoring mechanism is assumed to be non-informative, i.e., F and G share no parameters. Within the Cox model, the conditional hazard

rate $\lambda(x|z)$ for a subject with covariate vector $z \in \mathbb{R}^p$, is related to the corresponding covariate by

$$\lambda(t|z) = \lambda_0(t)\, e^{\beta_0' z}, \quad t \in \mathbb{R}^+,$$

where λ_0 represents the baseline hazard function, corresponding to a subject with $z = 0$, and $\beta_0 \in \mathbb{R}^p$ is the vector of the regression coefficients. We refer to the quantity

$$\Lambda_0(t) = \int_0^t \lambda_0(u)\, du,$$

as the cumulative baseline hazard. Then, for the conditional cumulative hazard $\Lambda(t|z) = -\log[1 - F(t|z)]$ we have $\Lambda(t|z) = \Lambda_0(t)\exp(\beta_0' z)$. It follows that the likelihood function is given by

$$\begin{aligned}
&\prod_{i=1}^n \{f(T_i|Z_i)\,[1 - G(T_i|Z_i)]\}^{\Delta_i}\,\{g(T_i|Z_i)\,[1 - F(T_i|Z_i)]\}^{1-\Delta_i} \\
&= \prod_{i=1}^n \lambda(T_i|Z_i)^{\Delta_i} \exp\{-\Lambda(T_i|Z_i)\} \prod_{i=1}^n g(T_i|Z_i)^{1-\Delta_i}\,[1 - G(T_i|Z_i)]^{\Delta_i}\,.
\end{aligned}$$

Hence, we essentially need to maximize the following pseudo log-likelihood function

$$l(\beta, \lambda_0) = \sum_{i=1}^n \left[\Delta_i \log \lambda_0(T_i) + \Delta_i \beta' Z_i - e^{\beta' Z_i} \Lambda_0(T_i)\right] \tag{1.1.5}$$

over $\beta \in \mathbb{R}$ and λ_0. Such a maximum does not exist if Λ_0 is only restricted to be absolutely continuous because we can always choose some function λ_0 with fixed values at the T_i while letting $\lambda_0(T_i)$ go to infinity for some T_i with $\Delta_i = 0$.

However, if one is interested only on the effect of the covariates on survival, the proportional hazard property of the Cox model allows estimation of β_0 while leaving the baseline hazard completely unspecified. Indeed, by now it seems to be rather a standard choice estimating β_0 by $\hat{\beta}_n$, the maximizer of the partial likelihood function

$$\beta \mapsto \prod_{i=1}^m \frac{e^{\beta' Z_{(i)}}}{\sum_{j=1}^n \mathbb{1}_{\{T_j \geqslant X_{(i)}\}} e^{\beta' Z_j}}, \tag{1.1.6}$$

as proposed in Cox, 1972. Here $X_{(1)}, \ldots, X_{(m)}$ denote the ordered observed event times. Note that each factor of the partial likelihood function can be seen as the ratio between the hazard at time $X_{(i)}$ of the individual that failed at that time and the total hazard of all the individuals that were alive

at time $X_{(i)}$. The event times are actually not being used, only their ranking is important.

The partial likelihood is a profile likelihood if the parameter space is restricted to

$\{(\beta, \lambda_0) : \beta \in \mathbb{R}$ and Λ_0 is increasing step function with jumps only at T_i's$\}$

Let $\lambda_0(t)$ be the jumpsize at time t. The modified loglikelihood is of the form

$$l(\beta, \lambda_0) = \sum_{i=1}^{n} \left\{ \Delta_i \log \lambda_0(T_i) + \Delta_i \beta' Z_i - e^{\beta' Z_i} \sum_{j=1}^{n} \lambda_0(T_j) \mathbb{1}_{\{T_j \leqslant T_i\}} \right\}.$$

and the maximum of $l(\beta, \lambda_0)$ w.r.t. $\lambda_0(T_i)$ is obtained at

$$\hat{\lambda}_n(T_i) = \frac{\Delta_i}{\sum_{j=1}^{n} e^{\beta' Z_j} \mathbb{1}_{\{T_j \geqslant T_i\}}}.$$

Substituting the above values for $\{\lambda_0(T_i)\}_{i=1,\ldots,n}$ into the expression of $l(\beta, \lambda_0)$, we obtain the profile log-likelihood function for β

$$l(\beta) = \sum_{i=1}^{m} \left\{ \beta' Z_{(i)} - \log \sum_{j=1}^{n} e^{\beta' Z_j} \mathbb{1}_{\{T_j \geqslant X_{(i)}\}} \right\}$$

which is exactly the log-partial likelihood function.

On the other hand, when one is interested for instance in the absolute time to event, estimation of λ_0 is required. A piecewise constant estimator was suggested in Breslow and Crowley, 1974, which results in

$$\Lambda_n(t) = \sum_{i: X_{(i)} \leqslant t} \frac{d_i}{\sum_{j=1}^{n} e^{\hat{\beta}_n' Z_j} \mathbb{1}_{\{T_j \geqslant X_{(i)}\}}}, \tag{1.1.7}$$

where $\hat{\beta}_n$ is the maximum partial likelihood estimator of β and d_i is the number of events at time $X_{(i)}$. In particular, in the case of no covariates, the Breslow estimator reduces to the Nelson-Aalen estimator in (1.1.3).

Although the most attractive property of this approach is that it does not assume any fixed shape on the hazard curve, there are several cases where order restrictions better match the practical expectations (e.g., see van Geloven et al., 2013 for an example of a decreasing hazard in a large clinical trial for patients with acute coronary syndrome). Estimation of the baseline hazard function under monotonicity constraints has been studied in Chung and Chang, 1994 and Lopuhaä and Nane, 2013.

Assume we want to estimate λ_0, subject to the constraint that it is increasing (the case of a decreasing hazard is analogous). For a fixed β, the constrained nonparametric maximum likelihood estimator is of the form

$$\hat{\lambda}_n(x;\beta) = \begin{cases} 0 & x < T_{(1)} \\ \hat{\lambda}_i & T_{(i)} \leqslant x < T_{(i+1)} \\ \infty & x \geqslant T_{(n)}, \end{cases}$$

where $\hat{\lambda}_i$ is the left derivative of the greatest convex minorant at the points P_i of the cumulative sum diagram consisting of points $P_0 = (0,0)$,

$$P_j = \left(W_n(T_{(j+1)};\beta), V_n(T_{(j+1)})\right), \qquad j = 1, 2 \ldots, n-1,$$

for

$$W_n(x;\beta) = \int \left(e^{\beta' z} \int_{T_{(1)}}^{x} \mathbb{1}_{u \geqslant s} \, ds \right) d\mathbb{P}_n(u,\delta,z), \quad x \geqslant T_{(1)},$$

$$V_n(x;\beta) = \int \delta \mathbb{1}_{\{u \leqslant x\}} \, d\mathbb{P}_n(u,\delta,z).$$

Then, Lopuhaä and Nane, 2013 propose $\hat{\lambda}_n(x) = \hat{\lambda}_n(x, \hat{\beta}_n)$ as an estimator of λ_0, where $\hat{\beta}_n$ is an estimator of β_0. The standard choice for $\hat{\beta}_n$ is the maximum partial likelihood estimator.

On the other hand, taking the Breslow estimator as a naive estimator of the cumulative baseline hazard, the Grenander estimator $\tilde{\lambda}_n$ of λ_0 is the left slope of the greatest convex minorant $\widehat{\Lambda}_n$ of Λ_n.

1.1.4 *Distributional results*

The isotonic estimators discussed in the previous sections are consistent in the interior of the support if the function we are estimating is continuous. Inconsistency at the boundaries or at discontinuity points has been shown in Anevski and Hössjer, 2002; Balabdaoui et al., 2011; Woodroofe and Sun, 1993. On the other hand, in the interior of the support, the pointwise asymptotic behavior of isotonic estimators is typically characterized by a cube-root n rate of convergence instead of the more common $\sqrt{n}$-rate. The reason behind this is explained in Kim and Pollard, 1990 for argmax type of estimators. Again we illustrate the main ideas for the monotone density model. Isotonic estimators are connected to argmax estimators through the inverse process. The inverse process related to the Grenander-type estimator of a nonincreasing density is given by

$$\widehat{U}_n(a) = \operatorname*{argmax}_{s \geqslant 0} \{F_n(s) - as\}, \tag{1.1.8}$$

It is a sort of inverse of $\hat{f}_n$ in the sense that it satisfies the switching relation

$$\hat{f}_n(t) \geqslant a \Longleftrightarrow \widehat{U}_n(a) \geqslant t, \qquad a \in \mathbb{R}, \quad t > 0. \tag{1.1.9}$$

Since $t_n = \widehat{U}_n(a)$ maximizes $\Gamma_n(t) = F_n(t) - at$ we have

$$\Gamma_n(t_n) - \Gamma_n(t_0) \geqslant 0, \tag{1.1.10}$$

where $t_0 = f^{-1}(a)$ is the maximizer of the deterministic version $\Gamma(t) = F(t) - at$. We can write

$$\begin{aligned}\Gamma_n(t) - \Gamma_n(t_0) &= \{(\Gamma_n - \Gamma)(t) - (\Gamma_n - \Gamma)(t_0)\} + \{\Gamma(t) - \Gamma(t_0)\} \\ &= I + II.\end{aligned}$$

By a Taylor expansion it follows that $II \sim -\frac{1}{2}|f'(t_0)|(t-t_0)^2$ while the first term is normally distributed with mean zero and variance

$$\frac{1}{n}\left(F(t) - F(t_0)\right)\left(1 - (F(t) - F(t_0))\right) \approx \frac{1}{n} f(t_0)|t - t_0|.$$

Hence $I = O_P\left(n^{-1/2}\sqrt{|t-t_0|}\right)$. In order to have (1.1.10) we need that I and II have the same order, i.e.,

$$n^{-1/2}\sqrt{|t-t_0|} \sim (t-t_0)^2,$$

which means that $\widehat{U}_n$ converges at rate $n^{1/3}$. By the switching relation, it can be shown that also $\hat{f}_n$ converges at the same rate.

The pointwise asymptotic distribution of these type of estimators was first obtained in Prakasa Rao, 1969, 1970 and reproved in Groeneboom, 1983, who introduced a more accessible approach based on inverses. For every $t_0 > 0$ such that $f'(t_0) < 0$ we have

$$n^{1/3}\left(\frac{1}{4f(t_0)|f'(t_0)|}\right)^{1/3}\left\{\hat{f}_n(t_0) - f(t_0)\right\} \xrightarrow{d} \operatorname*{argmax}_{t\in\mathbb{R}}\left\{W(t) - t^2\right\}, \tag{1.1.11}$$

where W_t denotes a two-sided Brownian motion. The limit distribution is known as the Chernoff distribution (Chernoff, 1964). It is the distribution of the value at zero of the process

$$X(a) = \operatorname*{argmax}_{t\in\mathbb{R}}\left\{W(t) - (t-a)^2\right\}, \tag{1.1.12}$$

which was introduced and investigated by Groeneboom, 1983; Groeneboom, 1989 and plays a key role in the asymptotic behavior of isotonic estimators.

The main steps for proving (1.1.11) can be found, for example, in Durot and Lopuhaä, 2018. Apart from the direct approach, using (1.1.9), the asymptotic distribution of $\hat{f}_n$ can also be obtained through the more tractable process $\widehat{U}_n$.

By defining the inverse process appropriately, this approach can be used to derive the limit distribution of isotonic estimators even in other models, given that we have a convex minorant (concave majorant) characterization as in the previous examples. It initiated a stream of research on isotonic estimators, e.g., see Huang and Zhang, 1994 for density estimation with censored observations, Huang and Wellner, 1995 for estimation of a monotone hazard in the right censoring model, Lopuhaä and Nane, 2013 for estimation of the baseline hazard in the Cox model. A more general theory on consistency and limit distribution of estimators that are constructed as left derivatives of the least concave majorant (greatest convex minorant) of a cumulative sum diagram can be found in Anevski and Hössjer, 2006; Westling and Carone, 2018.

1.2 SMOOTH ISOTONIC ESTIMATION

Traditional isotonic estimators, such as maximum likelihood estimators and Grenander-type estimators are step functions which, in case of heavily censored data and small or moderate sample sizes, tend to have only a few jumps of large size. In such situations smooth estimators would be preferred to piecewise constant ones because they have a nice graphical representation and are more accurate. Indeed, a long stream of research has shown that, at the price of additional smoothness assumptions on the function of interest, smooth estimators achieve a faster rate of convergence (with respect to the cube-root n rate) to a Gaussian distributional law. Smooth estimation has received considerable attention in the literature, also because it is needed to prove that a bootstrap method works (see for instance, Kosorok, 2008; Sen, Banerjee, and Woodroofe, 2010). Moreover, it provides a straightforward estimate of the derivative of the function of interest, which is of help when constructing confidence intervals (see for instance, Nane, 2015).

Various approaches can be used to obtain smooth shape constrained estimators. Typically, these estimators are constructed by combining an isotonization step with a smoothing step. It essentially depends on the methods of both isotonization and smoothing and on the order of operations. Estimators constructed by smoothing followed by an isotonization step have been considered in Cheng and Lin, 1981, Wright, 1982, Friedman and Tibshirani, 1984, and Ramsay, 1998, for the regression setting, in van der Vaart and

van der Laan, 2003 for estimating a monotone density, and in Eggermont and LaRiccia, 2000, who consider maximum smoothed likelihood estimators for monotone densities. Methods that interchange the smoothing step and the isotonization step, can be found in Mukerjee, 1988, Durot, Groeneboom, and Lopuhaä, 2013. Comparisons between isotonized smooth estimators and smoothed isotonic estimators are made in Mammen, 1991 for the regression setting, in Groeneboom, Jongbloed, and Witte, 2010 for the current status model and in Groeneboom and Jongbloed, 2013 for estimating a monotone hazard rate. Other references for combining shape constraints and smoothness can be found in Chapter 8 in Groeneboom and Jongbloed, 2014.

Smooth estimation under monotonicity constraints for the baseline hazard in the Cox model was introduced in Nane, 2013. By combining an isotonization step with a smoothing step and alternating the order of smoothing and isotonization, four different estimators can be constructed. Two of them are kernel smoothed versions of the maximum likelihood estimator and the Grenander-type estimator from Lopuhaä and Nane, 2013. The third estimator is a maximum smoothed likelihood estimator obtained by first smoothing the loglikelihood of the Cox model and then finding the maximizer of the smoothed likelihood among all decreasing baseline hazards. The forth one is a Grenander-type estimator based on the smooth Breslow estimator for the cumulative hazard.

1.2.1 *Characterization of smooth isotonic estimators*

Smoothing approaches include kernel, spline and penalized likelihood methods. Here we consider kernel smoothing which is probably the most popular method due to its simplicity and intuitive nature.

Let k be an m-orthogonal kernel function for some $m \geqslant 1$, which means that

$$\int |k(u)||u|^m \, du < \infty \quad \text{and} \quad \int k(u)u^j \, du = 0,$$

for $j = 1, \ldots, m-1$, if $m \geqslant 2$. We assume that

$$\begin{aligned} &k \text{ has bounded support } [-1,1] \text{ and is such that } \int_{-1}^{1} k(y)\, dy = 1; \\ &k \text{ is differentiable with a uniformly bounded derivative.} \end{aligned} \tag{1.2.1}$$

We denote by k_b its scaled version $k_b(u) = b^{-1}k(u/b)$. Here $b = b_n$ is a bandwidth that depends on the sample size, in such a way that $0 < b_n \to 0$ and $nb_n \to \infty$, as $n \to \infty$. From now on, we will simply write b instead

of b_n. Note that if $m > 2$, the kernel function k necessarily attains negative values and as a result also the kernel smoothed estimators may be negative and monotonicity might not be preserved. To avoid this, one could restrict oneself to $m = 2$. In that case, the most common choice is to let k be a symmetric probability density.

Consider estimating a function $\lambda : [0,1] \to \mathbb{R}$ subject to the constraint that it is non-increasing. Suppose that on the basis of n observations we have at hand a cadlag step estimator Λ_n of

$$\Lambda(t) = \int_0^t \lambda(u)\,du, \quad t \in [0,1].$$

Then, at a point $t \in [0,1]$, the standard kernel estimator of λ is given by

$$\tilde{\lambda}_n^s(t) = \int_{(t-b)\vee 0}^{(t+b)\wedge 1} k_b(t-u)\,d\Lambda_n(u). \tag{1.2.2}$$

This estimator is smooth but not necessarily monotone. On the other hand, we can construct, an isotonic estimator $\hat{\lambda}_n$ which is piecewise constant for example by a Grenander-type procedure. In order to obtain an estimator that is smooth and monotone at the same time we can smooth $\hat{\lambda}_n$ or isotonize $\tilde{\lambda}_n^s$. The first procedure produces the smoothed isotonic estimator

$$\tilde{\lambda}_n^{SI}(t) = \int_{(t-b)\vee 0}^{(t+b)\wedge 1} k_b(t-u)\hat{\lambda}_n(u)\,du = \int_{(t-b)\vee 0}^{(t+b)\wedge 1} k_b(t-u)\,d\widehat{\Lambda}_n(u), \tag{1.2.3}$$

where $\widehat{\Lambda}_n$ is the least concave majorant of Λ_n. Note that, if k is a symmetric probability density, smoothing preserves monotonicity on $(b, 1-b)$. Indeed, for a decreasing $\hat{\lambda}_n$ and $b < t < s < 1-b$, by a change of variable we have

$$\tilde{\lambda}_n^{SI}(s) - \tilde{\lambda}_n^{SI}(t) = \int_{-1}^{1} k(y)\left[\hat{\lambda}_n(s-by) - \hat{\lambda}_n(t-by)\right]\,dy \leqslant 0.$$

On the other hand, the isotonized kernel estimator is constructed as follows. First we smooth the piecewise constant estimator Λ_n by means of a kernel function,

$$\Lambda_n^s(t) = \int k_b(t-u)\Lambda_n(u)\,du, \quad \text{for } t \in [0,1].$$

Next, we define a continuous monotone estimator $\tilde{\lambda}_n^{IS}$ of λ as the left-hand slope of the least concave majorant $\widehat{\Lambda_n^s}$ of Λ_n^s on $[0,1]$. In this way, we define a sort of Grenander-type estimator based on a smoothed naive estimator for Λ. We use the superscript IS to indicate that smoothing is performed first, followed by isotonization. Note that $\tilde{\lambda}_n^{IS}$ is usually less smooth than $\tilde{\lambda}_n^{SI}$ since only continuity is guaranteed.

For both methods, in practice one also has to choose the kernel function and the bandwidth parameter. The choice of the kernel does not effect much the resulting estimates. The standard choice is a symmetric probability density function. On the contrary, the choice of the bandwidth is crucial. In particular, in situations when the data are not equally distributed, the fixed bandwidth kernel tends to oversmooth in dense regions and undersmooth in sparse ones. Therefore, data-dependent rules for determination of the bandwidth would be more appropriate for achieving a uniform degree of smoothing (see for example Hess, Serachitopol, and Brown, 1999, Müller and Wang, 1990). However, for simplicity reasons, usually a global bandwidth is used. Determining rigorous rules for optimal bandwidth selection is a difficult task. Through the asymptotic analysis of the estimators one can determine the optimal order of the bandwidth which regulates the trade-off between the bias and the variance but this does not solve the problem for finite sample sizes. Various methods of bandwidth selection have been proposed in the literature such as cross-validation, plug-in techniques, bootstrap etc. However, selecting the right amount of smoothness remains problematic.

1.2.2 *Distributional results*

Distribution theory of smooth isotonic estimators was first studied by Mukerjee, 1988, who established asymptotic normality for a kernel smoothed least squares regression estimator, but this result is limited to a rectangular kernel and the rate of convergence is slower than the usual rate for kernel estimators. Later on, a smoothed isotonic estimator and an isotonized smooth estimator are considered in Mammen, 1991 and their asymptotic equivalence is derived. In van der Vaart and van der Laan, 2003 it is shown that the isotonized kernel density estimator has the same limit normal distribution at the usual rate $n^{m/(2m+1)}$ as the ordinary kernel density estimator, when the density is m times continuously differentiable. Similar results were obtained by Groeneboom, Jongbloed, and Witte, 2010 for the smoothed maximum likelihood estimator and the maximum smoothed likelihood estimator, and by Groeneboom and Jongbloed, 2013 for a smoothed Grenander-type estimator. This long stream of research shows that the asymptotic behavior of smooth isotonic estimators resembles the one of the standard kernel estimator (in terms of rate of convergence and limit distribution). Next we illustrate why this is indeed the case.

We consider first the standard kernel estimator $\tilde{\lambda}_n^s$ in (1.2.2). Let $t \in (0,1)$. For n large enough, we have $t \in (b, 1-b)$. Thus, for now, we don't have to

worry about being in the boundary region. For simplicity, we also assume $m = 2$, i.e., the function λ is two times continuously differentiable and we use a 2-orthogonal kernel function. In this case, the optimal bandwidth is of order $n^{-1/5}$ and the kernel estimator converges at rate $n^{2/5}$ to a Gaussian distribution. Indeed, we have

$$\begin{aligned} n^{2/5}\left\{\tilde{\lambda}_n^s(t) - \lambda(t)\right\} = n^{2/5}\left\{\int k_b(t-u)\,d\Lambda(u) - \lambda(t)\right\} \\ + n^{2/5}\int k_b(t-u)\,d(\Lambda_n - \Lambda)(u). \end{aligned} \tag{1.2.4}$$

The first (deterministic) term on the right hand side of (1.2.4) gives us the asymptotic bias by using a change of variables, a Taylor expansion, and the properties of the kernel:

$$\begin{aligned} & n^{2/5}\left\{\int_{t-b}^{t+b} k_b(t-u)\,\lambda(u)\,du - \lambda(t)\right\} \\ & = n^{2/5}\int_{-1}^{1} k(y)\,\{\lambda(t-by) - \lambda(t)\}\,dy \\ & = n^{2/5}\int_{-1}^{1} k(y)\left\{-\lambda'(t)by + \frac{1}{2}\lambda''(\xi_n)b^2y^2\right\}dy, \\ & \to \frac{1}{2}c^2\,\lambda''(t)\int_{-1}^{1} y^2\,k(y)\,dy. \end{aligned} \tag{1.2.5}$$

where $|\xi_n - t| < b|y| \leqslant b \to 0$ and $n^{1/5}b \to c > 0$. In general, if the function λ was m-times continuously differentiable and the kernel was m-orthogonal then the rate of convergence would be $n^{m/(2m+1)}$ and the optimal bandwidth of order $n^{-1/(2m+1)}$.

On the other hand, for proving the asymptotic normality of the second term in (1.2.4), the idea is to write it in terms of the empirical process

$$n^{2/5}\int k_b(t-u)\,d(\Lambda_n - \Lambda)(u) = \int g_{t,n}(u)\,d\sqrt{n}(\mathbb{P}_n - \mathbb{P})(u)$$

for some function $g_{t,n}$, and then use results from empirical process theory to show that it converges to a normal distribution with mean zero and certain variance.

Now we want to show that the smoothed isotonic estimator $\tilde{\lambda}_n^{SI}$ defined in (1.2.3) has the same asymptotic behavior as the kernel estimator. By definition we can write

$$\tilde{\lambda}_n^{SI}(t) - \tilde{\lambda}_n^s(t) = \int_{t-b}^{t+b} k_b(t-u)\,d\left(\widehat{\Lambda}_n - \Lambda_n\right)(u).$$

Then, integration by parts and a change of variable yield

$$\begin{aligned}\tilde{\lambda}_n^{SI}(t) - \tilde{\lambda}_n^{s}(t) &= \frac{1}{b^2}\int_{t-b}^{t+b}\left\{\widehat{\Lambda}_n(u) - \Lambda_n(u)\right\} k'\left(\frac{t-u}{b}\right) du \\ &= \frac{1}{b}\int_{-1}^{1}\left\{\widehat{\Lambda}_n(t-by) - \Lambda_n(t-by)\right\} k'(y)\, dy.\end{aligned}$$

For the density model it was shown in Kiefer and Wolfowitz, 1976 that $\sup_t |\widehat{F}_n(t) - F_n(t)| = O_P\left(n^{-2/3}(\log n)^{2/3}\right)$. The result was later on extended to a more general setting which includes various statistical models by Durot and Lopuhaä, 2014. For such models, using

$$\sup_{t\in[0,1]} |\widehat{\Lambda}_n(t) - \Lambda_n(t)| = O_P\left(n^{-2/3}(\log n)^{2/3}\right), \tag{1.2.6}$$

we obtain

$$\tilde{\lambda}_n^{SI}(t) - \tilde{\lambda}_n^{s}(t) = O_P\left(b^{-1}n^{-2/3}(\log n)^{2/3}\right).$$

Hence

$$n^{2/5}\left\{\tilde{\lambda}_n^{SI}(t) - \tilde{\lambda}_n^{s}(t)\right\} \xrightarrow{\mathbb{P}} 0,$$

which means that the two estimators are asymptotically equivalent.

The approach for dealing with the isotonized kernel estimator $\tilde{\lambda}_n^{IS}$ is different. It follows from Lemma 1 in Groeneboom and Jongbloed, 2010 (in the case of a decreasing function), that $\tilde{\lambda}_n^{IS}$ is the unique minimizer of

$$\psi(\lambda) = \frac{1}{2}\int_0^1 \left(\lambda(t) - \tilde{\lambda}_n^{naive}(t)\right)^2 dt$$

over all nonincreasing functions λ, where $\tilde{\lambda}_n^{naive}(t) = d\Lambda_n^s(t)/dt$. Note that, for $t \in [b, 1-b]$, from integration by parts we get

$$\tilde{\lambda}_n^{naive}(t) = \frac{1}{b^2}\int_{t-b}^{t+b} k'\left(\frac{t-u}{b}\right)\Lambda_n(u)\, du = \int_{t-b}^{t+b} k_b(t-u)\, d\Lambda_n(u), \tag{1.2.7}$$

i.e., $\tilde{\lambda}_n^{naive}$ coincides with the standard kernel estimator of λ on the interval $[b, 1-b]$. Then, the idea is to show that for large n, the kernel estimator is monotone with large probability and as a result, for every $0 < \epsilon < M < 1$,

$$\mathbb{P}(\tilde{\lambda}_n^{s}(t) = \tilde{\lambda}_n^{IS}(t) \text{ for all } t \in [\epsilon, M]) \to 1. \tag{1.2.8}$$

Hence, the asymptotic distribution of $\tilde{\lambda}_n^{IS}$ is the same as for the kernel estimator.

1.2.3 *Boundary problems*

If the function we are estimating has bounded support, kernel estimates are inconsistent at the boundary regions. This effect is due to the fact that the support of the kernel exceeds the range of the data. Indeed, if $t = cb$ for $c \in [0, 1)$, then the bias of the kernel estimator is

$$\begin{aligned}
&\left\{\int_0^{t+b} k_b(t-u)\,\lambda(u)\,du - \lambda(t)\right\} \\
&= \int_{-1}^{c} k(y)\,\{\lambda(t-by) - \lambda(t)\}\,dy - \lambda(t)\int_c^1 k(y)\,dy \\
&= -b\lambda'(t)\int_{-1}^{c} k(y)y\,dy + \frac{1}{2}b^2\lambda''(t)\int_{-1}^{c} k(y)y^2\,dy - \lambda(t)\int_c^1 k(y)\,dy + o(b^2).
\end{aligned} \tag{1.2.9}$$

We do not get the usual bias of order b^2 because $\int_{-1}^{t/b} k(y)\,dy \neq 0$ and $\int_{-1}^{c} k(y)y\,dy \neq 0$.

In order to prevent inconsistency problems at the boundaries of the support, different approaches have been tried, including penalization (see for instance, Groeneboom and Jongbloed, 2013) and boundary corrections (see for instance, Albers, 2012). However, no method performs strictly better than the others. We choose to apply some boundary correction which allows the shape of the kernel to change in the boundaries. It is constructed by a linear combinations of $k(u)$ and $uk(u)$ with coefficients depending on the value near the boundary (see Durot, Groeneboom, and Lopuhaä, 2013; Zhang and Karunamuni, 1998). To be precise, we define, for instance, the smoothed isotonic estimator $\hat{\lambda}_n^{SI}$ by

$$\hat{\lambda}_n^{SI}(t) = \int_{(t-b)\vee 0}^{(t+b)\wedge 1} k_b^{(t)}(t-u)\,\tilde{\lambda}_n(u)\,du \tag{1.2.10}$$

with $k_b^{(t)}(u)$ denoting the rescaled kernel $b^{-1}k^{(t)}(u/b)$ and

$$k^{(t)}(u) = \begin{cases} \psi_1(\frac{t}{b})k(u) + \psi_2(\frac{t}{b})uk(u) & t \in [0, b], \\ k(u) & t \in (b, 1-b) \\ \psi_1(\frac{1-t}{b})k(u) - \psi_2(\frac{1-t}{b})uk(u) & t \in [1-b, 1], \end{cases} \tag{1.2.11}$$

where $k(u)$ is a standard kernel satisfying (1.2.1). For $s \in [-1, 1]$, the coefficients $\psi_1(s)$, $\psi_2(s)$ are determined by

$$\begin{aligned} \psi_1(s) \int_{-1}^{s} k(u)\,du + \psi_2(s) \int_{-1}^{s} uk(u)\,du &= 1, \\ \psi_1(s) \int_{-1}^{s} uk(u)\,du + \psi_2(s) \int_{-1}^{s} u^2 k(u)\,du &= 0. \end{aligned} \tag{1.2.12}$$

The coefficients ψ_1 and ψ_2 are not only well defined, but they are also continuously differentiable if the kernel k is assumed to be continuous (see Durot, Groeneboom, and Lopuhaä, 2013). Furthermore, it can be easily seen that, for each $t \in [0, b]$, equations (1.2.12) lead to

$$\int_{-1}^{t/b} k^{(t)}(u)\,du = 1 \qquad \text{and} \qquad \int_{-1}^{t/b} uk^{(t)}(u)\,du = 0. \tag{1.2.13}$$

Hence, by reasoning as in (1.2.9), we get a bias of order b^2 even in the boundary regions.

Note that boundary corrected kernel smoothed estimator coincides with the standard kernel smoothed estimator on $[b, 1-b]$.

1.3 GLOBAL ERRORS OF ESTIMATES

A lot of attention has been given in the literature to the pointwise asymptotic behavior of smooth and/or monotone estimators. However, for example for goodness of fit tests, global errors of estimates are needed instead of pointwise results. There are various measures of the global error of the estimators but the most common ones are the L_p-errors. Among L_p-errors, the L_1, L_2 and L_∞-errors are very popular. The L_1 and L_∞-errors are more natural because they can easily be visualized as the area and the maximum pointwise distance between the curves. While the L_1-distance can be used in hypothesis testing, the supremum distance is more useful for constructing confidence bands. Another widely used global measure of departure from the true parameter of interest is the Hellinger distance. It is a convenient metric in maximum likelihood problems, which goes back to Le Cam, 1973; Le Cam, 1970, and it has nice connections with Bernstein norms and empirical process theory methods to obtain rates of convergence, due fundamentally to Birgé and Massart, 1993, Wong and Shen, 1995, and others (see Section 3.4 of van der Vaart and Wellner, 1996 or Chapter 4 in van de Geer, 2000 for a more detailed overview).

For the Grenander estimator of a monotone density, a central limit theorem for the L_1-error was formulated in Groeneboom, 1983 and proven

rigorously in Groeneboom, Hooghiemstra, and Lopuhaä, 1999. A similar result was established in Durot, 2002 for the regression context. Extensions to general L_p-errors can be found in Kulikov and Lopuhaä, 2005 and in Durot, 2007, where the latter provides a unified approach that applies to a variety of statistical models. For the same general setup, an extremal limit theorem for the supremum distance has been obtained in Durot, Kulikov, and Lopuhaä, 2012. Consistency in Hellinger distance of shape constrained maximum likelihood estimators has been investigated in Pal, Woodroofe, and Meyer, 2007, Seregin and Wellner, 2010, and Doss and Wellner, 2016, whereas rates on Hellinger risk measures have been obtained in Seregin and Wellner, 2010, Kim and Samworth, 2016, and Kim, Guntuboyina, and Samworth, 2016. On the other hand, central limit theorems for L_p-errors of regular kernel density estimators have been obtained in Csörgő and Horváth, 1988 and Csörgő, Gombay, and Horváth, 1991. However there is no distribution theory for the global errors of smooth and isotonic estimators.

Compared to the pointwise behavior, deriving asymptotics for global errors requires some additional techniques such as strong approximations. Once more, we illustrate the main ideas for the monotone density model. Let $\hat{f}_n$ be the Grenander estimator of a decreasing density $f : [0,1] \to [0,\infty)$. Then, for $1 \leqslant p < 2.5$, we have (see Theorem 1.1 in Kulikov and Lopuhaä, 2005 or Theorem 2 in Durot, 2007)

$$n^{1/6}\left\{n^{p/3}\int_0^1 \left|\hat{f}_n(t) - f(t)\right|^p \, dt - m_p\right\} \xrightarrow{d} N(0,\sigma_p^2), \tag{1.3.1}$$

where $m_p = \mathbb{E}\left[|X(0)|^p\right]\int_0^1 |4f'(t)f(t)|^{p/3}\,dt$ with $X(a)$ as in (1.1.12) and the variance

$$\sigma_p^2 = 8k_p\int_0^1 |4f'(t)f(t)|^{2(p-1)/3} f(t)\,dt$$

depends on

$$k_p = \int_0^\infty \operatorname{cov}\left(|X(0)|^p, |X(a)-a|^p\right)\,da. \tag{1.3.2}$$

For $k > 2.5$, the inconsistency of $\hat{f}_n$ at the boundaries starts to dominate the behavior of the L_p-error, so the result is no longer true. As in the pointwise case, the central limit theorem in (1.3.1) is obtained through the more tractable inverse process $\widehat{U}_n$ defined in (1.1.8). First, using the switching relation (1.1.9), it can be shown that the L_p-error

$$\mathscr{J}_n := n^{p/3}\int_0^1 |\hat{f}_n(t) - f(t)|^p\,dt$$

can be approximated by

$$\mathscr{J}_n \approx n^{p/3} \int_{f(1)}^{f(0)} |\widehat{U}_n(a) - g(a)|^p |g'(a)|^{1-p} \, da,$$

where $g(a) = f^{-1}(a)$. Using properties of the argmax function we obtain

$$n^{1/3} \left\{ \widehat{U}_n(a) - g(a) \right\} = \operatorname*{argmax}_t \mathbb{Z}_n(t),$$

where

$$\begin{aligned} \mathbb{Z}_n(t) = {} & n^{2/3} \left\{ (F_n - F)(g(a) + n^{-1/3}t) + (F_n - F)(g(a)) \right\} \\ & + n^{2/3} \left\{ F((g(a) + n^{-1/3}t)) - F(g(a)) - f(g(a))n^{-1/3}t \right\}. \end{aligned} \tag{1.3.3}$$

By a Taylor expansion, the second term in (1.3.3) can be approximated by $\frac{1}{2}f'(t_0)t^2$. Moreover, from the embedding in Komlós, Major, and Tusnády, 1975, the empirical process $G_n(t) = n^{1/2}[F_n(t) - F(t)]$ can uniformly be approximated by $B_n(F(t))$, where B_n is a standard Brownian bridge constructed on the same probability space as F_n. It follows that the process $\mathbb{Z}_n(t)$ converges in distribution to the process

$$\mathbb{Z}(t) = W(at) + \frac{1}{2}f'(t_0)t^2.$$

The convergence in distribution of $\mathbb{Z}_n$ to $\mathbb{Z}$ is sufficient when dealing with pointwise limit results as in Section 1.1.4. For the global errors, since the speed of convergence in (1.3.1) is faster than in the pointwise case, we also need the convergence of $\mathbb{Z}_n$ to $\mathbb{Z}$ to be sufficiently fast. Till now, this has been obtained through strong embedding results that approximate the relevant process Λ_n by a Brownian motion or Brownian bridge (see for example Komlós, Major, and Tusnády, 1975; Major and Rejtő, 1988). Afterwards, for the approximating process, asymptotic normality follows from a big-blocks-small-blocks procedure. The idea is to partition the interval $[0, 1]$ in big blocks and small blocks. The small blocks do not contribute to the limit distribution and they separate the big blocks in such a way that the integrals over the big blocks become independent. This follows from the independence of the increments of the Brownian motion and the fact that the argmax of a Brownian motion with parabolic drift can actually be localized. Then, since the sum of the integrals over the big blocks is a sum of independent random variables, the central limit theorem can be used to derive the limit distribution.

The situation is slightly different when dealing with the kernel estimator, mainly because there is no need to go through the inverse process and

argmax approximations. In this case, the L_p-error of the kernel estimator $\tilde{f}_n^s$ of a smooth density on $\mathbb{R}$ (without monotonicity constraints) satisfies the following central limit theorem

$$b^{-1/2}\left\{(nb)^{p/2}\int_{\mathbb{R}}\left|\tilde{f}_n^s(t)-f(t)\right|^p\,dt-m_n(p)\right\}\xrightarrow{d}N(0,\sigma^2(p))$$

for some $m_n(p)$ and $\sigma^2(p)$ (see Csörgő and Horváth, 1988) that depend on whether $nb^5\to 0$ or $nb^5\to C>0$. Again the main idea is to use the embedding in Komlós, Major, and Tusnády, 1975

$$F_n(t)=F(t)+n^{-1/2}B_n(F(t))+O\left(n^{-1}\log n\right),$$

which holds uniformly in t, for approximating

$$\tilde{f}_n^s(t)-f(t)=\int k_b(t-u)\,d(F_n-F)(u)$$

by the corresponding Gaussian process $\Gamma_n(t)=n^{-1/2}\int k_b(t-u)\,dB_n(F(u))$. Then, if B_n was a Brownian motion, the asymptotic normality of the L_p-norm of Γ_n follows by a big-blocks-small-blocks procedure. The last step consists in showing that even if B_n is a Brownian bridge, the asymptotic behavior remains the same as for a Brownian motion. Note that no boundary problems are present here since the density is smooth on the whole real line.

1.4 OUTLINE

The thesis consists of two main parts. The first part establishes results on the pointwise behavior of smooth isotonic estimators in the right censoring and Cox regression model. The second part deals with asymptotics for global errors of estimates in a general setting which includes estimation of a probability density, a regression function or a failure rate under monotonicity constraints.

We start in Chapter 2 by considering kernel smoothed Grenander-type estimators for a monotone hazard rate and a monotone density in the presence of randomly right censored data. This is a relatively simple model since the limit distribution can be derived using a Kiefer-Wolfowitz type of result as in (1.2.6). We show that the estimators converge at rate $n^{2/5}$ and that the limit distribution at a fixed point is Gaussian with explicitly given mean and variance. To avoid inconsistency problems of standard kernel smoothing, we use some boundary correction. The obtained results are

used to construct pointwise confidence intervals and their performance is investigated through a simulation study.

The right censoring model is a special case of the Cox regression model that will be treated in Chapter 3. The lack of a Kiefer-Wolfowitz type of result makes the asymptotic analysis in this model more challenging and techniques different from the ones described in Section 1.2.2 need to be developed. We consider four smooth isotonic estimators for a monotone baseline hazard rate λ_0. Two of them are obtained by kernel smoothing the constrained maximum likelihood estimator and the Grenander-type estimator whereas the other two are a maximum smooth likelihood estimator and an isotonized kernel estimator. We analyze their asymptotic behavior and show that they are asymptotically normal at rate $n^{m/(2m+1)}$, when λ_0 is $m \geqslant 2$ times continuously differentiable. It turns out that the isotonized kernel estimator is asymptotically equivalent to the kernel smoothed isotonic estimators, while the maximum smoothed likelihood estimator exhibits the same asymptotic variance but a different bias. Finally, we present numerical results on pointwise confidence intervals that illustrate the comparable behavior of the four methods.

Once results for the pointwise behavior are established the interest naturally moves towards global errors of the estimators. Even if our main motivation was the Cox regression model, we decided to start with a simpler situation for two main reasons. First, there exist no results on the global errors of smooth and isotonic estimators even in more common models such as density or regression function estimation. Second, dealing directly with the Cox model is more challenging since no strong approximation result is available for the Breslow estimator. Hence, here we focus on a general setup, considered also in Durot, 2007 and Durot and Lopuhaä, 2014. It includes estimation of a monotone density, regression function and hazard rate. The main assumption is that there exist a strong approximation of Λ_n by a Gaussian process.

In Chapter 4 we consider Grenander type estimators for a monotone function $\lambda : [0,1] \to \mathbb{R}^+$, obtained as the slope of a concave (convex) estimate of the primitive of λ. Our main result is a central limit theorem for the Hellinger loss of this estimator. Moreover, we also propose a test for exponentiality knowing that the density is decreasing based on the Hellinger distance between a parametric estimator and the Grenander-type estimator. Its performance is investigated through simulation studies.

We proceed in Chapter 5 considering the process $\widehat{\Lambda}_n - \Lambda_n$, where Λ_n is a cadlag step estimator for the primitive Λ of a nonincreasing function λ on $[0,1]$, and $\widehat{\Lambda}_n$ is the least concave majorant of Λ_n. We extend the re-

sults in Kulikov and Lopuhaä, 2006, 2008 to the general setting considered in Durot, 2007. Under this setting we prove that a suitably scaled version of $\widehat{\Lambda}_n - \Lambda_n$ converges in distribution to the corresponding process for two-sided Brownian motion with parabolic drift and we establish a central limit theorem for the L_p-distance between $\widehat{\Lambda}_n$ and Λ_n. Such result is needed in the next chapter for dealing with smoothed Grenander-type estimators.

Finally, the asymptotic behavior of the L_p-distance between a monotone function on a compact interval and a smooth estimator of this function is investigated in Chapter 6. Our main result is a central limit theorem for the L_p-error of smooth isotonic estimators obtained by smoothing a Grenander-type estimator or isotonizing the ordinary kernel estimator. As a preliminary result we establish a similar result for ordinary kernel estimators. We also perform a simulation study for testing monotonicity on the basis of the L_2-distance between the kernel estimator and the smoothed Grenander-type estimator.

The last part of the thesis consists in some supplemental material containing proofs and additional technicalities.

Part II

LOCAL ASYMPTOTIC INFERENCE

2

SMOOTH ISOTONIC ESTIMATION UNDER RIGHT CENSORING

In this chapter we consider kernel smoothed Grenander-type estimators for the hazard function and the probability density in the presence of randomly right censored data. The results presented are based on:

Lopuhaä, H. P. and Musta E. (2017) "Smooth estimation of a monotone hazard and a monotone density under random censoring". *Statictica Neerlandica 71.1*, pp. 58-82.

The Grenander estimators of a monotone hazard and a monotone density in the random censorship model (see Example 1.1.1) were considered in Huang and Wellner, 1995. They established consistency and the limit distribution, together with the asymptotic equivalence with the maximum likelihood estimators. On the other hand, the smoothed maximum likelihood estimator of a monotone hazard under right censoring was investigated in Groeneboom and Jongbloed, 2014 and its limit distribution is stated in their Theorem 11.8. Hence, it seems quite natural to address the problem of the smoothed Grenander-type estimator.

We derive the limit distribution of the smoothed Grenander-type estimators using a a rather short and direct argument that relies on the method developed in Groeneboom and Jongbloed, 2013 (for non-censored observations) together with a Kiefer-Wolfowitz type of result derived in Durot and Lopuhaä, 2014. Both Theorem 2.2.2 and Theorem 2.3.4, highlight the fact that also after applying smoothing techniques, the constrained maximum likelihood estimator and the Grenander estimator remain asymptotically equivalent. In order to be able to apply the Kiefer-Wolfowitz type of result in Durot and Lopuhaä, 2014 we need to consider only estimation on a restricted interval $[0, \tau^*]$ that does not contain the end point of the support. Furthermore, boundary kernels are used to avoid inconsistency problems at the boundaries.

The chapter is organized as follows. In Section 2.1 we briefly introduce the Grenander estimator in the random censorship model and recall some results needed in the sequel. The smoothed estimator of a monotone hazard

function is described in Section 2.2 and it is shown to be asymptotically normally distributed. Moreover, a smooth estimator based on boundary kernels is studied and uniform consistency is derived. Using the same approach, in Section 2.3 we deal with the problem of estimating a smooth monotone density function. Section 2.4 is devoted to numerical results on pointwise confidence intervals.

2.1 THE RIGHT CENSORING MODEL

Suppose we have an i.i.d. sample $X_1, \ldots, X_n$ with distribution function F and density f, representing the survival times. Let $C_1, \ldots, C_n$ be the i.i.d. censoring variables with a distribution function G and density g. Under the random censorship model, we assume that the survival time X and the censoring time C are independent and the observed data consists of i.i.d. pairs of random variables $(T_1, \Delta_1), \ldots, (T_n, \Delta_n)$, where T denotes the follow-up time $T = \min(X, C)$ and $\Delta = \mathbb{1}_{\{X \leqslant C\}}$ is the censoring indicator.

Let H and H^{uc} denote the distribution function of the follow-up time and the sub-distribution function of the uncensored observations, respectively, i.e., $H^{uc}(t) = \mathbb{P}(T \leqslant t, \Delta = 1)$. Note that $H^{uc}(t)$ and $H(t)$ are differentiable with derivatives

$$h^{uc}(t) = f(t)\,(1 - G(t))$$

and

$$h(t) = f(t)(1 - G(t)) + g(t)(1 - F(t))$$

respectively. We also assume that $\tau_H = \tau_G < \tau_F \leqslant \infty$, where τ_F, τ_G and τ_H are the end points of the support of F, G and H.

First, we aim at estimating the hazard function λ (see (1.1.1)), subject to the constraint that it is increasing (the case of a decreasing hazard is analogous), on the basis of n observations $(T_1, \Delta_1), \ldots, (T_n, \Delta_n)$. Since we want to derive asymptotic normality using the Kiefer-Wolfowitz type of result in Durot and Lopuhaä, 2014, which holds only on intervals $[0, \tau^*]$ for $\tau^* < \tau_H$, we consider only estimation on $[0, \tau^*]$. A similar approach was considered in Groeneboom and Jongbloed, 2013, when estimating a monotone hazard of uncensored observations. Let Λ_n be the Nelson-Aalen estimator of the cumulative hazard function Λ defined in (1.1.3). Fix $0 < \tau^* < \tau_H$. The Grenander-type estimator $\tilde{\lambda}_n$ of λ is defined as the left-hand slope of the greatest convex minorant $\hat{\Lambda}_n$ of Λ_n on $[0, \tau^*]$. In practice we might not even know τ_H, so the choice of an estimation interval is necessary. It is reasonable to take as τ^* the 95%-empirical quantile of the follow-up times, because this

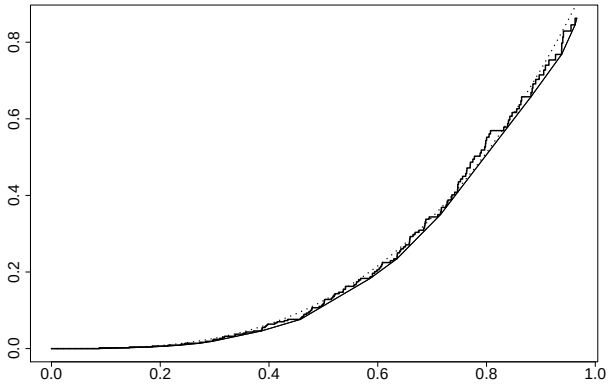


Figure 1: The Nelson-Aalen estimator (piecewise constant solid line) of the cumulative hazard (dotted line) and its greatest convex minorant (solid line).

converges to the theoretical 95%-quantile, which is strictly smaller than τ_H. Otherwise one can choose $\tau^* < T_{(n)}$.

Figure 1 shows the Nelson-Aalen estimator and its greatest convex minorant for a sample of $n = 500$ from a Weibull distribution with shape parameter 3 and scale parameter 1 for the event times and the uniform distribution on $(0, 1.3)$ for the censoring times. We consider only the data up to the last observed time before the 90% quantile of H. The resulting Grenander-type estimator can be seen in Figure 2.

In Huang and Wellner, 1995, the same type of estimator is considered without the restriction on $[0, \tau^*]$. Note that, in practice this requires the knowledge of τ_H. They show that the Grenander estimator of a nondecreasing hazard rate satisfies the following pointwise consistency result

$$\lambda(t-) \leqslant \liminf_{n\to\infty} \tilde{\lambda}_n(t) \leqslant \limsup_{n\to\infty} \tilde{\lambda}_n(t) \leqslant \lambda(t+), \tag{2.1.1}$$

with probability one and for all $0 < t < \tau_H$, where $\lambda(t-)$ and $\lambda(t+)$ denote the left and right limit at t. For our version of the estimator such result would hold for all $0 < t < \tau^*$.

Moreover, we will also make use of the fact that for any $0 < M < \tau_H$,

$$\sqrt{n} \sup_{u\in[0,M]} |\Lambda_n(u) - \Lambda(u)| = O_P(1), \tag{2.1.2}$$

(see for instance, Lopuhaä and Nane, 2013, Theorem 5, in the case $\beta = 0$, or van der Vaart and Wellner, 1996, Example 3.9.19).

It becomes useful to introduce

$$\Phi(t) = \int \mathbb{1}_{[t,\infty)}(y)\, d\mathbb{P}(y,\delta) = 1 - H(t) \tag{2.1.3}$$

and

$$\Phi_n(t) = \int \mathbb{1}_{[t,\infty)}(y)\, d\mathbb{P}_n(y,\delta), \tag{2.1.4}$$

where $\mathbb{P}$ is the probability distribution of (T,Δ) and $\mathbb{P}_n$ is the empirical measure of the pairs (T_i,Δ_i), $i = 1,\dots,n$. From Lemma 4 in Lopuhaä and Nane, 2013 we have,

$$\sup_{t\in[0,\tau_H]} |\Phi_n(t) - \Phi(t)| \to 0, \text{ a.s. and } \sqrt{n} \sup_{t\in[0,\tau_H]} |\Phi_n(t) - \Phi(t)| = O_P(1). \tag{2.1.5}$$

Let us notice that, with these notations, we can also write

$$\Lambda_n(t) = \int \frac{\delta \mathbb{1}_{\{u\leqslant t\}}}{\Phi_n(u)}\, d\mathbb{P}_n(u,\delta), \qquad \Lambda(t) = \int \frac{\delta \mathbb{1}_{\{u\leqslant t\}}}{\Phi(u)}\, d\mathbb{P}(u,\delta). \tag{2.1.6}$$

Our second objective is to estimate a monotone (e.g., increasing) density function f . In this case the Grenander-type estimator $\tilde{f}_n$ of f is defined as the left-hand slope of the greatest convex minorant $\hat{F}_n$ of the Kaplan-Meier estimator F_n (see (1.1.4)) restricted on $[0,\tau^*]$. Again, for the non-restricted version, pointwise consistency of the Grenander estimator of a nondecreasing density:

$$f(t-) \leqslant \liminf_{n\to\infty} \tilde{f}_n(t) \leqslant \limsup_{n\to\infty} \tilde{f}_n(t) \leqslant f(t+), \tag{2.1.7}$$

with probability one, for all $0 < t < \tau_H$, where $f(t-)$ and $f(t+)$ denote the left and right limit at t, is proved in Huang and Wellner, 1995. For any $0 < M < \tau_H$, it holds

$$\sqrt{n} \sup_{u\in[0,M]} |F_n(u) - F(u)| = O_P(1), \tag{2.1.8}$$

(see for instance, Breslow and Crowley, 1974, Theorem 5). Moreover, by Theorem 2 in Major and Rejtő, 1988, for each $0 < M < \tau_H$ and $x \geqslant 0$, we have the following strong approximation

$$\mathbb{P}\left\{ \sup_{t\in[0,M]} n \left| F_n(t) - F(t) - n^{-1/2}(1-F(t))W\circ L(t) \right| > x + K_1 \log n \right\} \leqslant K_2 e^{-K_3 x}, \tag{2.1.9}$$

where K_1, K_2, K_3 are positive constants, W is a Brownian motion and

$$L(t) = \int_0^t \frac{\lambda(u)}{1-H(u)}\, du. \tag{2.1.10}$$

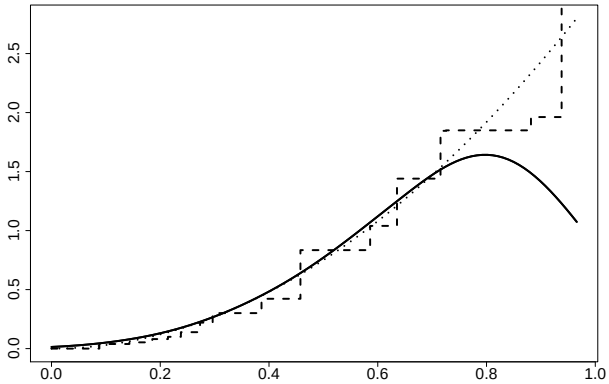


Figure 2: The Grenander-type estimator (dashed line) of the hazard rate (dotted line) and the kernel smoothed version (solid line).

2.2 SMOOTH ESTIMATION OF A MONOTONE HAZARD

Next, we introduce the smoothed Grenander-type estimator $\tilde{\lambda}_n^{SG}$ of an increasing hazard. Let k be a standard kernel, i.e.,

$$k \text{ is a symmetric probability density with support } [-1, 1]. \tag{2.2.1}$$

For a fixed $t \in [0, \tau^*]$, the smoothed Grenander-type estimator $\tilde{\lambda}_n^{SG}$ is defined by

$$\tilde{\lambda}_n^{SG}(t) = \int_{(t-b)\vee 0}^{(t+b)\wedge\tau^*} k_b(t-u)\,\tilde{\lambda}_n(u)\,du = \int k_b(t-u)\,d\hat{\Lambda}_n(u), \tag{2.2.2}$$

where k_b is the rescaled kernel function as in Section 1.2.1. Figure 2 shows the Grenander-type estimator together with the kernel smoothed version for the same sample as in Figure 1. We used the triweight kernel function

$$k(u) = \frac{35}{32}(1-u^2)^3 \mathbb{1}_{\{|u|\leqslant 1\}}$$

and the bandwidth $b = c_{opt}\, n^{-1/5}$, where c_{opt} is the asymptotically MSE-optimal constant (see (2.2.5)) calculated in the point $t_0 = 0.5$. Actually, the choice of the kernel function does not seem to effect the estimator.

The following result is rather standard when dealing with kernel smoothed isotonic estimators (see for instance, Nane, 2013, Chapter 5). For completeness, we provide a rigorous proof.

Theorem 2.2.1. *Let k be a kernel function satisfying* (2.2.1) *and let $\tilde{\lambda}^{SG}$ be the smoothed Grenander-type estimator defined in* (2.2.2)*. Suppose that the hazard function λ is nondecreasing and continuous. Then for each $0 < \epsilon < \tau^*$, it holds*

$$\sup_{t \in [\epsilon, \tau^* - \epsilon]} |\tilde{\lambda}_n^{SG}(t) - \lambda(t)| \to 0$$

with probability one.

Proof. First, note that for a fixed $t \in (0, \tau^*)$ and sufficiently large n, we have $0 < t - b < t + b < \tau^*$. We start by writing

$$\tilde{\lambda}_n^{SG}(t) - \lambda(t) = \tilde{\lambda}_n^{SG}(t) - \lambda_{(n)}(t) + \lambda_{(n)}(t) - \lambda(t),$$

where $\lambda_{(n)}(t) = \int k_b(t-u)\,\lambda(u)\,du$. Then, a change of variable yields

$$\lambda_{(n)}(t) - \lambda(t) = \int_{-1}^{1} k(y)\,\{\lambda(t - by) - \lambda(t)\}\,dy.$$

Using the continuity of λ and applying the dominated convergence theorem, we obtain that, for each $t \in (0, \tau^*)$,

$$\lambda_{(n)}(t) \to \lambda(t), \quad \text{as } n \to \infty. \tag{2.2.3}$$

On the other hand,

$$\tilde{\lambda}_n^{SG}(t) - \lambda_{(n)}(t) = \int_{-1}^{1} k(y)\,\left\{\tilde{\lambda}_n(t - by) - \lambda(t - by)\right\}\,dy.$$

Choose $\epsilon > 0$. Then by continuity of λ, we can find $\delta > 0$, such that $0 < t - \delta < t + \delta < \tau^*$ and $|\lambda(t+\delta) - \lambda(t-\delta)| < \epsilon$. Then, there exists N such that, for all $n \geqslant N$ and for all $y \in [-1, 1]$, it holds $|by| < \delta$. Hence, by the monotonicity of the hazard rate, we get

$$\tilde{\lambda}_n(t-\delta) - \lambda(t+\delta) \leqslant \tilde{\lambda}_n(t - by) - \lambda(t - by) \leqslant \tilde{\lambda}_n(t+\delta) - \lambda(t-\delta).$$

It follows from (2.1.1) and (1.2.1) that

$$-\epsilon \leqslant \liminf_{n\to\infty} \tilde{\lambda}_n^{SG}(t) - \lambda_{(n)}(t) \leqslant \limsup_{n\to\infty} \tilde{\lambda}_n^{SG}(t) - \lambda_{(n)}(t) \leqslant \epsilon,$$

with probability one. Since $\epsilon > 0$ is arbitrary, together with (2.2.3), this proves the strong pointwise consistency at each fixed $t \in (0, \tau^*)$. Finally, uniform consistency in $[\epsilon, \tau^* - \epsilon]$ follows from the fact that we have a sequence of monotone functions converging pointwise to a continuous, monotone function on a compact interval. □

It is worth noticing that, if one is willing to assume that λ is twice differentiable with uniformly bounded first and second derivatives, and that k is differentiable with a bounded derivative, then we get a more precise result on the order of convergence

$$\sup_{t\in[\epsilon,\tau^*-\epsilon]} |\tilde{\lambda}_n^{SG}(t) - \lambda(t)| = O_P(b^{-1}n^{-1/2} + b^2).$$

Such extra assumptions are considered in Theorem 5.2 in Nane, 2013 for the Cox model and the right censoring model is just a particular case with regression parameters $\beta = 0$. Furthermore, in a similar way, it can be proved that also the estimator for the derivative of the hazard is uniformly consistent in $[\epsilon, \tau^* - \epsilon]$, provided that λ is continuously differentiable and the kernel is differentiable with bounded derivative.

The pointwise asymptotic normality of the smoothed Grenander estimator is stated in the next theorem. Its proof is inspired by the approach used in Groeneboom and Jongbloed, 2013. The key step consists in using a Kiefer-Wolfowitz type of result for the Nelson-Aalen estimator, which has recently been obtained by Durot and Lopuhaä, 2014.

Theorem 2.2.2. *Let λ be a nondecreasing and twice continuously differentiable hazard such that λ and λ' are strictly positive. Let k satisfy* (2.2.1) *and suppose that it is differentiable with a uniformly bounded derivative. If $bn^{1/5} \to c \in (0, \infty)$, then for each $t \in (0, \tau^*)$,*

$$n^{2/5}\big(\tilde{\lambda}_n^{SG}(t) - \lambda(t)\big) \xrightarrow{d} N(\mu, \sigma^2),$$

where

$$\mu = \frac{1}{2}c^2\lambda''(t)\int u^2 k(u)\,du \quad \textit{and} \quad \sigma^2 = \frac{\lambda(t)}{c\,(1-H(t))}\int k^2(u)\,du. \tag{2.2.4}$$

For a fixed $t \in (0, \tau^*)$, the asymptotically MSE-optimal bandwidth b for $\tilde{\lambda}^{SG}$ is given by $c_{opt}(t)n^{-1/5}$, where

$$c_{opt}(t) = \left\{\lambda(t)\int k^2(u)\,du\right\}^{1/5}\left\{(1-H(t))\lambda''(t)^2\left(\int u^2\,k(u)\,du\right)^2\right\}^{-1/5}. \tag{2.2.5}$$

Proof. Once again we fix $t \in (0, \tau^*)$. Then, for sufficiently large n, we have $0 < t - b < t + b \leqslant \tau^*$. We write

$$\begin{aligned}\tilde{\lambda}_n^{SG}(t) = &\int k_b(t-u)\,d\Lambda(u) + \int k_b(t-u)\,d(\Lambda_n - \Lambda)(u)\\ &+ \int k_b(t-u)\,d(\hat{\Lambda}_n - \Lambda_n)(u).\end{aligned} \tag{2.2.6}$$

As in (1.2.5), the first (deterministic) term on the right hand side of (2.2.6) gives us the asymptotic bias:

$$n^{2/5}\left\{\int_{t-b}^{t+b} k_b(t-u)\,\lambda(u)\,du-\lambda(t)\right\}\to\frac{1}{2}\,c^2\,\lambda''(t)\int_{-1}^{1} y^2\,k(y)\,dy.$$

On the other hand, the last term on the right hand side of (2.2.6) converges to zero in probability. Indeed, integration by parts formula enables us to write

$$\begin{aligned}
&n^{2/5}\int_{t-b}^{t+b} k_b(t-u)\,d(\hat{\Lambda}_n-\Lambda_n)(u)\\
&=n^{2/5}\int_{t-b}^{t+b}\left\{\hat{\Lambda}_n(u)-\Lambda_n(u)\right\}\frac{1}{b^2}\,k'\left(\frac{t-u}{b}\right)du\\
&=\frac{n^{2/5}}{b}\int_{-1}^{1}\left\{\hat{\Lambda}_n(t-by)-\Lambda_n(x-by)\right\}k'(y)\,dy,
\end{aligned}$$

and then we use $\sup_{t\in[0,M]}|\hat{\Lambda}_n(t)-\Lambda_n(t)|=O_p(n^{-2/3}(\log n)^{2/3})$ (see Durot and Lopuhaä, 2014, Corollary 3.4) together with the boundedness of k'.

What remains is to prove that

$$n^{2/5}\int k_b(t-u)\,d(\Lambda_n-\Lambda)(u)\xrightarrow{d}N(0,\sigma^2),$$

where σ^2 is defined in (2.2.4). Let us start by writing

$$n^{2/5}\int_{t-b}^{t+b} k_b(t-u)\,d(\Lambda_n-\Lambda)(u)=\frac{1}{\sqrt{bn^{1/5}}}\int_{-1}^{1} k(y)\,d\hat{W}_n(y),$$

where, for each $y\in[-1,1]$, we define

$$\begin{aligned}
\hat{W}_n(y)&=\sqrt{\frac{n}{b}}\{\Lambda_n(t-by)-\Lambda_n(t)-\Lambda(t-by)+\Lambda(t)\}\\
&=\sqrt{\frac{n}{b}}\int\frac{\delta}{\Phi_n(u)}\left\{\mathbb{1}_{[0,t-by]}(u)-\mathbb{1}_{[0,t]}(u)\right\}d\mathbb{P}_n(u,\delta)\\
&\quad-\sqrt{\frac{n}{b}}\int\frac{\delta}{\Phi(u)}\left\{\mathbb{1}_{[0,t-by]}(u)-\mathbb{1}_{[0,t]}(u)\right\}d\mathbb{P}(u,\delta)\\
&=b^{-1/2}\int\frac{\delta}{\Phi(u)}\left\{\mathbb{1}_{[0,t-by]}(u)-\mathbb{1}_{[0,t]}(u)\right\}d\sqrt{n}(\mathbb{P}_n-\mathbb{P})(u,\delta)\\
&\quad+\sqrt{\frac{n}{b}}\int\delta\left\{\mathbb{1}_{[0,t-by]}(u)-\mathbb{1}_{[0,t]}(u)\right\}\left\{\frac{1}{\Phi_n(u)}-\frac{1}{\Phi(u)}\right\}d\mathbb{P}_n(u,\delta).
\end{aligned}\tag{2.2.7}$$

Here we took advantage of the representations in (2.1.6). The last term in the right hand side of (2.2.7) is bounded in absolute value by

$$\sqrt{\frac{n}{b}}\frac{1}{\Phi_n(M)\,\Phi(M)}\int \delta\left|\mathbb{1}_{[0,t-by]}(u)-\mathbb{1}_{[0,t]}(u)\right||\Phi(u)-\Phi_n(u)|\,d\mathbb{P}_n(u,\delta)=o_P(1).$$

Indeed, by using (2.1.5), we obtain that $1/\Phi_n(M) = O_P(1)$ and then it suffices to prove that

$$b^{-1/2}\int \delta\left|\mathbb{1}_{[0,t-by]}(u)-\mathbb{1}_{[0,t]}(u)\right|\,d\mathbb{P}_n(u,\delta)=o_P(1).$$

To do so, we write the left hand side as

$$\begin{aligned}
&b^{-1/2}\int \delta\left|\mathbb{1}_{[0,t-by]}(u)-\mathbb{1}_{[0,t]}(u)\right|\,d\mathbb{P}(u,\delta)\\
&\quad+b^{-1/2}\int \delta\left|\mathbb{1}_{[0,t-by]}(u)-\mathbb{1}_{[0,t]}(u)\right|\,d(\mathbb{P}_n-\mathbb{P})(u,\delta)\\
&=b^{-1/2}\left|H^{uc}(t-by)-H^{uc}(t)\right|+O_p(b^{-1/2}n^{-1/2})=o_P(1).
\end{aligned}\tag{2.2.8}$$

Here we have used that H^{uc} is continuously differentiable and that the class of indicators of intervals forms a VC-class, and is therefore Donsker (see van der Vaart and Wellner, 1996, Theorem 2.6.7 and Theorem 2.5.2).

The last step consists in showing that

$$\begin{aligned}
&\frac{1}{\sqrt{b}}\int\frac{\delta}{\Phi(u)}\left\{\mathbb{1}_{[0,t-by]}(u)-\mathbb{1}_{[0,t]}(u)\right\}\,d\sqrt{n}(\mathbb{P}_n-\mathbb{P})(u,\delta)\\
&\qquad\xrightarrow{d}\sqrt{\frac{\lambda(t)}{1-H(t)}}\,W(y),
\end{aligned}$$

where W is a two sided Brownian motion. This follows from Theorem 2.11.23 in van der Vaart and Wellner, 1996. Indeed, we can consider the functions

$$f_{n,y}(u,\delta)=b^{-1/2}\frac{\delta}{\Phi(u)}\left\{\mathbb{1}_{[0,t-by]}(u)-\mathbb{1}_{[0,t]}(u)\right\},\qquad y\in[-1,1].$$

with envelopes $F_n(u,\delta) = b^{-1/2}\Phi(M)^{-1}\delta\mathbb{1}_{[t-b,t+b]}(u)$. It can be easily checked that

$$\|F_n\|^2_{L_2(\mathbb{P})}=\frac{1}{b\,\Phi^2(M)}\int_{t-b}^{t+b}f(u)(1-G(u))\,du=O(1),$$

and that for all $\eta>0$,

$$\frac{1}{b\Phi^2(M)}\int_{\{u:b^{-1/2}\Phi(M)^{-1}\mathbb{1}_{[t-b,t+b]}(u)>\eta\sqrt{n}\}}f(u)(1-G(u))\,du\to 0.$$

Moreover, for every sequence $\delta_n \downarrow 0$, we have

$$\frac{1}{b\Phi^2(M)} \sup_{|s-r|<\delta_n} \int_{t-b(s\vee r)}^{t-b(s\wedge r)} f(u)(1-G(u))\,du \to 0.$$

Since $f_{n,y}$ are sums and products of bounded monotone functions, the bracketing number is bounded (see van der Vaart and Wellner, 1996, Theorem 2.7.5) by

$$\log N_{[]}\left(\epsilon\|F_n\|_{L_2(\mathbb{P})}, \mathcal{F}_n, \|\cdot\|_{L_2(\mathbb{P})}\right) \lesssim \log\left(1/\epsilon\,\|F_n\|_{L_2(\mathbb{P})}\right).$$

Hence, since $\|F_n\|_{L_2(\mathbb{P})}$ is bounded we obtain

$$\int_0^{\delta_n} \sqrt{\log N_{[]}\left(\epsilon\|F_n\|_{L_2(\mathbb{P})}, \mathcal{F}_n, \|\cdot\|_{L_2(\mathbb{P})}\right)}\,d\epsilon \lesssim \delta_n + \int_0^{\delta_n} \sqrt{\log(1/\epsilon)}\,d\epsilon \to 0.$$

Finally, as in (2.2.8), for any $s \in [-1,1]$,

$$\mathbb{P}f_{n,s} = b^{-1/2}\{H^{uc}(t-bs) - H(t)\} \to 0.$$

Furthermore, for $0 \leqslant s \leqslant r \leqslant 1$, we have

$$\begin{aligned}\mathbb{P}f_{n,s}f_{n,r} &= b^{-1}\int_{t-bs}^{t} \frac{f(u)(1-G(u))}{\Phi^2(u)}\,du = b^{-1}\int_{t-bs}^{t} \frac{\lambda(u)}{1-H(u)}\,du \\ &\to \frac{\lambda(t)}{1-H(t)}s.\end{aligned}$$

Similarly, for $-1 \leqslant r \leqslant s \leqslant 0$, $\mathbb{P}f_{n,s}f_{n,r} \to -s\lambda(t)/(1-H(t))$, whereas $\mathbb{P}f_{n,s}f_{n,r} = 0$, for $sr < 0$. It follows that

$$\mathbb{P}f_{n,s}f_{n,r} - \mathbb{P}f_{n,s}\mathbb{P}f_{n,r} \to \begin{cases} \dfrac{\lambda(t)}{1-H(t)}(|s|\wedge|r|) & \text{, if } sr \geqslant 0; \\ 0 & \text{, if } sr < 0. \end{cases} \tag{2.2.9}$$

Consequently, according to Theorem 2.11.23 in van der Vaart and Wellner, 1996, the sequence of stochastic processes $\sqrt{n}(\mathbb{P}_n - \mathbb{P})f_{n,y}$ converges in distribution to a tight Gaussian process $\mathbb{G}$ with mean zero and covariance given on the right hand side of (2.2.9). Note that this is the covariance function of $\sqrt{\lambda(t)/[1-H(t)]}W$, where W is a two sided Brownian motion. We conclude that

$$\begin{aligned}&n^{2/5}\int_{t-b}^{t+b} k_b(t-u)\,d(\Lambda_n - \Lambda)(u) \\ &\quad= \frac{1}{\sqrt{bn^{1/5}}}\int_{-1}^{1} k(y)\,d\hat{W}_n(y) \\ &\quad\xrightarrow{d} \left(\frac{\lambda(t)}{c(1-H(t))}\right)^{1/2}\int_{-1}^{1} k(y)\,dW(y) \stackrel{d}{=} N\left(0, \frac{\lambda(t)}{c\,(1-H(t))}\int_{-1}^{1} k^2(y)\,dy\right).\end{aligned}$$

This proves the first part of the theorem.

The optimal c is then obtained by minimizing the asymptotic mean squared error

$$\mathrm{AMSE}(\tilde{\lambda}^{SG}, c) = \frac{1}{4}\, c^4\, \lambda''(t)^2 \left(\int u^2\, k(u)\, du\right)^2 + \frac{\lambda(t)}{c\,(1-H(t))} \int k^2(u)\, du$$

with respect to c. □

This result is in line with Theorem 11.8 in Groeneboom and Jongbloed, 2014 on the asymptotic distribution of the SMLE under the same model, which highlights the fact that even after applying a smoothing technique the MLE and the Grenander-type estimator remain asymptotically equivalent.

Standard kernel density estimators lead to inconsistency problems at the boundary. In order to prevent those, here we consider a boundary corrected kernel. To be precise, for $t \in [0, \tau^*]$, we define the smoothed Grenander-type estimator $\hat{\lambda}_n^{SG}$ by

$$\hat{\lambda}_n^{SG}(t) = \int_{(t-b)\vee 0}^{(t+b)\wedge\tau^*} k_b^{(t)}(t-u)\, \tilde{\lambda}_n(u)\, du = \int_{(t-b)\vee 0}^{(t+b)\wedge\tau^*} k_b^{(t)}(t-u)\, d\hat{\Lambda}_n(u), \tag{2.2.10}$$

with $k^{(t)}(u)$ as in (1.2.11). In this case, we obtain a stronger uniform consistency result which is stated in the next theorem.

Theorem 2.2.3. *Let $\hat{\lambda}_n^{SG}$ be defined by* (2.2.10) *and suppose that λ is nondecreasing and uniformly continuous. Assume that k satisfies* (1.2.1) *and is differentiable with a uniformly bounded derivative and that $bn^\alpha \to c \in (0, \infty)$. If $0 < \alpha < 1/2$, then*

$$\sup_{t\in[0,\tau^*]} \left|\hat{\lambda}_n^{SG}(t) - \lambda(t)\right| \to 0$$

in probability.

Proof. Write

$$\hat{\lambda}_n^{SG}(t) - \lambda(t) = \left(\hat{\lambda}_n^{SG}(t) - \lambda_{(n)}(t)\right) + \left(\lambda_{(n)}(t) - \lambda(t)\right),$$

where

$$\lambda_{(n)}(t) = \int_{(t-b)\vee 0}^{(t+b)\wedge\tau^*} k_b^{(t)}(t-u)\, \lambda(u)\, du.$$

We have to distinguish between three cases. First, we consider the case $t \in [0, b]$. By means of (1.2.13) and the fact that λ is uniformly continuous, a change of variable yields

$$\sup_{t\in[0,b]} |\lambda_{(n)}(t) - \lambda(t)| \leqslant \sup_{t\in[0,b]} \int_{-1}^{t/b} k^{(t)}(y)\big|\lambda(t-by) - \lambda(t)\big|\, dy \to 0. \tag{2.2.11}$$

On the other hand, integration by parts and a change of variable give

$$\begin{aligned}\hat{\lambda}_n^{SG}(t) - \lambda_{(n)}(t) &= \int_0^{t+b} k_b^{(t)}(t-u)\, d\big(\hat{\Lambda}_n - \Lambda\big)(u) \\ &= -\int_0^{t+b} \frac{\partial}{\partial u} k_b^{(t)}(t-u)\big(\hat{\Lambda}_n(u) - \Lambda(u)\big)\, du \\ &= \frac{1}{b}\int_{-1}^{t/b} \frac{\partial}{\partial y} k_b^{(t)}(y)\big(\hat{\Lambda}_n(t-by) - \Lambda(t-by)\big)\, dy.\end{aligned} \tag{2.2.12}$$

Consequently, we obtain

$$\sup_{t\in[0,b]} \left|\hat{\lambda}_n^{SG}(t) - \lambda_{(n)}(t)\right| \lesssim \frac{1}{b} \sup_{u\in[0,\tau^*]} \left|\hat{\Lambda}_n(u) - \Lambda(u)\right| = O_p(n^{-1/2+\alpha}),$$

because of (2.1.2) and the boundedness of the coefficients ϕ, ψ and of $k(u)$ and $k'(u)$. Together with (2.2.11) and since $0 < \alpha < 1/2$, this proves that,

$$\sup_{t\in[0,b]} \left|\hat{\lambda}_n^{SG}(t) - \lambda(t)\right| = o_P(1).$$

Similarly we also find

$$\sup_{t\in[b,\tau^*-b]} \left|\hat{\lambda}_n^{SG}(t) - \lambda(t)\right| = o_P(1).$$

When $t \in (b, \tau^* - b)$, by a change of variable and uniform continuity of λ, it follows that

$$\sup_{t\in(b,\tau^*-b)} \left|\lambda_{(n)}(t) - \lambda(t)\right| \leqslant \int_{-1}^{1} k(y)|\lambda(t-by) - \lambda(t)|\, dy \to 0. \tag{2.2.13}$$

Furthermore,

$$\hat{\lambda}_n^{SG}(t) = \int_{t-b}^{t+b} k_b(t-u)\, \tilde{\lambda}_n(u)\, du,$$

so that, arguing as in (2.2.12), we find that

$$\sup_{t\in(b,\tau^*-b)} \left|\hat{\lambda}_n^{SG}(t) - \lambda_{(n)}(t)\right| = O_p(n^{-1/2+\alpha}),$$

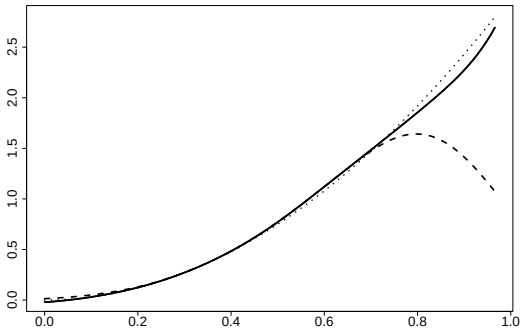


Figure 3: The kernel smoothed versions (standard-dashed line and boundary corrected-solid line) of the hazard rate (dotted line).

which, together with (2.2.13), proves that

$$\sup_{t\in(b,\tau^*-b)} \left|\hat{\lambda}_n^{SG}(t) - \lambda(t)\right| = o_P(1).$$

This proves the theorem. □

Note that for the unconstrained smoothed Grenander-type estimator on $[0, \tau_H]$, even with the boundary kernels, one can not avoid inconsistency problems at the end point of the support, i.e., uniform consistency would hold only on intervals $[0, M]$ for $M < \tau_H$. Although a bit surprising, this is to be expected because we can only control the distance between the Nelson-Aalen estimator and the cumulative hazard on intervals that stay away from the right boundary (see (2.1.2)). Figure 3 illustrates that boundary corrections improve the performance of the smooth estimator constructed with the standard kernel.

2.3 SMOOTHED ESTIMATION OF A MONOTONE DENSITY

This section is devoted to the smoothed Grenander-type estimator $\tilde{f}_n^{SG}$ of an increasing density f. Let k be a kernel function satisfying (2.2.1). For a fixed $t \in [0, \tau^*]$, $\tilde{f}_n^{SG}$ is defined by

$$\tilde{f}_n^{SG}(t) = \int_{(t-b)\vee 0}^{(t+b)\wedge\tau^*} k_b(t-u)\, \tilde{f}_n(u)\, du = \int k_b(t-u)\, d\hat{F}_n(u). \qquad (2.3.1)$$

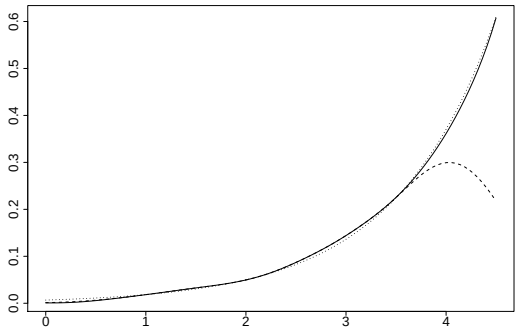


Figure 4: The kernel smoothed versions (standard-dashed line and boundary corrected-solid line) of the density function (dotted line).

We also consider the boundary corrected version of the smoothed Grenander-type estimator, defined by

$$\hat{f}_n^{SG}(t) = \int_{(t-b)\vee 0}^{(t+b)\wedge\tau^*} k_b^{(t)}(t-u)\,\tilde{f}_n(u)\,du = \int_{(t-b)\vee 0}^{(t+b)\wedge\tau^*} k_b^{(t)}(t-u)\,d\hat{F}_n(u), \tag{2.3.2}$$

with $k^{(t)}(u)$ as in (1.2.11). The following results can be proved in exactly the same way as Theorem 2.2.1 and Theorem 2.2.3.

Theorem 2.3.1. *Let k be a kernel function satisfying* (2.2.1) *and let $\tilde{f}^{SG}$ be the smoothed Grenander-type estimator defined in* (2.3.1)*. Suppose that the density function f is nondecreasing and continuous. Then for each $0 < \epsilon < \tau^*$, it holds*

$$\sup_{t\in[\epsilon,\tau^*-\epsilon]} |\tilde{f}_n^{SG}(t) - f(t)| \to 0$$

with probability one.

Theorem 2.3.2. *Let $\hat{f}_n^{SG}$ be defined by* (2.2.10) *and suppose that f is nondecreasing and uniformly continuous. Assume that k satisfies* (2.2.1) *and is differentiable with uniformly bounded derivatives and that $b\,n^{\alpha} \to c \in (0,\infty)$. If $0 < \alpha < 1/2$, then*

$$\sup_{t\in[0,\tau^*]} \left|\hat{f}_n^{SG}(t) - f(t)\right| \to 0$$

in probability.

Figure 4 shows the smooth isotonic estimators of an increasing density up to the 90% quantile of H for a sample of size $n = 500$. We choose

$$f(x) = (e^5 - 1)^{-1} e^x \mathbb{1}_{[0,5]}(x) \tag{2.3.3}$$

and

$$g(x) = (e^{5/2} - 1)^{-1} e^{x/2} \mathbb{1}_{[0,5]}(x) \tag{2.3.4}$$

for the censoring times. The bandwidth used is $b = c_{opt}\, n^{-1/5}$, where $c_{opt} = 4$ is the asymptotically MSE-optimal constant (see (2.3.10)) corresponding to $t_0 = 2.5$.

In order to derive the asymptotic normality of the smoothed Grenander-type estimator $\tilde{f}_n^{SG}$ we first provide a Kiefer-Wolfowitz type of result for the Kaplan-Meier estimator.

Lemma 2.3.3. *Let f be a nondecreasing and continuously differentiable density such that f and f' are strictly positive. Then we have*

$$\sup_{t \in [0,\tau^*]} |\hat{F}_n(t) - F_n(t)| = O_P\left(\frac{\log n}{n}\right)^{2/3},$$

where F_n is the Kaplan-Meier estimator and $\hat{F}_n$ is the greatest convex minorant of F_n on $[0, \tau^*]$.

Proof. We consider f on the interval $[0, \tau^*]$ and apply Theorem 2.2 in Durot and Lopuhaä, 2014. The density f satisfies condition (A1) of this theorem with $[a, b] = [0, \tau^*]$. Condition (2) of Theorem 2.2 in Durot and Lopuhaä, 2014 is provided by the strong approximation (2.1.9), with L defined in (2.1.10), $\gamma_n = O(n^{-1} \log n)^{2/3}$, and

$$B(t) = \Big(1 - F(L^{-1}(t))\Big) W(t), \quad t \in [L(0), L(\tau^*)]$$

where W is a Brownian motion. It remains to show that B satisfies conditions (A2)-(A3) of Theorem 2.2 in Durot and Lopuhaä, 2014 with $\tau = 1$. In order to check these conditions for the process B, let $x \in [L(0), L(\tau^*)] = [0, L(\tau^*)]$, $u \in (0, 1]$ and $\nu > 0$. Then

$$\begin{aligned}
&\mathbb{P}\left(\sup_{|x-y| \leqslant u} |B(x) - B(y)| > \nu\right) \\
&\leqslant \mathbb{P}\left(\sup_{|x-y| \leqslant u,\, y \in [0, L(\tau^*)]} |B(x) - B(y)| > \nu\right) \\
&\leqslant 2\mathbb{P}\left(\sup_{|x-y| \leqslant u} |W(x) - W(y)| > \frac{\nu}{3}\right) \\
&\qquad + \mathbb{P}\left(\sup_{|x-y| \leqslant u,\, y \in [0, L(\tau^*)]} |F(L^{-1}(x)) - F(L^{-1}(y))|\, |W(x)| > \frac{\nu}{3}\right).
\end{aligned} \tag{2.3.5}$$

Note that from the proof of Corollary 3.1 in Durot and Lopuhaä, 2014 it follows that W satisfies condition (A2) in Durot and Lopuhaä, 2014. This means that there exist $K_1, K_2 > 0$, such that the first probability on the right hand side of (2.3.5) is bounded by $K_1 \exp(-K_2 v^2 u^{-1})$. Furthermore, since $u \in (0, 1]$ and

$$|F(L^{-1}(x)) - F(L^{-1}(y))| \leqslant \frac{\sup_{u\in[0,\tau^*]} f(u)}{\inf_{u\in[0,\tau^*]} L'(u)} |x-y| \leqslant \frac{\sup_{u\in[0,\tau^*]} f(u)}{\inf_{u\in[0,\tau^*]} \lambda(u)} |x-y|,$$

the second probability on the right hand side of (2.3.5) is bounded by

$$\mathbb{P}\left(|W(x)| > \frac{v}{K_3\sqrt{u}}\right) \leqslant \mathbb{P}\left(\sup_{t\in[0,L(M)]} |W(t)| > \frac{v}{K_3\sqrt{u}}\right),$$

for some $K_3 > 0$. Hence, by the maximal inequality for Brownian motion, we conclude that there exist $K_1', K_2' > 0$ such that

$$\mathbb{P}\left(\sup_{|x-y|\leqslant u} |B(x) - B(y)| > v\right) \leqslant K_1' \exp\left(-K_2' v^2 u^{-1}\right)$$

which proves condition (A2) in Durot and Lopuhaä, 2014.

Let us now consider (A3). For all $x \in [0, L(\tau^*)]$, $u \in (0, 1]$, and $v > 0$, we obtain

$$\begin{aligned}
&\mathbb{P}\left(\sup_{u\leqslant z\leqslant x} \left\{B(x-z) - B(x) - vz^2\right\} > 0\right) \\
&\leqslant \mathbb{P}\left(\sup_{u\leqslant z\leqslant x} \left\{W(x-z) - W(x) - \frac{vz^2}{3}\right\} > 0\right) \\
&\quad + \mathbb{P}\left(\sup_{u\leqslant z\leqslant x} \left\{F(L^{-1}(x-z))[W(x) - W(x-z)] - \frac{vz^2}{3}\right\} > 0\right) \\
&\quad + \mathbb{P}\left(\sup_{u\leqslant z\leqslant x} \left\{\left(F(L^{-1}(x)) - F(L^{-1}(x-z))\right) W(x) - \frac{vz^2}{3}\right\} > 0\right).
\end{aligned} \tag{2.3.6}$$

Again, from the proof of Corollary 3.1 in Durot and Lopuhaä, 2014 it follows that W satisfies condition (A3) in Durot and Lopuhaä, 2014, which means that there exist $K_1, K_2 > 0$, such that the first probability on the right hand side of (2.3.6) is bounded by $K_1 \exp(-K_2 v^2 u^3)$. We establish the same bound for the remaining two probabilities. By the time reversal of the Brownian

motion, the process $W'(z) = W(x) - W(x-z)$ is also a Brownian motion on the interval $[u, x]$. Then, using the change of variable $u/z = t$ and the fact that $\widetilde{W}(t) = tu^{-1/2}W'(u/t)$, for $t > 0$, is again a Brownian motion, the second probability on the right hand side of (2.3.6) is bounded by

$$\mathbb{P}\left(\sup_{t\in(0,1]} \frac{t}{\sqrt{u}}\left|W'\left(\frac{u}{t}\right)\right| > \frac{\nu u^{3/2}}{3t}\right) = \mathbb{P}\left(\sup_{t\in[0,1]} |\widetilde{W}(t)| > \frac{\nu u^{3/2}}{3}\right). \quad (2.3.7)$$

Finally,

$$\sup_{u\leqslant z\leqslant x}\left|\frac{F(L^{-1}(x)) - F(L^{-1}(x-z))}{z}\right| \leqslant \frac{\sup_{u\in[0,\tau^*]} f(u)}{\inf_{u\in[0,\tau^*]} \lambda(u)},$$

so that the third probability on the right hand side of (2.3.6) is bounded by

$$\begin{aligned}\mathbb{P}&\left(\sup_{u\leqslant z\leqslant x}\left|\frac{F(L^{-1}(x)) - F(L^{-1}(x-z))}{z}\right| |W(x)| > \frac{\nu u}{3}\right)\\ &\leqslant \mathbb{P}\left(|W(x)| > \frac{\nu u^{3/2}}{K_3}\right) \leqslant \mathbb{P}\left(\sup_{t\in[0,L(\tau^*)]} |W(t)| > \frac{\nu u^{3/2}}{K_3}\right),\end{aligned} \quad (2.3.8)$$

for some $K_3 > 0$. By applying the maximal inequality for Brownian motion to the right hand sides of (2.3.7) and (2.3.8), we conclude that there exist $K_1', K_2' > 0$, such that

$$\mathbb{P}\left(\sup_{u\leqslant z\leqslant x} \{B(x-z) - B(x) - \nu z^2\} > 0\right) \leqslant K_1' \exp(-K_2'\nu^2 u^3),$$

which proves condition (A3) in Durot and Lopuhaä, 2014. □

Theorem 2.3.4. *Let f be a nondecreasing and twice continuously differentiable density such that f and f' are strictly positive. Let k satisfy (1.2.1) and suppose that it is differentiable with a uniformly bounded derivative. If $bn^{1/5} \to c \in (0, \infty)$, then for each $t \in (0, \tau^*)$, it holds*

$$n^{2/5}\left(\tilde{f}_n^{SG}(t) - f(t)\right) \xrightarrow{d} N(\mu, \sigma^2),$$

where

$$\mu = \frac{1}{2}c^2 f''(t)\int u^2 k(u)\,du \quad \textit{and} \quad \sigma^2 = \frac{f(t)}{c\,(1-G(t))}\int k^2(u)\,du. \quad (2.3.9)$$

For a fixed $t \in (0, \tau^*)$, the asymptotically MSE-optimal bandwidth b for $\tilde{\lambda}^{SG}$ is given by $c_{opt}(x)n^{-1/5}$, where

$$c_{opt} = \left\{f(t)\int k^2(u)\,du\right\}^{1/5}\left\{(1-G(t))f''(t)^2\left(\int u^2 k(u)\,du\right)^2\right\}^{-1/5}. \quad (2.3.10)$$

Proof. Fix $t \in (0, \tau^*)$. Then, for sufficiently large n, we have $0 < t - b < t + b \leqslant \tau^*$. Following the proof of 2.2.2, we write

$$\begin{aligned} \tilde{f}_n^{SG}(t) = \int k_b(t-u)\, dF(u) + \int k_b(t-u)\, d(F_n - F)(u) \\ + \int k_b(t-u)\, d(\hat{F}_n - F_n)(u). \end{aligned} \tag{2.3.11}$$

Again the first (deterministic) term on the right hand side of (2.3.11) gives us the asymptotic bias:

$$n^{2/5} \left\{ \int_{t-b}^{t+b} k_b(t-u) f(u)\, du - f(t) \right\} \to \frac{1}{2} c^2 f''(t) \int_{-1}^{1} y^2 k(y)\, dy,$$

and by the Kiefer-Wolfowitz type of result in Lemma 2.3.3, the last term on the right hand side of (2.3.11) converges to 0 in probability. What remains is to prove that

$$n^{2/5} \int k_b(t-u)\, d(F_n - F)(u) \xrightarrow{d} N(0, \sigma^2),$$

where σ^2 is defined in (2.3.9). We write

$$n^{2/5} \int_{t-b}^{t+b} k_b(t-u)\, d(F_n - F)(u) = \frac{1}{\sqrt{b\, n^{1/5}}} \int_{-1}^{1} k(y)\, d\hat{W}_n(y),$$

where, for each $y \in [-1, 1]$, we define

$$\begin{aligned} \hat{W}_n(y) &= \sqrt{\frac{n}{b}} \Big\{ F_n(t-by) - F_n(t) - F(t-by) + F(t) \Big\} \\ &= \sqrt{\frac{n}{b}} \Big\{ F_n(t-by) - F(t-by) - n^{-1/2}(1 - F(t-by)) W \circ L(t-by) \Big\} \\ &\quad - \sqrt{\frac{n}{b}} \Big\{ F_n(t) - F(t) - n^{-1/2}(1 - F(t)) W \circ L(t) \Big\} \\ &\quad + \frac{1}{\sqrt{b}} \Big(1 - F(t) \Big) \Big\{ W \circ L(t-by) - W \circ L(t) \Big\} \\ &\quad + \frac{1}{\sqrt{b}} \Big(F(t) - F(t-by) \Big) W \circ L(t-by). \end{aligned} \tag{2.3.12}$$

Using the strong approximation (2.1.9), we obtain

$$\begin{aligned} &\mathbb{P} \left(\sup_{u \in [0, \tau^*]} \sqrt{\frac{n}{b}} \Big| F_n(u) - F(u) - n^{-1/2}(1 - F(u)) W \circ L(u) \Big| > \epsilon \right) \\ &\qquad \leqslant K_1 \exp \Big\{ -K_2 (\epsilon \sqrt{nb} - K_3 \log n) \Big\} \to 0, \end{aligned}$$

and it then follows that the first two terms on the right hand side of (2.3.12) converge to zero in probability uniformly in y. For the last term, we get

$$\mathbb{P}\left(\sup_{y\in[-1,1]}\left|\frac{1}{\sqrt{b}}\Big(F(t)-F(t-by)\Big)W\circ L(t-by)\right|>\epsilon\right)$$
$$\leqslant \mathbb{P}\left(\sqrt{b}\sup_{u\in[0,\tau^*]}f(u)\sup_{u\in[0,\|L\|_\infty]}|W(u)|>\epsilon\right)\leqslant K_1\exp\left(-\frac{K_2\epsilon^2}{b}\right)\to 0.$$

For the third term on the right hand side of (2.3.12), note that

$$y\mapsto b^{-1/2}(W\circ L(x-by)-W\circ L(x)),\qquad y\in[-1,1],$$

has the same distribution as the process

$$y\mapsto \widetilde{W}\left(\frac{L(t)-L(t-by)}{b}\right),\quad \text{for } y\in[-1,1], \tag{2.3.13}$$

where $\widetilde{W}$ is a two-sided Brownian motion. By uniform continuity of the two-sided Brownian motion on compact intervals, the sequence of stochastic processes in (2.3.13) converges to the process $\{\widetilde{W}(L'(t)y):y\in[-1,1]\}$:

$$\sup_{y\in[-1,1]}\left|\widetilde{W}\left(\frac{L(t)-L(t-by)}{b}\right)-\widetilde{W}\left(L'(t)y\right)\right|\xrightarrow{\mathbb{P}}0.$$

As a result

$$\frac{1}{\sqrt{bn^{1/5}}}\int_{-1}^{1}k(y)\,d\hat{W}_n(y)\xrightarrow{d}\sqrt{\frac{f(t)}{c(1-G(t))}}\int_{-1}^{1}k(y)\,d\widetilde{W}(y)$$
$$\sim N\left(0,\frac{f(t)}{c(1-G(t))}\int_{-1}^{1}k^2(y)\,dy\right).$$

The optimal c is then obtained by minimizing the asymptotic mean squared error

$$\mathrm{AMSE}(\tilde{f}^{SG},c)=\frac{1}{4}c^4f''(t)^2\left(\int u^2\,k(u)\,du\right)^2+\frac{f(t)}{c(1-G(t))}\int k^2(u)\,du,$$

with respect to c. □

2.4 POINTWISE CONFIDENCE INTERVALS

In this section we construct pointwise confidence intervals for the hazard rate and the density based on the asymptotic distributions derived in Theorem 2.2.2 and Theorem 2.3.4 and compare them to confidence intervals

constructed using Grenander-type estimators without smoothing. According to Theorem 2.1 and Theorem 2.2 in Huang and Wellner, 1995, for a fixed $t_0 \in (0, \tau^*)$,

$$n^{1/3} \left| \frac{1 - G(t_0)}{4f(t_0)f'(t_0)} \right|^{1/3} \left(\tilde{f}_n(t_0) - f(t_0) \right) \xrightarrow{d} \mathbb{Z},$$

and

$$n^{1/3} \left| \frac{1 - H(t_0)}{4\lambda(t_0)\lambda'(t_0)} \right|^{1/3} \left(\tilde{\lambda}_n(t_0) - \lambda(t_0) \right) \xrightarrow{d} \mathbb{Z},$$

where W is a two-sided Brownian motion starting from zero and $\mathbb{Z} = \operatorname{argmin}_{t \in \mathbb{R}} \left\{ W(t) + t^2 \right\}$. This yields $100(1-\alpha)\%$-confidence intervals for $f(t_0)$ and $\lambda(t_0)$ of the following form

$$C^1_{n,\alpha} = \tilde{f}_n(t_0) \pm n^{-1/3} \hat{c}_{n,1}(t_0) q(\mathbb{Z}, 1 - \alpha/2),$$

and

$$C^2_{n,\alpha} = \tilde{\lambda}_n(t_0) \pm n^{-1/3} \hat{c}_{n,2}(t_0) q(\mathbb{Z}, 1 - \alpha/2),$$

where $q(\mathbb{Z}, 1 - \alpha/2)$ is the $(1 - \alpha/2)$ quantile of the distribution $\mathbb{Z}$ and

$$\hat{c}_{n,1}(t_0) = \left| \frac{4\tilde{f}_n(t_0)\tilde{f}'_n(t_0)}{1 - G_n(t_0)} \right|^{1/3}, \qquad \hat{c}_{n,2}(t_0) = \left| \frac{4\tilde{\lambda}_n(t_0)\tilde{\lambda}'_n(t_0)}{1 - H_n(t_0)} \right|^{1/3}.$$

Here, H_n is the empirical distribution function of T and in order to avoid the denominator taking the value zero, instead of the natural estimator of G, we consider a slightly different version as in Marron and Padgett, 1987:

$$G_n(t) = \begin{cases} 0 & \text{if } 0 \leqslant t < T_{(1)}, \\ 1 - \prod_{i=1}^{k-1} \left(\frac{n-i+1}{n-i+2} \right)^{1-\Delta_i} & \text{if } T_{(k-1)} \leqslant t < T_{(k)}, \quad k = 2, \ldots, n, \\ 1 - \prod_{i=1}^{n} \left(\frac{n-i+1}{n-i+2} \right)^{1-\Delta_i} & \text{if } t \geqslant T_{(n)}. \end{cases} \tag{2.4.1}$$

Furthermore, as an estimate for $\tilde{f}'_n(t_0)$ we choose

$$\frac{\tilde{f}_n(\tau_m) - \tilde{f}_n(\tau_{m-1})}{\tau_m - \tau_{m-1}},$$

where τ_{m-1} and τ_m are two succeeding points of jump of $\tilde{f}_n$ such that $t_0 \in (\tau_{m-1}, \tau_m]$, and $\tilde{\lambda}'_n(t_0)$ is estimated similarly. The quantiles of the distribution $\mathbb{Z}$ have been computed in Groeneboom and Wellner, 2001 and we will use $q(\mathbb{Z}, 0.975) = 0.998181$.

The pointwise confidence intervals based on the smoothed Grenander-type estimators are constructed from Theorem 2.2.2 and Theorem 2.3.4. We find

$$\widetilde{C}^1_{n,\alpha} = \tilde{f}^{SG}_n(t_0) \pm n^{-2/5}\left(\hat{\sigma}_{n,1}(t_0)q_{1-\alpha/2} + \hat{\mu}_{n,1}(t_0)\right),$$

and

$$\widetilde{C}^2_{n,\alpha} = \tilde{\lambda}^{SG}_n(t_0) \pm n^{-2/5}(\hat{\sigma}_{n,2}(t_0)q_{1-\alpha/2} + \hat{\mu}_{n,2}(t_0)),$$

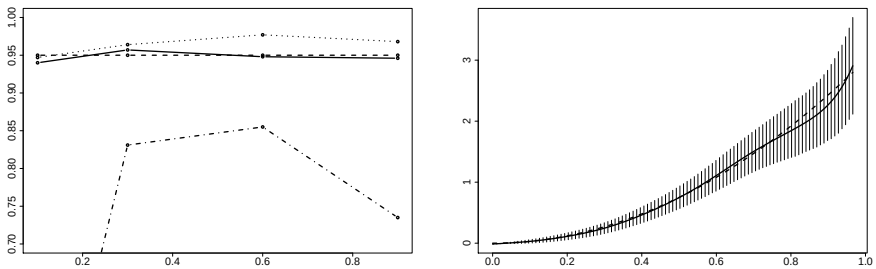
where $q_{1-\alpha/2}$ is the $(1-\alpha/2)$ quantile of the standard normal distribution. The estimators $\hat{\sigma}_{n,1}(t_0)$ and $\hat{\mu}_{n,1}(t_0)$ are obtained by plugging $\tilde{f}^{SG}_n$ and its second derivative for f and f'', respectively, and G_n and $c_{opt}(t_0)$ for G and c, respectively, in (2.3.9), and similarly $\hat{\sigma}_{n,2}(t_0)$ and $\hat{\mu}_{n,2}(t_0)$ are obtained from (2.2.4). Estimating the bias seems to be a hard problem because it depends on the second derivative of the function of interest. As discussed, for example in Hall, 1992, one can estimate the bias by using a bandwidth of a different order for estimating the second derivative or one can use undersmoothing (in that case the bias is zero and we do not need to estimate the second derivative). We tried both methods and it seems that undersmoothing performs better, which is in line with other results available in the literature (see for instance, Hall, 1992; Groeneboom and Jongbloed, 2015; Cheng, Hall, and Tu, 2006).

When estimating the hazard rate, we choose a Weibull distribution with shape parameter 3 and scale parameter 1 for the event times and the uniform distribution on $(0, 1.3)$ for the censoring times. Confidence intervals are calculated at the point $t_0 = 0.5$ with $\tau^* = H^{-1}(0.9)$, using 1000 sets of data and the bandwidth in the case of undersmoothing is $b = c_{opt}(t_0)n^{-1/4}$, where $c_{opt}(t_0) = 1.2$. In case of bias estimation we use $b = c_{opt}(t_0)n^{-5/17}$ to estimate the hazard and $b_1 = c_{opt}(t_0)n^{-1/17}$ to estimate its second derivative (as suggested in Hall, 1992). Table 1 shows the performance, for various sample sizes, of the confidence intervals based on the asymptotic distribution (AD) of the Grenander-type estimator and of the smoothed Grenander estimator (for both undersmoothing and bias estimation). The poor performance of the Grenander-type estimator seems to be related to the crude estimate of the derivative of the hazard with the slope of the correspondent segment. On the other hand, it is obvious that smoothing leads to significantly better results in terms of both average length and coverage probabilities. As expected, when using undersmoothing, as the sample size increases we get shorter confidence intervals and coverage probabilities that are closer to the nominal level of 95%. By estimating the bias, we obtain coverage probabilities that are higher than 95%, because the confidence intervals are bigger compared to the average length when using undersmoothing. Another way to compare the performance of the different methods is to take a fixed sample size $n = 500$ and different points of the support of the hazard

	Grenander		SG-undersmoothing		SG-bias estimation	
n	AL	CP	AL	CP	AL	CP
100	0.930	0.840	0.648	0.912	0.689	0.955
500	0.560	0.848	0.366	0.948	0.383	0.975
1000	0.447	0.847	0.283	0.954	0.295	0.977
5000	0.255	0.841	0.155	0.953	0.157	0.978

Table 1: The average length (AL) and the coverage probabilities (CP) for 95% pointwise confidence intervals of the hazard rate at the point $t_0 = 0.5$.

function. Figure 5 shows that confidence intervals based on undersmooth-



(a) (b)

Figure 5: Left panel: Actual coverage of 95% confidence intervals for the hazard rate. Dashed line-nominal level; dotdashed line-Grenander-type; solid line-SG undersmoothing; dotted line-SG bias estimation. Right panel: 95% confidence intervals for the hazard rate using undersmoothing.

ing behave well also at the boundaries in terms of coverage probabilities, but the length increases as we move to the left end point of the support. In order to maintain good visibility of the performance of the smooth estimators, we left out the poor performance of the Grenander estimator at point $x = 0.1$.

It is also of interest to compare our confidence intervals with other competing methods, in particular with those obtained via the inversion of the likelihood ratio statistic proposed in Banerjee, 2008. We consider their setting of simulation A.2 (heavy censoring case), where the event times come from a Weibull distribution with shape parameter 2 and scale $\sqrt{2}$ and the censoring times are uniform in $(0, 1.5)$. Again, we recorded the average

length of nominal 95% confidence intervals for the hazard rate at the point $t_0 = \sqrt{2\log 2}$ obtained using undersmoothing and their coverage probabilities. The results for various sample sizes and choices of the constant c in the definition of the bandwidth are displayed in Table 2. These observations show that the performance of the confidence intervals strongly depends on the choice of c. Most of the time the confidence intervals are shorter compared to those using likelihood ratio and for $c = 1$ or $c = 1.2$ our method produces also better coverage probabilities. On the other hand, if c is too small, e.g. $c = 0.8$, the confidence intervals become quite conservative (or anticonservative for large c). Of course the likelihood ratio method has the advantage of not requiring estimation of nuisance parameters or choosing the bandwidth but our results confirm that with the right choice of the bandwidth smooth estimation performs much better. The importance of the choice of the smoothing parameter is well-known and different methods has been proposed in the literature to find the optimal one (see for example Cheng, Hall, and Tu, 2006 and González-Manteiga, Cao, and Marron, 1996). However, it is beyond the scope of this paper to investigate methods of bandwidth selection.

	LR		$c = 0.8$		$c = 1$		$c = 1.2$	
n	AL	CP	AL	CP	AL	CP	AL	CP
50	3.110	0.911	2.951	0.962	2.700	0.946	2.463	0.925
100	2.408	0.917	2.213	0.964	1.991	0.951	1.820	0.935
200	1.684	0.929	1.680	0.972	1.512	0.947	1.392	0.930
500	1.073	0.932	1.202	0.970	1.070	0.952	0.981	0.927
1000	0.782	0.936	0.935	0.975	0.836	0.958	0.764	0.936
1500	0.653	0.941	0.809	0.982	0.720	0.965	0.663	0.944

Table 2: The average length (AL) and the coverage probabilities (CP) for 95% pointwise confidence intervals of the hazard rate at $t_0 = \sqrt{2\log 2}$ using likelihood ratio (LR) and undersmoothing with various choices of c .

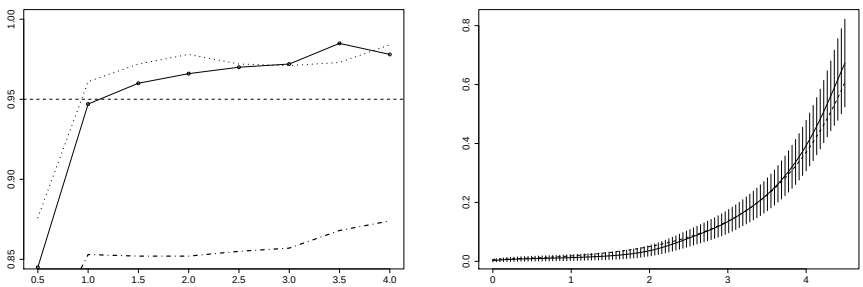
Finally, we consider estimation of the density. We simulate the event times and the censoring times from the density functions in (2.3.3) and (2.3.4). Confidence intervals are calculated at the point $t_0 = 2.5$ with $\tau^* = H^{-1}(0.9)$, using 1000 sets of data and the bandwidth in the case of undersmoothing is $b = c_{opt}(2.5)n^{-1/4}$, where $c_{opt}(2.5) = 4$. When estimating the bias we use $b = c_{opt}(2.5)n^{-5/17}$ to estimate the hazard and $b_1 = c_{opt}(2.5)n^{-1/17}$ to estimate its second derivative (as suggested in Hall, 1992). Table 3 shows the performance, for various sample sizes, of the confidence intervals based

on the asymptotic distribution (AD) of the Grenander-type estimator and of the smoothed Grenander estimator (for both undersmoothing and bias estimation).

	Grenander		SG-undersmoothing		SG-bias estimation	
n	AL	CP	AL	CP	AL	CP
50	0.157	0.822	0.129	0.948	0.136	0.929
100	0.127	0.856	0.101	0.954	0.109	0.959
500	0.073	0.859	0.056	0.971	0.064	0.979
1000	0.058	0.864	0.043	0.979	0.050	0.976
5000	0.032	0.845	0.023	0.965	0.029	0.965

Table 3: The average length (AL) and the coverage probabilities (CP) for 95% pointwise confidence intervals of the density function at the point $t_0 = 2.5$.

Confidence intervals based on the Grenander-type estimator have a poor coverage, while, by considering the smoothed version, we usually obtain high coverage probabilities. Again, undersmoothing behaves slightly better. The performance of these three methods for a fixed sample size $n = 500$ and different points of the support is illustrated in Figure 6.



(a) (b)

Figure 6: Left panel: Actual coverage of 95% confidence intervals for the density. Dashed line-nominal level; dotdashed line-Grenander-type; solid line-SG undersmoothing; dotted line-SG bias estimation. Right panel: 95% confidence intervals for the density using undersmoothing.

3

SMOOTH ISOTONIC ESTIMATION IN THE COX MODEL

In this chapter we consider smooth estimation under monotonicity constraints of the baseline hazard rate in the Cox regression model. The results presented are based on:

Lopuhaä, H. P. and Musta E. (2018) "Smoothed isotonic estimators of a monotone baseline hazard in the Cox model". *Scandinavian Journal of Statistics 45.3*, pp. 753-791.

Lopuhaä, H. P. and Musta E. (2017) "Isotonized smooth estimators of a monotone baseline hazard in the Cox model". *Journal of Statistical Planning and Inference 191*, pp. 43-67.

Nonparametric estimation under monotonicity constraints of the baseline hazard function in the Cox regression model (see Example 1.1.2) has been studied in Chung and Chang, 1994 and Lopuhaä and Nane, 2013 while smooth isotonic estimators were introduced in Nane, 2013. By combining an isotonization step with a smoothing step and alternating the order of smoothing and isotonization, four different estimators can be constructed.

Two of them are kernel smoothed versions of the maximum likelihood estimator and the Grenander-type estimator from Lopuhaä and Nane, 2013. The third estimator is a maximum smoothed likelihood estimator obtained by first smoothing the loglikelihood of the Cox model and then finding the maximizer of the smoothed likelihood among all decreasing baseline hazards. By first smoothing the loglikelihood, one avoids the discrete behavior of the traditional MLE. This approach is similar to the methods in Eggermont and LaRiccia, 2000 for monotone densities and in Groeneboom, Jongbloed, and Witte, 2010 for the current status model. The forth estimator we consider is isotonized kernel estimator. It is a Grenander-type estimator based on the smooth Breslow estimator for the cumulative hazard. Grenander-type estimators for a nondecreasing curve are obtained as the left-derivative of the greatest convex minorant of a naive nonparametric estimator for the integrated curve of interest, see Grenander, 1956 and also Durot, 2007 among others. For our setup, the smoothed Breslow estimator serves as an estimator for the cumulative baseline hazard. By smoothing

the Breslow estimator, one avoids the discrete behavior of the left-derivative of its least concave majorant. This approach is similar to the methods considered in Cheng and Lin, 1981, Wright, 1982, Friedman and Tibshirani, 1984, and van der Vaart and van der Laan, 2003, and to one of the two methods studied in Mammen, 1991.

We establish asymptotic normality at rate $n^{m/(2m+1)}$, where m denotes the level of smoothness of the baseline hazard, for the four estimators. The isotonized kernel estimator (GS) is shown to be asymptotically equivalent to the smoothed Grenander-type estimator (SG) and the smoothed maximum likelihood estimator (SMLE). In particular, this means that the order of smoothing and isotonization is irrelevant, which is in line with the findings in Mammen, 1991. On the other hand, the maximum smoothed likelihood estimator (MSLE) exhibits the same limit variance as the previous ones but has a different asymptotic bias, a phenomenon that was also encountered in Groeneboom, Jongbloed, and Witte, 2010. As a result, from the theoretical point of view, there is no reason to prefer one estimator with respect to the other (apart from the fact that the kernel smoothed estimators are differentiable while the other two are usually only continuous).

Deriving asymptotic normality of the two kernel smoothed isotonic estimators is particularly challenging for the Cox model, because the existing approaches to these type of problems do not apply to the Cox model. The smoothed Grenander-type estimator in the ordinary right censoring model without covariates was investigated in Chapter 2. Following the approach in Groeneboom and Jongbloed, 2013, asymptotic normality was established by using a Kiefer-Wolfowitz type of result, recently derived in Durot and Lopuhaä, 2014. Unfortunately, the lack of a Kiefer-Wolfowitz type of result for the Breslow estimator provides a strong limitation towards extending the previous approach to the more general setting of the Cox model. Recently, Groeneboom and Jongbloed, 2014 developed a different method for finding the limit distribution of smoothed isotonic estimators, which is mainly based on uniform L_2-bounds on the distance between the non-smoothed isotonic estimator and the true function, and also uses that the maximal distance between succeeding points of jump of the isotonic estimator is of the order $O_p(n^{-1/3}\log n)$. A sketch of proof in the right censoring model is given in Section 11.6 of Groeneboom and Jongbloed, 2014. However, these two key ingredients heavily depend on having exponential bounds for tail probabilities of the so-called inverse process, or rely on a strong embedding for the relevant sum process. Exponential bounds for tail probabilities of the inverse process are difficult to obtain in the Cox model and a strong embedding for the Breslow estimator is not available. Nevertheless, inspired by

the approach in Groeneboom and Jongbloed, 2014, we obtain polynomial bounds, which are suboptimal but sufficient for our purposes.

The method used to establish asymptotic normality for isotonized smooth estimators is quite different from the previous one because the isotonization step was performed after a smoothing step. Here we rely on techniques developed in Groeneboom, Jongbloed, and Witte, 2010 and the key idea is that the isotonized smooth estimator can be represented as a least squares projection of a naive smooth estimator. The latter estimator is not monotone, but much simpler to analyze and it is shown to be asymptotically equivalent to the smooth isotonic estimator. As a consequence, the resulting estimators are asymptotically equivalent to corresponding naive estimators that are combinations of ordinary kernel type estimators, to which standard techniques apply.

Furthermore, we also investigated the finite sample performance of these estimators by constructing pointwise confidence intervals. First, making use of the theoretical results, we construct pointwise confidence intervals based on the limit distribution with undersmoothing to avoid bias estimation. Results confirm the comparable behavior of the four methods and favor the use of the smoothed isotonic estimators instead of the unsmoothed Grenander-type estimator or the kernel estimator. However, coverage probabilities are far from the nominal level and strongly depend on the choice of the bandwidth and the accuracy in the estimation of the regression coefficient β_0. Since most of the well-known methods to overcome these problems do not seem to work in our setting, a thorough investigation is still needed for improving the performance of the confidence intervals. Instead, we choose to exploit pointwise confidence intervals based on smooth bootstrap procedures as proposed by Burr, 1994 and Xu, Sen, and Ying, 2014. It turns out, the simple percentile bootstrap works better than the studentized one. Such a phenomenon was also observed in Burr, 1994. The four estimators again exhibit comparable behavior but the smoothed maximum likelihood estimator and the maximum smoothed likelihood estimator have slightly better coverage probabilities. The performance is satisfactory, but still further investigation is required for bandwidth selection and correcting the asymptotic bias, which might improve the results.

The chapter is organized as follows. In Section 3.1 we specify the Cox regression model and provide some background information that will be used in the sequel. The kernel smoothed versions of the Grenander-type estimator and of the maximum likelihood estimator of a non-decreasing baseline hazard function are considered in Section 3.2.We only consider the case of a non-decreasing baseline hazard. The same results can be obtained similarly for a non-increasing hazard. The maximum smoothed likelihood

estimator is considered in Section 3.3 and the isotonized kernel estimator in Section 3.4. The results of a simulation study on pointwise confidence intervals are reported in Section 3.5. In order to keep the exposition clear and simple, most of the proofs are delayed until Section 3.6, and remaining technicalities have been put in the Supplementary Material A.1.

3.1 THE COX REGRESSION MODEL

Let $X_1, \ldots, X_n$ be an i.i.d. sample representing the survival times of n individuals, which can be observed only on time intervals $[0, C_i]$ for some i.i.d. censoring times $C_1, \ldots, C_n$. One observes i.i.d. triplets $(T_1, \Delta_1, Z_1), \ldots, (T_n, \Delta_n, Z_n)$, where $T_i = \min(X_i, C_i)$ denotes the follow up time, $\Delta_i = \mathbb{1}_{\{X_i \leqslant C_i\}}$ is the censoring indicator and $Z_i \in \mathbb{R}^p$ is a time independent covariate vector. Given the covariate vector Z, the event time X and the censoring time C are assumed to be independent. Furthermore, conditionally on $Z = z$, the event time is assumed to be a nonnegative r.v. with an absolutely continuous distribution function $F(x|z)$ and density $f(x|z)$. Similarly the censoring time is assumed to be a nonnegative r.v. with an absolutely continuous distribution function $G(x|z)$ and density $g(x|z)$. The censoring mechanism is assumed to be non-informative, i.e. F and G share no parameters. Within the Cox model, the conditional hazard rate $\lambda(x|z)$ for a subject with covariate vector $z \in \mathbb{R}^p$, is related to the corresponding covariate by

$$\lambda(t|z) = \lambda_0(t)\, e^{\beta_0' z}, \quad t \in \mathbb{R}^+,$$

where λ_0 represents the baseline hazard function, corresponding to a subject with $z = 0$, and $\beta_0 \in \mathbb{R}^p$ is the vector of the regression coefficients.

Let H and H^{uc} denote respectively the distribution function of the follow-up time and the sub-distribution function of the uncensored observations, i.e.,

$$H^{uc}(t) = \mathbb{P}(T \leqslant t, \Delta = 1) = \int \delta \mathbb{1}_{\{u \leqslant t\}}\, d\mathbb{P}(u, \delta, z), \tag{3.1.1}$$

where $\mathbb{P}$ is the distribution of (T, Δ, Z). We also require the following assumptions, some of which are common in large sample studies of the Cox model (e.g. see Lopuhaä and Nane, 2013):

(A1) Let τ_F, τ_G and τ_H be the end points of the support of F, G and H. Then

$$\tau_H = \tau_G < \tau_F \leqslant \infty.$$

(A2) There exists $\epsilon > 0$ such that

$$\sup_{|\beta-\beta_0|\leqslant\epsilon} \mathbb{E}\left[|Z|^2 e^{2\beta' Z}\right] < \infty.$$

(A3) There exists $\epsilon > 0$ such that

$$\sup_{|\beta-\beta_0|\leqslant\epsilon} \mathbb{E}\left[|Z|^2 e^{4\beta' Z}\right] < \infty.$$

Let us briefly comment on these assumptions. While the first one tells us that, at the end of the study, there is at least one subject alive, the other two are somewhat hard to justify from a practical point of view. One can think of (A2) and (A3) as conditions on the boundedness of the second moment of the covariates, uniformly for β in a neighborhood of β_0.

By now, it seems to be rather a standard choice estimating β_0 by $\hat{\beta}_n$, the maximizer of the partial likelihood function in (1.1.6), as proposed by Cox, 1972. The asymptotic behavior was first studied by Tsiatis, 1981. We aim at estimating λ_0, subject to the constraint that it is increasing (the case of a decreasing hazard is analogous), on the basis of n observations $(T_1, \Delta_1, Z_1), \ldots, (T_n, \Delta_n, Z_n)$. By introducing

$$\Phi(t;\beta) = \int \mathbb{1}_{\{u\geqslant t\}} e^{\beta' z} \, d\mathbb{P}(u, \delta, z), \tag{3.1.2}$$

we have

$$\lambda_0(t) = \frac{h(t)}{\Phi(t;\beta_0)}, \tag{3.1.3}$$

where $h(t) = dH^{uc}(t)/dt$ (e.g., see (9) in Lopuhaä and Nane, 2013). For $\beta \in \mathbb{R}^p$ and $x \in \mathbb{R}$, the function $\Phi(t;\beta)$ can be estimated by

$$\Phi_n(t;\beta) = \int \mathbb{1}_{\{u\geqslant t\}} e^{\beta' z} \, d\mathbb{P}_n(u, \delta, z), \tag{3.1.4}$$

where $\mathbb{P}_n$ is the empirical measure of the triplets (T_i, Δ_i, Z_i) with $i = 1, \ldots, n$. Moreover, in Lemma 4 of Lopuhaä and Nane, 2013 it is shown that

$$\sup_{t\in\mathbb{R}} |\Phi_n(t;\beta_0) - \Phi(t;\beta_0)| = O_p(n^{-1/2}). \tag{3.1.5}$$

It will be often used throughout the paper that a stochastic bound of the same order holds also for the distance between the cumulative hazard Λ_0 and the Breslow estimator

$$\Lambda_n(t) = \int \frac{\delta \mathbb{1}_{\{u\leqslant t\}}}{\Phi_n(t;\hat{\beta}_n)} \, d\mathbb{P}_n(u, \delta, z), \tag{3.1.6}$$

but only on intervals staying away of the right boundary, i.e.,

$$\sup_{t\in[0,M]} |\Lambda_n(t) - \Lambda_0(t)| = O_p(n^{-1/2}), \qquad \text{for all } 0 < M < \tau_H, \tag{3.1.7}$$

(see Theorem 5 in Lopuhaä and Nane, 2013).

Smoothing is done by means of an m-orthogonal kernel function satisfying (1.2.1). Note that if $m > 2$, the kernel function k necessarily attains negative values and as a result also the smooth estimators of the baseline hazard defined in Sections 3.2 may be negative and monotonicity might not be preserved. To avoid this, one could restrict oneself to $m = 2$. In that case, the most common choice is to let k be a symmetric probability density.

3.2 SMOOTHED ISOTONIC ESTIMATORS

We consider smoothed versions of two isotonic estimators for λ_0, i.e, the maximum likelihood estimator $\hat{\lambda}_n$ and the Grenander-type estimator $\tilde{\lambda}_n$, introduced in Lopuhaä and Nane, 2013. The maximum likelihood estimator of a nondecreasing baseline hazard rate λ_0 is of the form

$$\hat{\lambda}_n(t) = \begin{cases} 0 & t < T_{(1)} \\ \hat{\lambda}_i & T_{(i)} \leqslant t < T_{(i+1)}, \quad \text{for } i = 1, \ldots, n-1, \\ \infty & t \geqslant T_{(n)}, \end{cases}$$

where $\hat{\lambda}_i$ is the left derivative at point P_i of the greatest convex minorant of the cumulative sum diagram consisting of points $P_0 = (0,0)$ and $P_j = \big(\hat{W}_n(T_{(j+1)}), V_n(T_{(j+1)})\big)$, for $j = 1, \ldots, n-1$, where $\hat{W}_n$ and V_n are defined as

$$\begin{aligned} \hat{W}_n(t) &= \int \left(e^{\hat{\beta}_n' z} \int_{T_{(1)}}^{t} \mathbb{1}_{\{u \geqslant s\}} \, ds \right) d\mathbb{P}_n(u, \delta, z), \quad t \geqslant T_{(1)}, \\ V_n(t) &= \int \delta \mathbb{1}_{\{u<t\}} \, d\mathbb{P}_n(u, \delta, z), \end{aligned} \tag{3.2.1}$$

with $\hat{\beta}_n$ being the partial maximum likelihood estimator (see Lemma 1 in Lopuhaä and Nane, 2013). For a fixed $t \in [0, \tau_H]$, the smoothed maximum likelihood estimator $\hat{\lambda}_n^{SM}$ of a nondecreasing baseline hazard rate λ_0, was defined in Nane, 2013 by

$$\hat{\lambda}_n^{SM}(t) = \int_{(t-b)\vee 0}^{(t+b)\wedge \tau_H} k_b(t-u) \, \hat{\lambda}_n(u) \, du. \tag{3.2.2}$$

The Grenander-type estimator $\tilde{\lambda}_n$ of a nondecreasing baseline hazard rate λ_0 is defined as the left hand slope of the greatest convex minorant (GCM) $\hat{\Lambda}_n$ of the Breslow estimator Λ_n. For a fixed $t_0 \in [0, \tau_H]$, we consider the smoothed Grenander-type estimator $\tilde{\lambda}_n^{SG}$, which is defined by

$$\tilde{\lambda}_n^{SG}(t) = \int_{(t-b)\vee 0}^{(t+b)\wedge \tau_H} k_b(t-u)\tilde{\lambda}_n(u)\,du. \tag{3.2.3}$$

Uniform strong consistency on compact intervals in the interior of the support $[\epsilon, M] \subset [0, \tau_H]$ is provided by Theorem 5.2 of Nane, 2013,

$$\sup_{t\in[\epsilon,M]} \left|\tilde{\lambda}_n^{SG}(t) - \lambda_0(t)\right| \to 0, \quad \text{with probability one.} \tag{3.2.4}$$

Strong pointwise consistency of $\hat{\lambda}_n^{SM}$ in the interior of the support is established in Theorem 5.1 in Nane, 2013. Under additional smoothness assumptions on λ_0, one can obtain uniform strong consistency for $\hat{\lambda}_n^{SM}$ similar to (3.2.4). Inconsistency at the boundaries is can be partially avoided by using a boundary corrected kernel. It can be proved, exactly as it is done in Chapter 2, that uniform consistency holds on $[0, M] \subset [0, \tau_H]$.

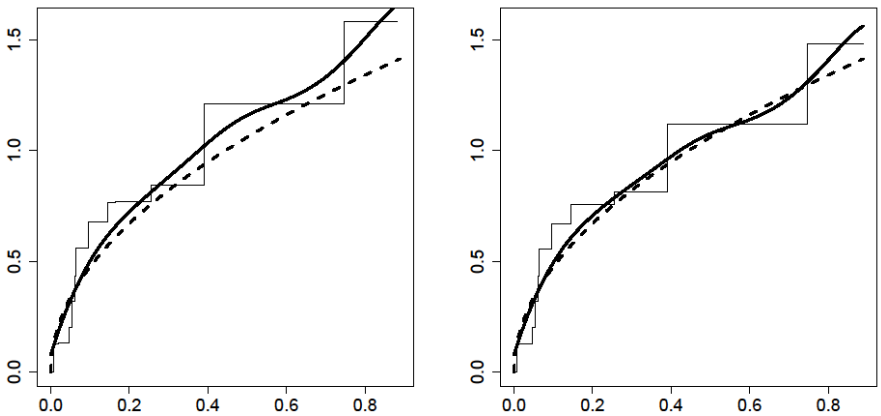


Figure 7: Left panel: The MLE (piecewise constant solid line) of the baseline hazard (dashed) together with the smoothed MLE (solid). Right panel: The Grenander estimator (piecewise constant solid line) of the baseline hazard (dashed) together with the smoothed Grenander estimator (solid).

Figure 7 shows the smoothed maximum likelihood estimator (left) and the smoothed Grenander-type estimator (right) for a sample of size $n = 500$ from a Weibull baseline distribution with shape parameter 1.5 and scale 1. For simplicity, we assume that the real valued covariate and the censoring

times are uniformly $(0,1)$ distributed and we take $\beta_0 = 0.5$. We used a boundary corrected triweight kernel function

$$k(u) = (35/32)(1-u^2)^3 \mathbb{1}_{\{|u|\leqslant 1\}}$$

and bandwidth $b = n^{-1/5}$. Note that, even though for deriving the asymptotic normality we do not need to restrict the estimation interval as in Chapter 2, for finite sample sizes restriction on $[0,\tau^*] \subset [0,\tau_H]$ gives better results. For the figures we used as τ^* the 95% empirical quantile of the follow-up times.

In the remainder of this section we will derive the pointwise asymptotic distribution of both smoothed isotonic estimators, in (3.2.2) and (3.2.3). As already mentioned, our approach is inspired by techniques introduced in Section 11.6 of Groeneboom and Jongbloed, 2014. We briefly describe this approach for the smoothed Grenander estimator, for which the computations are more complicated. We start by writing

$$\tilde{\lambda}_n^{SG}(t) = \int k_b(t-u)\, d\Lambda_0(u) + \int k_b(t-u)\, d(\hat{\Lambda}_n - \Lambda_0)(u). \tag{3.2.5}$$

The first (deterministic) term on the right hand side of (3.2.5) gives us the asymptotic bias. The method applied in Chapter 2 for the right censoring model continues by decomposing the second term in two parts

$$\int k_b(t-u)\, d(\hat{\Lambda}_n - \Lambda_n)(u) + \int k_b(t-u)\, d(\Lambda_n - \Lambda_0)(u),$$

and then uses the Kiefer-Wolfowitz type of result

$$\sup_{t\in[0,M]} |\hat{\Lambda}_n(t) - \Lambda_n(t)| = O_P\left(n^{-2/3}(\log n)^{2/3}\right), \tag{3.2.6}$$

to show that $\int k_b(t-u)\, d(\hat{\Lambda}_n - \Lambda_n)(u)$ converges to zero. Finally, results from empirical process theory are used to show the asymptotic normality of $\int k_b(t-u)\, d(\Lambda_n - \Lambda_0)(u)$. This approach cannot be followed in our case because of the lack of a Kiefer-Wolfowitz type of result as in (3.2.6) for the Cox model.

Alternatively, we proceed by describing the main steps of the L_2-bounds approach introduced in Groeneboom and Jongbloed, 2014. On an event E_n with probability tending to one, we will approximate

$$\int k_b(t-u)\, d(\hat{\Lambda}_n - \Lambda_0)(u) \tag{3.2.7}$$

by $\int \theta_{n,t}(u,\delta,z)\, d\mathbb{P}(u,\delta,z)$, for some suitable function $\theta_{n,t}$ (see Lemma 3.2.1), whose piecewise constant modification $\overline{\theta}_{n,t}$ integrates to zero with respect

to the empirical measure $\mathbb{P}_n$ (see Lemma 3.2.2). This enables us to approximate (3.2.7) by

$$\int \overline{\theta}_{n,t}(u,\delta,z)\,d(\mathbb{P}_n-\mathbb{P})(u,\delta,z)+\int\left(\overline{\theta}_{n,t}(u,\delta,z)-\theta_{n,t}(u,\delta,z)\right)\,d\mathbb{P}(u,\delta,z). \tag{3.2.8}$$

Then, the key step is to bound the second integral in (3.2.8) by means of L_2-bounds on the distance between the ordinary Grenander estimator and the true baseline hazard (see Lemma 3.2.3). The last step consists of replacing $\overline{\theta}_{n,t}$ by a deterministic function $\eta_{n,t}$ (see Lemma 3.2.4) and use empirical process theory to show that

$$\int \eta_{n,t}(u,\delta,z)\,d(\mathbb{P}_n-\mathbb{P})(u,\delta,z)$$

is asymptotically normal.

Before we proceed to our first main result, we will formulate the steps described above in a series of lemmas. Let $t\in(0,\tau_H)$, define

$$a_{n,t}(u)=\frac{k_b(t-u)}{\Phi(u;\beta_0)},\qquad u\in(0,\tau_H), \tag{3.2.9}$$

where $\Phi(u;\beta_0)$ is defined in (3.1.2). Note that $a_{n,t}(u)=0$ for $u\notin(t-b,t+b)$. We then have the following approximation for (3.2.7). The proof can be found in Section 3.6.

Lemma 3.2.1. *Suppose that (A1)–(A2) hold. Let $a_{n,t}$ be defined by* (3.2.9) *and let $\hat{\beta}_n$ be the partial MLE for β_0. There exists an event E_n, with $\mathbb{1}_{E_n}\to 1$ in probability, such that for*

$$\theta_{n,t}(u,\delta,z)=\mathbb{1}_{E_n}\left\{\delta\,a_{n,t}(u)-e^{\hat{\beta}_n' z}\int_0^u a_{n,t}(v)\,d\hat{\Lambda}_n(v)\right\}, \tag{3.2.10}$$

it holds

$$\int \theta_{n,t}(u,\delta,z)\,d\mathbb{P}(u,\delta,z)=-\mathbb{1}_{E_n}\int k_b(t-u)\,d(\hat{\Lambda}_n-\Lambda_0)(u)+O_p(n^{-1/2}).$$

Next, we consider a piecewise constant modification $\overline{a}_{n,t}\overline{\Phi}_n$ of $a_{n,t}\Phi_n$, which is constant on the same intervals as $\tilde{\lambda}_n$. Since the smoothed Grenander estimator is not affected by changing the values of $\tilde{\lambda}_n$ at the jump points, for the proof we consider the right-continuous version of $\tilde{\lambda}_n$. Let $\tau_0=0$, $\tau_{m+1}=\tau_H$ and let $(\tau_i)_{i=1}^m$ be successive points of jump of $\tilde{\lambda}_n$. Then, for $u\in[\tau_i,\tau_{i+1})$, we choose

$$\overline{a}_{n,t}\overline{\Phi}_n(u;\hat{\beta}_n)=a_{n,t}(\hat{\Lambda}_n(u))\Phi_n(\hat{\Lambda}_n(u);\hat{\beta}_n), \tag{3.2.11}$$

where for $u \in [\tau_i, \tau_{i+1})$,

$$\hat{A}_n(u) = \begin{cases} \tau_i, & \text{if } \lambda_0(s) > \tilde{\lambda}_n(\tau_{i+1}), \text{ for all } s \in [\tau_i, \tau_{i+1}), \\ s, & \text{if } \lambda_0(s) = \tilde{\lambda}_n(s), \text{ for some } s \in [\tau_i, \tau_{i+1}), \\ \tau_{i+1}, & \text{if } \lambda_0(s) < \tilde{\lambda}_n(\tau_{i+1}), \text{ for all } s \in [\tau_i, \tau_{i+1}). \end{cases} \tag{3.2.12}$$

Furthermore, let E_n be the event from Lemma 3.2.1 and define

$$\overline{\Psi}_{n,t}(u) = \frac{\overline{a}_{n,t}\overline{\Phi}_n(u; \hat{\beta}_n)}{\Phi_n(u; \hat{\beta}_n)} \mathbb{1}_{E_n}, \quad u \leqslant T_{(n)}, \tag{3.2.13}$$

and $\overline{\Psi}_{n,x}(u) = 0$ otherwise. Let

$$j_{n1} = \max\{j : \tau_j \leqslant t - b\}, \qquad j_{n2} = \min\{j : \tau_j \geqslant t + b\} \tag{3.2.14}$$

be the last (first) jump point of $\tilde{\lambda}_n$ before (after) $t - b$ $(t + b)$. Note that, from the definition of $a_{n,t}$ and of $\hat{A}_n(u)$, it follows that $\overline{\Psi}_{n,t}(u) = 0$ for $u \notin [\tau_{j_{n1}}, \tau_{j_{n2}}]$. Now, define the following piecewise constant modification of $\theta_{n,t}$, by

$$\overline{\theta}_{n,t}(u, \delta, z) = \delta\, \overline{\Psi}_{n,t}(u) - e^{\hat{\beta}_n' z} \int_0^u \overline{\Psi}_{n,t}(v)\, d\hat{\Lambda}_n(v). \tag{3.2.15}$$

We then have the following property. The proof can be found in Section 3.6.

Lemma 3.2.2. *Let* $\overline{\theta}_{n,t}$ *be defined in* (3.2.15)*. Then*

$$\int \overline{\theta}_{n,t}(u, \delta, z)\, d\mathbb{P}_n(u, \delta, z) = 0. \tag{3.2.16}$$

At this point it is important to discuss in some detail how we will obtain suitable bounds for the second integral in (3.2.8). In order to do so, we first introduce the inverse process $\tilde{U}_n$. It is defined by

$$\hat{U}_n(a) = \underset{s \in [0, T_{(n)}]}{\operatorname{argmin}} \{\Lambda_n(s) - as\}. \tag{3.2.17}$$

and it satisfies the switching relation, $\tilde{\lambda}_n(t) \leqslant a$ if and only if $\tilde{U}_n(a) \geqslant t$, for $t \leqslant T_{(n)}$. In their analysis of the current status model, Groeneboom, Jongbloed, and Witte, 2010 encounter an integral that is similar to the second integral in (3.2.8). They bound this integral using the fact that the maximal distance between succeeding points of jump of the isotonic estimator is of the order $O_p(n^{-1/3} \log n)$. Such a property typically relies on the exponential bounds for the tail probabilities of $\hat{U}_n(a)$, obtained either by using a suitable exponential martingale (e.g., see Lemma 5.9 in Groeneboom and

Wellner, 1992), or by an embedding of the relevant sum process into Brownian motion or Brownian bridge (e.g., see Lemma 5.1 in Durot, Kulikov, and Lopuhaä, 2012). Unfortunately, an embedding of the process Λ_n is not available and in our current situation the martingale approach only yields to polynomial bounds for tail probabilities of $\hat{U}_n(a)$. A polynomial bound was also found by Durot, 2007 (see her Lemma 2) leading to

$$\sup_{t \in I_n} \mathbb{E}\left[\left(\tilde{\lambda}_n(t) - \lambda_0(t)\right)^p\right] \leqslant K n^{-p/3}, \tag{3.2.18}$$

for $p \in [1, 2)$ and some interval I_n (see her Theorem 1). By intersecting with the event E_n from Lemma 3.2.1 we extend (3.2.18) to a similar bound for $p = 2$. Groeneboom and Jongbloed, 2014 provide an alternative approach to bound the second integral in (3.2.8), based on bounds for (3.2.18) with $p = 2$. Unfortunately, they still make use of the fact that the maximum distance between succeeding points of jump of the isotonic estimator is of the order $O_p(n^{-1/3} \log n)$ to obtain a result similar to (3.2.21). Nevertheless, we do follow the approach in Groeneboom and Jongbloed, 2014, but instead of using the maximum distance between succeeding points of jump of $\tilde{\lambda}_n$, we use a bound on

$$\mathbb{E}\left[\sup_{t \in [\epsilon, M]} \left(\tilde{\lambda}_n(t) - \lambda_0(t)\right)^2\right], \tag{3.2.19}$$

for $0 < \epsilon < M < \tau_H$. Exponential bounds for the tail probabilities of $\hat{U}_n(a)$ would yield the same bound for (3.2.19) as the one in (3.2.18) apart from a factor $\log n$. Since we can only obtain polynomial bounds on the tail probabilities of $\hat{U}_n(a)$, we establish a bound for (3.2.19) of the order $O(n^{-4/9})$. This is probably not optimal, but this will turn out to be sufficient for our purposes and leads to the following intermediate result, the proof of which can be found in Section 3.6.

Lemma 3.2.3. *Suppose that (A1)–(A2) hold. Fix* $t \in (0, \tau_H)$ *and let* $\theta_{n,t}$ *and* $\bar{\theta}_{n,t}$ *be defined by* (3.2.10) *and* (3.2.15)*, respectively. Assume that* λ_0 *is differentiable, such that* λ_0' *is uniformly bounded above and below by strictly positive constants. Assume that* $x \mapsto \Phi(x; \beta_0)$ *is differentiable with a bounded derivative in a neighborhood of* t *and let* k *satisfy* (1.2.1)*. Then, it holds*

$$\int \left\{\overline{\theta}_{n,t}(u, \delta, z) - \theta_{n,t}(u, \delta, z)\right\} \, d\mathbb{P}(u, \delta, z) = O_p(b^{-1} n^{-2/3}).$$

The last step is to replace $\overline{\theta}_{n,t}$ in the first integral of (3.2.8) with a deterministic approximation. This is done in the next lemma, the proof of which can be found in Section 3.6.

Lemma 3.2.4. *Suppose that (A1)–(A3) hold. Fix* $t \in (0, \tau_H)$ *and take* $0 < \epsilon < t < M' < M < \tau_H$. *Assume that* λ_0 *is differentiable, such that* λ_0' *is uniformly bounded above and below by strictly positive constants. Assume that* $x \mapsto \Phi(x; \beta_0)$ *is differentiable with a bounded derivative in a neighborhood of* t. *Let* $\bar{\theta}_{n,t}$ *be defined in* (3.2.15) *and define*

$$\eta_{n,t}(u, \delta, z) = \mathbb{1}_{E_n} \left(\delta\, a_{n,t}(u) - e^{\beta_0' z} \int_0^u a_{n,t}(v)\, d\Lambda_0(v) \right), \quad u \in [0, \tau_H]. \tag{3.2.20}$$

where $a_{n,t}$ *is defined in* (3.2.9) *and* E_n *is the event from Lemma 3.2.1. Let* k *satisfy* (1.2.1). *Then, it holds*

$$\begin{aligned} &\int \left\{ \overline{\theta}_{n,t}(u, \delta, z) - \eta_{n,t}(u, \delta, z) \right\} d(\mathbb{P}_n - \mathbb{P})(u, \delta, z) \\ &\quad = O_p(b^{-3/2} n^{-13/18}) + O_p(n^{-1/2}). \end{aligned} \tag{3.2.21}$$

We are now in the position to state our first main result.

Theorem 3.2.5. *Suppose that (A1)–(A3) hold. Fix* $t \in (0, \tau_H)$. *Assume that* λ_0 *is* $m \geqslant 2$ *times continuously differentiable in* t, *such that* λ_0' *is uniformly bounded above and below by strictly positive constants. Moreover, assume that* $x \mapsto \Phi(x; \beta_0)$ *is differentiable with a bounded derivative in a neighborhood of* t *and let* k *satisfy* (1.2.1). *Let* $\tilde{\lambda}^{SG}$ *be defined in* (3.2.3) *and assume that* $n^{1/(2m+1)} b \to c > 0$. *Then, it holds*

$$n^{m/(2m+1)} \left(\tilde{\lambda}_n^{SG}(t) - \lambda_0(t) \right) \xrightarrow{d} N(\mu, \sigma^2),$$

where

$$\mu = \frac{(-c)^m}{m!} \lambda_0^{(m)}(t) \int_{-1}^{1} k(y) y^m\, dy \quad \textit{and} \quad \sigma^2 = \frac{\lambda_0(t)}{c\Phi(t; \beta_0)} \int k^2(u)\, du. \tag{3.2.22}$$

Furthermore,

$$n^{m/(2m+1)} \left(\tilde{\lambda}_n^{SG}(t) - \hat{\lambda}_n^{SM}(t) \right) \to 0, \tag{3.2.23}$$

in probability, where $\hat{\lambda}_n^{SM}(t)$ *is defined in* (3.2.2), *so that* $\tilde{\lambda}_n^{SM}(t)$ *has the same limiting distribution as* $\tilde{\lambda}_n^{SG}(t)$.

Proof. Choose $0 < \epsilon < t < M' < M < \tau_H$, so that for n sufficiently large, we have $\epsilon < t - b \leqslant t + b \leqslant M'$. Consider the event E_n from Lemma 3.2.1 and choose $\xi_1, \xi_2 > 0$ and ξ_3, such that it satisfies (3.6.19). We write

$$\begin{aligned}\tilde{\lambda}_n^{SG}(t) &= \int k_b(t-u)\, d\hat{\Lambda}_n(u) \\ &= \int k_b(t-u)\, d\Lambda_0(u) + \mathbb{1}_{E_n} \int k_b(t-u)\, d(\hat{\Lambda}_n - \Lambda_0)(u) \\ &\quad + \mathbb{1}_{E_n^c} \int k_b(t-u)\, d(\hat{\Lambda}_n - \Lambda_0)(u).\end{aligned} \tag{3.2.24}$$

Because $\mathbb{1}_{E_n^c} \to 0$ in probability, the third term on the right hand side tends to zero in probability. For the first term, we obtain from a change of variable, a Taylor expansion, and the properties of the kernel:

$$\begin{aligned}&n^{m/(2m+1)} \left\{ \int k_b(t-u)\, \lambda_0(u)\, du - \lambda_0(t) \right\} \\ &= n^{m/(2m+1)} \int_{-1}^{1} k(y) \{\lambda_0(t-by) - \lambda_0(t)\}\, dy \\ &= n^{m/(2m+1)} \int_{-1}^{1} k(y) \Big\{ -\lambda_0'(t) by + \cdots \\ &\qquad + \frac{\lambda_0^{(m-1)}(t)}{(m-1)!}(-by)^{m-1} + \frac{\lambda_0^{(m)}(\xi_n)}{m!}(-by)^m \Big\}\, dy \\ &\to \frac{(-c)^m}{m!} \lambda_0^{(m)}(t) \int_{-1}^{1} k(y) y^m\, dy,\end{aligned} \tag{3.2.25}$$

with $|\xi_n - t| < b|y|$. Finally, for the second term on the right hand side of (3.2.24), Lemmas 3.2.1 to 3.2.4 yield that

$$\begin{aligned}&n^{m/(2m+1)} \mathbb{1}_{E_n} \int k_b(t-u)\, d(\hat{\Lambda}_n - \Lambda_0)(u) \\ &= n^{m/(2m+1)} \int \eta_{n,t}(u,\delta,z)\, d(\mathbb{P}_n - \mathbb{P})(u,\delta,z) + o_p(1).\end{aligned} \tag{3.2.26}$$

For the first term on the right hand side of (3.2.26) we can write

$$\begin{aligned}&n^{m/(2m+1)} \int \eta_{n,t}(u,\delta,z)\, d(\mathbb{P}_n - \mathbb{P})(u,\delta,z) \\ &= n^{m/(2m+1)} \mathbb{1}_{E_n} \int \frac{\delta k_b(t-u)}{\Phi(u;\beta_0)}\, d(\mathbb{P}_n - \mathbb{P})(u,\delta,z) \\ &\quad - n^{m/(2m+1)} \mathbb{1}_{E_n} \int e^{\beta_0' z} \int_0^u a_{n,t}(v)\, d\Lambda_0(v)\, d(\mathbb{P}_n - \mathbb{P})(u,\delta,z).\end{aligned} \tag{3.2.27}$$

We will show that the first term on the right hand is asymptotically normal and the second term tends to zero in probability. Define

$$Y_{n,i} = n^{-(m+1)/(2m+1)} \Delta_i k_b(t - T_i)/\Phi(T_i; \beta_0),$$

so that the first term on the right hand side of (3.2.27) can be written as

$$\mathbb{1}_{E_n} n^{m/(2m+1)} \int \frac{\delta k_b(t-u)}{\Phi(u;\beta_0)} \, d(\mathbb{P}_n - \mathbb{P})(u, \delta, z) = \mathbb{1}_{E_n} \sum_{i=1}^{n} \left(Y_{n,i} - \mathbb{E}\left[Y_{n,i}\right] \right).$$

Using (3.1.3), together with a Taylor expansion and the boundedness assumptions on the derivatives of λ_0 and $\Phi(t;\beta_0)$, we have

$$\begin{aligned}
&\sum_{i=1}^{n} \mathrm{Var}(Y_{n,i}) \\
&= n^{-1/(2m+1)} \left\{ \int \frac{k_b^2(t-u)}{\Phi(u;\beta_0)^2} \, dH^{uc}(u) - \left(\int \frac{k_b(t-u)}{\Phi(u;\beta_0)} \, dH^{uc}(u) \right)^2 \right\} \\
&= n^{-1/(2m+1)} \left\{ \frac{1}{b} \int_{-1}^{1} k^2(y) \frac{\lambda_0(t-by)}{\Phi(t-by;\beta_0)} \, dy - \left(\int k_b(t-u) \, \lambda_0(u) \, du \right)^2 \right\} \\
&= \frac{\lambda_0(t)}{c\Phi(t;\beta_0)} \int_{-1}^{1} k^2(y) \, dy - n^{-\frac{1}{2m+1}} \int_{-1}^{1} y k^2(y) \left[\frac{d}{dt} \frac{\lambda_0(t)}{\Phi(t;\beta_0)} \right]_{t=\xi_y} dy + o(1) \\
&= \frac{\lambda_0(t)}{c\Phi(t;\beta_0)} \int_{-1}^{1} k^2(y) \, dy + o(1).
\end{aligned} \tag{3.2.28}$$

Moreover,

$$|Y_{n,i}| \leqslant n^{-(m+1)/(2m+1)} \Phi(M;\beta_0)^{-1} \sup_{x \in [-1,1]} k(x),$$

so that $\sum_{i=1}^{n} \mathbb{E}\left[|Y_{n,i}|^2 \mathbb{1}_{\{|Y_{n,i}|>\epsilon\}}\right] \to 0$, for any $\epsilon > 0$, since $\mathbb{1}_{\{|Y_{n,i}|>\epsilon\}} = 0$, for n sufficiently large. Consequently, by Lindeberg central limit theorem, and the fact that $\mathbb{1}_{E_n} \to 1$ in probability, we obtain

$$\mathbb{1}_{E_n} n^{m/(2m+1)} \int \frac{\delta k_b(t-u)}{\Phi(u;\beta_0)} \, d(\mathbb{P}_n - \mathbb{P})(u, \delta, z) \to N(0, \sigma^2). \tag{3.2.29}$$

For the second term on the right hand side of (3.2.27), write

$$n^{\frac{m}{2m+1}} \int e^{\beta_0' z} \int_0^u a_{n,t}(v) \, d\Lambda_0(v) \, d(\mathbb{P}_n - \mathbb{P})(u, \delta, z) = \sum_{i=1}^{n} \left(\widetilde{Y}_{n,i} - \mathbb{E}[\widetilde{Y}_{n,i}] \right).$$

where

$$\widetilde{Y}_{n,i} = n^{-(m+1)/(2m+1)} e^{\beta_0' Z_i} \int_0^{T_i} \frac{k_b(t-v)}{\Phi(v;\beta_0)} \, d\Lambda_0(v).$$

We have

$$\sum_{i=1}^{n} \mathrm{Var}(\widetilde{Y}_{n,i}) \leqslant \sum_{i=1}^{n} \mathbb{E}\left[\widetilde{Y}_{n,i}^2\right]$$
$$\leqslant n^{-1/(2m+1)} \int e^{2\beta_0' z} \left(\int_0^u \frac{k_b(t-v)}{\Phi(v;\beta_0)}\, d\Lambda_0(v)\right)^2 d\mathbb{P}(u,\delta,z),$$

where the integral on the right hand side is bounded by

$$\left(\int_{t-b}^{t+b} \frac{k_b(t-v)}{\Phi(v;\beta_0)}\, d\Lambda_0(v)\right)^2 \Phi(0;2\beta_0)$$
$$\leqslant \frac{\Phi(0;2\beta_0)}{\Phi^2(M;\beta_0)} \left(\int_{t-b}^{t+b} k_b(t-v)\, d\Lambda_0(v)\right)^2 = O(1).$$

Hence, the second term on the right hand side of (3.2.27) tends to zero in probability. Together with (3.2.24), (3.2.25), and (3.2.29), this proves the first part of the theorem.

For the smoothed maximum likelihood estimator, we can follow the same approach and obtain similar results as those in Lemmas 3.2.1 to 3.2.4. The arguments are more or less the same as those used to prove Lemmas 3.2.1 to 3.2.4. We briefly sketch the main differences. First, $\hat{\Lambda}_n$, we will now be replaced by

$$\hat{\Lambda}_n(t) = \int_0^t \hat{\lambda}_n(u)\, du$$

in (3.2.7). Then, since the maximum likelihood estimator is defined as the left slope of the greatest convex minorant of a cumulative sum diagram that is different from the one corresponding to the Grenander-type estimator, Lemmas 3.2.1 and 3.2.2 will hold with a different event $\widehat{E}_n$ and $\overline{\Psi}_{n,t}$ will have a simpler form (see Lemmas A.2.1-A.2.2 and definition (A.2.4) in A.2. Similar to the proof of Lemma 3.2.3, the proof of its counterpart for the maximum likelihood estimator (see Lemma A.2.10) is quite technical and involves bounds on the tail probabilities of the inverse process corresponding to $\hat{\lambda}_n$ (see Lemma A.2.5), used to obtain the analogue of (3.2.19) (see Lemma A.2.6). Moreover, the inverse process related to the maximum likelihood estimator is defined by

$$\hat{U}_n(a) = \underset{s\in[T_{(1)},T_{(n)}]}{\mathrm{argmin}} \left\{V_n(s) - a\hat{W}_n(s)\right\}, \tag{3.2.30}$$

where V_n and $\hat{W}_n$ are defined in (3.2.1), and we get a slightly different bound on the tail probabilities of $\hat{U}_n$ (compare Lemmas 3.6.3 and A.2.5). The

reason is that the martingale decomposition of $V_n(s) - a\hat{W}_n(s)$ has a simper form. The counterpart of Lemma 3.2.4 (see Lemma A.2.11) is established in the same way, replacing $\tilde{\lambda}_n$ by $\hat{\lambda}_n$. For details we refer to Section A.2.

From (3.2.24) and (3.2.26), we have that

$$\begin{aligned} n^{m/(2m+1)}\tilde{\lambda}_n^{SG}(t) &= n^{m/(2m+1)} \int k_b(t-u)\, d\Lambda_0(u) \\ &\quad + n^{m/(2m+1)} \int \eta_{n,t}(u,\delta,z)\, d(\mathbb{P}_n - \mathbb{P})(u,\delta,z) + o_p(1) \end{aligned} \tag{3.2.31}$$

where $\eta_{n,t}$ is defined in (3.2.20) and where

$$n^{m/(2m+1)} \int \eta_{n,t}(u,\delta,z)\, d(\mathbb{P}_n - \mathbb{P})(u,\delta,z) \to N(0,\sigma^2). \tag{3.2.32}$$

Similarly, from the results in Section A.2, we have that there exists an event $\widehat{E}_n$, such that

$$\begin{aligned} n^{m/(2m+1)}\hat{\lambda}_n^{SM}(t) &= n^{m/(2m+1)} \int k_b(t-u)\, d\Lambda_0(u) \\ &\quad + n^{m/(2m+1)} \int \widehat{\eta}_{n,t}(u,\delta,z)\, d(\mathbb{P}_n - \mathbb{P})(u,\delta,z) + o_p(1) \end{aligned} \tag{3.2.33}$$

where $\widehat{\eta}_{n,t}$ is defined in (3.2.20) with $\widehat{E}_n$ instead of E_n, where $\mathbb{1}_{\widehat{E}_n} \to 1$ in probability, and where

$$n^{m/(2m+1)} \int \widehat{\eta}_{n,t}(u,\delta,z)\, d(\mathbb{P}_n - \mathbb{P})(u,\delta,z) \to N(0,\sigma^2). \tag{3.2.34}$$

Together with (3.2.32) and (3.2.34), this means that

$$\begin{aligned} &n^{m/(2m+1)} \left(\tilde{\lambda}_n^{SG}(t) - \hat{\lambda}_n^{SM}(t)\right) \\ &= \left(\mathbb{1}_{\widehat{E}_n^c}\mathbb{1}_{E_n} - \mathbb{1}_{E_n^c}\mathbb{1}_{\widehat{E}_n}\right) \\ &\quad \times n^{\frac{m}{2m+1}} \int \left\{\delta a_{n,t}(u) - e^{\beta_0' z}\int_0^u a_{n,t}(v)\, d\Lambda_0(v)\right\} d(\mathbb{P}_n - \mathbb{P})(u,\delta,z) + o_p(1) \\ &= \mathbb{1}_{\widehat{E}_n^c} O_p(1) - \mathbb{1}_{E_n^c} O_p(1) + o_p(1) = o_p(1), \end{aligned}$$

because $\mathbb{1}_{\widehat{E}_n^c} \to 0$ and $\mathbb{1}_{E_n^c} \to 0$ in probability. □

Note that in the special case $\beta_0 = 0$ and $m = 2$, we recover Theorem 2.2.2 and Theorem 11.8 in Groeneboom and Jongbloed, 2014, for the right

censoring model without covariates. The fact that $\tilde{\lambda}_n^{SG}(t)$ and $\hat{\lambda}_n^{SM}(t)$ are asymptotically equivalent does not come as a surprise, since for the corresponding isotonic estimators according to Theorem 2 in Lopuhaä and Nane, 2013, for $t \in (0, \tau_H)$ fixed, $n^{1/3}\big(\tilde{\lambda}_n(t) - \hat{\lambda}_n(t)\big) \to 0$, in probability. However, we have not been able to exploit this fact, and we have established the asymptotic equivalence in (3.2.23) by obtaining the expansions in (3.2.31) and (3.2.33) separately for each estimator.

Remark 3.2.6. The estimators considered in Theorem 3.2.5 are based on the partial maximum likelihood estimator $\hat{\beta}_n$, which defines the Breslow estimator, see (1.1.7), and the cumulative sum diagram from which the SMLE is determined, see (3.2.1). However, Theorem 3.2.5 remains true, if $\hat{\beta}_n$ is any estimator that satisfies

$$\hat{\beta}_n - \beta_0 \to 0, \text{ a.s.}, \quad \text{and} \quad \sqrt{n}(\hat{\beta}_n - \beta_0) = O_p(1) \tag{3.2.35}$$

In particular, this holds for the partial MLE for β_0. See, e.g., Theorems 3.1 and 3.2 in Tsiatis, 1981. When proving consistency of the bootstrap, we are not able to establish bootstrap versions of Theorems 3.1 and 3.2 in Tsiatis, 1981, but, in view of this remark, it is sufficient to assume the bootstrap version of (3.2.35).

3.3 MAXIMUM SMOOTH LIKELIHOOD ESTIMATOR

Maximum smoothed likelihood estimation is studied in Eggermont and LaRiccia, 2000, who obtain L_1-error bounds for the maximum smoothed likelihood estimator of a monotone density. This method was also considered in Groeneboom, Jongbloed, and Witte, 2010 for estimating the distribution function of interval censored observations. The approach is to smooth the loglikelihood and then maximize the smoothed loglikelihood over all monotone functions of interest. For a fixed β, the (pseudo) loglikelihood for the Cox model can be expressed as

$$l_\beta(\lambda_0) = \int \left(\delta \log \lambda_0(t) - e^{\beta' z} \int_0^t \lambda_0(u)\, du \right) d\mathbb{P}_n(t, \delta, z), \tag{3.3.1}$$

(see (1.1.5)). To construct the maximum smoothed likelihood estimator (MSLE) we replace $\mathbb{P}_n$ in the previous expression with the smoothed empirical measure (in the time direction),

$$d\tilde{\mathbb{P}}_n(t, \delta, z) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{(\Delta_i, Z_i)}(\delta, z)\, k_b(t - T_i)\, dt,$$

and then maximize the smoothed (pseudo) loglikelihood

$$\ell^s_\beta(\lambda_0) = \int \left(\delta \log \lambda_0(t) - e^{\beta' z} \int_0^t \lambda_0(u)\, du \right) d\tilde{\mathbb{P}}_n(t, \delta, z). \tag{3.3.2}$$

The characterization of the MSLE is similar to that of the ordinary MLE. It involves the following processes. Fix $\beta \in \mathbb{R}^p$ and let

$$\begin{aligned} w_n(t;\beta) &= \frac{1}{n} \sum_{i=1}^n e^{\beta' Z_i} \int_t^\infty k_b(u - T_i)\, du, \\ v_n(t) &= \frac{1}{n} \sum_{i=1}^n \Delta_i k_b(t - T_i). \end{aligned} \tag{3.3.3}$$

The next lemma characterizes the maximizer of ℓ^s_β. The proof can be found in Section 3.6.2.

Lemma 3.3.1. *Let ℓ^s_β, w_n and v_n be defined by* (3.3.2) *and* (3.3.3)*, respectively. The unique maximizer of ℓ^s_β over all nondecreasing positive functions λ_0 can be described as the slope of the greatest convex minorant (GCM) of the continuous cumulative sum diagram*

$$t \mapsto \left(\int_0^t w_n(x;\beta)\, dx, \int_0^t v_n(x)\, dx \right), \qquad t \in [0, \tau_\beta], \tag{3.3.4}$$

where $\tau_\beta = \sup\{t \geqslant 0 : w_n(t;\beta) > 0\}$.

For a fixed β, let $\hat{\lambda}^s_n(\cdot;\beta)$ be the unique maximizer of $\ell^s_\beta(\lambda_0)$ over all nondecreasing positive functions λ_0. We define the MSLE by

$$\hat{\lambda}^{MS}_n(t) = \hat{\lambda}^s_n(t; \hat{\beta}_n), \tag{3.3.5}$$

where $\hat{\beta}_n$ denotes the maximum partial likelihood estimator for β_0. It can be seen that under appropriate smoothness assumptions,

$$\begin{aligned} \int_0^t w_n(u;\hat{\beta}_n)\, du &= \int \hat{W}_n(s) k_b(t-s)\, ds + O_p(n^{-1/2}) + O_p(b), \\ \int_0^t v_n(u)\, du &= \int V_n(s) k_b(t-s)\, ds + O_p(b), \end{aligned}$$

where the processes V_n and $\hat{W}_n$, as defined in (3.2.1), determine the cumulative sum diagram corresponding to the ordinary MLE. This means that the cumulative sumdiagram that characterizes the MSLE, is asymptotically equivalent to a kernel smoothed version of the cumulative sumdiagram that characterizes the ordinary MLE.

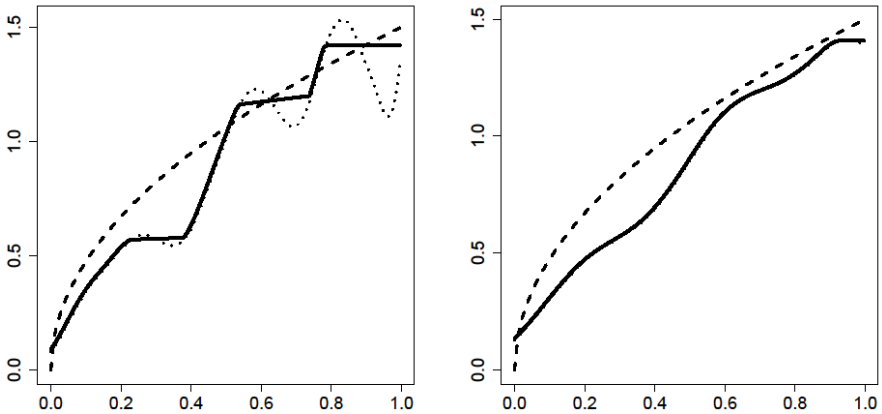


Figure 8: Left panel: The MSLE (solid) and the naive estimator (dotted) of the hazard function (dashed) using bandwidth $b_n = 0.5n^{-1/5}$. Right panel: The same but using bandwidth $b_n = n^{-1/5}$.

As can be seen from the proof of Lemma 3.3.1, the MSLE minimizes

$$\psi(\lambda) = \frac{1}{2}\int \left(\lambda(t) - \frac{v_n(t)}{w_n(t;\hat{\beta}_n)}\right)^2 w_n(t;\beta)\,dt, \tag{3.3.6}$$

over all nondecreasing functions λ. This suggests

$$\hat{\lambda}_n^{\text{naive}}(t) = \frac{v_n(t)}{w_n(t;\hat{\beta}_n)} \tag{3.3.7}$$

as a naive estimator for λ_0. The naive estimator is the ratio of two smooth functions, being the derivatives of the vertical and horizontal processes in the continuous cumulative sum diagram in (3.3.4). The naive estimator is smooth, but not necessarily monotone and its weighted least squares projection is the MSLE. Figure 8 illustrates the MSLE and the naive estimator for a sample of size $n = 500$ from a Weibull baseline distribution with shape parameter 1.5 and scale 1. For simplicity, the covariate and the censoring time are chosen to be uniformly $(0, 1)$ distributed and we take $\beta_0 = 0.5$. We used the triweight kernel function $k(u) = (35/32)(1-u^2)^3 \mathbb{1}_{\{|u|\leqslant 1\}}$ and bandwidth $b = n^{-1/5}$. Note that if we use bandwidth $b_n = 0.5n^{-1/5}$, the naive estimator is not monotone, but the distance to the MSLE (which is the isotonic version of $\hat{\lambda}_n^{\text{naive}}$) is very small. On the other hand, for bandwidth $b_n = n^{-1/5}$ isotonization is not needed and the two estimators coincide. Indeed, following the reasoning in Groeneboom, Jongbloed, and Witte, 2010, the derivation of the asymptotic distribution of $\hat{\lambda}_n^{MS}$ is based on showing that with probability converging to one, the naive estimator will be monotone and equal to $\hat{\lambda}_n^{MS}$ on large intervals. Consequently, it will be sufficient

to find the asymptotic distribution of the naive estimator. The advantage of this approach is that in this way we basically have to deal with the naive estimator, which is a more tractable process.

This approach applies more generally. The situation for the MSLE is a special case of the more general situation, where the isotonic estimator is the derivative $d\widehat{Y}_n/dX_n$ of the greatest convex minorant $\{(X_n(t), \widehat{Y}_n(t)) : t \in [0, \hat{\tau}]\}$ of the graph $\{(X_n(t), Y_n(t)) : t \in [0, \hat{\tau}]\}$, for some $0 < \hat{\tau} < \tau_H$, where X_n and Y_n are differentiable processes in a cumulative sumdiagram, whereas the naive estimator is the ratio dY_n/dX_n of the derivatives of X_n and Y_n. The MSLE and the corresponding naive estimator from (3.3.7) form a special case, with $X_n = \tilde{W}_n$, $Y_n = \tilde{V}_n$, where

$$\tilde{W}_n(t) = \int_0^t w_n(u; \hat{\beta}_n)\, du, \qquad \tilde{V}_n(t) = \int_0^t v_n(u)\, du, \tag{3.3.8}$$

and $\hat{\tau} = \sup\{t \geqslant 0 : w_n(t; \hat{\beta}_n) > 0\}$. The following result considers the general setup and shows that, in that case, the isotonic estimator and the corresponding naive estimator coincide on large intervals with probability tending to one.

Lemma 3.3.2. *Let X_n and Y_n be differentiable processes and let $\{(X_n(t), \widehat{Y}_n(t)) : t \in [0, \hat{\tau}]\}$ be the greatest convex minorant of the graph $\{(X_n(t), Y_n(t)) : t \in [0, \hat{\tau}]\}$, for some $0 < \hat{\tau} < \tau_H$. Let $\hat{\lambda}_n^{IS}(t) = d\widehat{Y}_n(t)/dX_n(t)$ and $\hat{\lambda}_n^{naive}(t) = dY_n(t)/dX_n(t)$, for $t \in [0, \hat{\tau}]$. Suppose that*

(a) $X_n(s) \leqslant X_n(t)$, for $0 \leqslant s \leqslant t \leqslant \hat{\tau}$;

(b) for every $t \in (0, \hat{\tau})$ fixed, $\hat{\lambda}_n^{naive}(t) \to \lambda_0(t)$, in probability;

(c) for all $0 < \ell < M < \hat{\tau}$ fixed, $\mathbb{P}\big(\hat{\lambda}_n^{naive}$ is increasing on $[\ell, M]\big) \to 1$;

(d) there exists processes X_0 and Y_0, such that

$$\sup_{t \in [0, \hat{\tau}]} |X_n(t) - X_0(t)| \xrightarrow{P} 0, \qquad \sup_{t \in [0, \hat{\tau}]} |Y_n(t) - Y_0(t)| \xrightarrow{P} 0.$$

Moreover, the process X_0 is absolutely continuous with a strictly positive nonincreasing derivative x_0, and X_0 and Y_0 are related by

$$Y_0(t) = \int_0^t \lambda_0(u)\, dX_0(u).$$

Then, for all $0 < \ell < M < \hat{\tau}$, $\mathbb{P}\left(\hat{\lambda}_n^{naive}(t) = \hat{\lambda}_n^{MS}(t), \text{ for all } t \in [\ell, M]\right) \to 1$.

The proof of Lemma 3.3.2 can be found in the Appendix 3.6.2. We will apply Lemma 3.3.2 to the MSLE and the naive estimator from (3.3.7). Recall that $\hat{\lambda}_n^{MS}$ and $\hat{\lambda}_n^{naive}$ are defined on $[0, \hat{\tau}_n]$, where

$$\hat{\tau}_n = \sup\{t \geqslant 0 : w_n(t; \hat{\beta}_n) > 0\},$$

and note that $\hat{\tau}_n \to \tau_H$ with probability one. Condition (a) of Lemma 3.3.2 is trivially fulfilled with $X_n = \tilde{W}_n$ defined in (3.3.8). A first key result is that for each $0 < \ell < M < \tau_H$, it holds

$$\begin{aligned} \sup_{t\in[\ell,M]} |v_n(t) - h(t)| &= O(b^m) + O_p(b^{-1}n^{-1/2}), \\ \sup_{t\in[\ell,M]} |w_n(t; \hat{\beta}_n) - \Phi(t; \beta_0)| &= O(b^m) + O_p(b^{-1}n^{-1/2}), \end{aligned} \tag{3.3.9}$$

where v_n, w_n and Φ are defined in (3.3.3) and (3.1.2), see Lemma 3.6.8. A direct consequence of (3.3.9) is the fact that the naive estimator converges to λ_0 uniformly on compact intervals within the support, as long as $b \to 0$ and $1/b = o(n^{1/2})$, see Lemma 3.6.9. In particular, this will ensure condition (b) of Lemma 3.3.2. A second key result is that, under suitable smoothness conditions, for each $0 < \ell < M < \tau_H$, it holds

$$\begin{aligned} \sup_{t\in[\ell,M]} |v_n'(t) - h'(t)| &\xrightarrow{\mathbb{P}} 0, \\ \sup_{t\in[\ell,M]} |w_n'(t; \hat{\beta}_n) - \Phi'(t; \beta_0)| &\xrightarrow{\mathbb{P}} 0, \end{aligned} \tag{3.3.10}$$

see Lemma 3.6.8. This will imply that the naive estimator is increasing on large intervals with probability tending to one, see Lemma 3.6.11, which yields condition (c) of Lemma 3.3.2. Finally, condition (d) of Lemma 3.3.2 is shown to hold with $X_0 = H^{uc}$ from (3.1.1) and $Y_0 = W_0$, defined by

$$W_0(t) = \int_0^t \Phi(u; \beta_0)\, du. \tag{3.3.11}$$

In view of (3.3.9) and (3.1.3), this is to be expected, and it is made precise in Lemma 3.6.12. Hence, Lemma 3.3.2 applies to the MSLE and the naive estimator from (3.3.7). Therefore we have the following corollary.

Corollary 3.3.3. *Suppose that (A1)-(A2) hold. Let* $H^{uc}(t)$ *and* $\Phi(t; \beta_0)$ *be defined in* (3.1.1) *and* (3.1.2)*, and let* $h(t) = dH^{uc}(t)/dt$*, satisfying* (3.1.3)*. Suppose that* h *and* $t \mapsto \Phi(t; \beta_0)$ *are continuously differentiable, and that* λ_0' *is uniformly bounded from below by a strictly positive constant. Let* k *satisfy* (1.2.1) *and let* $\hat{\lambda}_n^{naive}$ *be defined in* (3.3.7)*. If* $b \to 0$ *and* $1/b = O(n^\alpha)$*, for some* $\alpha \in (0, 1/4)$*, then for each* $0 < \ell < M < \tau_H$*,*

$$\mathbb{P}\left(\hat{\lambda}_n^{naive}(t) = \hat{\lambda}_n^{MS}(t), \textit{ for all } t \in [\ell, M]\right) \to 1.$$

Consequently, for all $t \in (0, \tau_H)$*, the asymptotic distributions of* $\hat{\lambda}_n^{\text{naive}}(t)$ *and* $\hat{\lambda}_n^{MS}(t)$ *are the same.*

Under similar smoothness conditions as needed in Lemma 3.6.8 to obtain (3.3.9), one can show that

$$\begin{aligned} \sup_{t\in[\ell,M]} |v_n'(t) - h'(t)| &= O(b^{m-1}) + O_p(b^{-2}n^{-1/2}), \\ \sup_{t\in[\ell,M]} |w_n'(t;\hat{\beta}_n) - \Phi'(t;\beta_0)| &= O(b^{m-1}) + O_p(b^{-1}n^{-1/2}). \end{aligned} \tag{3.3.12}$$

In that case, it can also be proved that

$$\sup_{t\in[\ell,M]} \left| \frac{d}{dt}\hat{\lambda}_n^{\text{naive}}(t) - \lambda_0'(t) \right| = O(b^{m-1}) + O_p(b^{-2}n^{-1/2}) = o_P(1),$$

as long as $b \to 0$ and $1/b^2 = o(n^{1/2})$.

From Corollary 3.3.3 and the fact that the naive estimator converges to λ_0 uniformly on compact intervals within the support, see Lemma 3.6.9, another consequence of Lemma 3.3.2 is the following corollary concerning uniform convergence of the MSLE.

Corollary 3.3.4. *Suppose that (A1)-(A2) hold. Let* $H^{uc}(t)$ *and* $\Phi(t;\beta_0)$ *be defined in* (3.1.1) *and* (3.1.2)*, and let* $h(t) = dH^{uc}(t)/dt$*, satisfying* (3.1.3)*. Suppose that* h *and* $t \mapsto \Phi(t;\beta_0)$ *are* $m \geqslant 1$ *times continuously differentiable, and that* λ_0' *is uniformly bounded from below by a strictly positive constant. Let* k *be* m*-orthogonal satisfying* (1.2.1)*. Then, the maximum smooth likelihood estimator is uniformly consistent on compact intervals* $[\ell, M] \subset (0, \tau_H)$*:*

$$\sup_{t\in[\ell,M]} \left| \hat{\lambda}_n^{MS}(t) - \lambda_0(t) \right| = O(b^m) + O_p(b^{-1}n^{-1/2}).$$

Proof. The result follows immediately from Corollary 3.3.3 and Lemma 3.6.9. □

To obtain the asymptotic distribution of $\hat{\lambda}_n^{MS}(t)$, we first obtain the asymptotic distribution of $\hat{\lambda}_n^{\text{naive}}(t)$. To this end we establish the joined asymptotic distribution of the vector $(w_n(t;\hat{\beta}_n), v_n(t))$, see Lemma 3.6.13. Then an application of the delta-method yields the limit distribution of $\hat{\lambda}_n^{\text{naive}}$ as well as that of $\hat{\lambda}_n^{MS}$, due to Corollary 3.3.3.

Theorem 3.3.5. *Suppose that (A1)-(A2) hold. Let* $H^{uc}(t)$ *and* $\Phi(t;\beta_0)$ *be defined in* (3.1.1) *and* (3.1.2)*, and let* $h(t) = dH^{uc}(t)/dt$*, satisfying* (3.1.3)*. Suppose that* h *and* $t \mapsto \Phi(t;\beta_0)$ *are* $m \geqslant 2$ *times continuously differentiable and let* k *be*

m-orthogonal satisfying (1.2.1). *Let $\hat{\lambda}_n^{MS}(t)$ be defined in* (3.3.5) *and assume that $n^{1/(2m+1)}b \to c > 0$. Then, for each $t \in (0, \tau_H)$, the following holds*

$$n^{m/(2m+1)}\left(\hat{\lambda}_n^{MS}(t) - \lambda_0(t)\right) \xrightarrow{d} N(\tilde{\mu}, \sigma^2),$$

where

$$\begin{aligned} \tilde{\mu} &= \frac{(-c)^m}{m!}\frac{h^{(m)}(t) - \lambda_0(t)\Phi^{(t)}(x;\beta_0)}{\Phi(t;\beta_0)}\int_{-1}^{1} k(y)y^m\,dy; \\ \sigma^2 &= \frac{\lambda_0(t)}{c\Phi(t;\beta_0)}\int_{-1}^{1} k^2(y)\,dy. \end{aligned} \tag{3.3.13}$$

This also holds if we replace $\hat{\lambda}_n^{MS}(t)$ with $\hat{\lambda}_n^{naive}(t)$, as defined in (3.3.7).

The proof of Theorem 3.3.5 can be found in Section 3.6.2. Theorem 3.3.5 is comparable to Theorem 3.2.5. The limiting variance is the same, but the asymptotic mean is shifted. A natural question is whether $\hat{\lambda}_n^{MS}(x)$ is asymptotically equivalent to these estimators, if we correct for the difference in the asymptotic mean. The next theorem shows that this is indeed the case. The proof can be found in Section 3.6.2. In order to use results from Section 3.2, apart from conditions (A1) and (A2), we have to assume also (A3).

Theorem 3.3.6. *Suppose that (A1)-(A3) hold. Suppose that λ_0 and $t \mapsto \Phi(t;\beta_0)$ are $m \geqslant 2$ times continuously differentiable, with λ_0' uniformly bounded from below by a strictly positive constant, and let k be m-orthogonal satisfying* (1.2.1). *Let $\hat{\lambda}_n^{MS}(t)$ be the maximum smoothed likelihood estimator and let $\hat{\lambda}^{SM}(t)$ be the smoothed maximum likelihood estimator, defined in* (3.2.2). *Let $\widetilde{\mu}$ and μ be defined in* (3.3.13) *and* (3.2.22), *respectively. Then, for each $t \in (0, \tau_H)$, the following holds*

$$n^{m/(2m+1)}\left(\hat{\lambda}_n^{MS}(t) - \hat{\lambda}_n^{SM}(t)\right) - (\widetilde{\mu} - \mu) \to 0$$

in probability, and similarly if we replace $\hat{\lambda}_n^{SM}(t)$ by the smoothed Grenander-type estimator $\tilde{\lambda}_n^{SG}(t)$, defined in (3.2.3).

3.4 ISOTONIZED KERNEL ESTIMATOR

The fourth method that we consider is an isotonized kernel estimator. Let Λ_n^s be the smoothed Breslow estimator defined by

$$\Lambda_n^s(t) = \int k_b(t-u)\Lambda_n(u)\,du. \tag{3.4.1}$$

In order to avoid problems at the right end of the support, we fix $0 < \tau^* < \tau_H$ and consider estimation only on $[0, \tau^*]$. A similar approach was considered in Groeneboom and Jongbloed, 2013, when estimating a monotone

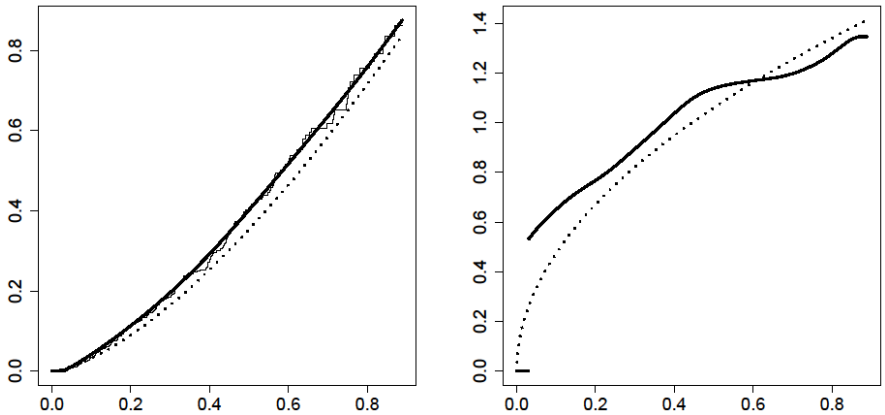


Figure 9: Left panel: The smoothed version (solid) of the Breslow estimator (solid-step function) for the cumulative baseline hazard (dotted) and the greatest convex minorant (dashed). Right panel: The Grenander-type smoothed estimator (solid) of the baseline hazard (dotted).

hazard of uncensored observations. The main reason in our setup is that in order to exploit the representation in (3.1.3), we must have $t < \tau_H$, because $\Phi(t; \beta_0) = 0$ otherwise. The isotonized kernel estimator of a nondecreasing baseline hazard is a Grenander-type estimator, as being defined as the left derivative of the greatest convex minorant of Λ_n^s on $[0, \tau^*]$. We denote this estimator by $\tilde{\lambda}_n^{GS}$.

Note that this type of estimator was defined also in Nane, 2013 without the restriction on $[0, \tau^*]$. Strong pointwise consistency was proved and uniform consistency on intervals $[\epsilon, \tau_H - \epsilon] \subset [0, \tau_H]$ follows immediately from the monotonicity and the continuity of λ_0. These results also illustrate that there are consistency problems at the end point of the support. Since in practice we do not even know τ_H, the choice of τ^* might be an issue. Since one wants τ^* to be close to τ_H, one reasonable choice would be to take as τ^* the 95%-empirical quantile of the follow-up times, because this converges to the theoretical 95%-quantile, which is strictly smaller than τ_H. Note that we cannot choose $T_{(n)}$, because it converges to τ_H, i.e., for large n, it will be greater than any fixed $\tau^* < \tau_H$.

Figure 9 shows the smoothed Breslow estimator and the isotonized kernel estimator for the same sample as in Figure 8. To avoid problems at the boundary we use the boundary corrected version of the kernel function and consider the data up to the 95%-empirical quantile of the follow-up times. The bandwidth is $b_n = n^{-1/5}$. Similar to the proof of Lemma 3.3.1,

it follows from Lemma 1 in Groeneboom and Jongbloed, 2010, that $\tilde{\lambda}_n^{GS}$ is continuous and is the unique maximizer of

$$\psi(\lambda) = \frac{1}{2}\int_0^{\tau^*} \left(\lambda(t) - \tilde{\lambda}_n^s(t)\right)^2 \, dt$$

over all nondecreasing functions λ, where

$$\tilde{\lambda}_n^s(t) = \frac{d}{dt}\Lambda_n^s(t) = \int k_b'(t-u)\Lambda_n(u)\, du. \tag{3.4.2}$$

This suggests

$$\tilde{\lambda}_n^{\text{naive}}(t) = \tilde{\lambda}_n^s(t) \tag{3.4.3}$$

as another naive estimator for $\lambda_0(t)$. This naive estimator is the derivative of the smoothed Breslow. Again, it is smooth but not necessarily monotone and its least squares projection is the isotonized kernel estimator. Note that by means of integration by parts, we can also write

$$\tilde{\lambda}_n^s(t) = \int k_b(t-u)\, d\Lambda_n(u).$$

Hence, the naive estimator from (3.4.2) is equal to the ordinary Rosenblatt-Parzen kernel estimator for the baseline hazard. Asymptotic normality for this estimator under random censoring has been proven by Ramlau-Hansen, 1983 and Tanner and Wong, 1983. A similar result in a general counting processes setup, that includes the Cox model, is stated in Wells, 1994, but only the idea of the proof is provided. We will establish asymptotic normality for the naive estimator from (3.4.2) in our current setup of the Cox model, see the proof of Theorem 3.4.3.

Then, similar to the approach used in Section 3.3, the derivation of the asymptotic distribution of $\tilde{\lambda}_n^{GS}$ is based on showing that it is equal to the naive estimator in (3.4.3) on large intervals with probability converging to one. The isotonized kernel estimator is a special case of Lemma 3.3.2, with $X_n(t) = t$, $Y_n(t) = \Lambda_n^s(t)$, and $\hat{\tau} = \tau^*$. As before, condition (a) of Lemma 3.3.2 is trivial and condition (b) is fairly straightforward, see (3.6.69) for details. condition (c) of Lemma 3.3.2 is established in Lemma 3.6.14 and condition (d) is also straightforward, see (3.6.70) for details. Hence, Lemma 3.3.2 applies to the isotonized kernel estimator and the naive estimator from (3.4.3), which leads to the following corollary.

Corollary 3.4.1. *Suppose that (A1)-(A2) hold. Let λ_0 be continuously differentiable, with λ_0' uniformly bounded from below by a strictly positive constant, and let k satisfy* (1.2.1). *If $b \to 0$ and $1/b = O(n^\alpha)$, for some $\alpha \in (0, 1/4)$, then for each $0 < \ell < M < \tau^*$, it holds*

$$\mathbb{P}\left(\tilde{\lambda}_n^{\text{naive}}(t) = \tilde{\lambda}_n^{GS}(t) \textit{ for all } t \in [\ell, M]\right) \to 1.$$

Consequently, for all $t \in (0, \tau^*)$*, the asymptotic distributions of* $\tilde{\lambda}_n^{\text{naive}}(t)$ *and* $\tilde{\lambda}_n^{GS}(t)$ *are the same.*

The proof of Corollary 3.4.1 can be found in Section 3.6.3.

Remark 3.4.2. Note that in case the kernel function is strictly positive on $(-1, 1)$ and the baseline hazard is strictly increasing, one can easily check that

$$t \mapsto \int_{-1}^{t/b} k(y)\lambda_0(t - by)\, dy$$

is a continuously differentiable, strictly increasing function on $[0, M]$ and as a result we obtain that

$$\begin{aligned} \frac{d}{dt}\tilde{\lambda}_n^{\text{naive}}(t) &= \frac{d}{dt}\left(\int_{-1}^{t/b} k(y)\lambda_0(t - by)\, dy\right) \\ &\quad + \frac{1}{b^2}\int k'\left(\frac{t-u}{b}\right) d\left(\Lambda_n - \Lambda_0\right)(u) \\ &\geqslant C + o_P(1). \end{aligned} \tag{3.4.4}$$

This implies that $\tilde{\lambda}_n^{\text{naive}}$ is increasing on $[0, M]$.

Finally, consistency and the asymptotic distribution of $\tilde{\lambda}_n^{GS}(t)$ is provided by the next theorem. Its proof can be found in Section 3.6.3.

Theorem 3.4.3. *Suppose that (A1)-(A2) hold. Fix* $t \in (0, \tau_H)$ *and* $\tau^* \in (t, \tau_H)$*. Assume that* λ_0 *is* $m \geqslant 2$ *times continuously differentiable, with* λ_0' *uniformly bounded from below by a strictly positive constant. Assume that* $s \mapsto \Phi(s; \beta_0)$ *is continuous in a neighborhood of* t *and let* k *be* m*-orthogonal satisfying* (1.2.1)*. Let* $\tilde{\lambda}_n^{GS}$ *be the left derivative of the greatest convex minorant on* $[0, \tau^*]$ *of* Λ_n^s *defined in* (3.4.1) *and suppose that* $n^{1/(2m+1)}b \to c > 0$*. Then, for all* $0 < \ell < M < \tau^*$*,*

$$\sup_{s \in [\ell, M]} \left|\tilde{\lambda}_n^{GS}(s) - \lambda_0(s)\right| = O(b^m) + O_p(b^{-1}n^{-1/2}),$$

in probability, and it holds that

$$n^{m/(2m+1)}\left(\tilde{\lambda}_n^{GS}(t) - \lambda_0(t)\right) \xrightarrow{d} N(\mu, \sigma^2),$$

where

$$\mu = \frac{(-c)^m}{m!}\lambda_0^{(m)}(t)\int_{-1}^{1} k(y)y^m\, dy \quad \textit{and} \quad \sigma^2 = \frac{\lambda_0(t)}{c\Phi(t; \beta_0)}\int_{-1}^{1} k(y)^2\, dy.$$

According to Corollary 3.4.1, the naive estimator from (3.4.3) has the same limiting distribution described in Theorem 3.4.3. In this case we recover a result similar to Theorem 3.2 in Wells, 1994. As can be seen, the limiting distribution of the isotonized kernel estimator in Theorem 3.4.3 is completely the same the one for the smoothed MLE and smoothed Grenander-type estimator in Theorem 3.2.5. The following theorem shows that $\tilde{\lambda}_n^{GS}(t)$ is in fact asymptotically equivalent to both these estimators. In particular, this means that the order of smoothing and isotonization for the Grenander-type estimator does not affect the limit behavior. This is in line with the findings in Mammen, 1991 and van der Vaart and van der Laan, 2003.

Theorem 3.4.4. *Suppose that (A1)-(A3) hold. Fix* $t \in (0, \tau_h)$ *and* $\tau^* \in (t, \tau_H)$. *Assume that* λ_0 *is* $m \geqslant 2$ *times continuously differentiable, with* λ_0' *uniformly bounded from below by a strictly positive constant. Assume that* $s \mapsto \Phi(s; \beta_0)$ *is differentiable with a bounded derivative in a neighborhood of* t *and let* k *be* m-*orthogonal satisfying* (1.2.1). *Let* $\tilde{\lambda}_n^{GS}$ *be the left derivative of the greatest convex minorant on* $[0, \tau^*]$ *of* Λ_n^s *defined in* (3.4.1) *and suppose that* $n^{1/(2m+1)}b \to c > 0$. *Let* $\tilde{\lambda}_n^{SG}$ *be the smoothed Grenander-type estimator defined in* (3.2.3). *Then*

$$n^{m/(2m+1)} \left(\hat{\lambda}_n^{GS}(t) - \tilde{\lambda}_n^{SG}(t) \right) \to 0,$$

in probability, and similarly if we replace $\tilde{\lambda}_n^{SG}(t)$ *by the smoothed maximum likelihood estimator* $\hat{\lambda}_n^{SM}(t)$, *defined in* (3.2.2). *This also holds if we replace* $\tilde{\lambda}_n^{GS}(t)$ *with* $\tilde{\lambda}_n^{naive}(t)$, *defined in* (3.4.3).

The proof of Theorem 3.4.4 can be found in Section 3.6.3.

3.5 NUMERICAL RESULTS FOR POINTWISE CONFIDENCE INTERVALS

In this section we illustrate the finite sample performance of the four estimators considered in Sections 3.2, 3.3 and 3.4 by constructing pointwise confidence intervals for the baseline hazard rate. We consider two different procedures: the first one relies on the limit distribution and the second one is a bootstrap based method. In all the simulations we use the triweight kernel function, which means that the degree of smoothness is $m = 2$. The reason for choosing a second-order kernel is that higher order kernels may also take negative values, which then might lead to non monotone estimators for the baseline hazard.

3.5.1 *Asymptotic confidence intervals*

From Theorems 3.2.5, 3.3.5 and 3.4.3, it can be seen that the asymptotic $100(1-\alpha)$%-confidence intervals at the point $t_0 \in (0, \tau_H)$ are of the form

$$\lambda_n^i(t_0) - n^{-2/5}\left\{\widehat{\mu}_n(t_0) \pm \widehat{\sigma}_n(t_0) q_{1-\alpha/2}\right\},$$

where $q_{1-\alpha/2}$ is the $(1-\alpha/2)$ quantile of the standard normal distribution, $\lambda_n^i(t_0)$ is the smooth isotonic estimator at hand ($i \in \{SG, SMLE, GS, MSLE\}$), and $\widehat{\mu}_n(t_0)$, $\widehat{\sigma}_n(t_0)$ are corresponding plug-in estimators of the asymptotic mean and standard deviation, respectively. However, from the expression of the asymptotic mean in Theorems 3.2.5, 3.3.5 and 3.4.3 for $m = 2$, it is obvious that obtaining the plug-in estimators requires estimation of the second derivative of λ_0. Since accurate estimation of derivatives is a hard problem, we choose to avoid it by using undersmoothing. This procedure is to be preferred above bias estimation, because it is computationally more convenient and leads to better results (see also Hall, 1992, Groeneboom and Jongbloed, 2015, Cheng, Hall, and Tu, 2006). Undersmoothing consists of using a bandwidth of a smaller order than the optimal one (in our case $n^{-1/5}$). As a result, the bias of $n^{2/5}(\lambda_n^i(t_0) - \lambda_0(t_0))$, which is of the order $n^{2/5}b^2$ (see (3.2.25)), will converge to zero. On the other hand, the asymptotic variance is $n^{-1/5}b^{-1}\sigma^2$ (see (3.2.28) with $m = 2$). For example, with $b = n^{-1/4}$, asymptotically $n^{2/5}(\lambda_n^{SI}(t_0) - \lambda_0(t_0))$ behaves like a normal distribution with mean of the order $n^{-1/10}$ and variance $n^{1/20}\sigma^2$. Hence, the confidence interval becomes

$$\lambda_n^{SI}(t_0) \pm n^{-3/8}\widehat{\sigma}_n(t_0) q_{1-\alpha/2}, \tag{3.5.1}$$

where

$$\widehat{\sigma}_n(t_0) = \frac{\lambda_n^{SI}(t_0)}{c\Phi_n(t_0; \hat{\beta}_n)} \int_{-1}^{1} k(y)^2 \, dy. \tag{3.5.2}$$

Note that undersmoothing leads to confidence intervals of asymptotic length $O_P(n^{-3/8})$, while the optimal ones would be of length $O_P(n^{-2/5})$. In our simulations, the event times are generated from a Weibull baseline distribution with shape parameter 1.5 and scale 1. The real valued covariate and the censoring time are chosen to be uniformly distributed on the interval $(0, 1)$ and we take $\beta_0 = 0.5$. We note that this setup corresponds to around 35% uncensored observations. Confidence intervals are calculated at the point $t_0 = 0.5$ using 10 000 sets of data and we take bandwidth $b = cn^{-1/4}$, with $c = 1$, and kernel function $k(u) = (35/32)(1-u^2)^3 \mathbb{1}_{\{|u| \leqslant 1\}}$.

It is important to note that the performance depends strongly on the choice of the constant c, because the asymptotic length is inversely proportional to c (see (3.5.2)). This means that, by choosing a smaller c we get wider

	SG		SMLE		MSLE		GS	
n	AL	CP	AL	CP	AL	CP	AL	CP
100	0.996	0.740	1.101	0.796	1.035	0.779	1.007	0.740
500	0.545	0.824	0.563	0.857	0.562	0.840	0.554	0.841
1000	0.421	0.852	0.430	0.883	0.424	0.846	0.428	0.843
5000	0.232	0.910	0.234	0.916	0.234	0.892	0.234	0.888

Table 4: The average length (AL) and the coverage probabilities (CP) for asymptotic 95% pointwise confidence intervals of the baseline hazard rate at $t_0 = 0.5$.

confidence intervals and as a result higher coverage probabilities. Unfortunately, it is not clear which would be the optimal choice of such a constant. This is actually a common problem in the literature (see for example Cheng, Hall, and Tu, 2006 and González-Manteiga, Cao, and Marron, 1996). As indicated in Müller and Wang, 1990, cross-validation methods that consider a trade-off between bias and variance suffer from the fact that the variance of the estimator increases as one approaches the endpoint of the support. This is even enforced in our setting, because the bias is also decreasing when approaching the endpoint of the support. We tried a locally adaptive choice of the bandwidth, as proposed in Müller and Wang, 1990, by minimizing an estimator of the Mean Squared Error, but in our setting this method did not lead to better results. A simple choice is to take c equal to the range of the data (see Groeneboom and Jongbloed, 2015), which in our case corresponds to $c = 1$. Table 4 shows the performance of the four smooth isotonic estimators. For each of them we report the average length (AL) and the coverage probabilities (CP) of the confidence intervals given in (3.5.1) for various sample sizes. Results indicate that the SMLE behaves slightly better, but as the sample size increases its behavior becomes comparable to that of the other estimators. Even though the coverage probabilities are below the nominal level of 95%, smoothing leads to significantly more accurate results in comparison with the non-smoothed Grenander-type estimator given in the last two columns of Table 5. The confidence intervals for the Grenander-type estimator are constructed on the basis of Theorem 2 in Lopuhaä and Nane, 2013, i.e., they are of the form $\tilde{\lambda}_n(t_0) \pm n^{-1/3}\hat{C}_n(t_0)q_{1-\alpha/2}(\mathbb{Z})$, where

$$\hat{C}_n(t_0) = \left(\frac{4\tilde{\lambda}_n(t_0)\tilde{\lambda}_n'(t_0)}{\Phi_n(t_0;\hat{\beta}_n)}\right)^{1/3},$$

$q_\alpha(\mathbb{Z})$ is the α-quantile of the distribution of $\mathbb{Z} = \operatorname{argmin}_{t\in\mathbb{R}}\{W(t) + t^2\}$, with W a standard two-sided Brownian motion starting from zero. In particular, $q_{0.975}(\mathbb{Z}) = 0.998181$. The main advantage of using the non-smoothed

	Kernel		Grenander	
n	AL	CP	AL	CP
100	1.055	0.756	0.757	0.500
500	0.560	0.822	0.449	0.615
1000	0.429	0.845	0.359	0.657
5000	0.234	0.884	0.215	0.764

Table 5: The average length (AL) and the coverage probabilities (CP) for asymptotic 95% pointwise confidence intervals of the baseline hazard rate at $t_0 = 0.5$.

Grenander-type estimator is that it does not involve the choice of a tuning parameter. However, the performance is not satisfactory, because we still need to estimate the derivative of λ_0, which is difficult if the estimator of λ_0 is a step function. Here we use the slope of the segment $[\tilde{\lambda}_n(T_{(i)}, \tilde{\lambda}_n(T_{i+1})]$ on the interval $[T_{(i)}, T_{(i+1)}]$ that contains t_0.

We also compare the performance of the smooth isotonic estimators with that of the ordinary (non-monotone) kernel estimator

$$\tilde{\lambda}_n^s(t_0) = \int k_b(t_0 - u)\, d\Lambda_n(u),$$

which is shown in the first two columns of Table 5. We note that the kernel estimator coincides with the naive estimator that approximates the isotonized kernel estimator, see Section 3.4. In the proof of Theorem 3.4.3, it is shown that $\tilde{\lambda}_n^s$ exhibits a limit distribution which coincides with the one of the smoothed estimators in Theorem 3.2.5. Also the numerical results in Table 5 confirm that the performance of the kernel estimator is comparable with that of the smooth isotonic estimators.

More importantly, estimation of the parameter β_0 has a great effect on the accuracy of the results. Table 6 shows that if we use the true value of β_0 in the computation of the estimators, the coverage probabilities increase significantly. However, in this case the confidence intervals for the SMLE and the MSLE become too conservative. Things are illustrated for the isotonized kernel estimator in Figure 10, which shows the kernel densities of the values of the GS estimator and the corresponding lengths of the confidence intervals, computed using the true parameter β_0 and the partial ML estimator $\hat{\beta}_n$, for 1000 samples of size $n = 500$. We conclude that the use of $\hat{\beta}_n$ leads to underestimation or overestimation of both $\lambda_0(t_0)$ as well as the corresponding length of the confidence interval. In fact, underestimation of both goes hand in hand, since the variance of the GS estimator is proportional to $\lambda_0(t_0)$, and similarly for overestimation. As can be seen in Table 6, estimation of β_0

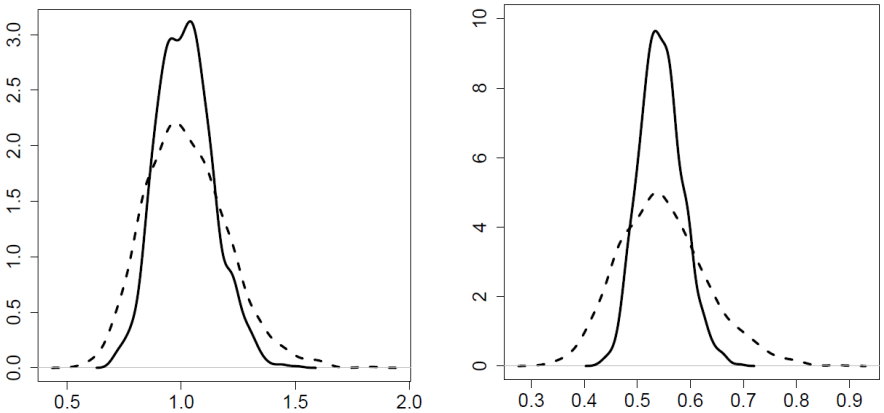


Figure 10: Left panel: Values of the GS estimator computed using the true parameter β_0 (solid line) and the Cox's partial MLE $\hat{\beta}_n$ (dashed line). Right panel: Values of the length of the confidence interval computed using the true parameter β_0 (solid line) and the Cox's partial MLE $\hat{\beta}_n$ (dashed line).

does not seem to effect the length of the confidence interval. However, the coverage probabilities change significantly. When $\lambda_0(t_0)$ is underestimated, the midpoint of the confidence interval lies below $\lambda_0(t_0)$ and the simultaneous underestimation of the length even stronger prevents the confidence interval to cover $\lambda_0(t_0)$. When $\lambda_0(t_0)$ is overestimated, the midpoint of the confidence interval lies above $\lambda_0(t_0)$, but the simultaneous overestimation of the length does not compensate this, so that the confidence interval too often fails to cover $\lambda_0(t_0)$.

Although the partial ML estimator $\hat{\beta}_n$ is a standard estimator for the regression coefficients, the efficiency results are only asymptotic. As pointed out in Cox and Oakes, 1984 and Ren and Zhou, 2011, for finite samples the use of the partial likelihood leads to a loss of accuracy. Recently, Ren and Zhou, 2011 introduced the MLE for β_0 obtained by joint maximization of the loglikelihood in (3.3.1) over both β and λ_0. It was shown that for small and moderate sample sizes, the joint MLE for β_0 performs better than $\hat{\beta}_n$. However, in our case, using this estimator instead of $\hat{\beta}_n$, does not bring any essential difference in the coverage probabilities. Pointwise confidence intervals, for a fixed sample size $n = 500$, at different points of the support are illustrated in Figure 11. The results are again comparable and the common feature is that the length increases as we move to the left boundary. This is due to the fact that the length is proportional to the asymptotic standard deviation, which in this case turns out to be increasing,

	$SMLE_0$		SG_0		$MSLE_0$		GS_0	
n	AL	CP	AL	CP	AL	CP	AL	CP
100	1.057	0.958	0.984	0.941	1.050	0.970	0.981	0.911
500	0.559	0.977	0.538	0.949	0.556	0.979	0.547	0.940
1000	0.430	0.979	0.419	0.957	0.429	0.976	0.424	0.953
5000	0.234	0.981	0.232	0.969	0.234	0.975	0.248	0.960

Table 6: The average length (AL) and the coverage probabilities (CP) for asymptotic 95% pointwise confidence intervals of the baseline hazard rate at the point $t_0 = 0.5$ using β_0.

$\sigma^2(t) = 1.5\sqrt{t}/(c\Phi(t;\beta_0))$. Note that $\Phi(t;\beta_0)$ defined in (3.1.2) is decreasing.

3.5.2 *Bootstrap confidence intervals*

An alternative to confidence intervals based on the asymptotic distribution relies on bootstrap. Studies on bootstrap confidence intervals in the Cox model are investigated in Burr, 1994 and Xu, Sen, and Ying, 2014. In the latter paper, the authors investigate several bootstrap procedures for the Cox model. We will use one (method M5) of the two proposals for a smooth bootstrap that had the best performance and were recommended by the authors.

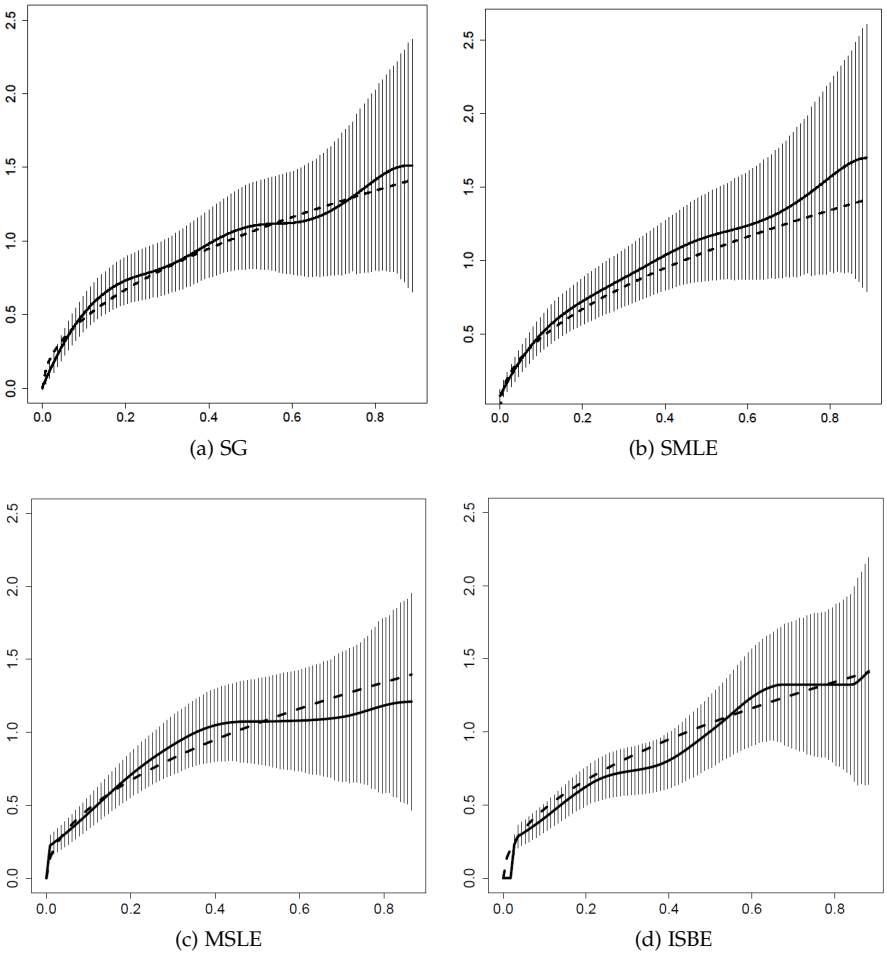
We fix the covariates and we generate the event time X_i^* from a smooth estimate for the cdf of X conditional on Z_i:

$$\hat{F}_n(t|Z_i) = 1 - \exp\left\{-\Lambda_n^s(t)e^{\hat{\beta}_n' Z_i}\right\},$$

where Λ_n^s is the smoothed Breslow estimator

$$\Lambda_n^s(t) = \int k_b(t-u)\Lambda_n(u)\,du.$$

The censoring times C_i^* are generated from the Kaplan-Meier estimate $\hat{G}_n$. Then we take $T_i^* = \min(X_i^*, C_i^*)$ and $\Delta_i^* = \mathbb{1}_{\{X_i^* \leqslant C_i^*\}}$. For constructing the confidence intervals, we take 1000 bootstrap samples (T_i^*, Δ_i^*, Z_i) and for each bootstrap sample we compute the smooth isotonic estimates $\tilde{\lambda}_n^{SG,*}(t_0)$, $\hat{\lambda}_n^{SM,*}(t_0)$, $\hat{\lambda}_n^{MS,*}(t_0)$ and $\hat{\lambda}_n^{GS,*}(t_0)$. Here the kernel function is the same as



(a) SG (b) SMLE

(c) MSLE (d) ISBE

Figure 11: 95% pointwise confidence intervals based on the asymptotic distribution for the baseline hazard rate using undersmoothing.

before and the bandwidth is taken to be $b = n^{-1/5}$. Then, the $100(1-\alpha)\%$ confidence interval is given by

$$\left[q^*_{\alpha/2}(t_0), q^*_{1-\alpha/2}(t_0)\right], \tag{3.5.3}$$

where $q^*_\alpha(t_0)$ is the α-percentile of the 1000 values of the estimates $\tilde{\lambda}^{i,*}_n(t_0)$ of the corresponding smooth isotonic estimator ($i \in \{SG, SMLE, GS, MSLE\}$).

We investigate the behavior of the four estimators in the following two different settings

	Model 1	**Model 2**
X	Weibull $(1.5, 1)$	Weibull $(3, 1)$
C	Uniform $(0, 1)$	Uniform $(0, 2)$
Z	Uniform $(0, 1)$	Bernoulli (0.5)
β_0	0.5	0.1
x_0	0.5	0.5

Note that Model 1 is the same as in the previous simulation. The main differences between the two models are the following: the baseline hazard rate is slightly increasing in Model 1 and strongly increasing in Model 2, the covariates have a smaller effect on the hazard rate in Model 2, and Model 1 corresponds to 35% uncensored observations, while in the Model 2 we have about 50% uncensored observations. It is also worthy noticing that, for Model 1, we calculate the confidence intervals at the middle point of the support $t_0 = 0.5$ in order to avoid boundary problems, while, in Model 2 we again consider $t_0 = 0.5$, because the estimation becomes more problematic on the interval $[1, 2]$. This is probably due to the fact that we only have a few observations in this time interval, on which the hazard rate is strongly increasing.

The average length and the empirical coverage for 1000 iterations and different sample sizes are reported in Table 7 and Table 8 . We observe that bootstrap confidence intervals behave better that confidence intervals constructed on the basis of the asymptotic distribution, i.e., the coverage probabilities are closer to the nominal level of 95%. Results also indicate that the SMLE and MSLE behave slightly better than the other two estimators, in general, it leads to shorter confidence intervals and better coverage probabilities.

	SMLE		SG		MSLE		GS	
n	AL	CP	AL	CP	AL	CP	AL	CP
100	1.870	0.948	1.376	0.899	1.553	0.943	1.625	0.914
500	0.730	0.942	0.660	0.892	0.701	0.947	0.726	0.941
1000	0.521	0.960	0.487	0.902	0.512	0.949	0.527	0.963

Table 7: The average length (AL) and the coverage probabilities (CP) for the 95% bootstrap confidence intervals of the baseline hazard rate at $t_0 = 0.5$ for Model 1.

	SMLE		SG		MSLE		GS	
n	AL	CP	AL	CP	AL	CP	AL	CP
100	0.899	0.966	0.730	0.960	0.766	0.970	0.858	0.975
500	0.387	0.955	0.376	0.969	0.362	0.959	0.395	0.951
1000	0.279	0.954	0.275	0.960	0.271	0.959	0.286	0.962

Table 8: The average length (AL) and the coverage probabilities (CP) for the 95% bootstrap confidence intervals of the baseline hazard rate at $t_0 = 0.5$ for Model 2.

In order to provide some theoretical evidence for the consistency of the method, we would like to establish that, given the data $(T_1, \Delta_1, Z_1), \ldots, (T_n, \Delta_n, Z_n)$, it holds

$$n^{2/5}\left(\lambda_n^{SI,*}(t) - \lambda_n^{SI}(t)\right) \xrightarrow{d} N(\tilde{\mu}, \sigma^2), \tag{3.5.4}$$

for some $\bar{\mu} \in \mathbb{R}$ (possibly different from μ in Theorem 3.2.5) and σ^2 as in (3.2.22), where λ_n^{SI} is one of the smoothed isotonic estimators at hand and $\lambda_n^{SI,*}$ is the same estimator computed for the bootstrap sample. We focus here on the smoothed Grenander-type estimator. In view of Remark 3.2.6, we are able to obtain (3.5.4) for the smoothed Grenander estimator, if $\hat{\beta}_n^* - \hat{\beta}_n \to 0$, for almost all sequences (T_i^*, Δ_i^*, Z_i), $i = 1, 2, \ldots$, conditional on the sequence (T_i, Δ_i, Z_i), $i = 1, 2, \ldots$, and $\sqrt{n}(\hat{\beta}_n^* - \hat{\beta}_n) = O_p^*(1)$. By the latter we mean that for all $\epsilon > 0$, there exists $M > 0$ such that

$$\limsup_{n\to\infty} P_n^*\left(\sqrt{n}|\hat{\beta}_n^* - \hat{\beta}_n| > M\right) < \epsilon, \qquad \mathbb{P} - \text{almost surely.}$$

where P_n^* is the measure corresponding to the distribution of (T^*, Δ^*, Z) conditional on the data $(T_1, \Delta_1, Z_1), \ldots, (T_n, \Delta_n, Z_n)$, with $T^* = (\min(X^*, C^*)$ and $\Delta^* = \mathbb{1}_{\{X^* \leqslant C^*\}}, Z)$, where X^* conditional on Z has distribution function $\hat{F}_n(t \mid Z)$ and C^* has distribution function $\hat{G}_n$. To prove (3.5.4), we mimic the proof of Theorem 3.2.5, which means that one needs to establish the bootstrap versions of Lemmas 3.2.1-3.2.4. A brief sketch of the arguments is provided in Appendix A.3.

Then, we can approximate the distribution of $n^{2/5}(\lambda_0(t_0) - \lambda_n^{SI}(t_0))$ by the distribution of $n^{2/5}(\lambda_n^{SI,*}(t_0) - \lambda_n^{SI}(t_0)) - (\tilde{\mu} + \mu)$. Consequently, we can write

$$\begin{aligned} &P_n^*\{q_{\alpha/2}^*(t_0) \leqslant \lambda_n^{SI,*}(t) \leqslant q_{1-\alpha/2}^*(t_0)\} \\ &= P_n^* \left\{\lambda_0(t_0) \in \left[q_{\alpha/2}^*(t_0) - n^{-2/5}(\tilde{\mu} + \mu), q_{1-\alpha/2}^*(t_0) - n^{-2/5}(\tilde{\mu} + \mu)\right]\right\} \end{aligned}$$

This means that we should actually take

$$[q_{\alpha/2}^*(t_0), q_{1-\alpha/2}^*(t_0)] - n^{-2/5}(\tilde{\mu} + \mu)$$

instead of (3.5.3). The use of (3.5.3) avoids bias estimation. However, since the effect of the bias is of the order $n^{-2/5}$, the results are still satisfactory. In order to further reduce the effect of the bias, we also investigated the possibility of constructing bootstrap confidence intervals with undersmoothing, i.e, we repeat the previous procedure with bandwidth $b = n^{-1/4}$. Results for Model 1 are shown in Table 9. We notice that the length of the confidence interval increases and the coverage probabilities improve. To summarize, also the bootstrap confidence intervals are affected by the choice of the bandwidth, but the results are more satisfactory in comparison with the ones in Table 4.

	SMLE		SG		MSLE		GS	
n	AL	CP	AL	CP	AL	CP	AL	CP
100	1.901	0.954	1.415	0.900	1.659	0.951	1.620	0.922
500	0.749	0.951	0.672	0.918	0.740	0.954	0.735	0.941
1000	0.540	0.950	0.501	0.924	0.543	0.951	0.541	0.939

Table 9: The average length (AL) and the coverage probabilities (CP) for the 95% bootstrap confidence intervals of the baseline hazard rate at $t_0 = 0.5$, using $b = n^{-1/4}$.

3.6 PROOFS

3.6.1 *Proofs for Section 3.2*

Proof of Lemma 3.2.1. Define $D_n^{(1)}(x;\beta) = \partial\Phi_n(x;\beta)/\partial\beta$ and let $D_{nj}^{(1)}(x;\beta)$ be the jth component of $D_n^{(1)}(x;\beta)$, for $j = 1,\dots,p$. Then according to the proof of Lemma 3(iv) in Lopuhaä and Nane, 2013, for any sequence β_n^*, such that $\beta_n^* \to \beta_0$ almost surely, it holds

$$\limsup_{n\to\infty} \sup_{x\in\mathbb{R}} |D_n^{(1)}(x;\beta_n^*)| < \infty.$$

In fact, from its proof, it can be seen that

$$\begin{aligned}\sup_{x\in\mathbb{R}} |D_{nj}^{(1)}(x;\beta_n^*)| &\leqslant \sum_{I_k\subseteq I}\left[\frac{1}{n}\sum_{i=1}^n |Z_i|\, e^{\gamma_k' Z_i}\right] \\ &\to \sum_{I_k\subseteq I} \mathbb{E}\left[|Z|e^{\gamma_k' Z}\right] < 2^p \sup_{|\beta-\beta_0|\leqslant\epsilon} \mathbb{E}\left[|Z|e^{\beta' Z}\right] < \infty\end{aligned}$$

with probability 1, where the summations are over all subsets $I_k = \{i_1,\dots,i_k\}$ of $I = \{1,\dots,p\}$, and γ_k is the vector consisting of coordinates $\gamma_{kj} = \beta_{0j} + \epsilon/(2\sqrt{p})$, for $j \in I_k$, and $\gamma_{kj} = \beta_{0j} - \epsilon/(2\sqrt{p})$, for $j \in I \setminus I_k$. Therefore,

$$\sup_{x\in\mathbb{R}} |D_n^{(1)}(x;\beta_n^*)| \leqslant \sqrt{p}\sum_{I_k\subseteq I}\left(\frac{1}{n}\sum_{i=1}^n |Z_i|\, e^{\gamma_k' Z_i}\right) \to \sqrt{p}\sum_{I_k\subseteq I}\mathbb{E}\left[|Z|e^{\gamma_k' Z}\right]$$

with probability one. Hence, if for some $\xi_1 > 0$,

$$E_{n,1} = \left\{\left|\sqrt{p}\sum_{I_k\subseteq I}\left(\frac{1}{n}\sum_{i=1}^n |Z_i|\, e^{\gamma_k' Z_i}\right) - \sqrt{p}\sum_{I_k\subseteq I}\mathbb{E}\left[|Z|e^{\gamma_k' Z}\right]\right| \leqslant \xi_1\right\}, \tag{3.6.1}$$

then $\mathbb{1}_{E_{n,1}} \to 1$ in probability. Moreover, on this event, we have

$$\sup_{x\in\mathbb{R}} |D_n^{(1)}(x;\beta_n^*)| \leqslant \sqrt{p}\sum_{I_k\subseteq I}\mathbb{E}\left[|Z|e^{\gamma_k' Z}\right] + \xi_1, \tag{3.6.2}$$

i.e., $\sup_{x\in\mathbb{R}} |D_n^{(1)}(x;\beta_n^*)|$ is bounded uniformly in n. For $\xi_2, \xi_3, \xi_4 > 0$ and $0 < M < \tau_H$ define

$$E_{n,2} = \left\{ n^{2/3}|\hat{\beta}_n - \beta_0|^2 < \xi_2 \right\}, \qquad E_{n,3} = \left\{ \sup_{t\in[0,M]} |\hat{\Lambda}_n(t) - \Lambda_0(t)| < \xi_3 \right\},$$
$$E_{n,4} = \left\{ n^{1/3} \sup_{t\in\mathbb{R}} |\Phi_n(t;\beta_0) - \Phi(t;\beta_0)| \leqslant \xi_4 \right\}, \qquad E_{n,5} = \left\{ T_{(n)} > M \right\} \tag{3.6.3}$$

where $T_{(n)}$ denotes the last observed time. Because $\sqrt{n}(\hat{\beta}_n - \beta_0) = O_p(1)$ (see Theorem 3.2 in Tsiatis, 1981), together with (3.1.7) and Lemma 4 in Lopuhaä and Nane, 2013, it follows that $\mathbb{1}_{E_n} \to 1$ in probability, for the event $E_n = E_{n,1} \cap E_{n,2} \cap E_{n,3} \cap E_{n,4} \cap E_{n,5}$.

From the definitions of $a_{n,t}$, $\theta_{n,t}$ and H^{uc}, in (3.2.9), (3.2.10), and (3.1.1), respectively, we have

$$\begin{aligned}
&\int \theta_{n,t}(u,\delta,z)\, d\mathbb{P}(u,\delta,z) \\
&= \mathbb{1}_{E_n} \left\{ \int a_{n,t}(u)\, dH^{uc}(u) - \int e^{\hat{\beta}_n' z} \int_{v=0}^{u} a_{n,t}(v)\, d\hat{\Lambda}_n(v)\, d\mathbb{P}(u,\delta,z) \right\}.
\end{aligned}$$

Then, by applying Fubini's theorem, together with (3.1.3), we obtain

$$\begin{aligned}
&\int \theta_{n,t}(u,\delta,z)\, d\mathbb{P}(u,\delta,z) \\
&= \mathbb{1}_{E_n} \left\{ \int a_{n,t}(u)\, dH^{uc}(u) - \int a_{n,t}(v) \int_{u=v}^{\infty} e^{\hat{\beta}_n' z}\, d\mathbb{P}(u,\delta,z)\, d\hat{\Lambda}_n(v) \right\} \\
&= \mathbb{1}_{E_n} \left\{ \int a_{n,t}(u)\, dH^{uc}(u) - \int a_{n,t}(v)\, \Phi(v;\hat{\beta}_n)\, d\hat{\Lambda}_n(v) \right\} \\
&= \mathbb{1}_{E_n} \left\{ \int \frac{k_b(t-u)}{\Phi(u;\beta_0)}\, dH^{uc}(u) - \int k_b(t-u)\, \frac{\Phi(u;\hat{\beta}_n)}{\Phi(u;\beta_0)}\, d\hat{\Lambda}_n(u) \right\} \\
&= \mathbb{1}_{E_n} \left\{ \int k_b(t-u) \left(1 - \frac{\Phi(u;\hat{\beta}_n)}{\Phi(u;\beta_0)} \right) d\hat{\Lambda}_n(u) - \int k_b(t-u)\, d(\hat{\Lambda}_n - \Lambda_0)(u) \right\}.
\end{aligned}$$

The mean value theorem yields

$$\begin{aligned}
&\int k_b(t-u) \left| 1 - \frac{\Phi(u;\hat{\beta}_n)}{\Phi(u;\beta_0)} \right| d\hat{\Lambda}_n(u) \\
&= \int k_b(t-u)\, \frac{|\Phi(u;\beta_0) - \Phi(u;\hat{\beta}_n)|}{\Phi(u;\beta_0)}\, d\hat{\Lambda}_n(u) \\
&\leqslant |\hat{\beta}_n - \beta_0| \sup_{y\in\mathbb{R}} \left| \frac{\partial \Phi(y;\beta^*)}{\partial \beta} \right| \frac{\tilde{\lambda}_n^{SG}(t)}{\Phi(t+b;\beta_0)},
\end{aligned}$$

with $|\beta^* - \beta_0| \leqslant |\hat{\beta}_n - \beta_0|$. According to Lemma 3(iii) in Lopuhaä and Nane, 2013, for $\epsilon > 0$ from (A2),

$$\sup_{y\in\mathbb{R}} \left| \frac{\partial \Phi(y;\beta^*)}{\partial \beta} \right| < \sup_{y\in\mathbb{R}} \sup_{|\beta-\beta_0|<\epsilon} \left| \frac{\partial \Phi(y;\beta)}{\partial \beta} \right| < \infty.$$

Furthermore, there exists $M < \tau_H$, such that for sufficiently large n we have $t + b \leqslant M$. This yields the following bound $\Phi(t+b;\beta_0) \geqslant \Phi(M;\beta_0) > 0$. Moreover, according to (3.2.4), $\tilde{\lambda}_n^{SG}(t) \to \lambda_0(t)$ with probability one. Since $|\hat{\beta}_n - \beta_0| = O_p(n^{-1/2})$ (see Theorem 3.1 in Tsiatis, 1981), it follows that

$$\mathbb{1}_{E_n} \int k_b(t-u) \left| 1 - \frac{\Phi(u;\hat{\beta}_n)}{\Phi(u;\beta_0)} \right| d\tilde{\Lambda}_n(u) = O_p(n^{-1/2}),$$

which finishes the proof. □

Proof of Lemma 3.2.2. By means of Fubini's theorem

$$\begin{aligned}
&\int \overline{\theta}_{n,t}(u,\delta,z)\, d\mathbb{P}_n(u,\delta,z) \\
&= \int \delta\, \overline{\Psi}_{n,t}(u)\, d\mathbb{P}_n(u,\delta,z) - \int e^{\hat{\beta}_n' z} \int_{v=0}^{u} \overline{\Psi}_{n,t}(v)\, d\hat{\Lambda}_n(v)\, d\mathbb{P}_n(u,\delta,z) \\
&= \int \delta\, \overline{\Psi}_{n,t}(u)\, d\mathbb{P}_n(u,\delta,z) - \int_0^{T_{(n)}} \overline{\Psi}_{n,t}(v) \int \mathbb{1}_{\{u\geqslant v\}} e^{\hat{\beta}_n' z}\, d\mathbb{P}_n(u,\delta,z)\, d\hat{\Lambda}_n(v) \\
&= \mathbb{1}_{E_n} \left\{ \int \delta\, \frac{\overline{a}_{n,t}\overline{\Phi}_n(u;\hat{\beta}_n)}{\Phi_n(u;\hat{\beta}_n)}\, d\mathbb{P}_n(u,\delta,z) \right. \\
&\qquad \left. - \int_0^{\tau_H} \overline{a}_{n,t}\overline{\Phi}_n(v;\hat{\beta}_n)\, d\hat{\Lambda}_n(v) \right\} \\
&= \mathbb{1}_{E_n} \sum_{i=0}^{m} \overline{a}_{n,t}\overline{\Phi}_n(\tau_i;\hat{\beta}_n) \left\{ \int \frac{\mathbb{1}_{[\tau_i,\tau_{i+1})}(u)\delta}{\Phi_n(u;\hat{\beta}_n)}\, d\mathbb{P}_n(u,\delta,z) \right. \\
&\qquad \left. - \Big(\hat{\Lambda}_n(\tau_{i+1}) - \hat{\Lambda}_n(\tau_i) \Big) \right\}.
\end{aligned}$$

Then, (3.2.16) follows immediately from the characterization of the Breslow estimator in (3.1.6). □

To obtain suitable bounds for (3.2.19), we will establish bounds on the tail probabilities of $\hat{U}_n(a)$ defined in (3.2.17). To this end we consider a suitable martingale that will approximate the process $\Lambda_n - \Lambda_0$. For $i = 1, 2, \ldots, n$, let $N_i(t) = \mathbb{1}_{\{X_i \leqslant t\}}\Delta_i$ be the right continuous counting process for the number of observed failures on $(0, t]$ and $Y_i(t) = \mathbb{1}_{\{T_i \geqslant t\}}$ be the

at-risk process. Then, for each $i = 1, 2, \ldots, n$, $M_i(t) = N_i(t) - A_i(t)$, with $A_i(t) = \int_0^t Y_i(s) e^{\beta_0' Z_i} \, d\Lambda_0(s)$, is a mean zero martingale with respect to the filtration

$$\mathcal{F}_t^n = \sigma\left\{ \mathbb{1}_{\{X_i \leqslant s\}} \Delta_i, \mathbb{1}_{\{T_i \geqslant s\}}, Z_i : 1 \leqslant i \leqslant n, 0 \leqslant s \leqslant t \right\}.$$

(e.g., see Kalbfleisch and Prentice, 2002). Furthermore, it is square integrable, since

$$\begin{aligned} \mathbb{E}\left[M_i(t)^2\right] &\leqslant 2 + 2\int_0^t \mathbb{E}\left[\mathbb{1}_{\{T_i \geqslant s\}} e^{2\beta_0' Z_i}\right] \lambda_0^2(s) \, ds \\ &\leqslant 2 + 2\tau_H \, \lambda_0^2(\tau_H) \Phi(0; 2\beta_0) < \infty. \end{aligned}$$

Finally, it has predictable variation process $\langle M_i \rangle = A_i(t)$ (e.g., see Gill, 1984 or Theorem 2 of Appendix B in Shorack and Wellner, 1986). For each $n \geqslant 1$, define

$$\mathbb{N}_n(t) = \sum_{i=1}^n N_i(t), \qquad \mathbb{A}_n(t) = \sum_{i=1}^n A_i(t), \qquad \mathbb{M}_n(t) = \mathbb{N}_n(t) - \mathbb{A}_n(t). \tag{3.6.4}$$

Then $\mathbb{M}_n(t)$ is a mean zero square integrable martingale with predictable variation process

$$\langle \mathbb{M}_n \rangle(t) = \sum_{i=1}^n \langle M_i \rangle(t) = \sum_{i=1}^n \int_0^t \mathbb{1}_{\{T_i \geqslant s\}} \, e^{\beta_0' Z_i} \, d\Lambda_0(s) = \int_0^t n\Phi_n(s; \beta_0) \, d\Lambda_0(s),$$

where Φ_n is defined in (3.1.4).

Lemma 3.6.1. *Suppose that (A1)–(A2) hold. Let $0 < M < \tau_H$ and let Φ be defined in* (3.1.2)*. Then, the process*

$$\mathbb{B}_n(t) = \int_0^{t \wedge M} \frac{1}{n\Phi(s; \beta_0)} \, d\mathbb{M}_n(s) \tag{3.6.5}$$

is a mean zero, square integrable martingale with respect to the filtration $\mathcal{F}_t^n$, Moreover, $\mathbb{B}_n$ has predictable variation process

$$\langle \mathbb{B}_n \rangle(t) = \int_0^{t \wedge M} \frac{\lambda_0(s) \Phi_n(s; \beta_0)}{n\Phi^2(s; \beta_0)} \, ds.$$

Proof. Write

$$\mathbb{B}_n(t) = \int_0^t Y_n(s) \, d\mathbb{M}_n(s), \quad \text{where } Y_n(s) = \frac{\mathbb{1}_{\{s \leqslant M\}}}{n\Phi(s; \beta_0)},$$

and $\mathbb{M}_n = \mathbb{N}_n - \mathbb{A}_n$. We apply Theorem B.3.1c in Shorack and Wellner, 1986 with Y, H, M, N, and A, replaced by $\mathbb{B}_n$, Y_n, $\mathbb{M}_n$, $\mathbb{N}_n$, and $\mathbb{A}_n$, respectively. In order to check the conditions of this theorem, note that Y_n is a predictable process satisfying $|Y_n(t)| < \infty$, almost surely, for all $t \geqslant 0$, and that

$$\begin{aligned}\int_0^t Y_n(s)\,\mathrm{d}A_n(s) &= \sum_{i=1}^n \int_0^t \frac{\mathbb{1}_{\{s\leqslant M\}}}{n\Phi(s;\beta_0)}\mathbb{1}_{\{T_i\geqslant s\}}e^{\beta_0' Z_i}\,\mathrm{d}\Lambda_0(s)\\ &= \int_0^t \frac{\mathbb{1}_{\{s\leqslant M\}}}{\Phi(s;\beta_0)}\Phi_n(s;\beta_0)\,\mathrm{d}\Lambda_0(s) < \infty, \quad \text{a.s..}\end{aligned}$$

Moreover, since for $s \leqslant M$ we have $\Phi(s;\beta_0) \geqslant \Phi(M;\beta_0) > 0$, it follows that

$$\begin{aligned}\mathbb{E}\left[\int_0^\infty Y_n^2(s)\,\mathrm{d}\langle\mathbb{M}_n\rangle(s)\right] &= \mathbb{E}\left[\int_0^\infty \frac{\mathbb{1}_{\{s\leqslant M\}}}{n\Phi^2(s;\beta_0)}\Phi_n(s;\beta_0)\,\mathrm{d}\Lambda_0(s)\right]\\ &\leqslant \frac{\lambda_0(\tau_H)M}{n^2\Phi^2(M;\beta_0)}\sum_{i=1}^n \mathbb{E}\left[e^{\beta_0' Z_i}\right] < \infty,\end{aligned}$$

because of the assumption (A2). It follows from Theorem B.3.1c in Shorack and Wellner, 1986, that $\mathbb{B}_n$ is a square integrable martingale with mean zero and predictable variation process

$$\langle\mathbb{B}_n\rangle(t) = \int_0^t Y_n^2(s)\,\mathrm{d}\langle\mathbb{M}_n\rangle(s) = \int_0^t \frac{\mathbb{1}_{\{s\leqslant M\}}}{n\Phi^2(s;\beta_0)}\Phi_n(s;\beta_0)\,\mathrm{d}\Lambda_0(s),$$

where Φ and Φ_n are defined in (3.1.2) and (3.1.4), respectively. □

It is straightforward to verify that for $t \in [0, M]$ and $M < T_{(n)}$,

$$\Lambda_n(t) - \Lambda_0(t) = \mathbb{B}_n(t) + R_n(t), \tag{3.6.6}$$

where

$$R_n(t) = \int_0^t \frac{\Phi_n(s;\beta_0)}{\Phi(s;\beta_0)}\,\mathrm{d}\Lambda_0(s) - \Lambda_0(t) + \int_0^t \left(\frac{1}{\Phi_n(s;\hat\beta_n)} - \frac{1}{\Phi(s;\beta_0)}\right)\mathrm{d}H_n^{uc}(s), \tag{3.6.7}$$

with

$$H_n^{uc}(s) = \int \delta\mathbb{1}_{\{t\leqslant s\}}\,\mathrm{d}\mathbb{P}_n(t,\delta,z). \tag{3.6.8}$$

For establishing suitable bounds on the tail probabilities of $\hat{U}_n(a)$, we need the following result for the process $\mathbb{B}_n$, which is comparable to condition (A2) in Durot, 2007.

Lemma 3.6.2. *Suppose that (A1)–(A2) hold. Let* $0 < M < \tau_H$ *and let* $\mathbb{B}_n$ *be defined as in* (3.6.5)*. Then, there exists a constant* $C > 0$ *such that, for all* $x > 0$ *and* $t \in [0, M]$*,*

$$\mathbb{E}\left[\sup_{u\in[0,M],|t-u|\leqslant x}\left(\mathbb{B}_n(u)-\mathbb{B}_n(t)\right)^2\right] \leqslant \frac{Cx}{n}.$$

Proof. The proof is similar to that of Theorem 3 in Durot, 2007. First consider the case $t \leqslant u \leqslant t+x$. According to Lemma 3.6.1, $\mathbb{B}_n$ is a martingale. Hence, by Doob's inequality, we have

$$\begin{aligned}
&\mathbb{E}\left[\sup_{u\in[0,M],\, t\leqslant u\leqslant t+x}(\mathbb{B}_n(u)-\mathbb{B}_n(t))^2\right]\\
&\leqslant 4\mathbb{E}\left[\left(\mathbb{B}_n\left((t+x)\wedge M\right)-\mathbb{B}_n(t)\right)^2\right]\\
&= 4\mathbb{E}\left[\mathbb{B}_n\left((t+x)\wedge M\right)^2-\mathbb{B}_n(t)^2\right]\\
&= 4\mathbb{E}\left[\int_t^{(t+x)\wedge M}\frac{\Phi_n(s;\beta_0)\lambda_0(s)}{n\Phi^2(s;\beta_0)}\,ds\right]\\
&\leqslant \frac{4\lambda(M)x}{n\Phi^2(M;\beta_0)}\mathbb{E}\left[\Phi_n(0;\beta_0)\right],
\end{aligned} \tag{3.6.9}$$

where according to (A2),

$$\mathbb{E}\left[\Phi_n(0;\beta_0)\right] = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left[e^{\beta_0' Z_i}\right] \leqslant C,$$

for some $C > 0$. This proves the lemma for the case $t \leqslant u \leqslant t+x$.

For the case $t-x \leqslant u \leqslant t$, we can write

$$\begin{aligned}
&\mathbb{E}\left[\sup_{u\in[0,M],t-x\leqslant u\leqslant t}\left(\mathbb{B}_n(u)-\mathbb{B}_n(t)\right)^2\right]\\
&= \mathbb{E}\left[\sup_{0\vee(t-x)\leqslant u\leqslant t}\left(\mathbb{B}_n(u)-\mathbb{B}_n(t)\right)^2\right]\\
&\leqslant 2\mathbb{E}\left[\left(\mathbb{B}_n(t)-\mathbb{B}_n(0\vee(t-x))\right)^2\right]\\
&\quad + 2\mathbb{E}\left[\sup_{0\vee(t-x)\leqslant u<t}\left(\mathbb{B}_n(u)-\mathbb{B}_n(0\vee(t-x))\right)^2\right].
\end{aligned}$$

Then similar to (3.6.9), the right hand side is bounded by

$$
\begin{aligned}
&2\mathbb{E}\left[\left(\mathbb{B}_n(t)-\mathbb{B}_n(0\vee(t-x))\right)^2\right]+8\mathbb{E}\left[\left(\mathbb{B}_n(t)-\mathbb{B}_n(0\vee(t-x))\right)^2\right]\\
&=10\mathbb{E}\left[\mathbb{B}_n(t)^2-\mathbb{B}_n(0\vee(t-x))^2\right]=10\mathbb{E}\left[\int_{0\vee(t-x)}^{t}\frac{\Phi_n(s;\beta_0)\,\lambda_0(s)}{n\Phi^2(s;\beta_0)}\,ds\right]\\
&\leqslant\frac{10\,\lambda(M)\,x}{n\Phi^2(M;\beta_0)}\,\mathbb{E}\left[\Phi_n(0;\beta_0)\right]\leqslant\frac{Cx}{n},
\end{aligned}
$$

for some $C>0$. This concludes the proof. □

In what follows, let $0<M<\tau_H$. Moreover, let U be the inverse of λ_0 on $[\lambda_0(0),\lambda_0(M)]$, i.e.,

$$
U(a)=\begin{cases}0 & a<\lambda_0(0);\\ \lambda_0^{-1}(a) & a\in[\lambda_0(0),\lambda_0(M)];\\ M & a>\lambda_0(M).\end{cases}
\tag{3.6.10}
$$

Note that U is continuous and differentiable on $(\lambda_0(0),\lambda_0(M))$, but it is different from the inverse of λ_0 on the entire interval $[\lambda_0(0),\lambda_0(\tau_H)]$.

Lemma 3.6.3. *Suppose that (A1)–(A2) hold. Let $0<M<\tau_H$ and let $\hat{U}_n$ and U be defined in* (3.2.17) *and* (3.6.10)*, respectively. Suppose that H^{uc}, defined in* (3.1.1)*, has a bounded derivative h on $[0,M]$ and that λ_0' is bounded below by a strictly positive constant. Then, there exists an event E_n, such that $\mathbb{1}_{E_n}\to 1$ in probability, and a constant K such that, for every $a\geqslant 0$ and $x>0$,*

$$
\mathbb{P}\left(\left\{|\hat{U}_n(a)-U(a)|\geqslant x\right\}\cap E_n\cap\left\{\hat{U}_n(a)\leqslant M\right\}\right)\leqslant K\max\left\{\frac{1}{nx^3},\frac{1}{n^3x^5}\right\},
\tag{3.6.11}
$$

for n sufficiently large.

Note that Lemma 3.6.2 and Lemma 3.6.3 correspond to Theorem 3(i) and Lemma 2 in Durot, 2007. It is useful to spend some words on the restriction to the event $E_n\cap\{\hat{U}_n(a)\leqslant M\}$. The event $\{\hat{U}_n(a)\leqslant M\}$ is implicit in Durot, 2007, because there the Grenander-type estimator is defined by only considering Λ_n on a compact interval not containing the end point of the support. The event E_n is needed in our setup because of the presence of the covariates, which lead to more complicated processes, and because we require (3.2.18) for $p=2$.

Proof of Lemma 3.6.3. First, we note that, from the definition of U and the fact that $\hat{U}_n$ is increasing, it follows that

$$
|\hat{U}_n(a)-U(a)|\leqslant|\hat{U}_n(\lambda_0(0))-U(\lambda_0(0))|,
$$

if $a \leqslant \lambda_0(0)$, and

$$\mathbb{1}_{\{\hat{U}_n(a)\leqslant M\}}|\hat{U}_n(a) - U(a)| \leqslant \mathbb{1}_{\{\hat{U}_n(a)\leqslant M\}}|\hat{U}_n(\lambda_0(M)) - U(\lambda_0(M))|,$$

if $a \geqslant \lambda_0(M)$. Hence, it suffices to prove (3.6.11) only for $a \in [\lambda_0(0), \lambda_0(M)]$. Let E_n be the event from Lemma 3.2.1. We start by writing

$$\begin{aligned}
&\mathbb{P}\left(\left\{|\hat{U}_n(a) - U(a)| \geqslant x\right\} \cap E_n \cap \left\{\hat{U}_n(a) \leqslant M\right\}\right) \\
&\quad = \mathbb{P}\left(\left\{U(a) + x \leqslant \hat{U}_n(a) \leqslant M\right\} \cap E_n\right) + \mathbb{P}\left(\left\{\hat{U}_n(a) \leqslant U(a) - x\right\} \cap E_n\right).
\end{aligned} \tag{3.6.12}$$

First consider the first probability on the right hand side of (3.6.12). It is zero, if $U(a) + x > M$. Otherwise, if $U(a) + x \leqslant M$, then $x \leqslant M$ and

$$\begin{aligned}
&\mathbb{P}\left(\left\{U(a) + x \leqslant \hat{U}_n(a) \leqslant M\right\} \cap E_n\right) \\
&\leqslant \mathbb{P}\Big(\{\Lambda_n(y) - ay \leqslant \Lambda_n(U(a)) - aU(a), \text{for some } y \in [U(a) + x, M]\} \cap E_n\Big) \\
&\leqslant \mathbb{P}\left(\left\{\inf_{y\in[U(a)+x,M]}\Big(\Lambda_n(y) - ay - \Lambda_n(U(a)) + a\,U(a)\Big) \leqslant 0\right\} \cap E_n\right).
\end{aligned}$$

From Taylor's expansion, we obtain

$$\Lambda_0(y) - \Lambda_0(U(a)) \geqslant \big(y - U(a)\big)a + c\big(y - U(a)\big)^2,$$

where $c = \inf_{t\in[0,\tau_F)} \lambda_0'(t)/2 > 0$, so that with (3.6.6), the probability on the right hand side is bounded by

$$\begin{aligned}
\mathbb{P}\Bigg(\Bigg\{\inf_{y\in[U(a)+x,M]}\Big(&\mathbb{B}_n(y) - \mathbb{B}_n(U(a)) + R_n(y) - R_n(U(a)) \\
&+ c(y - U(a))^2\Big) \leqslant 0\Bigg\} \cap E_n\Bigg).
\end{aligned}$$

Let $i \geqslant 0$ be such that $M - U(a) \in [x2^i, x2^{i+1})$ and note that, on the event E_n one has $T_{(n)} \geqslant M$. Therefore, if $U(a) < y \leqslant M$, then $y \leqslant T_{(n)}$ and $U(a) < T_{(n)}$. It follows that the previous probability can be bounded by

$$\sum_{k=0}^{i} \mathbb{P}\left(\left\{\sup_{y\in I_k}\Big(\big|\mathbb{B}_n(y) - \mathbb{B}_n(U(a))\big| + |R_n(y) - R_n(U(a))|\Big) \geqslant c\,x^2\,2^{2k}\right\} \cap E_n\right).$$

where the supremum is taken over $y \in [0, M]$, such that $y - U(a) \in [x2^k, x2^{k+1})$. Using that $\mathbb{P}(X + Y \geqslant \epsilon) \leqslant \mathbb{P}(X \geqslant \epsilon/2) + \mathbb{P}(Y \geqslant \epsilon/2)$, together with the Markov inequality, we can bound this probability by

$$4 \sum_{k=0}^{i} \left(c^2 x^4 2^{4k}\right)^{-1} \mathbb{E}\left[\sup_{\substack{y \leqslant M \\ y - U(a) \in [x2^k, x2^{k+1})}} \left|\mathbb{B}_n(y) - \mathbb{B}_n(U(a))\right|^2\right]$$
$$+ 8 \sum_{k=0}^{i} \left(c^3 x^6 2^{6k}\right)^{-1} \mathbb{E}\left[\sup_{\substack{y < M \\ y - U(a) \in [x2^k, x2^{k+1})}} \mathbb{1}_{E_n} \left|R_n(y) - R_n(U(a))\right|^3\right]. \tag{3.6.13}$$

We have

$$\mathbb{E}\left[\sup_{\substack{y < M \\ y - U(a) \in [x2^k, x2^{k+1})}} \mathbb{1}_{E_n} \left|R_n(y) - R_n(U(a))\right|^3\right]$$
$$\leqslant 4\mathbb{E}\left[\sup_{\substack{y < M \\ y - U(a) \in [x2^k, x2^{k+1})}} \mathbb{1}_{E_n} \left|\int_{U(a)}^{y} \left(\frac{\Phi_n(s;\beta_0)}{\Phi(s;\beta_0)} - 1\right) \lambda_0(s) ds\right|^3\right]$$
$$+ 4\mathbb{E}\left[\sup_{\substack{y < M \\ y - U(a) \in [x2^k, x2^{k+1})}} \mathbb{1}_{E_n} \left|\int_{U(a)}^{y} \left(\frac{1}{\Phi_n(s;\hat{\beta}_n)} - \frac{1}{\Phi(s;\beta_0)}\right) dH_n^{uc}(s)\right|^3\right] \tag{3.6.14}$$

For the first term in the right hand side of (3.6.14) we have

$$\mathbb{E}\left[\sup_{\substack{y < M \\ y - U(a) \in [x2^k, x2^{k+1})}} \mathbb{1}_{E_n} \left|\int_{U(a)}^{y} \left(\frac{\Phi_n(s;\beta_0)}{\Phi(s;\beta_0)} - 1\right) \lambda_0(s) ds\right|^3\right]$$
$$\leqslant \mathbb{E}\left[\mathbb{1}_{E_n} \left(\int_{U(a)}^{(U(a)+x2^{k+1}) \wedge M} \frac{|\Phi_n(s;\beta_0) - \Phi(s;\beta_0)|}{\Phi(s;\beta_0)} \lambda_0(s)\, ds\right)^3\right]$$
$$\leqslant \frac{x^3 2^{3(k+1)} \lambda_0^3(M)}{\Phi(M;\beta_0)^3} \mathbb{E}\left[\mathbb{1}_{E_n} \sup_{s \in [0,M]} |\Phi_n(s;\beta_0) - \Phi(s;\beta_0)|^3\right]$$
$$\leqslant \frac{x^3 2^{3(k+1)} \lambda_0^3(M) \xi_4}{n \Phi(M;\beta_0)^3},$$

where we have used (3.6.3). In order to bound the second term on the right hand side of (3.6.14), note that on the event E_n,

$$\begin{aligned}\sup_{s\in\mathbb{R}}|\Phi_n(s;\hat{\beta}_n)-\Phi(s;\beta_0)| &\leqslant \sup_{s\in\mathbb{R}}|\Phi_n(s;\hat{\beta}_n)-\Phi_n(s;\beta_0)| \\ &\quad+\sup_{s\in\mathbb{R}}|\Phi_n(s;\beta_0)-\Phi(s;\beta_0)| \\ &\leqslant |\hat{\beta}_n-\beta_0|\sup_{s\in\mathbb{R}}|D_n^{(1)}(s;\beta^*)|+\frac{\xi_4}{n^{1/3}} \\ &\leqslant \frac{\sqrt{\xi_2}L+\xi_4}{n^{1/3}}.\end{aligned} \tag{3.6.15}$$

In particular, for sufficiently large n we have

$$\sup_{s\in\mathbb{R}}\left|\Phi_n(s;\hat{\beta}_n)-\Phi(s;\beta_0)\right| \leqslant \Phi(M;\beta_0)/2,$$

which yields that, for $s\in[0,M]$,

$$\Phi_n(s;\hat{\beta}_n)\geqslant\Phi(s;\beta_0)-\frac{1}{2}\Phi(M;\beta_0)\geqslant\frac{1}{2}\Phi(M;\beta_0). \tag{3.6.16}$$

Using (3.6.15), on the event E_n, for n sufficiently large, we can write

$$\begin{aligned}\sup_{s\in[0,M]}\left|\frac{1}{\Phi_n(s;\hat{\beta}_n)}-\frac{1}{\Phi(s;\beta_0)}\right| &\leqslant \sup_{s\in[0,M]}\frac{\left|\Phi_n(s;\hat{\beta}_n)-\Phi(s;\beta_0)\right|}{\Phi_n(s;\hat{\beta}_n)\,\Phi(s;\beta_0)} \\ &\leqslant \frac{2}{\Phi^2(M;\beta_0)}\sup_{s\in[0,M]}\left|\Phi_n(s;\hat{\beta}_n)-\Phi(s;\beta_0)\right| \\ &\leqslant Cn^{-1/3},\end{aligned}$$

for some $C>0$. Consequently, for the second term in the right hand side of (3.6.14) we obtain

$$\begin{aligned}&\mathbb{E}\left[\sup_{\substack{y<M\\ y-U(a)\in[x2^k,x2^{k+1})}}\mathbb{1}_{E_n}\left|\int_{U(a)}^{y}\left(\frac{1}{\Phi_n(s;\hat{\beta}_n)}-\frac{1}{\Phi(s;\beta_0)}\right)dH_n^{uc}(s)\right|^3\right] \\ &\leqslant\frac{C^3}{n}\mathbb{E}\left[\mathbb{1}_{E_n}\left(\frac{1}{n}\sum_{i=1}^{n}\Delta_i\mathbb{1}_{\{T_i\in[U(a),(U(a)+x2^{k+1})\wedge M)]\}}\right)^3\right]\leqslant\frac{C^3}{n^4}\mathbb{E}\left[N^3\right],\end{aligned}$$

where N is a binomial distribution with probability of success

$$\gamma=H^{uc}((U(a)+x2^{k+1})\wedge M))-H^{uc}(U(a))\leqslant\sup_{s\in[0,M]}|h(s)|x2^{k+1}.$$

Furthermore,

$$\mathbb{E}\left[N^3\right] = n\gamma(1-3\gamma+3n\gamma+2\gamma^2-3n\gamma^2+n^2\gamma^2) \leqslant \begin{cases} 7n\gamma & \text{, if } n\gamma \leqslant 1; \\ 7n^3\gamma^3 & \text{, if } n\gamma > 1. \end{cases}$$

Using Lemma 3.6.2 and the bound in (3.6.13), for the first probability on the right hand side of (3.6.12), it follows that there exist $K_1, K_2 > 0$, such that for all $a \geqslant 0$, $n \geqslant 1$ and $x > 0$,

$$\begin{aligned} &\mathbb{P}\left(\left\{U(a)+x \leqslant \hat{U}_n(a) \leqslant M\right\} \cap E_n\right) \\ &\quad \leqslant K_1 \sum_{k=0}^{i} \frac{x2^{k+1}}{nx^4 2^{4k}} + K_2 \sum_{k=0}^{i} \max\left\{\frac{x2^{k+1}}{n^3x^6 2^{6k}}, \frac{x^3 2^{3(k+1)}}{nx^6 2^{6k}}\right\} \\ &\quad \leqslant \frac{2K_1}{nx^3} \sum_{k=0}^{\infty} 2^{-3k} + \max\left\{\frac{2K_2}{n^3x^5} \sum_{k=0}^{\infty} 2^{-5k}, \frac{8K_2}{nx^3} \sum_{k=0}^{\infty} 2^{-3k}\right\} \\ &\quad \leqslant K \max\left\{\frac{1}{nx^3}, \frac{1}{n^3x^5}\right\}. \end{aligned} \tag{3.6.17}$$

We proceed with the second probability on the right hand side of (3.6.12). We can assume $x \leqslant U(a)$, because otherwise $\mathbb{P}\big(\hat{U}_n(a) \leqslant U(a)-x\big) = 0$. We have

$$\begin{aligned} &\mathbb{P}\left(\left\{\hat{U}_n(a) \leqslant U(a)-x\right\} \cap E_n\right) \\ &\leqslant \mathbb{P}\left(\left\{\inf_{y\in[0,U(a)-x]} \left[\Lambda_n(y) - ay - \Lambda_n(U(a)) + a\,U(a)\right] \leqslant 0\right\} \cap E_n\right). \end{aligned}$$

Let $i \geqslant 0$ be such that $U(a) \in [x2^i, x2^{i+1})$. By a similar argument used to obtain the bound (3.6.13), this probability is bounded by

$$\begin{aligned} &4\sum_{k=0}^{i} \left(c^2x^4 2^{4k}\right)^{-1} \mathbb{E}\left[\sup_{\substack{y\leqslant U(a) \\ U(a)-y\in[x2^k, x2^{k+1})}} \left|\mathbb{B}_n(y) - \mathbb{B}_n(U(a))\right|^2\right] \\ &+ 8\sum_{k=0}^{i} \left(c^3x^6 2^{6k}\right)^{-1} \mathbb{E}\left[\sup_{\substack{y\leqslant U(a) \\ U(a)-y\in[x2^k, x2^{k+1})}} \mathbb{1}_{E_n}\left|R_n(y) - R_n(U(a))\right|^3\right]. \end{aligned} \tag{3.6.18}$$

In the same way as in the first case, we also have

$$\mathbb{E}\left[\sup_{\substack{y\leqslant U(a)\\ U(a)-y\in[x2^k,x2^{k+1})}} \mathbb{1}_{E_n}\left|R_n(y)-R_n(U(a))\right|^3\right]$$
$$\leqslant K_2\max\left\{\frac{x2^{k+1}}{n^3},\frac{x^3 2^{3(k+1)}}{n}\right\}.$$

Exactly as in (3.6.17), Lemma 3.6.2 and (3.6.18) imply that

$$\mathbb{P}\left(\left\{\hat{U}_n(a)\leqslant U(a)-x\right\}\cap E_n\right)\leqslant K\max\left\{\frac{1}{nx^3},\frac{1}{n^3x^5}\right\},$$

for some positive constant K. Together with (3.6.12) and (3.6.17), this finishes the proof. □

Lemma 3.6.4. *Suppose that (A1)–(A2) hold. Let $0<\epsilon<M'<M<\tau_H$ and suppose that H^{uc}, defined in* (3.1.1)*, has a bounded derivative h on $[0,M]$. Let $\tilde{\lambda}_n$ be the Grenander-type estimator of a nondecreasing baseline hazard rate λ_0, which is differentiable with λ_0' bounded above and below by strictly positive constants. Let E_n be the event from Lemma 3.2.1 and take ξ_3 in* (3.6.3) *such that*

$$0<\xi_3<\frac{1}{8}\min\left\{(M-M')^2,\epsilon^2\right\}\inf_{t\in[0,\tau_H]}\lambda_0'(t). \tag{3.6.19}$$

Then, there exists a constant C such that, for n sufficiently large,

$$\sup_{t\in[\epsilon,M']}\mathbb{E}\left[n^{2/3}\mathbb{1}_{E_n}\left(\lambda_0(t)-\tilde{\lambda}_n(t)\right)^2\right]\leqslant C.$$

Proof. It is sufficient to prove that there exist some constants $C_1,C_2>0$, such that for each $n\in\mathbb{N}$ and each $t\in(\epsilon,M']$, we have

$$\mathbb{E}\left[n^{2/3}\mathbb{1}_{E_n}\left\{(\tilde{\lambda}_n(t)-\lambda_0(t))_+\right\}^2\right]\leqslant C_1, \tag{3.6.20}$$

$$\mathbb{E}\left[n^{2/3}\mathbb{1}_{E_n}\left\{(\lambda_0(t)-\tilde{\lambda}_n(t))_+\right\}^2\right]\leqslant C_2. \tag{3.6.21}$$

Lets first consider (3.6.20). We will make use of the following result

$$\begin{aligned}
&\mathbb{E}\left[n^{2/3}\mathbb{1}_{E_n}\left\{(\tilde{\lambda}_n(t)-\lambda_0(t))_+\right\}^2\right] \\
&= 2\int_0^\infty \mathbb{P}\left(n^{1/3}\mathbb{1}_{E_n}(\tilde{\lambda}_n(t)-\lambda_0(t)) \geqslant x\right) x\,dx \\
&= 2\int_0^{2\eta} \mathbb{P}\left(n^{1/3}\mathbb{1}_{E_n}(\tilde{\lambda}_n(t)-\lambda_0(t)) \geqslant x\right) x\,dx \\
&\quad + 2\int_{2\eta}^\infty \mathbb{P}\left(n^{1/3}\mathbb{1}_{E_n}(\tilde{\lambda}_n(t)-\lambda_0(t)) \geqslant x\right) x\,dx \\
&\leqslant 4\eta^2 + 2\int_{2\eta}^\infty \mathbb{P}\left(n^{1/3}\mathbb{1}_{E_n}(\tilde{\lambda}_n(t)-\lambda_0(t)) > x/2\right) x\,dx \\
&\leqslant 4\eta^2 + 4\int_{\eta}^\infty \mathbb{P}\left(n^{1/3}\mathbb{1}_{E_n}(\tilde{\lambda}_n(t)-\lambda_0(t)) > x\right) x\,dx
\end{aligned}$$

for a fixed $\eta > 0$. We distinguish between the cases $a + n^{-1/3}x \leqslant \lambda_0(M)$ and $a + n^{-1/3}x > \lambda_0(M)$, where $a = \lambda_0(t)$. We prove that, in the first case, there exist a positive constant C such that for all $t \in (\epsilon, M']$ and $n \in \mathbb{N}$,

$$\mathbb{P}\big(n^{1/3}\mathbb{1}_{E_n}(\tilde{\lambda}_n(t)-\lambda_0(t)) > x\big) \leqslant C/x^3,$$

for all $x \geqslant \eta$, and in the second case $\mathbb{P}\big(n^{1/3}\mathbb{1}_{E_n}(\tilde{\lambda}_n(t)-\lambda_0(t)) > x\big) = 0$. Then (3.6.20) follows immediately.

First, assume $a + n^{-1/3}x \leqslant \lambda_0(M)$. By the switching relation, we get

$$\begin{aligned}
\mathbb{P}\left(n^{1/3}\mathbb{1}_{E_n}(\tilde{\lambda}_n(t)-\lambda_0(t)) > x\right) &= \mathbb{P}\left(\left\{\tilde{\lambda}_n(t) > a + n^{-1/3}x\right\} \cap E_n\right) \\
&= \mathbb{P}\left(\left\{\hat{U}_n(a + n^{-1/3}x) < t\right\} \cap E_n\right).
\end{aligned}$$

Because $a + n^{-1/3}x \leqslant \lambda_0(M)$, we have $U(a + n^{-1/3}x) \geqslant M > t$. Furthermore, $\{\hat{U}_n(a + n^{-1/3}x) < t\} \subset \{\hat{U}_n(a + n^{-1/3}x) < M\}$. Hence, together with Lemma 3.6.3, we can write

$$\begin{aligned}
&\mathbb{P}\left(\left\{\hat{U}_n(a + n^{-1/3}x) < t\right\} \cap E_n\right) \\
&\quad \leqslant \mathbb{P}\Bigg(\left\{\left|U(a + n^{-1/3}x) - \hat{U}_n(a + n^{-1/3}x)\right| > U(a + n^{-1/3}x) - t\right\} \\
&\qquad\qquad\qquad\qquad \cap E_n \cap \left\{\hat{U}_n(a + n^{-1/3}x) < M\right\}\Bigg) \\
&\quad \leqslant K \max\left\{\frac{1}{n\big(U(a + n^{-1/3}x) - t\big)^3}, \frac{1}{n^3\big(U(a + n^{-1/3}x) - t\big)^5}\right\} \leqslant \frac{C}{x^3},
\end{aligned} \tag{3.6.22}$$

because $U(a+n^{-1/3}x)-t = U'(\xi_n)n^{-1/3}x$, for some $\xi_n \in (a, a+n^{-1/3}x)$, where $U'(\xi_n) = \lambda_0'(\lambda_0^{-1}(\xi_n))^{-1} \geqslant 1/\sup_{t\in[0,\tau_H]}\lambda_0'(t) > 0$.

Next, consider the case $a+n^{-1/3}x > \lambda_0(M)$. Note that, we cannot argue as in the previous case, because for $a+n^{-1/3}x > \lambda_0(M)$ we always have $U(a+n^{-1/3}x) = M$, so that we loose the dependence on x. However, if $n^{1/3}(\tilde{\lambda}_n(t)-\lambda_0(t)) > x$, then for each $y > t$, we have

$$\tilde{\Lambda}_n(y)-\tilde{\Lambda}_n(t) \geqslant \tilde{\lambda}_n(t)\,(y-t) > (a+n^{-1/3}x)\,(y-t),$$

where $a = \lambda_0(t)$. In particular for $y = \tilde{M} = M' + (M-M')/2$, we obtain

$$\begin{aligned}
&\mathbb{P}\left\{n^{1/3}\mathbb{1}_{E_n}(\tilde{\lambda}_n(t)-\lambda_0(t)) > x\right\}\\
&\leqslant \mathbb{P}\left(\left\{\tilde{\Lambda}_n(\tilde{M})-\tilde{\Lambda}_n(t) > \left(a+n^{-1/3}x\right)(\tilde{M}-t)\right\}\cap E_n\right)\\
&\leqslant \mathbb{P}\Bigg(\Bigg\{\tilde{\Lambda}_n(\tilde{M})-\tilde{\Lambda}_n(t)-\left(\Lambda_0(\tilde{M})-\Lambda_0(t)\right) >\\
&\qquad\qquad \left(a+n^{-1/3}x\right)(\tilde{M}-t)-\left(\Lambda_0(\tilde{M})-\Lambda_0(t)\right)\Bigg\}\cap E_n\Bigg)\\
&\leqslant \mathbb{P}\left(\left\{2\sup_{s\in[0,M]}|\tilde{\Lambda}_n(s)-\Lambda_0(s)| > \left(a+n^{-1/3}x-\lambda_0(\tilde{M})\right)(\tilde{M}-t)\right\}\cap E_n\right),
\end{aligned} \tag{3.6.23}$$

also using that $\Lambda_0(\tilde{M})-\Lambda_0(t) \geqslant \lambda_0(\tilde{M})(\tilde{M}-t)$. Furthermore, since $a+n^{-1/3}x > \lambda_0(M)$, it follows from (3.6.19) that

$$\left(a+n^{-1/3}x-\lambda_0(\tilde{M})\right)(\tilde{M}-t) \geqslant \frac{1}{4}(M-M')^2 \inf_{x\in[0,\tau_H]}\lambda_0'(x) \geqslant 2\xi_3, \tag{3.6.24}$$

so that, by the definition of ξ_3 in (3.6.3), the probability on the right hand side (3.6.23) is zero. This concludes the proof of (3.6.20).

Next, we have to deal with (3.6.21). Arguing as in the proof of (3.6.20), we obtain

$$\begin{aligned}
&\mathbb{E}\left[n^{2/3}\mathbb{1}_{E_n}\left\{(\lambda_0(t)-\tilde{\lambda}_n(t))_+\right\}^2\right]\\
&\leqslant \eta^2 + 2\int_\eta^\infty \mathbb{P}\left(n^{1/3}\mathbb{1}_{E_n}(\lambda_0(t)-\tilde{\lambda}_n(t)) \geqslant x\right)x\,dx,
\end{aligned}$$

for a fixed $\eta > 0$, where

$$\mathbb{P}\left(n^{1/3}\mathbb{1}_{E_n}(\lambda_0(t)-\tilde{\lambda}_n(t)) \geqslant x\right) = \mathbb{P}\left(\left\{\hat{U}_n(a-n^{-1/3}x) \geqslant t\right\}\cap E_n\right),$$

with $a = \lambda_0(t)$. First of all, we can assume that $a - n^{-1/3}x \geqslant 0$, because otherwise $\mathbb{P}\{\tilde{\lambda}_n(t) \leqslant a - n^{-1/3}x\} = 0$. Since $t = U(a)$, as before, we write

$$\begin{aligned}
&\mathbb{P}\left(\left\{\hat{U}_n(a - n^{-1/3}x) \geqslant t\right\} \cap E_n\right) \\
&\quad \leqslant \mathbb{P}\left(\left\{\left|\hat{U}_n(a - n^{-1/3}x) - U(a - n^{-1/3}x)\right| \geqslant t - U(a - n^{-1/3}x)\right\} \cap E_n\right).
\end{aligned}$$

In order to apply Lemma 3.6.3, we intersect with the event $\hat{U}_n(a - n^{-1/3}x) \leqslant M$. Note that

$$\mathbb{P}\left(\left\{\hat{U}_n(a - n^{-1/3}x) > M\right\} \cap E_n\right) \leqslant \mathbb{P}\left(\left\{\tilde{\lambda}_n(M) \leqslant a - n^{-1/3}x\right\} \cap E_n\right) = 0.$$

This can be seen as follows. If $\tilde{\lambda}_n(M) \leqslant a - n^{-1/3}x$, then for each $y < M$, we have

$$\tilde{\Lambda}_n(M) - \tilde{\Lambda}_n(y) \leqslant \tilde{\lambda}_n(M)(M - y) \leqslant (a - n^{-1/3}x)(M - y).$$

In particular for $y = \tilde{M} = M' + (M - M')/2$, similar to (3.6.23), we obtain

$$\begin{aligned}
&\mathbb{P}\left(\left\{\tilde{\lambda}_n(M) \leqslant a - n^{-1/3}x\right\} \cap E_n\right) \leqslant \\
&\mathbb{P}\left(\left\{2 \sup_{s \in [0,M]} |\tilde{\Lambda}_n(s) - \Lambda_0(s)| \geqslant \left(-a + n^{-1/3}x + \lambda_0(\tilde{M})\right)\left(M - \tilde{M}\right)\right\} \cap E_n\right).
\end{aligned}$$

Because $a = \lambda_0(t) \leqslant \lambda_0(M')$, we can argue as in (3.6.24) and conclude that the probability on the right hand side is zero. It follows that

$$\begin{aligned}
&\mathbb{P}\left(\left\{\hat{U}_n(a - n^{-1/3}x) \geqslant t\right\} \cap E_n\right) \\
&\quad \leqslant \mathbb{P}\Bigg(\left\{\left|\hat{U}_n(a - n^{-1/3}x) - U(a - n^{-1/3}x)\right| \geqslant t - U(a - n^{-1/3}x)\right\} \\
&\qquad\qquad\qquad\qquad \cap E_n \cap \left\{\hat{U}_n(a - n^{-1/3}x) \leqslant M\right\}\Bigg) \\
&\quad \leqslant K \max\left\{\frac{1}{n\left(t - U(a - n^{-1/3}x)\right)^3}, \frac{1}{n^3\left(t - U(a - n^{-1/3}x)\right)^5}\right\}.
\end{aligned}$$

To bound the right hand side, we have to distinguish between the cases $a - n^{-1/3}x > \lambda_0(0)$ and $a - n^{-1/3}x \leqslant \lambda_0(0)$. If $a - n^{-1/3}x > \lambda_0(0)$, then the right hand side is bounded by K/x^3, because $t - U(a - n^{-1/3}x) = U'(\xi_n)n^{-1/3}x$, for some $\xi_n \in (a - n^{-1/3}x, a)$, where

$$U'(\xi_n) = \lambda_0'(\lambda_0^{-1}(\xi_n))^{-1} \geqslant \frac{1}{\sup_{t \in [0,\tau_H]} \lambda_0'(t)} > 0.$$

Otherwise, if $a - n^{-1/3}x \leqslant \lambda_0(0)$, then we are done because then

$$\mathbb{P}\big(n^{1/3}\mathbb{1}_{E_n}(\lambda_0(t) - \tilde{\lambda}_n(t)) \geqslant x\big) = 0.$$

This can be seen as follows. When $a - n^{-1/3}x \leqslant \lambda_0(0)$, then for each $y < t$, we have

$$\hat{\Lambda}_n(t) - \hat{\Lambda}_n(y) \leqslant \hat{\Lambda}_n(t)(t-y) \leqslant (a - n^{-1/3}x)(t-y).$$

In particular, for $y = \epsilon' = \epsilon/2$, we obtain

$$\begin{aligned}
&\mathbb{P}\left(n^{1/3}\mathbb{1}_{E_n}\left(\lambda_0(t) - \tilde{\lambda}_n(t)\right) \geqslant x\right) \\
&\leqslant \mathbb{P}\left(\left\{\hat{\Lambda}_n(t) - \hat{\Lambda}_n(\epsilon') \leqslant \left(a - n^{-1/3}x\right)(t-\epsilon')\right\} \cap E_n\right) \\
&\leqslant \mathbb{P}\Bigg(\Bigg\{\hat{\Lambda}_n(t) - \hat{\Lambda}_n(\epsilon') - \left(\Lambda_0(t) - \Lambda_0(\epsilon')\right) \\
&\qquad\qquad \leqslant \left(a - n^{-1/3}x\right)(t-\epsilon') - \left(\Lambda_0(t) - \Lambda_0(\epsilon')\right)\Bigg\} \cap E_n\Bigg) \\
&\leqslant \mathbb{P}\left(\left\{2 \sup_{s\in[0,M]} |\hat{\Lambda}_n(s) - \Lambda_0(s)| \geqslant \left(-a + n^{-1/3}x + \lambda_0(\epsilon')\right)(t-\epsilon')\right\} \cap E_n\right).
\end{aligned}$$

Because $a - n^{-1/3}x \leqslant \lambda_0(0)$, we can argue as in (3.6.24),

$$\begin{aligned}
\left(-a + n^{-1/3}x + \lambda_0(\epsilon')\right)(t-\epsilon') &\geqslant \left(\lambda_0(\epsilon') - \lambda_0(0)\right)(\epsilon - \epsilon') \\
&\geqslant \frac{1}{4}\epsilon^2 \inf_{x\in[0,\tau_H]} \lambda_0'(x) \geqslant 2\xi_3.
\end{aligned} \tag{3.6.25}$$

and conclude that the probability on the right hand side is zero. This concludes the proof of (3.6.21). □

We also need a slightly stronger version of Lemma 3.6.4 which is used to prove Lemma 3.2.4 and to show that the distance between the jump times is of smaller order than b (see Lemma 3.6.6). Note that, in order to have the uniform result in (3.6.26), we loose a factor $n^{2/9}$ with respect to the bound in Lemma 3.6.4. This might not be optimal, but it is sufficient for our purposes.

Lemma 3.6.5. *Suppose that (A1)–(A2) hold. Let $0 < \epsilon < M' < M < \tau_H$ and suppose that H^{uc}, defined in (3.1.1), has a bounded derivative h on $[0, M]$. Let $\tilde{\lambda}_n$ be the Grenander-type estimator of a nondecreasing baseline hazard rate λ_0, which is differentiable with λ_0' bounded above and below by strictly positive constants. Let*

E_n *be the event from Lemma 3.2.1 and assume that* ξ_3 *satisfies* (3.6.19). *Then, there exists a constant* $C > 0$, *such that, for each* $n \in \mathbb{N}$,

$$\mathbb{E}\left[n^{4/9}\mathbb{1}_{E_n}\sup_{t\in(\epsilon,M']}\left(\lambda_0(t)-\tilde{\lambda}_n(t)\right)^2\right] \leqslant C. \tag{3.6.26}$$

Proof. We decompose $(\epsilon, M']$ in m intervals $(c_k, c_{k+1}]$, where

$$c_k = \epsilon + k\frac{M'-\epsilon}{m}, \quad \text{for } k = 0, 1, \ldots, m,$$

and $m = (M'-\epsilon)n^{2/9}$. Then, we have

$$\sup_{t\in(\epsilon,M']}\left(\lambda_0(t)-\tilde{\lambda}_n(t)\right)^2 = \max_{0\leqslant k\leqslant m-1}\sup_{t\in(c_k,c_{k+1}]}\left(\lambda_0(t)-\tilde{\lambda}_n(t)\right)^2.$$

On the other hand, the fact that λ_0 is differentiable with bounded derivative implies that

$$\begin{aligned}
&\sup_{t\in(c_k,c_{k+1}]}\left(\lambda_0(t)-\tilde{\lambda}_n(t)\right)^2 \\
&\leqslant 2\sup_{t\in(c_k,c_{k+1}]}\left(\lambda_0(t)-\lambda_0(c_{k+1})\right)^2 + 2\sup_{t\in(c_k,c_{k+1}]}\left(\lambda_0(c_{k+1})-\tilde{\lambda}_n(t)\right)^2 \\
&\leqslant 2\left(\sup_{u\in[0,M']}\lambda_0'(u)\right)^2\left(c_k-c_{k+1}\right)^2 \\
&\qquad + 2\max\left\{\left(\lambda_0(c_{k+1})-\tilde{\lambda}_n(c_k)\right)^2, \left(\lambda_0(c_{k+1})-\tilde{\lambda}_n(c_{k+1})\right)^2\right\} \\
&\leqslant 2\left(\sup_{u\in[0,M']}\lambda_0'(u)\right)^2\frac{(M'-\epsilon)^2}{m^2} \\
&\qquad + \max_{0\leqslant k\leqslant m}\left(\lambda_0(c_k)-\tilde{\lambda}_n(c_k)\right)^2 + \max_{0\leqslant k\leqslant m-1}\left(\lambda_0(c_{k+1})-\lambda_0(c_k)\right)^2.
\end{aligned}$$

Here we used that $\tilde{\lambda}_n$ is non-decreasing, and therefore

$$\sup_{t\in(c_k,c_{k+1}]}\left(\lambda_0(c_{k+1})-\tilde{\lambda}_n(t)\right)^2$$

is achieved either at $t = c_k$ or $t = c_{k+1}$, for $k = 0, 1, \ldots, m-1$. Hence,

$$\begin{aligned}
&\sup_{t\in(\epsilon,M']}\left(\lambda_0(t)-\tilde{\lambda}_n(t)\right)^2 \\
&\leqslant 4\max_{0\leqslant k\leqslant m}\left(\lambda_0(c_k)-\tilde{\lambda}_n(c_k)\right)^2 + 6\left(\sup_{u\in[0,M']}\lambda_0'(u)\right)^2\frac{(M'-\epsilon)^2}{m^2} \\
&\leqslant 4\max_{0\leqslant k\leqslant m}\left(\lambda_0(c_k)-\tilde{\lambda}_n(c_k)\right)^2 + C_1 n^{-4/9},
\end{aligned}$$

where $C_1 = 6\big(\sup_{u\in[\epsilon,M']}\lambda_0'(u)\big)^2$. Consequently, using Lemma 3.6.4, we derive

$$\begin{aligned}
&\mathbb{E}\left[n^{4/9}\mathbb{1}_{E_n}\sup_{t\in(\epsilon,M']}\big(\lambda_0(t)-\tilde{\lambda}_n(t)\big)^2\right]\\
&\leqslant 4\mathbb{E}\left[n^{4/9}\mathbb{1}_{E_n}\max_{0\leqslant k\leqslant m}\big(\lambda_0(c_k)-\tilde{\lambda}_n(c_k)\big)^2\right]+C_1\\
&\leqslant 4n^{-2/9}\sum_{k=0}^{m}\mathbb{E}\left[n^{2/3}\mathbb{1}_{E_n}\big(\lambda_0(c_k)-\tilde{\lambda}_n(c_k)\big)^2\right]+C_1\\
&\leqslant 4\left(M'-\epsilon+1\right)C+C_1.
\end{aligned}$$

This concludes the proof of (3.6.26). □

Lemma 3.6.6. *Under the assumption of Lemma 3.6.5, if $\tau_1,\ldots,\tau_m$ are jump times of $\tilde{\lambda}_n$ on the interval $[\epsilon,M']$ then*

$$\max_{i=1,\ldots,m-1}|\tau_i-\tau_{i+1}| = O_P(n^{-2/9}).$$

Proof. Since $\mathbb{1}_{E_n^c}\max_{i=1,\ldots,m-1}|\tau_i-\tau_{i+1}| = o_P(1)$, it is sufficient to consider $\mathbb{1}_{E_n}\max_{i=1,\ldots,m-1}|\tau_i-\tau_{i+1}|$. By the Markov inequality we have

$$\begin{aligned}
&\mathbb{P}\left(n^{2/9}\mathbb{1}_{E_n}\max_{i=1,\ldots,m-1}|\tau_i-\tau_{i+1}|>K\right)\\
&\leqslant\frac{1}{K^2}\mathbb{E}\left[n^{4/9}\mathbb{1}_{E_n}\max_{i=1,\ldots,m-1}|\tau_i-\tau_{i+1}|^2\right].
\end{aligned}\tag{3.6.27}$$

From the mean value theorem and the boundedness of λ_0', it follows that,

$$|\tau_i-\tau_{i+1}|\leqslant C|\lambda_0(\tau_i)-\lambda_0(\tau_{i+1})|,\quad i=1,\ldots,m-1.$$

Now we consider three possible cases separately. First, if $\tilde{\lambda}_n(\tau_{i+1})<\lambda_0(u)$ for each $u\in(\tau_i,\tau_{i+1}]$, then

$$|\lambda_0(\tau_i)-\lambda_0(\tau_{i+1})|\leqslant|\lambda_0(\tau_{i+1})-\tilde{\lambda}_n(\tau_{i+1})|$$

and as a result

$$|\tau_i-\tau_{i+1}|\leqslant C|\lambda_0(\tau_{i+1})-\tilde{\lambda}_n(\tau_{i+1})|.$$

On the other hand, if $\tilde{\lambda}_n(\tau_{i+1})>\lambda_0(u)$ for each $u\in(\tau_i,\tau_{i+1}]$, then

$$|\lambda_0(\tau_i)-\lambda_0(\tau_{i+1})|\leqslant\lim_{t\downarrow\tau_i}|\lambda_0(t)-\tilde{\lambda}_n(t)|$$

and consequently

$$|\tau_i-\tau_{i+1}|\leqslant C\lim_{t\downarrow\tau_i}|\lambda_0(t)-\tilde{\lambda}_n(t)|.$$

The last case is when $\tilde{\lambda}_n(s) > \lambda_0(s)$ for some $s \in (\tau_i, \tau_{i+1}]$. Then we have

$$|\lambda_0(\tau_i) - \lambda_0(\tau_{i+1})| \leqslant |\lambda_0(\tau_{i+1}) - \tilde{\lambda}_n(\tau_{i+1})| + \lim_{t \downarrow \tau_i} |\lambda_0(t) - \tilde{\lambda}_n(t)|$$

and it follows that

$$|\tau_i - \tau_{i+1}| \leqslant C \left\{ |\lambda_0(\tau_{i+1}) - \tilde{\lambda}_n(\tau_{i+1})| + \lim_{t \downarrow \tau_i} |\lambda_0(t) - \tilde{\lambda}_n(t)| \right\}.$$

In other words, it always holds

$$|\tau_i - \tau_{i+1}| \leqslant 2C \sup_{t \in [\epsilon, M']} |\lambda_0(t) - \tilde{\lambda}_n(t)|. \tag{3.6.28}$$

Therefore, (3.6.27) and Lemma 3.6.5 yield

$$\begin{aligned}
&\mathbb{P}\left(n^{2/9} \mathbb{1}_{E_n} \max_{i=1,\ldots,m-1} |\tau_i - \tau_{i+1}| > K \right) \\
&\leqslant \frac{4C^2}{K^2} \mathbb{E}\left[n^{4/9} \mathbb{1}_{E_n} \sup_{t \in [\epsilon, M']} |\lambda_0(t) - \tilde{\lambda}_n(t)| \right] \\
&\leqslant \frac{C'}{K},
\end{aligned}$$

which concludes the proof. □

Lemma 3.6.7. *Suppose that (A1)–(A2) hold. Fix* $t \in (0, \tau_h)$. *Let* $0 < \epsilon < x < M' < M < \tau_H$ *and suppose that* H^{uc}, *defined in* (3.1.1), *has a bounded derivative* h *on* $[0, M]$. *Let* $\tilde{\lambda}_n$ *be the Grenander-type estimator of a nondecreasing baseline hazard rate* λ_0, *which is differentiable with* λ_0' *bounded above and below by strictly positive constants. Let* E_n *be the event from Lemma 3.2.1 and assume that* ξ_3 *satisfies* (3.6.19). *Let* j_{n1} *and* j_{n2} *be defined as in* (3.2.14). *Then*

$$\mathbb{1}_{E_n} \int_{\tau_{j_{n1}}}^{\tau_{j_{n2}}} (\lambda_0(u) - \tilde{\lambda}_n(u))^2 du = O_p(bn^{-2/3}).$$

Proof. Define the event $A_n = \{|(t-b) - \tau_{j_{n1}}| < b \wedge |\tau_{j_{n2}} - (t+b)| < b\}$. From the definition of $\tau_{j_{n1}}$ and $\tau_{j_{n2}}$ and Lemma 3.6.6, it follows that $\mathbb{P}(A_n^c) \to 0$. We start by writing

$$\begin{aligned}
&\mathbb{P}\left(b^{-1} n^{2/3} \mathbb{1}_{E_n} \int_{\tau_{j_{n1}}}^{\tau_{j_{n2}}} (\lambda_0(u) - \tilde{\lambda}_n(u))^2 \, du > K \right) \\
&\leqslant \mathbb{P}\left(b^{-1} n^{2/3} \mathbb{1}_{E_n \cap A_n} \int_{\tau_{j_{n1}}}^{\tau_{j_{n2}}} (\lambda_0(u) - \tilde{\lambda}_n(u))^2 \, du > K \right) + \mathbb{P}(A_n^c) \\
&\leqslant \mathbb{P}\left(b^{-1} n^{2/3} \mathbb{1}_{E_n} \int_{t-2b}^{t+2b} (\lambda_0(u) - \tilde{\lambda}_n(u))^2 \, du > K \right) + o(1).
\end{aligned}$$

Moreover, Markov's inequality and Fubini, yield

$$\begin{aligned}
&\mathbb{P}\left(b^{-1}n^{2/3}\mathbb{1}_{E_n}\int_{t-2b}^{t+2b}(\lambda_0(u)-\tilde{\lambda}_n(u))^2\,du > K\right)\\
&\leqslant \frac{1}{K}\mathbb{E}\left[b^{-1}n^{2/3}\mathbb{1}_{E_n}\int_{t-2b}^{t+2b}(\lambda_0(u)-\tilde{\lambda}_n(u))^2\,du\right]\\
&\leqslant \frac{2}{K}\sup_{u\in[t-2b,t+2b]}\mathbb{E}\left[n^{2/3}\mathbb{1}_{E_n}(\lambda_0(u)-\tilde{\lambda}_n(u))^2\right].
\end{aligned}$$

For n sufficiently large $[t-2b, t+2b] \subset [\epsilon, M']$, so that according to Lemma 3.6.4, the right hand side is bounded by $2C/K$, for some constant $C > 0$. This proves the lemma. □

Proof of Lemma 3.2.3. Take $t < M < \tau_H$ and n sufficiently large such that $t + b \leqslant M$. With $\overline{a}_{n,t}\overline{\Phi}_n$ defined in (3.2.11), we have

$$\begin{aligned}
&\int\left\{\overline{\theta}_{n,t}(u,\delta,z)-\theta_{n,t}(u,\delta,z)\right\}\,d\mathbb{P}(u,\delta,z)\\
&\quad= \mathbb{1}_{E_n}\int \mathbb{1}_{[\tau_{j_{n1}},\tau_{j_{n2}}]}(u)\delta\left(\frac{\overline{a}_{n,t}\overline{\Phi}_n(u;\hat{\beta}_n)}{\Phi_n(u;\hat{\beta}_n)}-a_{n,t}(u)\right)d\mathbb{P}(u,\delta,z)\\
&\qquad-\mathbb{1}_{E_n}\int e^{\hat{\beta}_n' z}\int_0^u\left(\overline{\Psi}_{n,t}(v)-a_{n,t}(v)\right)d\,\hat{\Lambda}_n(v)\,d\mathbb{P}(u,\delta,z)
\end{aligned}$$

Using Fubini, the definition (3.1.2) of Φ, and the fact that $\overline{a}_{n,t}\overline{\Phi}_n$ and $a_{n,t}$ are zero outside $[\tau_{j_{n1}}, \tau_{j_{n2}}]$, we obtain

$$\begin{aligned}
&\int\left\{\overline{\theta}_{n,t}(u,\delta,z)-\theta_{n,t}(u,\delta,z)\right\}\,d\mathbb{P}(u,\delta,z)\\
&= \mathbb{1}_{E_n}\int_{\tau_{j_{n1}}}^{\tau_{j_{n2}}}\left(\frac{\overline{a}_{n,t}\overline{\Phi}_n(u;\hat{\beta}_n)}{\Phi_n(u;\hat{\beta}_n)}-a_{n,t}(u)\right)dH^{uc}(u)\\
&\qquad-\mathbb{1}_{E_n}\int_{\tau_{j_{n1}}}^{\tau_{j_{n2}}}\left(\frac{\overline{a}_{n,t}\overline{\Phi}_n(v;\hat{\beta}_n)}{\Phi_n(v;\hat{\beta}_n)}-a_{n,t}(v)\right)\Phi(v;\hat{\beta}_n)\,d\hat{\Lambda}_n(v).
\end{aligned}\tag{3.6.29}$$

Write $\widehat{\Phi}_n(u) = \Phi_n(u;\hat{\beta}_n)$, $\widehat{\Phi}(u) = \Phi(u;\hat{\beta}_n)$, and $\Phi_0(u) = \Phi(u;\beta_0)$. Then the right hand side can be written as

$$\mathbb{1}_{E_n}\int_{\tau_{j_{n1}}}^{\tau_{j_{n2}}}\frac{a_{n,t}(\hat{\Lambda}_n(u))\widehat{\Phi}_n(\hat{\Lambda}_n(u))-a_{n,t}(u)\widehat{\Phi}_n(u)}{\widehat{\Phi}_n(u)}\left(\Phi_0(u)\lambda_0(u)-\widehat{\Phi}(u)\tilde{\lambda}_n(u)\right)du$$

where $\hat{A}_n(u)$ is defined in (3.2.12). The Cauchy-Schwarz inequality then yields

$$\begin{aligned}
&\left|\int \left\{\overline{\theta}_{n,t}(u,\delta,z) - \theta_{n,t}(u,\delta,z)\right\} d\mathbb{P}(u,\delta,z)\right| \\
&\leqslant \mathbb{1}_{E_n} \left\| \frac{(a_{n,t}\circ\hat{A}_n)(\widehat{\Phi}_n\circ\hat{A}_n) - a_{n,t}\widehat{\Phi}_n}{\widehat{\Phi}_n} \mathbb{1}_{[\tau_{j_{n1}},\tau_{j_{n2}}]} \right\|_{\mathcal{L}_2} \\
&\quad \times \left\| \left(\Phi_0\lambda_0 - \widehat{\Phi}\tilde{\lambda}_n\right) \mathbb{1}_{[\tau_{j_{n1}},\tau_{j_{n2}}]} \right\|_{\mathcal{L}_2}.
\end{aligned} \tag{3.6.30}$$

Furthermore,

$$\begin{aligned}
&\mathbb{1}_{E_n} \left\| \frac{(a_{n,t}\circ\hat{A}_n)(\widehat{\Phi}_n\circ\hat{A}_n) - a_{n,t}\widehat{\Phi}_n}{\widehat{\Phi}_n} \mathbb{1}_{[\tau_{j_{n1}},\tau_{j_{n2}}]} \right\|_{\mathcal{L}_2} \\
&\leqslant 2\mathbb{1}_{E_n} \int_{\tau_{j_{n1}}}^{\tau_{j_{n2}}} \left(\frac{k_b(t-\hat{A}_{n,t}(u))}{\Phi_0(\hat{A}_n(u))} - \frac{k_b(t-u)}{\Phi_0(u)} \right)^2 du \\
&\quad + 2\mathbb{1}_{E_n} \int_{\tau_{j_{n1}}}^{\tau_{j_{n2}}} \left(\frac{k_b(t-\hat{A}_n(u))}{\Phi_0(\hat{A}_n(u))} \right)^2 \frac{\left(\widehat{\Phi}_n(\hat{A}_n(u)) - \widehat{\Phi}_n(u)\right)^2}{\widehat{\Phi}_n(u)^2} du
\end{aligned}$$

Then, using the boundedness of k', $d\Phi(x;\beta_0)/dx$ and $1/\Phi(x;\beta_0)$ on $[0,\tau_{j_n^2}] \subseteq [0,M]$, we obtain

$$\begin{aligned}
&\mathbb{1}_{E_n} \left\| \frac{(a_{n,t}\circ\hat{A}_n)(\widehat{\Phi}_n\circ\hat{A}_n) - a_{n,t}\widehat{\Phi}_n}{\widehat{\Phi}_n} \mathbb{1}_{[\tau_{j_{n1}},\tau_{j_{n2}}]} \right\|_{\mathcal{L}_2} \\
&\leqslant 2\mathbb{1}_{E_n} \int_{\tau_{j_{n1}}}^{\tau_{j_{n2}}} \left(\frac{d}{dy} \frac{k_b(t-y)}{\Phi_0(y)} \bigg|_{y=\xi_u} \right)^2 \left(\hat{A}_n(u) - u\right)^2 du \\
&\quad + \mathbb{1}_{E_n} \frac{c_1}{b^2\widehat{\Phi}_n(M)^2} \int_{\tau_{j_{n1}}}^{\tau_{j_{n2}}} \left(\widehat{\Phi}_n(\hat{A}_n(u)) - \widehat{\Phi}_n(u)\right)^2 du \\
&\leqslant \mathbb{1}_{E_n} \frac{c_2}{b^4} \int_{\tau_{j_{n1}}}^{\tau_{j_{n2}}} \left(\hat{A}_n(u) - u\right)^2 du \\
&\quad + \mathbb{1}_{E_n} \frac{c_1}{b^2\widehat{\Phi}_n(M)^2} \int_{\tau_{j_{n1}}}^{\tau_{j_{n2}}} \left(\widehat{\Phi}_n(\hat{A}_n(u)) - \widehat{\Phi}_n(u)\right)^2 du,
\end{aligned} \tag{3.6.31}$$

for some constants $c_1, c_2 > 0$. Then, since $\lambda_0(u) - \lambda_0(\hat{A}_n(u)) = \lambda_0'(\xi)(u - \hat{A}_n(u))$ and λ_0' is bounded and strictly positive on $[0,M] \supseteq [\tau_{j_n^1}, \tau_{j_n^2}]$, there exists a constant $K > 0$ such that

$$|u - \hat{A}_n(u)| \leqslant K\,|\lambda_0(u) - \lambda_0(\hat{A}_n(u))|.$$

If $u \in [\tau_i, \tau_{i+1})$ and $\hat{A}_n(u) < \tau_{i+1}$, then $\tilde{\lambda}_n(u) = \tilde{\lambda}_n(\hat{A}_n(u))$ and we obtain

$$\begin{aligned} |u - \hat{A}_n(u)| &\leqslant K|\lambda_0(u) - \tilde{\lambda}_n(u)| + K|\tilde{\lambda}_n(\hat{A}_n(u)) - \lambda_0(\hat{A}_n(u))| \\ &\leqslant 2K|\lambda_0(u) - \tilde{\lambda}_n(u)|. \end{aligned} \tag{3.6.32}$$

This holds also in the case $\hat{A}_n(u) = \tau_{i+1}$, simply because $|\lambda_0(u) - \lambda_0(\hat{A}_n(u))| \leqslant |\lambda_0(u) - \tilde{\lambda}_n(u)|$. As a result, using Lemma 3.6.7, for the first term on the right hand side of (3.6.31), we derive that

$$\begin{aligned} \mathbb{1}_{E_n} \frac{1}{b^4} \int_{\tau_{j_{n1}}}^{\tau_{j_{n2}}} \left(\hat{A}_n(u) - u\right)^2 du &\leqslant \frac{C}{b^4} \mathbb{1}_{E_n} \int_{\tau_{j_{n1}}}^{\tau_{j_{n2}}} \left(\lambda_0(u) - \tilde{\lambda}_n(u)\right)^2 du \\ &= O_p(b^{-3} n^{-2/3}). \end{aligned} \tag{3.6.33}$$

For the second term on the right hand side of (3.6.31), we find

$$\begin{aligned} &\left|\Phi_n(\hat{A}_n(u); \hat{\beta}_n) - \Phi_n(u; \hat{\beta}_n)\right| \\ &\leqslant 2 \sup_{x \in \mathbb{R}} \left|\Phi_n(x; \hat{\beta}_n) - \Phi_n(x; \beta_0)\right| + \left|\Phi(\hat{A}_n(u); \beta_0) - \Phi(u; \beta_0)\right| \\ &\quad + 2 \sup_{x \in \mathbb{R}} |\Phi_n(x; \beta_0) - \Phi(x; \beta_0)| \\ &\leqslant 2|\hat{\beta}_n - \beta_0| \sup_{x \in \mathbb{R}} \left|D_n^{(1)}(x; \beta^*)\right| + |\Phi'(\xi; \beta_0)||\hat{A}_n(u) - u| \\ &\quad + 2 \sup_{x \in \mathbb{R}} |\Phi_n(x; \beta_0) - \Phi(x; \beta_0)|, \end{aligned}$$

which, using (3.6.2), (3.1.5) and $|\hat{\beta}_n - \beta_0| = O_p(n^{-1/2})$ (see Theorem 3.2 in Tsiatis, 1981), leads to

$$\begin{aligned} &b^{-2} \mathbb{1}_{E_n} \int_{\tau_{j_{n1}}}^{\tau_{j_{n2}}} \left(\Phi_n(\hat{A}_n(u); \hat{\beta}_n) - \Phi_n(u; \hat{\beta}_n)\right)^2 du \\ &\leqslant 8 O_P(b^{-1}) \mathbb{1}_{E_n} |\hat{\beta}_n - \beta_0|^2 \sup_{x \in \mathbb{R}} \left|D_n^{(1)}(x; \beta^*)\right|^2 \\ &\quad + 2b^{-2} \mathbb{1}_{E_n} \sup_{s \in [\epsilon, M]} \Phi'(s; \beta_0) \int_{\tau_{j_{n1}}}^{\tau_{j_{n2}}} \left(\hat{A}_n(u) - u\right)^2 du \\ &\quad + 8 O_P(b^{-1}) \sup_{x \in \mathbb{R}} |\Phi_n(x; \beta_0) - \Phi(x; \beta_0)|^2 \\ &= O_p(b^{-1} n^{-1}) + O_p(b^{-1} n^{-2/3}) + O_p(b^{-1} n^{-1}) = O_p(b^{-1} n^{-2/3}). \end{aligned}$$

Consequently, from (3.6.31) together with (3.6.16), for the first term on the right hand side of (3.6.30), we obtain

$$\begin{aligned}
&\mathbb{1}_{E_n} \left\| \frac{(a_{n,t} \circ \hat{A}_n)(\widehat{\Phi}_n \circ \hat{A}_n) - a_{n,t}\widehat{\Phi}_n}{\widehat{\Phi}_n} \mathbb{1}_{[\tau_{j_{n1}}, \tau_{j_{n2}}]} \right\|_{\mathcal{L}_2}^2 \\
&= O_p(b^{-3}n^{-2/3}) + O_p(b^{-1}n^{-2/3}) \\
&= O_p(b^{-3}n^{-2/3}).
\end{aligned}$$

For the second term on the right hand side of (3.6.30), we first write

$$\begin{aligned}
&\left\| \left(\Phi_0\lambda_0 - \widehat{\Phi}\tilde{\lambda}_n \right) \mathbb{1}_{[\tau_{j_{n1}}, \tau_{j_{n2}}]} \right\|_{\mathcal{L}_2} \\
&\leqslant \left\| \left(\Phi_0 - \widehat{\Phi} \right) \lambda_0 \mathbb{1}_{[\tau_{j_{n1}}, \tau_{j_{n2}}]} \right\|_{\mathcal{L}_2} + \left\| \left(\lambda_0 - \tilde{\lambda}_n \right) \widehat{\Phi} \mathbb{1}_{[\tau_{j_{n1}}, \tau_{j_{n2}}]} \right\|_{\mathcal{L}_2}
\end{aligned} \tag{3.6.34}$$

On the event E_n, we find

$$\begin{aligned}
&\mathbb{1}_{E_n} \left\| \left(\Phi_0 - \widehat{\Phi} \right) \lambda_0 \mathbb{1}_{[\tau_{j_{n1}}, \tau_{j_{n2}}]} \right\|_{\mathcal{L}_2}^2 \\
&\leqslant 2O_P(b)\mathbb{1}_{E_n} |\hat{\beta}_n - \beta_0|^2 \sup_{x \in \mathbb{R}} \left| D_n^{(1)}(x; \beta^*) \right|^2 \sup_{u \in [0,M]} \lambda_0(u) \\
&= O_p(bn^{-1}),
\end{aligned}$$

and

$$\begin{aligned}
\mathbb{1}_{E_n} \left\| \left(\lambda_0 - \tilde{\lambda}_n \right) \widehat{\Phi} \mathbb{1}_{[\tau_{j_{n1}}, \tau_{j_{n2}}]} \right\|_{\mathcal{L}_2}^2 &\leqslant \Phi(0, \hat{\beta}_n)\mathbb{1}_{E_n} \int_{\tau_{j_{n1}}}^{\tau_{j_{n2}}} \left(\lambda_0(u) - \tilde{\lambda}_n(u) \right)^2 du \\
&= O_p(bn^{-2/3}),
\end{aligned}$$

due to Lemma 3.6.7. It follows that

$$\mathbb{1}_{E_n} \left\| \left(\Phi_0\lambda_0 - \widehat{\Phi}\tilde{\lambda}_n \right) \mathbb{1}_{[t-b, t+b]} \right\|_{\mathcal{L}_2} = O_p(b^{1/2}n^{-1/3}).$$

Together with (3.6.30), this concludes the proof. □

Proof of Lemma 3.2.4. Let n be sufficiently large such that $t + b \leqslant M' < M < \tau_H$. Denote by R_n the left hand side of (3.2.21). Write $\mathbb{G}_n = \sqrt{n}(\mathbb{P}_n - \mathbb{P})$ and decompose $R_n = R_{n1} + R_{n2}$, where

$$\begin{aligned}
R_{n1} &= \frac{1}{\sqrt{n}} \mathbb{1}_{E_n} \int \frac{\delta \mathbb{1}_{[\tau_{j_{n1}}, \tau_{j_{n2}}]}(u)}{\Phi_n(u; \hat{\beta}_n)} \left(\bar{a}_{n,t}\bar{\Phi}_n(u; \hat{\beta}_n) - a_{n,t}(u)\Phi_n(u; \hat{\beta}_n) \right) d\mathbb{G}_n(u, \delta, z), \\
R_{n2} &= \frac{1}{\sqrt{n}} \mathbb{1}_{E_n} \int \mathbb{1}_{\{u > \tau_{j_{n1}}\}} \Bigg[e^{\hat{\beta}_n' z} \int_{\tau_{j_{n1}}}^{u \wedge \tau_{j_{n2}}} \frac{\bar{a}_{n,t}\bar{\Phi}_n(v; \hat{\beta}_n)}{\Phi_n(v; \hat{\beta}_n)} d\hat{\Lambda}_n(v) \\
&\qquad - e^{\beta_0' z} \int_{t-b}^{u \wedge (t+b)} a_{n,t}(v)\, d\Lambda_0(v) \Bigg] d\mathbb{G}_n(u, \delta, z).
\end{aligned}$$

Choose $\eta > 0$. We prove separately that there exists $\nu > 0$, such that

$$\begin{aligned} \limsup_{n\to\infty} \mathbb{P}\left(b^{3/2}n^{13/18}|R_{n1}| > \nu\right) &\leqslant \eta \\ \limsup_{n\to\infty} \mathbb{P}\left(n^{1/2}|R_{n2}| > \nu\right) &\leqslant \eta. \end{aligned} \tag{3.6.35}$$

We consider the following events.

$$\begin{aligned} \mathcal{A}_{n1} &= \left\{\tilde{\lambda}_n(M) > K_1\right\}, \\ \mathcal{A}_{n2} &= \left\{\sup_{s\in[\epsilon,M']} \left|\lambda_0(s) - \tilde{\lambda}_n(s)\right| > K_2 n^{-2/9}\right\}, \end{aligned} \tag{3.6.36}$$

where $K_1, K_2 > 0$, and let $\mathcal{A}_n = \mathcal{A}_{n1} \cup \mathcal{A}_{n2}$. Since $\tilde{\lambda}_n(M) = O_p(1)$ we can choose $K_1 > 0$, such that $\mathbb{P}(\mathcal{A}_{n1}) \leqslant \eta/3$ and from Lemma 3.6.5 we find that we can choose $K_2 > 0$, such that $\mathbb{P}(\mathcal{A}_{n2}) \leqslant \eta/3$, so that $\mathbb{P}(\mathcal{A}_n) \leqslant 2\eta/3$. First consider R_{n1}. Since

$$\begin{aligned} \mathbb{P}\left(b^{3/2}n^{13/18}|R_{n1}| > \nu\right) &\leqslant \mathbb{P}\left(\mathcal{A}_n\right) + \mathbb{P}\left(\left\{b^{3/2}n^{13/18}|R_{n1}| > \nu\right\} \cap \mathcal{A}_n^c\right) \\ &\leqslant 2\eta/3 + b^{3/2}n^{13/18}\nu^{-1}\mathbb{E}\left[|R_{n1}|\mathbb{1}_{\mathcal{A}_n^c}\right], \end{aligned} \tag{3.6.37}$$

it suffices to show that there exists $\nu > 0$, such that

$$b^{3/2}n^{13/18}\nu^{-1}\mathbb{E}\left[|R_{n1}|\mathbb{1}_{\mathcal{A}_n^c}\right] \leqslant \eta/3,$$

for all n sufficiently large. Write

$$\begin{aligned} w(u) &= \frac{1}{\Phi_n(u;\hat{\beta}_n)}\left(a_{n,t}(\hat{A}_n(u))\Phi_n(\hat{A}_n(u);\hat{\beta}_n) - a_{n,t}(u)\Phi_n(u;\hat{\beta}_n)\right) \\ &= \frac{a_{n,t}(\hat{A}_n(u))}{\Phi_n(u;\hat{\beta}_n)}\left(\Phi_n(\hat{A}_n(u);\hat{\beta}_n) - \Phi(\hat{A}_n(u);\beta_0)\right) \\ &\quad + \frac{a_{n,t}(u)}{\Phi_n(u;\hat{\beta}_n)}\left(\Phi(u;\beta_0) - \Phi_n(u;\hat{\beta}_n)\right) \\ &\quad + \frac{1}{\Phi_n(u;\hat{\beta}_n)}\left(a_{n,t}(\hat{A}_n(u))\Phi(\hat{A}_n(u);\beta_0) - a_{n,t}(u)\Phi(u;\beta_0)\right). \end{aligned}$$

We will argue that the function $W_n = b^2 n^{2/9} w$ is uniformly bounded and of bounded variation. Because of (3.6.15), we have that

$$n^{1/3}(\Phi_n(\hat{A}_n(u);\hat{\beta}_n) - \Phi(\hat{A}_n(u);\beta_0)) \quad \text{and} \quad n^{1/3}(\Phi(u;\beta_0) - \Phi_n(u;\hat{\beta}_n))$$

are uniformly bounded. Moreover, they are of bounded variation, as being the difference of two monotone functions. Similarly, $1/\Phi_n(u;\hat{\beta}_n)$ is of

bounded variation and on the event E_n it is also uniformly bounded. Furthermore, by the definition of $a_{n,t}$, we have

$$a_{n,t}(\hat{A}_n(u))\Phi(\hat{A}_n(u);\beta_0) - a_{n,t}(u)\Phi(u;\beta_0) = k_b\left(t - \hat{A}_n(u)\right) - k_b(t-u).$$

This is a function of bounded variation, such that multiplied by $b^2 n^{2/9}$ it is uniformly bounded on the event $\mathcal{A}_n^c$, because using (3.6.32), we obtain

$$\begin{aligned}\left|k_b\left(t - \hat{A}_n(u)\right) - k_b(t-u)\right| &\leqslant b^{-2}|\hat{A}_n(u) - u| \sup_{x\in[-1,1]} |k'(x)| \\ &\leqslant 2Kb^{-2}|\tilde{\lambda}_n(u) - \lambda_0(u)| \sup_{x\in[-1,1]} |k'(x)| \\ &\leqslant b^{-2}n^{-2/9}2KK_2 \sup_{x\in[-1,1]} |k'(x)|.\end{aligned} \tag{3.6.38}$$

Finally, $ba_{n,t}(u) = bk_b(t-u)/\Phi(u;\beta_0)$ is also a function of bounded variation, as being the product of a function of bounded variation $bk_b(t-u)$ with the monotone function $1/\Phi(u;\beta_0)$, and it is uniformly bounded. Then, since $ba_{n,t}(\hat{A}_n(u))$ is the composition of an increasing function with a function of bounded variation that is uniformly bounded, it is also a function of bounded variation and uniformly bounded. As a result, being the sum and product of functions of bounded variation that are uniformly bounded, $W_n = b^2 n^{2/9} w$ belongs to the class $\mathcal{B}_{\tilde{K}}$ of functions of bounded variation, uniformly bounded by some constant $\tilde{K}$. Moreover, from Lemma 3.6.6, it follows that on the event $\mathcal{A}_{n2}^c$ the distance between the jump times of $\tilde{\lambda}_n$ on $[\epsilon, M']$ is of smaller order than b. In particular, this means that, $|\tau_{j_{n1}} - (t-b)| < b$ and $|\tau_{j_{n2}} - (t+b)| < b$. Consequently, on $\mathcal{A}_n^c$, it holds

$$\begin{aligned}R_{n1} &= n^{-1/2}\mathbb{1}_{E_n}\int \delta\mathbb{1}_{[t-2b,t+2b]}(u)w(u)\,d\sqrt{n}(\mathbb{P}_n - \mathbb{P})(u,\delta,z) \\ &= b^{-2}n^{-13/18}\mathbb{1}_{E_n}\int \delta\mathbb{1}_{[t-2b,t+2b]}(u)W_n(u)\,d\sqrt{n}(\mathbb{P}_n - \mathbb{P})(u,\delta,z).\end{aligned}$$

Let $\mathcal{B}_{\tilde{K}}$ be the class of functions of bounded variation on $[0,M]$, that are uniformly bounded by $\tilde{K} > 0$, and let $\mathcal{G}_n = \{\zeta_{B,n} : B \in \mathcal{B}_{\tilde{K}}\}$, where $\zeta_{B,n}(u,\delta) = \delta\mathbb{1}_{[t-2b,t+2b]}(u)B(u)$. Then, $\delta\mathbb{1}_{[t-2b,t+2b]}W_n$ is a member of the class $\mathcal{G}_n$, which has envelope $F_n(u,\delta) = \tilde{K}\delta\mathbb{1}_{[t-2b,t+2b]}(u)$. Furthermore, if $J(\delta,\mathcal{G}_n)$ is the corresponding entropy-integral (see Section 2.14 in van der Vaart and Wellner, 1996), then according to Lemma A.1.1, $J(\delta,\mathcal{G}_n) \leqslant \int_0^\delta \sqrt{1 + C/\epsilon}\,d\epsilon$,

for some $C > 0$. Consequently, together with Theorem 2.14.1 in van der Vaart and Wellner, 1996, we obtain that

$$\begin{aligned}\mathbb{E}\left[|R_{n1}|\mathbb{1}_{\mathcal{A}_n^c}\right] &\leqslant b^{-2}n^{-13/18}\mathbb{E}\sup_{\zeta\in\mathcal{G}_n}\left|\int \zeta_{B,n}(u,\delta,z)\,d\sqrt{n}(\mathbb{P}_n-\mathbb{P})(u,\delta,z)\right| \\ &\leqslant KJ(1,\mathcal{G}_n)\|F_n\|_{L_2(\mathbb{P})}b^{-2}n^{-13/18} \\ &\leqslant K'\left(H^{uc}(t+2b)-H^{uc}(t-2b)\right)^{1/2}b^{-2}n^{-13/18} \\ &\leqslant K''b^{-3/2}n^{-13/18},\end{aligned}$$

because H^{uc} is absolutely continuous. As a result, for sufficiently large ν

$$b^{3/2}n^{13/18}\nu^{-1}\mathbb{E}\left[|R_{n1}|\mathbb{1}_{\mathcal{A}_n^c}\right] \leqslant \frac{K''}{\nu} \leqslant \eta/3.$$

This proves the first part of (3.6.35).

We proceed with R_{n2}. Similar to (3.6.37),

$$\mathbb{P}\left(n^{1/2}|R_{n2}| > \nu\right) \leqslant 2\eta/3 + n^{1/2}\nu^{-1}\mathbb{E}\left[|R_{n2}|\mathbb{1}_{\mathcal{A}_n^c}\right]. \tag{3.6.39}$$

and it suffices to show that there exists $\nu > 0$, such that

$$n^{1/2}\nu^{-1}\mathbb{E}\left[|R_{n2}|\mathbb{1}_{\mathcal{A}_n^c}\right] \leqslant \eta/3,$$

for all n sufficiently large. We write

$$n^{1/2}R_{n2} = \mathbb{1}_{E_n}\int\left(e^{\hat{\beta}_n' z}r_{1,n}(u) - e^{\beta_0' z}r_{2,n}(u)\right)\,d\sqrt{n}(\mathbb{P}_n-\mathbb{P})(u,\delta,z),$$

where

$$\begin{aligned}r_{1,n}(u) &= \mathbb{1}_{\{u>\tau_{j_{n1}}\}}\int_{\tau_{j_{n1}}}^{u\wedge\tau_{j_{n2}}}\frac{a_{n,t}(\hat{A}_n(v))}{\Phi_n(v;\hat{\beta}_n)}\Phi_n(\hat{A}_n(v);\hat{\beta}_n)\tilde{\lambda}_n(v)\,dv, \\ r_{2,n}(u) &= \mathbb{1}_{\{u>t-b\}}\int_{t-b}^{u\wedge(t+b)} a_{n,t}(v)\,\lambda_0(v)\,dv,\end{aligned}$$

are both monotone functions, uniformly bounded by some constant C on the event $\mathcal{A}_n^c \cap E_n$. Let $\mathcal{M}_C$ be the class of monotone functions bounded uniformly by $C > 0$ and let $\mathcal{G}_n = \{\zeta_{r,\beta}(u,z) : r \in \mathcal{M}_C, \beta \in \mathbb{R}^p, |\beta-\beta_0| \leqslant \xi_2\}$, where ξ_2 is chosen as in (3.6.3) and $\zeta_{r,\beta}(u,z) = r(u)e^{\beta' z}$. Then $e^{\hat{\beta}_n' z}r_{1,n}(u)$ is a member of the class $\mathcal{G}_n$, which has envelope

$$F_n(u,z) = C\exp\left\{\sum_{j=1}^{p}(\beta_{0,j}-\sigma_n)z_j \vee (\beta_{0,j}+\sigma_n)z_j\right\},$$

with $\sigma_n = \sqrt{\xi_2 n^{-2/3}}$ is the envelope of $\mathcal{G}_n$. If $J_{[]}(\delta, \mathcal{G}_n, L_2(\mathbb{P}))$ is the bracketing integral (see Section 2.14 in van der Vaart and Wellner, 1996), then according to Lemma A.1.2, $J_{[]}(\delta, \mathcal{G}_n, L_2(\mathbb{P})) \leqslant \int_0^\delta \sqrt{1 + C/\epsilon}\, d\epsilon$, for some $C > 0$. Hence, together with Theorem 2.14.2 in van der Vaart and Wellner, 1996, we obtain

$$\begin{aligned}
&\mathbb{E}\left[\left|\mathbb{1}_{A_n^c \cap E_n} \int e^{\hat{\beta}_n' z} r_{1,n}(u)\, d\sqrt{n}(\mathbb{P}_n - \mathbb{P})(u, \delta, z)\right|\right] \\
&\quad \leqslant \mathbb{E} \sup_{\zeta \in \mathcal{G}_n} \left|\int \zeta_{r,\beta}(u, z)\, d\sqrt{n}(\mathbb{P}_n - \mathbb{P})(u, \delta, z)\right| \\
&\quad \leqslant K J_{[]}(1, \mathcal{G}_n, L_2(\mathbb{P})) \|F_n\|_{L_2(\mathbb{P})} \leqslant K',
\end{aligned}$$

for some $K' > 0$. We conclude that,

$$\nu^{-1}\mathbb{E}\left[\left|\mathbb{1}_{A_n^c \cap E_n} \int e^{\hat{\beta}_n' z} r_{1,n}(u)\, d\sqrt{n}(\mathbb{P}_n - \mathbb{P})(u, \delta, z)\right|\right] \leqslant \frac{K'}{\nu} \leqslant \eta/6,$$

for sufficiently large ν. In the same way, it can also be proved

$$\nu^{-1}\mathbb{E}\left[\left|\mathbb{1}_{A_n^c \cap E_n} \int e^{\beta_0 z} r_{2,n}(u)\, d\sqrt{n}(\mathbb{P}_n - \mathbb{P})(u, \delta, z)\right|\right] \leqslant \frac{K}{\nu} \leqslant \eta/6$$

for sufficiently large ν, concluding the proof of (3.6.39) and therefore the second part of (3.6.35). □

3.6.2 *Proofs for Section 3.3*

Proof of Lemma 3.3.1. We start by writing

$$\begin{aligned}
\ell_\beta^s(\lambda_0) &= \frac{1}{n}\sum_{i=1}^n \Delta_i \int_0^\infty \log \lambda_0(t) k_b(t - T_i)\, dt \\
&\quad - \frac{1}{n}\sum_{i=1}^n e^{\beta' Z_i} \int_0^\infty \left(\int_0^t \lambda_0(u)\, du\right) k_b(t - T_i)\, dt \\
&= \int_0^\infty \log \lambda_0(t) \left(\frac{1}{n}\sum_{i=1}^n \Delta_i k_b(t - T_i)\right) dt \\
&\quad - \int_0^\infty \lambda_0(u) \left(\frac{1}{n}\sum_{i=1}^n e^{\beta' Z_i} \int_u^\infty k_b(t - T_i)\, dt\right) du,
\end{aligned}$$

which is equal to

$$\begin{aligned}&\int_0^\infty \Big\{ v_n(t)\log\lambda_0(t) - w_n(t;\beta)\lambda_0(t)\Big\}\,dt\\ &= \int_0^\infty \left\{ \frac{v_n(t)}{w_n(t;\beta)}\log\lambda_0(t) - \lambda_0(t)\right\} w_n(t;\beta)\,dt,\end{aligned}$$

with v_n and w_n defined in (3.3.3). Maximizing the right hand side over nondecreasing λ_0 is equivalent to minimizing

$$\int \Delta_\Phi\left(\frac{v_n(t)}{w_n(t;\beta)}, \lambda(t)\right) w_n(t;\beta)\,dt \tag{3.6.40}$$

over nondecreasing λ, where $\Delta_\Phi(u,v) = \Phi(u) - \Phi(v) - (u-v)\phi(v)$, with $\Phi(u) = u\log u$. Theorem 1 in Groeneboom and Jongbloed, 2010 provides a characterization of the minimizer $\hat\lambda_n^s(t;\beta)$ of (3.6.40), and hence of the maximizer of ℓ_β^s. It is the unique solution of a generalized continuous isotonic regression problem, i.e., it is continuous and it is the minimizer of

$$\psi(\lambda) = \frac{1}{2}\int \left(\lambda(t) - \frac{v_n(t)}{w_n(t;\beta)}\right)^2 w_n(t;\beta)\,dt,$$

over all nondecreasing functions λ and can be described as the slope of the GCM of the graph defined by (3.3.4). □

Proof of Lemma 3.3.2. It is enough to prove that for an arbitrarily fixed $\epsilon > 0$ and for sufficiently large n

$$\mathbb{P}\left(\lambda_n^{\text{naive}}(t) = \lambda_n^{\text{IS}}(t),\ \text{for all } t \in [\ell, M]\right) \geqslant 1 - \epsilon.$$

Recall that $\lambda_n^{\text{IS}}(t)$ is defined as the slope of the greatest convex minorant $\big\{\big(X_n(t), \widehat{Y}_n(t)\big), t \in [0,\hat\tau]\big\}$ of the graph $\big\{\big(X_n(t), Y_n(t)\big)\big), t \in [0,\hat\tau]\big\}$. We consider Y_n on the interval $[\ell, M]$ and define the linearly extended version of Y_n on $[0,\hat\tau]$ by

$$Y_n^*(t) = \begin{cases} Y_n(\ell) + \big(X_n(t) - X_n(\ell)\big)\lambda_n^{\text{naive}}(\ell), & \text{for } t \in [0,\ell),\\ Y_n(t), & \text{for } t \in [\ell, M],\\ Y_n(M) + \big(X_n(t) - X_n(M)\big)\lambda_n^{\text{naive}}(M), & \text{for } t \in (M,\hat\tau].\end{cases}$$

It suffices to prove that, for sufficiently large n,

$$\mathbb{P}\left(\big\{\big(X_n(t), Y_n^*(t)\big) : t \in [0,\hat\tau]\big\}\ \text{is convex}\right) \geqslant 1 - \epsilon/2, \tag{3.6.41}$$

and

$$\mathbb{P}\left(Y_n^*(t) \leqslant Y_n(t),\ \text{for all } t \in [0,\hat\tau]\right) \geqslant 1 - \epsilon/2. \tag{3.6.42}$$

Indeed, if (3.6.41) and (3.6.42) hold, then with probability greater than or equal to $1-\epsilon$, the curve $\{(X_n(t), Y_n^*(t)) : t \in [0, \hat{\tau}]\}$ is a convex curve lying below the graph $\{(X_n(t), Y_n(t)) : t \in [0, \hat{\tau}]\}$, with $Y_n^*(t) = Y_n(t)$ for all $t \in [\ell, M]$. Hence, $Y_n(t) = Y_n^*(t) \leqslant \widehat{Y}_n(t) \leqslant Y_n(t)$, for all $t \in [\ell, M]$. It follows that, for sufficiently large n,

$$\mathbb{P}\left(\lambda_n^{\text{naive}}(t) = \frac{dY_n(t)}{dX_n(t)} = \frac{d\widehat{Y}_n(t)}{dX_n(t)} = \lambda_n^{IS}(t), \text{ for all } t \in [\ell, M]\right) \geqslant 1-\epsilon.$$

To prove (3.6.41), define the event

$$A_n = \left\{\lambda_n^{\text{naive}} \text{ is increasing on } [\ell-\eta_1, M+\eta_2]\right\},$$

for $\eta_1 \in (0, \ell)$ and $\eta_2 \in (0, \hat{\tau} - M)$. Note that on the intervals $[0, \ell)$ and $(M, \hat{\tau}]$, the curve $\{(X_n(t), Y_n^*(t)) : t \in [0, \hat{\tau}]\}$ is the tangent line of the graph $\{(X_n(t), Y_n(t)) : t \in [0, \hat{\tau}]\}$ at the points $(X_n(\ell), Y_n(\ell))$ and $(X_n(M), Y_n(M))$. As a result, on the event A_n the curve is convex, so that together with condition (c), for sufficiently large n

$$\mathbb{P}\left(\{(X_n(t), Y_n^*(t)) : t \in [0, \hat{\tau}]\} \text{ is convex }\right) \geqslant \mathbb{P}(A_n) \geqslant 1-\epsilon/2.$$

To prove (3.6.42), we split the interval $[0, \hat{\tau}]$ in five different intervals $I_1 = [0, \ell-\eta_1)$, $I_2 = [\ell-\eta_1, \ell)$, $I_3 = [\ell, M]$, $I_4 = (M, M+\eta_2]$, and $I_5 = (M+\eta_2, \hat{\tau}]$, and show that

$$\mathbb{P}\left(Y_n^*(t) \leqslant Y_n(t), \text{ for all } t \in I_i\right) \geqslant 1-\epsilon/10, \quad i = 1, \ldots, 5. \tag{3.6.43}$$

For $t \in I_3$, $Y_n^*(t) = Y_n(t)$ and thus (3.6.43) is trivial. For $t \in I_2$, by the mean value theorem,

$$Y_n(t) - Y_n(\ell) = \left(X_n(t) - X_n(\ell)\right) \lambda_n^{\text{naive}}(\xi_t),$$

for some $\xi_t \in [t, \ell]$. Thus, since $X_n(t) \leqslant X_n(\ell)$ according to condition (a),

$$\begin{aligned}
&\mathbb{P}\left(Y_n^*(t) \leqslant Y_n(t), \text{ for all } t \in I_2\right) \\
&= \mathbb{P}\left(\left(X_n(t) - X_n(\ell)\right)\left(\lambda_n^{\text{naive}}(\xi_t) - \lambda_n^{\text{naive}}(\ell)\right) \geqslant 0, \text{ for all } t \in I_2\right) \\
&= \mathbb{P}\left(\left(\lambda_n^{\text{naive}}(\xi_t) - \lambda_n^{\text{naive}}(\ell)\right) \leqslant 0, \text{ for all } t \in I_2\right) \\
&\geqslant \mathbb{P}(A_n) \geqslant 1-\epsilon/10,
\end{aligned} \tag{3.6.44}$$

for n sufficiently large, according to condition (c). The argument for $t \in I_4$ is exactly the same. Furthermore, making use of condition (d), for each $t \in I_1$, we obtain

$$\begin{aligned}
Y_0(t) - Y_0(\ell) - \lambda_0(\ell)\left(X_0(t) - X_0(\ell)\right) &= \int_t^\ell \left(\lambda_0(\ell) - \lambda_0(u)\right) dX_0(u) \\
&\geqslant \int_{\ell-\eta_1}^\ell \left(\lambda_0(\ell) - \lambda_0(u)\right) dX_0(u).
\end{aligned}$$

This implies that

$$\begin{aligned}
Y_n(t) - Y_n^*(t) &= Y_n(t) - Y_n(\ell) - \big(X_n(t) - X_n(\ell)\big)\lambda_n^{\text{naive}}(\ell) \\
&\geqslant Y_n(t) - Y_0(t) + Y_0(\ell) - Y_n(\ell) + \lambda_0(\ell)\big(X_0(t) - X_n(t)\big) \\
&\quad + \lambda_0(\ell)\big(X_n(\ell) - X_0(\ell)\big) + \big(\lambda_n^{\text{naive}}(\ell) - \lambda_0(\ell)\big)\big(X_n(\ell) - X_n(t)\big) \\
&\quad + \int_{\ell-\eta_1}^{\ell} \big(\lambda_0(\ell) - \lambda_0(u)\big) dX_0(u) \\
&\geqslant -2 \sup_{s\in[0,\ell]} |Y_n(s) - Y_0(s)| - 2\lambda_0(\ell) \sup_{s\in[0,\ell]} |X_n(s) - X_0(s)| \\
&\quad - 2|\lambda_n^{\text{naive}}(\ell) - \lambda_0(\ell)| \sup_{s\in[0,\ell]} |X_n(s)| \\
&\quad + \int_{\ell-\eta_1}^{\ell} \big(\lambda_0(\ell) - \lambda_0(u)\big) dX_0(u).
\end{aligned}$$

According to conditions (b) and (c), the first three terms on the right hand side tend to zero in probability. This means that the probability on the left hand side of (3.6.43) for $i = 1$, is bounded from below by

$$\mathbb{P}\left(Z_n \leqslant \int_{\ell-\eta_1}^{\ell} \big(\lambda_0(\ell) - \lambda_0(u)\big) dX_0(u)\right),$$

where $Z_n = o_p(1)$. This probability is greater than $1 - \epsilon/10$ for n sufficiently large, since

$$\int_{\ell-\eta_1}^{\ell} \big(\lambda_0(\ell) - \lambda_0(u)\big)\, dX_0(u) \geqslant x_0(\ell) \int_{\ell-\eta_1}^{\ell} \big(\lambda_0(\ell) - \lambda_0(u)\big)\, du > 0,$$

using that λ_0 is strictly increasing. For I_5 we can argue exactly in the same way. □

Lemma 3.6.8. *Suppose that (A1)-(A2) hold. Let* $H^{uc}(t)$ *and* $\Phi(t;\beta_0)$ *be defined in* (3.1.1) *and* (3.1.2)*, and let* $h(t) = dH^{uc}(t)/dt$*. Suppose that* h *and* $t \mapsto \Phi(t;\beta_0)$ *are* $m \geqslant 1$ *times continuously differentiable and let* k *be* m*-orthogonal satisfying* (1.2.1)*. Then, for each* $0 < \ell < M < \tau_H$*, it holds*

$$\begin{aligned}
\sup_{t\in[\ell,M]} |v_n(t) - h(t)| &= O(b^m) + O_p(b^{-1}n^{-1/2}), \\
\sup_{t\in[\ell,M]} |w_n(t;\hat\beta_n) - \Phi(t;\beta_0)| &= O(b^m) + O_p(b^{-1}n^{-1/2}),
\end{aligned} \tag{3.6.45}$$

where v_n*,* w_n *and* Φ *are defined in* (3.3.3) *and* (3.1.2)*.*

Proof. To obtain the first result in (3.6.45), we write

$$v_n(t) - h(t) = v_n(t) - h_s(t) + h_s(t) - h(t),$$

where

$$h_s(t) = \int k_b(t-u)\, h(u)\, du. \tag{3.6.46}$$

By a change of variable and a Taylor expansion, using that k is m-orthogonal, we deduce that

$$\begin{aligned} &h_s(t) - h(t) \\ &= \int_{-1}^{1} k(y) \left\{ -h'(t)by + \cdots + \frac{h^{(m-1)}(t)}{(m-1)!}(-by)^{m-1} + \frac{h^{(m)}(\xi_{ty})}{m!}(-by)^m \right\} dy \\ &= \frac{(-b)^m}{m!} \int_{-1}^{1} h^{(m)}(\xi_{ty}) k(y) y^m \, dy, \end{aligned} \tag{3.6.47}$$

for some $|\xi_{ty} - t| < |by|$. It follows that

$$\sup_{t\in[\ell,M]} |h_s(t) - h(t)| \leqslant \frac{b^m}{m!} \sup_{t\in[0,\tau_H]} \left| h^{(m)}(t) \right| \int_{-1}^{1} |k(y)||y|^m \, dy = O(b^m). \tag{3.6.48}$$

Let H_n^{uc} be the empirical sub-distribution function of the uncensored observations, defined by

$$H_n^{uc}(t) = \int \delta \mathbb{1}_{\{u \leqslant t\}} \, d\mathbb{P}_n(u, \delta, z).$$

Then integration by parts yields

$$\begin{aligned} v_n(t) - h_s(t) &= \int k_b(t-u)\, d(H_n^{uc} - H^{uc})(u) \\ &= -\int \frac{\partial}{\partial u} k_b(t-u)\, (H_n^{uc} - H^{uc})(u)\, du \\ &= \frac{1}{b} \int_{-1}^{1} k'(y)\, (H_n^{uc} - H^{uc})(t - by)\, dy. \end{aligned} \tag{3.6.49}$$

Note that $H_n^{uc}(x) - H^{uc}(x) = \int \delta \mathbb{1}_{\{u \leqslant x\}} \, d(\mathbb{P}_n - \mathbb{P})(u, \delta, z)$. Because the class of functions $\mathcal{F} = \{f(\cdot; x) : x \in [0, \tau_H]\}$, with $f(u; x) = \mathbb{1}_{\{u \leqslant x\}}$, is a VC-class (e.g., see Example 2.6.1 in van der Vaart and Wellner, 1996), also the class of functions $\{\mathcal{G} = \delta f : f \in \mathcal{F}\}$ is a VC-class, according to Lemma 2.6.18 in van der Vaart and Wellner, 1996. It follows that the class $\mathcal{G}$ is Donsker, i.e., the process $\sqrt{n}(H_n^{uc} - H^{uc})$ converges weakly, see Theorems 2.6.8 and 2.5.2 in van der Vaart and Wellner, 1996. It follows by the continuous mapping theorem that

$$\sqrt{n} \sup_{t\in[0,\tau_H]} |H_n^{uc}(t) - H^{uc}(t)| = O_p(1). \tag{3.6.50}$$

Hence, we get

$$\begin{aligned}\sup_{t\in[\ell,M]}|v_n(t)-h_s(t)| &\leqslant \frac{1}{b}\sup_{t\in[\ell,M]}|H_n^{uc}(t)-H^{uc}(t)|\int_{-1}^{1}|k'(y)|\,dy \\ &= O_p(b^{-1}n^{-1/2}).\end{aligned} \tag{3.6.51}$$

Together with (3.6.48), this proves the first result in (3.6.45).

To prove the second result in (3.6.45), note that from (3.3.3) and (3.1.4) we have

$$w_n(t;\beta) = \frac{1}{n}\sum_{i=1}^{n} e^{\beta' Z_i}\int_t^{\infty} k_b(u-T_i)\,du = \int_{-1}^{1} k(y)\Phi_n(t-by;\beta)\,dy. \tag{3.6.52}$$

Consequently, we can write

$$\begin{aligned}w_n(t;\hat\beta_n)-\Phi(t;\beta_0) &= \int_{-1}^{1} k(y)\left\{\Phi_n(t-by;\hat\beta_n)-\Phi(t;\beta_0)\right\}\,dy \\ &= \int_{-1}^{1} k(y)\left\{\Phi_n(t-by;\hat\beta_n)-\Phi(t-by;\beta_0)\right\}\,dy \\ &\quad+\int_{-1}^{1} k(y)\{\Phi(t-by;\beta_0)-\Phi(t;\beta_0)\}\,dy.\end{aligned}$$

Similar to (3.6.47) and (3.6.48), for the second term on the right hand side, we obtain

$$\sup_{t\in[\ell,M]}\left|\int_{-1}^{1} k(y)\{\Phi(t-by;\beta_0)-\Phi(t;\beta_0)\}\,dy\right| = O(b^m).$$

Hence, by means of the triangular inequality,

$$\begin{aligned}\sup_{t\in[\ell,M]}\left|w_n(t;\hat\beta_n)-\Phi(t;\beta_0)\right| &\leqslant \sup_{t\in\mathbb{R}}\left|\Phi_n(t;\hat\beta_n)-\Phi(t;\beta_0)\right| + O(b^m) \\ &= O_p(n^{-1/2})+O(b^m),\end{aligned}$$

according to Lemma 4 in Lopuhaä and Nane, 2013. □

Lemma 3.6.9. *Let $\hat\lambda_n^{\text{naive}}$ be defined in* (3.3.7)*. Then, under the assumptions of Lemma 3.6.8, for each $0<\ell<M<\tau_H$,*

$$\sup_{t\in[\ell,M]}|\hat\lambda_n^{\text{naive}}(t)-\lambda_0(t)| = O(b^m)+O_p(b^{-1}n^{-1/2}).$$

Proof. By (3.1.3) and the definition of $\hat{\lambda}_n^{\text{naive}}$, we have

$$\sup_{t\in[\ell,M]} |\hat{\lambda}_n^{\text{naive}}(t) - \lambda_0(t)| = \sup_{t\in[\ell,M]} \left| \frac{v_n(t)}{w_n(t;\hat{\beta}_n)} - \frac{h(t)}{\Phi(t;\beta_0)} \right|$$
$$\leqslant \frac{\sup_{t\in[\ell,M]} \left| v_n(t)\Phi(t;\beta_0) - h(t)w_n(t;\hat{\beta}_n) \right|}{|w_n(M;\hat{\beta}_n)|\Phi(M;\beta_0)}.$$

The triangular inequality and Lemma 3.6.8 yield

$$\sup_{t\in[\ell,M]} \left| v_n(t)\Phi(t;\beta_0) - h(t)w_n(t;\hat{\beta}_n) \right| = O(b^m) + O_p(b^{-1}n^{-1/2}).$$

and $w_n(M;\hat{\beta}_n)^{-1} = O_p(1)$. The statement follows immediately. □

Lemma 3.6.10. *Suppose that (A1)-(A2) hold. Let* $H^{uc}(t)$ *and* $\Phi(t;\beta_0)$ *be defined in* (3.1.1) *and* (3.1.2)*, and let* $h(t) = dH^{uc}(t)/dt$*. Suppose that* h *and* $t \mapsto \Phi(t;\beta_0)$ *are* $m \geqslant 1$ *times continuously differentiable and let* k *be* m*-orthogonal satisfying* (1.2.1)*. If* $b \to 0$ *and* $1/b = O(n^\alpha)$*, for some* $\alpha \in (0, 1/4)$*, then for each* $0 < \ell < M < \tau_H$*, it holds*

$$\sup_{t\in[\ell,M]} |v_n'(t) - h'(t)| \xrightarrow{\mathbb{P}} 0, \qquad \sup_{t\in[\ell,M]} |w_n'(t;\hat{\beta}_n) - \Phi'(t;\beta_0)| \xrightarrow{\mathbb{P}} 0, \tag{3.6.53}$$

where v_n*,* w_n *and* Φ *are defined in* (3.3.3) *and* (3.1.2)*.*

Proof. Let us consider the first statement of (3.6.53). We write

$$v_n'(t) - h'(t) = v_n'(t) - h_s'(t) + h_s'(t) - h'(t),$$

where h_s is defined in (3.6.46). For the second term we have

$$\sup_{t\in[\ell,M]} \left| h_s'(t) - h'(t) \right| \leqslant \sup_{t\in[\ell,M]} \int_{-1}^{1} |k(y)| \left| h'(t-by) - h'(t) \right| dy \to 0,$$

by the uniform continuity of h'. Moreover, similar to (3.6.49) and (3.6.51),

$$\sup_{t\in[\ell,M]} \left| v_n'(t) - h_s'(t) \right|$$
$$\leqslant \frac{1}{b^2} \sup_{t\in[\ell,M]} |H_n^{uc}(t) - H^{uc}(t)| \int_{-1}^{1} |k''(y)|\, dy$$
$$= O_p(n^{2\alpha-1/2}),$$

which tends to zero in probability, as $\alpha < 1/4$. To obtain the second statement of (3.6.53), first note that from (3.3.3),

$$w_n'(t;\hat{\beta}_n) = \int k_b'(t-u)\Phi_n(u;\hat{\beta}_n)\, du, \tag{3.6.54}$$

and write

$$w_n'(t;\hat{\beta}_n) - \Phi'(t;\beta_0) = w_n'(t;\hat{\beta}_n) - w_s'(t;\beta_0) + w_s'(t;\beta_0) - \Phi'(t;\beta_0), \tag{3.6.55}$$

where

$$w_s(t;\beta_0) = \int k_b(t-u)\Phi(u;\beta_0)\,du.$$

For the second difference on the right hand side of (3.6.55) we have

$$\begin{aligned}
&\sup_{t\in[\ell,M]} \left|w_s'(t;\beta_0) - \Phi'(t;\beta_0)\right| \\
&= \sup_{t\in[\ell,M]} \left|\int_{-1}^{1} k(y)\Phi'(t-by;\beta_0)\,dy - \Phi'(t;\beta_0)\right| \\
&\leqslant \sup_{t\in[\ell,M]} \int_{-1}^{1} |k(y)| \left|\Phi'(t-by;\beta_0) - \Phi'(t;\beta_0)\right| dy \to 0,
\end{aligned} \tag{3.6.56}$$

by uniform continuity of Φ'. Furthermore, with (3.6.54), we obtain

$$\sup_{t\in[\ell,M]} \left|w_n'(t;\hat{\beta}_n) - w_s'(t;\beta_0)\right| \leqslant \frac{1}{b} \sup_{t\in\mathbb{R}} |\Phi_n(t;\hat{\beta}_n) - \Phi(t;\beta_0)| \int_{-1}^{1} |k'(y)|\,dy,$$

which converges to zero because

$$\begin{aligned}
\sup_{t\in\mathbb{R}} |\Phi_n(t;\hat{\beta}_n) - \Phi(t;\beta_0)| &\leqslant \sup_{t\in\mathbb{R}} |\Phi_n(t;\hat{\beta}_n) - \Phi_n(t;\beta_0)| \\
&\quad + \sup_{t\in\mathbb{R}} |\Phi_n(t;\beta_0) - \Phi(t;\beta_0)| \\
&\leqslant \sup_{t\in\mathbb{R}} \left|\frac{\partial\Phi_n(t;\beta_n^*)}{\partial\beta}\right| (\hat{\beta}_n - \beta_0) + O_p(n^{-1/2}) \\
&= O_p(n^{-1/2}),
\end{aligned} \tag{3.6.57}$$

due to Lemmas 3 and 4 in Lopuhaä and Nane, 2013. Together with (3.6.56) this proves the last result. □

Lemma 3.6.11. *Suppose that (A1)-(A2) hold. Let* $H^{uc}(t)$ *and* $\Phi(t;\beta_0)$ *be defined in* (3.1.1) *and* (3.1.2)*, and let* $h(t) = dH^{uc}(t)/dt$*, satisfying* (3.1.3)*. Suppose that* h *and* $t \mapsto \Phi(t;\beta_0)$ *are continuously differentiable, and that* λ_0' *is uniformly bounded from below by a strictly positive constant. Let* k *satisfy* (1.2.1) *and let* $\hat{\lambda}_n^{\text{naive}}$ *be defined in* (3.3.7)*. If* $b \to 0$ *and* $1/b = O(n^{\alpha})$*, for some* $\alpha \in (0, 1/4)$*, then for each* $0 < \ell < M < \tau_H$*, it holds*

$$\mathbb{P}\big(\hat{\lambda}_n^{\text{naive}} \text{ is increasing on } [\ell, M]\big) \to 1.$$

Proof. Note that $w_n(x, \hat{\beta}_n) = 0$ if and only if $T_i \leqslant x - b$, for all $i = 1, \ldots, n$, which happens with probability $H(x-b)^n \leqslant H(M)^n \to 0$. This means that with probability tending to one, $w_n(x, \hat{\beta}_n) > 0$ for all $x \in [\ell, M]$. Thus with probability tending to one, $\hat{\lambda}_n^{\text{naive}}$ is well defined on $[\ell, M]$ and

$$\frac{d}{dt}\hat{\lambda}_n^{\text{naive}}(t) = \frac{v_n'(t)\, w_n(t; \hat{\beta}_n) - v_n(t)\, w_n'(t; \hat{\beta}_n)}{w_n(t; \hat{\beta}_n)^2}. \tag{3.6.58}$$

In order to prove that $\hat{\lambda}_n^{\text{naive}}$ is increasing on $[\ell, M]$ with probability tending to one, it suffices to show that

$$\mathbb{P}\left(\inf_{t\in[\ell,M]} \left\{v_n'(t) w_n(t; \hat{\beta}_n) - v_n(t) w_n'(t; \hat{\beta}_n)\right\} \leqslant 0\right) \to 0. \tag{3.6.59}$$

We can write

$$\begin{aligned}
&v_n'(t) w_n(t; \hat{\beta}_n) - v_n(t) w_n'(t; \hat{\beta}_n) \\
&= w_n(t; \hat{\beta}_n)\left(v_n'(t) - h'(t)\right) + v_n(t)\left(\Phi'(t; \beta_0) - w_n'(t; \hat{\beta}_n)\right) \\
&\quad + h'(t)\left(w_n(t; \hat{\beta}_n) - \Phi(t; \beta_0)\right) + \Phi'(t; \beta_0)\left(h(t) - v_n(t)\right) \\
&\quad + h'(t)\Phi(t; \beta_0) - \Phi'(t; \beta_0) h(t),
\end{aligned}$$

where the right hand side can be bounded from below by

- $\sup_{t\in[\ell,M]} |v_n'(t) - h'(t)| \sup_{t\in[\ell,M]} |w_n(t; \hat{\beta}_n)|$
- $\sup_{t\in[\ell,M]} |\Phi'(t; \beta_0) - w_n'(t; \hat{\beta}_n)| \sup_{t\in[\ell,M]} |v_n(t)|$
- $\sup_{t\in[\ell,M]} |w_n(t; \hat{\beta}_n) - \Phi(t; \beta_0)| \sup_{t\in[\ell,M]} |h'(t)|$
- $\sup_{t\in[\ell,M]} |h(t) - v_n(t)| \sup_{t\in[\ell,M]} |\Phi'(t; \beta_0)| + h'(t)\Phi(t; \beta_0) - \Phi'(t; \beta_0) h(t)$.

From the proof of Lemma 3.6.12 we have that

$$\sup_{t\in[\ell,M]} |v_n(t)| = O_P(1) \quad \text{and} \quad \sup_{t\in[\ell,M]} w_n(t; \hat{\beta}_n) = O_P(1),$$

so that from Lemmas 3.6.10 and 3.6.8 (with $m = 1$), it follows that the first four terms on the right hand side tend to zero in probability. Therefore, the probability in (3.6.59) is bounded by

$$\mathbb{P}\left(X_n \geqslant \inf_{t\in[\ell,M]} \left(h'(t)\Phi(t; \beta_0) - \Phi'(t; \beta_0) h(t)\right)\right),$$

where $X_n = o_p(1)$. This probability tends to zero, because with (3.1.3), we have

$$\begin{aligned}
\inf_{t\in[\ell,M]} \left(h'(t)\Phi(t; \beta_0) - \Phi'(t; \beta_0) h(t)\right) &= \inf_{t\in[\ell,M]} \lambda_0'(t)\Phi(t; \beta_0)^2 \\
&\geqslant \Phi(M; \beta_0)^2 \inf_{t\in[0,\tau_H]} \lambda_0'(t) > 0.
\end{aligned}$$

□

Lemma 3.6.12. *Let $\tilde{V}_n$, $\tilde{V}_n$, and W_0 be defined by* (3.3.8) *and* (3.3.11)*, and let H^{uc} be defined in* (3.1.1)*. If $b \to 0$ and $1/b = O(n^{-1/2})$, then, under the assumptions of Lemma 3.6.8, it holds*

$$\sup_{t\in[0,\tau_H]} \left|\tilde{V}_n(t) - H^{uc}(t)\right| \xrightarrow{\mathbb{P}} 0, \qquad \sup_{t\in[0,\tau_H]} \left|\tilde{W}_n(t) - W_0(t)\right| \xrightarrow{\mathbb{P}} 0. \quad (3.6.60)$$

Proof. To prove the first result in (3.6.60), we take $0 < \epsilon < \tau_H$ arbitrarily and write

$$\begin{aligned}
&\sup_{t\in[0,\tau_H]} |\tilde{V}_n(t) - H^{uc}(t)| \leqslant \int_0^{\tau_H} |v_n(u) - h(u)|\,du \\
&= \int_0^{\epsilon} |v_n(u) - h(u)|\,du + \int_{\epsilon}^{\tau_H-\epsilon} |v_n(u) - h(u)|\,du + \int_{\tau_H-\epsilon}^{\tau_H} |v_n(u) - h(u)|\,du \\
&\leqslant 2\epsilon \sup_{u\in[0,\tau_H]} |v_n(u)| + 2\epsilon \sup_{u\in[0,\tau_H]} |h(u)| + (\tau_H - 2\epsilon) \sup_{u\in[\epsilon,\tau_H-\epsilon]} |v_n(u) - h(u)|.
\end{aligned} \quad (3.6.61)$$

Since h is bounded and the last term tends to zero in probability, according to Lemma 3.6.8, it suffices to prove that $\sup_{u\in[0,\tau_H]} |v_n(u)| = O_p(1)$. By definition and the triangular inequality we have

$$\begin{aligned}
|v_n(t)| &= \left|\int k_b\,(t-u)\;dH_n^{uc}(u)\right| \\
&\leqslant b^{-1}\,|H_n^{uc}((t+b)\wedge\tau_H) - H_n^{uc}((t-b)\vee 0)| \sup_{y\in[-1,1]} |k(y)| \\
&\leqslant b^{-1}\Bigg\{\Big|H_n^{uc}((t+b)\wedge\tau_H) - H^{uc}((t+b)\wedge\tau_H) \\
&\qquad - H_n^{uc}((t-b)\vee 0) + H^{uc}((t-b)\vee 0)\Big| \\
&\qquad + H^{uc}((t+b)\wedge\tau_H) - H^{uc}((t-b)\vee 0)\Bigg\} \sup_{y\in[-1,1]} |k(y)| \\
&\leqslant 2\left\{\frac{1}{b} \sup_{y\in[0,\tau_H]} |H_n^{uc}(y) - H^{uc}(y)| + 2 \sup_{u\in[0,\tau_H]} |h(u)|\right\} \sup_{y\in[-1,1]} |k(y)|.
\end{aligned}$$

Using (3.6.50), it follows that the right hand side of the previous inequality is bounded in probability. For the second result in (3.6.60), similar to (3.6.61) we have

$$\begin{aligned}\sup_{t\in[0,\tau_H]}\left|\tilde{W}_n(t)-W_0(t)\right| &\leqslant \int_0^{\tau_H}\left|w_n(u;\hat{\beta}_n)-\Phi(u;\beta_0)\right|\,du\\ &\leqslant 2\epsilon\sup_{u\in[0,\tau_H]}|w_n(u;\hat{\beta}_n)|+2\epsilon\sup_{u\in[0,\tau_H]}|\Phi(u;\beta_0)|\\ &\quad+(\tau_H-2\epsilon)\sup_{u\in[\epsilon,\tau_H-\epsilon]}\left|w_n(u;\hat{\beta}_n)-\Phi(u;\beta_0)\right|.\end{aligned}$$

By using Lemma 3.6.8 and the fact that $\Phi(u;\beta_0)$ is bounded, it remains to handle the first term on right hand side. Since

$$\left|\int_t^\infty k_b(s-u)\,ds\right|=\left|\int_{(t-u)/b}^\infty k(y)\,dy\right|\leqslant 2\sup_{y\in[-1,1]}|k(y)|, \tag{3.6.62}$$

and $k_b(t-u)=0$, for $u<t-b$, we have

$$\begin{aligned}|w_n(t;\hat{\beta}_n)| &= \left|\int e^{\hat{\beta}_n' z}\int_t^\infty k_b(s-u)\,ds\,d\mathbb{P}_n(u,\delta,z)\right|\\ &\leqslant 2\sup_{y\in[-1,1]}|k(y)|\int_{t-b}^\infty e^{\hat{\beta}_n' z}\,d\mathbb{P}_n(u,\delta,z)\\ &= 2\Phi_n(t-b;\hat{\beta}_n)\sup_{y\in[-1,1]}|k(y)|,\end{aligned}$$

whereas Lemma 3 in Lopuhaä and Nane, 2013 gives that $\sup_{t\in\mathbb{R}}\Phi_n(t;\hat{\beta}_n)=O_p(1)$. This establishes the second result in (3.6.60). □

Proof of Corollary 3.3.3. According to Lemmas 3.6.9, 3.6.11 and 3.6.12, together with (3.1.3), conditions (b)-(d) of Lemma 3.3.2 are satisfied, with $X_n=\tilde{W}_n$, $Y_n=\tilde{V}_n$, and $\hat{\tau}=\sup\{t\geqslant 0: w_n(t;\hat{\beta}_n)>0\}$, and condition (a) of Lemma 3.3.2 is trivially fulfilled with $X_n=\tilde{W}_n$. Hence, the corollary follows from Lemma 3.3.2. □

Lemma 3.6.13. *Suppose that (A1)-(A2) hold. Fix* $t\in(0,\tau_H)$. *Let* H^{uc} *and* Φ *be defined in* (3.1.1) *and* (3.1.2), *and let* $h(s)=dH^{uc}(s)/ds$, *satisfying* (3.1.3). *Suppose that* h *and* $s\mapsto\Phi(s;\beta_0)$ *are* $m\geqslant 2$ *times continuously differentiable and that* λ_0' *is uniformly bounded from below by a strictly positive constant. Let* k *be* m*-orthogonal satisfying* (1.2.1). *Let* v_n *and* w_n *be defined in* (3.3.3) *and suppose that* $n^{1/(2m+1)}b\to c>0$. *Then*

$$n^{m/(2m+1)}\left(\begin{bmatrix}w_n(t;\hat{\beta}_n)\\ v_n(t)\end{bmatrix}-\begin{bmatrix}\Phi(t;\beta_0)\\ h(t)\end{bmatrix}\right)\to N\left(\begin{bmatrix}\mu_1\\ \mu_2\end{bmatrix},\begin{bmatrix}0 & 0\\ 0 & \sigma^2\end{bmatrix}\right)$$

where

$$\begin{bmatrix}\mu_1\\ \mu_2\end{bmatrix} = \frac{(-c)^m}{m!}\int_{-1}^{1} k(y)y^m\,dy \begin{bmatrix}\Phi^{(m)}(t;\beta_0)\\ h^{(m)}(t)\end{bmatrix}, \qquad \sigma^2 = \frac{h(t)}{c}\int_{-1}^{1} k^2(y)\,dy.$$

Proof. First we show that

$$n^{m/(2m+1)}(w_n(t;\hat\beta_n) - w_n(t;\beta_0)) \to 0$$

in probability, which enables us to replace $w_n(t;\hat\beta_n)$ with $w_n(t;\beta_0)$ in the statement. From (3.3.3), together with (3.6.62), we find

$$\begin{aligned}\left|w_n(t;\hat\beta_n) - w_n(t;\beta_0)\right| &\leqslant \frac{1}{n}\sum_{i=1}^{n}\left|e^{\hat\beta_n' Z_i} - e^{\beta_0' Z_i}\right|\left|\int_t^\infty k_b(u-T_i)\,du\right| \\ &\leqslant 2\sup_{y\in[-1,1]}|k(y)|\frac{1}{n}\sum_{i=1}^{n}|Z_i|e^{\tilde\beta_{n,i}' Z_i}\left|\hat\beta_n - \beta_0\right|,\end{aligned}$$

for some $|\tilde\beta_{n,i} - \beta_0| \leqslant |\hat\beta_n - \beta_0| = O_p(n^{-1/2})$. Furthermore, for all $M > 0$,

$$\begin{aligned}\mathbb{P}\left(\frac{1}{n}\sum_{i=1}^{n}|Z_i|e^{\tilde\beta_{n,i}' Z_i} \geqslant M\right) &\leqslant \frac{1}{nM}\sum_{i=1}^{n}\mathbb{E}\left[|Z_i|e^{\tilde\beta_{n,i}' Z_i}\right] \\ &\leqslant \frac{1}{M}\sup_{|\beta-\beta_0|\leqslant\epsilon}\mathbb{E}\left[|Z|e^{\beta' Z}\right],\end{aligned}$$

where $\sup_{|\beta-\beta_0|\leqslant\epsilon}\mathbb{E}[|Z|e^{\beta' Z}] < \infty$ according to assumption (A2). It follows that

$$n^{m/(2m+1)}\left(w_n(t;\hat\beta_n) - w_n(x;\beta_0)\right) = O_p(n^{-1/(4m+2)}).$$

Now, define

$$Y_{ni} = \begin{bmatrix}Y_{ni,1}\\ Y_{ni,2}\end{bmatrix} = n^{-(m+1)/(2m+1)}\begin{bmatrix}e^{\beta_0' Z_i}\int_t^\infty k_b(s-T_i)\,ds\\ k_b(t-T_i)\Delta_i\end{bmatrix}.$$

By a Taylor expansion, using that h is m times continuously differentiable and that k is m-orthogonal, as in (3.6.47) we obtain

$$\begin{aligned}\mathbb{E}\left[Y_{ni,2}\right] &= n^{-(m+1)/(2m+1)}\int_{-1}^{1} k(y)h(t-by)\,dy \\ &= n^{-\frac{m+1}{2m+1}}\left(h(t) + \frac{(-b)^m}{m!}h^{(m)}(t)\int_{-1}^{1} k(y)y^m\,dy + o(b^m)\right).\end{aligned} \tag{3.6.63}$$

Similarly, with Fubini we get

$$\begin{aligned}
\mathbb{E}\left[Y_{ni,1}\right] &= n^{-(m+1)/(2m+1)} \int e^{\beta_0' z} \int_t^\infty k_b(s-u)\,ds\,d\mathbb{P}(u,\delta,z) \\
&= n^{-(m+1)/(2m+1)} \int_{-1}^{1} \left(\int e^{\beta_0' z} \mathbb{1}_{\{u \geqslant t-by\}}\,d\mathbb{P}(u,\delta,z) \right) k(y)\,dy \\
&= n^{-(m+1)/(2m+1)} \int_{-1}^{1} k(y)\Phi(t-by;\beta_0)\,dy \\
&= n^{-\frac{m+1}{2m+1}} \left(\Phi(t;\beta_0) + \frac{(-b)^m}{m!}\Phi^{(m)}(t;\beta_0) \int_{-1}^{1} k(y)y^m\,dy + o(b^m) \right).
\end{aligned} \tag{3.6.64}$$

Hence, we have

$$\mathbb{E}\left[Y_{ni}\right] = n^{-(m+1)/(2m+1)} \begin{bmatrix} \Phi(t;\beta_0) \\ h(t) \end{bmatrix} + n^{-1} \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} + o(n^{-1}),$$

and we can write

$$n^{-\frac{m+1}{2m+1}} \left(\begin{bmatrix} w_n(t;\hat{\beta}_n) \\ v_n(t) \end{bmatrix} - \begin{bmatrix} \Phi(t;\beta_0) \\ h(t) \end{bmatrix} \right) = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} + \sum_{i=1}^{n} \left(Y_{ni} - \mathbb{E}\left[Y_{ni}\right] \right) + o(1).$$

It remains to show that $\sum_{i=1}^{n} \left(Y_{ni} - \mathbb{E}\left[Y_{ni}\right] \right)$ converges in distribution to a bivariate normal distribution with mean zero. From (3.6.64) we have,

$$\begin{aligned}
&\mathrm{Var}(Y_{ni,1}) = \mathbb{E}\left[Y_{ni,1}^2\right] + O(n^{-2(m+1)/(2m+1)}) \\
&= n^{-2\frac{m+1}{2m+1}} \int e^{2\beta_0' z} \left(\int_t^\infty k_b(s-u)\,ds \right)^2 d\mathbb{P}(u,\delta,z) + O\left(n^{-2\frac{m+1}{2m+1}}\right) \\
&= O\left(n^{-2(m+1)/(2m+1)}\right),
\end{aligned} \tag{3.6.65}$$

using that, with (3.6.62),

$$\begin{aligned}
&\int e^{2\beta_0' z} \left(\int_t^\infty k_b(s-u)\,ds \right)^2 d\mathbb{P}(u,\delta,z) \\
&\leqslant \left(2 \sup_{y\in[-1,1]} |k(y)| \right)^2 \int e^{2\beta_0' z}\,d\mathbb{P}(u,\delta,z) \\
&= \left(2 \sup_{y\in[-1,1]} |k(y)| \right)^2 \Phi(0;2\beta_0) < \infty.
\end{aligned}$$

Moreover,

$$\begin{aligned}
&\mathrm{Cov}(Y_{ni,1}, Y_{ni,2}) = \mathbb{E}\left[Y_{ni,1}Y_{ni,2}\right] + O\left(n^{-2(m+1)/(2m+1)}\right) \\
&= n^{-2\frac{m+1}{2m+1}} \int \delta e^{\beta_0' z} \left(\int_t^\infty k_b(s-u)\,ds\right) k_b(t-u)\,d\mathbb{P}(u,\delta,z) + O_p\left(n^{-2\frac{m+1}{2m+1}}\right) \\
&= o(n^{-1}) + O\left(n^{-2(m+1)/(2m+1)}\right),
\end{aligned}$$

because, with (3.6.62),

$$\begin{aligned}
&\left| b \int \delta e^{\beta_0' z} \left(\int_t^\infty k_b(s-u)\,ds\right) k_b(t-u)\,d\mathbb{P}(u,\delta,z)\right| \\
&\leqslant 2 \sup_{y\in[-1,1]} |k(y)| \int \mathbb{1}_{\{x-b\leqslant u\leqslant x+b\}} e^{\beta_0' z} \left| k\left(\frac{x-u}{b}\right)\right| d\mathbb{P}(u,\delta,z) \\
&\leqslant 2 \left(\sup_{y\in[-1,1]} |k(y)|\right)^2 \left(\Phi(t-b;\beta_0) - \Phi(t+b;\beta_0)\right) \to 0.
\end{aligned}$$

Once again, by a Taylor expansion, from (3.6.63), we obtain

$$\begin{aligned}
\mathrm{Var}(Y_{ni,2}) &= \mathbb{E}\left[Y_{ni,2}^2\right] + O\left(n^{-2(m+1)/(2m+1)}\right) \\
&= n^{-2\frac{m+1}{2m+1}} b^{-1} \int_{-1}^{1} k^2(y) h(t-by)\,dy + O\left(n^{-2\frac{m+1}{2m+1}}\right) \qquad (3.6.66) \\
&= n^{-1}\sigma^2 + o(n^{-1}).
\end{aligned}$$

It follows that

$$\sum_{i=1}^n \mathrm{Cov}(Y_{ni}) = \begin{bmatrix} 0 & 0 \\ 0 & \sigma^2 \end{bmatrix} + o(1).$$

Furthermore, since

$$|Y_{ni}|^2 = n^{-2(m+1)/(2m+1)} \left(e^{2\beta_0' Z_i} \left(\int_t^\infty k_b(s-T_i)\,ds\right)^2 + k_b^2(t-T_i)\Delta_i\right),$$

with (3.6.62), we obtain

$$\begin{aligned}
\sum_{i=1}^n \mathbb{E}\left[|Y_{ni}|^2 \mathbb{1}_{\{|Y_{ni}|>\epsilon\}}\right] &\leqslant \left(2 \sup_{y\in[-1,1]} |k(y)|\right)^2 n^{-1/(2m+1)} \mathbb{E}\left[e^{2\beta_0' Z}\right] \\
&\quad + n^{-2\frac{m+1}{2m+1}} b^{-2} \sup_{y\in[-1,1]} |k(y)| \sum_{i=1}^n \mathbb{P}\left(|Y_{ni}| > \epsilon\right),
\end{aligned}$$

where the right hand side tends to zero, because $\mathbb{E}[e^{2\beta_0' Z}] = \Phi(0; 2\beta_0) < \infty$ and, with (3.6.65) and (3.6.66), we have

$$\sum_{i=1}^n \mathbb{P}\left(|Y_{ni}| > \epsilon\right) \leqslant \epsilon^{-2} \sum_{i=1}^n \mathbb{E}|Y_{ni}|^2 = O(1).$$

By the multivariate Lindeberg-Feller central limit theorem, we get

$$\sum_{i=1}^{n} \left(Y_{ni} - \mathbb{E}\left[Y_{ni}\right]\right) \xrightarrow{d} N\left(\begin{bmatrix}\mu_1\\ \mu_2\end{bmatrix}, \begin{bmatrix}0 & 0\\ 0 & \sigma^2\end{bmatrix}\right),$$

which finishes the proof. □

Proof of Theorem 3.3.5. By definition of $\hat{\lambda}_n^{\text{naive}}(t)$ in (3.3.7) together with (3.1.3), we can write

$$\hat{\lambda}_n^{\text{naive}}(t) - \lambda_0(t) = \varphi\Big(w_n(t;\hat{\beta}_n), v_n(t)\Big) - \varphi\Big(\Phi(t;\beta_0), \lambda_0(t)\Phi(t;\beta_0)\Big)$$

with $\varphi(w,v) = v/w$. The asymptotic distribution of $\hat{\lambda}_n^{\text{naive}}(t)$ then follows from an application of the delta method to the result in Lemma 3.6.13. Then, by Corollary 3.3.3, this also gives the asymptotic distribution of $\hat{\lambda}_n^{MS}(t)$. □

Proof of Theorem 3.3.6. First note that by means of (3.1.3), it follows from the assumptions of the theorem that $h(s) = dH^{uc}(s)/ds$ is $m \geqslant 2$ times continuously differentiable. We write

$$\begin{aligned}
&n^{m/(2m+1)}\left(\hat{\lambda}_n^{MS}(t) - \hat{\lambda}_n^{SM}(t)\right)\\
&= n^{m/(2m+1)}\left(\hat{\lambda}_n^{\text{naive}}(t) - \hat{\lambda}_n^{SM}(t)\right) + n^{m/(2m+1)}\left(\hat{\lambda}_n^{MS}(t) - \hat{\lambda}_n^{\text{naive}}(t)\right).
\end{aligned}$$

By Corollary 3.3.3, the second term on the right hand side converges to zero in probability. Furthermore, as can be seen from the proof of Theorem 3.2.5

$$n^{\frac{m}{2m+1}}\left(\hat{\lambda}_n^{SM}(t) - \lambda_0(t)\right) = \mu + n^{\frac{m}{2m+1}} \int \frac{\delta k_b(t-u)}{\Phi(u;\beta_0)}\, d(\mathbb{P}_n - \mathbb{P})(u,\delta,z) + o_p(1),$$

with μ from (3.2.22). From the proof of Lemma 3.6.13, we have

$$\hat{\lambda}_n^{\text{naive}}(t) - \lambda_0(t) = \phi\Big(w_n(t;\hat{\beta}_n), v_n(t)\Big) - \phi\Big(\Phi(t;\beta_0), \lambda_0(t)\Phi(t;\beta_0)\Big),$$

where $\phi(w,v) = v/w$ and

$$n^{m/(2m+1)}\left(\begin{bmatrix} w_n(t;\hat{\beta}_n)\\ v_n(t)\end{bmatrix} - \begin{bmatrix}\Phi(t;\beta_0)\\ h(t)\end{bmatrix}\right) = \begin{bmatrix}\mu_1\\ \mu_2\end{bmatrix} + \begin{bmatrix}Z_{n1}\\ Z_{n2}\end{bmatrix} + o(1),$$

with $Z_{n1} = o_P(1)$ and

$$\begin{bmatrix}\mu_1\\ \mu_2\end{bmatrix} = \frac{(-c)^m}{m!}\int_{-1}^{1} k(y)y^m\, dy \begin{bmatrix}\Phi^{(m)}(t;\beta_0)\\ h^{(m)}(t)\end{bmatrix}.$$

Then with a Taylor expansion it follows that

$$\begin{aligned}
&n^{m/(2m+1)}\left(\hat{\lambda}_n^{\text{naive}}(t)-\lambda_0(t)\right)\\
&=\begin{bmatrix}-\dfrac{\lambda_0(t)}{\Phi(t;\beta_0)} & \dfrac{1}{\Phi(t;\beta_0)}\end{bmatrix}\left(\begin{bmatrix}\mu_1\\ \mu_2\end{bmatrix}+\begin{bmatrix}Z_{n1}\\ Z_{n2}\end{bmatrix}\right)+o_p(1)\\
&=\widetilde{\mu}+\frac{Z_{n2}}{\Phi(t;\beta_0)}+o_p(1),
\end{aligned}$$

where $\widetilde{\mu}$ is from Theorem 3.3.5. Moreover, from the proof of Lemma 3.6.13 it can be seen that

$$\begin{aligned}
\frac{Z_{n2}}{\Phi(t;\beta_0)}&=\frac{1}{\Phi(t;\beta_0)}n^{m/(2m+1)}\int \delta k_b(t-u)\,(\mathbb{P}_n-\mathbb{P})(u,\delta,z)+o_P(1)\\
&=n^{m/(2m+1)}\int \frac{\delta k_b(t-u)}{\Phi(u;\beta_0)}\,(\mathbb{P}_n-\mathbb{P})(u,\delta,z)\\
&\quad+n^{\frac{m}{2m+1}}\int \delta k_b(t-u)\left(\frac{1}{\Phi(t;\beta_0)}-\frac{1}{\Phi(u;\beta_0)}\right)(\mathbb{P}_n-\mathbb{P})(u,\delta,z)\\
&\quad+o_P(1)\\
&=n^{m/(2m+1)}\int \frac{\delta k_b(t-u)}{\Phi(u;\beta_0)}\,(\mathbb{P}_n-\mathbb{P})(u,\delta,z)+o_P(1),
\end{aligned}$$

because

$$\begin{aligned}
&n^{m/(2m+1)}\int \delta k_b(t-u)\left(\frac{1}{\Phi(t;\beta_0)}-\frac{1}{\Phi(u;\beta_0)}\right)(\mathbb{P}_n-\mathbb{P})(u,\delta,z)\\
&=\sum_{i=1}^{n}\left(X_{ni}-\mathbb{E}\left[X_{ni}\right]\right)
\end{aligned}$$

with

$$X_{ni}=n^{-(m+1)/(2m+1)}\Delta_i k_b(t-T_i)\left(\frac{1}{\Phi(t;\beta_0)}-\frac{1}{\Phi(T_i;\beta_0)}\right),$$

where similar to the proof of Lemma 3.6.13,

$$\begin{aligned}
\mathbb{E}\left[X_{ni}^2\right]&=n^{-2(m+1)/(2m+1)}\int k_b^2(t-u)\left(\frac{1}{\Phi(t;\beta_0)}-\frac{1}{\Phi(u;\beta_0)}\right)^2 h(u)\,du\\
&=n^{-2\frac{m+1}{2m+1}}b^{-1}\int k^2(y)\left(\frac{1}{\Phi(t;\beta_0)}-\frac{1}{\Phi(t-by;\beta_0)}\right)^2 h(t-by)\,dy\\
&=o(n^{-1}).
\end{aligned}$$

We conclude that

$$\begin{aligned}
n^{\frac{m}{2m+1}}\left(\hat{\lambda}_n^{\text{naive}}(t)-\lambda_0(t)\right)&=\widetilde{\mu}+n^{\frac{m}{2m+1}}\int \frac{\delta k_b(t-u)}{\Phi(u;\beta_0)}\,(\mathbb{P}_n-\mathbb{P})(u,\delta,z)+o_P(1)\\
&=\widetilde{\mu}-\mu+n^{\frac{m}{2m+1}}\left(\tilde{\lambda}_n^{SM}(t)-\lambda_0(t)\right)+o_p(1)
\end{aligned}$$

which proves the first statement in the theorem. The second statement is immediate using the asymptotic equivalence in (3.2.23). □

3.6.3 *Proofs for Section 3.4*

Lemma 3.6.14. *Suppose that (A1)-(A2) hold. Let λ_0 be continuously differentiable, with λ_0' uniformly bounded from below by a strictly positive constant, and let k satisfy (1.2.1). If $b \to 0$ and $1/b = O(n^\alpha)$, for some $\alpha \in (0, 1/4)$, then for each $0 < \ell < M < \tau^*$, it holds*

$$\mathbb{P}\left(\tilde{\lambda}_n^{\text{naive}} \textit{ is increasing on } [\ell, M]\right) \to 1.$$

Proof. From (3.4.3), it follows with integration by parts that

$$\begin{aligned}\tilde{\lambda}_n^{\text{naive}}(t) &= \int k_b'(t-u)\Lambda_0(u)\,du + \int k_b'(t-u)\left(\Lambda_n(u) - \Lambda_0(u)\right)\,du \\ &= \lambda_0(t) + \int k_b(t-u)\{\lambda_0(u) - \lambda_0(t)\}\,du \\ &\quad + \int k_b(t-u)\,d\left(\Lambda_n - \Lambda_0\right)(u),\end{aligned} \tag{3.6.67}$$

so that

$$\begin{aligned}\frac{d}{dt}\tilde{\lambda}_n^{\text{naive}}(t) = \lambda_0'(t) &+ \int_{-1}^{1} k(y)\{\lambda_0'(t-by) - \lambda_0'(t)\}\,dy \\ &+ \frac{1}{b^2}\int k'\left(\frac{t-u}{b}\right)\,d\left(\Lambda_n - \Lambda_0\right)(u).\end{aligned} \tag{3.6.68}$$

By assumption, the first term on the right hand side of (3.6.68) is bounded from below by a strictly positive constant and the second term converges to zero because of the continuity of λ_0'. Moreover, for n sufficiently large $M + b < \tau^*$. Then, the second term on the right hand side of (3.6.68) is bounded from above in absolute value by

$$\frac{1}{b^2}\sup_{s\in[0,\tau^*]}|\Lambda_n(s) - \Lambda_0(s)| \sup_{y\in[-1,1]}|k''(y)| = O_p(n^{2\alpha-1/2}) = o_p(1),$$

according to (3.1.7) and the fact that $\alpha < 1/4$. We conclude that $\tilde{\lambda}_n^{\text{naive}}$ is increasing on $[\ell, M]$ with probability tending to one. □

Proof of Corollary 3.4.1. We apply Lemma 3.3.2. Condition (a) is trivial with $X_n(t) = t$. Furthermore, for every fixed $t \in (0, \tau^*)$, we have for sufficiently large n, that $t \in (b, \tau^* - b)$ and

$$\begin{aligned}
&\tilde{\lambda}_n^{\text{naive}}(t) - \lambda_0(t) \\
&= \int k_b(t-u)\lambda_0(u)\,du - \lambda_0(t) + \int k_b(t-u)\,d\left(\Lambda_n(u) - \Lambda_0(u)\right) \\
&= \int_{-1}^{1} k(y)\left\{\lambda_0(t-by) - \lambda_t(t)\right\}dy + \frac{1}{b}\int_{-1}^{1}\left(\Lambda_n(t-by) - \Lambda_0(t-by)\right)k'(y)\,dy \\
&= o_p(1),
\end{aligned} \tag{3.6.69}$$

by continuity of λ_0 and (3.1.7), which proves condition (b) of Lemma 3.3.2. Condition (c) follows from Lemma 3.6.14. Finally, for $t \in [0, \tau^*]$,

$$\begin{aligned}
\left|\hat{\Lambda}_n^s(t) - \Lambda_0(t)\right| &= \left|\int_{(t-b)\vee 0}^{(t+b)\wedge\tau^*} k_b(t-u)\left(\Lambda_n(u) - \Lambda_0(u)\right)du\right. \\
&\qquad \left. + \int_{(t-b)\vee 0}^{(t+b)\wedge\tau^*} k_b(t-u)\Lambda_0(u)\,du - \Lambda_0(t)\right| \\
&\leqslant \sup_{x\in[0,\tau^*]}\left|\Lambda_n(t) - \Lambda_0(t)\right|\int_{-1}^{1}|k(y)|\,dy \\
&\qquad + \left|\int_{-1\vee(t-\tau^*)/b}^{t/b\wedge 1} k(y)\Lambda_0(t-by)\,dy - \Lambda_0(t)\right|.
\end{aligned}$$

According to (3.1.7), the first term on the right hand side is of the order $O_p(n^{-1/2})$. For the second term we distinguish between $t < b$, $t \in [b, \tau^* - b]$ and $t > \tau^* - b$. When $t \in [b, \tau^* - b]$, then with (1.2.1),

$$\begin{aligned}
\left|\int_{-1}^{1} k(y)\Lambda_0(t-by)\,dy - \Lambda_0(t)\right| &\leqslant \int_{-1}^{1}|k(y)|\left|\Lambda_0(t-by) - \Lambda_0(t)\right|dy \\
&= b\sup_{t\in[0,\tau_H]}|\lambda_0(t)|\int_{-1}^{1}|k(y)|\,dy \to 0,
\end{aligned}$$

uniformly for $t \in [b, \tau^* - b]$. When $t < b$, then again with (1.2.1), we can write

$$\begin{aligned}
&\left|\int_{-1}^{t/b} k(y)\Lambda_0(t-by)\,dy - \Lambda_0(t)\right| \\
&\leqslant \int_{-1}^{t/b}|k(y)|\left|\Lambda_0(t-by) - \Lambda_0(t)\right|dy + \Lambda_0(t)\int_{t/b}^{1}|k(y)|\,dy \\
&\leqslant O(b) + b\lambda_0(b)\int_{-1}^{1}|k(y)|\,dy \to 0,
\end{aligned}$$

uniformly for $t \in [0, b]$. Similarly, we deal with the case $t \in (\tau^* - b, \tau^*]$. It follows that

$$\sup_{t \in [0,\tau^*]} \left|\hat{\Lambda}_n^s(t) - \Lambda_0(t)\right| = o_P(1), \tag{3.6.70}$$

which proves condition (d) of Lemma 3.6.11. The result now follows from Lemma 3.6.14. □

Proof of Theorem 3.4.3. From (3.6.69), similar to (3.6.47), we find

$$\begin{aligned}
\tilde{\lambda}_n^{\text{naive}}(t) - \lambda_0(t) &= \int_{-1}^{1} k(y)\{\lambda_0(t - by) - \lambda_0(t)\}\, dy \\
&\quad + \frac{1}{b}\int_{-1}^{1} (\Lambda_n(t - by) - \Lambda_0(t - by))\, k'(y)\, dy \\
&= \frac{(-b)^m}{m!}\int_{-1}^{1} \lambda_0^{(m)}(\xi_{ty}) k(y) y^m\, dy \\
&\quad + \frac{1}{b}\int_{-1}^{1} (\Lambda_n(t - by) - \Lambda_0(t - by))\, k'(y)\, dy,
\end{aligned}$$

for some $|\xi_{ty} - x| \leqslant |by|$. It follows that

$$\begin{aligned}
\sup_{t \in [\ell, M]} \left|\tilde{\lambda}_n^{\text{naive}}(t) - \lambda_0(t)\right| &\leqslant \frac{b^m}{m!} \sup_{t \in [0,\tau_H]} \left|\lambda_0^{(m)}(t)\right| \int_{-1}^{1} |k(y)||y|^m\, dy \\
&\quad + b^{-1} \sup_{x \in [\ell, M]} |\Lambda_n(t) - \Lambda_0(t)| \int_{-1}^{1} |k'(y)|\, dy \\
&= o_p(1).
\end{aligned}$$

Similar to (3.6.48), the first term on the right hand side is of the order $O(b^m)$, and according to (3.1.7) the second term is of the order $O_p(b^{-1}n^{-1/2})$. The first statement now follows directly from Corollary 3.4.1.

To obtain the asymptotic distribution, note that from (3.6.67), (3.1.3) and (3.1.7), we have

$$\begin{aligned}
&n^{m/(2m+1)} \left(\tilde{\lambda}_n^{\text{naive}}(t) - \lambda_0(t)\right) \\
&= n^{m/(2m+1)} \left(\int k_b(t - u)\, \lambda_0(u)\, du - \lambda_0(t)\right) \\
&\quad + n^{m/(2m+1)} \int k_b(t - u) \frac{\delta}{\Phi(u; \beta_0)} d(\mathbb{P}_n - \mathbb{P})(u, \delta, z) \\
&\quad + n^{-(m+1)/(2m+1)} \sum_{i=1}^{n} k_b(t - T_i) \Delta_i \left(\frac{1}{\Phi_n(T_i; \hat{\beta}_n)} - \frac{1}{\Phi(T_i; \beta_0)}\right).
\end{aligned} \tag{3.6.71}$$

We find that, the first term in the right hand side of (3.6.71) converges to μ, since

$$\begin{aligned} & n^{m/(2m+1)} \left(\int k_b(t-u)\, \lambda_0(u)\, du - \lambda_0(t) \right) \\ & = n^{m/(2m+1)} \int_{-1}^{1} k(y) \{\lambda_0(t-by) - \lambda_0(t)\}\, dy \\ & = n^{m/(2m+1)} \frac{(-b)^m}{m!} \int_{-1}^{1} \lambda_0^{(m)}(\xi_{ty}) k(y) y^m\, dy \\ & \to \frac{(-c)^m}{m!} \lambda_0^{(m)}(t) \int_{-1}^{1} k(y) y^m\, dy, \end{aligned} \tag{3.6.72}$$

for some $|\xi_{ty} - x| \leqslant |by|$. Let $0 < M < M' < \tau_H$, so that $t + b \leqslant M'$ for sufficiently large n. Because $1/\Phi_n(M'; \hat{\beta}_n) = O_p(1)$, similar to (3.6.57)

$$\begin{aligned} \sup_{u \in [0, M']} \left| \frac{1}{\Phi_n(u; \hat{\beta}_n)} - \frac{1}{\Phi(u; \beta_0)} \right| & \leqslant \sup_{u \in [0, M']} \left| \frac{\Phi_n(u; \hat{\beta}_n) - \Phi(u; \beta_0)}{\Phi_n(M'; \hat{\beta}_n) \Phi(M'; \beta_0)} \right| \\ & = O_p(n^{-1/2}), \end{aligned}$$

and similar to (3.6.66)

$$\operatorname{Var}\left(n^{-(m+1)/(2m+1)} \sum_{i=1}^{n} |k_b(t - T_i)|\, \Delta_i \right) = O(n^{-1}),$$

so that the last term on the right hand side of (3.6.71) converges to zero in probability. The second term on the right hand side of (3.6.71) can be written as

$$\sum_{i=1}^{n} \left(Y_{ni} - \mathbb{E}\left[Y_{ni} \right] \right), \quad Y_{ni} = n^{-(m+1)/(2m+1)} k_b(t - T_i) \frac{\Delta_i}{\Phi(T_i; \beta_0)}.$$

where similar to (3.6.66),

$$\begin{aligned} \operatorname{Var}(Y_{ni}) & = \mathbb{E}\left[Y_{ni}^2 \right] + O(n^{-2(m+1)/(2m+1)}) \\ & = n^{-2\frac{m+1}{2m+1}} \frac{1}{b} \int_{-1}^{1} \frac{k^2(y) h(t - by)}{\Phi^2(t - by; \beta_0)}\, dy + O\left(n^{-2\frac{m+1}{2m+1}} \right) \\ & = n^{-1} \sigma^2 + o(n^{-1}). \end{aligned} \tag{3.6.73}$$

Moreover,

$$\sum_{i=1}^{n} \mathbb{E}\left[|Y_{ni}|^2 \mathbb{1}_{\{|Y_{ni}| > \epsilon\}} \right] \leqslant n^{-2\frac{m+1}{2m+1}} \frac{1}{b^2} \sup_{y \in [-1,1]} |k(y)| \sum_{i=1}^{n} \mathbb{P}\left(|Y_{ni}| > \epsilon \right),$$

where the right hand side tends to zero, because with (3.6.73),

$$\sum_{i=1}^{n} \mathbb{P}\left(|Y_{ni}| > \epsilon\right) \leqslant \sum_{i=1}^{n} \frac{\mathbb{E}|Y_{ni}|^2}{\epsilon^2} = O(1).$$

By Lindeberg-Feller central limit theorem, we obtain

$$\sum_{i=1}^{n} \left(Y_{ni} - \mathbb{E}\left[Y_{ni}\right]\right) \xrightarrow{d} N(0, \sigma^2),$$

which determines the asymptotic distribution of $\tilde{\lambda}_n^{\text{naive}}(t)$. Then, by Corollary 3.4.1, this also gives the asymptotic distribution of $\tilde{\lambda}_n^{GS}(t)$. □

Proof of Theorem 3.4.4. We write

$$\begin{aligned} &n^{m/(2m+1)}\left(\tilde{\lambda}_n^{SG}(t) - \tilde{\lambda}_n^{GS}(t)\right) \\ &= n^{m/(2m+1)}\left(\tilde{\lambda}_n^{SG}(t) - \tilde{\lambda}_n^{naive}(t)\right) + n^{m/(2m+1)}\left(\tilde{\lambda}_n^{naive}(t) - \tilde{\lambda}_n^{GS}(t)\right). \end{aligned}$$

By Corollary 3.4.1, the second term on the right hand side converges to zero in probability. Furthermore, as can be seen from the proof of Theorem 3.2.5,

$$\begin{aligned} n^{m/(2m+1)}\tilde{\lambda}_n^{SG}(t) &= n^{m/(2m+1)} \int k_b(t-u)\, d\Lambda_0(u) \\ &\quad + n^{m/(2m+1)} \int \frac{\delta k_b(t-u)}{\Phi(u;\beta_0)}\, d(\mathbb{P}_n - \mathbb{P})(u,\delta,z) + o_p(1). \end{aligned}$$

Similarly, from the proof of Theorem 3.4.3, we have

$$\begin{aligned} n^{m/(2m+1)}\tilde{\lambda}_n^{naive}(t) &= n^{m/(2m+1)} \int k_b'(t-u)\Lambda_n(u)\, du \\ &= n^{m/(2m+1)} \int k_b(t-u)\, d\Lambda_n(u) \\ &= n^{\frac{m}{2m+1}} \int k_b(t-u)\, d\Lambda_0(u) + n^{\frac{m}{2m+1}} \int k_b(t-u)\, d(\Lambda_n(u) - \Lambda_0(u)) \\ &= n^{\frac{m}{2m+1}} \int k_b(t-u)\, d\Lambda_0(u) + n^{\frac{m}{2m+1}} \int \frac{\delta k_b(t-u)}{\Phi(u;\beta_0)}\, d(\mathbb{P}_n - \mathbb{P})(u,\delta,z) + o_p(1). \end{aligned}$$

From this it immediately follows that

$$n^{m/(2m+1)}\left(\tilde{\lambda}_n^{SG}(t) - \tilde{\lambda}_n^{naive}(t)\right) = o_p(1),$$

and hence, with Corollary 3.4.1, also

$$n^{m/(2m+1)}\left(\tilde{\lambda}_n^{SG}(t) - \tilde{\lambda}_n^{GS}(t)\right) = o_p(1).$$

The second statement about $\hat{\lambda}_n^{SM}(t)$, is immediate using the asymptotic equivalence in (3.2.23). □

Part III

GLOBAL ASYMPTOTIC INFERENCE

4

THE HELLINGER LOSS OF GRENANDER-TYPE ESTIMATORS

In this chapter we consider the Hellinger loss of Grenander type estimators for a monotone function $\lambda : [0, 1] \to \mathbb{R}^+$. The results presented are based on:

Lopuhaä, H. P. and Musta E. (2018) "A central limit theorem for the Hellinger loss of Grenander-type estimators". *To appear in Statistica Neerlandica*.

The Hellinger distance is a convenient metric in maximum likelihood problems, which goes back to Le Cam, 1973; Le Cam, 1970, and it has nice connections with Bernstein norms and empirical process theory methods to obtain rates of convergence, due fundamentally to Birgé and Massart, 1993, Wong and Shen, 1995, and others, see Section 3.4 of van der Vaart and Wellner, 1996 or Chapter 4 in van de Geer, 2000 for a more detailed overview. Consistency in Hellinger distance of shape constrained maximum likelihood estimators has been investigated in Pal, Woodroofe, and Meyer, 2007, Seregin and Wellner, 2010, and Doss and Wellner, 2016, whereas rates on Hellinger risk measures have been obtained in Seregin and Wellner, 2010, Kim and Samworth, 2016, and Kim, Guntuboyina, and Samworth, 2016. However, there is no distribution theory available for the Hellinger loss of shape constrained nonparametric estimators.

We present a first result in this direction, i.e., a central limit theorem for the Hellinger loss of Grenander type estimators for a monotone function λ. The result applies to statistical models that satisfy the setup in Durot, 2007, which includes estimation of a probability density, a regression function, or a failure rate under monotonicity constraints.

In fact, after approximating the squared Hellinger distance by a weighted L_2-distance, a central limit theorem can be obtained by mimicking the approach introduced in Durot, 2007. An interesting feature of our main result is that in the monotone density model, the variance of the limiting normal distribution for the Hellinger distance does not depend on the underlying density. This phenomena was also encountered for the L_1-distance in Groeneboom, 1983; Groeneboom, Hooghiemstra, and Lopuhaä, 1999.

In Section 4.1 we define the setup and approximate the squared Hellinger loss by a weighted L_2-distance. A central limit theorem for the Hellinger distance is established in Section 4.2. A short discussion on the consequences for particular statistical models can be found in Section 4.3. The chapter ends with a a simulation study on testing exponentiality against a non-increasing density by means of the Hellinger distance (Section 4.4).

4.1 RELATION TO THE L_2-DISTANCE

Consider the problem of estimating a non-increasing (or non-decreasing) function $\lambda : [0,1] \to \mathbb{R}^+$ on the basis of n observations. Suppose that we have at hand a cadlag step estimator Λ_n for

$$\Lambda(t) = \int_0^t \lambda(u)\,du, \qquad t \in [0,1].$$

If λ is non-increasing, then the Grenander-type estimator $\hat{\lambda}_n$ for λ is defined as the left-hand slope of the least concave majorant (LCM) of Λ_n, with $\hat{\lambda}_n(0) = \lim_{t\downarrow 0} \hat{\lambda}_n(t)$. If λ is non-decreasing, then the Grenander-type estimator $\hat{\lambda}_n$ for λ is defined as the left-hand slope of the greatest convex minorant (GCM) of Λ_n, with $\hat{\lambda}_n(0) = \lim_{t\downarrow 0} \hat{\lambda}_n(t)$. We aim at proving the asymptotic normality of the Hellinger distance between $\hat{\lambda}_n$ and λ defined by

$$H(\hat{\lambda}_n, \lambda) = \left(\frac{1}{2} \int_0^1 \left(\sqrt{\hat{\lambda}_n(t)} - \sqrt{\lambda(t)} \right)^2 dt \right)^{1/2}. \tag{4.1.1}$$

We will consider the same general setup as in Durot, 2007, i.e., we will assume the following conditions

(A1) λ is monotone and differentiable on $[0,1]$ with

$$0 < \inf_t |\lambda'(t)| \leqslant \sup_t |\lambda'(t)| < \infty.$$

(A2′) Let $M_n = \Lambda_n - \Lambda$. There exist $C > 0$ such that for all $x > 0$ and $t \in \{0,1\}$,

$$\mathbb{E}\left[\sup_{u\in[0,1],\, x/2 \leqslant |t-u| \leqslant x} (M_n(u) - M_n(t))^2 \right] \leqslant \frac{Cx}{n}. \tag{4.1.2}$$

In Durot, 2007, an additional condition (A2) is considered in order to obtain bounds on p-th moments without requiring (A4) (see Theorem 1 and

Corollary 1 in Durot, 2007). Here, as in Theorem 2 in Durot, 2007, we need (A4) to get a central limit theorem type of result. Hence, condition (A2′) is sufficient.

(A3) $\hat{\lambda}_n(0)$ and $\hat{\lambda}_n(1)$ are stochastically bounded.

(A4) Let B_n be either a Brownian bridge or a Brownian motion. There exists $q > 12$, $C_q > 0$, $L : [0,1] \mapsto \mathbb{R}$ and versions of $M_n = \Lambda_n - \Lambda$ and B_n, such that

$$P\left(n^{1-1/q} \sup_{t\in[0,1]} \left|M_n(t) - n^{-1/2} B_n \circ L(t)\right| > x\right) \leqslant C_q x^{-q}$$

for $x \in (0, n]$. Moreover, L is increasing and twice differentiable on $[0,1]$ with $\sup_t |L''(t)| < \infty$ and $\inf_t L'(t) > 0$.

In Durot, 2007 a variety of statistical models are discussed for which the above assumptions are satisfied, such as estimation of a monotone probability density, a monotone regression function, and a monotone failure rate under right censoring. In Section 4.3, we briefly discuss the consequence of our main result for these models. We restrict ourselves to the case of a non-increasing function λ. The case of non-decreasing λ can be treated similarly. Note that, even if this may not be a natural assumption, for example in the regression setting, we need to assume that λ is positive for the Hellinger distance to be well-defined.

The reason that one can expect a central limit theorem for the Hellinger distance is the fact that the squared Hellinger distance can be approximated by a weighted squared L_2-distance. This can be seen as follows,

$$\begin{aligned}\int_0^1 \left(\sqrt{\hat{\lambda}_n(t)} - \sqrt{\lambda(t)}\right)^2 dt &= \int_0^1 \left(\hat{\lambda}_n(t) - \lambda(t)\right)^2 \left(\sqrt{\hat{\lambda}_n(t)} + \sqrt{\lambda(t)}\right)^{-2} dt \\ &\approx \int_0^1 \left(\hat{\lambda}_n(t) - \lambda(t)\right)^2 (4\lambda(t))^{-1}\, dt.\end{aligned} \tag{4.1.3}$$

Since L_2-distances for Grenander-type estimators obey a central limit theorem (e.g., see Durot, 2007; Kulikov and Lopuhaä, 2005), similar behavior might be expected for the squared Hellinger distance. An application of the delta-method will then do the rest.

The next lemma makes the approximation in (4.1.3) precise.

Lemma 4.1.1. *Assume (A1), (A2′), (A3), and (A4). Moreover, suppose that there exist* $C' > 0$ *and* $s > 3/4$ *with*

$$|\lambda'(t) - \lambda'(x)| \leqslant C'|t - x|^s, \quad \textit{for all } t, x \in [0,1]. \tag{4.1.4}$$

If λ is strictly positive, we have that

$$\int_0^1 \left(\sqrt{\hat{\lambda}_n(t)} - \sqrt{\lambda(t)}\right)^2 dt = \int_0^1 \left(\hat{\lambda}_n(t) - \lambda(t)\right)^2 (4\lambda(t))^{-1}\, dt + o_P(n^{-5/6}).$$

In order to prove Lemma 4.1.1, we need the preparatory lemma below. To this end, we introduce the inverse of $\hat{\lambda}_n$, defined by

$$\hat{U}_n(a) = \operatorname*{argmax}_{u\in[0,1]} \left\{\Lambda_n^+(u) - au\right\}, \quad \text{for all } a \in \mathbb{R}, \tag{4.1.5}$$

where

$$\Lambda_n^+(t) = \max\left\{\Lambda_n(t), \lim_{u\uparrow t} \Lambda_n(u)\right\}.$$

Note that

$$\hat{\lambda}_n(t) \geqslant a \Rightarrow \hat{U}_n(a) \geqslant t. \tag{4.1.6}$$

Furthermore, let g denote the inverse of λ. We then have the following result.

Lemma 4.1.2. *Under the conditions of Lemma 4.1.1, it holds*

$$\int_0^1 |\hat{\lambda}_n(t) - \lambda(t)|^3\, dt = o_P\left(n^{-5/6}\right).$$

Proof. We follow the line of reasoning in the first step of the proof of Theorem 2 in Durot, 2007 with $p = 3$. For completeness we briefly sketch the main steps. We will first show that

$$\int_0^1 |\hat{\lambda}_n(t) - \lambda(t)|^3\, dt = \int_{\lambda(0)}^{\lambda(1)} |\hat{U}_n(b) - g(b)|^3 \lambda'(g(b))^2\, db + o_P(n^{-5/6}).$$

To this end, consider

$$I_1 = \int_0^1 \left(\hat{\lambda}_n(t) - \lambda(t)\right)_+^3 dt, \qquad I_2 = \int_0^1 \left(\lambda(t) - \hat{\lambda}_n(t)\right)_+^3 dt,$$

where $x_+ = \max\{x, 0\}$. We approximate I_1 by

$$J_1 = \int_0^1 \int_0^{(\lambda(0)-\lambda(t))^3} \mathbb{1}_{\{\hat{\lambda}_n(t) \geqslant \lambda(t) + a^{1/3}\}}\, da\, dt.$$

From the reasoning on page 1092 of Durot, 2007, we deduce that

$$0 \leqslant I_1 - J_1 \leqslant \int_0^{n^{-1/3}\log n} \left(\hat{\lambda}_n(t) - \lambda(t)\right)_+^3 dt + |\hat{\lambda}_n(0) - \lambda(1)|^3 \mathbb{1}_{\{n^{1/3}\hat{U}_n(\lambda(0)) > \log n\}}.$$

Since the $\hat{\lambda}_n(0)$ is stochastically bounded and $\lambda(1)$ is bounded, together with Lemma 4 in Durot, 2007, the second term is of the order $o_p(n^{-5/6})$. Furthermore, for the first term we can choose $p' \in [1,2)$ such that the first term on the right hand side is bounded by

$$|\hat{\lambda}_n(0) - \lambda(1)|^{3-p'} \int_0^{n^{-1/3}\log n} |\hat{\lambda}_n(t) - \lambda(t)|^{p'}\, dt.$$

As in Durot, 2007, we get

$$\mathbb{E}\left[\int_0^{n^{-1/3}\log n} |\hat{\lambda}_n(t) - \lambda(t)|^{p'}\, dt\right] \leqslant K n^{-(1+p')/3} \log n = o(n^{-5/6}),$$

by choosing $p' \in (3/2, 2)$. It follows that $I_1 = J_1 + o_P(n^{-5/6})$. By a change of variable $b = \lambda(t) + a^{1/3}$, we find

$$I_1 = \int_{\lambda(1)}^{\lambda(0)} \int_{g(b)}^{\hat{U}_n(b)} 3(b - \lambda(t))^2 \mathbb{1}_{\{g(b) < \hat{U}_n(b)\}}\, dt\, db + o_p(n^{-5/6}).$$

Then, by a Taylor expansion, (A1) and (4.1.4), there exists a $K > 0$, such that

$$\left|(b - \lambda(t))^2 - \left\{ (g(b) - t)\, \lambda'(g(b))\right\}^2\right| \leqslant K\, (t - g(b))^{2+s}\,, \tag{4.1.7}$$

for all $b \in (\lambda(1), \lambda(0))$ and $t \in (g(b), 1]$. We find

$$\begin{aligned} I_1 = &\int_{\lambda(1)}^{\lambda(0)} \int_{g(b)}^{\hat{U}_n(b)} 3(t - g(b))^2 \lambda'(g(b))^2 \mathbb{1}_{\{g(b) < \hat{U}_n(b)\}}\, dt\, db \\ &+ R_n + o_p(n^{-5/6}), \end{aligned} \tag{4.1.8}$$

where

$$\begin{aligned} |R_n| &\leqslant \int_{\lambda(1)}^{\lambda(0)} \int_{g(b)}^{\hat{U}_n(b)} 3K(t - g(b))^{2+s} \mathbb{1}_{\{g(b) < \hat{U}_n(b)\}}\, dt\, db \\ &\leqslant \frac{3K}{3+s} \int_{\lambda(1)}^{\lambda(0)} |\hat{U}_n(b) - g(b)|^{3+s}\, db = O_p(n^{-(3+s)/3}) = o_p(n^{-5/6}), \end{aligned}$$

by using (23) from Durot, 2007, i.e., for every $q' < 3(q-1)$, there exists $K_{q'} > 0$ such that

$$\mathbb{E}\left[\left(n^{1/3}|\hat{U}_n(a) - g(a)|\right)^{q'}\right] \leqslant K_{q'}, \quad \text{for all } a \in \mathbb{R}. \tag{4.1.9}$$

It follows that

$$I_1 = \int_{\lambda(1)}^{\lambda(0)} \left(\hat{U}_n(b) - g(b)\right)^3 \lambda'(g(b))^2 \mathbb{1}_{\{g(b) < \hat{U}_n(b)\}}\, db + o_p(n^{-5/6}).$$

In the same way, one finds

$$I_2 = \int_{\lambda(1)}^{\lambda(0)} \left(g(b) - \hat{U}_n(b)\right)^3 \lambda'(g(b))^2 \mathbb{1}_{\{g(b) > \hat{U}_n(b)\}} \, db + o_p(n^{-5/6}),$$

and it follows that

$$\begin{aligned}\int_0^1 |\hat{\lambda}_n(t) - \lambda(t)|^3 \, dt &= I_1 + I_2 \\ &= \int_{\lambda(1)}^{\lambda(0)} |\hat{U}_n(b) - g(b)|^3 \lambda'(g(b))^2 \, db + o_p(n^{-5/6}).\end{aligned}$$

Now, since λ' is bounded, by Markov's inequality, for each $\epsilon > 0$, we can write

$$\begin{aligned}&\mathbb{P}\left(n^{5/6} \int_{\lambda(0)}^{\lambda(1)} |\hat{U}_n(b) - g(b)|^3 \lambda'(g(b))^2 \, db > \epsilon\right) \\ &\quad \leqslant \frac{1}{c\epsilon n^{1/6}} \int_{\lambda(0)}^{\lambda(1)} \mathbb{E}\left[n|\hat{U}_n(b) - g(b)|^3\right] db \leqslant K n^{-1/6} \to 0.\end{aligned}$$

For the last inequality we again used (4.1.9) with $q' = 3$. It follows that

$$\int_{\lambda(0)}^{\lambda(1)} |\hat{U}_n(b) - g(b)|^3 \lambda'(g(b))^2 \, db = o_p(n^{-5/6}), \tag{4.1.10}$$

which concludes the proof. □

Proof of Lemma 4.1.1. Similar to (4.1.3), we write

$$\int_0^1 \left(\sqrt{\hat{\lambda}_n(t)} - \sqrt{\lambda(t)}\right)^2 dt = \int_0^1 \left(\hat{\lambda}_n(t) - \lambda(t)\right)^2 (4\lambda(t))^{-1} \, dt + R_n,$$

where

$$R_n = \int_0^1 \left(\hat{\lambda}_n(t) - \lambda(t)\right)^2 \left\{\left(\sqrt{\hat{\lambda}_n(t)} + \sqrt{\lambda(t)}\right)^{-2} - (4\lambda(t))^{-1}\right\} dt.$$

Write

$$\begin{aligned}4\lambda(t) - \left(\sqrt{\hat{\lambda}_n(t)} + \sqrt{\lambda(t)}\right)^2 &= \lambda(t) - \hat{\lambda}_n(t) - 2\sqrt{\lambda(t)}\left(\sqrt{\hat{\lambda}_n(t)} - \sqrt{\lambda(t)}\right) \\ &= \left(\lambda(t) - \hat{\lambda}_n(t)\right)\left(1 + \frac{2\sqrt{\lambda(t)}}{\sqrt{\hat{\lambda}_n(t)} + \sqrt{\lambda(t)}}\right).\end{aligned}$$

Since $0 < \lambda(1) \leqslant \lambda(t) \leqslant \lambda(0) < \infty$, this implies that

$$|R_n| \leqslant \int_0^1 \left(\hat{\lambda}_n(t) - \lambda(t)\right)^2 \frac{\left|4\lambda(t) - \left(\sqrt{\hat{\lambda}_n(t)} + \sqrt{\lambda(t)}\right)^2\right|}{4\lambda(t)\left(\sqrt{\hat{\lambda}_n(t)} + \sqrt{\lambda(t)}\right)^2} \, dt$$

$$\leqslant C \int_0^1 \left|\hat{\lambda}_n(t) - \lambda(t)\right|^3 \, dt,$$

for some positive constant C only depending on $\lambda(0)$ and $\lambda(1)$. Then, from Lemma 4.1.2, it follows that $n^{5/6} R_n = o_P(1)$. □

4.2 A CENTRAL LIMIT THEOREM

The following theorem gives a central limit theorem for the squared Hellinger loss. Note that the limit distribution depends on the process X defined in (1.1.12).

Theorem 4.2.1. *Assume (A1), (A2′), (A3), (A4), and* (4.1.4)*. Moreover, suppose that λ is strictly positive. Then, the following holds*

$$n^{1/6} \left\{ n^{2/3} \int_0^1 \left(\sqrt{\hat{\lambda}_n(t)} - \sqrt{\lambda(t)}\right)^2 dt - \mu^2 \right\} \to N(0, \sigma^2),$$

where

$$\mu^2 = \mathbb{E}\left[|X(0)|^2\right] \int_0^1 \frac{|\lambda'(t)\, L'(t)|^{2/3}}{2^{2/3}\lambda(t)} \, dt,$$

and

$$\sigma^2 = 2^{1/3} k_2 \int_0^1 \frac{|\lambda'(t)\, L'(t)|^{2/3} L'(t)}{\lambda(t)^2} \, dt,$$

with k_2 defined in (1.3.2)*.*

Proof. According to Lemma 4.1.1, it is sufficient to show that

$$n^{1/6}\left(n^{2/3} I_n - \mu^2\right) \to N(0, \sigma^2),$$

with

$$I_n = \int_0^1 \left(\hat{\lambda}_n(t) - \lambda(t)\right)^2 (4\lambda(t))^{-1} \, dt.$$

Again, we follow the same line of reasoning as in the proof of Theorem 2 in Durot, 2007. We briefly sketch the main steps of the proof. We first express

I_n in terms of the inverse process $\hat{U}_n$, defined in (4.1.5). To this end, similar to the proof of Lemma 4.1.2, consider

$$\tilde{I}_1 = \int_0^1 \left(\hat{\lambda}_n(t) - \lambda(t)\right)_+^2 (4\lambda(t))^{-1}\, dt, \quad \tilde{I}_2 = \int_0^1 \left(\lambda(t) - \hat{\lambda}_n(t)\right)_+^2 (4\lambda(t))^{-1}\, dt.$$

For the first integral, we can now write

$$\tilde{I}_1 = \int_0^1 \int_0^\infty \mathbb{1}_{\left\{\hat{\lambda}_n(t) \geqslant \lambda(t) + \sqrt{4a\lambda(t)}\right\}}\, da\, dt.$$

Then, if we introduce

$$\tilde{J}_1 = \int_0^1 \int_0^{(\lambda(0)-\lambda(t))^2/4\lambda(t)} \mathbb{1}_{\left\{\hat{\lambda}_n(t) \geqslant \lambda(t) + \sqrt{4a\lambda(t)}\right\}}\, da\, dt, \tag{4.2.1}$$

we obtain

$$\begin{aligned} 0 \leqslant \tilde{I}_1 - \tilde{J}_1 &\leqslant \int_0^{\hat{U}_n(\lambda(0))} \int_{(\lambda(0)-\lambda(t))^2/4a\lambda(t)}^\infty \mathbb{1}_{\left\{\hat{\lambda}_n(t) \geqslant \lambda(t) + \sqrt{4a\lambda(t)}\right\}}\, da\, dt \\ &\leqslant \frac{1}{4\lambda(1)} \int_0^{\hat{U}_n(\lambda(0))} \left(\hat{\lambda}_n(t) - \lambda(t)\right)_+^2\, dt. \end{aligned}$$

Similar to the reasoning in the proof of Lemma 4.1.2, we conclude that

$$\tilde{I}_1 = \tilde{J}_1 + o_p(n^{-5/6}).$$

Next, the change of variable $b = \lambda(t) + \sqrt{4a\lambda(t)}$ yields

$$\begin{aligned} \tilde{J}_1 &= \int_{\lambda(1)}^{\lambda(0)} \int_{g(b)}^{\hat{U}_n(b)} \frac{b - \lambda(t)}{2\lambda(t)} \mathbb{1}_{\{\hat{U}_n(b) > g(b)\}}\, dt\, db \\ &= \int_{\lambda(1)}^{\lambda(0)} \int_{g(b)}^{\hat{U}_n(b)} \frac{b - \lambda(t)}{2b} \mathbb{1}_{\{\hat{U}_n(b) > g(b)\}}\, dt\, db \\ &\quad + \int_{\lambda(1)}^{\lambda(0)} \int_{g(b)}^{\hat{U}_n(b)} \frac{(b - \lambda(t))^2}{2b\lambda(t)} \mathbb{1}_{\{\hat{U}_n(b) > g(b)\}}\, dt\, db. \end{aligned} \tag{4.2.2}$$

Let us first consider the second integral on the right hand side of (4.2.2). We then have

$$\begin{aligned}
&\int_{\lambda(1)}^{\lambda(0)}\int_{g(b)}^{\hat{U}_n(b)} \frac{(b-\lambda(t))^2}{2b\lambda(t)} \mathbb{1}_{\{\hat{U}_n(b)>g(b)\}}\,dt\,db \\
&\quad\leqslant \frac{1}{2\lambda(1)^2}\int_{\lambda(1)}^{\lambda(0)}\int_{g(b)}^{\hat{U}_n(b)} (b-\lambda(t))^2\,\mathbb{1}_{\{\hat{U}_n(b)>g(b)\}}\,dt\,db \\
&\quad\leqslant \frac{1}{2\lambda(1)^2}\sup_{t\in[0,1]}|\lambda'(t)|\int_{\lambda(1)}^{\lambda(0)} \mathbb{1}_{\{\hat{U}_n(b)>g(b)\}}\int_{g(b)}^{\hat{U}_n(b)} (t-g(b))^2\,dt\,db \\
&\quad= \frac{1}{6\lambda(1)^2}\sup_{t\in[0,1]}|\lambda'(t)|\int_{\lambda(1)}^{\lambda(0)} \mathbb{1}_{\{\hat{U}_n(b)>g(b)\}}\left(\hat{U}_n(b)-g(b)\right)^3\,db.
\end{aligned}$$

Again by using (4.1.9) with $q'=3$ we obtain

$$\int_{\lambda(1)}^{\lambda(0)}\int_{g(b)}^{\hat{U}_n(b)} \frac{(b-\lambda(t))^2}{2b\lambda(t)} \mathbb{1}_{\{\hat{U}_n(b)>g(b)\}}\,dt\,db = o_P(n^{-5/6}).$$

Then consider the first integral on the right hand side of (4.2.2). Similar to (4.1.7), there exists $K>0$ such that, for all $b\in(\lambda(1),\lambda(0))$ and $t\in(g(b),1]$,

$$\left|(b-\lambda(t)-(g(b)-t)\lambda'(g(b)))\right| \leqslant K(t-g(b))^{1+s}.$$

Taking into account that $\lambda'(g(b))<0$, similar to (4.1.8), it follows that

$$\begin{aligned}
\tilde{I}_1 = &\int_{\lambda(1)}^{\lambda(0)}\int_{g(b)}^{\hat{U}_n(b)} \frac{|\lambda'(g(b))|}{2b}(t-g(b))\,\mathbb{1}_{\{\hat{U}_n(b)>g(b)\}}\,dt\,db \\
&\qquad + \tilde{R}_n + o_p(n^{-5/6}),
\end{aligned}$$

where

$$\begin{aligned}
|\tilde{R}_n| &\leqslant \int_{\lambda(1)}^{\lambda(0)}\int_{g(b)}^{\hat{U}_n(b)} \frac{K}{2\lambda(1)}(t-g(b))^{1+s}\mathbb{1}_{\{g(b)<\hat{U}_n(b)\}}\,dt\,db \\
&\leqslant \frac{K}{2\lambda(1)(2+s)}\int_{\lambda(1)}^{\lambda(0)}|\hat{U}_n(b)-g(b)|^{2+s}\,db \\
&= O_p(n^{-(2+s)/3}) = o_p(n^{-5/6}),
\end{aligned}$$

by using (4.1.9) once more, and the fact that $s>3/4$. It follows that

$$\tilde{I}_1 = \int_{\lambda(1)}^{\lambda(0)} \frac{|\lambda'(g(b))|}{4b}\left(\hat{U}_n(b)-g(b)\right)^2\mathbb{1}_{\{\hat{U}_n(b)>g(b)\}}\,db + o_p(n^{-5/6}).$$

In the same way

$$\tilde{I}_2 = \int_{\lambda(1)}^{\lambda(0)} \frac{|\lambda'(g(b))|}{4b}\left(\hat{U}_n(b)-g(b)\right)^2\mathbb{1}_{\{\hat{U}_n(b)<g(b)\}}\,db + o_p(n^{-5/6}),$$

so that

$$I_n = \tilde{I}_1 + \tilde{I}_2 = \int_{\lambda(1)}^{\lambda(0)} \left(\hat{U}_n(b) - g(b)\right)^2 \frac{|\lambda'(g(b))|}{4b}\, db + o_P(n^{-5/6}).$$

We then mimic step 2 in the proof of Theorem 2 in Durot, 2007. Let B_n be the process in condition (A4). Consider the representation

$$B_n(t) = W_n(t) - \xi_n t,$$

where W_n is a standard Brownian motion, $\xi_n = 0$ if B_n is Brownian motion, and ξ_n is a standard normal random variable independent of B_n, if B_n is a Brownian bridge. Then, define

$$\mathbb{W}_t(u) = n^{1/6}\left\{W_n(L(t) + n^{-1/3}u) - W_n(L(t))\right\}, \quad \text{for } t \in [0,1],$$

which has the same distribution as a standard Brownian motion. Now, for $t \in [0,1]$, let $d(t) = |\lambda'(t)|/(2L'(t)^2)$ and define

$$\tilde{V}(t) = \underset{|u| \leqslant \log n}{\operatorname{argmax}} \left\{\mathbb{W}_t(u) - d(t)u^2\right\}. \tag{4.2.3}$$

Then similar to (26) in Durot, 2007, we will obtain

$$n^{2/3} I_n = \int_0^1 \left|\tilde{V}(t) - n^{-1/6}\frac{\xi_n}{2d(t)}\right|^2 \left|\frac{\lambda'(t)}{L'(t)}\right|^2 \frac{1}{4\lambda(t)}\, dt + o_P(n^{-1/6}). \tag{4.2.4}$$

To prove (4.2.4), by using the approximation

$$\hat{U}_n(a) - g(a) \approx \frac{L(\hat{U}_n(a)) - L(g(a))}{L'(g(a))}$$

and a change of variable $a^\xi = a - n^{1/2}\xi_n L'(g(a))$, we first obtain

$$\begin{aligned} n^{2/3} I_n = n^{2/3} &\int_{\lambda(1)+\delta_n}^{\lambda(0)-\delta_n} \left|L(\hat{U}_n(a^\xi)) - L(g(a^\xi))\right|^2 \frac{|\lambda'(g(a))|}{(L'(g(a)))^2}\frac{1}{4a}\, da \\ &+ o_p(n^{-1/6}), \end{aligned}$$

where $\delta_n = n^{-1/6}/\log n$. Apart from the factor $1/4a$, the integral on the right hand side is the same as in the proof of Theorem 2 in Durot, 2007 for $p = 2$. This means that we can apply the same series of succeeding approximations for $L(\hat{U}_n(a^\xi)) - L(g(a^\xi))$ as in Durot, 2007, which yields

$$\begin{aligned} n^{2/3} I_n = n^{2/3} &\int_{\lambda(1)+\delta_n}^{\lambda(0)-\delta_n} \left|\tilde{V}(g(a)) - n^{-1/6}\frac{\xi_n}{2d(g(a))}\right|^2 \frac{|\lambda'(g(a))|}{(L'(g(a)))^2}\frac{1}{4a}\, da \\ &+ o_p(n^{-1/6}). \end{aligned}$$

Finally, because the integrals over $[\lambda(1), \lambda(1) + \delta_n]$ and $[\lambda(0) - \delta_n, \lambda(0)]$ are of the order $o_P(n^{-1/6})$, this yields (4.2.4) by a change of variables $t = g(a)$.

The next step is to show that the term with ξ_n can be removed from (4.2.4). This can be done exactly as in Durot, 2007, since the only difference with the corresponding integral in Durot, 2007 is the factor $1/4\lambda(t)$, which is bounded and does not influence the argument in Durot, 2007. We find that

$$n^{2/3} I_n = \int_0^1 |\tilde{V}(t)|^2 \left| \frac{\lambda'(t)}{L'(t)} \right|^2 \frac{1}{4\lambda(t)} \, dt + o_P(n^{-1/6}).$$

Then define

$$Y_n(t) = \left(|\tilde{V}(t)|^2 - \mathbb{E}\left[|\tilde{V}(t)|^2 \right] \right) \left| \frac{\lambda'(t)}{L'(t)} \right|^2 \frac{1}{4\lambda(t)}. \tag{4.2.5}$$

By approximating $\tilde{V}(t)$ by

$$V(t) = \operatorname*{argmax}_{u \in \mathbb{R}} \left\{ W_t(u) - d(t) u^2 \right\},$$

and using that, by Brownian scaling, $d(t)^{2/3} V(t)$ has the same distribution as $X(0)$, see Durot, 2007 for details, we have that

$$\begin{aligned}
&\int_0^1 \mathbb{E}\left[|\tilde{V}(t)|^2 \right] \left| \frac{\lambda'(t)}{L'(t)} \right|^2 \frac{1}{4\lambda(t)} \, dt \\
&= \mathbb{E}\left[|X(0)|^2 \right] \int_0^1 d(t)^{-4/3} \left| \frac{\lambda'(t)}{L'(t)} \right|^2 \frac{1}{4\lambda(t)} \, dt + o(n^{-1/6}) \\
&= \mu^2 + o(n^{-1/6}).
\end{aligned}$$

It follows that

$$n^{1/6}(I_n - \mu^2) = n^{1/6} \int_0^1 Y_n(t) \, dt + o_P(1).$$

We then first show that

$$\operatorname{Var}\left(n^{1/6} \int_0^1 Y_n(t) \, dt \right) \to \sigma^2. \tag{4.2.6}$$

Once more, following the proof in Durot, 2007 we have

$$\begin{aligned}
\nu_n &= \operatorname{Var}\left(\int_0^1 Y_n(t) \, dt \right) \\
&= 2 \int_0^1 \int_s^1 \left| \frac{\lambda'(t)}{L'(t)} \frac{\lambda'(s)}{L'(s)} \right|^2 \frac{1}{4\lambda(t)} \frac{1}{4\lambda(s)} \operatorname{cov}(|\tilde{V}(t)|^2, |\tilde{V}(s)|^2) \, dt \, ds.
\end{aligned}$$

After the same sort of approximations as in Durot, 2007, we get

$$v_n = 2\int_0^1 \int_s^{\min(1,s+c_n)} \left|\frac{\lambda'(s)}{L'(s)}\right|^4 \frac{\operatorname{cov}(|V_t(s)|^2, |V_s(s)|^2)}{(4\lambda(s))^2}\,dt\,ds + o(n^{-1/3}),$$

where $c_n = 2n^{-1/3}\log n/\inf_t L'(t)$ and where, for all s and t,

$$V_t(s) = \operatorname*{argmax}_{u\in\mathbb{R}} \left\{\mathbb{W}_t(u) - d(s)u^2\right\}.$$

Then use that $d(s)^{2/3}V_t(s)$ has the same distribution as

$$X\big(n^{1/3}d(s)^{2/3}\big(L(t)-L(s)\big)\big) - n^{1/3}d(s)\big(L(t)-L(s)\big),$$

so that the change of variable $a = n^{1/3}d(s)^{2/3}(L(t)-L(s))$ in v_n leads to

$$\begin{aligned} n^{1/3}v_n &\to 2\int_0^1\int_0^\infty \left|\frac{\lambda'(s)}{L'(s)}\right|^4 \frac{1}{(4\lambda(s))^2}\frac{1}{d(s)^{10/3}L'(s)}\operatorname{cov}(|X(a)|^2, |X(0)|^2)\,da\,ds \\ &\to 2k_2\int_0^1 \left|\frac{\lambda'(s)}{L'(s)}\right|^4 \frac{1}{(4\lambda(s))^2}\frac{2^{10/3}|L'(s)|^{17/3}}{|\lambda'(s)|^{10/3}}\,ds = \sigma^2, \end{aligned}$$

which proves (4.2.6).

Finally, asymptotic normality of $n^{1/6}\int_0^1 Y_n(t)\,dt$ follows by Bernstein's method of big blocks and small blocks in the same way as in step 6 of the proof of Theorem 2 in Durot, 2007. □

Corollary 4.2.2. *Assume (A1), (A2′), (A3), (A4), and* (4.1.4) *and let* $H(\hat\lambda_n, \lambda)$ *be the Hellinger distance defined in* (4.1.1)*. Moreover, suppose that* λ *is strictly positive. Then,*

$$n^{1/6}\left\{n^{1/3}H(\hat\lambda_n,\lambda) - \tilde\mu\right\} \to N(0, \tilde\sigma^2),$$

where $\tilde\mu = 2^{-1/2}\mu$ *and* $\tilde\sigma^2 = \sigma^2/8\mu^2$*, where* μ^2 *and* σ^2 *are defined in Theorem 4.2.1.*

Proof. This follows immediately by applying the delta method with $\phi(x) = 2^{-1/2}\sqrt{x}$ to the result in Theorem 4.2.1. □

4.3 EXAMPLES

The type of scaling for the Hellinger distance in Corollary 4.2.2 is similar to that in the central limit theorem for L_p-distances. This could be expected in view of the approximation in terms of a weighted squared L_2-distance, see

Lemma 4.1.1, and the results, e.g., in Kulikov and Lopuhaä, 2005 and Durot, 2007. Actually, this is not always the case. The phenomenon of observing different speeds of convergence for the Hellinger distance from those for the L_1 and L_2 norms is considered in Birgé, 1986. In fact, this is related to the existence of a lower bound for the function we are estimating. If the function of interest is bounded from below, which is the case considered in this paper, then the approximation (4.1.3) holds, see Birgé, 1986 for an explanation.

When we insert the expressions for μ^2 and σ^2 from Theorem 4.2.1, then we get

$$\tilde{\sigma}^2 = \frac{k_2}{4\mathbb{E}\left[|X(0)|^2\right]} \frac{\int_0^1 |\lambda'(t)L'(t)|^{2/3} L'(t)\lambda(t)^{-2}\,dt}{\int_0^1 |\lambda'(t)L'(t)|^{2/3}\lambda(t)^{-1}\,dt},$$

where k_2 is defined in (1.3.2). This means that in statistical models where $L = \Lambda$ in condition (A4), and hence $L' = \lambda$, the limiting variance $\tilde{\sigma}^2 = k_2/(4\mathbb{E}[|X(0)|^2])$ does not depend on λ.

One such a model is estimation of the common monotone density λ on $[0,1]$ of independent random variables $X_1, \ldots, X_n$. Then, Λ_n is the empirical distribution function of $X_1, \ldots, X_n$ and $\hat{\lambda}_n$ is Grenander's estimator (Grenander, 1956). In that case, if $\inf_t \lambda(t) > 0$, the conditions of Corollary 4.2.2 are satisfied with $L = \Lambda$ (see Theorem 6 in Durot, 2007), so that the limiting variance of the Hellinger loss for the Grenander estimator does not depend on the underlying density. This behavior was conjectured in Wellner, 2015 and coincides with that of the limiting variance in the central limit theorem for the L_1-error for the Grenander estimator, first discovered by Groeneboom, 1983 (see also Durot, 2002; Groeneboom, Hooghiemstra, and Lopuhaä, 1999 and Durot, 2007; Kulikov and Lopuhaä, 2005).

Another example is when we observe independent identically distributed inhomogeneous Poisson processes $N_1, \ldots, N_n$ with common mean function Λ on $[0,1]$ with derivative λ, for which $\Lambda(1) < \infty$. Then Λ_n is the restriction of $(N_1 + \cdots + N_n)/n$ to $[0,1]$. Also in that case, the conditions of Corollary 4.2.2 are satisfied with $L = \Lambda$ (see Theorem 4 in Durot, 2007), so that the limiting variance of the Hellinger loss for $\hat{\lambda}_n$ does not depend on the common underlying intensity λ. However, note that for this model, the L_1-loss for $\hat{\lambda}_n$ is asymptotically normal according to Theorem 2 in Durot, 2007, but with limiting variance depending on the value $\Lambda(1) - \Lambda(0)$.

Consider the monotone regression model $y_{i,n} = \lambda(i/n) + \epsilon_{i,n}$, for $i = 1, \ldots, n$, where the $\epsilon_{i,n}$'s are i.i.d. random variables with mean zero and variance $\sigma^2 > 0$. Let

$$\Lambda_n(t) = \frac{1}{n} \sum_{i \leqslant nt} y_{i,n}, \quad t \in [0, 1],$$

be the empirical distribution function. Then $\hat{\lambda}_n$ is (a slight modification of) Brunk's estimator from Brunk, 1958. Under appropriate moment conditions on the $\epsilon_{i,n}$, the conditions of Corollary 4.2.2 are satisfied with $L(t) = t\sigma^2$ (see Theorem 5 in Durot, 2007). In this case, the limiting variance of the Hellinger loss for $\hat{\lambda}_n$ depends on both λ and σ^2, whereas the L_1-loss for $\hat{\lambda}_n$ is asymptotically normal according to Theorem 2 in Durot, 2007, but with limiting variance only depending on σ^2.

Suppose we observe a right-censored sample $(X_1, \Delta_1), \ldots, (X_n, \Delta_n)$, where $X_i = \min(T_i, Y_i)$ and $\Delta_i = \mathbb{1}_{\{T_i \leqslant Y_i\}}$, with the T_i's being nonnegative i.i.d. failure times and the Y_i's are i.i.d. censoring times independent of the T_i's. Let F be the distribution function of the T_i's with density f and let G be the distribution function of the Y_i's. The parameter of interest is the monotone failure rat $\lambda = f/(1 - F)$ on $[0, 1]$. In this case, Λ_n is the restriction of he Nelson-Aalen estimator to $[0, 1]$. If we assume (A1) and $\inf_t \lambda(t) > 0$, then under suitable assumptions on F and G the conditions of Corollary 4.2.2 hold with

$$L(t) = \int_0^t \frac{\lambda(u)}{(1 - F(u)))(1 - G(u))} \, du, \quad t \in [0, 1],$$

(see Theorem 3 in Durot, 2007). This means that the limiting variance of the Hellinger loss depends on λ, F and G, whereas the limiting variance of the L_1-loss depends only on their values at 0 and 1. In particular, in the case of nonrandom censoring times, $L = (1 - F)^{-1} - 1$, the limiting variance of the Hellinger loss depends on λ and F, whereas the limiting variance of the L_1-loss depends only on the value $F(1)$.

4.4 TESTING EXPONENTIALITY AGAINST A NON-DECREASING DENSITY

In this section we investigate a possible application of Theorem 4.2.1, i.e., testing for an exponential density against a non-increasing alternative by means of the Hellinger loss. The exponential distribution is one of the most used and well-known distributions. It plays a very important role in reliability, survival analysis, and in renewal process theory, when modeling random times until some event. As a result, a lot of attention has been given in the literature to testing for exponentiality against a wide variety of alternatives, by making use of different properties and characterizations of

the exponential distribution (see Meintanis, 2007, Alizadeh Noughabi and Arghami, 2011, Jammalamadaka and Taufer, 2003, Haywood and Khmaladze, 2008). In this section we consider a test for exponentiality, assuming that data comes from a decreasing density. The test is based on the Hellinger distance between the parametric estimator of the exponential density and the Grenander-type estimator of a general decreasing density. In order to be able to apply the result of Corollary 4.2.2, we first investigate a test whether the data is exponentially distributed with a fixed parameter $\lambda_0 > 0$. Since such a test may not be very interesting from a practical point of view, we also investigate testing exponentiality leaving the parameter $\lambda > 0$ unspecified.

4.4.1 *Testing a simple null hypothesis of exponentiality*

Let $f_\lambda(x) = \lambda e^{-\lambda x} \mathbb{1}_{\{x \geqslant 0\}}$ be the exponential density with parameter $\lambda > 0$. Assume we have a sample of i.i.d. observations $X_1, \ldots, X_n$ from some distribution with density f and for $\lambda_0 > 0$ fixed, we want to test

$$H_0 : f = f_{\lambda_0} \quad \text{against} \quad H_1 : f \text{ is non-increasing.}$$

Under the alternative hypothesis we can estimate f on an interval $[0, \tau]$ by the Grenander-type estimator $\hat{f}_n$ from Section 4.1. Then as a test statistic we take $T_n = H(\hat{f}_n, f_{\lambda_0})$, the Hellinger distance on $[0, \tau]$ between $\hat{f}_n$ and f_{λ_0}, and at level α, we reject the null hypothesis if $T_n > c_{n,\alpha,\lambda_0}$, for some critical value $c_{n,\alpha,\lambda_0} > 0$.

According to Corollary 4.2.2, it follows that T_n is asymptotically normally distributed, but the mean and the variance depend on the constant k_2 defined in (1.3.2). To avoid computation of k_2, we estimate the mean and the variance of T_n empirically. We generate $B = 10\,000$ samples from f_{λ_0}. For each of these samples we compute the Grenander estimator $\hat{f}_{n,i}$ and the Hellinger distance $T_{n,i} = H(\hat{f}_{n,i}, f_{\lambda_0})$, for $i = 1, 2, \ldots, B$. Finally, we compute the mean $\bar{T}$ and the variance s_T of the values $T_{n,1}, \ldots, T_{n,B}$. For the critical value of the test we take $c_{n,\alpha,\lambda_0} = \bar{T} + q_{1-\alpha} s_T$, where $q_{1-\alpha}$ is the $100(1-\alpha)\%$ quantile of the standard normal distribution. Note that, even if in the density model the asymptotic variance is independent of the underlying distribution, the asymptotic mean does depend on λ_0, i.e., the test is not distribution free. Another possibility, instead of the normal approximation, is to take as a critical value $\tilde{c}_{n,\alpha,\lambda_0}$ the empirical $100(1-\alpha)\%$ quantile of the values $T_{n,1}, \ldots, T_{n,B}$.

To investigate the performance of the test, we generate $N = 10\,000$ samples from f_{λ_0}. For each sample we compute the value of the test statistic

$T_n = H(\hat{f}_n, f_{\lambda_0})$ and we reject the null hypothesis if $T_n > c_{n,\alpha,\lambda_0}$ (or if $T_n > \tilde{c}_{n,\alpha,\lambda_0}$). The percentage of rejections gives an approximation of the level of the test. Table 10 shows the results of the simulations for different sample sizes n and two values of λ_0 and $\alpha = 0.01, 0.05, 0.10$. Here we take $\tau = 5$, since the mass of the exponential distribution with parameter one or five outside the interval $[0, 5]$ is negligible. We observe that the percentage of rejections is close to the nominal level if we use $\tilde{c}_{n,\alpha,\lambda_0}$ as critical value for the test, but it is a bit higher if we use c_{n,α,λ_0}. This is due to the fact that for small sample sizes, the normal approximation of Corollary 4.2.2 is not very precise. Moreover, to investigate the power, we generate a sample

	$\lambda_0 = 1$			$\lambda_0 = 5$		
	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.10$	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.10$
$n = 20$	0.0229	0.0680	0.1016	0.0310	0.0791	0.1127
	0.0118	0.0498	0.0971	0.0117	0.0533	0.1058
$n = 50$	0.0244	0.0684	0.1123	0.0243	0.0659	0.1086
	0.0106	0.0469	0.0923	0.0103	0.0494	0.0964
$n = 100$	0.0190	0.0589	0.1021	0.0236	0.0673	0.1126
	0.0106	0.0531	0.1063	0.0091	0.0453	0.0951

Table 10: Simulated levels of T_n using c_{n,α,λ_0} (top) and $\tilde{c}_{n,\alpha,\lambda_0}$ (bottom), with $\alpha = 0.01, 0.05, 0.10$, under the null hypothesis varying the sample size n and the parameter λ_0.

from the Weibull distribution with shape parameter ν and scale parameter λ_0^{-1}. Recall that $\text{Weibull}(1, \lambda_0^{-1})$ corresponds to the exponential distribution with parameter λ_0 and that a Weibull distribution with $\nu < 1$ has a decreasing density. We compute the Hellinger distance $T_n = H(\hat{f}_n, f_{\lambda_0})$ and we reject the null hypothesis if $T_n > c_{n,\alpha,\lambda_0}$ (or if $T_n > \tilde{c}_{n,\alpha,\lambda_0}$). After repeating the procedure $N = 10\,000$ times, we compute the percentage of times that we reject the null hypothesis, which gives an approximation of the power of the test.

The results of the simulations, done with $n = 100$, $\lambda_0 = 1$, $\alpha = 0.05$ and alternatives for which ν varies between 0.4 and 1 by steps of 0.05, are shown in Figure 12. As a benchmark, we compute the power of the likelihood ratio test statistic for each ν. As expected, our test is less powerful with respect to the LR test, which is designed to test against a particular al-

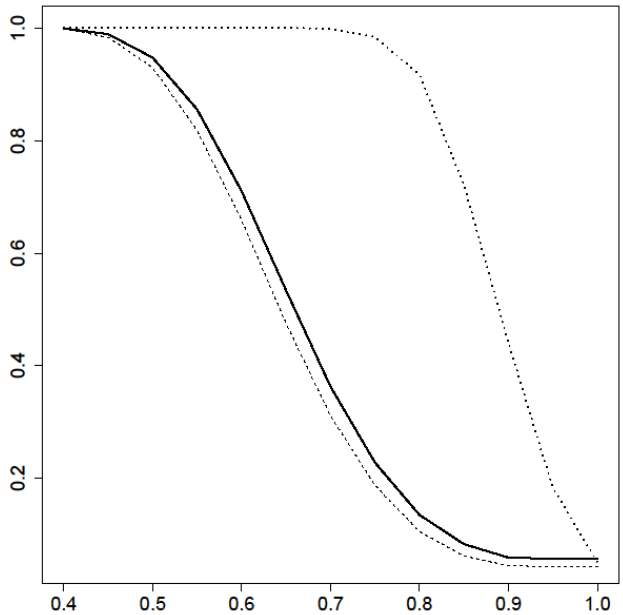


Figure 12: Simulated powers using c_{n,α,λ_0} (solid) and $\tilde{c}_{n,\alpha,\lambda_0}$ (dashed), with $\alpha = 0.05$, of T_n and the power of the LR test (dotted) for $\lambda = 1$, $\nu = 0.4, 0.45, \ldots, 1$ and $n = 100$.

ternative. However, as the sample size increases, the performance improves significantly and the difference of the results when using c_{n,α,λ_0} or $\tilde{c}_{n,\alpha,\lambda_0}$ becomes smaller.

4.4.2 *Testing a composite null hypothesis of exponentiality*

Assume we have a sample of i.i.d.observations $X_1, \ldots, X_n$ from some distribution with density f and we want to test

$$H_0 : f = f_\lambda, \text{ for some } \lambda > 0, \quad \text{against} \quad H_1 : f \text{ is non-increasing.}$$

Under the null hypothesis, we can construct a parametric estimator of the density which is given by $f_{\hat{\lambda}_n}$, where $\hat{\lambda}_n = n/\sum_{i=1}^n X_i$ is the MLE of λ. On the other hand, under the alternative hypothesis we can estimate f on an interval $[0, \tau]$ by the Grenander-type estimator $\hat{f}_n$ from Section 4.1. Then as a test statistic we take $R_n = H(\hat{f}_n, f_{\hat{\lambda}_n})$, the Hellinger distance on $[0, \tau]$ between the two estimators, and at level α, we reject the null hypothesis if $R_n > d_{n,\alpha}$ for some critical value $d_{n,\alpha} > 0$. Because the limit distribution

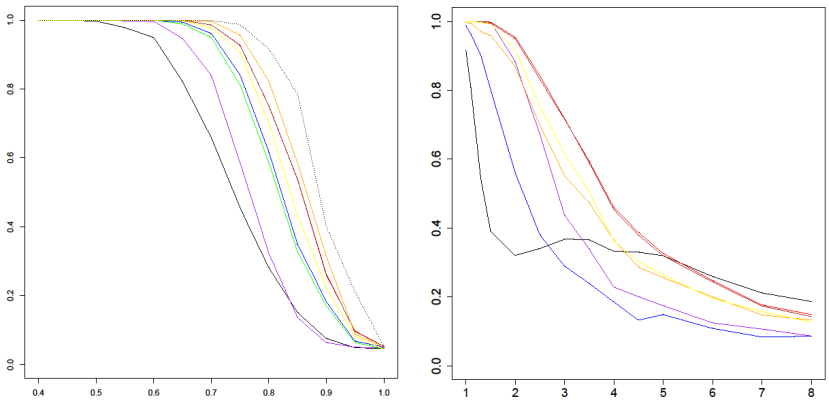
of the test statistic is not known we use a bootstrap procedure to calibrate the test. We generate $B = 1000$ bootstrap samples of size n from $f_{\hat{\lambda}_n}$ and for each of them we compute the estimators $f_{\hat{\lambda}^*_{n,i}}$, $\hat{f}^*_{n,i}$ and the test statistic $R^*_{n,i} = H(f_{\hat{\lambda}^*_{n,i}}, \hat{f}^*_{n,i})$, for $i = 1, 2, \ldots, B$. Then we determine the 100α-th upper-percentile $d^*_{n,\alpha}$ of the values $R^*_{n,1}, \ldots, R^*_{n,B}$. Finally we reject the null hypothesis if $R_n > d^*_{n,\alpha}$.

To investigate the level of the test, for $\alpha = 0.05$ and $\lambda > 0$ fixed, we start with a sample from an exponential distribution with parameter λ and repeat the above procedure $N = 10\,000$ times. We count the number of times we reject the null hypothesis, i.e., the number of times the value of the test statistics exceeds the corresponding 5th upper-percentile. Dividing this number by N gives an approximation of the level. Table 11 shows the results of the simulations for different sample sizes n and different values of λ. The rejection probabilities are close to 0.05 for all the values of λ, which shows that the test performs well in the different scenarios (slightly and strongly decreasing densities).

λ	0.05	0.1	0.5	1	2	3	4	5
$n = 50$	0.051	0.052	0.049	0.049	0.05	0.053	0.051	0.054
$n = 100$	0.049	0.047	0.050	0.052	0.054	0.047	0.049	0.050
$n = 500$	0.052	0.049	0.049	0.049	0.053	0.052	0.053	0.048
$n = 1000$	0.053	0.046	0.049	0.051	0.049	0.048	0.048	0.052

Table 11: Simulated levels of R_n under the null hypothesis varying the sample size n and the parameter λ.

To investigate the power, for $\alpha = 0.05$ and fixed $0 < \nu < 1$ and $\lambda > 0$, we now start with a sample from a Weibull distribution with shape parameter ν and scale parameter λ^{-1} and compute the value $R_n = H(f_{\hat{\lambda}_n}, \hat{f}_n)$. In order to calibrate the test, we treat this sample as if it were an exponential sample, and estimate λ by $\hat{\lambda}_n = n/\sum_{i=1}^n X_i$. Next, we generate $B = 1000$ bootstrap samples of size n from the exponential density with parameter $\hat{\lambda}_n$. For each bootstrap sample we compute the test statistic $R^*_{n,i} = H(f_{\hat{\lambda}^*_{n,i}}, \hat{f}^*_{n,i})$, for $i = 1, 2, \ldots, B$, and we determine the 5th upper-percentile $d^*_{n,0.05}$ of the values $R^*_{n,1}, \ldots, R^*_{n,B}$. Finally, we reject the null hypothesis if $R_n > d^*_{n,0.05}$. After repeating the above procedure $N = 10\,000$ times, each time starting



(a) Weibull (b) Beta

Figure 13: Simulated powers of the Hellinger distance test (black solid) and some other competitor tests: T_1 (blue), T_2 (green), ω_n^2 (yellow), S_n (brown), EP_n (red), KL_{mn} (purple), CO_n (orange), and the power of the LR test (black dotted) for $n = 100$, $\lambda = 1$, $0.4 \leqslant \nu \leqslant 1$ (left) and $1 \leqslant \beta \leqslant 8$ (right).

with a Weibull sample, we compute the percentage of times that we reject the null hypothesis, which gives an approximation of the power of the test.

We compare the Hellinger distance test to some of the tests from Alizadeh Noughabi and Arghami, 2011, which are designed to test exponentiality against all the possible alternatives, i.e., not only against decreasing densities. These test are all distribution free, which means that their critical values can be computed independently of λ. Then, for each of the Weibull samples generated before, we count the percentage of times that the tests T_1, T_2, ω_n^2, S_n, EP_n, KL_{mn}, CO_n (see Alizadeh Noughabi and Arghami, 2011 for a precise definition) reject the null hypothesis. Finally, we also compare the power of our test with the likelihood ratio test for each ν.

The results of the simulations, done with $n = 100$, $\lambda = 1$, and alternatives for which ν varies between 0.4 and 1, are shown in the left panel in Figure 13. Actually, we also investigated the power for different choices of λ and we observed similar behavior as for $\lambda = 1$. The figure shows that the test based on the Hellinger distance performs worse that the other tests. In this case, the test of Cox and Oakes CO_n has greater power. However, Alizadeh Noughabi and Arghami, 2011 concluded that none of the tests is uniformly most powerful with respect to the others.

We repeated the experiment taking, instead of the Weibull distribution, the beta distribution with parameters $\alpha = 1$ and $1 \leqslant \beta \leqslant 8$ as alternative.

Note that it has a non-increasing density on $[0, 1]$ proportional to $(1-x)^{\beta-1}$ and the extreme case $\beta = 1$ corresponds to the uniform distribution. Results are shown in the right panel in Figure 13. We observe that for small values of β the Hellinger distance test again behaves worse than the others and in this case R_n and EP_n have greater power. However, for larger β the Hellinger distance test outperforms all the others.

5

THE DISTANCE BETWEEN A NAIVE CUMULATIVE ESTIMATOR AND ITS LEAST CONCAVE MAJORANT

In this chapter we consider the process $\widehat{\Lambda}_n - \Lambda_n$, where Λ_n is a cadlag step estimator for the primitive Λ of a nonincreasing function λ on $[0, 1]$, and $\widehat{\Lambda}_n$ is the least concave majorant of Λ_n. The results presented are based on:

Lopuhaä, H. P. and Musta E. (2018) "The distance between a naive cumulative estimator and its least concave majorant". *Statistics and Probability Letters 139*, pp. 119-128.

A large part of the literature is devoted to investigating properties of Grenander-type estimators for monotone curves, and somewhat less attention is paid to properties of the difference between the corresponding naive estimator for the primitive of the curve and its LCM. Kiefer and Wolfowitz, 1976 show that $\sup_t |\widehat{F}_n - F_n| = O_p((n^{-1} \log n)^{2/3})$. Although the first motivation for this type of result has been asymptotic optimality of shape constrained estimators, it has several important statistical applications. The Kiefer-Wolfowitz result was a key argument in Sen, Banerjee, and Woodroofe, 2010 to prove that the m out of n bootstrap from $\widehat{F}_n$ works. Mammen, 1991 suggested to use the result to make an asymptotic comparison between a smoothed Grenander-type estimator and an isotonized kernel estimator in the regression context. See also Wang and Woodroofe, 2007 for a similar application of their Kiefer-Wolfowitz comparison theorem. An extension to a more general setting was established in Durot and Lopuhaä, 2014, which has direct applications in Durot, Groeneboom, and Lopuhaä, 2013 to prove that a smoothed bootstrap from a Grenander-type estimator works for k-sample tests, and in Groeneboom and Jongbloed, 2013 and in Chapter 2 of this thesis to extract the pointwise limit behavior of smoothed Grenander-type estimators for a monotone hazard from that of ordinary kernel estimators. However, to approximate the L_p-error of smoothed Grenander-type estimators by that of ordinary kernel estimators, such as in Csörgő and Horváth, 1988 for kernel density estimators, a Kiefer-Wolfowitz type result no longer suffices. In that case, results on the L_p-distance, between $\widehat{F}_n$ and F_n are more appropriate, such as the ones in Durot and Tocquet, 2003 and Kulikov and Lopuhaä, 2008.

We extend the results in Kulikov and Lopuhaä, 2006, 2008 to the general setting considered in Durot, 2007. Under this setting we prove that a suitably scaled version of $\widehat{\Lambda}_n - \Lambda_n$ converges in distribution to the corresponding process for two-sided Brownian motion with parabolic drift and we establish a central limit theorem for the L_p-distance between $\widehat{\Lambda}_n$ and Λ_n. In this paper, we extend the results in Durot and Tocquet, 2003 and Kulikov and Lopuhaä, 2008 to the general setting of Durot, 2007. Our main result is a central limit theorem for the L_p-distance between $\widehat{\Lambda}_n$ and Λ_n, where Λ_n is a naive estimator for the primitive Λ of a monotone curve λ and $\widehat{\Lambda}_n$ is the LCM of Λ_n. As special cases we recover Theorem 5.2 in Durot and Tocquet, 2003 and Theorem 2.1 in Kulikov and Lopuhaä, 2008. Our approach requires another preliminary result, which might be of interest in itself, i.e., a limit process for a suitably scaled difference between $\widehat{\Lambda}_n$ and Λ_n. As special cases we recover Theorem 1 in Wang, 1994, Theorem 4.1 in Durot and Tocquet, 2003, and Theorem 1.1 in Kulikov and Lopuhaä, 2006.

5.1 PROCESS CONVERGENCE

We consider the general setting in Durot, 2007. Let $\lambda : [0,1] \to \mathbb{R}$ be non-increasing and assume that we have at hand a cadlag step estimator Λ_n of

$$\Lambda(t) = \int_0^t \lambda(u)\,du, \quad t \in [0,1].$$

In the sequel we will make use of the following assumptions.

(A1) λ is strictly decreasing and twice continuously differentiable on $[0,1]$ with $\inf_t |\lambda'(t)| > 0$.

(A2) Let B_n be either a Brownian motion or a Brownian bridge. There exists $q > 6$, $C_q > 0$, $L : [0,1] \to \mathbb{R}$, and versions of $M_n = \Lambda_n - \Lambda$ and B_n such that

$$\mathbb{P}\left(n^{1-1/q} \sup_{t\in[0,1]} \left| M_n(t) - n^{-1/2} B_n \circ L(t) \right| > x \right) \leqslant C_q x^{-q}$$

for all $x \in (0,n]$. Moreover, L is increasing and twice differentiable on $[0,1]$, with $\sup_t |L''(t)| < \infty$ and $\inf_t |L'(t)| > 0$.

Note that this setup includes several statistical models, such as monotone density, monotone regression, and the monotone hazard model under random censoring, see Section 3 in Durot, 2007.

We consider the distance between Λ_n and its least concave majorant $\widehat{\Lambda}_n = \mathrm{CM}_{[0,1]}\Lambda_n$, where CM_I maps a function $h : \mathbb{R} \to \mathbb{R}$ into the least concave majorant of h on the interval $I \subset \mathbb{R}$. Consider the process

$$A_n(t) = n^{2/3}\left(\widehat{\Lambda}_n(t) - \Lambda_n(t)\right), \qquad t \in [0,1], \tag{5.1.1}$$

and define

$$Z(t) = W(t) - t^2, \qquad \zeta(t) = [\mathrm{CM}_{\mathbb{R}} Z](t) - Z(t), \tag{5.1.2}$$

where W denotes a standard two-sided Brownian motion originating from zero. For each $t \in (0,1)$ fixed and $t + c_2(t)sn^{-1/3} \in (0,1)$, define

$$\zeta_{nt}(s) = c_1(t)A_n\left(t + c_2(t)sn^{-1/3}\right), \tag{5.1.3}$$

where

$$c_1(t) = \left(\frac{|\lambda'(t)|}{2L'(t)^2}\right)^{1/3}, \qquad c_2(t) = \left(\frac{4L'(t)}{|\lambda'(t)|^2}\right)^{1/3}. \tag{5.1.4}$$

Our first result is the following theorem, which extends Theorem 1.1 in Kulikov and Lopuhaä, 2006.

Theorem 5.1.1. *Suppose that assumptions (A1)-(A2) are satisfied. Let ζ_{nt} and ζ be defined in* (5.1.3) *and* (5.1.2)*. Then the process $\{\zeta_{nt}(s) : s \in \mathbb{R}\}$ converges in distribution to the process $\{\zeta(s) : s \in \mathbb{R}\}$ in $D(\mathbb{R})$, the space of cadlag function on $\mathbb{R}$.*

Note that as a particular case $\zeta_{nt}(0)$ converges weakly to $\zeta(0)$. In this way, we recover Theorem 1 in Wang, 1994 and Theorem 4.1 in Durot and Tocquet, 2003. The proof of Theorem 5.1.1 follows the line of reasoning in Kulikov and Lopuhaä, 2006.

Let us briefly sketch the argument to prove Theorem 5.1.1. Note that $A_n = D_{[0,1]}[n^{2/3}\Lambda_n]$ and $\zeta = D_{\mathbb{R}}[Z]$, where $D_I h = \mathrm{CM}_I h - h$, for $h : \mathbb{R} \to \mathbb{R}$. Since D_I is a continuous mapping, the main idea is to apply the continuous mapping theorem to properly scaled approximations of the processes Λ_n and Z on a suitable chosen fixed interval I. The first step is to determine the weak limit of Λ_n, which is given in the following lemma.

Lemma 5.1.2. *Suppose that assumptions (A1)-(A2) are satisfied. Then for fixed $t \in (0,1)$, the process*

$$X_{nt}(s) = n^{2/3}\left(\Lambda_n(t + sn^{-1/3}) - \Lambda_n(t) - \left(\Lambda(t + sn^{-1/3}) - \Lambda(t)\right)\right)$$

converges in distribution to the process $\{W(L'(t)s) : s \in \mathbb{R}\}$.

Since

$$n^{2/3}(\Lambda(t+sn^{-1/3})-\Lambda(t)) \approx n^{1/3}\lambda(t)s+\lambda'(t)s^2/2$$

and D_I is invariant under addition of linear functions, it follows that the process A_n can be approximated by a Brownian motion with a parabolic drift. The idea now is to use continuity of D_I, for a suitably chosen interval $I=[-d,d]$, to show that $D_I E_{nt}$ converges to $D_I Z_t$, where

$$\begin{aligned} E_{nt}(s) &= n^{2/3}\Lambda_n(t+sn^{-1/3}) \\ Z_t(s) &= W(L'(t)s)+\lambda'(t)s^2/2. \end{aligned} \tag{5.1.5}$$

In order to relate this to the processes ζ_{nt} and ζ in Theorem 5.1.1, note that $A_n(t+sn^{-1/3}) = [D_{I_{nt}}E_{nt}](s)$, where $I_{nt}=[-tn^{1/3},(1-t)n^{1/3}]$, and by Brownian scaling, the process $Z(s)$ has the same distribution as the process $c_1(t)Z_t(c_2(t)s)$. This means that we must compare the concave majorants of E_{nt} on the intervals I_{nt} and I, as well as the concave majorants of Z_t on the interval I and $\mathbb{R}$. Lemma 1.2 in Kulikov and Lopuhaä, 2006 shows that, locally, with high probability, both concave majorants of the process Z_t coincide on $[-d/2,d/2]$, for large $d>0$. A similar result is established for the concave majorants of the process E_{nt} in Lemma 5.1.3, which is analogous to Lemma 1.3 in Kulikov and Lopuhaä, 2006. As a preparation for Theorem 5.2.1, the lemma also contains a similar result for a Brownian motion version of E_{nt}.

Let B_n be as in assumption (A2) and let ξ_n be a $N(0,1)$ distributed random variable independent of B_n, if B_n is a Brownian bridge, and $\xi_n=0$, when B_n is a Brownian motion. Define versions W_n of a Brownian motion by $W_n(t)=B_n(t)+\xi_n t$, for $t\in[0,1]$, and define

$$A_n^W = n^{2/3}\left(CM_{[0,1]}\Lambda_n^W-\Lambda_n^W\right) \tag{5.1.6}$$

where $\Lambda_n^W(t)=\Lambda(t)+n^{-1/2}W_n(L(t))$, with L as in assumption (A2). Furthermore, define $E_n=\sqrt{n}(\Lambda_n-\Lambda)$, $\Lambda_n^E=\Lambda_n$, $A_n^E=A_n$. The superscripts E and W refer to the empirical and Brownian motion version. For $d>0$, let $I_{nt}(d)=[0,1]\cap[t-dn^{-1/3},t+dn^{-1/3}]$ and, for $J=E,W$, define the event

$$N_{nt}^J(d)=\left\{[CM_{[0,1]}\Lambda_n^J](s)=[CM_{I_{nt}(d)}\Lambda_n^J](s),\ \text{for all } s\in I_{nt}(d/2)\right\}. \tag{5.1.7}$$

Let $I_{nt}=I_{nt}(\log n)$ and $N_{nt}^J=N_{nt}^J(\log n)$.

Lemma 5.1.3. *Assume that assumptions (A1)-(A2) hold. For* $d>0$, *let* $N_{nt}^J(d)$ *be the event defined in* (5.1.7). *There exists* $C>0$, *independent of* n, t, d, *such that*

$$\begin{aligned} \mathbb{P}\left((N_{nt}^W(d))^c\right) &= O\left(e^{-Cd^3}\right) \\ \mathbb{P}\left((N_{nt}^E(d))^c\right) &= O\left(n^{1-q/3}d^{-2q}+e^{-Cd^3}\right), \end{aligned}$$

where q *is from assumption (A2).*

The proof of Theorem 5.1.1 now follows the same line of reasoning as that of Theorem 1.1 in Kulikov and Lopuhaä, 2006, see Section 5.3 for more details.

5.2 CLT FOR THE L_p-DISTANCE

The next step is to deal with the L_p norm. Our main result is the following.

Theorem 5.2.1. *Suppose that assumptions (A1)-(A2) are satisfied and let* A_n *and* ζ *be defined by* (5.1.1) *and* (5.1.2)*, respectively. Let* μ *be a measure on the Borel sets of* $\mathbb{R}$*, such that*

(A3) $d\mu(t) = w(t)\,dt$*, where* $w(t) \geqslant 0$ *is differentiable with bounded derivative on* $[0,1]$*.*

Then, for all $1 \leqslant p < \min(q, 2q-7)$*, (with* q *as in assumption (A2)),*

$$n^{1/6}\left(\int_0^1 A_n(t)^p\,d\mu(t) - m\right) \xrightarrow{d} N(0,\sigma^2),$$

where

$$m = \mathbb{E}\left[\zeta(0)^p\right]\int_0^1 \frac{2^{p/3}L'(t)^{2p/3}}{|\lambda'(t)|^{p/3}}\,d\mu(t)$$

and

$$\sigma^2 = \int_0^1 \frac{2^{(2p+5)/3}L'(t)^{(4p+1)/3}}{|\lambda'(t)|^{(2p+2)/3}} w^2(t)\,dt \int_0^\infty \operatorname{cov}\left(\zeta(0)^p, \zeta(s)^p\right)\,ds.$$

For the special cases that λ is a probability density or a regression function, we recover Theorem 2.1 in Kulikov and Lopuhaä, 2008 and Theorem 5.2 inDurot and Tocquet, 2003, respectively. In order to prove Theorem 5.2.1 we first need some preliminary results. We aim at approximating the L_p-norm of A_n by that of the Brownian motion version A_n^W and then finding the asymptotic distribution for the latter one. To this end, we first need to relate the moments of A_n to those of A_n^W. We start by showing that, for $J = E, W$, a rescaled version of Λ_n^J can be approximated by the same process Y_{nt} plus a linear term. This result corresponds to Lemma 4.1 in Kulikov and Lopuhaä, 2008.

Lemma 5.2.2. *Suppose that assumptions (A1)-(A2) are satisfied. Then, for fixed* $t \in (0,1)$ *, for* $J = E, W$*, and* $s \in [-tn^{1/3}, (1-t)n^{1/3}]$*, it holds*

$$n^{2/3}\Lambda_n^J(t + n^{-1/3}s) = Y_{nt}(s) + L_{nt}^J(s) + R_{nt}^J(s),$$

where $L^J_{nt}(s)$ *is linear in* s *and*

$$Y_{nt}(s) = n^{1/6}\left\{W_n(L(t+n^{-1/3}s)) - W_n(L(t))\right\} + \frac{1}{2}\lambda'(t)s^2.$$

Moreover, for all $p \geqslant 1$,

$$\mathbb{E}\left[\sup_{|s|\leqslant \log n}\left|R^W_{nt}(s)\right|^p\right] = O\left(n^{-p/3}(\log n)^{3p}\right),$$

uniformly in $t \in (0,1)$. *If, in addition* $1 \leqslant p < q$ *(with* q *as in assumption (A2)), then*

$$\mathbb{E}\left[\sup_{|s|\leqslant \log n}\left|R^E_{nt}(s)\right|^p\right] = O\left(n^{-p/3+p/q}\right)$$

uniformly in $t \in (0,1)$.

Since the map D_I is invariant under addition of linear terms, Lemma 5.2.2 allows us to approximate the moments of $A^J_n(t) = n^{2/3}D_{[0,1]}\Lambda^J_n$ by those of $[D_{H_{nt}}Y_{nt}](0)$ for some interval H_{nt}, as in Lemma 4.2 in Kulikov and Lopuhaä, 2008.

Lemma 5.2.3. *Suppose that assumptions (A1)-(A2) are satisfied. and let* Y_{nt} *be the process defined in Lemma 5.2.2. Define*

$$H_{nt} = [-n^{1/3}t, n^{1/3}(1-t)] \cap [-\log n, \log n].$$

Then for all $p \geqslant 1$, *it holds*

$$\mathbb{E}\left[A^W_n(t)^p\right] = \mathbb{E}\left[[D_{H_{nt}}Y_{nt}](0)^p\right] + o\left(n^{-1/6}\right),$$

uniformly for $t \in (0,1)$. *If, in addition* $1 \leqslant p < \min(q, 2q-7)$, *with* q *from condition (A2), then also*

$$\mathbb{E}\left[A^E_n(t)^p\right] = \mathbb{E}\left[[D_{H_{nt}}Y_{nt}](0)^p\right] + o\left(n^{-1/6}\right),$$

uniformly for $t \in (0,1)$.

The process Y_{nt} has the same distribution as

$$\tilde{Y}_{nt} = W\left(n^{1/3}\left(L\left(t+n^{-1/3}s\right) - L(t)\right)\right) + \frac{1}{2}\lambda'(t)s^2, \tag{5.2.1}$$

which is close to the process Z_t in (5.1.5) by continuity of Brownian motion. Lemma 4.3 in Kulikov and Lopuhaä, 2008 is then used to show that the concave majorants at zero are sufficiently close. Note that, with by Brownian scaling, the process $c_1(t)Z_t(c_2(t)s)$ has the same distribution as the process $Z(s)$. As a consequence of Lemma 5.2.3 the moments of $A^J_n(t)$ can be related to those of the process ζ. This formulated in the next lemma, which corresponds to Lemma 4.4 in Kulikov and Lopuhaä, 2008.

Lemma 5.2.4. *Suppose that assumptions (A1)-(A2) hold. Then, for all* $p \geqslant 1$,

$$\mathbb{E}\left[A_n^W(t)^p\right] = \left(\frac{2L'(t)^2}{|\lambda'(t)|}\right)^{p/3} \mathbb{E}\left[\zeta(0)^p\right] + o\left(n^{-1/6}\right)$$

uniformly in $t \in (n^{-1/3}\log n, 1 - n^{-1/3}\log n)$ *and*

$$\mathbb{E}\left[A_n^W(t)^p\right] \leqslant \left(\frac{2L'(t)^2}{|\lambda'(t)|}\right)^{p/3} \mathbb{E}\left[\zeta(0)^p\right] + o\left(n^{-1/6}\right)$$

uniformly in $t \in (0,1)$. *If, in addition* $1 \leqslant p < \min(q, 2q-7)$, *where* q *is from assumption (A2), then the same (in)equalities hold for* $A_N^E(t)$.

In Lemmas 5.2.3 and 5.2.4 the moments of A_n^E and A_n^W are approximated by the moments of the same process. This suggests that the difference between them is of smaller order than $n^{-1/6}$. Indeed, on the events N_{nt}^J, where $A_n^J = n^{2/3} D_{I_{nt}} \Lambda_n^J$, we make use of Lemma 5.2.3 and the fact that D_I is invariant under addition of linear functions to obtain that

$$\begin{aligned}
&\sup_{t\in(0,1)} \left| n^{2p/3}[D_{I_{nt}}\Lambda_n^E](t) - n^{2p/3}[D_{I_{nt}}\Lambda_n^W](t) \right| \\
&\leqslant \sup_{t\in(0,1)} \sup_{|s|\leqslant \log n} \left\{ |R_{nt}^E(s)| + |R_{nt}^W(s)| \right\},
\end{aligned}$$

where the processes R_{nt}^J converge to zero sufficiently fast. On the other hand, on $(N_{nt}^J)^c$ we just need the boundedness of the moments of A_n^J, which follows by Lemma 5.2.4 and the fact that the probability of these events is very small (Lemma 5.1.3).

Lemma 5.2.5. *Suppose that assumptions (A1)-(A2) hold. Then, for* $1 \leqslant p < \min(q, 2q-7)$, *with* q *from assumption (A2), it holds*

$$\begin{aligned}
\mathbb{E}\left[\left|A_n^E(t)^p - A_n^W(t)^p\right|\right] &= o\left(n^{-1/6}\right) \\
\mathbb{E}\left[\left|A_n^E(t) - A_n^W(t)\right|^p\right] &= o\left(n^{-1/6}\right)
\end{aligned}$$

uniformly in $t \in (0,1)$.

From Lemma 5.2.4 it follows that

$$n^{1/6}\left| m - \int_0^1 \mathbb{E}\left[A_n^W(t)^p\right] dt \right| \to 0,$$

where m is the asymptotic mean in Theorem 5.2.1. Moreover, Lemma 5.2.5 implies that

$$n^{1/6}\left|\int_0^1 A_n^E(t)^p\,dt-\int_0^1 A_n^W(t)^p\,dt\right|\leqslant n^{1/6}\int_0^1\left|A_n^E(t)^p-A_n^W(t)^p\right|\,dt\to 0.$$

As a consequence, in order to prove Theorem 5.2.1, it suffices to prove asymptotic normality of its Brownian motion version

$$T_n^W=n^{1/6}\int_0^1\left(A_n^W(t)^p-\mathbb{E}\left[A_n^W(t)^p\right]\right)\,d\mu(t).$$

The proof of this is completely similar to that of Theorem 2.1 in Kulikov and Lopuhaä, 2008. First, by using Theorem 5.1.1 for a Brownian version of ζ_{nt} and the mixing property of A_n^W (this can be obtained in the same way as Lemma 4.6 in Kulikov and Lopuhaä, 2008), we derive the asymptotic variance of T_n^W in the following lemma.

Lemma 5.2.6. *Suppose that assumptions (A1)-(A3) are satisfied. Then, for every* $p\geqslant 1$,

$$\begin{aligned}&\mathrm{Var}\left(n^{1/6}\int_0^1 A_n^W(t)^p\,d\mu(t)\right)\\&\to\int_0^1\frac{2^{(2p+5)/3}L'(t)^{(4p+1)/3}}{|\lambda'(t)|^{(2p+2)/3}}w^2(t)\,dt\int_0^\infty \mathrm{cov}\left(\zeta(0)^p,\zeta(s)^p\right)\,ds.\end{aligned}$$

The last step is proving the asymptotic normality of T_n^W. This is done by a big-blocks small-blocks argument, where the contribution of the small blocks to the asymptotic distribution is negligible, while the mixing property of A_n^W allows us to approximate the sum over the big blocks by a sum of independent random variables which satisfy the assumptions of Lindeberg central limit theorem.

5.3 PROOFS

Proof of Lemma 5.1.2. The proof is completely similar to that of Lemma 1.1 in Kulikov and Lopuhaä, 2006, but this time $E_n=\sqrt{n}(\Lambda_n-\Lambda)$ and

$$\sup_{t\in[0,1]}|E_n(t)-B_n\circ L(t)|=O_p(n^{-1/2+1/q}),$$

according to (A2). Similar to the proof of Lemma 1.1 in Kulikov and Lopuhaä, 2006, this means that

$$\begin{aligned} X_{nt}(s) &= n^{1/6}\left(W_n(L(t+sn^{-1/3}))-W_n(L(t))\right)+O_p(n^{-1/3+1/q}) \\ &\stackrel{d}{=} W(L'(t)s)+R_n(s), \end{aligned}$$

where $\sup_{s\in I}|R_n(s)|\to 0$ in probability for compact $I\subset\mathbb{R}$. From here on the proof is the same as that of Lemma 1.1 in Kulikov and Lopuhaä, 2006. □

Proof Lemma 5.1.3. Let $\widehat{\lambda}_n^W$ be the left derivative of $\widehat{\Lambda}_n^W = CM_{[0,1]}\Lambda_n^W$. Define the inverse process

$$U_n^W(a) = \operatorname*{argmax}_{t\in[0,1]}\left\{\Lambda_n^W(t)-at\right\}$$

and

$$V_n^W(a) = n^{1/3}\left(L(U_n^W(a))-L(g(a))\right),$$

where g denotes the inverse of λ. As in the proof of Lemma 1.3 in Kulikov and Lopuhaä, 2006 [see (2.2)], we get

$$\begin{aligned} \mathbb{P}\left((N_{nt}^W(d))^c\right) \leqslant\ & \mathbb{P}\left(\widehat{\lambda}_n^W(t-n^{-1/3}d)=\widehat{\lambda}_n^W(t-n^{-1/3}d/2)\right) \\ & +\mathbb{P}\left(\widehat{\lambda}_n^W(t+n^{-1/3}d)=\widehat{\lambda}_n^W(t+n^{-1/3}d/2)\right). \end{aligned} \tag{5.3.1}$$

Then, with $s=t-dn^{-1/3}/2$, $x=d/2$, and $\epsilon_n=\inf_{t\in[0,1]}|\lambda'(t)|dn^{-1/3}/8$, it holds (see (2.3) in Kulikov and Lopuhaä, 2006),

$$\begin{aligned} &\mathbb{P}\left(\widehat{\lambda}_n^W(t-n^{-1/3}d)=\widehat{\lambda}_n^W(t-n^{-1/3}d/2)\right) \\ &\leqslant \mathbb{P}\left(\widehat{\lambda}_n^W(s+n^{-1/3}x)-\lambda(s+n^{-1/3}x)\geqslant\epsilon_n\right) \\ &\quad+\mathbb{P}\left(\widehat{\lambda}_n^W(s)-\lambda(s)<-\epsilon_n\right). \end{aligned} \tag{5.3.2}$$

Moreover, using the switching relation $\widehat{\lambda}_n^W(t)\geqslant a \Leftrightarrow U_n^W(a)\geqslant t$, we rewrite this probability as

$$\begin{aligned} &\mathbb{P}\left\{U_n^W\left(\lambda\left(s+n^{-1/3}x\right)+\epsilon_n\right)\geqslant s+n^{-1/3}x\right\} \\ &=\mathbb{P}\left\{V_n^W\left(\lambda(s+n^{-1/3}x)+\epsilon_n\right)\geqslant\right. \\ &\qquad\left. n^{1/3}\left(L\left(s+n^{-1/3}x\right)-L\left(g\left(\lambda\left(s+n^{-1/3}x\right)+\epsilon_n\right)\right)\right)\right\} \\ &=\mathbb{P}\left\{V_n^W\left(\lambda\left(s+n^{-1/3}x\right)+\epsilon_n\right)\geqslant\frac{\inf_{t\in[0,1]}|\lambda'(t)|\inf_{t\in[0,1]}L'(t)d}{8\sup_{t\in[0,1]}|\lambda'(t)|}\right\}. \end{aligned}$$

It suffices to show that there exists positive constants C_1, C_2 such that

$$\mathbb{P}\left(V_n^W(a) \geqslant x\right) \leqslant C_1 e^{-C_2 x^3} \tag{5.3.3}$$

because then it follows that

$$\mathbb{P}\left(V_n^W\left(\lambda\left(s+n^{-1/3}x\right)+\epsilon_n\right) \geqslant \frac{\inf_t |\lambda'(t)| \inf_t L'(t) d}{8 \sup_t |\lambda'(t)|}\right) \leqslant \tilde{C}_1 e^{-\tilde{C}_2 d^3}.$$

Similarly we can also bound the second probabilities in (5.3.1) and (5.3.2). Then the statement of the lemma follows immediately.

Now we prove (5.3.3). First write

$$\begin{aligned} V_n^W(a) &= n^{1/3}\left(L\left(\operatorname*{argmax}_{t\in[0,1]}\left\{W(L(t))+\sqrt{n}(\Lambda(t)-at)\right\}\right)-L(g(a))\right) \\ &= n^{1/3}\left(\operatorname*{argmax}_{s\in[L(0),L(1)]}\left\{W(s)+\sqrt{n}\left(\Lambda\left(L^{-1}(s)\right)-aL^{-1}(s)\right)\right\}-L(g(a))\right). \end{aligned}$$

Using properties of the argmax functional we obtain that the right hand side is equal to the argmax of the process

$$\begin{aligned} &n^{1/6}\left\{W\left(n^{-1/3}s+L(g(a))\right)-W\left(L(g(a))\right)\right\} \\ &+n^{2/3}\left\{\Lambda\left(L^{-1}\left(n^{-1/3}s+L(g(a))\right)\right)-\Lambda(g(a))\right. \\ &\qquad\left.-aL^{-1}\left(n^{-1/3}s+L(g(a))\right)+ag(a)\right\} \end{aligned}$$

for

$$s \in I_n(a) = [n^{1/3}(L(0)-L(g(a))), n^{1/3}(L(1)-L(g(a)))].$$

By Brownian motion scaling, $V_n^W(a)$ is equal in distribution to

$$\operatorname*{argmax}_{t\in I_n(a)}\{W(t)-D_{a,n}(t)\},$$

where W is a standard two-sided Brownian motion originating from zero and

$$\begin{aligned} D_{a,n}(s) = -n^{2/3}\Big\{&\Lambda\left(L^{-1}\left(n^{-1/3}s+L(g(a))\right)\right)-\Lambda(g(a)) \\ &-aL^{-1}\left(n^{-1/3}s+L(g(a))\right)+ag(a)\Big\}. \end{aligned}$$

By Taylor's formula and the assumptions on λ and L, one can show that there exist a constant $c_0 > 0$, independent of n, a and t, such that $D_{a,n}(t) \geqslant$

$c_0 t^2$. Then (5.3.3) follows from Theorem 4 in Durot, 2002, which proves the first statement.

To continue with the second statement, let $\widehat{\lambda}_n$ be the left derivative of $\widehat{\Lambda}_n$ and define the inverse process

$$U_n(a) = \underset{t\in[0,1]}{\operatorname{argmax}}\{\Lambda_n(t) - at\}, \quad \text{and} \quad V_n(a) = n^{1/3}\left(U_n(a) - g(a)\right),$$

where g denotes the inverse of λ. As in (5.3.1), we get

$$\begin{aligned}\mathbb{P}\left((N^E_{nt}(d))^c\right) \leqslant \mathbb{P}\left(\widehat{\lambda}_n(t - n^{-1/3}d) = \widehat{\lambda}_n(t - n^{-1/3}d/2)\right) \\ + \mathbb{P}\left(\widehat{\lambda}_n(t + n^{-1/3}d) = \widehat{\lambda}_n(t + n^{-1/3}d/2)\right).\end{aligned} \tag{5.3.4}$$

where similar to (5.3.2),

$$\begin{aligned}&\mathbb{P}\left(\widehat{\lambda}_n(t - n^{-1/3}d) = \widehat{\lambda}_n(t - n^{-1/3}d/2)\right) \\ &\leqslant \mathbb{P}\left(\widehat{\lambda}_n(s + n^{-1/3}x) - \lambda(s + n^{-1/3}x) > \epsilon_n\right) \\ &\quad + \mathbb{P}\left(\widehat{\lambda}_n(s) - \lambda(s) < -\epsilon_n\right).\end{aligned} \tag{5.3.5}$$

Then using the switching relation $\widehat{\lambda}_n(t) \leqslant a \Leftrightarrow U_n(a) \leqslant t$, we rewrite the first probability in (5.3.5) as

$$\mathbb{P}\left(V_n\left(\lambda\left(s + n^{-1/3}x\right) + \epsilon_n\right) > \frac{\inf_{t\in[0,1]}|\lambda'(t)|d}{8\sup_{t\in[0,1]}|\lambda'(t)|}\right).$$

According to Lemma 6.4 in Durot, Kulikov, and Lopuhaä, 2012, there exists positive constants $C_1, C_2 > 0$, independent of n, a, and x, such that

$$\mathbb{P}\left(V_n(a) > x\right) \leqslant \frac{C_1 n^{1-q/3}}{x^{2q}} + 2e^{-C_2 x^3}.$$

It follows that

$$\mathbb{P}\left(V_n\left(\lambda(s + n^{-1/3}x) + \epsilon_n\right) > \frac{\inf_{t\in[0,1]}|\lambda'(t)|d}{8\sup_{t\in[0,1]}|\lambda'(t)|}\right) \leqslant \frac{\tilde{C}_1 n^{1-q/3}}{d^{2q}} + 2e^{-\tilde{C}_2 d^3}.$$

Similarly we can also bound the second probabilities in (5.3.4) and (5.3.5). Then the statement of the lemma follows immediately. □

Proof Theorem 5.1.1. The proof is similar to the proof of Theorem 1.1 in Kulikov and Lopuhaä, 2006. We briefly sketch the main steps. Arguing as in the proof of Theorem 1.1 inKulikov and Lopuhaä, 2006, it suffices to

show that for any compact $K \subset \mathbb{R}$, the process $\{A_n(t + sn^{-1/3}) : s \in K\}$ converges in distribution to the process $\{[D_{\mathbb{R}} Z_t](s) : s \in K\}$ on $D(K)$, the space of cadlag functions on K, where Z_t is defined in (5.1.5). By definition $A_n(t + sn^{-1/3}) = [D_{I_{nt}} E_{nt}](s)$, for $s \in I_{nt} = [-tn^{1/3}, (1-t)n^{1/3}]$, where E_{nt} is defined in (5.1.5). To prove convergence in distribution, we show that for any bounded continuous function $g : D(K) \to \mathbb{R}$,

$$|\mathbb{E}\left[g\left(D_{I_{nt}} E_{nt}\right)\right] - \mathbb{E}\left[g\left(D_{\mathbb{R}} Z_t\right)\right]| \to 0. \tag{5.3.6}$$

To this end, we choose $d > 0$ sufficiently large, such that $K \subset [-d/2, d/2] \subset [-d, d] = I$ and take n sufficiently large so that $I \subset I_{nt}$. Then, similar to inequality (2.7) in Kulikov and Lopuhaä, 2006, the triangular inequality yields

$$\begin{aligned} |\mathbb{E}\left[g\left(D_{I_{nt}} E_{nt}\right)\right] - \mathbb{E}\left[D_{\mathbb{R}} Z_t\right]| \leqslant\ & |\mathbb{E}[g(D_{I_{nt}} E_{nt})] - \mathbb{E}[D_I E_{nt}]| \\ & + |\mathbb{E}[g(D_I E_{nt})] - \mathbb{E}[D_I Z_t]| \\ & + |\mathbb{E}[g(D_I Z_t)] - \mathbb{E}[D_{\mathbb{R}} Z_t]|. \end{aligned} \tag{5.3.7}$$

In the same way as in Kulikov and Lopuhaä, 2006, the three terms on the right hand side are shown to go to zero. For the last term on the right hand side of (5.3.7), the argument is exactly the same and makes use of their Lemma 1.2. The first term on the right hand side of (5.3.7) is bounded similar to their inequality (2.9) and then uses Lemma 5.1.3. For the second term on the right hand side of (5.3.7), note that from Lemma 5.1.2, it follows that

$$\begin{aligned} Z_{nt}(s) = n^{2/3} & \left(\Lambda_n(t + sn^{-1/3}) - \Lambda_n(t) - \left(\Lambda(t + sn^{-1/3}) - \Lambda(t)\right)\right) \\ & + \frac{1}{2}\lambda'(t)s^2, \end{aligned}$$

converges in distribution to Z_t. Therefore, because of the continuity of the mapping D_I, we get

$$|\mathbb{E}[h(D_I Z_{nt})] - \mathbb{E}[h(D_I Z_t)]| \to 0,$$

for any $h : D(I) \to \mathbb{R}$ bounded and continuous. Moreover, we now have

$$E_{nt}(s) = Z_{nt}(s) + n^{2/3}\Lambda_n(t) + \lambda(t)sn^{1/3} + R_{nt}(s),$$

where

$$R_{nt}(s) = n^{2/3}\left(\Lambda(t + sn^{-1/3}) - \Lambda(t) - \lambda(t)sn^{-1/3} - \frac{1}{2}\lambda'(t)s^2 n^{-2/3}\right).$$

Similar to the argument leading up to (2.11) in Kulikov and Lopuhaä, 2006, from the continuity of D_I, its invariance under addition of linear functions, and continuity of λ', it follows that $|\mathbb{E}[g(D_I Z_{nt})] - \mathbb{E}[g(D_I E_{nt})]| \to 0$. This establishes (5.3.6) and finishes the proof. $\square$

Proof of Lemma 5.2.2. By a Taylor expansion, together with (5.1.6), we can write

$$n^{2/3}\Lambda_n^W(t+n^{-1/3}s) = Y_{nt}(s) + L_{nt}^W(s) + R_{nt}^W(s),$$

where

$$L_{nt}^W(s) = n^{2/3}\Lambda(t) + n^{1/6}W_n(L(t)) + n^{1/3}\lambda(t)s$$

and

$$\begin{aligned} R_{nt}^W(s) &= n^{2/3}\left(\Lambda(t+n^{-1/3}s) - \Lambda(t) - n^{-1/3}\lambda(t)s - \frac{1}{2}n^{-2/3}\lambda'(t)s^2\right) \\ &= \frac{1}{6}n^{-1/3}\lambda''(\theta_1)s^3 \end{aligned}$$

for some $|\theta_1 - t| \leqslant n^{-1/3}|s|$. Then, from the assumptions (A1)-(A2), it follows that

$$\sup_{|s|\leqslant \log n}\left|R_{nt}^W(s)\right|^p = O\left(n^{-p/3}(\log n)^{3p}\right),$$

uniformly in $t \in (0,1)$. Similarly, we also obtain

$$\begin{aligned} n^{2/3}\Lambda_n^E(t+n^{-1/3}s) &= n^{2/3}\Lambda_n^W(t+n^{-1/3}s) - n^{1/6}\zeta_n\left(L(t) + L'(t)n^{-1/3}s\right) \\ &\quad + n^{1/6}\left(E_n(t+n^{-1/3}s) - B_n(L(t+n^{-1/3}s))\right) \\ &\quad - n^{1/6}\zeta_n\left(L(t+n^{-1/3}s) - L(t) - L'(t)n^{-1/3}s\right) \\ &= Y_{nt}(s) + L_{nt}^E(s) + R_{nt}^E(s), \end{aligned}$$

where $L_{nt}^E(s) = L_{nt}^W(s) - n^{1/6}\zeta_n L(t) - n^{-1/6}\zeta_n L'(t)s$ and

$$\begin{aligned} R_{nt}^E(s) &= R_{nt}^W(s) + n^{1/6}\left(E_n(t+n^{-1/3}s) - B_n(L(t+n^{-1/3}s))\right) \\ &\quad - \frac{1}{2}n^{-1/2}\zeta_n L''(\theta_2)s^2, \end{aligned}$$

for some $|\theta_2 - t| \leqslant n^{-1/3}|s|$. Let $S_n = \sup_{s\in[0,1]}|E_n(s) - B_n(L(s))|$. From assumption (A2) we have $\mathbb{P}(S_n > n^{-1/2+1/q}x) \leqslant C_q x^{-q}$ and it follows that

$$\begin{aligned} \mathbb{E}\left[S_n^p\right] &= \int_0^\infty \mathbb{P}\left(S_n^p \geqslant x\right)\,dx = p\int_0^\infty y^{p-1}\mathbb{P}\left(S_n \geqslant y\right)\,dy \\ &= pn^{-p/2+p/q}\int_0^\infty x^{p-1}\mathbb{P}\left(S_n \geqslant n^{-1/2+1/q}x\right)\,dx \\ &\leqslant pn^{-p/2+p/q}\left\{\int_0^1 x^{p-1}\,dx + C_q\int_1^\infty x^{p-1-q}\,dx\right\} \\ &= O\left(n^{-p/2+p/q}\right), \end{aligned} \tag{5.3.8}$$

if $p < q$. Consequently $\mathbb{E}\left[\sup_{|s|\leqslant\log n}\left|R_{nt}^E(s)\right|^p\right] = O\left(n^{-p/3+p/q}\right)$. □

Proof of Lemma 5.2.3. Note that we can write

$$A_n^J(t)\mathbb{1}_{N_{nt}^J} = n^{2/3}[D_{I_{nt}}\Lambda_n^J](t)\mathbb{1}_{N_{nt}^J}.$$

We have

$$\begin{aligned}\mathbb{E}\left[A_n^J(t)^p\right] &= n^{2p/3}\mathbb{E}\left[[D_{I_{nt}}\Lambda_n^J](t)^p\right] \\ &\quad + \mathbb{E}\left[\left(A_n^J(t)^p - n^{2p/3}[D_{I_{nt}}\Lambda_n^J](t)^p\right)\mathbb{1}_{(N_{nt}^J)^c}\right].\end{aligned}$$

To bound the second term on the right hand side, first note that

$$\left|A_n^J(t)^p - n^{2p/3}[D_{I_{nt}}\Lambda_n^J](t)^p\right| \leqslant 2A_n^J(t)^p, \tag{5.3.9}$$

because the LCM on $[0,1]$ always lies above the LCM over I_{nt}. Since Λ is concave, we have that

$$\begin{aligned}\left|CM_{[0,1]}\Lambda_n^E - \Lambda_n^E\right| &\leqslant \left|CM_{[0,1]}\Lambda_n^E - [CM_{[0,1]}\Lambda\right| + \left|\Lambda_n^E - \Lambda\right| + \left|CM_{[0,1]}\Lambda - \Lambda\right| \\ &= \left|CM_{[0,1]}\Lambda_n^E - [CM_{[0,1]}\Lambda\right| + \left|\Lambda_n^E - \Lambda\right| \\ &\leqslant 2\sup_{s\in[0,1]}\left|\Lambda_n^E(s) - \Lambda(s)\right|,\end{aligned}$$

which means that

$$0 \leqslant A_n^E(t)^p \leqslant 2^p n^{2p/3}\sup_{s\in[0,1]}\left|\Lambda_n^E(s) - \Lambda(s)\right|^p.$$

Furthermore,

$$0 \leqslant A_n^W(t)^p \leqslant 2^p n^{2p/3}\left\{\Lambda(1) + n^{-1/2}\sup_{s\in[0,1]}|W_n(s)|\right\}^p.$$

In contrast to Kulikov and Lopuhaä, 2008 it is more convenient to treat both cases separately. For the case $J = E$, with (5.3.9), we find that

$$\begin{aligned}&\mathbb{E}\left[\left(A_n^E(t)^p - n^{2p/3}[D_{I_{nt}}\Lambda_n^E](t)^p\right)\mathbb{1}_{(N_{nt}^E)^c}\right] \\ &\leqslant 2^{p+1}n^{2p/3}\mathbb{E}\left[\sup_{s\in[0,1]}\left|\Lambda_n^E(s) - \Lambda(s)\right|^p\mathbb{1}_{(N_{nt}^J)^c}\right],\end{aligned}$$

where

$$\begin{aligned}\sup_{s\in[0,1]}\left|\Lambda_n^E(s) - \Lambda(s)\right|^p \leqslant 2^p\Bigg\{&\sup_{s\in[0,1]}\left|\Lambda_n^E(s) - \Lambda(s) - n^{-1/2}W_n(L(s))\right|^p \\ &+ n^{-p/2}\sup_{s\in[0,1]}|W_n(L(s))|^p\Bigg\}.\end{aligned}$$

For the first term on the right hand side we get with Hölder's inequality

$$
\begin{aligned}
&n^{2p/3}\mathbb{E}\left[\sup_{s\in[0,1]}\left|\Lambda_n^E(s)-\Lambda(s)-n^{-1/2}W_n(L(s))\right|^p \mathbb{1}_{(N_{nt}^E)^c}\right]\\
&\leqslant n^{2p/3}\mathbb{E}\left[\sup_{s\in[0,1]}\left|\Lambda_n^E(s)-\Lambda(s)-n^{-1/2}W_n(L(s))\right|^{p\ell}\right]^{1/\ell}\mathbb{P}\left((N_{nt}^E)^c\right)^{1/\ell'}\\
&=n^{2p/3}O\left(n^{-p+p/q}\right)O\left(n^{1-q/3}(\log n)^{-2q}+e^{-C(\log n)^3}\right)^{1/\ell'},
\end{aligned}
$$

for any $\ell,\ell'>1$ such that $1/\ell+1/\ell'=1$, according to (5.3.8) and Lemma 5.1.3. When $q>6$, then the right hand side is of the order $o(n^{-1/6})$. For the second term, with Hölder's inequality

$$
\begin{aligned}
&n^{2p/3}\mathbb{E}\left[\sup_{s\in[0,1]}|W_n(L(s))|^p \mathbb{1}_{(N_{nt}^E)^c}\right]\\
&\leqslant n^{2p/3}\mathbb{E}\left[\sup_{s\in[0,L(1)]}|W_n(s)|^{p\ell}\right]^{1/\ell}\mathbb{P}\left((N_{nt}^E)^c\right)^{1/\ell'}
\end{aligned}
$$

for any $\ell,\ell'>1$ such that $1/\ell+1/\ell'=1$. Since $\sup_{s\in[0,L(1)]}|W_n(s)|$ has finite moments of any order, it follows from Lemma 5.1.3 that the right hand side is of the order

$$
\begin{aligned}
&n^{2p/3}\mathbb{E}\left[\sup_{s\in[0,1]}|W_n(L(s))|^p \mathbb{1}_{(N_{nt}^E)^c}\right]\\
&\leqslant n^{2p/3}O\left(n^{1-q/3}(\log n)^{-2q}+e^{-C(\log n)^3}\right)^{1/\ell'}.
\end{aligned}
$$

Hence, because $q>6$ and $p<2q-7$, it follows that

$$\left|A_n^E(t)^p-n^{2p/3}[D_{I_{nt}}\Lambda_n^E](t)\right|=o(n^{-1/6}).$$

Next, consider the case $J=W$. Then with (5.3.9) and Cauchy-Schwarz, we find

$$
\begin{aligned}
&\mathbb{E}\left[\left(A_n^W(t)^p-n^{2p/3}[D_{I_{nt}}\Lambda_n^W](t)^p\right)\mathbb{1}_{(N_{nt}^W)^c}\right]\\
&\leqslant 2^{p+1}n^{2p/3}\mathbb{E}\left[\left(\Lambda(1)+n^{-1/2}\sup_{s\in[0,1]}|W_n(s)|\right)^{2p}\right]^{1/2}\left\{\mathbb{P}\left((N_{nt}^W)^c\right)\right\}^{1/2}.
\end{aligned}
$$

Again using that all moments of $\sup_{s\in[0,L(1)]}|W_n(s)|$ are finite, according to Lemma 5.1.3, the right hand side is of the order

$$n^{2p/3}O(e^{-C(\log n)^3})=o(n^{-1/6}).$$

It follows that for $J = E, W$,

$$\mathbb{E}\left[A_n^J(t)^p\right] = n^{2p/3}\mathbb{E}\left[[D_{I_{nt}}\Lambda_n^J](t)^p\right] + o\left(n^{-1/6}\right).$$

Moreover, Lemma 5.2.2 implies that

$$n^{2/3}\left[D_{I_{nt}}\Lambda_n^J\right](t) = [D_{H_{nt}}Y_{nt}](0) + \Delta_{nt},$$

where

$$\Delta_{nt} = \left[D_{H_{nt}}(Y_{nt} + R_{nt}^J)\right](0) - [D_{H_{nt}}Y_{nt}](0).$$

From Lemma 5.2.2, we have

$$\mathbb{E}|\Delta_{nt}|^p \leqslant 2^p\mathbb{E}\left[\sup_{|s|\leqslant \log n}\left|R_{nt}^J(s)\right|^p\right] = O\left(n^{-p/3+p/q}\right). \tag{5.3.10}$$

Then as in Lemma 4.2 in Kulikov and Lopuhaä, 2008, one can show that

$$\begin{aligned}\mathbb{E}\left[A_n^J(t)^p\right] &= \mathbb{E}\left[[D_{H_{nt}}Y_{nt}(0)^p] + \epsilon_{nt} + o\left(n^{-1/6}\right)\right.\\ &= \mathbb{E}\left[[D_{H_{nt}}Y_{nt}(0)^p] + O\left(n^{-1/3+1/q}(\log n)^{2p-2}\right) + o\left(n^{-1/6}\right)\right.\\ &= \mathbb{E}\left[[D_{H_{nt}}Y_{nt}(0)^p] + o\left(n^{-1/6}\right).\right.\end{aligned}$$

This concludes the proof. □

Proof of Lemma 5.2.4. The proof is exactly the same as the one for Lemma 4.4 in Kulikov and Lopuhaä, 2008. Define

$$J_{nt} = \left[n^{1/3}\frac{L(a_{nt}) - L(t)}{L'(t)}, n^{1/3}\frac{L(b_{nt}) - L(t)}{L'(t)}\right],$$

where $a_{nt} = \max(0, t - n^{-1/3}\log n)$ and $b_{nt} = \min(1, t + n^{-1/3}\log n)$. Furthermore, here we take

$$\phi_{nt}(s) = \frac{n^{1/3}\left(L(t + n^{-1/3}s) - L(t)\right)}{L'(t)}.$$

As in the proof of Lemma 4.4 in Kulikov and Lopuhaä, 2008, it follows that $1 - \alpha_n \leqslant \phi_{nt}(s)/s \leqslant 1 + \alpha_n$, for $s \in H_{nt}$, the interval from Lemma 5.2.3, and $\alpha_n = C_1 n^{-1/3}\log n$, with $C_1 > 0$ only depending on L'. Let Z_t be the process in (5.1.5). Then

$$(Z_t \circ \phi_{nt})(s) = \tilde{Y}_{nt} + \frac{1}{2}\lambda'(t)s^2\left(\frac{\phi_{nt}(s)^2}{s^2} - 1\right),$$

where $\tilde{Y}_{nt}$ is defined in (5.2.1). Lemma 4.3 in Kulikov and Lopuhaä, 2008, then allows us to approximate the moments of $[D_{H_{nt}}\tilde{Y}_{nt}](0)$ by the moments of $[D_{J_{nt}}Z_t](0)$. Completely similar to the proof of Lemma 4.4 in Kulikov and Lopuhaä, 2008, the result now follows from Lemma 5.2.3 and Brownian scaling. □

Proof of Lemma 5.2.5. Let I_{nt} and N^J_{nt} be as in Lemma 5.2.3 and define $K_{nt} = N^E_{nt} \cap N^W_{nt}$. Then

$$\mathbb{E}\left|A^E_n(t)^p - A^W_n(t)^p\right| = n^{2p/3}\mathbb{E}\left|[D_{I_{nt}}\Lambda^E_n](t)^p - [D_{I_{nt}}\Lambda^W_n](t)^p\right| \mathbb{1}_{K_{nt}} + \mathbb{E}\left|A^E_n(t)^p - A^W_n(t)^p\right| \mathbb{1}_{K^c_{nt}}. \tag{5.3.11}$$

We bound the two terms on the right hand side, following the same line of reasoning as in Lemma 4.5 in Kulikov and Lopuhaä, 2008. Using that according to Lemma 5.1.3,

$$\mathbb{P}(K^c_{nt}) \leqslant \mathbb{P}\left((N^E_{nt})^c\right) + \mathbb{P}\left((N^W_{nt})^c\right) = O\left(n^{1-q/3}(\log n)^{-2q} + e^{-C(\log n)^3}\right),$$

the second term on the right hand side of (5.3.11) is of the order

$$O\left(\mathbb{P}(K^c_{nt})^{1/2}\right) = o\left(n^{-1/6}\right),$$

because $q > 6$. On the other hand, the first term on the right hand side of (5.3.11) can be bounded by

$$p\mathbb{E}\left[\left(A^E_n(t)^{p-1} + A^W_n(t)^{p-1}\right)^2\right]^{\frac{1}{2}} \mathbb{E}\left[\left(\sup_{|s|\leqslant \log n}|R^E_{nt}| + \sup_{|s|\leqslant \log n}|R^W_{nt}|\right)^2\right]^{\frac{1}{2}},$$

where the right hand side is of the order $O\left(n^{-1/3+1/q}\right) = o\left(n^{-1/6}\right)$, according to Lemmas 5.2.2 and 5.2.4.

In the same way, we have

$$\mathbb{E}\left[\left|A^E_n(t) - A^W_n(t)\right|^p \mathbb{1}_{K^c_{nt}}\right] = O\left(\mathbb{P}(K^c_{nt})^{1/2}\right) = o\left(n^{-1/6}\right)$$

and

$$n^{2p/3}\mathbb{E}\left[\left|[D_{I_{nt}}\Lambda^E_n](t) - [D_{I_{nt}}\Lambda^W_n](t)\right|^p \mathbb{1}_{K_{nt}}\right] \leqslant \mathbb{E}\left[\left(\sup_{|s|\leqslant \log n}|R^E_{nt}| + \sup_{|s|\leqslant \log n}|R^W_{nt}|\right)^p\right]$$

which is of the order $O\left(n^{-p/3+p/q}\right) = o\left(n^{-1/6}\right)$, according to Lemma 5.2.2. □

Proof of Lemma 5.2.6. The proof is completely similar to the one of Lemma 4.7 in Kulikov and Lopuhaä, 2008. For $t \in (0,1)$ fixed, and $t + c_2(t)sn^{-1/3} \in (0,1)$, let

$$\zeta_{nt}(s) = c_1(t)A_n^W\left(t + c_2(t)sn^{-1/3}\right),$$

where A_n^W is defined in (5.1.6) and $c_1(t)$ and $c_2(t)$ are defined in (5.1.4). According to Theorem 5.1.1, ζ_{nt} converges in distribution to ζ, as defined in (5.1.2). As in the proof of Lemma 4.7 in Kulikov and Lopuhaä, 2008, Lemma 5.2.4 yields that, for s, t, and k fixed, the sequence $\zeta_{nt}^W(s)^k$ is uniformly integrable, so that the moments of $(\zeta_{nt}^W(0)^k, \zeta_{nt}^W(s)^k)$ converge to the corresponding moments of $(\zeta(0)^k, \zeta(s)^k)$.

Furthermore, the process $\{A_n^W(t) : t \in (0,)\}$ is strong mixing, i.e., for $d > 0$,

$$\sup |P(A \cap B) - P(A)P(B)| = \alpha_n(d) = 48e^{-Cnd^3} \tag{5.3.12}$$

where $C > 0$ only depends on λ and L from (A2), and where the supremum is taken over all sets

$$A \in \sigma\left\{A_n^W(s) : 0 \leqslant s \leqslant t\right\} \quad \text{and} \quad B \in \sigma\left\{A_n^W(s) : t + d \leqslant s < 1\right\}.$$

This can be obtained by arguing completely the same as in the proof of Lemma 4.6 in Kulikov and Lopuhaä, 2008. The rest of the proof is the same as that of Lemma 4.7 in Kulikov and Lopuhaä, 2008. □

Proof of Theorem 5.2.1. The proof is completely similar to the proof of Theorem 2.1 in Kulikov and Lopuhaä, 2008, by using the method of big-blocks small-blocks and the exponential decreasing mixing function α_n from (5.3.12). □

6

ON THE L_p-ERROR OF SMOOTH ISOTONIC ESTIMATORS

In this chapter we investigate the L_p-error of smooth isotonic estimators obtained by kernel smoothing the Grenander-type estimator or by isotonizing the ordinary kernel estimator. The results presented are based on:

Lopuhaä, H. P. and Musta E. (2018) "Central limit theorems for global errors of smooth isotonic estimators". *Submitted to Electronic Journal of Statistics*.

We consider the same general setup as in Durot, 2007, which includes estimation of a probability density, a regression function, or a failure rate under monotonicity constraints (see Section 3 in Durot, 2007 for more details on these models). An essential assumption in this setup is that the observed process of interest can be approximated by a Brownian motion or a Brownian bridge. Our main results are central limit theorems for the L_p-error of smooth isotonic estimators for a monotone function on a compact interval. However, since the behavior of these estimators is closely related to the behavior of ordinary kernel estimators, we first establish a central limit theorem for the L_p-error of ordinary kernel estimators for a monotone function on a compact interval. This extends the work by Csörgő and Horváth, 1988 on the L_p-error of densities that are smooth on the whole real line, but is also of interest by itself. The fact that we no longer have a smooth function on the whole real line, leads to boundary effects. Unexpectedly, different from Csörgő and Horváth, 1988, we find that the limit variance of the L_p-error changes, depending on whether the approximating process is a Brownian motion or a Brownian bridge. Such a phenomenon has also not been observed in other isotonic problems, where a similar embedding assumption was made. Usually, both approximations lead to the same asymptotic results (e.g., see Durot, 2007 and Kulikov and Lopuhaä, 2005).

After establishing a central limit theorem for the L_p-error of ordinary kernel estimators, we transfer this result to the smoothed Grenander estimator (SG). The key ingredient here is the behavior of the process obtained as the difference between a naive estimator and its least concave majorant.

For this we use results from Chapter 5. As an intermediate result, we show that the L_p-distance between the smoothed Grenander-type estimator and the ordinary kernel estimator converges at rate $n^{2/3}$ to some functional of two-sided Brownian motion minus a parabolic drift.

The situation for the isotonized kernel estimator (GS) is much easier, because it can be shown that this estimator coincides with the ordinary kernel estimator on large intervals in the interior of the support, with probability tending to one. However, since the isotonization step is performed last, the estimator is inconsistent at the boundaries. For this reason, we can only obtain a central limit theorem for the L_p-error on a sub-interval that approaches the whole support, as n diverges to infinity. Finally, the results on the L_p-error can be applied immediately to obtain a central limit theorem for the Hellinger loss.

The chapter is organized as follows. In Section 6.1 we describe the model, the assumptions and fix some notation that will be used throughout the paper. A central limit theorem for the L_p-error of the kernel estimator is obtained in Section 6.2. This result is used in Section 6.3 and 6.4 to obtain the limit distribution of the L_p-error of the SG and GS estimators. Section 6.5 is dedicated to corresponding asymptotics for the Hellinger distance. In Section 6.6 we provide a possible application of our results by considering a test for monotonicity on the basis of the L_2-distance between the kernel estimator and the smoothed Grenander-type estimator. Details of some of the proofs are delayed to Section 6.7 and additional technicalities have been put in Appendix B.

6.1 ASSUMPTIONS AND NOTATIONS

Consider estimating a function $\lambda : [0,1] \to \mathbb{R}$ subject to the constraint that it is non-increasing. Suppose that on the basis of n observations we have at hand a cadlag step estimator Λ_n of

$$\Lambda(t) = \int_0^t \lambda(u)\,du, \quad t \in [0,1].$$

A typical example is the estimation of a monotone density λ on a compact interval. In that case, Λ_n is the empirical distribution function. Hereafter M_n denotes the process $M_n = \Lambda_n - \Lambda$, μ is a measure on the Borel sets of $\mathbb{R}$, and

k is a twice differentiable symmetric probability density with support $[-1,1]$. (6.1.1)

The rescaled kernel is defined as $k_b(u) = b^{-1}k(u/b)$ where the bandwidth $b = b_n \to 0$, as $n \to \infty$. In the sequel we will make use of the following assumptions.

(A1) λ is decreasing and twice continuously differentiable on $[0,1]$ with $\inf_t |\lambda'(t)| > 0$.

(A2) Let B_n be either a Brownian motion or a Brownian bridge. There exists $q > 5/2$, $C_q > 0$, $L : [0,1] \to \mathbb{R}$ and versions of M_n and B_n such that

$$\mathbb{P}\left(n^{1-1/q} \sup_{t\in[0,1]} \left|M_n(t) - n^{-1/2}B_n \circ L(t)\right| > x\right) \leqslant C_q x^{-q}$$

for all $x \in (0,n]$. Moreover, L is increasing and twice differentiable on $[0,1]$ with $\sup_t |L''(t)| < \infty$ and $\inf_t |L'(t)| > 0$.

(A3) $d\mu(t) = w(t)\,dt$, where $w(t) \geqslant 0$ is continuous on $[0,1]$.

In particular, the approximation of the process M_n by a Gaussian process, as in assumption (A2), is required also in Durot, 2007. It corresponds to a general setting which includes estimation of a probability density, regression function or a failure rate under monotonicity constraints (see Section 3 in Durot, 2007 for more details on these models).

First we introduce some notation. We partly adopt the one used in Csörgő and Horváth, 1988 and briefly explain their appearance. Let $\tilde{\lambda}_n^s$ be the standard kernel estimator of λ, i.e.

$$\tilde{\lambda}_n^s(t) = \int_{t-b}^{t+b} k_b(t-u)\,d\Lambda_n(u), \quad \text{for } t \in [b, 1-b]. \tag{6.1.2}$$

As usual we decompose into a random term and a bias term:

$$(nb)^{1/2}\left(\tilde{\lambda}_n^s(t) - \lambda(t)\right) = (nb)^{1/2}\int k_b(t-u)\,d(\Lambda_n - \Lambda)(u) + g_{(n)}(t) \tag{6.1.3}$$

where

$$g_{(n)}(t) = (nb)^{1/2}\left(\lambda_{(n)}(t) - \lambda(t)\right), \qquad \lambda_{(n)}(t) = \int k_b(t-u)\lambda(u)\,du. \tag{6.1.4}$$

When $nb^5 \to C_0 > 0$, then $g_{(n)}(t)$ converges to

$$g(t) = \frac{1}{2}C_0\lambda''(t)\int k(y)y^2\,dy. \tag{6.1.5}$$

After separating the bias term, the first term on the right hand side of (6.1.3) involves an integral of $k_b(t-u)$ with respect to the process M_n. Due to (A2),

this integral will be approximated by an integral with respect to a Gaussian process. For this reason, the limiting moments of the L_p-error involve integrals with respect to Gaussian densities, such as

$$\begin{aligned}\phi(x) &= (2\pi)^{-1/2}\exp(-x^2/2),\\ \psi(u,x,y) &= \frac{1}{2\pi\sqrt{1-u^2}}\exp\left(-\frac{x^2-2uxy+y^2}{2(1-u^2)}\right)\\ &= \frac{1}{\sqrt{1-u^2}}\phi\left(\frac{x-uy}{\sqrt{1-u^2}}\right)\phi(y),\end{aligned} \tag{6.1.6}$$

and a Taylor expansion of $k_b(t-u)$ yields the following constants involving the kernel function:

$$D^2 = \int k(y)^2\,dy, \qquad r(s) = \frac{\int k(z)k(s+z)\,dz}{\int k^2(z)\,dz}. \tag{6.1.7}$$

For example, the limiting means of the L_p-error and a truncated version are given by:

$$\begin{aligned}m_n(p) &= \int_{\mathbb{R}}\int_0^1 \left|\sqrt{L'(t)}Dx + g_{(n)}(t)\right|^p w(t)\phi(x)\,dt\,dx,\\ m_n^c(p) &= \int_{\mathbb{R}}\int_b^{1-b} \left|\sqrt{L'(t)}Dx + g_{(n)}(t)\right|^p w(t)\phi(x)\,dt\,dx,\end{aligned} \tag{6.1.8}$$

where D and $g_{(n)}$ are defined in (6.1.7) and (6.1.4). Depending on the rate at which $b \to 0$, the limiting variance of the L_p-error has a different form. When $nb^5 \to 0$, the limiting variance turns out to be

$$\sigma^2(p) = \sigma_1 D^{2p}\int_0^1 \left|L'(u)\right|^p w(u)^2\,du, \tag{6.1.9}$$

where

$$\sigma_1 = \int_{\mathbb{R}}\left\{\int_{\mathbb{R}}\int_{\mathbb{R}} |xy|^p\psi(r(s),x,y)\,dx\,dy - \int_{\mathbb{R}}\int_{\mathbb{R}} |xy|^p\phi(x)\phi(y)\,dx\,dy\right\}ds, \tag{6.1.10}$$

with σ_1 representing p-th moments of bivariate Gaussian vectors, where D, ψ, and ϕ are defined in (6.1.7) and (6.1.6). When $nb^5 \to C_0 > 0$ and B_n in (A2) is a Brownian motion, the limiting variance of the L_p-error is

$$\begin{aligned}\theta^2(p) = \int_0^1\int_{\mathbb{R}^3} &\left|g(u)^2 + g(u)(x+y)\sqrt{L'(u)}D + D^2L'(u)xy\right|^p\\ &w^2(u)\Big(\psi(r(s),x,y) - \phi(x)\phi(y)\Big)\,ds\,dy\,dx\,du,\end{aligned} \tag{6.1.11}$$

where g, D, ψ, and ϕ are defined in (6.1.5), (6.1.7) and (6.1.6), whereas, if B_n in (A2) is a Brownian bridge, the limiting variance is slightly different,

$$\tilde{\theta}^2(p) = \theta^2(p) - \frac{\theta_1^2(p)}{D^2 L(1)}, \tag{6.1.12}$$

with

$$\theta_1(p) = \int_0^1 \int_{\mathbb{R}} \left| \sqrt{L'(t)} Dx + g(t) \right|^p x\phi(x)\, dx \sqrt{L'(t)} w(t) dt. \tag{6.1.13}$$

Finally, the following inequality will be used throughout this chapter:

$$\begin{aligned} \int_A^B \left| |q(t)|^p - |h(t)|^p \right| d\mu(t) &\leqslant p2^{p-1} \int_A^B |q(t) - h(t)|^p \, d\mu(t) \\ &\quad + p2^{p-1} \left(\int_A^B |h(t)|^p \, d\mu(t) \right)^{1-\frac{1}{p}} \left(\int_A^B |q(t) - h(t)|^p \, d\mu(t) \right)^{\frac{1}{p}}, \end{aligned} \tag{6.1.14}$$

where $p \in [1, \infty)$, $-\infty \leqslant A < B \leqslant \infty$ and $q, h \in L_p(A, B)$.

6.2 KERNEL ESTIMATOR OF A DECREASING FUNCTION

We extend the results of Csörgő and Horváth, 1988 and Csörgő, Gombay, and Horváth, 1991 to the case of a kernel estimator of a decreasing function with compact support. Note that, since the function of interest cannot be twice differentiable on $\mathbb{R}$ (not even continuous), the kernel estimator is inconsistent at zero and one. Moreover we show that the contribution of the boundaries to the L_p-error is not negligible, so in order to avoid the L_p-distance to explode we have to restrict ourselves to the interval $[b, 1-b]$ or apply some boundary correction.

6.2.1 *A modified* L_p*-distance*

Let $\tilde{\lambda}_n^s$ be the standard kernel estimator of λ defined in (6.1.2). In order to avoid boundary problems, we start by finding the asymptotic distribution of a modification of the L_p-distance

$$J_n^c(p) = \int_b^{1-b} \left| \tilde{\lambda}_n^s(t) - \lambda(t) \right|^p d\mu(t), \tag{6.2.1}$$

instead of

$$J_n(p) = \int_0^1 \left| \tilde{\lambda}_n^s(t) - \lambda(t) \right|^p d\mu(t). \tag{6.2.2}$$

Theorem 6.2.1. *Assume that (A1)-(A3) hold. Let* k *satisfy* (6.1.1) *and let* J_n^c *be defined in* (6.2.1). *Suppose* $p \geqslant 1$ *and* $nb \to \infty$.

i) If $nb^5 \to 0$, *then*

$$(b\sigma^2(p))^{-1/2}\left\{(nb)^{p/2}J_n^c(p) - m_n^c(p)\right\} \xrightarrow{d} N(0,1);$$

ii) If $nb^5 \to C_0^2 > 0$, *and* B_n *in Assumption* (A2) *is a Brownian motion, then*

$$(b\theta^2(p))^{-1/2}\left\{(nb)^{p/2}J_n^c(p) - m_n^c(p)\right\} \xrightarrow{d} N(0,1);$$

iii) If $nb^5 \to C_0^2 > 0$, *and* B_n *in Assumption* (A2) *is a Brownian bridge, then*

$$(b\tilde{\theta}^2(p))^{-1/2}\left\{(nb)^{p/2}J_n^c(p) - m_n^c(p)\right\} \xrightarrow{d} N(0,1),$$

where $m_n^c(p)$, $\sigma^2(p)$, $\theta^2(p)$, $\tilde{\theta}^2(p)$ *are defined in* (6.1.8), (6.1.9), (6.1.11), *and* (6.1.12), *respectively.*

The proof goes along the same lines as in the one for the case of the L_p-norms for kernel density estimators on the whole real line (see Csörgő and Horváth, 1988 and Csörgő, Gombay, and Horváth, 1991). The main idea is that by means of assumption (A2), it is sufficient to prove the central limit theorem for the approximating process. When B_n in (A2) is a Brownian motion, the latter one can be obtained by a big-blocks-small-blocks procedure using the independence of the increments of the Brownian motion. When B_n in (A2) is a Brownian bridge, we can still obtain a central limit theorem, but the limiting variance turns out to be different. The latter result differs from what is stated in Csörgő and Horváth, 1988. In Csörgő and Horváth, 1988, the complete proof for both Brownian motion and Brownian bridge, is only given for the case $nb^5 \to 0$, and it is shown that the random variables obtained by using the Brownian motion and the Brownian bridge as approximating processes are asymptotically equivalent (see their Lemma 6). In fact, when dealing with a Brownian bridge, the rescaled L_p-error is asymptotically equivalent to the L_p-error that corresponds to the Brownian motion process plus an additional term which is equal to $CW(L(1))$, for a constant C proportional on $\theta_1(p)$ defined in (6.1.13). When the bandwidth is small, i.e., $nb^5 \to 0$, the bias term $g(t)$ in the definition of $\theta_1(p)$ disappears. Hence, by the symmetry property of the standard normal density, $\theta_1(p) = 0$ and as a consequence $C = 0$. This means that the additional term resulting from the fact that we are dealing with a Brownian bridge converges to zero. For details, see the proof of Lemma 6.7.1. When $nb^5 \to C_0^2 > 0$, only a sketch

of the proof is given in Csörgő and Horváth, 1988 for B_n being a Brownian motion and it is claimed that again the limit distribution would be the same for B_n being a Brownian bridge. However, we find that the limit variances are different.

Proof of Theorem 6.2.1. From the definition of $J_n^c(p)$ we have

$$(nb)^{p/2} J_n^c(p) = \int_b^{1-b} \left| (nb)^{1/2} \int k_b(t-u)\, d(\Lambda_n - \Lambda)(u) + g_{(n)}(t) \right|^p d\mu(t).$$

Let $(W_t)_{t\in\mathbb{R}}$ be a Wiener process and define

$$\Gamma_n^{(1)}(t) = \int k\left(\frac{t-u}{b}\right) dW(L(u)), \tag{6.2.3}$$

Hence, if B_n in assumption (A2) is a Brownian motion, then according to (6.1.14),

$$\begin{aligned}
&\left| (nb)^{p/2} J_n^c(p) - \int_b^{1-b} \left| b^{-1/2}\Gamma_n^{(1)}(t) + g_{(n)}(t) \right|^p d\mu(t) \right| \\
&\leqslant p2^{p-1} b^{-p/2} \int_b^{1-b} \left| \int k\left(\frac{t-u}{b}\right) d(B_n \circ L(u) - n^{1/2} M_n(u)) \right|^p d\mu(t) \\
&+ p2^{p-1} \left(b^{-p/2} \int_b^{1-b} \left| \int k\left(\frac{t-u}{b}\right) d(B_n \circ L - n^{1/2} M_n)(u) \right|^p d\mu(t) \right)^{1/p} \cdot \\
&\qquad \cdot \left(\int_b^{1-b} \left| b^{-1/2}\Gamma_n^{(1)}(t) + g_{(n)}(t) \right|^p d\mu(t) \right)^{1-1/p}
\end{aligned}$$

We can write

$$\begin{aligned}
&\left| \int k\left(\frac{t-u}{b}\right) d(B_n \circ L - n^{1/2} M_n)(u) \right| \\
&= \left| \int_{-1}^{1} k(y)\, d(B_n \circ L - n^{1/2} M_n)(t - by) \right| \\
&= \left| \int_{-1}^{1} (B_n \circ L - n^{1/2} M_n)(t - by)\, dk(y) \right| \\
&\leqslant C \sup_{t\in[0,1]} \left| B_n \circ L(t) - n^{1/2} M_n(t) \right|.
\end{aligned} \tag{6.2.4}$$

According to assumption (A2), the right hand side of (6.2.4) is of the order $O_P(n^{-1/2+1/q})$, and because

$$b^{-1/2} O_P(n^{-1/2+1/q}) = (nb^5)^{3/10} O_P(n^{-2/5+1/q}) = o_P(1)$$

we derive that

$$\left|(nb)^{p/2}J_n^c(p) - \int_b^{1-b} \left|b^{-1/2}\Gamma_n^{(1)}(t) + g_{(n)}(t)\right|^p \, d\mu(t)\right| = o_P(1).$$

As a result, the statement follows from the fact that

$$(b\sigma^2(p))^{-1/2}\left\{\int_b^{1-b} \left|b^{-1/2}\Gamma_n^{(1)}(t) + g_{(n)}(t)\right|^p \, d\mu(t) - m_n^c(p)\right\} \xrightarrow{d} N(0,1),$$

where $g_{(n)}$ and $m_n^c(p)$ are defined in (6.1.4) and (6.1.8), respectively. This result is a generalization of Lemmas 1-5 in Csörgő and Horváth, 1988 and the proof goes in the same way. However, for completeness we give all the details in Appendix B (see Lemma B.1.1).

Finally, if B_n is a Brownian bridge on $[0, L(1)]$, we use the representation $B_n(t) = W(t) - tW(L(1))/L(1)$. By replacing $\Gamma_n^{(1)}$ with

$$\Gamma_n^{(2)}(t) = \int k\left(\frac{t-u}{b}\right) d\left(W(L(u)) - \frac{L(u)}{L(1)}W(L(1))\right) \tag{6.2.5}$$

in the previous reasoning, the statement follows from Lemma 6.7.1. □

When $nb^4 \to 0$, the centering constant $m_n(p)$ can be replaced by a quantity that does not depend on n.

Theorem 6.2.2. *Assume that (A1)-(A3) hold. Let* k *satisfy* (6.1.1) *and let* J_n^c *be defined in* (6.2.1). *Suppose* $p \geqslant 1$ *and* $nb \to \infty$, *such that* $nb^4 \to 0$. *Then*

$$(b\sigma^2(p))^{-1/2}\left\{(nb)^{p/2}J_n^c(p) - m(p)\right\} \xrightarrow{d} N(0,1),$$

where $\sigma^2(p)$ *is defined in* (6.1.9) *and*

$$m(p) = \int_{\mathbb{R}} |x|^p \phi(x)\, dx \left(\int k^2(t)\, dt\right)^{p/2} \int_0^1 |L'(t)|^{p/2}\, d\mu(t).$$

Proof. If $|m_n^c(p) - m(p)| = o(b^{1/2})$, the statement follows from Theorem 6.2.1. First we note that

$$\int_0^b |L'(t)|^{p/2}\, d\mu(t) = o(b^{1/2}) \quad \text{and} \quad \int_{1-b}^1 |L'(t)|^{p/2}\, d\mu(t) = o(b^{1/2}).$$

Moreover, according to (6.1.14), for each $x \in \mathbb{R}$, we have

$$\begin{aligned}
&\int_b^{1-b} \left| \left| \sqrt{L'(t)} Dx + g_{(n)}(t) \right|^p - \left| \sqrt{L'(t)} Dx \right|^p \right| d\mu(t) \\
&\leqslant p2^{p-1} \int_b^{1-b} \left| g_{(n)}(t) \right|^p d\mu(t) \\
&\quad + p2^{p-1} \left(\int_b^{1-b} \left| \sqrt{L'(t)} Dx \right|^p d\mu(t) \right)^{1-1/p} \left(\int_b^{1-b} \left| g_{(n)}(t) \right|^p d\mu(t) \right)^{1/p},
\end{aligned}$$

where $g_{(n)}(t)$ is defined in (6.1.4). Hence, it suffices to prove

$$b^{-p/2} \int_b^{1-b} \left| g_{(n)}(t) \right|^p d\mu(t) = o(1).$$

This follows, since $\sup_{t \in [0,1]} \left| g_{(n)}(t) \right| = O((nb)^{1/2} b^2)$ and

$$b^{-p/2} (nb)^{p/2} b^{2p} = (nb^4)^{p/2} \to 0.$$

□

6.2.2 *Boundary problems*

We show that, actually, we cannot extend the results of Theorem 6.2.1 to the whole interval $[0, 1]$, because then the inconsistency at the boundaries dominates the L_p-error. A similar phenomenon was also observed in the case of the Grenander-type estimator (see Durot, 2007 and Kulikov and Lopuhaä, 2005), but only for $p \geqslant 2.5$. In our case the contribution of the boundaries to the L_p-error is not negligible for all $p \geqslant 1$. This mainly has to do with the fact that the functions $g_{(n)}$, defined in (6.1.4), diverge to infinity. As a result, all the previous theory, which relies on the fact that $g_{(n)} = O(1)$ does not hold. For example, for $t \in [0, b)$, we have

$$\begin{aligned}
g_{(n)}(t) &= (nb)^{1/2} \int_0^{t+b} k_b(t-u)\, d\Lambda(u) - \lambda(t) \\
&= (nb)^{1/2} \int_{-1}^{t/b} k(y)[\lambda(t-by) - \lambda(t)]\, dy - (nb)^{1/2} \lambda(t) \int_{t/b}^{1} k(y)\, dy \\
&= (nb)^{1/2} \left\{ \int_{-1}^{t/b} k(y)[\lambda(t-by) - \lambda(t)]\, dy - \lambda(t) \int_{t/b}^{1} k(y)\, dy \right\}.
\end{aligned} \tag{6.2.6}$$

For the first term within the brackets, we have

$$\left|\int_{-1}^{t/b} k(y)[\lambda(t-by)-\lambda(t)]\,dy\right| \leqslant b \sup_{t\in[0,1]} |\lambda'(t)| \left|\int_{-1}^{t/b} k(y)y\,dy\right| = O(b), \tag{6.2.7}$$

whereas for any $0 < c < 1$ and $t \in [0, cb]$,

$$0 < \inf_{t\in[0,1]} \lambda(t) \int_c^1 k(y)\,dy \leqslant \lambda(t) \int_{t/b}^1 k(y)\,dy \leqslant \lambda(0). \tag{6.2.8}$$

Because $nb \to \infty$, this would mean that

$$\sup_{t\in[0,cb]} g_{(n)}(t) \to -\infty. \tag{6.2.9}$$

What would solve the problem is to assume that λ is twice differentiable as a function defined on $\mathbb{R}$ (see Csörgő and Horváth, 1988 and Csörgő, Gombay, and Horváth, 1991). This is not the case, because here we are considering a function which is positive and decreasing on $[0, 1]$ and usually is zero outside this interval. This means that as a function on $\mathbb{R}$, λ is not monotone anymore and has at least one discontinuity point.

The following results indicate that inconsistency at the boundaries dominates the L_p-error, i.e., the expectation and the variance of the integral close to the end points of the support diverge to infinity. We cannot even approach the boundaries at a rate faster than b (as in the case of the Grenander-type estimator), because the kernel estimator is inconsistent on the whole interval $[0, b)$ (and $(1-b, 1]$).

Proposition 6.2.3. *Assume that (A1)-(A3) hold and let* $\tilde{\lambda}_n^s$ *be defined in* (6.1.2). *Let* k *satisfy* (6.1.1). *Suppose that* $p \geqslant 1$ *and* $nb \to \infty$.

i) When $nb^3 \to \infty$, *then for each* $p \geqslant 1$,

$$(nb)^{p/2}\mathbb{E}\left[\int_0^b \left|\tilde{\lambda}_n^s(t)-\lambda(t)\right|^p\,d\mu(t)\right] \to \infty;$$

ii) If $bn^{1-1/p} \to 0$, *then*

$$b^{-1/2}\left\{\int_0^b (nb)^{p/2}\left|\tilde{\lambda}_n^s(t)-\lambda(t)\right|^p\,d\mu(t) - \int_0^b \left|g_{(n)}(t)\right|^p\,d\mu(t)\right\} \to 0,$$

where $g_{(n)}$ *is defined in* (6.1.4);

iii) Let

$$Y_n(t) = b^{1/2}\int_0^{t+b} k_b(t-u)\,dB_n(L(u)), \qquad t \in [0, b]. \tag{6.2.10}$$

If $b^{-1}n^{-1+1/q} = O(1)$ *and* $b^{p-1}n^{p-2+2/q} \to 0$, *then*

$$b^{-1/2}\left|\int_0^b (nb)^{p/2}\left|\tilde{\lambda}_n^s(t)-\lambda(t)\right|^p d\mu(t) - \int_0^b \left|Y_n(t)+g_{(n)}(t)\right|^p d\mu(t)\right| \tag{6.2.11}$$

converges to zero in probability. Moreover, when $bn^{1-1/p} \to \infty$, *then for all* $0 < c < 1$,

$$b^{-1}\mathrm{Var}\left(\int_0^{cb}\left|Y_n(t)+g_{(n)}(t)\right|^p d\mu(t)\right) \to \infty,$$

where $g_{(n)}$ *is defined in* (6.1.4).

The previous results also hold if we consider the integral on $(1-b, 1]$ *instead of* $[0, b)$.

The proof can be found in Appendix B.

Remark 6.2.4. Note that, if $b \sim n^{-\alpha}$, for some $0 < \alpha < 1$, then for $\alpha < 1/3$, Proposition 6.2.3(i) shows that for all $p \geqslant 1$, the expectation of the boundary regions in the L_p-error tends to infinity. This holds in particular for the optimal choice $\alpha = 1/5$. For $p < 1/(1-\alpha)$, Proposition 6.2.3(ii) allows us to include the boundary regions in the central limit theorem for the L_p-error of the kernel estimator,

$$(b\sigma^2(p))^{-1/2}\left\{(nb)^{p/2}J_n(p) - \bar{m}_n(p)\right\} \xrightarrow{d} N(0,1),$$

with $J_n(p)$ defined in (6.2.2) and $\bar{m}_n(p) = \int_0^1 \left|g_{(n)}(t)\right|^p d\mu(t)$. However, the bias term $\bar{m}_n(p)$ is not bounded anymore. On the other hand, if $p > 1/(1-\alpha)$, Proposition 6.2.3(iii) shows that the boundary regions in the L_p-error behave asymptotically as random variables whose variance tends to infinity.

Remark 6.2.5. The choice of the measure μ instead of the Lebesgue measure, in Csörgő and Horváth, 1988 and Csörgő, Gombay, and Horváth, 1991, is motivated by the fact that, for a particular $\mu(t) = w(t)dt$, the normalizing constants $m(p)$ and $\sigma(p)$ in the CLT will not depend on the unknown function. In our case, a proper choice for μ can also be used to get rid of the boundary problems. This happens when μ puts less mass on the boundary regions in order to compensate the inconsistency of the kernel estimator. For example, if $\mu(t) = t^{2p}(1-t)^{2p}dt$, then

$$\int_0^b |g_{(n)}(t)|^p\, d\mu(t) + \int_{1-b}^1 |g_{(n)}(t)|^p\, d\mu(t) \to 0$$

and, as a result, Theorem 6.2.1 also holds if we replace $J_n^c(p)$ with $J_n(p)$, defined in (6.2.2).

6.2.3 *Kernel estimator with boundary correction*

One way to overcome the inconsistency problems of the standard kernel estimator is to apply some boundary correction. Let now $\hat{\lambda}_n^s$ be the 'corrected' kernel estimator of λ, i.e.

$$\hat{\lambda}_n^s(t) = \int_{t-b}^{t+b} k_b^{(t)}(t-u)\, d\Lambda_n(u), \quad \text{for } t \in [0,1], \tag{6.2.12}$$

where $k_b^{(t)}(u)$ denotes the rescaled kernel $b^{-1}k^{(t)}(u/b)$, with $k^{(t)}(u)$ as in (1.2.11).

We aim at showing that in this case, Theorem 6.2.1 holds for the L_p-error on the whole support, i.e., with $J_n(p)$ instead of $J_n^c(p)$. Note that boundary corrected kernel estimator coincides with the standard kernel estimator on $[b, 1-b]$. Hence the behavior of the L_p-error on $[b, 1-b]$ will be the same. We just have to deal with the boundary regions $[0, b]$ and $[1-b, 1]$.

Proposition 6.2.6. *Assume that (A1)-(A3) hold and let* $\hat{\lambda}_n^s$ *be defined in* (6.2.12). *Let* k *satisfy* (6.1.1) *and suppose* $p \geqslant 1$ *and* $nb \to \infty$. *Then*

$$b^{-1/2}(nb)^{p/2} \int_0^b \left|\hat{\lambda}_n^s(t) - \lambda(t)\right|^p d\mu(t) \xrightarrow{\mathbb{P}} 0.$$

The previous result also holds if we consider the integral on $(1-b, 1]$ *instead of* $[0, b)$.

The proof can be found in Appendix B.

Corollary 6.2.7. *Assume that (A1)-(A3) hold and let* $J_n(p)$ *be defined in* (6.2.2). *Let* k *satisfy* (6.1.1) *and suppose* $p \geqslant 1$ *and* $nb \to \infty$. *Then*

i) if $nb^5 \to 0$, *then it holds*

$$(b\sigma^2(p))^{-1/2}\left\{(nb)^{p/2}J_n(p) - m_n(p)\right\} \xrightarrow{d} N(0,1);$$

ii) If $nb^5 \to C_0^2 > 0$ *and* B_n *in Assumption (A2) is a Brownian motion, then it holds*

$$(b\theta^2(p))^{-1/2}\left\{(nb)^{p/2}J_n(p) - m_n(p)\right\} \xrightarrow{d} N(0,1);$$

iii) If $nb^5 \to C_0^2 > 0$ *and* B_n *in Assumption (A2) is a Brownian bridge, then it holds*

$$(b\tilde{\theta}^2(p))^{-1/2}\left\{(nb)^{p/2}J_n(p) - m_n(p)\right\} \xrightarrow{d} N(0,1),$$

where σ^2, θ^2, $\tilde{\theta}^2$ *and* m_n *are defined respectively in* (6.1.9), (6.1.11), (6.1.12) *and* (6.1.8).

Proof. It follows from combining Theorem 6.2.1 and Proposition 6.2.6, together with the fact that

$$b^{-1/2} \int_{\mathbb{R}} \int_0^b \left| \sqrt{L'(t)} D x + g_{(n)}(t) \right|^p w(t) \phi(x) \, dt \, dx \to 0,$$

where D and $g_{(n)}$ are defined in (6.1.7) and (6.1.4). □

6.3 SMOOTHED GRENANDER-TYPE ESTIMATOR

The smoothed Grenander-type estimator is defined by

$$\tilde{\lambda}_n^{SG}(t) = \int_{0 \vee (t-b)}^{1 \wedge (t+b)} k_b^{(t)}(t-u) \, d\hat{\Lambda}_n(u), \qquad \text{for } t \in [0,1], \tag{6.3.1}$$

where $\hat{\Lambda}_n$ is the least concave majorant of Λ_n. We are interested in the asymptotic distribution of the L_p-error of this estimator:

$$I_n^{SG}(p) = \int_0^1 \left| \hat{\lambda}_n^{SG}(t) - \lambda(t) \right|^p \, d\mu(t). \tag{6.3.2}$$

We will compare the behavior of the L_p-error of $\tilde{\lambda}_n^{SG}$ with that of the regular kernel estimator $\hat{\lambda}_n^s$ from (6.2.12). Because

$$\tilde{\lambda}_n^{SG}(t) - \hat{\lambda}_n^s(t) = \int k_b^{(t)}(t-u) \, d(\hat{\Lambda}_n - \Lambda_n)(u),$$

we will make use of the behavior of $\hat{\Lambda}_n - \Lambda_n$, which has been investigated in Chapter 5, extending similar results from Durot and Tocquet, 2003 and Kulikov and Lopuhaä, 2008. The idea is to represent $\hat{\Lambda}_n - \Lambda_n$ in terms of the mapping CM_I that maps a function $h : \mathbb{R} \to \mathbb{R}$ into the least concave majorant of h on the interval $I \subset \mathbb{R}$, or equivalently by the mapping $D_h = CM_I h - h$.

Let B_n be as in assumption (A2) and ξ_n a $N(0,1)$ distributed r.v. independent of B_n. Define versions W_n of Brownian motion by

$$W_n(t) = \begin{cases} B_n(t) + \xi_n t & \text{if } B_n \text{ is a Brownian bridge} \\ B_n(t) & \text{if } B_n \text{ is a Brownian motion.} \end{cases}$$

Define

$$\begin{aligned} A_n^E &= n^{2/3}\left(CM_{[0,1]}\Lambda_n - \Lambda_n\right) = n^{2/3}D_{[0,1]}\Lambda_n, \\ A_n^W &= n^{2/3}\left(CM_{[0,1]}\Lambda_n^W - \Lambda_n^W\right) = n^{2/3}D_{[0,1]}\Lambda_n^W. \end{aligned} \tag{6.3.3}$$

where

$$\Lambda_n^W(t) = \Lambda(t) + n^{-1/2}W_n(L(t)), \tag{6.3.4}$$

with L as in Assumption (A2). We start with the following result on the L_p-distance between $\tilde{\lambda}_n^{SG}$ and $\hat{\lambda}_n^s$. In order to use results from Chapter 5, we need that $1 \leqslant p < \min(q, 2q-7)$, where q is from Assumption (A2). Moreover, in order to obtain suitable approximations in combination with results from Chapter 5, we require additional conditions on the rate at which $1/b$ tends to infinity. Also see Remark 6.3.2. For the optimal rate $b \sim n^{-1/5}$, the result in Theorem 6.3.1 is valid, as long as $p < 5$ and $q > 9$.

Theorem 6.3.1. *Assume that* $(A1) - (A2)$ *hold and let* μ *be a finite measure on* $(0,1)$. *Let* k *satisfy* (6.1.1) *and let* $\tilde{\lambda}_n^{SG}$ *and* $\hat{\lambda}_n^s$ *be defined in* (6.3.1) *and* (6.2.12), *respectively. If* $1 \leqslant p < \min(q, 2q-7)$ *and* $nb \to \infty$, *such that*

$$1/b = o\left(n^{1/3-1/q}\right), \qquad 1/b = o\left(n^{(q-3)/(6p)}\right)$$

and

$$1/b = o\left(n^{1/6+1/(6p)}(\log n)^{-(1/2+1/(2p))}\right),$$

then

$$n^{2/3}\left(\int_b^{1-b}\left|\tilde{\lambda}_n^{SG}(t) - \hat{\lambda}_n^s(t)\right|^p d\mu(t)\right)^{1/p} \xrightarrow{d} \alpha_0[D_{\mathbb{R}}Z](0),$$

where $Z(t) = W(t) - t^2$, *with* W *being a two-sided Brownian motion originating from zero, and*

$$\alpha_0 = \left(\int_0^1\left|\frac{c_1'(t)}{c_1(t)^2}\right|^p d\mu(t)\right)^{1/p}, \qquad c_1(t) = \left|\frac{\lambda'(t)}{2L'(t)^2}\right|^{1/3}.$$

Proof. We write

$$n^{2/3}\left(\int_b^{1-b}\left|\tilde{\lambda}_n^{SG}(t) - \hat{\lambda}_n^s(t)\right|^p d\mu(t)\right)^{1/p} = b^{-1}\left(\int_b^{1-b}|Y_n(t)|^p\, d\mu(t)\right)^{1/p},$$

where

$$Y_n(t) = bn^{2/3}\left(\int_{t-b}^{t+b} k_b(t-u)\, d(\hat{\Lambda}_n - \Lambda_n)(u)\right), \qquad t \in (b, 1-b). \tag{6.3.5}$$

We first show that

$$b^{-p}\int_b^{1-b}|Y_n(t)|^p\,d\mu(t)\xrightarrow{d}\alpha_0^p[D_{\mathbb{R}}Z](0)^p, \tag{6.3.6}$$

and then the result would follow from the continuous mapping theorem. Note that integration by parts yields

$$Y_n(t)=\frac{1}{b}\int_{-1}^{1}k'\left(\frac{t-v}{b}\right)A_n^E(v)\,dv.$$

The proof consists of several succeeding approximations of A_n^E. For details, see Lemmas 6.7.2 to 6.7.6. First we replace A_n^E in the previous integral by A_n^W. The approximation of $Y_n(t)$ by

$$Y_n^{(1)}(t)=\frac{1}{b}\int_{-1}^{1}k'\left(\frac{t-v}{b}\right)A_n^W(v)\,dv. \tag{6.3.7}$$

where A_n^W is defined in (6.3.3), is possible thanks to Assumption (A2). According to (6.1.14),

$$\begin{aligned}&\left|\int_b^{1-b}|Y_n(t)|^p\,d\mu(t)-\int_b^{1-b}|Y_n^{(1)}(t)|^p\,d\mu(t)\right|\\&\leqslant p2^{p-1}\int_b^{1-b}|Y_n(t)-Y_n^{(1)}(t)|^p\,d\mu(t)\\&\quad+p2^{p-1}\left(\int_b^{1-b}|Y_n(t)-Y_n^{(1)}(t)|^p\,d\mu(t)\right)^{\frac{1}{p}}\left(\int_b^{1-b}|Y_n^{(1)}(t)|^p\,d\mu(t)\right)^{1-\frac{1}{p}}.\end{aligned} \tag{6.3.8}$$

According to Lemma 6.7.2, $b^{-p}\int_b^{1-b}|Y_n(t)-Y_n^{(1)}(t)|^p\,d\mu(t)=o_P(1)$. Consequently, in view of (6.3.8), if we show that

$$b^{-p}\int_b^{1-b}|Y_n^{(1)}(t)|^p\,d\mu(t)\xrightarrow{d}\alpha_0{}^p[D_{\mathbb{R}}Z](0)^p, \tag{6.3.9}$$

then we obtain

$$b^{-p}\int_b^{1-b}|Y_n(t)|^p\,d\mu(t)=b^{-p}\int_b^{1-b}|Y_n^{(1)}(t)|^p\,d\mu(t)+o_P(1), \tag{6.3.10}$$

and (6.3.6) follows.

In order to prove (6.3.9), we replace A_n^W by $n^{2/3}D_{I_{nv}}\Lambda_n^W$, i.e., we approximate $Y_n^{(1)}$ by

$$Y_n^{(2)}(t)=\frac{1}{b}\int_{t-b}^{t+b}k'\left(\frac{t-v}{b}\right)n^{2/3}[D_{I_{nv}}\Lambda_n^W](v)\,dv. \tag{6.3.11}$$

where $I_{n\nu} = [0,1] \cap [\nu - n^{-1/3}\log n, \nu + n^{-1/3}\log n]$ and Λ^W is defined in (6.3.4). From Lemma 6.7.3, we have that

$$b^{-p}\int_b^{1-b} |Y_n^{(1)}(t) - Y_n^{(2)}(t)|^p \, d\mu(t) = o_P(1).$$

Hence, similar to the argument that leads to (6.3.10), if we show that

$$b^{-p}\int_b^{1-b} |Y_n^{(2)}(t)|^p \, d\mu(t) \xrightarrow{d} \alpha_0{}^p [D_{\mathbb{R}} Z](0)^p, \tag{6.3.12}$$

then, together with (6.1.14), it follows that

$$b^{-p}\int_b^{1-b} |Y_n^{(1)}(t)|^p \, d\mu(t) = b^{-p}\int_b^{1-b} |Y_n^{(2)}(t)|^p \, d\mu(t) + o_P(1).$$

Consequently, (6.3.9) is equivalent to (6.3.12).

In order to prove (6.3.12), let

$$Y_{n\nu}(s) = n^{1/6}\left[W_n(L(\nu + n^{-1/3}s)) - W_n(L(\nu))\right] + \frac{1}{2}\lambda'(\nu)s^2. \tag{6.3.13}$$

Let $H_{n\nu} = [-n^{1/3}\nu, n^{1/3}(1-\nu)] \cap [-\log n, \log n]$ and

$$\Delta_{n\nu} = n^{2/3}[D_{I_{n\nu}}\Lambda_n^W](\nu) - [D_{H_{n\nu}} Y_{n\nu}](0).$$

We approximate $Y_n^{(2)}$ by

$$Y_n^{(3)}(t) = \frac{1}{b}\int_{t-b}^{t+b} k'\left(\frac{t-\nu}{b}\right)[D_{H_{n\nu}} Y_{n\nu}](0)\, d\nu. \tag{6.3.14}$$

From Lemma 6.7.4, we have that

$$b^{-p}\int_b^{1-b} |Y_n^{(2)}(t) - Y_n^{(3)}(t)|^p \, d\mu(t) = o_P(1).$$

Again, similar to the argument that leads to (6.3.10), if we show that

$$b^{-p}\int_b^{1-b} |Y_n^{(3)}(t)|^p \, d\mu(t) \xrightarrow{d} \alpha_0{}^p [D_{\mathbb{R}} Z](0)^p. \tag{6.3.15}$$

then, together with (6.1.14), it follows that

$$b^{-p}\int_b^{1-b} |Y_n^{(2)}(t)|^p \, d\mu(t) = b^{-p}\int_b^{1-b} |Y_n^{(3)}(t)|^p \, d\mu(t) + o_P(1),$$

which would prove (6.3.12).

We proceed with proving (6.3.15). Let W be a two sided Brownian motion originating from zero. We have that

$$n^{1/6}\left[W_n(L(v+n^{-1/3}s))-W_n(L(v))\right] \stackrel{d}{=} W\left(n^{1/3}(L(v+n^{-1/3}s)-L(v))\right)$$

as a process in s. Consequently,

$$Y_n^{(3)}(t) \stackrel{d}{=} \frac{1}{b}\int_{t-b}^{t+b} k'\left(\frac{t-v}{b}\right)[D_{H_{nv}}\tilde{Y}_{nv}](0)\,dv$$

where

$$\tilde{Y}_{nv}(s) = W(n^{1/3}(L(v+n^{-1/3}s)-L(v)))+\frac{1}{2}\lambda'(v)s^2. \tag{6.3.16}$$

Now define

$$Z_{nv}(s) = W(L'(v)s)+\frac{1}{2}\lambda'(v)s^2. \tag{6.3.17}$$

and

$$J_{nv} = \left[n^{1/3}\frac{L(a_{nv})-L(v)}{L'(v)}, n^{1/3}\frac{L(b_{nv})-L(v)}{L'(v)}\right],$$

where $a_{nv} = \max(0, v-n^{-1/3}\log n)$ and $b_{nv} = \min(1, v+n^{-1/3}\log n)$. We approximate $\tilde{Y}_{nv}$ by Z_{nv}, i.e., we approximate $Y_n^{(3)}$ by

$$Y_n^{(4)}(t) = \frac{1}{b}\int_{t-b}^{t+b} k'\left(\frac{t-v}{b}\right)[D_{J_{nv}}Z_{nv}](0)\,dv, \tag{6.3.18}$$

Lemma 6.7.5 yields

$$b^{-p}\int_b^{1-b}|Y_n^{(3)}(t)-Y_n^{(4)}(t)|^p\,d\mu(t) = o_P(1).$$

Once more, similar to the argument that leads to (6.3.10), if we show that

$$b^{-p}\int_b^{1-b}|Y_n^{(4)}(t)|^p\,d\mu(t) \xrightarrow{d} \alpha_0{}^p[D_{\mathbb{R}}Z](0)^p, \tag{6.3.19}$$

then, together with (6.1.14), it follows that

$$b^{-p}\int_b^{1-b}|Y_n^{(3)}(t)|^p\,d\mu(t) = b^{-p}\int_b^{1-b}|Y_n^{(4)}(t)|^p\,d\mu(t)+o_P(1),$$

and as a result, also (6.3.15) holds.

As a final step, we prove (6.3.19). Since $c_1(v)W(L'(v)c_2(v)s) \stackrel{d}{=} W(s)$ as a process in s, where

$$c_1(v) = \left(\frac{|\lambda'(v)|}{2L'(v)^2}\right)^{1/3}, \qquad c_2(v) = \left(\frac{4L'(v)}{|\lambda'(v)|^2}\right)^{1/3}$$

we obtain that

$$Y_n^{(4)}(t) \stackrel{d}{=} \frac{1}{b}\int_{t-b}^{t+b} k'\left(\frac{t-v}{b}\right)\frac{1}{c_1(v)}[D_{I_{nv}}Z](0)\,dv$$

where $I_{nv} = c_2(v)^{-1}J_{nv}$ and $Z(t) = W(t) - t^2$. We approximate $D_{I_{nv}}$ by $D_{\mathbb{R}}$, i.e., we approximate $Y_n^{(4)}$ by

$$Y_n^{(5)}(t) = [D_{\mathbb{R}}Z](0)\frac{1}{b}\int_{t-b}^{t+b} k'\left(\frac{t-v}{b}\right)\frac{1}{c_1(v)}\,dv. \tag{6.3.20}$$

It remains to show that

$$b^{-p}\int_b^{1-b} |Y_n^{(5)}(t)|^p\,d\mu(t) \xrightarrow{d} \alpha_0^p[D_{\mathbb{R}}Z](0)^p, \tag{6.3.21}$$

because then, it follows that

$$b^{-p}\int_b^{1-b} |Y_n^{(4)}(t)|^p\,d\mu(t) = b^{-p}\int_b^{1-b} |Y_n^{(5)}(t)|^p\,d\mu(t) + o_P(1)$$

so that (6.3.19) holds. Since

$$\frac{1}{b}\int_{t-b}^{t+b} k'\left(\frac{t-v}{b}\right)\frac{1}{c_1(t)}\,dv = \frac{1}{c_1(t)}\int_{-1}^{1} k'(y)\,dy = 0.$$

we can write

$$\begin{aligned}\frac{1}{b}\int_{t-b}^{t+b} k'\left(\frac{t-v}{b}\right)\frac{1}{c_1(v)}\,dv &= \frac{1}{b}\int_{t-b}^{t+b} k'\left(\frac{t-v}{b}\right)\left(\frac{1}{c_1(v)} - \frac{1}{c_1(t)}\right)dv \\ &= \int_{-1}^{1} k'(y)\left(\frac{1}{c_1(t-by)} - \frac{1}{c_1(t)}\right)dy.\end{aligned}$$

Assumptions (A1) and (A2) imply that $t \mapsto c_1(t)$ is strictly positive and differentiable with bounded derivative, so by a Taylor expansion we get

$$\int_{-1}^{1} k'(y)\left(\frac{1}{c_1(t-by)} - \frac{1}{c_1(t)}\right)dy = \frac{c_1'(t)}{c_1(t)^2}b\int_{-1}^{1} k'(y)\,y\,dy + O(b^2).$$

Hence,

$$\begin{aligned}b^{-p}\int_b^{1-b} |Y_n^{(5)}(t)|^p\,d\mu(t) &= [D_{\mathbb{R}}Z](0)^p b^{-p}\int_b^{1-b}\left|\frac{c_1'(t)b}{c_1(t)^2}\right|^p d\mu(t) + o_P(1) \\ &= [D_{\mathbb{R}}Z](0)^p\int_0^1\left|\frac{c_1'(t)}{c_1(t)^2}\right|^p d\mu(t) + o_P(1)\end{aligned}$$

which concludes the proof of (6.3.21) and concludes the proof of the theorem. □

Remark 6.3.2. Note that the assumption

$$1/b = o\left(n^{1/6+1/(6p)}(\log n)^{-(1+1/p)}\right)$$

of the previous theorem puts a restriction on p, when b has the optimal rate $n^{-1/5}$. This is due to the approximation of $Y_n^{(4)}(t)$ by $Y_n^{(5)}(t)$ for $t \in (b, 1-b)$. This restriction on p can be avoided if we consider the L_p-error on the smaller interval $(b + n^{-1/3}\log n, 1 - b - n^{-1/3}\log n)$.

Remark 6.3.3. For $p > 1$, the boundary regions cannot be included in the CLT of Theorem 6.3.1. For example, for $t \in (0, b)$, it can be shown that there exists a universal constant $K > 0$, such that

$$n^{2p/3}\int_0^b \left|\tilde{\lambda}_n^{SG}(t) - \tilde{\lambda}_n^{s}(t)\right|^p \, d\mu(t) > Kb^{-p+1}[D_{\mathbb{R}}Z](0)^p + o_P(b^{-p+1}),$$

which is not bounded in probability for $p > 1$. For details see Appendix B The same result also holds for $t \in (1-b, 1)$.

In the special case $p = 1$, for $t \in (0, b)$ we have

$$\begin{aligned} &n^{2/3}\int_0^b \left|\tilde{\lambda}_n^{SG}(t) - \tilde{\lambda}_n^{s}(t)\right| \, d\mu(t) \\ &= [D_{\mathbb{R}}Z](0)\frac{1}{b}\int_0^b \left|\frac{1}{c_1(t)}\int_{-1}^{t/b}\frac{d}{dy}k^{(t)}(y)\, dy\right| d\mu(t) + o_P(1). \end{aligned}$$

If (A3) holds, then

$$\begin{aligned} &\frac{1}{b}\int_0^b \left|\frac{1}{c_1(t)}\int_{-1}^{t/b}\frac{d}{dy}k^{(t)}(y)\, dy\right| d\mu(t) \\ &\to \frac{w(0)}{c_1(0)}\int_0^1 |\psi_1(y)\,k(y) + \psi_2(y)\,yk(y)|\, dy. \end{aligned}$$

Similarly, we can deal with the case $t \in (1-b, 1)$. It follows that

$$n^{2/3}\int_0^1 \left|\tilde{\lambda}_n^{SG}(t) - \tilde{\lambda}_n^{s}(t)\right| \, d\mu(t) \xrightarrow{d} \tilde{\alpha}_0[D_{\mathbb{R}}Z](0)$$

with

$$\tilde{\alpha}_0 = \alpha_0 + \left(\frac{w(0)}{c_1(0)} + \frac{w(1)}{c_1(1)}\right)\int_0^1 |\psi_1(y)\,k(y) + \psi_2(y)\,yk(y)|\, dy.$$

We are now ready to formulate the CLT for the smoothed Grenander-type estimator. The result will follow from combining Corollary 6.2.7 with

Theorem 6.3.1. Because we now deal with the L_p-error between $\tilde{\lambda}_n^{SG}$ and λ, the contribution of the integrals over the boundary regions $(0, 2b)$ and $(1-2b, 1)$ can be shown to be negligible. This means we no longer need the third requirement in Theorem 6.3.1 on the rate of $1/b$.

Theorem 6.3.4. *Assume that* $(A1)-(A3)$ *hold and let* k *satisfy* (6.1.1). *Let* I_n^{SG} *be defined in* (6.3.2). *If* $1 \leqslant p < \min(q, 2q-7)$ *and* $nb \to \infty$, *such that*

$$1/b = o\left(n^{1/3-1/q}\right) \quad \textit{and} \quad 1/b = o\left(n^{(q-3)/(6p)}\right).$$

i) If $nb^5 \to 0$, *then*

$$(b\sigma^2(p))^{-1/2}\left\{(nb)^{p/2} I_n^{SG}(p) - m_n(p)\right\} \xrightarrow{d} N(0,1);$$

ii) If $nb^5 \to C_0^2 > 0$, *and* B_n *in assumption (A2) is a Brownian motion, then*

$$(b\theta^2(p))^{-1/2}\left\{(nb)^{p/2} I_n^{SG}(p) - m_n(p)\right\} \xrightarrow{d} N(0,1);$$

iii) If $nb^5 \to C_0^2 > 0$, *and* B_n *in assumption (A2) is a Brownian bridge, then*

$$(b\tilde{\theta}^2(p))^{-1/2}\left\{(nb)^{p/2} I_n^{SG}(p) - m_n(p)\right\} \xrightarrow{d} N(0,1),$$

where I_n^{SG}, m_n, σ^2, θ^2, *and* $\tilde{\theta}^2$ *are defined in* (6.3.2), (6.1.8), (6.1.9), (6.1.11), *and* (6.1.12), *respectively.*

Proof. Define

$$\gamma^2(p) = \begin{cases} \sigma^2(p) & \text{if} \quad nb^5 \to 0 \\ \theta^2(p) & \text{if} \quad nb^5 \to C_0^2. \end{cases} \tag{6.3.22}$$

By Corollary 6.2.7, we already have that

$$(b\gamma^2(p))^{-1/2}\left\{(nb)^{p/2}\int_0^1 \left|\hat{\lambda}_n^s(t) - \lambda(t)\right|^p \, d\mu(t) - m_n(p)\right\} \xrightarrow{d} N(0,1),$$

for $\hat{\lambda}_n^s$ defined in (6.2.12). Hence it is sufficient to show that

$$b^{-1/2}(nb)^{p/2}\left|\int_0^1 \left|\tilde{\lambda}_n^{SG}(t) - \lambda(t)\right|^p \, d\mu(t) - \int_0^1 \left|\hat{\lambda}_n^s(t) - \lambda(t)\right|^p \, d\mu(t)\right| \xrightarrow{\mathbb{P}} 0,$$

in all three cases (i)-(iii). First we show that

$$b^{-1/2}(nb)^{p/2}\left|\int_0^{2b} \left|\tilde{\lambda}_n^{SG}(t) - \lambda(t)\right|^p \, d\mu(t) - \int_0^{2b} \left|\hat{\lambda}_n^s(t) - \lambda(t)\right|^p \, d\mu(t)\right| \xrightarrow{\mathbb{P}} 0. \tag{6.3.23}$$

Indeed, by (6.1.14), we get

$$\begin{aligned}
&\left|\int_0^{2b} \left|\tilde{\lambda}_n^{SG}(t)-\lambda(t)\right|^p \, d\mu(t) - \int_0^{2b} \left|\hat{\lambda}_n^{s}(t)-\lambda(t)\right|^p \, d\mu(t)\right| \\
&\leqslant p2^{p-1} \int_0^{2b} \left|\tilde{\lambda}_n^{SG}(t)-\hat{\lambda}_n^{s}(t)\right|^p \, d\mu(t) \\
&+ p2^{p-1} \left(\int_0^{2b} \left|\tilde{\lambda}_n^{SG}(t)-\hat{\lambda}_n^{s}(t)\right|^p \, d\mu(t)\right)^{\frac{1}{p}} \left(\int_0^{2b} \left|\hat{\lambda}_n^{s}(t)-\lambda(t)\right|^p \, d\mu(t)\right)^{1-\frac{1}{p}}.
\end{aligned}$$

Moreover, by integration by parts and the Kiefer-Wolfowitz type of result in Corollary 3.1 in Durot and Lopuhaä, 2014, it follows that

$$\begin{aligned}
\sup_{t\in[0,1]} \left|\tilde{\lambda}_n^{SG}(t)-\hat{\lambda}_n^{s}(t)\right| &= \sup_{t\in[0,1]} \left|\int k_b^{(t)}(t-u)\, d(\hat{\Lambda}_n-\Lambda_n)(u)\right| \\
&\leqslant Cb^{-1} \sup_{t\in[0,1]} |\hat{\Lambda}_n(t)-\Lambda_n(t)| \\
&= O_P\left(b^{-1}\left(\frac{\log n}{n}\right)^{2/3}\right).
\end{aligned} \tag{6.3.24}$$

Hence

$$\int_0^{2b} \left|\tilde{\lambda}_n^{SG}(t)-\hat{\lambda}_n^{s}(t)\right|^p \, d\mu(t) = O_P\left(b^{1-p}\left(\frac{\log n}{n}\right)^{2p/3}\right). \tag{6.3.25}$$

Together with Proposition 6.2.6 this implies (6.3.23). Similarly, we also have

$$b^{-\frac{1}{2}}(nb)^{\frac{p}{2}} \left|\int_{1-2b}^{1} \left|\tilde{\lambda}_n^{SG}(t)-\lambda(t)\right|^p \, d\mu(t) - \int_{1-2b}^{1} \left|\hat{\lambda}_n^{s}(t)-\lambda(t)\right|^p \, d\mu(t)\right| \xrightarrow{\mathbb{P}} 0.$$

Thus, it remains to prove

$$b^{-\frac{1}{2}}(nb)^{\frac{p}{2}} \left|\int_{2b}^{1-2b} \left|\tilde{\lambda}_n^{SG}(t)-\lambda(t)\right|^p \, d\mu(t) - \int_{2b}^{1-2b} \left|\hat{\lambda}_n^{s}(t)-\lambda(t)\right|^p \, d\mu(t)\right| \xrightarrow{\mathbb{P}} 0. \tag{6.3.26}$$

Again, from (6.1.14), we have

$$\begin{aligned}
&\left|\int_{2b}^{1-2b}\left|\tilde{\lambda}_n^{SG}(t)-\lambda(t)\right|^p\,d\mu(t)-\int_{2b}^{1-2b}\left|\hat{\lambda}_n^{s}(t)-\lambda(t)\right|^p\,d\mu(t)\right| \\
&\leqslant p2^{p-1}\int_{2b}^{1-2b}\left|\tilde{\lambda}_n^{SG}(t)-\hat{\lambda}_n^{s}(t)\right|^p\,d\mu(t) \\
&\quad+p2^{p-1}\left(\int_{2b}^{1-2b}\left|\tilde{\lambda}_n^{SG}(t)-\hat{\lambda}_n^{s}(t)\right|^p\,d\mu(t)\right)^{1/p} \\
&\qquad\qquad\left(\int_{2b}^{1-2b}\left|\hat{\lambda}_n^{s}(t)-\lambda(t)\right|^p\,d\mu(t)\right)^{1-1/p}.
\end{aligned} \tag{6.3.27}$$

Because $b^{-1}=o(n^{1/3-1/q})$ implies that

$$(2b,1-2b)\subset(b+n^{-1/3}\log n,1-b-n^{-1/3}\log n),$$

from Theorem 6.3.1, in particular Remark 6.3.2, we have

$$\int_{2b}^{1-2b}\left|\tilde{\lambda}_n^{SG}(t)-\hat{\lambda}_n^{s}(t)\right|^p\,d\mu(t)=O_P(n^{-2p/3})=o_P(n^{-p/2}). \tag{6.3.28}$$

Then, (6.3.26) follows immediately from (6.3.27) and the fact that, according to Theorem 6.2.1,

$$\int_{2b}^{1-2b}\left|\hat{\lambda}_n^{s}(t)-\lambda(t)\right|^p\,d\mu(t)=O_P((nb)^{-p/2}).$$

This proves the theorem. □

Remark 6.3.5. Note that, if $b=cn^{-\alpha}$, for some $0<\alpha<1$, the proof is simple and short in case $\alpha<p/(3(1+p))$ because the Kiefer-Wolfowitz type of result in Corollary 3.1 in Durot and Lopuhaä, 2014 is sufficient to prove (6.3.28). Indeed, from (6.3.24), it follows that

$$\begin{aligned}
\int_{2b}^{1-2b}\left|\tilde{\lambda}_n^{SG}(t)-\hat{\lambda}_n^{s}(t)\right|^p\,d\mu(t)&=O_P\left(b^{-p}\left(\frac{\log n}{n}\right)^{2p/3}\right)\\
&=o_P\left(b^{1/2}(nb)^{-p/2}\right).
\end{aligned}$$

However, this assumption on α is quite restrictive because for example if $\alpha=1/5$ then the theorem holds only for $p>3/2$ (not for the L_1-loss) and if $\alpha=1/4$ then the theorem holds only for $p>3$.

6.4 ISOTONIZED KERNEL ESTIMATOR

The isotonized kernel estimator is defined as follows. First, we smooth the piecewise constant estimator Λ_n by means of a boundary corrected kernel function, i.e., let

$$\Lambda_n^s(t) = \int_{(t-b)\vee 0}^{(t+b)\wedge 1} k_b^{(t)}(t-u)\Lambda_n(u)\,du, \quad \text{for } t \in [0,1], \tag{6.4.1}$$

where $k^{(t)}(u)$ defined as in (1.2.11). Next, we define a continuous monotone estimator $\tilde{\lambda}_n^{GS}$ of λ as the left-hand slope of the least concave majorant $\widehat{\Lambda}_n^s$ of Λ_n^s on $[0,1]$. In this way we define a sort of Grenander estimator based on a smoothed naive estimator for Λ. For this reason we use the superscript GS.

We are interested in the asymptotic distribution of the L_p-error of this estimator:

$$I_n^{GS}(p) = \int_0^1 \left|\tilde{\lambda}_n^{GS}(t) - \lambda(t)\right|^p \, d\mu(t).$$

It follows from Lemma 1 in Groeneboom and Jongbloed, 2010 (in the case of a decreasing function), that $\tilde{\lambda}_n^{GS}$ is continuous and is the unique minimizer of

$$\psi(\lambda) = \frac{1}{2}\int_0^1 \left(\lambda(t) - \tilde{\lambda}_n^s(t)\right)^2 dt$$

over all nonincreasing functions λ, where $\tilde{\lambda}_n^s(t) = d\Lambda_n^s(t)/dt$. This suggests $\tilde{\lambda}_n^s(t)$ as a naive estimator for $\lambda_0(t)$. Note that, for $t \in [b, 1-b]$, from integration by parts we get

$$\tilde{\lambda}_n^s(t) = \frac{1}{b^2}\int_{t-b}^{t+b} k'\left(\frac{t-u}{b}\right)\Lambda_n(u)\,du = \int_{t-b}^{t+b} k_b(t-u)\,d\Lambda_n(u), \tag{6.4.2}$$

i.e., $\tilde{\lambda}_n^s$ coincides with the usual kernel estimator of λ on $[b, 1-b]$.

Let $0 < \gamma < 1$. It can be shown that

$$\mathbb{P}(\tilde{\lambda}_n^s(t) = \tilde{\lambda}_n^{GS}(t) \text{ for all } t \in [b^\gamma, 1-b^\gamma]) \to 1. \tag{6.4.3}$$

See Corollary B.2.2 in Appendix B. Hence, their L_p-error between $\tilde{\lambda}_n^{GS}$ and $\tilde{\lambda}_n^s$ will exhibit the same behavior in the limit. Note that this holds for every $\gamma < 1$, which means that the interval we are considering is approaching $(b, 1-b)$. Consider a modified L_p-error of the isotonized kernel estimator defined by

$$I_{n,\gamma}^{GS,c}(p) = \int_{b^\gamma}^{1-b^\gamma} \left|\tilde{\lambda}_n^{GS}(t) - \lambda(t)\right|^p \, d\mu(t). \tag{6.4.4}$$

We then have the following result.

Theorem 6.4.1. *Assume that (A1)-(A3) hold and let* $I_{n,\gamma}^{GS,c}(p)$ *be defined in* (6.4.4). *Let* k *satisfy* (6.1.1) *and let* L *be as in Assumption (A2). Assume* $b \to 0$ *and* $1/b = o(n^{1/4})$ *and let* $1/2 < \gamma < 1$.

i) If $nb^5 \to 0$, *then*

$$(b\sigma^2(p))^{-1/2}\left\{(nb)^{p/2} I_{n,\gamma}^{GS,c}(p) - m_n(p)\right\} \xrightarrow{d} N(0,1);$$

ii) If $nb^5 \to C_0^2 > 0$ *and* B_n *in assumption (A2) is a Brownian motion, then*

$$(b\theta^2(p))^{-1/2}\left\{(nb)^{p/2} I_{n,\gamma}^{GS,c}(p) - m_n(p)\right\} \xrightarrow{d} N(0,1);$$

iii) If $nb^5 \to C_0^2 > 0$ *and* B_n *in assumption (A2) is a Brownian bridge, then*

$$(b\tilde{\theta}^2(p))^{-1/2}\left\{(nb)^{p/2} I_{n,\gamma}^{GS,c}(p) - m_n(p)\right\} \xrightarrow{d} N(0,1),$$

where σ^2, θ^2, $\tilde{\theta}^2$ *and* m_n *are defined respectively in* (6.1.9), (6.1.11), (6.1.12) *and* (6.1.8).

Proof. It follows from Theorem 6.2.1 and (6.4.3). Note that the results of Theorem 6.2.1 do not change if we consider the interval $[b^\gamma, 1-b^\gamma]$ instead of $[b, 1-b]$ and that $b^{-1/2}|m_n^c(p) - m_n(p)| \to 0$. □

6.5 HELLINGER ERROR

In this section we investigate the global behavior of estimators by means of a weighted Hellinger distance

$$H(\hat{\lambda}_n, \lambda) = \left(\frac{1}{2}\int_0^1 \left(\sqrt{\hat{\lambda}_n(t)} - \sqrt{\lambda(t)}\right)^2 d\mu(t)\right)^{1/2}, \tag{6.5.1}$$

where $\hat{\lambda}_n$ is the estimator at hand. This metric is convenient in maximum likelihood problems, which goes back to Birgé and Massart, 1993; Le Cam, 1973; Le Cam, 1970. Consistency in Hellinger distance of shape constrained maximum likelihood estimators has been investigated in Pal, Woodroofe, and Meyer, 2007, Seregin and Wellner, 2010, and Doss and Wellner, 2016, whereas rates on Hellinger risk measures have been obtained in Seregin and Wellner, 2010, Kim and Samworth, 2016, and Kim, Guntuboyina, and Samworth, 2016. The first central limit theorem type of result for the Hellinger distance was presented in Chapter 4 for Grenander type estimators of a

monotone function. We deal with the smooth (isotonic) estimators following the same approach.

Note that, for the Hellinger distance to be well defined we need to assume that λ takes only positive values. We follow the same line of argument as in Chapter 4. We first establish that

$$\int_0^1 \left(\sqrt{\hat{\lambda}_n^s(t)} - \sqrt{\lambda(t)}\right)^2 d\mu(t) = \int_0^1 \left(\hat{\lambda}_n^s(t) - \lambda(t)\right)^2 (4\lambda(t))^{-1} d\mu(t) + O_P\left((nb)^{-3/2}\right),$$

which shows that the squared Hellinger loss can be approximated by a weighted squared L_2-distance. For details, see Lemma B.3.1 in Appendix B, which is the corresponding version of Lemma 4.1.1. Hence, a central limit theorem for squared the Hellinger loss follows directly from the central limit theorem for the weighted L_2-distance (see Theorem B.3.2, which corresponds to Theorem 4.2.1. An application of the delta method will then lead to the following result.

Theorem 6.5.1. *Assume (A1)-(A3) hold. Let $\tilde{\lambda}_n^s$ be defined in* (6.1.2)*, with k satisfying* (6.1.1)*, and let H be defined in* (6.5.1)*. Suppose that $nb \to \infty$ and that λ is strictly positive.*

i) If $nb^5 \to 0$, then

$$\left(b\frac{\tau^2(2)}{8\mu_n(2)}\right)^{-1/2} \left\{(nb)^{1/2} H(\hat{\lambda}_n^s, \lambda) - 2^{-1/2}\mu_n(2)^{1/2}\right\} \xrightarrow{d} N(0,1).$$

ii) If $nb^5 \to C_0^2 > 0$ and B_n in Assumption (A2) is a Brownian motion, then

$$\left(b\frac{\kappa^2(2)}{8\mu_n(2)}\right)^{-1/2} \left\{(nb)^{1/2} H(\hat{\lambda}_n^s, \lambda) - 2^{-1/2}\mu_n(2)^{1/2}\right\} \xrightarrow{d} N(0,1),$$

iii) If $nb^5 \to C_0^2 > 0$ and B_n in Assumption (A2) is a Brownian bridge, then

$$\left(b\frac{\tilde{\kappa}^2(2)}{8\mu_n(2)}\right)^{-1/2} \left\{(nb)^{1/2} H(\hat{\lambda}_n^s, \lambda) - 2^{-1/2}\mu_n(2)^{1/2}\right\} \xrightarrow{d} N(0,1),$$

where τ^2, κ^2, $\tilde{\kappa}^2$ and μ_n are defined as in (6.1.9), (6.1.11), (6.1.12) *and* (6.1.8)*, respectively, by replacing $w(t)$ with $w(t)(4\lambda(t))^{-1}$.*

(iv) Under the conditions of Theorem 6.3.4, results (i)-(iii) also hold when replacing $\hat{\lambda}_n^s$ by the smoothed Grenander-type estimator $\tilde{\lambda}_n^{SG}$, defined in (6.3.1)*.*

Proof. The proof consists of an application of the delta-method in combination with Theorem B.3.2. According to part (i) of Theorem B.3.2,

$$b^{-1/2}\left(2nbH(\hat{\lambda}_n^s, \lambda) - \mu_n(2)\right) \xrightarrow{d} Z$$

where Z is a mean zero normal random variable with variance $\tau^2(2)$. Therefore, in order to obtain part (i) of Theorem 6.5.1, we apply the delta method with the mapping $\phi(x) = 2^{-1/2}x^{1/2}$. Parts (ii)-(iv) are obtained in the same way. □

To be complete, note that from Corollary B.2.2, the previous central limit theorems also hold for the isotonized kernel estimator $\tilde{\lambda}_n^{GS}$, defined in Section 6.4, when considering a Hellinger distance corresponding to the interval $(b^\gamma, 1 - b^\gamma)$ instead of $(0, 1)$ in (6.5.1).

6.6 TESTING MONOTONICITY

In this section we investigate a possible application of the results obtained in Section 6.3 for testing monotonicity. For example, Theorem 6.3.4 could be used to construct a test for the single null hypothesis $H_0 : \lambda = \lambda_0$, for some known monotone function λ_0. Instead, we investigate a nonparametric test for monotonicity on the basis of the L_p-distance between the smoothed Grenander-type estimator and the kernel estimator, see Theorem 6.3.1.

The problem of testing a nonparametric null hypothesis of monotonicity has gained a lot of interest in the literature (see for example Kulikov and Lopuhaä, 2004 for the density setting, Hall and Van Keilegom, 2005, Groeneboom and Jongbloed, 2012 for the hazard rate, Akakpo, Balabdaoui, and Durot, 2014, Birke and Dette, 2007, Birke and Neumeyer, 2013, Gijbels et al., 2000 for the regression function).

We consider a regression model with deterministic design points

$$Y_i = \lambda\left(\frac{i}{n}\right) + \epsilon_i, \qquad i \in \{1, \ldots, n\}, \tag{6.6.1}$$

where the ϵ_i's are independent normal random variables with mean zero and variance σ^2. Such a model satisfies Assumption (A2) with $q = +\infty$ and $\Lambda_n(t) = n^{-1}\sum_{i \leqslant nt} Y_i$, for $t \in [0, 1]$ (see Theorem 5 in Durot, 2007).

Assume we have a sample of n observations $Y_1, \ldots, Y_n$. Let $\mathcal{D}$ be the space of decreasing functions on $[0, 1]$. We want to test $H_0 \colon \lambda \in \mathcal{D}$ against $H_1 \colon \lambda \notin \mathcal{D}$. Under the null hypothesis we can estimate λ by the smoothed

Grenander-type estimator $\tilde{\lambda}_n^{SG}$ defined as in (6.3.1). On the other hand, under the alternative hypothesis we can estimate λ by the kernel estimator with boundary corrections $\hat{\lambda}_n^s$ defined in (6.2.12). Then, as a test statistic we take

$$T_n = n^{2/3} \left(\int_b^{1-b} \left| \tilde{\lambda}_n^{SG}(t) - \hat{\lambda}_n^s(t) \right|^2 dt \right)^{1/2},$$

and at level α, we reject the null hypothesis if $T_n > c_{n,\alpha}$ for some critical value $c_{n,\alpha} > 0$.

In order to use the asymptotic quantiles of the limit distribution in Theorem 6.3.1, we need to estimate the constant C_0 which depends on the derivatives of λ. To avoid this, we choose to determine the critical value by a bootstrap procedure. We generate $B = 1000$ samples of size n from the model (6.6.1) with λ replaced by its estimator $\tilde{\lambda}_n^{SG}$ under the null hypothesis. For each of these samples we compute the estimators $\tilde{\lambda}_n^{SG,*}$, $\hat{\lambda}_n^{s,*}$ and the test statistic

$$T_{n,j}^* = n^{2/3} \left(\int_b^{1-b} \left| \tilde{\lambda}_n^{SG,*}(t) - \hat{\lambda}_n^{s,*}(t) \right|^2 dt \right)^{1/2}, \qquad j = 1, \dots, B.$$

Then we take as a critical value, the 100α-th upper-percentile of the values $T_{n,1}^*, \dots, T_{n,B}^*$. We repeat this procedure $N = 1000$ times and we count the percentage of rejections. This gives an approximation of the level (or the power) of the test if we start with a sample for which the true λ is decreasing (or non-decreasing).

We investigate the performance of the test by comparing it to tests proposed in Akakpo, Balabdaoui, and Durot, 2014, Baraud, Huet, and Laurent, 2005 and in Gijbels et al., 2000. For a power comparison, Akakpo, Balabdaoui, and Durot, 2014 and Baraud, Huet, and Laurent, 2005 consider the following functions

$$\begin{aligned}
&\lambda_1(x) = -15(x-0.5)^3 \mathbb{1}_{\{x \leqslant 0.5\}} - 0.3(x-0.5) + \exp\left(-250(x-0.25)^2\right), \\
&\lambda_2(x) = 16\sigma x, \quad \lambda_3(x) = 0.2\exp\left(-50(x-0.5)^2\right), \quad \lambda_4(x) = -0.1\cos(6\pi x), \\
&\lambda_5(x) = -0.2x + \lambda_3(x), \quad \lambda_6(x) = -0.2x + \lambda_4(x), \\
&\lambda_7(x) = -(1+x) + 0.45\exp\left(-50(x-0.5)^2\right),
\end{aligned}$$

We denote by T_B the local mean test of Baraud, Huet, and Laurent, 2005 and S_n^{reg} the test proposed in Akakpo, Balabdaoui, and Durot, 2014 on the basis of the distance between the least concave majorant of Λ_n and Λ_n. The result of the simulations for $n = 100$, $\alpha = 0.05$, $b = 0.1$, are given in Table 12.

Function	λ_1	λ_2	λ_3	λ_4	λ_5	λ_6	λ_7
σ^2	0.01	0.01	0.01	0.01	0.004	0.006	0.01
T_n	1	1	1	1	1	1	0.99
T_B	0.99	0.99	1	0.99	0.99	0.98	0.76
S_n^{reg}	0.99	1	0.98	0.99	0.99	0.99	0.68

Table 12: Simulated power of T_n, T_B and S_n^{reg} for $n = 100$.

We see that, apart from the last case, all the three tests perform very well and they are comparable. However, our test behaves much better for the function λ_7, which is more difficult to detect than the others.

The second model that we consider is taken from Akakpo, Balabdaoui, and Durot, 2014 and Gijbels et al., 2000, which is a regression function given by

$$\lambda_a(x) = -(1+x) + a\exp\left(-50(x-0.5)^2\right), \qquad x \in [0,1].$$

The results of the simulation, again for $n = 100$, $\alpha = 0.05$, $b = 0.1$ and various values of a and σ^2 are given in Table 13. We denote by S_n^{reg} the test of Akakpo, Balabdaoui, and Durot, 2014 and by T_{run} the test of Gijbels et al., 2000. Note that when $a = 0$, the regression function is decreasing so H_0

	$a = 0$			$a = 0.25$			$a = 0.45$		
σ	0.025	0.05	0.1	0.025	0.05	0.1	0.025	0.05	0.1
T_n	0.012	0.025	0.022	0.927	0.497	0.219	1	1	0.992
T_{run}	0	0	0	0.106	0.037	0.014	1	1	0.805
S_n^{reg}	0	0.002	0.013	0.404	0.053	0.007	1	1	0.683

Table 13: Simulated power of T_n, T_{run} and S_n^{reg} for $n = 100$.

is satisfied. We observe that our test rejects the null hypothesis more often than T_{run} and S_n^{reg} so the price we pay for getting higher power is higher level. As the value of a increases, the monotonicity of λ_a is perturbed. For $a = 0.25$ our test performs significantly better than the other two and, as expected, the power decreases as the variance of the errors increases. When $a = 0.45$ and σ^2 not too large, the three tests have power one but, when σ^2 increases, T_n outperforms T_{run} and S_n^{reg}.

We note that the test performs the same way if, instead of the L_2-distance between $\tilde{\lambda}_n^{SG}$ and $\hat{\lambda}_n^s$, we use the L_1-distance on $(0,1)$. Indeed, in Remark 6.3.3 we showed that , for $p = 1$, the limit theorem holds on the whole interval

$(0,1)$. Moreover, we did not investigate the choice of the bandwidth. We take $b = 0.1$, which seems to be a reasonable one considering that the whole interval has length one.

6.7 AUXILIARY RESULTS AND PROOFS

6.7.1 *Proofs for Section 6.2*

Lemma 6.7.1. *Let* $L : [0,1] \to \mathbb{R}$ *be strictly positive and twice differentiable, such that* $\inf_{t\in[0,1]} L'(t) > 0$ *and* $\sup_{t\in[0,1]} |L''(t)| < \infty$. *Let* $\Gamma_n^{(2)}$, $g_{(n)}$, *and* $m_n^c(p)$ *be defined in* (6.2.5), (6.1.4), *and* (6.1.8), *respectively. Assume that (A1) and (A3) hold.*

1. *If* $nb^5 \to 0$, *then*

$$(b\sigma^2(p))^{-\frac{1}{2}} \left\{ \int_b^{1-b} \left| b^{-\frac{1}{2}} \Gamma_n^{(2)}(t) + g_{(n)}(t) \right|^p \, d\mu(t) - m_n^c(p) \right\} \xrightarrow{d} N(0,1),$$

   where $\sigma^2(p)$ *is defined in* (6.1.9).

2. *If* $nb^5 \to C_0^2$, *then*

$$(b\tilde{\theta}^2(p))^{-\frac{1}{2}} \left\{ \int_b^{1-b} \left| b^{-\frac{1}{2}} \Gamma_n^{(2)}(t) + g_{(n)}(t) \right|^p \, d\mu(t) - m_n^c(p) \right\} \xrightarrow{d} N(0,1),$$

   where $\tilde{\theta}^2(p)$ *is defined in* (6.1.12).

Proof. From the properties of the kernel function and L we have

$$\begin{aligned}
\Gamma_n^{(2)}(t) &= \int k\left(\frac{t-u}{b}\right) dW(L(u)) - \frac{W(L(1))}{L(1)} \int k\left(\frac{t-u}{b}\right) L'(u)\, du \\
&= \int k\left(\frac{t-u}{b}\right) dW(L(u)) - b\frac{W(L(1))}{L(1)} L'(t) + O_P(b^3),
\end{aligned}$$

where the O_P term is uniformly for $t \in [0,1]$. Hence, (6.1.14) implies that

$$\begin{aligned}
&\int_b^{1-b} \left| b^{-1/2} \Gamma_n^{(2)}(t) + g_{(n)}(t) \right|^p d\mu(t) = \\
&\int_b^{1-b} \left| b^{-\frac{1}{2}} \int k\left(\frac{t-u}{b}\right) dW(L(u)) + g_{(n)}(t) - b^{\frac{1}{2}} \frac{W(L(1))}{L(1)} L'(t) \right|^p d\mu(t) \\
&\quad + O(b^3).
\end{aligned}$$

Therefore, it is sufficient to prove a CLT for

$$\int_b^{1-b} \left| b^{-1/2} \int k\left(\frac{t-u}{b}\right) dW(L(u)) + g_{(n)}(t) - b^{1/2} \frac{W(L(1))}{L(1)} L'(t) \right|^p d\mu(t). \tag{6.7.1}$$

Let

$$X_{n,t} = b^{-1/2} \int k\left(\frac{t-u}{b}\right) dW(L(u)) + g_{(n)}(t). \tag{6.7.2}$$

Then $X_{nt} \sim N(g_{(n)}(t), \sigma_n^2(t))$, where

$$\sigma_n^2(t) = \frac{1}{b} \int k^2\left(\frac{t-u}{b}\right) L'(u)\, du. \tag{6.7.3}$$

We can then write

$$\begin{aligned}
&b^{-1/2} \left\{ \int_b^{1-b} \left| b^{-1/2} \Gamma_n^{(2)}(t) + g_{(n)}(t) \right|^p d\mu(t) - m_n^c(p) \right\} \\
&= b^{-1/2} \left\{ \int_b^{1-b} \left| X_{n,t} - b^{1/2} \frac{W(L(1))}{L(1)} L'(t) \right|^p d\mu(t) - m_n^c(p) \right\} + o(1) \\
&= b^{-1/2} \left\{ \int_b^{1-b} |X_{n,t}|^p \, d\mu(t) - m_n^c(p) \right\} \\
&\qquad - p \frac{W(L(1))}{L(1)} \int_b^{1-b} |X_{n,t}|^{p-1} \operatorname{sgn}\{X_{n,t}\} L'(t) w(t)\, dt \\
&\qquad + b^{-1/2} \int_b^{1-b} O\left(b W(L(1))^2 \right) dt + o(1),
\end{aligned} \tag{6.7.4}$$

where we use

$$|x|^p = |y|^p + p(x-y)|y|^{p-1} \operatorname{sgn}(y) + O((x-y)^2) \tag{6.7.5}$$

for the first term in the integrand on the right hand side of the first equality in (6.7.4). The third term on the right hand side of (6.7.4) converges to zero in probability, so it suffices to deal with the first two terms. To establish a central limit theorem for the first term, one can mimic the approach in Csörgő and Horváth, 1988 using a big-blocks-small-blocks procedure. See Lemmas B.1.1 and B.1.2 for details. It can be shown that

$$b^{-1/2} \left\{ \int_b^{1-b} |X_{n,t}|^p \, d\mu(t) - m_n^c(p) \right\} = b^{1/2} \sum_{i=1}^{M_3} \zeta_i + o_P(1),$$

where $\zeta_i = \sum_{j=c_i}^{d_i} \xi_j$, with

$$c_i = (i-1)(M_2+2)+1 \quad \text{and} \quad d_i = (i-1)(M_2+2)+M_2,$$

for some $0 < \nu < 1$, $M_2 = [(M_1 - 1)^\nu]$,

$$M_1 = \left[\frac{1}{b} - 1\right] \quad \text{and} \quad M_3 = \left[\frac{M_1 - 1}{M_2 + 2}\right],$$

and

$$\xi_i = b^{-1} \int_{ib}^{ib+b} \left\{ |X_{n,t}|^p - \int_{-\infty}^{+\infty} \left| \sqrt{L'(t)} Dx + g_{(n)}(t) \right|^p \phi(x)\, dx \right\} w(t)\, dt.$$

The random variables ζ_i are independent and satisfy

$$b^{1/2} \sum_{i=1}^{M_3} \zeta_i \xrightarrow{d} N(0, \gamma^2(p)), \tag{6.7.6}$$

where $\gamma^2(p)$ is defined in (6.3.22).

Next, consider the second term in the right hand side of (6.7.4). We have

$$\begin{aligned}
&\mathbb{E}\left[\int_b^{1-b} |X_{n,t}|^{p-1} \operatorname{sgn}\{X_{n,t}\} L'(t) w(t)\, dt\right] \\
&= \int_b^{1-b} \int_{\mathbb{R}} \left|\sigma_n(t)x + g_{(n)}(t)\right|^{p-1} \operatorname{sgn}\left\{\sigma_n(t)x + g_{(n)}(t)\right\} \phi(x) dx L'(t) w(t) dt \\
&\to \int_0^1 \int_{\mathbb{R}} \left|\sqrt{L'(t)} Dx + g(t)\right|^{p-1} \operatorname{sgn}\left\{\sqrt{L'(t)} Dx + g(t)\right\} \phi(x)\, dx\, L'(t) w(t)\, dt,
\end{aligned}$$

where D and $\sigma_n(t)$ are defined in (6.1.7) and (6.7.3), respectively, and ϕ denotes the standard normal density. Note that

$$\frac{d}{dx} \left|\sqrt{L'(t)} Dx + g(t)\right|^p = p \left|\sqrt{L'(t)} Dx + g(t)\right|^{p-1} \operatorname{sgn}\left\{\sqrt{L'(t)} Dx + g(t)\right\}.$$

Hence, integration by parts gives

$$\begin{aligned}
&\int_0^1 \int_{\mathbb{R}} \left|\sqrt{L'(t)} Dx + g(t)\right|^{p-1} \operatorname{sgn}\left\{\sqrt{L'(t)} Dx + g(t)\right\} \phi(x)\, dx\, L'(t) w(t)\, dt \\
&\qquad = \frac{\theta_1(p)}{Dp},
\end{aligned}$$

where θ_1 is defined in (6.1.13). We conclude

$$\mathbb{E}\left[\int_b^{1-b} |X_{n,t}|^{p-1} \operatorname{sgn}\{X_{n,t}\} L'(t) w(t)\, dt\right] \to \frac{\theta_1(p)}{Dp}.$$

Moreover,

$$\begin{aligned}
&\operatorname{Var}\left(\int_b^{1-b} |X_{n,t}|^{p-1} \operatorname{sgn}\{X_{n,t}\} L'(t) w(t)\, dt\right) \\
&= \int_b^{1-b}\int_b^{1-b} \operatorname{Covar}\left(|X_{n,t}|^{p-1} \operatorname{sgn}\{X_{n,t}\}, |X_{n,s}|^{p-1} \operatorname{sgn}\{X_{n,s}\}\right) \\
&\qquad\qquad \cdot L'(t)L'(s)w(t)w(s)\, dt\, ds \\
&= \int_b^{1-b}\int_b^{1-b} \mathbb{1}_{\{|t-s|\leqslant 2b\}} \operatorname{Covar}\left(|X_{n,t}|^{p-1} \operatorname{sgn}\{X_{n,t}\}, |X_{n,s}|^{p-1} \operatorname{sgn}\{X_{n,s}\}\right) \\
&\qquad\qquad \cdot L'(t)L'(s)w(t)w(s)\, dt\, ds,
\end{aligned}$$

because for $|t-s| > 2b$, $X_{n,t}$ is independent of $X_{n,s}$. As a result, using that $X_{n,t}$ has bounded moments, we obtain

$$\operatorname{Var}\left(\int_b^{1-b} |X_{n,t}|^{p-1} \operatorname{sgn}\{X_{n,t}\} L'(t) w(t)\, dt\right) \to 0.$$

This means that

$$\int_b^{1-b} |X_{n,t}|^{p-1} \operatorname{sgn}\{X_{n,t}\} L'(t) w(t)\, dt \to \frac{\theta_1(p)}{Dp},$$

in probability and

$$-p\frac{W(L(1))}{L(1)}\int_b^{1-b} |X_{n,t}|^{p-1} \operatorname{sgn}\{X_{n,t}\} L'(t) w(t)\, dt = CW(L(1)) + o_P(1),$$

where

$$C = -\frac{\theta_1(p)}{DL(1)}. \tag{6.7.7}$$

Going back to (6.7.4), we conclude that

$$\begin{aligned}
&b^{-1/2}\left\{\int_b^{1-b} \left|b^{-1/2}\Gamma_n^{(2)}(t) + g_{(n)}(t)\right|^p d\mu(t) - m_n^c(p)\right\} \\
&\quad = b^{1/2}\sum_{i=1}^{M_3} \zeta_i + CW(L(1)) + o_P(1).
\end{aligned} \tag{6.7.8}$$

In the case $nb^5 \to 0$, we have $g(t) = 0$ in the definition of $\theta_1(p)$ in (6.1.13). Hence, by the symmetry of the standard normal distribution, it follows that $\theta_1(p) = 0$ and as a result $C = 0$. According to (6.7.6) and (6.7.8), this means that

$$b^{-1/2}\left\{\int_b^{1-b} \left|b^{-1/2}\Gamma_n^{(2)}(t) + g_{(n)}(t)\right|^p d\mu(t) - m_n^c(p)\right\}$$

converges in distribution to a mean zero normal random variable with variance $\sigma^2(p)$.

Then, consider the case $nb^5 \to C_0^2 > 0$. Note that ζ_i depends only on the Brownian motion on the interval $[c_i b - b, c_i b + b]$. These intervals are disjoint, because $c_{i+1} b - b = d_i b + b$. We write

$$W(L(1)) = \sum_{i=1}^{M_3} [W(t_{i+1}) - W(t_i)] + W(L(1)) - W(t_{M_3}),$$

where $t_i = L(c_i b - b)$, for $i = 1, \ldots, M_3$. Moreover, $W(L(1)) - W(t_{M_3}) \to 0$, in probability, since $t_{M_3} \sim L(1 + O(b)) \to L(1)$. Hence, the left hand side of (6.7.8), can be written as

$$\sum_{i=1}^{M_3} Y_i + o_P(1), \quad Y_i = b^{1/2}\zeta_i + C\left[W(t_{i+1}) - W(t_i)\right].$$

Since now we have a sum of independent random variables, we apply the Lindeberg-Feller central limit theorem. Using $\mathbb{E}[Y_i] = O(b^{5/2}M_2)$, it suffices to show that

$$\mathbb{E}\left[\left(\sum_{i=1}^{M_3} Y_i\right)^2\right] \to \tilde{\theta}^2(p) > 0, \tag{6.7.9}$$

and that the Lyapounov condition

$$\sum_{i=1}^{M_3} \mathbb{E}[Y_i^4] \left(\sum_{i=1}^{M_3} \mathbb{E}[Y_i^2]\right)^{-2} \to 0. \tag{6.7.10}$$

is satisfied. Once we have (6.7.9), condition (6.7.10) is equivalent to

$$\sum_{i=1}^{M_3} \mathbb{E}[Y_i^4] \to 0.$$

In order to prove this, we use that $\mathbb{E}[\zeta_i^4] = O(M_2^2)$, (see (B.1.4) in the proof of Lemma B.1.2. Then, we get

$$\begin{aligned}\sum_{i=1}^{M_3} \mathbb{E}[Y_i^4] &\leqslant O(b^2) \sum_{i=1}^{M_3} \mathbb{E}[\zeta_i^4] + O(1) \sum_{i=1}^{M_3} \mathbb{E}[(W(t_{i+1}) - W(t_i))^4] \\ &\leqslant O(M_3 b^2 M_2^2) + O(M_3 (t_{i+1} - t_i)^2) \\ &= o(1) + O(M_3 M_2^2 b^2) = o(1).\end{aligned}$$

Because $\mathbb{E}[Y_i] = O(b^{5/2}M_2)$, for (6.7.9) we have

$$\begin{aligned}\mathbb{E}\left[\left(\sum_{i=1}^{M_3} Y_i\right)^2\right] &= \sum_{i=1}^{M_3} \mathbb{E}\left[Y_i^2\right] + o(1)\\ &= b\sum_{i=1}^{M_3} \mathbb{E}\left[\zeta_i^2\right] + C^2 \sum_{i=1}^{M_3} (t_{i+1} - t_i)\\ &\quad + 2Cb^{1/2} \sum_{i=1}^{M_3} \mathbb{E}[\zeta_i\{W(t_{i+1}) - W(t_i)\}] + o(1).\end{aligned}$$

It can be shown that $b\sum_{i=1}^{M_3} \mathbb{E}\left[\zeta_i^2\right] \to 0$, see Lemma B.1.2 for details. Moreover,

$$\sum_{i=1}^{M_3} (t_{i+1} - t_i) = L((M_3 - 1)(M_2 + 2)b) - L(0) = L(1) + o(1).$$

Finally, since

$$\zeta_i = b^{-1} \int_{c_i b}^{d_i b} \left\{ |X_{n,t}|^p - \int_{-\infty}^{+\infty} \left|\sqrt{l(t)}Dx + g_{(n)}(t)\right|^p \phi(x)\, dx \right\} w(t)\, dt,$$

we can write

$$2Cb^{1/2} \sum_{i=1}^{M_3} \mathbb{E}[\zeta_i\{W(t_{i+1}) - W(t_i)\}] = 2C \sum_{i=1}^{M_3} \int_{c_i b}^{d_i b} \mathbb{E}\left[|X_{n,t}|^p Z_{n,t}\right] w(t)\, dt,$$

where $Z_{n,t} = b^{-1/2}\{W(t_{i+1}) - W(t_i)\}$. Note that

$$(X_{n,t}, Z_{n,t}) \sim N\left(\begin{bmatrix} g_{(n)}(t) \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_n^2(t) & \rho_n(t)\sigma_n(t)\tilde{\sigma}_n(t) \\ \rho_n(t)\sigma_n(t)\tilde{\sigma}_n(t) & \tilde{\sigma}_n^2(t)) \end{bmatrix}\right).$$

where $\sigma_n^2(t)$ is defined in (6.7.3) and

$$\begin{aligned}\tilde{\sigma}_n^2(t) &= b^{-1}[L(t+b) - L(t-b)],\\ \rho_n(t) &= \sigma_n(t)^{-1}\tilde{\sigma}_n(t)^{-1} b^{-1} \int k\left(\frac{t-u}{b}\right) l(u)\, du.\end{aligned}$$

Using

$$Z_{n,t} \mid X_{n,t} = x \sim N\left(\frac{\tilde{\sigma}_n(t)}{\sigma_n(t)}\rho_n(t)(x - g_{(n)}(t)), \left(1 - \rho_n^2(t)\right)\tilde{\sigma}_n^2(t)\right)$$

we obtain

$$\begin{aligned}
&\mathbb{E}\left[|X_{n,t}|^p Z_{n,t}\right] = \mathbb{E}\left[|X_{n,t}|^p\, \mathbb{E}[Z_{n,t} \mid X_{n,t}]\right] \\
&= \mathbb{E}\left[|X_{n,t}|^p \frac{\tilde{\sigma}_n(t)}{\sigma_n(t)} \rho_n(t) \left(X_{n,t} - g_{(n)}(t)\right)\right] \\
&= \frac{\tilde{\sigma}_n(t)}{\sigma_n(t)} \rho_n(t) \mathbb{E}\left[|X_{n,t}|^p \left(X_{n,t} - g_{(n)}(t)\right)\right] \\
&= \frac{\tilde{\sigma}_n(t)}{\sigma_n(t)} \rho_n(t) \int_{\mathbb{R}} |g_{(n)}(t) + \sigma_n(t)x|^p \sigma_n(t) x \phi(x)\, dx \\
&= \sigma_n(t)^{-1} b^{-1} \int k\left(\frac{t-u}{b}\right) l(u)\, du \int_{\mathbb{R}} |g_{(n)}(t) + \sigma_n(t)x|^p x\phi(x)\, dx.
\end{aligned}$$

Because $\sigma_n^2(t) \to D^2 l(t)$, where D is defined in (6.1.7), $g_{(n)}(t) \to g(t)$, as defined in (6.1.5), and $b^{-1} \int k\left(\frac{t-u}{b}\right) l(u)\, du \to l(t)$, we find that

$$\mathbb{E}\left[|X_{n,t}|^p Z_{n,t}\right] \to \frac{\sqrt{l(t)}}{D} \int_{\mathbb{R}} |g(t) + D\sqrt{l(t)}x|^p x\phi(x)\, dx.$$

Hence

$$\begin{aligned}
&\mathbb{E}\left[\left(\sum_{i=1}^{M_3} Y_i\right)^2\right] \\
&= \theta^2(p) + C^2 L(1) + \frac{2C}{D} \sum_{i=1}^{M_3} \int_{c_i b}^{d_i b} \int_{\mathbb{R}} |g(t) + \sqrt{l(t)}Dx|^p x\phi(x) dx \sqrt{l(t)} w(t) dt + o(1) \\
&= \theta^2(p) + C^2 L(1) + \frac{2C}{D} \int_0^1 \int_{\mathbb{R}} |g(t) + D\sqrt{l(t)}x|^p x\phi(x) dx \sqrt{l(t)} w(t)\, dt + o(1) \\
&= \theta^2(p) + C^2 L(1) + 2CD^{-1}\theta_1(p) + o(1) \\
&= \theta^2(p) - \frac{\theta_1^2(p)}{D^2 L(1)} + o(1),
\end{aligned}$$

applying the definitions of C and $\theta_1(p)$ in (6.7.7) and (6.1.13), respectively. It follows from the Lindeberg-Feller central limit theorem that $\sum_{i=1}^{M_3} Y_i \xrightarrow{d} N(0, \tilde{\theta}^2(p))$, where $\tilde{\theta}(p)$ is defined in (6.1.12). □

6.7.2 *Proofs for Section 6.3*

Lemma 6.7.2. *Let* Y_n *and* $Y_n^{(1)}$ *be defined in* (6.3.5) *and* (6.3.7)*, respectively. Assume that* (A1) – (A2) *hold. If* $1 \leqslant p < \min(q, 2q-7)$*,*

$$1/b = o\left(n^{1/3-1/q}\right) \quad \textit{and} \quad 1/b = o\left(n^{(q-3)/(6p)}\right),$$

then

$$b^{-p}\int_b^{1-b}|Y_n(t)-Y_n^{(1)}(t)|^p\,d\mu(t)=o_P(1).$$

Proof. We follow the same reasoning as in the proof of Lemma 5.2.5. Let $I_{nv}=[0,1]\cap[v-n^{-1/3}\log n, v+n^{-1/3}\log n]$ and for $J=E,W$, let

$$N_{nv}^J=\left\{[CM_{[0,1]}\Lambda_n^J](s)=[CM_{I_{nv}}\Lambda_n^J](s)\text{ for all } s\in I_{nv}\right\}. \tag{6.7.11}$$

Then according to Lemma 5.1.3, there exists $C>0$, independent of n,v,d, such that

$$\begin{aligned}\mathbb{P}\left((N_{nv}^W)^c\right)&=O(e^{-Cd^3})\\ \mathbb{P}\left((N_{nv}^E)^c\right)&=O(n^{1-q/3}d^{-2q}+e^{-Cd^3}).\end{aligned} \tag{6.7.12}$$

Let $K_{nv}=N_{nv}^E\cap N_{nv}^W$ and write

$$\begin{aligned}\mathbb{E}\left[\left|A_n^E(v)^p-A_n^W(v)\right|\right]&=\mathbb{E}\left[\left|A_n^E(v)^p-A_n^W(v)\right|\mathbb{1}_{K_{nv}^c}\right]\\ &\quad+n^{2p/3}\mathbb{E}\left[\left|[D_{I_{nv}}\Lambda_n](t)^p-[D_{I_{nv}}\Lambda_n^W](t)^p\right|\mathbb{1}_{K_{nv}}\right].\end{aligned}$$

From the proof of Lemma 5.2.5, using (6.7.12) with $d=\log n$, we have

$$\mathbb{E}\left[\left|A_n^E(v)^p-A_n^W(v)\right|\mathbb{1}_{K_{nv}^c}\right]=O_P(n^{1/2-q/6}(\log n)^{-q}+e^{-C(\log n)^3/2}/2)$$

and

$$n^{2p/3}\mathbb{E}\left[\left|[D_{I_{nv}}\Lambda_n](t)^p-[D_{I_{nv}}\Lambda_n^W](t)^p\right|\mathbb{1}_{K_{nv}}\right]=O_P\left(n^{-1/3+1/q}\right).$$

It follows that

$$\begin{aligned}&b^{-p}\int_b^{1-b}|Y_n(t)-Y_n^{(1)}(t)|^p\,d\mu(t)\\ &\leqslant Cb^{-p}\int_{-1}^{1}|A_n^E(t-by)-A_n^W(t-by)|^p\,dy\\ &=b^{-p}O_P\left(n^{-p/3+p/q}\right)+b^{-p}O_P\left(n^{1/2-q/6}(\log n)^{-q}+e^{-C(\log n)^3/2}\right).\end{aligned}$$

According to the assumptions on the order of b^{-1}, the right hand side is of order $o_P(1)$. □

Lemma 6.7.3. *Let* $Y_n^{(1)}$ *and* $Y_n^{(2)}$ *be defined in* (6.3.7) *and* (6.3.11)*, respectively. Assume that* (A1) – (A2) *hold. If* $b\to 0$*, such that* $nb\to\infty$*, then*

$$b^{-p}\int_b^{1-b}|Y_n^{(1)}(t)-Y_n^{(2)}(t)|^p\,d\mu(t)=o_P(1).$$

Proof. We have

$$\begin{aligned}
&\sup_{t\in(b,1-b)} \mathbb{E}\left[\left|Y_n^{(1)}(t)-Y_n^{(2)}(t)\right|^p\right] \\
&= \sup_{t\in(b,1-b)} \mathbb{E}\left[\left|\frac{1}{b}\int k'\left(\frac{t-v}{b}\right)\mathbb{1}_{(N_{nv}^W)^c}\left(A_n^W(v)-n^{\frac{2}{3}}[D_{I_{nv}}\Lambda_n^W](v)\right)dv\right|^p\right] \\
&\leqslant c\sup_{t\in(b,1-b)} \mathbb{E}\left[\sup_{v\in[0,1]}\left|A_n^W(v)-n^{\frac{2}{3}}[D_{I_{nv}}\Lambda_n^W](v)\right|^p\left(\frac{1}{b}\int_{t-b}^{t+b}\mathbb{1}_{(N_{nv}^W)^c}\,dv\right)^p\right],
\end{aligned}$$

where N_{nv}^W is defined in (6.7.11). Moreover, since

$$\sup_{v\in[0,1]}\left|A_n^W(v)-n^{\frac{2}{3}}[D_{I_{nv}}\Lambda_n^W](v)\right| \leqslant 4n^{\frac{2}{3}}\left\{\Lambda(1)+n^{-1/2}\sup_{s\in[0,L(1)]}|W_n(s)|\right\},$$

from the Cauchy-Schwartz inequality we obtain

$$\begin{aligned}
&\sup_{t\in(b,1-b)} \mathbb{E}\left[\sup_{v\in[0,1]}\left|A_n^W(v)-n^{2/3}[D_{I_{nv}}\Lambda_n^W](v)\right|^p\left(\frac{1}{b}\int_{t-b}^{t+b}\mathbb{1}_{(N_{nv}^W)^c}\,dv\right)^p\right] \\
&\quad\leqslant 4^p n^{2p/3}\mathbb{E}\left[\left\{\Lambda(1)+n^{-1/2}\sup_{s\in[0,L(1)]}|W_n(s)|\right\}^{2p}\right]^{1/2} \\
&\qquad\cdot\sup_{t\in(b,1-b)}\mathbb{E}\left[\left(\frac{1}{b}\int_{t-b}^{t+b}\mathbb{1}_{(N_{nv}^W)^c}\,dv\right)^{2p}\right]^{1/2}.
\end{aligned}$$

For the last term on the right hand side, we can use Jensen's inequality:

$$\left(\frac{1}{b-a}\int_a^b f(x)\,dx\right)^p \leqslant \frac{1}{b-a}\int_a^b f(x)^p\,dx,$$

for all $a<b$, $p\geqslant 1$, and $f(x)\geqslant 0$. Because all the moments of

$$\sup_{s\in[0,L(1)]}|W_n(s)|$$

are finite, together with (6.7.12), it follows that

$$\begin{aligned}
&\sup_{t\in(b,1-b)} \mathbb{E}\left[\left|Y_n^{(1)}(t)-Y_n^{(2)}(t)\right|^p\right] \\
&\leqslant Cn^{2p/3}\sup_{t\in(b,1-b)}\mathbb{E}\left[\frac{1}{b}\int_{t-b}^{t+b}\mathbb{1}_{(N_{nv}^W)^c}\,dv\right]^{1/2} \\
&= O\left(n^{2p/3}\exp\left(-C(\log n)^3/2\right)\right).
\end{aligned} \tag{6.7.13}$$

Since

$$b^{-p} n^{2p/3} \exp\left(-C(\log n)^3/2\right) = (nb)^{\frac{2p}{3} - \frac{C(\log n)^2}{2}} b^{-p - \frac{2p}{3} + \frac{C(\log n)^2}{2}} \to 0,$$

this concludes the proof. □

Lemma 6.7.4. *Let* $Y_n^{(2)}$ *and* $Y_n^{(3)}$ *be defined in* (6.3.11) *and* (6.3.14)*, respectively. Assume that* (A1) – (A2) *hold. If* $1/b = o\left(n^{1/3-1/q}\right)$*, then*

$$b^{-p} \int_b^{1-b} |Y_n^{(2)}(t) - Y_n^{(3)}(t)|^p \, d\mu(t) = o_P(1).$$

Proof. Let $H_{nv} = [-n^{1/3}v, n^{1/3}(1-v)] \cap [-\log n, \log n]$ and

$$\Delta_{nv} = n^{2/3}[D_{I_{nv}} \Lambda_n^W](v) - [D_{H_{nv}} Y_{nv}](0).$$

By definition, we have

$$\int_b^{1-b} |Y_n^{(2)}(t) - Y_n^{(3)}(t)|^p \, d\mu(t) = \int_b^{1-b} \left| \frac{1}{b} \int_{t-b}^{t+b} k'\left(\frac{t-v}{b}\right) \Delta_{nv} \, dv \right|^p d\mu(t).$$

Moreover, using

$$\sup_{t \in (0,1)} \mathbb{E}\left[|\Delta_{nt}|^p\right] = O\left(n^{-p/3+p/q}\right)$$

(see the proof of Lemma 5.2.3, we obtain

$$\begin{aligned}
&\sup_{t \in (b,1-b)} \mathbb{E}\left[\left| \frac{1}{b} \int_{t-b}^{t+b} k'\left(\frac{t-v}{b}\right) \Delta_{nv} \, dv \right|^p\right] \\
&\leqslant \sup_{u \in [-1,1]} |k'(u)|^p \sup_{t \in (b,1-b)} \mathbb{E}\left[\left| \frac{1}{b} \int_{t-b}^{t+b} \Delta_{nv} \, dv \right|^p\right] \\
&\leqslant C \sup_{t \in (b,1-b)} \frac{1}{b} \int_{t-b}^{t+b} \mathbb{E}\left[|\Delta_{nv}|^p\right] dv \leqslant 2C \sup_{v \in (0,1)} \mathbb{E}\left[|\Delta_{nv}|^p\right] \\
&= O\left(n^{-p/3+p/q}\right).
\end{aligned} \tag{6.7.14}$$

Because $1/b = o\left(n^{1/3-1/q}\right)$, this concludes the proof. □

Lemma 6.7.5. *Let* $Y_n^{(3)}$ *and* $Y_n^{(4)}$ *be defined in* (6.3.14) *and* (6.3.18)*, respectively. Assume that* (A1) – (A2) *hold. If* $1/b = o\left(n^{1/3-1/q}\right)$*, then*

$$b^{-p} \int_b^{1-b} |Y_n^{(3)}(t) - Y_n^{(4)}(t)|^p \, d\mu(t) = o_P(1).$$

Proof. Let $H_{n\nu}$ be defined as in the proof of Lemma 6.7.4 and let

$$J_{n\nu} = \left[n^{1/3} \frac{L(a_{n\nu}) - L(\nu)}{L'(\nu)}, n^{1/3} \frac{L(b_{n\nu}) - L(\nu)}{L'(\nu)} \right],$$

where $a_{n\nu} = \max(0, \nu - n^{-1/3} \log n)$ and $b_{n\nu} = \min(1, \nu + n^{-1/3} \log n)$. As in (4.31) in Kulikov and Lopuhaä, 2008 we have

$$\sup_{\nu \in (0,1)} \mathbb{E} \left[\left| [D_{H_{n\nu}} \tilde{Y}_{n\nu}](0) - [D_{J_{n\nu}} Z_{n\nu}](0) \right|^p \right] = O(n^{-p/3} (\log n)^{3p}),$$

where $\tilde{Y}_{n\nu}$ and $Z_{n\nu}$ are defined in (6.3.16) and (6.3.17). This means that,

$$\begin{aligned}
&\sup_{t \in (b, 1-b)} \mathbb{E} \left[\left| Y_n^{(3)}(t) - Y_n^{(4)}(t) \right|^p \right] \\
&\leqslant C \sup_{t \in (b, 1-b)} \mathbb{E} \left[\left| \frac{1}{b} \int_{t-b}^{t+b} \left\{ [D_{H_{n\nu}} \tilde{Y}_{n\nu}](0) - [D_{J_{n\nu}} Z_{n\nu}](0) \right\} d\nu \right|^p \right] \\
&\leqslant C \sup_{t \in (b, 1-b)} \frac{1}{b} \int_{t-b}^{t+b} \mathbb{E} \left[\left| [D_{H_{n\nu}} \tilde{Y}_{n\nu}](0) - [D_{J_{n\nu}} Z_{n\nu}](0) \right|^p \right] d\nu \\
&\leqslant C \sup_{\nu \in (b, 1-b)} \mathbb{E} \left[\left| [D_{H_{n\nu}} \tilde{Y}_{n\nu}](0) - [D_{J_{n\nu}} Z_{n\nu}](0) \right|^p \right] = O\left(n^{-p/3} (\log n)^{3p} \right).
\end{aligned} \tag{6.7.15}$$

Since $1/b = o\left(n^{1/3 - 1/q}\right)$, this concludes the proof. □

Lemma 6.7.6. *Let* $Y_n^{(4)}$ *and* $Y_n^{(5)}$ *be defined in* (6.3.18) *and* (6.3.20)*, respectively. Assume that* (A1) − (A2) *hold. If* $nb \to \infty$*, such that*

$$1/b = o(n^{1/6 + 1/(6p)} (\log n)^{-(1/2 + 1/(2p))}),$$

then

$$b^{-p} \int_b^{1-b} |Y_n^{(4)}(t) - Y_n^{(5)}(t)|^p \, d\mu(t) = o_P(1).$$

Proof. We argue as in the proof of Lemma 4.4 in Kulikov and Lopuhaä, 2008. When $\nu \in (n^{-1/3} \log n, 1 - n^{-1/3} \log n)$, there exists $M > 0$, depending only on λ, such that $[-M \log n, M \log n] \subset I_{n\nu}$, and on $[-M \log n, M \log n]$ we have that

$$CM_{[-M \log n, M \log n]} Z \leqslant CM_{I_{n\nu}} Z \leqslant CM_{\mathbb{R}} Z.$$

Let $N_{nM} = N(M \log n)$, where $N(d)$ is the event that $[CM_{[-d,d]} Z](s)$ is equal to $[CM_{\mathbb{R}} Z](s)$ for $s \in [-d/2, d/2]$. According to Lemma 1.2 in Kulikov and Lopuhaä, 2006, it holds that

$$\mathbb{P}(N(d)^c) \leqslant \exp(-d^3/2^7). \tag{6.7.16}$$

For convenience, write $\delta_n = n^{-1/3}\log n$. Because

$$[CM_{[-M\log n, M\log n]}Z](0) = [CM_{I_{nv}}Z](0) = [CM_{\mathbb{R}}Z](0)$$

on the event N_{nM}, we have by means of Cauchy-Schwartz, we find that

$$\begin{aligned}
&\sup_{v\in(\delta_n,1-\delta_n)} \mathbb{E}\left[|[D_{I_{nv}}Z](0) - [D_{\mathbb{R}}Z](0)|^p\right] \\
&= \sup_{v\in(\delta_n,1-\delta_n)} \mathbb{E}\left[|[D_{I_{nv}}Z](0) - [D_{\mathbb{R}}Z](0)|^p\right] \mathbb{1}_{N_{nM}^c} \\
&\leqslant 2^p \mathbb{E}\left[\left(\sup_{s\in\mathbb{R}}|Z(s)|\right)^p \mathbb{1}_{N_{nM}^c}\right] \\
&\leqslant 2^p \left(\mathbb{E}\left[\left(\sup_{s\in\mathbb{R}}|Z(s)|\right)^{2p}\right]\right)^{1/2} \mathbb{P}(N_{nM}^c)^{1/2}.
\end{aligned}$$

Because $\mathbb{E}[(\sup|Z|)^{2p}] < \infty$, together with (6.7.16), we find that

$$\sup_{v\in(\delta_n,1-\delta_n)} \mathbb{E}\left[|[D_{I_{nv}}Z](0) - [D_{\mathbb{R}}Z](0)|^p\right] = O\left(\exp(-C(\log n)^3)\right).$$

Note that

$$Y_n^{(4)}(t) - Y_n^{(5)}(t) = \frac{1}{b}\int_{t-b}^{t+b} k'\left(\frac{t-v}{b}\right)\frac{1}{c_1(v)}\left([D_{I_{nv}}Z](0) - [D_{\mathbb{R}}Z](0)\right)dv. \tag{6.7.17}$$

When $t \in (b+\delta_n, 1-b-\delta_n)$, then $v \in (t-b, t+b) \subset (\delta_n, 1-\delta_n)$, so after change of variables, it follows that

$$\begin{aligned}
&\sup_{t\in(b+\delta_n,1-b-\delta_n)} \mathbb{E}\left[\left|Y_n^{(4)}(t) - Y_n^{(5)}(t)\right|^p\right] \\
&\leqslant 2^p \frac{\sup_{u\in[-1,1]}|k'(u)|^p}{\inf_{v\in(0,1)} c_1(v)^p} \sup_{v\in(\delta_n,1-\delta_n)} \mathbb{E}\left[|[D_{I_{nv}}Z](0) - [D_{\mathbb{R}}Z](0)|^p\right] \\
&= O\left(\exp(-C(\log n)^3)\right).
\end{aligned} \tag{6.7.18}$$

Next, consider the case where $t \in (b, b+\delta_n)$. In this case we split the integral on the right hand side of (6.7.17) into an integral over $v \in (t-b, \delta_n)$

and an integral over $v \in (\delta_n, t+b)$. The latter integral can be bounded in the same way as in (6.7.18), whereas for the first integral we have

$$\begin{aligned}
&\left| \frac{1}{b} \int_{t-b}^{\delta_n} k'\left(\frac{t-v}{b}\right) \frac{1}{c_1(v)} \left([D_{I_{nv}} Z](0) - [D_{\mathbb{R}} Z](0)\right) dv \right| \\
&\quad \leqslant b^{-1} \delta_n \frac{\sup_{u \in [-1,1]} |k'(u)|}{\inf_{v \in (0,1)} c_1(v)} \left|[D_{I_{nv}} Z](0) - [D_{\mathbb{R}} Z](0)\right| \\
&\quad \leqslant b^{-1} \delta_n \frac{\sup_{u \in [-1,1]} |k'(u)|}{\inf_{v \in (0,1)} c_1(v)} [D_{\mathbb{R}} Z](0),
\end{aligned}$$

where we also use that $[D_{I_{nv}} Z](0) \leqslant [D_{\mathbb{R}} Z](0)$. Furthermore, since $[D_{\mathbb{R}} Z](0)$ has bounded moments of any order, for $t \in (b, b+\delta_n)$, we obtain

$$\begin{aligned}
&\sup_{t \in (b, b+\delta_n)} \mathbb{E}\left[\left|Y_n^{(4)}(t) - Y_n^{(5)}(t)\right|^p\right] \\
&\quad \leqslant b^{-p} \delta_n^p \frac{\sup_{u \in [-1,1]} |k'(u)|^p}{\inf_{v \in (0,1)} c_1(v)^p} \mathbb{E}\left[[D_{\mathbb{R}}] Z](0)^p\right] + O\left(\exp(-C(\log n)^3)\right) \\
&\quad = O_P\left(b^{-p} \delta_n^p\right) + O_P\left(\exp(-C(\log n)^3)\right).
\end{aligned} \tag{6.7.19}$$

A similar bound can be obtained for $t \in (1-b-\delta_n, 1-b)$. Putting things together yields,

$$\int_b^{1-b} \left|Y_n^{(4)}(t) - Y_n^{(5)}(t)\right|^p \, d\mu(t) = O_P\left(\exp(-C(\log n)^3)\right) + O_P\left(b^{-p} \delta_n^{p+1}\right).$$

Because $nb \to \infty$ implies

$$b^{-p} \exp(-C(\log n)^3) \to 0 \quad \text{and} \quad \frac{1}{b} = o\left(n^{1/6+1/(6p)} (\log n)^{-(1/2+1/(2p))}\right)$$

yields $b^{-2p} \delta_n^{p+1} \to 0$, this concludes the proof. □

Part IV

CONCLUSIONS

7 CONCLUSIONS AND REMARKS

This thesis deals with smooth estimation under monotonicity constraints of a real valued function on a compact interval. Functions of prime interest are probability densities, hazard rates and regression relationships. We considered different scenarios in which such an estimation problem is encountered and we analyzed various estimators obtained by a two step procedure: isotonization and smoothing (the order can be reversed). Isotonization is done through a constrained maximum likelihood or a Grenander-type procedure and smoothing through kernel functions. We were mainly interested in rates of convergence, which reflect the accuracy of the estimators, and large sample distributional properties, which are necessary for making statistical inference.

7.1 LOCAL BEHAVIOR OF ESTIMATORS

There is a wide literature on pointwise asymptotic results of various smooth monotone estimators in particular models (mostly for density and regression problems). Our objective was to apply and investigate smooth isotonic estimation in survival analyses, in particularly for the semi-parametric Cox regression model.

This model was initially proposed in biostatistics and quickly became broadly used to study the time to device failure in engineering, the effectiveness of a treatment in medicine, mortality in insurance problems, duration of unemployment in social sciences etc. One advantage of such a model is that it allows for the presence of censored data and covariates. Moreover, the proportional hazard property allows estimation of the regression coefficients while leaving the baseline hazard completely unspecified. However, when one is interested for instance in the absolute time to event, estimation of the baseline hazard is required. In practice, due to simplicity, a parametric assumption is commonly made on the baseline distribution but, in case of model misspecification, this can lead to incorrect inference. On the other hand, when one does not want to assume any particular functional form, nonparametric methods can be used (for example the Breslow estimator for the cumulative baseline hazard). However, it was shown that if a monotone

hazard is expected (for example when modeling survival times of patients after a successful medical treatment or failure times of mechanical components that degrade over time), using shape-constrained methods produces more accurate and reasonable estimates.

Motivated by these results, we aimed at further improving estimation of a monotone baseline hazard by means of smoothing, which in general leads to a faster rate of convergence and a nicer graphical representation. Indeed, if the true hazard function is twice differentiable and the bandwidth converges at an optimal rate, then we showed that the smooth isotonic estimators converge pointwise at rate $n^{2/5}$ (instead of $n^{1/3}$ for the isotonic ones) to a Gaussian distribution. The convergence could be faster if the true function was smoother but in that case higher order kernels need to be used which might lead to negative values of the estimates or deviations of monotonicity. We considered four different estimators and concluded that interchanging the order of smoothing and isotonization does not effect the asymptotic distribution. Moreover, smoothing preserves the asymptotic equivalence between the isotonic estimators. Hence, from the theoretical point of view, there is no reason to prefer one estimator with respect to the others. Such behavior was to be expected in view of existing results in other frameworks such as density or regression models. However, to our best knowledge, this is the first study that investigates and compares at the same time four different smooth isotonic estimators, varying both the method of isotonization and the order of the procedures.

From a methodological point of view, the Cox model is more challenging than standard density and regression models because it contains a parametric an a nonparametric component. The estimates of the nonparametric baseline hazard depend on the parametric estimator of the regression coefficients. To emphasize this, we first considered the simpler right random censoring model without covariates and showed that there, the asymptotic normality is obtained using a short and direct argument that relies on a Kiefer-Wolfowitz type of result. This approach does not apply to the Cox model because such result is not available. Hence, we followed a more technical method that is mainly based on tail probabilities for the inverse process and uniform L_2-bounds for the distance between the non-smoothed isotonic estimator and the true function.

As an application, we considered constructing pointwise confidence intervals using the asymptotic distribution or a bootstrap procedure. Numerical results show that smoothing improves the performance of the non-smooth isotonic estimators. Moreover, undersmoothing leads to more accurate asymptotic confidence intervals with respect to bias estimation. We observe that the performance of the confidence intervals depends on the

choice of the constant c in the definition of the bandwidth ($b = cn^{-1/5}$). However, with right choice of the smoothing parameter, the asymptotic confidence intervals based on the smooth isotonic estimators can behave even better than those obtained through inversion of likelihood ratio statistics which, on the other hand, has the advantage of being parameter free. The fact that estimation in the Cox model is more challenging is also reflected by the lower coverage probabilities compared to the ones for the right censoring model. For finite sample sizes, estimation of β_0 effects the performance of the estimators even if it does not influence the limit distribution. However, using a bootstrap procedure, instead of the normal approximation, leads to more satisfactory results. In particular, bootstrap confidence intervals with undersmoothing seem to have the most accurate coverage probabilities. Simulations confirm that the four smooth isotonic estimators that we considered (and the ordinary kernel estimator) have comparable behavior, with the smoothed maximum likelihood and the maximum smoothed likelihood being slightly more accurate.

Unfortunately, it is still not clear how to optimally choose the smoothing parameter. Various methods of bandwidth selection have been proposed in the literature such as cross-validation, plug-in techniques, bootstrap etc. For an increasing hazard estimation, cross-validation methods seem to suffer from the fact that the variance of the estimator increases as one approaches the endpoint of the support. We also tried a locally optimal bandwidth by minimizing the estimated asymptotic mean squared error but, in our setting, it did not improve the results. We did not investigate this problem in details but chose to prefer simplicity by using a constant c equal to the range of the data in the definition of the bandwidth. However, more insight into this problem would certainly be useful for the practical use of smooth monotone estimators.

7.2 GLOBAL BEHAVIOR OF ESTIMATORS

In contrast to pointwise asymptotic results of shape constrained estimators, the literature on global errors of estimates is more limited. Common measures are the L_p and the Hellinger distances. Central limit theorems for the L_p-errors were previously established for the Grenander-type estimator of a monotone density or a monotone function in a general setting which includes the density, regression and right censoring model. On the other hand, central limit theorems for the kernel estimator of a smooth density on $\mathbb{R}$ (without monotonicity constraints) were also available but there was no result on global errors of smooth isotonic estimators. Hence, instead of

investigating directly the global behavior of smooth isotonic estimators in the Cox model, which is quite challenging, we started with a simpler setup. We considered estimation of a general monotone function under a strong approximation assumption of the observed cumulative process by a Gaussian process. Such an assumption was shown to be satisfied in the density, regression and right censoring model, but does not hold in the Cox framework.

We first derived a central limit theorem for the Hellinger distance of the Grenander-type estimator by relating it to a weighted L_2-error. We observed that, in the density model, the limit variance of the Hellinger-error does not depend on the underlying density. Such a phenomenon was previously encountered for the L_1-distance. Moreover, we investigated the use of the Hellinger distance for goodness of fit tests. In particular, we considered testing exponentiality, which is of interest in practice, against a decreasing alternative. The critical region of the test is determined using a bootstrap procedure and the performance is investigated in different scenarios. In terms of level, the test behaves quite well but it is usually less powerful than existing tests for exponentiality (which is to be expected because they make use of properties of the exponential distribution). However, there are cases (Beta distribution) in which our test is more powerful than the other competing tests.

Afterwards we investigated the asymptotic behavior of the L_p-error of the smoothed Grenander-type estimator and the isotonized kernel estimator. First, we noticed that when dealing with global errors, a uniform bound on the distance between a naive cumulative estimator and its least concave majorant (Kiefer-Wolfowitz result) is no longer sufficient for connecting the smoothed Grenander-type estimator to the kernel estimator. Hence, as an intermediate step, we generalized previous results in the literature (only available for density functions) to our general setting. We proved the process convergence of $\hat{\Lambda}_n - \Lambda_n$ after a suitable rescaling and a central limit theorem for the L_p-distance between $\hat{\Lambda}_n$ and Λ_n. Such result was then used to show that the L_p-error of the smoothed Grenander-type estimator has the same asymptotic behavior as the L_p-error of the kernel estimator. We conclude that the smooth isotonic estimators have the same limit distribution as the kernel estimator even in terms of global distances.

For the L_p-error of the ordinary kernel estimator, we observe that not having a smooth density on the whole real line leads to boundary effects which can be overcome by using a boundary corrected kernel. In the latter case, the contribution of the boundary regions in the limit distribution of the L_p-error is negligible. An interesting fact is that we find a limit variance which changes according to whether the approximating process is a Brownian mo-

tion or a Brownian bridge. Such a phenomenon has not been observed before in isotonic problems and also contradicts the previously made claim in Csörgő and Horváth, 1988 that the limit variance of the L_p-error of the kernel density estimator (in this case the approximating process is a Brownian bridge) is the same as if the approximating process was a Brownian motion. Such difference is not encountered only when the bandwidth is taken of smaller order than $n^{-1/5}$.

As an application, we propose a test for monotonicity based on the L_2-distance between the smoothed Grenander estimator and the kernel estimator. We consider a regression model with deterministic design points and compute the critical region of the test by a bootstrap procedure instead of using the asymptotic distribution (to avoid estimation of unknown quantities). Compared to other tests available in the literature, this test performs better in terms of power. However, a more thorough investigation is still needed to show that the bootstrap method works and that the test also has the right level.

To conclude, it remains to address the global errors of isotonic and smooth isotonic estimators in the Cox model. This is more challenging since a strong approximation by a Gaussian process, that we assumed here, is not available for the Cox model. Such a problem is object of my current (and future) research. We expect that similar results to those established in Part III of this thesis would hold even in the Cox model. In contrast to the pointwise case, we expect that, for a bandwidth of order $n^{-1/5}$, $\hat{\beta}_n$ will effect the limit distribution of the L_p-error of the kernel estimator (and of smooth isotonic estimators).

Part V

SUPPLEMENTARY MATERIAL

A

SUPPLEMENT TO: SMOOTH ISOTONIC ESTIMATION IN THE COX MODEL

A.1 ENTROPY BOUNDS FOR THE SG ESTIMATOR

Lemma A.1.1. *Let $0 < t < M < \tau_H$ and let $\mathcal{B}_{\tilde{K}}$ be the class of functions of bounded variation on $[0, M]$, that are uniformly bounded by $\tilde{K} > 0$. Let $\mathcal{G}_n = \{\zeta_{B,n} : B \in \mathcal{B}_{\tilde{K}}\}$, where $\zeta_{B,n}(u, \delta) = \delta \mathbb{1}_{[t-2b,t+2b]}(u)B(u)$. For $\delta > 0$, let*

$$J(\delta, \mathcal{G}_n) = \sup_Q \int_0^\delta \sqrt{1 + \log N(\epsilon \|F_n\|_{L_2(Q)}, \mathcal{G}_n, L_2(Q))}\, d\epsilon$$

where $N(\epsilon, \mathcal{G}_n, L_2(Q))$ is the minimal number of $L_2(Q)$-balls of radius ϵ needed to cover $\mathcal{G}_n$, $F_n = \tilde{K}\delta \mathbb{1}_{[t-2b,t+2b]}$ is the envelope of $\mathcal{G}_n$, and the supremum is taken over measures Q on $[0, M] \times \{0, 1\} \times \mathbb{R}^p$, for which $Q([t - 2b, t + 2b] \times \{1\} \times \mathbb{R}^p) > 0$. Then $J(\delta, \mathcal{G}_n) \leqslant \int_0^\delta \sqrt{1 + C/\epsilon}\, d\epsilon$, for some $C > 0$.

Proof. We first bound the entropy of $\mathcal{G}_n$ with respect to any probability measure Q on $[0, M] \times \{0, 1\} \times \mathbb{R}^p$, such that

$$\|F_n\|^2_{L_2(Q)} = \tilde{K} Q\left([t - 2b, t + 2b] \times \{1\} \times \mathbb{R}^p\right) > 0. \tag{A.1.1}$$

Fix such a probability measure Q and let $\epsilon > 0$. Let Q' be the probability measure on $[0, M]$ defined by

$$Q'(S) = \frac{Q(S \cap [t - 2b, t + 2b] \times \{1\} \times \mathbb{R}^p)}{Q([t - 2b, t + 2b] \times \{1\} \times \mathbb{R}^p)}, \qquad S \in [0, M],$$

For a given $\epsilon' > 0$, select a minimal ϵ'-net $B_1, \ldots, B_N$ in $\mathcal{B}_{\tilde{K}}$ with respect to $L_2(Q')$, where $N = N(\epsilon', \mathcal{B}_{\tilde{K}}, L_2(Q'))$. Then, from (2.6) in Geer, 2000

$$\log N(\epsilon', \mathcal{B}_{\tilde{K}}, L_2(Q')) \leqslant \frac{K}{\epsilon'}, \tag{A.1.2}$$

for some constant $K > 0$. Then, consider the functions $\zeta_{B_1,n}, \ldots, \zeta_{B_N,n}$ corresponding to $B_1, \ldots, B_N$. For any $\zeta_{B,n} \in \mathcal{G}_n$, there exists a B_i in the ϵ'-net that is closest function to B, i.e., $\|B - B_i\|_{L_2(Q')} \leqslant \epsilon'$. Then, we find

$$\begin{aligned} \|\zeta_{B,n} - \zeta_{B_i,n}\|_{L_2(Q)} &= Q([t - 2b, t + 2b] \times \{1\} \times \mathbb{R}^p)^{1/2} \| B - B_i\|_{L_2(Q')} \\ &\leqslant Q([t - 2b, t + 2b] \times \{1\} \times \mathbb{R}^p)^{1/2} \epsilon'. \end{aligned} \tag{A.1.3}$$

Hence, if we take $\epsilon' = \epsilon/Q([t-2b, t+2b] \times \{1\} \times \mathbb{R}^p)^{1/2}$, it follows that $\zeta_{B_1,n}, \ldots, \zeta_{B_N,n}$ forms an ϵ-net in $\mathcal{G}_n$ with respect to $L_2(Q)$, and

$$N(\epsilon, \mathcal{G}_n, L_2(Q)) \leqslant N(\epsilon', \mathcal{B}_{\tilde{K}}, L_2(Q')).$$

Using (A.1.2), this implies that

$$\log N(\epsilon, \mathcal{G}_n, L_2(Q)) \leqslant \frac{K}{\epsilon'} = \frac{K \cdot Q([t-2b, t+2b] \times \{1\} \times \mathbb{R}^p)^{1/2}}{\epsilon}$$

where K does not depend on Q, and according to (A.1.1),

$$\begin{aligned} J(\delta, \mathcal{G}_n) &= \sup_Q \int_0^\delta \sqrt{1 + \log N(\epsilon \|F_n\|_{L_2(Q)}, \mathcal{G}_n, L_2(Q))}\, d\epsilon \\ &\leqslant \int_0^\delta \sqrt{1 + \frac{C}{\epsilon \tilde{K}^{1/2}}}\, d\epsilon, \end{aligned}$$

for some $C > 0$. □

Lemma A.1.2. *Let $\mathcal{M}_C$ be the class of monotone functions bounded uniformly by $C > 0$. Let $\mathcal{G}_n = \{\zeta_{r,\beta}(u,z) : r \in \mathcal{M}_C, \beta \in \mathbb{R}^P, |\beta - \beta_0| \leqslant \xi_2\}$, where ξ_2 is chosen as in* (3.6.3) *and $\zeta_{r,\beta}(u,z) = r(u)e^{\beta' z}$. For $\delta > 0$, let*

$$J_{[]}(\delta, \mathcal{G}_n, L_2(\mathbb{P})) = \int_0^\delta \sqrt{1 + \log N_{[]}(\epsilon \|F_n\|_{L_2(\mathbb{P})}, \mathcal{G}_n, L_2(\mathbb{P}))}\, d\epsilon.$$

where $N_{[]}(\epsilon, \mathcal{G}_n, L_2(\mathbb{P}))$ is the bracketing number and

$$F_n(u,z) = C \exp\left\{ \sum_{j=1}^p (\beta_{0,j} - \sigma_n) z_j \vee (\beta_{0,j} + \sigma_n) z_j \right\},$$

with $\sigma_n = \sqrt{\xi_2 n^{-2/3}}$ is the envelope of $\mathcal{G}_n$. Then

$$J_{[]}(\delta, \mathcal{G}_n, L_2(\mathbb{P})) \leqslant \int_0^\delta \sqrt{1 + C/\epsilon}\, d\epsilon,$$

for some $C > 0$.

Proof. The entropy with bracketing for the class of bounded monotone functions on $\mathbb{R}$ satisfies

$$\log N_{[]}(\gamma, \mathcal{M}_C, L_p(Q)) \leqslant \frac{K}{\gamma}, \tag{A.1.4}$$

for every probability measure Q, $\gamma > 0$ and $p \geqslant 1$ (e.g., see Theorem 2.7.5 in van der Vaart and Wellner, 1996). Define the probability measure Q on $\mathbb{R}$ by

$$Q(S) = \int \mathbb{1}_S(u)\, d\mathbb{P}(u, \delta, z), \qquad \text{for all } S \subseteq \mathbb{R},$$

and fix $\epsilon > 0$. For a given $\gamma > 0$, take a net of γ-brackets $\{(l_1, L_1) \ldots, (l_N, L_N)\}$ in $\mathcal{M}_C$ with respect to $L_4(Q)$, where $N = N_{[]}(\gamma, \mathcal{M}_C, L_4(Q))$. For every $j = 1, \ldots, p$, divide the interval $[\beta_{0,j} - \sigma_n, \beta_{0,j} + \sigma_n]$ in subintervals of length γ, i.e.,

$$\left[a_{k_j}, b_{k_j}\right] = \left[\beta_{0,j} - \sigma_n + k_j\gamma, \beta_{0,j} - \sigma_n + (k_j + 1)\gamma\right],$$

for $k_j = 0, \ldots, \bar{N} - 1$, where $\bar{N} = 2\sigma_n/\gamma$. Then, for $i = 1, \ldots, N$ and $k_j = 0, \ldots, \bar{N} - 1$, $j = 1, \ldots, p$, construct brackets $(l_{i,k_1,\ldots,k_p}, L_{i,k_1,\ldots,k_p})$, in the following way:

$$l_{i,k_1,\ldots,k_p}(u, z) = l_i(u) \exp\left\{\sum_{j=1}^{p} a_{k_j} z_j \wedge b_{k_j} z_j\right\},$$

$$L_{i,k_1,\ldots,k_p}(u, z) = L_i(u) \exp\left\{\sum_{j=1}^{p} a_{k_j} z_j \vee b_{k_j} z_j\right\}.$$

By construction, using (A.1.4), the number of these brackets is

$$N \cdot \bar{N}^p \leqslant \exp\{K/\gamma\} \cdot \left(\frac{2\sigma_n}{\gamma}\right)^p \leqslant \exp\{K/\gamma\} \cdot \exp\left\{2p\sqrt{\xi_2}/\gamma\right\} = e^{K_1/\gamma}, \tag{A.1.5}$$

for some $K_1 > 0$ independent of $n \geqslant 1$. For any $\zeta_{r,\beta} \in \mathcal{G}_n$, there exist a γ-bracket (l_i, L_i), such that $r \in [l_i, L_i]$, and intervals $[a_{k_j}, b_{k_j}]$, such that $\beta_j \in [a_{k_j}, b_{k_j}]$, for all $j = 1, \ldots, p$. It follows that there exists a bracket $(l_{i,k_1,\ldots,k_p}, L_{i,k_1,\ldots,k_p})$, such that $\zeta_{r,\beta} \in [l_{i,k_1,\ldots,k_p}, L_{i,k_1,\ldots,k_p}]$. Moreover, for each $i = 1, \ldots, n$ and $k_j = 0, \ldots, \bar{N} - 1$, for $j = 1, \ldots, p$,

$$\begin{aligned}
&\left\|l_{i,k_1,\ldots,k_p} - L_{i,k_1,\ldots,k_p}\right\|_{L_2(\mathbb{P})} \\
&\leqslant \left\|\exp\left\{\sum_{j=1}^{p} a_{k_j} z_j \wedge b_{k_j} z_j\right\}(l_i(u) - L_i(u))\right\|_{L_2(\mathbb{P})} \\
&\quad + C\left\|\exp\left\{\sum_{j=1}^{p} a_{k_j} z_j \wedge b_{k_j} z_j\right\} - \exp\left\{\sum_{j=1}^{p} a_{k_j} z_j \vee b_{k_j} z_j\right\}\right\|_{L_2(\mathbb{P})} \\
&\leqslant \left\|\exp\left\{\sum_{j=1}^{p} a_{k_j} z_j \wedge b_{k_j} z_j\right\}\right\|_{L_4(\mathbb{P})} \|l_i - L_i\|_{L_4(\mathbb{P})} \\
&\quad + C\left\|\exp\left\{\sum_{j=1}^{p} a_{k_j} z_j \wedge b_{k_j} z_j\right\} - \exp\left\{\sum_{j=1}^{p} a_{k_j} z_j \vee b_{k_j} z_j\right\}\right\|_{L_2(\mathbb{P})}.
\end{aligned}$$

For the first term on the right hand side, we use that for all $j = 1, \ldots, p$,

$$\left|a_{k_j} - \beta_{0,j}\right| \leqslant \sigma_n \quad \text{and} \quad \left|b_{k_j} - \beta_{0,j}\right| \leqslant \sigma_n, \tag{A.1.6}$$

so that, according to (A3), there exist a $K_1 > 0$, such that for n sufficiently large

$$\int \exp\left\{4\sum_{j=1}^{p} a_{k_j} z_j \wedge b_{k_j} z_j\right\} d\mathbb{P}(u,\delta,z) \leqslant \sup_{|\beta-\beta_0|\leqslant p\sigma_n} \mathbb{E}\left[e^{4\beta' Z}\right] \leqslant K_1.$$

For the second term we have $\|l_i - L_i\|_{L_4(Q)} \leqslant \gamma$, by construction, and by the mean value theorem, the third term on the right hand side can be bounded by

$$C\left(\int\left(\gamma\sum_{i=1}^{p}|z_i|\right)^2 e^{2\theta_z} d\mathbb{P}(u,\delta,z)\right)^{1/2} \leqslant C'\gamma\left(\int |z|^2 e^{2\theta_z} d\mathbb{P}(u,\delta,z)\right)^{1/2},$$

for some $\sum_{j=1}^{p} a_{k_j} z_j \wedge b_{k_j} z_j \leqslant \theta_z \leqslant \sum_{j=1}^{p} a_{k_j} z_j \vee b_{k_j} z_j$. Consequently, in view of (A.1.6), we find

$$\begin{aligned}\|l_{i,k_1,\ldots,k_p} - L_{i,k_1,\ldots,k_p}\|_{L_2(\mathbb{P})} &\leqslant \gamma K_1^{1/4} + C'\gamma\left(\sup_{|\beta-\beta_0|\leqslant p\sigma_n} \mathbb{E}\left[|Z|^2 e^{2\beta' Z}\right]\right)^{1/2} \\ &\leqslant K'\gamma,\end{aligned}$$

for some $K' > 0$, using (A2). Hence, if we take $\gamma = \epsilon/K'$, then the brackets $\{(l_{i,k_1,\ldots,k_p}, L_{i,k_1,\ldots,k_p})\}$, for $i = 1,\ldots,n$ and $k_j = 0,\ldots,\bar{N}-1$, for $j = 1,\ldots,p$, forms a net of ϵ-brackets, and according to (A.1.5), there exists a $K > 0$, such that

$$\log N_{[]}(\epsilon, \mathcal{G}_n, L_2(\mathbb{P})) \leqslant \frac{K}{\epsilon}.$$

As a result,

$$\begin{aligned}J_{[]}(1, \mathcal{G}_n, L_2(\mathbb{P})) &= \int_0^1 \sqrt{1 + \log N_{[]}(\epsilon\,\|F\|_{L_2(\mathbb{P})}, \mathcal{G}_n, L_2(\mathbb{P}))}\, d\epsilon \\ &= \int_0^1 \sqrt{1 + \frac{K}{\epsilon}}\, d\epsilon,\end{aligned}$$

for some $K > 0$. □

A.2 SMOOTH MAXIMUM LIKELIHOOD ESTIMATOR

To derive the pointwise asymptotic distribution of $\hat{\lambda}_n^{SM}$, we follow the same approach as the one used for $\tilde{\lambda}_n^{SG}$. We will go through the same line of reasoning as used to obtain Theorem 3.2.5. However, large parts of the proof are very similar, if not the same. We briefly sketch the main steps.

In this case we take

$$\hat{\Lambda}_n(t) = \int_0^t \hat{\lambda}_n(u)\,du,$$

where $\hat{\lambda}_n$ is the MLE for λ_0. First, the analogue of Lemma 3.2.1 still holds.

Lemma A.2.1. *Suppose that (A1)–(A2) hold. Let* $a_{n,t}$ *be defined by* (3.2.9) *and let* $\hat{\beta}_n$ *be the partial maximum likelihood estimator for* β_0*. Define*

$$\theta_{n,t}(u,\delta,z) = \mathbb{1}_{E_n}\left\{\delta a_{n,t}(u) - e^{\hat{\beta}_n' z}\int_0^u a_{n,t}(v)\,d\hat{\Lambda}_n(v)\right\}, \tag{A.2.1}$$

Then, there exists an event E_n*, with* $\mathbb{1}_{E_n} \to 1$ *in probability, such that*

$$\int \theta_{n,t}(u,\delta,z)\,d\mathbb{P}(u,\delta,z) = -\mathbb{1}_{E_n}\int k_b(t-u)\,d(\hat{\Lambda}_n - \Lambda_0)(u) + O_p(n^{-1/2}).$$

Proof. We modify the definition of the event E_n from Lemma 3.2.1 as follows. The events $E_{n,1}$, $E_{n,2}$, $E_{n,4}$, and $E_{n,5}$, from (3.6.1) and (3.6.3) remain the same. Replace $E_{n,3}$ in (3.6.3) by

$$E_{n,3} = \left\{\sup_{x\in[T_{(1)},T_{(n)}]} |V_n(x) - H^{uc}(x)| < \xi_3\right\}, \tag{A.2.2}$$

and let

$$E_{n,6} = \left\{T_{(1)} \leqslant \epsilon\right\}, \tag{A.2.3}$$

for some $\epsilon > 0$ and $\xi_3 > 0$. Then $\mathbb{P}(E_{n,6}) \to 1$, and also $\mathbb{P}(E_{n,3}) \to 1$ according to Lemma 5 in Lopuhaä and Nane, 2013. As before $E_n = \bigcap_{i=1}^6 E_{n,i}$. Similar to the proof of Lemma 3.2.1, we obtain

$$\begin{aligned}
&\int \theta_{n,t}(u,\delta,z)\,d\mathbb{P}(u,\delta,z)\\
&= \mathbb{1}_{E_n}\left\{\int a_{n,t}(u)\,dH^{uc}(u) - \int e^{\hat{\beta}_n' z}\int_{v=0}^u a_{n,t}(v)\,d\hat{\Lambda}_n(v)\,d\mathbb{P}(u,\delta,z)\right\}\\
&= \mathbb{1}_{E_n}\left\{\int k_b(t-u)\left(1 - \frac{\Phi(u;\hat{\beta}_n)}{\Phi(u;\beta_0)}\right)d\hat{\Lambda}_n(u) - \int k_b(t-u)\,d(\hat{\Lambda}_n - \Lambda_0)(u)\right\}
\end{aligned}$$

and

$$\mathbb{1}_{E_n}\int k_b(t-u)\left|1 - \frac{\Phi(u;\hat{\beta}_n)}{\Phi(u;\beta_0)}\right| d\hat{\Lambda}_n(u) = O_p(n^{-1/2}).$$

which proves the lemma. □

Next, we slightly change the definition of $\overline{\Psi}_{n,x}$ from (3.2.13), i.e.,

$$\overline{\Psi}_{n,t}(u) = \overline{a}_{n,t}(u)\mathbb{1}_{E_n} = a_{n,t}(\hat{A}_n(u))\mathbb{1}_{E_n} \tag{A.2.4}$$

where E_n is the event from Lemma A.2.1 and $\hat{A}_n$, as defined in (3.2.12), is now taken constant on $[\tau_i, \tau_{i+1})$ and consider

$$\bar{\theta}_{n,t}(u,\delta,z) = \delta\overline{\Psi}_{n,t}(u) - e^{\hat{\beta}'_n z}\int_0^u \overline{\Psi}_{n,t}(v)\,d\hat{\Lambda}_n(v). \tag{A.2.5}$$

Let

$$j_{n1} = \max\{j : \tau_j \leqslant t-b\}, \qquad j_{n2} = \min\{j : \tau_j \geqslant t+b\} \tag{A.2.6}$$

where $(\tau_j)_{j=1\ldots,m}$ are jump point of $\hat{\lambda}_n$. Note that, from the definition of $a_{n,t}$ and of $\hat{A}_n(u)$, it follows that $\overline{\Psi}_{n,t}(u) = 0$ for $u \notin [\tau_{j_{n1}}, \tau_{j_{n2}}]$. With this $\bar{\theta}_{n,t}$, we have the same property as in Lemma 3.2.2.

Lemma A.2.2. *Let* $\bar{\theta}_{n,t}$ *be defined in* (A.2.5). *Then*

$$\int \bar{\theta}_{n,t}(u,\delta,z)\,d\mathbb{P}_n(u,\delta,z) = 0.$$

Proof. Similar to the proof of Lemma 3.2.2, we have

$$\begin{aligned}
&\int \bar{\theta}_{n,t}(u,\delta,z)\,d\mathbb{P}_n(u,\delta,z)\\
&= \mathbb{1}_{E_n}\sum_{i=0}^{m}\overline{a}_{n,t}(\tau_i)\left\{\int \mathbb{1}_{[\tau_i,\tau_{i+1})}(u)\delta\,d\mathbb{P}_n(u,\delta,z) - \int_{\tau_i}^{\tau_{i+1}}\Phi_n(v;\hat{\beta}_n)\,d\hat{\Lambda}_n(v)\right\}\\
&= \mathbb{1}_{E_n}\sum_{i=0}^{m}\overline{a}_{n,t}(\tau_i)\left\{V_n(\tau_{i+1}) - V_n(\tau_i) - \hat{\lambda}_n(\tau_i)\left(\hat{W}_n(\tau_{i+1}) - \hat{W}_n(\tau_i)\right)\right\}\\
&= 0,
\end{aligned}$$

The last equality follows from the characterization of the maximum likelihood estimator. □

Furthermore, for $\bar{\theta}_{n,t}$ defined in (A.2.5), we also have

$$\int \left\{\bar{\theta}_{n,t}(u,\delta,z) - \theta_{n,t}(u,\delta,z)\right\}\,d\mathbb{P}(u,\delta,z) = O_p(b^{-1}n^{-2/3}), \tag{A.2.7}$$

see Lemma A.2.10, and

$$\begin{aligned}
&\int \left\{\bar{\theta}_{n,t}(u,\delta,z) - \eta_{n,t}(u,\delta,z)\right\}\,d(\mathbb{P}_n - \mathbb{P})(u,\delta,z)\\
&= O_p(b^{-3/2}n^{-13/18}) + O_p(n^{-1/2}),
\end{aligned} \tag{A.2.8}$$

see Lemma A.2.11, where $\eta_{n,t}$ is defined similar to (3.2.20), but with E_n taken from Lemma A.2.1. Similar to the proof of Lemma 3.2.3, the proof of Lemma A.2.10 is quite technical and involves bounds on the tail probabilities of the inverse process corresponding to $\hat{\lambda}_n$ (see Lemma A.2.5), used to obtain the analogue of (3.2.19) (see Lemma A.2.6). The inverse process defined as in (3.2.30) satisfies the switching relation $\hat{\lambda}_n(t) \leqslant a$ if and only if $\hat{U}_n(a) \geqslant t$. Let U be the inverse of λ_0 on $[\lambda_0(\epsilon), \lambda_0(M)]$, for some $0 < \epsilon < M < \tau_H$, i.e.,

$$U(a) = \begin{cases} \epsilon & a < \lambda_0(\epsilon); \\ \lambda_0^{-1}(a) & a \in [\lambda_0(\epsilon), \lambda_0(M)]; \\ M & a > \lambda_0(M). \end{cases} \tag{A.2.9}$$

In order to bound the tail probabilities of $\hat{U}_n(a)$ we first introduce a suitable martingale that will approximate the process $V_n(t) - a\hat{W}_n(t)$.

Lemma A.2.3. *Suppose that (A1)–(A2) hold. Define*

$$\bar{\mathbb{B}}_n(t) = V_n(t) - \int_0^t \Phi_n(s; \beta_0)\lambda_0(s)\, ds. \tag{A.2.10}$$

The process $\{(\bar{\mathbb{B}}_n(t), \mathcal{F}_t^n) : 0 \leqslant t < \tau_H\}$ *is a square integrable martingale with mean zero and predictable variation process*

$$\langle \bar{\mathbb{B}}_n \rangle(t) = \frac{1}{n} \int_0^t \Phi_n(s; \beta_0)\, d\Lambda_0(s).$$

Proof. Note that

$$\bar{\mathbb{B}}_n(t) = \frac{1}{n}\mathbb{M}_n(t) - \frac{1}{n}\sum_{i=1}^n \mathbb{1}_{\{T_i = t\}}\Delta_i$$

where $\mathbb{M}_n$ is defined in (3.6.4). Since H^{uc} is absolutely continuous, we have $\mathbb{1}_{\{T_i=t\}}\Delta_i = 0$ a.s., which means that $\bar{\mathbb{B}}_n = \mathbb{M}_n$ a.s.. Hence $\bar{\mathbb{B}}_n$ is a mean zero martingale and has the same predictable variation as $n^{-1}\mathbb{M}_n$. □

Lemma A.2.4. *Suppose that (A1)–(A2) hold. There exists a constant* $C > 0$ *such that, for all* $x > 0$ *and* $t \in [0, \tau_H]$,

$$\mathbb{E}\left[\sup_{u \in [0,\tau_H], |t-u| \leqslant x} \left(\bar{\mathbb{B}}_n(u) - \bar{\mathbb{B}}_n(t)\right)^2\right] \leqslant \frac{Cx}{n}.$$

Proof. First, consider the case $t \leqslant u \leqslant t + x$. Then, by Doob's inequality, we have

$$\begin{aligned}
&\mathbb{E}\left[\sup_{u\in[0,\tau_H],t\leqslant u\leqslant t+x}(\bar{\mathbb{B}}_n(u)-\bar{\mathbb{B}}_n(t))^2\right]\\
&\leqslant 4\mathbb{E}\left[\left(\bar{\mathbb{B}}_n\left((t+x)\wedge\tau_H\right)-\bar{\mathbb{B}}_n(t)\right)^2\right]\\
&= 4\mathbb{E}\left[\left(\bar{\mathbb{B}}_n\left((t+x)\wedge\tau_H\right)\right)^2-\left(\bar{\mathbb{B}}_n(t)\right)^2\right]\\
&= \frac{4}{n}\mathbb{E}\left[\int_t^{(t+x)\wedge\tau_H}\Phi_n(s;\beta_0)\lambda_0(s)\,ds\right]\\
&\leqslant \frac{4\lambda_0(\tau_H)x}{n}\mathbb{E}\left[\Phi_n(0;\beta_0)\right]\\
&= \frac{4\lambda_0(\tau_H)x}{n}\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left[e^{\beta_0' Z_i}\right]\leqslant\frac{Kx}{n}
\end{aligned}$$

for some $K > 0$, using (A2). For the case $t - x \leqslant u \leqslant t$, we can write

$$\begin{aligned}
&\mathbb{E}\left[\sup_{u\in[0,\tau_H],t-x\leqslant u\leqslant t}\left(\bar{\mathbb{B}}_n(u)-\bar{\mathbb{B}}_n(t)\right)^2\right]\\
&= \mathbb{E}\left[\sup_{0\vee(t-x)\leqslant u\leqslant t}\left(\bar{\mathbb{B}}_n(u)-\bar{\mathbb{B}}_n(t)\right)^2\right]\\
&\leqslant 2\mathbb{E}\left[\left(\bar{\mathbb{B}}_n(t)-\bar{\mathbb{B}}_n(0\vee(t-x))\right)^2\right]\\
&\quad+2\mathbb{E}\left[\sup_{0\vee(t-x)\leqslant u<t}\left(\bar{\mathbb{B}}_n(u)-\bar{\mathbb{B}}_n(0\vee(t-x))\right)^2\right].
\end{aligned}$$

Then similar, the right hand side is bounded by

$$\begin{aligned}
&2\mathbb{E}\left[\left(\bar{\mathbb{B}}_n(t)-\bar{\mathbb{B}}_n((t-x)_+)\right)^2\right]+8\mathbb{E}\left[\left(\bar{\mathbb{B}}_n(t)-\bar{\mathbb{B}}_n((t-x)_+)\right)^2\right]\\
&= 10\mathbb{E}\left[\bar{\mathbb{B}}_n(t)^2-\bar{\mathbb{B}}_n((t-x)_+)^2\right]=\frac{10}{n}\mathbb{E}\left[\int_{(t-x)_+}^{t}\Phi_n(s;\beta_0)\lambda_0(s)\,ds\right]\\
&\leqslant\frac{10\lambda_0(\tau_H)x}{n}\mathbb{E}\left[\Phi_n(0;\beta_0)\right]\leqslant\frac{Cx}{n},
\end{aligned}$$

for some $C > 0$. This concludes the proof. □

Lemma A.2.5. *Suppose that (A1)–(A2) hold. Let* $0 < \epsilon < M < \tau_H$ *and let* $\hat{U}_n(a)$ *and* U *be defined in* (3.2.30) *and* (A.2.9). *Suppose that* λ_0' *is uniformly bounded below by a strictly positive constant. Then, there exists an event* E_n, *such*

that $\mathbb{1}_{E_n} \to 1$ *in probability, and a constant* K *such that, for every* $a \geqslant 0$ *and* $x > 0$,

$$\mathbb{P}\left(\left\{|\hat{U}_n(a) - U(a)| \geqslant x\right\} \cap E_n \cap \left\{\epsilon \leqslant \hat{U}_n(a) \leqslant M\right\}\right) \leqslant \frac{K}{nx^3}. \tag{A.2.11}$$

Proof. Similar to the proof of Lemma 3.6.3, we start by writing

$$\begin{aligned}&\mathbb{P}\left(\left\{|\hat{U}_n(a) - U(a)| \geqslant x\right\} \cap E_n \cap \left\{\epsilon \leqslant \hat{U}_n(a) \leqslant M\right\}\right)\\ &= \mathbb{P}\left(\left\{U(a) + x \leqslant \hat{U}_n(a) \leqslant M\right\} \cap E_n\right) + \mathbb{P}\left(\left\{\epsilon \leqslant \hat{U}_n(a) \leqslant U(a) - x\right\} \cap E_n\right).\end{aligned} \tag{A.2.12}$$

The first probability is zero if $U(a) + x > M$. Otherwise, if $U(a) + x \leqslant M$, then $x \leqslant M$ and we get

$$\begin{aligned}&\mathbb{P}\left(\left\{U(a) + x \leqslant \hat{U}_n(a) \leqslant M\right\} \cap E_n\right) \leqslant\\ &\mathbb{P}\left(\left\{\inf_{y \in [U(a)+x, M]} \left(V_n(y) - a\hat{W}_n(y) - V_n(U(a)) + a\hat{W}_n(U(a))\right) \leqslant 0\right\} \cap E_n\right).\end{aligned}$$

Define

$$\bar{R}_n(t) = a\int_0^t \left(\Phi_n(s;\beta_0) - \Phi_n(s;\hat{\beta}_n)\right)\, ds. \tag{A.2.13}$$

Then, for $T_{(1)} < U(a) < y$,

$$\begin{aligned}&\bar{\mathbb{B}}_n(y) - \bar{\mathbb{B}}_n(U(a)) + \bar{R}_n(y) - \bar{R}_n(U(a))\\ &= V_n(y) - V_n(U(a)) - \int_{U(a)}^y \Phi_n(s;\beta_0)\lambda_0(s)\, ds\\ &\quad + a\int_{U(a)}^y \left(\Phi_n(s;\beta_0) - \Phi_n(s;\hat{\beta}_n)\right)\, ds\\ &= V_n(y) - V_n(U(a)) - a\int_{U(a)}^y \Phi_n(s;\hat{\beta}_n)\, ds - \int_{U(a)}^y \Phi_n(s;\beta_0)\lambda_0(s)\, ds\\ &\quad + a\int_{U(a)}^y \Phi_n(s;\beta_0)\, ds\\ &= V_n(y) - a\hat{W}_n(y) - V_n(U(a)) + a\hat{W}_n(U(a))\\ &\quad - \int_{U(a)}^y \Phi_n(s;\beta_0)(\lambda_0(s) - a)\, ds.\end{aligned}$$

On the event E_n, by Taylor expansion we find that

$$\begin{aligned}
&\int_{U(a)}^{y} \Phi_n(s;\beta_0)(\lambda_0(s)-a)\,ds \\
&= \int_{U(a)}^{y} \Phi_n(s;\beta_0)(\lambda_0(s)-\lambda_0(U(a)))\,ds \\
&= \int_{U(a)}^{y} \Phi_n(s;\beta_0)\left(\lambda_0'(\xi_s)(s-U(a))\right)\,ds \\
&\geqslant \inf_{t\in[0,\tau_H)} \lambda_0'(t)\left(\Phi(M;\beta_0)-\xi_4 n^{-1/3}\right)\frac{1}{2}(y-U(a))^2 \\
&\geqslant c(y-U(a))^2
\end{aligned}$$

for some $c>0$. Similar to the proof of Lemma 3.6.3, it follows that

$$\begin{aligned}
&\mathbb{P}\left(\left\{\inf_{y\in[U(a)+x,M]}\left(V_n(y)-a\hat{W}_n(y)-V_n(U(a))+a\hat{W}_n(U(a))\right)\leqslant 0\right\}\cap E_n\right) \\
&\leqslant \mathbb{P}\left(\left\{\inf_{y\in[U(a)+x,M]}\left(\bar{\mathbb{B}}_n(y)-\bar{\mathbb{B}}_n(U(a))+\bar{R}_n(y)-\bar{R}_n(U(a))\right.\right.\right. \\
&\qquad\qquad \left.\left.\left.+c(y-U(a))^2\right)\leqslant 0\right\}\cap E_n\right).
\end{aligned}$$

Hence, as before

$$\begin{aligned}
&\mathbb{P}\left(\left\{U(a)+x\leqslant \hat{U}_n(a)\leqslant M\right\}\cap E_n\right)\leqslant \\
&\sum_{k=0}^{i}\mathbb{P}\left(\left\{\sup_{I_k}\left(\left|\bar{\mathbb{B}}_n(y)-\bar{\mathbb{B}}_n(U(a))\right|+\left|\bar{R}_n(y)-\bar{R}_n(U(a))\right|\right)\geqslant cx^2 2^{2k}\right\}\cap E_n\right)
\end{aligned}$$

where the supremum runs over $y\leqslant M$, such that $y-U(a)\in[x2^k,x2^{k+1})$. With Markov, we can bound this probability by

$$\begin{aligned}
&4\sum_{k=0}^{i}\left(c^2x^4 2^{4k}\right)^{-1}\mathbb{E}\left[\sup_{\substack{y\leqslant M\\ y-U(a)\in[x2^k,x2^{k+1})}}\left|\bar{\mathbb{B}}_n(y)-\bar{\mathbb{B}}_n(U(a))\right|^2\right] \\
&+8\sum_{k=0}^{i}\left(c^3x^6 2^{6k}\right)^{-1}\mathbb{E}\left[\sup_{\substack{y\leqslant M\\ y-U(a)\in[x2^k,x2^{k+1})}}\mathbb{1}_{E_n}\left|\bar{R}_n(y)-\bar{R}_n(U(a))\right|^3\right].
\end{aligned} \tag{A.2.14}$$

As in the proof of Lemma 3.6.3, we will bound both expectations separately. We have

$$\begin{aligned}
&\mathbb{E}\left[\sup_{y<M,\, y-U(a)\in[x2^k,x2^{k+1})} \mathbb{1}_{E_n}\left|\bar{R}_n(y)-\bar{R}_n(U(a))\right|^3\right]\\
&\leqslant \mathbb{E}\left[\mathbb{1}_{E_n}\left(\int_{U(a)}^{(U(a)+x2^{k+1})\wedge M} a\left|\Phi_n(s;\beta_0)-\Phi_n(s;\hat{\beta}_n)\right|\,ds\right)^3\right]\\
&\leqslant x^3 2^{3(k+1)}\lambda_0(M)^3\mathbb{E}\left[\mathbb{1}_{E_n}\sup_{s\in[0,M]}\left|\Phi_n(s;\beta_0)-\Phi_n(s;\hat{\beta}_n)\right|^3\right]\\
&\leqslant x^3 2^{3(k+1)}\lambda_0(M)^3\mathbb{E}\left[\mathbb{1}_{E_n}|\hat{\beta}_n-\beta_0|^3\sup_{x\in\mathbb{R}}\left|D_n^{(1)}(\beta^*;x)\right|^3\right]\\
&\leqslant x^3 2^{3(k+1)}\lambda_0(M)^3\frac{L^3\xi_2^{3/2}}{n}\leqslant\frac{Cx^3 2^{3(k+1)}}{n},
\end{aligned}$$

for some $C>0$.

To bound the first expectation in (A.2.14), we use Lemma A.2.4 and we can argue as in the proof of Lemma 3.6.3 to obtain

$$\mathbb{P}\left(\left\{U(a)+x\leqslant\hat{U}_n(a)\leqslant M\right\}\cap E_n\right)\leqslant\frac{K}{nx^3}.$$

We can deal in the same way as in the proof of Lemma 3.6.3 with the second probability on the right hand side of (A.2.12), using the properties of $\bar{B}_n$ and $\bar{R}_n$. □

Note that on the event E_n from Lemma A.2.1, similar to (3.6.15), we have

$$\sup_{x\in\mathbb{R}}|\Phi_n(x;\hat{\beta}_n)-\Phi(x;\beta_0)|\leqslant\frac{C_\Phi}{n^{1/3}},\tag{A.2.15}$$

where $C_\Phi=\sqrt{\xi_2}L+\xi_4$, with L the upper bound from (3.6.2).

Lemma A.2.6. *Suppose that (A1)–(A2) hold. Take $0<\epsilon<\epsilon'<M'<M<\tau_H$. Let $\hat{\lambda}_n$ be the maximum likelihood estimator of a nondecreasing baseline hazard rate λ_0, which is differentiable with λ_0' uniformly bounded above and below by strictly positive constants. Let E_n be the event from Lemma A.2.1. Take $\xi_2>0$ and $\xi_4>0$ in* (3.6.3) *sufficiently small, such that*

$$C_\Phi<\frac{\Phi(M;\beta_0)}{2\lambda_0(M)}\min\left\{\epsilon'-\epsilon,M-M'\right\}\inf_{t\in[0,\tau_H]}\lambda_0'(t).\tag{A.2.16}$$

and take ξ_3 *in* (A.2.2) *sufficiently small, such that*

$$\xi_3 \leqslant \frac{1}{4}\left\{\frac{(M-M')}{2}\inf_{t\in[0,\tau_H]}\lambda_0'(t) - \frac{C_\Phi}{\Phi(M;\beta_0)}\lambda_0(M)\right\}(M-M')\Phi(M;\beta_0). \tag{A.2.17}$$

Then, there exists a constant C *such that, for each* $n \in \mathbb{N}$,

$$\sup_{t\in[\epsilon',M']} \mathbb{E}\left[n^{2/3}\mathbb{1}_{E_n}\left(\lambda_0(t)-\hat{\lambda}_n(t)\right)^2\right] \leqslant C.$$

Proof. It is sufficient to prove that there exist some constants C_1, C_2 such that for each $n \in \mathbb{N}$ and each $t \in [\epsilon', M']$, we have

$$\mathbb{E}\left[n^{2/3}\mathbb{1}_{E_n}\left\{(\hat{\lambda}_n(t)-\lambda_0(t))_+\right\}^2\right] \leqslant C_1, \tag{A.2.18}$$

$$\mathbb{E}\left[n^{2/3}\mathbb{1}_{E_n}\left\{(\lambda_0(t)-\hat{\lambda}_n(t))_+\right\}^2\right] \leqslant C_2. \tag{A.2.19}$$

Lets first consider (A.2.18). Then as in the proof of Lemma 3.6.4 we have

$$\begin{aligned}&\mathbb{E}\left[n^{2/3}\mathbb{1}_{E_n}\left\{(\hat{\lambda}_n(t)-\lambda_0(t))_+\right\}^2\right]\\ &\leqslant 4\eta^2 + 4\int_\eta^\infty \mathbb{P}\left(n^{1/3}\mathbb{1}_{E_n}(\hat{\lambda}_n(t)-\lambda_0(t)) > x\right)x\,dx,\end{aligned}$$

for a fixed $\eta > 0$, where

$$\mathbb{P}\left(n^{1/3}\,\mathbb{1}_{E_n}\,(\hat{\lambda}_n(t)-\lambda_0(t)) > x\right) = \mathbb{P}\left(\{\hat{U}_n(a+n^{-1/3}x) < t\}\cap E_n\right).$$

We distinguish between the cases

$$a+n^{-1/3}x \leqslant \lambda_0(M) \quad \text{and} \quad a+n^{-1/3}x > \lambda_0(M),$$

where $a = \lambda_0(t)$. We prove that, in the first case, there exist a positive constant C such that for all $t \in (\epsilon, M']$, and $n \in \mathbb{N}$,

$$\mathbb{P}\{n^{1/3}\mathbb{1}_{E_n}(\hat{\lambda}_n(t)-\lambda_0(t)) > x\} \leqslant \frac{C}{x^3},$$

for all $x \geqslant \eta$, and in the second case $\mathbb{P}\left(n^{1/3}\mathbb{1}_{E_n}(\hat{\lambda}_n(t)-\lambda_0(t)) > x\right) = 0$. Then (A.2.18) follows immediately.

First assume $a+n^{-1/3}x > \lambda_0(M)$. Note that, if $\hat{\lambda}_n(t) > a+n^{-1/3}x$, then for each $y > t$, we have

$$V_n(y)-V_n(t) \geqslant \hat{\lambda}_n(t)\left(\hat{W}_n(y)-\hat{W}_n(t)\right) > \left(a+n^{-1/3}x\right)\left(\hat{W}_n(y)-\hat{W}_n(t)\right).$$

In particular for $y = \tilde{M} = M' + (M - M')/2$, we obtain

$$\begin{aligned}
&\mathbb{P}\left(n^{1/3}\mathbb{1}_{E_n}(\hat{\lambda}_n(t) - \lambda_0(t)) > x\right) \\
&\leqslant \mathbb{P}\Big(\Big\{V_n(\tilde{M}) - V_n(t) - \left(H^{uc}(\tilde{M}) - H^{uc}(t)\right) \\
&\qquad > \left(a + n^{-1/3}x\right)\left(\hat{W}_n(\tilde{M}) - \hat{W}_n(t)\right) - \left(H^{uc}(\tilde{M}) - H^{uc}(t)\right)\Big\} \cap E_n\Big) \\
&\leqslant \mathbb{P}\Bigg(\Bigg\{2 \sup_{s\in[T_{(1)},T_{(n)}]} |V_n(s) - H^{uc}(s)| \\
&\qquad > \int_t^{\tilde{M}} \left\{\left(a + n^{-1/3}x\right)\Phi_n(s;\hat{\beta}_n) - \lambda_0(s)\Phi(s;\beta_0)\right\} ds\Bigg\} \cap E_n\Bigg).
\end{aligned} \tag{A.2.20}$$

Note that according to (A.2.15), $\Phi_n(s;\hat{\beta}_n) - \Phi(s,\beta_0) \geqslant -C_\phi$, and that $a + n^{-1/3}x > \lambda_0(M) > \lambda_0(\tilde{M}) \geqslant \lambda_0(s)$. Therefore, since $C_\phi \leqslant \Phi(M;\beta_0)$, from (A.2.17), we have

$$\begin{aligned}
&\int_t^{\tilde{M}} \left\{\left(a + n^{-1/3}x\right)\Phi_n(s;\hat{\beta}_n) - \lambda_0(s)\Phi(s;\beta_0)\right\} ds \\
&\geqslant \Phi(M;\beta_0)\int_t^{\tilde{M}} \left\{\left(a + n^{-1/3}x\right)\left(1 - \frac{C_\phi}{\Phi(M;\beta_0)}\right) - \lambda_0(s)\right\} ds \\
&> \Phi(M;\beta_0)(\tilde{M} - M')\left(\lambda_0(M)\left(1 - \frac{C_\phi}{\Phi(M;\beta_0)}\right) - \lambda_0(\tilde{M})\right) \\
&\geqslant \frac{1}{2}\left\{\frac{M - M'}{2} \inf_{x\in[0,\tau_H]} \lambda_0'(x) - \frac{C_\phi}{\Phi(M;\beta_0)}\lambda_0(M)\right\}(M - M')\Phi(M;\beta_0) \geqslant 2\xi_3.
\end{aligned} \tag{A.2.21}$$

Hence, similar to (3.6.24) and (3.6.25), we conclude that the probability on the right hand side of (A.2.20) is zero.

Then, consider the case $a + n^{-1/3}x \leqslant \lambda_0(M)$. Similar to (3.6.22), from Lemma A.2.5, we have

$$\mathbb{P}\left(\left\{\epsilon \leqslant \hat{U}_n(a + n^{-1/3}x) < t\right\} \cap E_n\right) \leqslant \frac{C}{x^3}$$

for some $C > 0$. Moreover, for the case $\hat{U}_n(a + n^{-1/3}x) < \epsilon$, we find

$$\mathbb{P}\left(\left\{\hat{U}_n(a + n^{-1/3}x) < \epsilon\right\} \cap E_n\right) = \mathbb{P}\left(\left\{\hat{\lambda}_n(\epsilon) > a + n^{-1/3}x\right\} \cap E_n\right).$$

Note that, if $\hat{\lambda}_n(\epsilon) > a + n^{-1/3}x$, then for each $y > \epsilon$, we have

$$\begin{aligned}
V_n(y) - V_n(\epsilon) &\geqslant \hat{\lambda}_n(\epsilon)\left(\hat{W}_n(y) - \hat{W}_n(\epsilon)\right) \\
&> \left(a + n^{-1/3}x\right)\left(\hat{W}_n(y) - \hat{W}_n(\epsilon)\right).
\end{aligned}$$

In particular for $y = \tilde{\epsilon} = \epsilon + (\epsilon' - \epsilon)/2$, similar to (A.2.20), we obtain

$$\begin{aligned}
&\mathbb{P}\left(\left\{\hat{\lambda}_n(\epsilon) > a + n^{-1/3}x\right\} \cap E_n\right) \\
&\leqslant \mathbb{P}\left(\left\{V_n(\tilde{\epsilon}) - V_n(\epsilon) > \left(a + n^{-1/3}x\right)\left(\hat{W}_n(y) - \hat{W}_n(\epsilon)\right)\right\} \cap E_n\right) \\
&\leqslant \mathbb{P}\Bigg(\Bigg\{2 \sup_{s \in [T_{(1)}, T_{(n)}]} |V_n(s) - H^{uc}(s)| \\
&\qquad > \int_{\epsilon}^{\tilde{\epsilon}} \left\{\left(a + n^{-1/3}x\right)\Phi_n(s; \hat{\beta}_n) - \lambda_0(s)\Phi(s; \beta_0)\right\} ds\Bigg\} \cap E_n\Bigg).
\end{aligned} \tag{A.2.22}$$

Then, similar to (A.2.21) and using $\tilde{\epsilon} > \epsilon'$, from (A.2.17) we obtain

$$\begin{aligned}
&\int_{\epsilon}^{\tilde{\epsilon}} \left\{\left(a + n^{-1/3}x\right)\Phi_n(s; \hat{\beta}_n) - \lambda_0(s)\Phi(s; \beta_0)\right\} ds \\
&\geqslant \frac{1}{2}\left\{\frac{\epsilon' - \epsilon}{2} \inf_{x \in [0, \tau_H]} \lambda_0'(x) - \frac{C_\phi}{\Phi(M, \beta_0)}\lambda_0(\epsilon')\right\}(\epsilon' - \epsilon)\,\Phi(M; \beta_0) \geqslant 2\xi_3,
\end{aligned}$$

and we conclude that the probability on the right hand side of (A.2.22) is zero. This concludes the proof of (A.2.18).

We proceed with (A.2.19). Arguing as in the proof of (A.2.18), we obtain

$$\begin{aligned}
&\mathbb{E}\left[n^{2/3}\mathbb{1}_{E_n}\left\{(\lambda_0(t) - \hat{\lambda}_n(t))_+\right\}^2\right] \\
&\leqslant \eta^2 + 2\int_{\eta}^{\infty} \mathbb{P}\left(n^{1/3}\mathbb{1}_{E_n}\left(\lambda_0(t) - \hat{\lambda}_n(t)\right) \geqslant x\right) x\,dx,
\end{aligned}$$

for a fixed $\eta > 0$, where

$$\mathbb{P}\left(n^{1/3}\mathbb{1}_{E_n}\left(\lambda_0(t) - \hat{\lambda}_n(t)\right) \geqslant x\right) = \mathbb{P}\left(\left\{\hat{U}_n(a - n^{-1/3}x) \geqslant t\right\} \cap E_n\right),$$

where $a = \lambda_0(t)$. First consider the case $0 < a - n^{-1/3}x \leqslant \lambda_0(\epsilon)$. For each $y < t$, we have

$$V_n(t) - V_n(y) \leqslant \hat{\lambda}_n(t)\left(\hat{W}_n(t) - \hat{W}_n(y)\right) \leqslant \left(a - n^{-1/3}x\right)\left(\hat{W}_n(t) - \hat{W}_n(y)\right).$$

In particular, for $y = \tilde{\epsilon} = \epsilon + (\epsilon' - \epsilon)/2$, similar to (A.2.22), we obtain

$$\begin{aligned}
&\mathbb{P}\left(n^{1/3}\mathbb{1}_{E_n}\left(\lambda_0(t) - \hat{\lambda}_n(t)\right) \geqslant x\right) \\
&\leqslant \mathbb{P}\Bigg(\Bigg\{2 \sup_{s \in [T_{(1)}, T_{(n)}]} |V_n(s) - H^{uc}(s)| \\
&\qquad \geqslant \int_{\tilde{\epsilon}}^{t} \left\{\left(-a + n^{-1/3}x\right)\Phi_n(s; \hat{\beta}_n) + \lambda_0(s)\Phi(s; \beta_0)\right\} ds\Bigg\} \cap E_n\Bigg).
\end{aligned} \tag{A.2.23}$$

As before, using that $-a + n^{-1/3}x + \lambda_0(s) > 0$ and $t - \tilde{\epsilon} \geqslant= \frac{1}{2}(\epsilon' - \epsilon)$, similar to (A.2.21), from (A.2.17) we have

$$\begin{aligned}
&\int_{\tilde{\epsilon}}^{t} \left\{ \left(-a + n^{-1/3}x\right) \Phi_n(s;\hat{\beta}_n) + \lambda_0(s)\Phi(s;\beta_0) \right\} ds \\
&\geqslant \frac{1}{2} \left\{ \frac{\epsilon' - \epsilon}{2} \inf_{s\in[0,\tau_H]} \lambda_0'(s) - \frac{C_\phi}{\Phi(M;\beta_0)} \lambda_0(\epsilon) \right\} (\epsilon' - \epsilon)\Phi(M;\beta_0) \geqslant 2\xi_3.
\end{aligned} \tag{A.2.24}$$

and we conclude that the probability on the right hand side of (A.2.23) is zero.

Next, suppose that $a - n^{-1/3}x > \lambda_0(\epsilon)$ and consider

$$\mathbb{P}\left(\left\{ \hat{U}_n(a - n^{-1/3}x) \geqslant t \right\} \cap E_n \right).$$

In order to use Lemma A.2.5, we must intersect with the event

$$\{\epsilon \leqslant \hat{U}_n(a - n^{-1/3}x) \leqslant M\}.$$

Since $t \in [\epsilon', M']$, $\hat{U}_n(a - n^{-1/3}x) \geqslant t$ implies $\hat{U}_n(a - n^{-1/3}x) \geqslant \epsilon$. Using Lemma A.2.5 and the mean value theorem, we obtain

$$\begin{aligned}
&\mathbb{P}\left(\left\{ t \leqslant \hat{U}_n(a - n^{-1/3}x) \leqslant M \right\} \cap E_n \right) \\
&= \mathbb{P}\Bigg(\left\{ \hat{U}_n(a - n^{-1/3}x) - U(a - n^{-1/3}x) \geqslant t - U(a - n^{-1/3}x) \right\} \\
&\qquad\qquad \cap \left\{ \epsilon \leqslant \hat{U}_n(a - n^{-1/3}x) \leqslant M \right\} \cap E_n \Bigg) \\
&\leqslant \mathbb{P}\Bigg(\left\{ |\hat{U}_n(a - n^{-1/3}x) - U(a - n^{-1/3}x)| \geqslant t - U(a - n^{-1/3}x) \right\} \\
&\qquad\qquad \cap \left\{ \epsilon \leqslant \hat{U}_n(a - n^{-1/3}x) \leqslant M \right\} \cap E_n \Bigg) \\
&\leqslant \frac{K}{n\left\{t - U(a - n^{-1/3}x)\right\}^3} \leqslant \frac{K}{(U'(\xi_n))^3 x^3} \leqslant \frac{C}{x^3}
\end{aligned}$$

where $t = U(a)$, $\xi_n \in (a - n^{-1/3}x, a)$, and $U'(\xi_n) = 1/\lambda_0'(\lambda_0^{-1}(\xi_n))$ is bounded. Finally, note that

$$\mathbb{P}\left(\left\{ \hat{U}_n(a - n^{-1/3}x) > M \right\} \cap E_n \right) \leqslant \mathbb{P}\left(\left\{ \hat{\lambda}_n(M) \leqslant a - n^{-1/3}x \right\} \cap E_n \right).$$

If $\hat{\lambda}_n(M) \leqslant a - n^{-1/3}x$, then for each $y < M$, we have

$$\begin{aligned}
V_n(M) - V_n(y) &\leqslant \hat{\lambda}_n(M) \left(\hat{W}_n(M) - \hat{W}_n(y) \right) \\
&\leqslant \left(a - n^{-1/3}x \right) \left(\hat{W}_n(M) - \hat{W}_n(y) \right).
\end{aligned}$$

In particular for $y = \tilde{M} = M' + (M - M')/2$, similar to (A.2.23), we obtain

$$\begin{aligned}
&\mathbb{P}\left(\left\{\hat{\lambda}_n(M) \leqslant a - n^{-1/3}x\right\} \cap E_n\right) \\
&\leqslant \mathbb{P}\Bigg(\Bigg\{2 \sup_{s\in[T_{(1)},T_{(n)}]} |V_n(s) - H^{uc}(s)| \\
&\qquad \geqslant \int_{\tilde{M}}^{M} \left\{\left(a - n^{-1/3}x\right)\Phi_n(s;\hat{\beta}_n) - \lambda_0(s)\Phi(s;\beta_0)\right\} ds\Bigg\} \cap E_n\Bigg)
\end{aligned} \tag{A.2.25}$$

As before, similar to (A.2.24), from (A.2.17) we have

$$\begin{aligned}
&\int_{\tilde{M}}^{M} \left\{\left(a - n^{-1/3}x\right)\Phi_n(s;\hat{\beta}_n) - \lambda_0(s)\Phi(s;\beta_0)\right\} ds \\
&\geqslant \frac{1}{2}\left\{\frac{M - M'}{2} \inf_{s\in[0,\tau_H]} \lambda_0'(s) - \frac{C_\phi}{\Phi(M;\beta_0)}\lambda_0(M')\right\}(M - M')\Phi(M;\beta_0) \geqslant 2\xi_3.
\end{aligned}$$

and we conclude that the probability on the right hand side of (A.2.25) is zero. This concludes the proof. □

To establish the analogue of Lemma 3.2.4 for $\hat{\lambda}_n$ and to show that the distance between jump times of $\hat{\lambda}_n$ is of smaller order than b, similar to Lemma 3.6.5, we need a stronger version of Lemma A.2.6. As before, we loose a factor $n^{2/9}$ with respect to the bound in Lemma A.2.6, which might not be optimal, but suffices for our purposes.

Lemma A.2.7. *Suppose that (A1)–(A2) hold. Take $0 < \epsilon < \epsilon' < M' < M < \tau_H$. Let $\hat{\lambda}_n$ be the maximum likelihood estimator of a nondecreasing baseline hazard rate λ_0, which is differentiable with λ_0' uniformly bounded above and below by strictly positive constants. Let E_n be the event from Lemma A.2.1 and choose C_ϕ and ξ_3 such that they satisfy* (A.2.16) *and* (A.2.17)*, respectively. Then, there exists a constant $C > 0$ such that, for each $n \in \mathbb{N}$,*

$$\mathbb{E}\left[n^{4/9}\mathbb{1}_{E_n} \sup_{t\in[\epsilon',M']} \left(\lambda_0(t) - \hat{\lambda}_n(t)\right)^2\right] \leqslant C. \tag{A.2.26}$$

Proof. The proof is exactly the same as the proof of Lemma 3.6.5 (replacing $\tilde{\lambda}_n$ with $\hat{\lambda}_n$). □

Lemma A.2.8. *Under the assumption of Lemma A.2.7, if $\tau_1, \ldots, \tau_m$ are jump times of $\hat{\lambda}_n$ on the interval $[\epsilon', M']$ then*

$$\max_{i=1,\ldots,m-1} |\tau_i - \tau_{i+1}| = O_P(n^{-2/9}).$$

Proof. The proof is similar to that of Lemma 3.6.6. This time we use Lemma A.2.7 instead of Lemma 3.6.5. □

Lemma A.2.9. *Suppose that (A1)–(A2) hold. Fix* $t \in (0, \tau_H)$ *and take* $0 < \epsilon < \epsilon' < M' < M < \tau_H$. *Let* $\hat{\lambda}_n$ *be the maximum likelihood estimator of a nondecreasing baseline hazard rate* λ_0 *which is differentiable with* λ_0' *uniformly bounded above and below by strictly positive constants. Let* E_n *be the event from Lemma A.2.1 and choose* C_Φ *and* ξ_3 *such that they satisfy* (A.2.16) *and* (A.2.17), *respectively. Let* $\tau_{j_{n1}}, \tau_{j_{n2}}$ *be defined as in* (A.2.6). *Then*

$$\mathbb{1}_{E_n} \int_{\tau_{j_{n1}}}^{\tau_{j_{n2}}} (\lambda_0(u) - \hat{\lambda}_n(u))^2 \, du = O_p(bn^{-2/3}).$$

Proof. The proof is exactly the same as the proof of Lemma 3.6.7 (replacing $\tilde{\lambda}_n$ with $\hat{\lambda}_n$). □

We are now in the position to establish the analogue (A.2.7) of Lemma 3.2.3.

Lemma A.2.10. *Suppose that (A1)–(A2) hold. Fix* $t \in (0, \tau_H)$ *and let* $\theta_{n,t}$ *and* $\bar{\theta}_{n,t}$ *be defined by* (A.2.1) *and* (A.2.5), *respectively. Assume that* λ_0 *is differentiable, such that* λ_0' *is uniformly bounded above and below by strictly positive constants and let* k *satisfy* (1.2.1). *Then, it holds*

$$\int \left\{\bar{\theta}_{n,t}(u,\delta,z) - \theta_{n,t}(u,\delta,z)\right\} d\mathbb{P}(u,\delta,z) = O_p(b^{-1}n^{-2/3}).$$

Proof. Take $0 < \epsilon < \epsilon' < t < M' < M < \tau_H$ and consider n sufficiently large such that $[\tau_{j_{n1}}, \tau_{j_{n2}}] \subset [\epsilon', M']$. Similar to (3.6.29), we have

$$\begin{aligned}
&\int \left\{\bar{\theta}_{n,t}(u,\delta,z) - \theta_{n,t}(u,\delta,z)\right\} d\mathbb{P}(u,\delta) \\
&= \mathbb{1}_{E_n} \int_{\tau_{j_{n1}}}^{\tau_{j_{n2}}} \left(a_{n,t}(\hat{A}_n(u)) - a_{n,t}(u)\right) \left(\Phi(u;\beta_0)\lambda_0(u) - \Phi(u;\hat{\beta}_n)\hat{\lambda}_n(u)\right) du
\end{aligned}$$

so that by Cauchy-Schwarz inequality

$$\begin{aligned}
&\left|\int \left\{\bar{\theta}_{n,t}(u,\delta,z) - \theta_{n,t}(u,\delta,z)\right\} d\mathbb{P}(u,\delta,z)\right| \\
&\leqslant \mathbb{1}_{E_n} \left\|\left(a_{n,t} \circ \hat{A}_n - a_{n,t}\right) \mathbb{1}_{[\tau_{j_{n1}},\tau_{j_{n2}}]}\right\|_{\mathcal{L}_2} \left\|\left(\Phi_0\lambda_0 - \hat{\Phi}_n\hat{\lambda}_n\right) \mathbb{1}_{[\tau_{j_{n1}},\tau_{j_{n2}}]}\right\|_{\mathcal{L}_2},
\end{aligned} \tag{A.2.27}$$

where $\Phi_0(u) = \Phi(u;\beta_0)$ and $\hat{\Phi}_n(u) = \Phi_n(u;\hat{\beta}_n)$. Similar to (3.6.31),

$$\mathbb{1}_{E_n} \left\|(a_{n,t}(\hat{A}_n) - a_{n,t})\mathbb{1}_{[\tau_{j_{n1}},\tau_{j_{n2}}]}\right\|_{\mathcal{L}_2}^2 \leqslant \frac{c}{b^4} \mathbb{1}_{E_n} \int_{\tau_{j_{n1}}}^{\tau_{j_{n2}}} (\hat{A}_n(u) - u)^2 \, du,$$

for some constant c, and by the same reasoning as in (3.6.32), for $u \in [\tau_i, \tau_{i+1})$ and $\hat{A}_n(u) < \tau_{i+1}$, we obtain

$$|u - \hat{A}_n(u)| \leqslant 2K|\lambda_0(u) - \hat{\lambda}_n(u)|,$$

which also holds in the case $\hat{A}_n(u) = \tau_{i+1}$ simply because

$$|\lambda_0(u) - \lambda_0(\hat{A}_n(u))| \leqslant |\lambda_0(u) - \hat{\lambda}_n(u)|.$$

As a result, using Lemma A.2.9, we derive that

$$\begin{aligned}\mathbb{1}_{E_n} \frac{1}{b^4} \int_{\tau_{j_{n1}}}^{\tau_{j_{n2}}} \left(\hat{A}_n(u) - u\right)^2 du &\leqslant \frac{C}{b^4} \mathbb{1}_{E_n} \int_{\tau_{j_{n1}}}^{\tau_{j_{n2}}} \left(\lambda_0(u) - \hat{\lambda}_n(u)\right)^2 du \\ &= O_p(b^{-3} n^{-2/3}).\end{aligned}$$

The argument for second factor in (A.2.27) is the same as for (3.6.34), and yields

$$\mathbb{1}_{E_n} \left\| \left(\Phi_0 \lambda_0 - \widehat{\Phi} \tilde{\lambda}_n\right) \mathbb{1}_{[\tau_{j_{n1}}, \tau_{j_{n2}}]} \right\|_{\mathcal{L}_2} = O_p(b^{1/2} n^{-1/3}).$$

Together with (A.2.27), this concludes the proof. □

Lemma A.2.11. *Suppose that (A1)–(A2) hold. Fix $t \in (0, \tau_H)$ and take $0 < \epsilon < \epsilon' < t < M' < M < \tau_H$. Assume that λ_0 is differentiable, and such that λ_0' is uniformly bounded above and below by strictly positive constants. Assume that $x \mapsto \Phi(x; \beta_0)$ is differentiable with a bounded derivative in a neighborhood of t. Let $\bar{\theta}_{n,t}$ be defined in* (A.2.5) *and let $\eta_{n,t}$ be defined by* (3.2.20)*, where E_n is the event from Lemma A.2.1. Let k satisfy* (1.2.1)*. Then, it holds*

$$\begin{aligned}&\int \left\{\bar{\theta}_{n,t}(u, \delta, z) - \eta_{n,t}(u, \delta, z)\right\} d(\mathbb{P}_n - \mathbb{P})(u, \delta, z) \\ &= O_p(b^{-3/2} n^{-13/18}) + O_p(n^{-1/2}).\end{aligned} \tag{A.2.28}$$

Proof. Let n be sufficiently large, such that $\epsilon' < \tau_{j_{n1}} < \tau_{j_{n2}} < M'$. Denote by R_n the left hand side of (A.2.28) and write $R_n = R_{n1} + R_{n2}$, where

$$R_{n1} = n^{-1/2} \mathbb{1}_{E_n} \mathbb{1}_{[\tau_{j_{n1}}, \tau_{j_{n2}}]}(u) \int \delta \{\bar{a}_{n,t}(u) - a_{n,t}(u)\} \, d\sqrt{n}(\mathbb{P}_n - \mathbb{P})(u, \delta, z),$$

$$\begin{aligned}R_{n2} = n^{-1/2} \mathbb{1}_{E_n} \int \mathbb{1}_{\{u > \tau_{j_{n1}}\}} \Bigg\{ &e^{\hat{\beta}_n' z} \int_{\tau_{j_{n1}}}^{u \wedge \tau_{j_{n2}}} \bar{a}_{n,t} n(u) \, d\hat{\Lambda}_n(v) \\ &- e^{\beta_0' z} \int_{t-b}^{u \wedge (t+b)} a_{n,t}(v) \, d\Lambda_0(v) \Bigg\} d\sqrt{n}(\mathbb{P}_n - \mathbb{P})(u, \delta, z).\end{aligned}$$

Choose $\eta > 0$. Consider similar two events as in (3.6.36):

$$\begin{aligned} \mathcal{A}_{n1} &= \left\{\hat{\lambda}_n(M) > K_1\right\}, \\ \mathcal{A}_{n2} &= \left\{\sup_{s\in[\epsilon',M']} \left|\lambda_0(s) - \hat{\lambda}_n(s)\right| > K_2 n^{-2/9}\right\}, \end{aligned} \tag{A.2.29}$$

where $K_1, K_2 > 0$, and let $\mathcal{A}_n = \mathcal{A}_{n1} \cup \mathcal{A}_{n2}$. From Lemma A.2.7 and the fact that $\hat{\lambda}_n(M) = O_p(1)$, it follows that we can choose $K_1, K_2 > 0$ such that $\mathbb{P}(\mathcal{A}_n) \leqslant 2\eta/3$. As in the proof of Lemma 3.2.4, it suffices to show that there exists $\nu > 0$, such that $b^{3/2}n^{13/18}\nu^{-1}\mathbb{E}\left[|R_{n1}|\mathbb{1}_{\mathcal{A}_n^c}\right] \leqslant \eta/3$ and $n^{1/2}\nu^{-1}\mathbb{E}\left[|R_{n2}|\mathbb{1}_{\mathcal{A}_n^c}\right] \leqslant \eta/3$, for all n sufficiently large.

Let us first consider R_{n1}. We have

$$\begin{aligned} a_{n,t}(\hat{A}_n(u)) - a_{n,t}(u) = &\frac{k_b(t-\hat{A}_n(u)) - k_b(t-u)}{\Phi(\hat{A}_n(u);\beta_0)} \\ &+ k_b(t-u)\frac{\Phi(u;\beta_0) - \Phi(\hat{A}_n(u);\beta_0)}{\Phi(\hat{A}_n(u);\beta_0)\Phi(u;\beta_0)}. \end{aligned} \tag{A.2.30}$$

Similar to (3.6.38),

$$\left|k_b(t-\hat{A}_n(u)) - k_b(t-u)\right| \leqslant b^{-2}n^{-2/9}K_2 \sup_{x\in[-1,1]} |k'(x)|,$$

for some $K_2 > 0$, and similarly, using that $x \mapsto \Phi(x;\beta_0)$ is differentiable with bounded derivative in a neighborhood of t,

$$b^{-1}|\Phi(u;\beta_0) - \Phi(\hat{A}_n(u);\beta_0)| \leqslant Kb^{-1}\,|\hat{A}_n(u) - u| \leqslant b^{-1}n^{-2/9}KK_3,$$

Consequently, as in the proof of Lemma 3.2.4, on the event $\mathcal{A}_n^c$ we can write R_{n1} as

$$R_{n1} = \mathbb{1}_{E_n} b^{-2}n^{-13/18}\int \mathbb{1}_{[t-2b,t+2b]}(u)\delta W_n(u)\, d\sqrt{n}(\mathbb{P}_n - \mathbb{P})(u,\delta,z),$$

where W_n is a function of bounded variation, uniformly bounded. Completely similar to the proof of Lemma 3.2.4, together with Lemma A.1.1 we find that

$$b^2 n^{13/18}\nu^{-1}\mathbb{E}\left[|R_{n1}|\mathbb{1}_{\mathcal{A}_n^c}\right] \leqslant \frac{K''}{\nu n^{2/90}} \leqslant \eta/3,$$

for sufficiently large ν. For R_{n2} we write

$$n^{1/2}R_{n2} = \mathbb{1}_{E_n}\int \left(e^{\hat{\beta}_n' z}r_{1,n}(u) - e^{\beta_0' z}r_{2,n}(u)\right)\, d\sqrt{n}(\mathbb{P}_n - \mathbb{P})(u,\delta,z),$$

where

$$r_{1,n}(u) = \mathbb{1}_{\{u>\tau_{j_{n1}}\}} \int_{\tau_{j_{n1}}}^{u\wedge\tau_{j_{n2}}} a_{n,t}(\hat{A}_n(v))\,\hat{\lambda}_n(v)\,dv,$$

$$r_{2,n}(u) = \mathbb{1}_{\{u>t-b\}} \int_{t-b}^{u\wedge(t+b)} a_{n,t}(v)\,\lambda_0(v)\,dv,$$

are both monotone functions, uniformly bounded by some constant C on the event $\mathcal{A}_n^c$. Once more, from here we follow exactly the same proof as the one for Lemma A.1.1. □

Theorem A.2.12. *Suppose that (A1)–(A2) hold. Fix* $t \in (0, \tau_H)$. *Assume that* λ_0 *is* $m \geqslant 2$ *times differentiable in* t, *such that* λ_0' *is uniformly bounded above and below by strictly positive constants. Moreover, that* $x \mapsto \Phi(x;\beta_0)$ *is differentiable with a bounded derivative in a neighborhood of* t, *and let* k *satisfy* (1.2.1). *Let* $\hat{\lambda}_n^{SM}$ *be defined in* (3.2.2) *and assume that* $n^{1/(2m+1)}b \to c > 0$. *Then, it holds*

$$n^{m/(2m+1)}\left(\hat{\lambda}_n^{SM}(t) - \lambda_0(t)\right) \xrightarrow{d} N(\mu, \sigma^2),$$

where

$$\mu = \frac{(-c)^m}{m!}\lambda_0^{(m)}(t)\int k(u)u^m\,du \quad \textit{and} \quad \sigma^2 = \frac{\lambda_0(t)}{c\Phi(t;\beta_0)}\int k^2(u)\,du.$$

Proof. The proof is completely analogous to that of Theorem 3.2.5 and is based on a similar decomposition as in (3.2.24). After using Lemmas A.2.1, A.2.2, A.2.10, and A.2.11, it remains to obtain the limit of (3.2.27), where E_n is the event from Lemma A.2.1. This is completely similar to the argument in the proof of Theorem 3.2.5. □

A.3 CONSISTENCY OF THE BOOTSTRAP

Instead of $\mathbb{P}_n$, we consider $\mathbb{P}_n^*$, the empirical measure corresponding to the bootstrap sample (T_1^*, Δ_1^*, Z_1),..., (T_n^*, Δ_n^*, Z_n), and instead of $\mathbb{P}$, we consider P_n^*, the measure corresponding to the bootstrap distribution of $(T^*, \Delta^*, Z) = (\min(X^*, C^*), \mathbb{1}_{\{X^*\leqslant C^*\}}, Z)$ conditional on the data (T_1, Δ_1, Z_1), ... ,(T_n, Δ_n, Z_n), where X^* conditional on Z has distribution function $\hat{F}_n(x \mid Z)$ and C^* has distribution function $\hat{G}_n$. To prove (3.5.4), we mimic the proof of Theorem 3.2.5, which means that one needs to establish the bootstrap versions of Lemmas 3.2.1-3.2.4.

In view of Remark 3.2.6, let $\hat{\beta}_n$ be an estimate for β_0 satisfying (3.2.35). Let $\hat{\beta}_n^*$ be the bootstrap version and suppose that $\hat{\beta}_n^* - \hat{\beta}_n \to 0$, for almost all sequences (T_i^*, Δ_i^*, Z_i), $i = 1, 2, \ldots$, conditional on the sequence

(T_i, Δ_i, Z_i), $i = 1, 2, \ldots$, and that $\sqrt{n}(\hat{\beta}_n^* - \hat{\beta}_n) = O_p^*(1)$, meaning that for all $\epsilon > 0$, there exists $M > 0$ such that

$$\limsup_{n\to\infty} P_n^* \left(\sqrt{n}|\hat{\beta}_n^* - \hat{\beta}_n| > M\right) < \epsilon, \qquad \mathbb{P} - \text{almost surely.}$$

Then, similar to (3.2.9) and(3.1.2), define

$$a_{n,t}^*(u) = \frac{k_b(t-u)}{\Phi^*(u;\hat{\beta}_n)} \quad \text{and} \quad \Phi^*(t;\hat{\beta}_n) = \int \mathbb{1}_{\{u \geqslant t\}} e^{\hat{\beta}_n' z} \, dP_n^*(u,\delta,z).$$

and let

$$\theta_{n,t}^*(u,\delta,z) = \mathbb{1}_{E_n^*} \left\{ \delta \, a_{n,t}^*(u) - e^{(\hat{\beta}_n^*)' z} \int_0^u a_{n,t}^*(v) \, d\hat{\Lambda}_n^*(v) \right\}.$$

Here $\hat{\Lambda}_n^*$ is the greatest convex minorant of the bootstrap Breslow estimator

$$\Lambda_n^*(t) = \int \frac{\delta \mathbb{1}_{\{u \leqslant t\}}}{\Phi_n^*(u;\hat{\beta}_n^*)} \, d\mathbb{P}_n^*(u,\delta,z),$$

with

$$\Phi_n^*(t;\beta) = \int \mathbb{1}_{\{u \geqslant t\}} e^{\beta' z} \, d\mathbb{P}_n^*(u,\delta,z),$$

and E_n^* is an event such that $\mathbb{1}_{E_n^*} = 1 + o_p^*(1)$, meaning that for all $\epsilon > 0$,

$$\limsup_{n\to\infty} P_n^*(|\mathbb{1}_{E_n^*} - 1| > \epsilon) = 0, \qquad \mathbb{P} - \text{almost surely.}$$

To obtain the bootstrap equivalent of Lemma 3.2.1, we first show that

$$\Lambda_0^*(t) := \int \frac{\delta \mathbb{1}_{\{u \leqslant t\}}}{\Phi^*(u;\hat{\beta}_n)} \, dP_n^*(u,\delta,z) = \Lambda_n^s(t).$$

For constructing of the event E_n^*, we prove that $\sqrt{n}\sup_{t\in[0,M]} |\hat{\Lambda}_n^*(t) - \Lambda_n^s(t)|$ and $\sqrt{n}\sup_{x\in\mathbb{R}} |\Phi_n^*(x;\hat{\beta}_n) - \Phi^*(x;\hat{\beta}_n)|$ are of the order $O_p^*(1)$. This yields the bootstrap equivalent of Lemma 3.2.1:

$$\int \theta_{n,t}^*(u,\delta,z) \, dP_n^*(u,\delta,z) = -\mathbb{1}_{E_n^*} \int k_b(t-u) \, d(\hat{\Lambda}_n^* - \Lambda_n^s)(u) + O_p^*(n^{-1/2}).$$

The proof of the bootstrap version of Lemma 3.2.2 is completely the same as that of Lemma 3.2.2. The bootstrap versions of Lemmas 3.6.1-3.6.7, which are preparatory for the bootstrap version of Lemma 3.2.3, can be obtained by means of similar arguments. This requires a suitable martingale that approximates the process $\Lambda_n^* - \Lambda_n^s$. To this end we define

$$\mathbb{M}_n^*(t) = \sum_{i=1}^n \left(\mathbb{1}_{\{X_i^* \leqslant t\}} \Delta_i^* - \int_0^t Y_i^*(u) e^{\hat{\beta}_n' Z_i} \, d\Lambda_n^s(u) \right)$$

and

$$\mathbb{B}_n^*(t) = \int_0^{t\wedge M} \frac{1}{n\Phi^*(s;\hat{\beta}_n)}\, d\mathbb{M}_n^*(s),$$

where the latter can be shown to be a mean zero square integrable martingale, that satisfies a bound similar to one in Lemma 3.6.2. Similar to Lemma 3.6.3, this leads to a suitable bound on the tail probabilities of the bootstrap inverse process, defined for $a \in [\tilde{\lambda}_n^s(\epsilon), \tilde{\lambda}_n^s(M)]$, for $0 < \epsilon < M < \tau_H$, by

$$U_n^*(a) = \begin{cases} \epsilon & a < \tilde{\lambda}_n^s(\epsilon); \\ (\lambda_n^s)^{-1}(a) & a \in [\tilde{\lambda}_n^s(\epsilon), \tilde{\lambda}_n^s(M)]; \\ M & a > \tilde{\lambda}_n^s(M). \end{cases} \tag{A.3.1}$$

This enables us to obtain L_2-bounds similar to Lemmas 3.6.1 and 3.6.2,

$$\sup_{t\in[\epsilon',M']} \mathbb{E}^* \left[n^{2/3} \mathbb{1}_{E_n^*} \left(\tilde{\lambda}_n^s(t) - \tilde{\lambda}_n^*(t)\right)^2 \right] \leqslant C;$$

$$\mathbb{1}_{E_n^*} \int_{t-b}^{t+b} (\tilde{\lambda}_n^s(u) - \tilde{\lambda}_n^*(u))^2 dt = O_p^*(bn^{-2/3}),$$

for $0 < \epsilon < \epsilon' < M' < M < \tau_H$, where $\mathbb{E}^*$ denotes the expectation with respect to P_n^*. Moreover, since the proof of Lemma 3.2.3 makes use of the derivative of $k_b(t-y)/\Phi(y;\beta_0)$, differentiation of its bootstrap counterpart $k_b(t-y)/\Phi^*(y;\hat{\beta}_n)$ has to be circumvented. This is done by a suitable differentiable approximation of $\Phi^*(y;\hat{\beta}_n)$, and we then obtain the bootstrap version of Lemma 3.2.3:

$$\int \left\{ \overline{\theta}_{n,t}^*(u,\delta,z) - \theta_{n,t}^*(u,\delta,z) \right\} dP_n^*(u,\delta,z) = O_p^*(b^{-1}n^{-2/3}),$$

Finally, after proving the bootstrap version of Lemma 3.6.5, i.e.,

$$\mathbb{E}^* \left[n^{4/9} \mathbb{1}_{E_n^*} \sup_{t\in(\epsilon',M']} \left(\tilde{\lambda}_n^s(t) - \tilde{\lambda}_n^*(t)\right)^2 \right] \leqslant C,$$

we obtain the bootstrap version of Lemma 3.2.4 for

$$\eta_{n,t}^*(u,\delta,z) = \mathbb{1}_{E_n^*} \left(\delta\, a_{n,t}^*(u) - e^{\hat{\beta}_n' z} \int_0^u a_{n,t}^*(v)\, d\Lambda_n^s(v) \right), \quad u \in [0,\tau_H],$$

by using arguments similar to those in the proof of Lemma 3.2.4. Next, the proof of (3.5.4) for $\tilde{\lambda}_n^{SG,*}$ is the same as that of Theorem 3.2.5 for $\tilde{\lambda}_n^{SG}$.

B

SUPPLEMENT TO: ON THE L_p-ERROR OF SMOOTH ISOTONIC ESTIMATORS

B.1 KERNEL ESTIMATOR OF A DECREASING FUNCTION

Lemma B.1.1. *Let* $l(t)$ *be a differentiable function on* $[0,1]$ *such that* $\inf_{[0,1]} l(t) > 0$ *and* $\sup_{[0,1]} |l'(t)| < \infty$. *Define* $L(t) = \int_0^t l(u)\,du$ *and let* $\Gamma_n^{(1)}$ *be as in* (6.2.3). *Assume that (A1) and (A3) hold. Then*

$$(b\gamma^2(p))^{-1/2}\left\{\int_b^{1-b}\left|b^{-1/2}\Gamma_n^{(1)}(t)+g_{(n)}(t)\right|^p d\mu(t)-m_n^c(p)\right\}\xrightarrow{d}N(0,1),$$

where $\gamma^2(p)$, $g_{(n)}$ *and* $m_n^c(p)$ *are defined respectively in* (6.3.22), (6.1.4) *and* (6.1.8).

Proof. With a change of variable we can write

$$\begin{aligned}
&\int_b^{1-b}\left|b^{-1/2}\Gamma_n^{(1)}(t)+g_{(n)}(t)\right|^p d\mu(t)-m_n^c(l,p)\\
&=b\int_1^{(1-b)/b}\left\{\left|b^{-1/2}\int_{t-1}^{t+1}k(t-y)\,dW(L(by))+g_{(n)}(tb)\right|^p w(tb)\right.\\
&\qquad\qquad\left.-\int_{\mathbb{R}}\left|l(tb)Dx+g_{(n)}(tb)\right|^p\phi(x)\,dx\right\}dt\\
&=b\left\{\sum_{i=1}^{M_1-1}\xi_i+\eta\right\},
\end{aligned}$$

where $M_1=[1/b-1]$,

$$\begin{aligned}
\xi_i=\int_i^{i+1}&\left\{\left|b^{-1/2}\int_{t-1}^{t+1}k(t-y)\,dW(L(by))+g_{(n)}(tb)\right|^p\right.\\
&\left.-\int_{-\infty}^{+\infty}\left|l(tb)Dx+g_{(n)}(tb)\right|^p\phi(x)\,dx\right\}w(tb)\,dt
\end{aligned}\tag{B.1.1}$$

and

$$\eta = \int_{M_1}^{(1-b)/b} \left\{ \left| b^{-1/2} \int_{t-1}^{t+1} k(t-y)\, dW(L(by)) + g_{(n)}(tb) \right|^p - \int_{-\infty}^{+\infty} \left| l(tb)Dx + g_{(n)}(tb) \right|^p \phi(x)\, dx \right\} w(tb)\, dt.$$

First, we show that η has no effect on the asymptotic distribution, i.e. is negligible. Using Jensen inequality and $(a+b)^p \leqslant 2^p(a^p + b^p)$ and the fact that l and w are bounded, we obtain

$$\begin{aligned}\eta^2 &\leqslant \int_{M_1}^{(1-b)/b} \left\{ \left| b^{-1/2} \int_{t-1}^{t+1} k(t-y)\, dW(L(by)) + g_{(n)}(tb) \right|^{2p} \right. \\ &\qquad \left. + \left(\int_{\mathbb{R}} \left| l(tb)Dx + g_{(n)}(tb) \right|^p \phi(x)\, dx \right)^2 \right\} w(tb)\, dt \\ &\leqslant C_1 \int_{M_1}^{\frac{1-b}{b}} \left\{ \left| b^{-\frac{1}{2}} \int_{t-1}^{t+1} k(t-y)\, dW(L(by)) \right|^{2p} + \left| g_{(n)}(tb) \right|^{2p} \right\} dt + C_2,\end{aligned}$$

for some positive constants C_1 and C_2. On the other hand,

$$\begin{aligned}\int_{M_1}^{\frac{1-b}{b}} \left| g_{(n)}(tb) \right|^{2p} dt &= (nb)^p \int_{M_1}^{(1-b)/b} \left| \lambda_{(n)}(tb) - \lambda(tb) \right|^{2p} dt \\ &= (nb)^p b^{-1} \int_{M_1 b}^{1-b} \left| \lambda_{(n)}(t) - \lambda(t) \right|^{2p} dt \\ &= (nb)^p b^{-1} \int_{M_1 b}^{1-b} \left| \int k(y)[\lambda(t-by) - \lambda(t)]\, dy \right|^{2p} dt \\ &\leqslant (nb)^p b^{4p} \sup_{t\in[0,1]} |\lambda''(t)|^{2p} \left| \int k(y) y^2\, dy \right|^{2p}\end{aligned}$$

Hence,

$$\begin{aligned}\mathbb{E}[\eta^2] &\leqslant C_1 \int_{M_1}^{\frac{1-b}{b}} \mathbb{E}\left[\left| b^{-\frac{1}{2}} \int_{t-1}^{t+1} k(t-y)\, dW(L(by)) \right|^{2p} \right] + 2C_3 (nb)^p b^{4p} + C_2 \\ &= O\left((nb)^p b^{4p} \right) = O(1).\end{aligned}$$

This means that $b\eta = o_P(1)$. The statement follows immediately from Lemma B.1.2. □

Lemma B.1.2. *Let* $l(t)$ *be a differentiable function on* $[0,1]$ *such that* $\inf_{[0,1]} l(t) > 0$ *and* $\sup_{[0,1]} |l'(t)| < \infty$. *Define* $L(t) = \int_0^t l(u)\,du$. *Assume that (A1) and (A3) hold. Let* ξ_i, *for* $i = 1, \ldots, M_1 - 1$, *be defined as in* (B.1.1). *Then we have*

$$b^{1/2}\gamma(p)^{-1} \sum_{i=1}^{M_1-1} \xi_i \to N(0,1),$$

where $\gamma^2(p)$ *is defined in* (6.3.22).

Proof. Let $\gamma \in (0,1)$ and $M_2 = [(M_1-1)^\gamma]$, $M_3 = [(M_1-1)/(M_2+2)]$. Define

$$\zeta_i = \sum_{j=(i-1)(M_2+2)+1}^{(i-1)(M_2+2)+M_2} \xi_j, \qquad i = 1, \ldots, M_3$$

$$\gamma_i = \xi_{iM_2+2i-1} + \xi_{iM_2+2i}, \qquad \gamma^* = \sum_{j=M_3(M_2+2)+1}^{M_1-1} \xi_j.$$

With this notation we can write

$$\sum_{i=1}^{M_1-1} \xi_i = \sum_{i=1}^{M_3} \zeta_i + \sum_{i=1}^{M_3} \gamma_i + \gamma^*$$

and we aim at showing that the first term in the right hand side of the previous equation determines the asymptotic distribution of $\sum_{i=0}^{M_1-1} \xi_i$.

Note that

$$b^{-1/2} \int_{t-1}^{t+1} k(t-y)\,dW(L(by)) \sim N\left(0, \sigma_t^2\right)$$

where

$$\sigma_t^2 = \int_{t-1}^{t+1} k^2(t-y) l(by)\,dy = D^2 l(bt) + O(b^2)$$

and

$$\begin{aligned}
&\mathbb{E}\left[\left| b^{-1/2} \int_{t-1}^{t+1} k(t-y)\,dW(L(by)) + g_{(n)}(tb) \right|^p\right] \\
&= \int_{-\infty}^{+\infty} \left|\sigma_t x + g_{(n)}(tb)\right|^p \phi(x)\,dx \\
&= \int_{-\infty}^{+\infty} \left|D\sqrt{l(tb)}x + g_{(n)}(tb)\right|^p \phi(x)\,dx + O(b^2).
\end{aligned}$$

Hence, we get $\mathbb{E}[\xi_i] = O(b^2)$ and $\mathbb{E}[\gamma_i] = O(b^2)$. Furthermore, and, as we did for η, it can be seen that $\mathbb{E}[\xi_i^2] = O(1)$ and $\mathbb{E}[\gamma_i^2] = O(1)$.

Since γ_i depends only on the Brownian motion on the interval

$$[L(b(iM_2 + 2i - 2)), L(b(iM_2 + 2i + 2))],$$

it follows that γ_i are independent (note that $M_2 > 2$). Moreover, γ^* is independent of γ_i, $i = 1, \dots, M_3 - 1$ and $\mathbb{E}[\gamma^*] = O(M_2 b^2)$. In addition, since ξ_i is independent of ξ_j for $|i - j| \geqslant 3$, we also have $\mathbb{E}[(\gamma^*)^2] \leqslant CM_2$. As a result

$$\mathbb{E}\left[\left(\sum_{i=1}^{M_3} \gamma_i + \gamma^*\right)^2\right] \leqslant c(M_3 + M_2) = o(1/b) \tag{B.1.2}$$

because $bM_2 \to 0$ and $bM_3 \to 0$. Indeed $M_2 \leqslant (T/b)^\gamma$ and

$$b\left[\frac{[(1-b)/b]}{[[(1-b)/b]^\gamma] + 2}\right] \leqslant \frac{1-b}{\left[\frac{1-b}{b}\right]^\gamma + 1} \leqslant \frac{1-b}{1 + \frac{(1-2b)^\gamma}{b^\gamma}} = \frac{b^\gamma}{(1-2b)^\gamma + b^\gamma} \to 0.$$

Consequently

$$b^{1/2}\left(\sum_{i=1}^{M_3} \gamma_i + \gamma^*\right) \xrightarrow{\mathbb{P}} 0.$$

Next, since ζ_i, $i = 1, \dots, M_3$ are independent, we apply the central limit theorem to conclude that

$$b^{1/2}\gamma(p)^{-1} \sum_{i=0}^{M_3} \zeta_i \to N(0, 1)$$

It suffices to show that

$$b\mathbb{E}\left[\left(\sum_{i=1}^{M_3} \zeta_i\right)^2\right] = b\sum_{i=0}^{M_3} \mathbb{E}[\zeta_i^2] \to \gamma^2(p). \tag{B.1.3}$$

and that they satisfy the Lyapunov's condition

$$\frac{\sum_i \mathbb{E}[\zeta_i^4]}{\left(\sum_i \mathbb{E}[\zeta_i^2]\right)^2} \to 0.$$

Note that, once we have (B.1.3), the Lyapunov's condition is equivalent to $b^2 \sum_i \mathbb{E}[\zeta_i^4] \to 0$. Using

$$\mathbb{E}[\zeta_i^4] = 4! \sum_{\substack{k,l,m,r \in I_i \\ k \leqslant l \leqslant m \leqslant r}} \mathbb{E}[\xi_k \xi_l \xi_m \xi_r],$$

for $I_i = \{(i-1)(M_2+2)+1, \dots, (i-1)(M_2+2)+M_2\}$, the fact that

$$\mathbb{E}[\xi_k \xi_l \xi_m \xi_r] = O(b^2)^4 \qquad \text{if} \qquad l \geqslant k+3 \text{ or } r \geqslant m+3$$

and that all the moments of the ξ_i's are finite, we obtain that

$$\mathbb{E}[\zeta_i^4] = O(M_2^2), \qquad \text{(uniformly w.r.t. } i). \tag{B.1.4}$$

Consequently $b^2 \sum_i \mathbb{E}[\zeta_i^4] = O(b^2 M_3 M_2^2) \to 0$ because $bM_2 \to 0$ and $bM_3M_2 = O(1)$. Indeed

$$bM_2M_3 \leqslant bM_2 \frac{M_1-1}{M_2+2} \leqslant bM_1 \leqslant 1.$$

In particular, it also follows that

$$b \sum_i \mathbb{E}[\zeta_i^2] = b\mathbb{E}\left[\left(\sum_{i=0}^{M_3} \zeta_i\right)^2\right] + bO(M_3^2 M_2^2 b^4) = O(bM_3M_2) = O(1). \tag{B.1.5}$$

Now we prove (B.1.3). From (B.1), it follows that

$$\mathrm{Var}\left(\int_b^{1-b} \left|b^{-1/2}\Gamma_n^{(1)}(t) + g_{(n)}(t)\right|^p \, d\mu(t)\right) = b^2 \mathrm{Var}\left(\sum_{i=1}^{M_1-1} \xi_i + \eta\right).$$

Moreover, since $\mathbb{E}[\xi_i] = O(b^2)$ for $i = 1, \dots, M_1 - 1$ and $\mathbb{E}[\eta] = 0$, we get

$$\begin{aligned}
&b^{-1}\mathrm{Var}\left(\int_b^{1-b} \left|l(t) b^{-1/2}\Gamma_n^{(1)}(t) + g_{(n)}(t)\right|^p \, d\mu(t)\right) \\
&= b\mathbb{E}\left[\left(\sum_{i=1}^{M_1-1} \xi_i + \eta\right)^2\right] + o(1) \\
&= b\mathbb{E}[\eta^2] + 2b\mathbb{E}\left[\left(\sum_{i=1}^{M_1-1} \xi_i\right)\eta\right] + b\mathbb{E}\left[\left(\sum_{i=1}^{M_1-1} \xi_i\right)^2\right] + o(1)
\end{aligned}$$

We have already shown in the proof of the previous lemma that $\mathbb{E}[\eta^2] = O(1)$, so the first term in the right hand side of the previous equation converges to zero. Furthermore,

$$\begin{aligned} b\mathbb{E}\left[\left(\sum_{i=1}^{M_1-1}\xi_i\right)^2\right] &= b\mathbb{E}\left[\left(\sum_{i=1}^{M_3}\zeta_i + \sum_{i=1}^{M_3}\gamma_i + \gamma^*\right)^2\right] \\ &= b\mathbb{E}\left[\left(\sum_{i=1}^{M_3}\zeta_i\right)^2\right] + b\mathbb{E}\left[\left(\sum_{i=1}^{M_3}\gamma_i + \gamma^*\right)^2\right] \\ &\quad + b\mathbb{E}\left[\left(\sum_{i=1}^{M_3}\zeta_i\right)\left(\sum_{i=1}^{M_3}\gamma_i + \gamma^*\right)\right]. \end{aligned}$$

Now, making use of (B.1.2), (B.1.5) and the fact that, by Cauchy-Schwartz,

$$\mathbb{E}\left[\left(\sum_{i=1}^{M_3}\zeta_i\right)\left(\sum_{i=1}^{M_3}\gamma_i + \gamma^*\right)\right] \leqslant \mathbb{E}\left[\left(\sum_{i=1}^{M_3}\zeta_i\right)^2\right]^{\frac{1}{2}} \mathbb{E}\left[\left(\sum_{i=1}^{M_3}\gamma_i + \gamma^*\right)^2\right]^{\frac{1}{2}}$$

we obtain

$$b\mathbb{E}\left[\left(\sum_{i=1}^{M_1-1}\xi_i\right)^2\right] = b\mathbb{E}\left[\left(\sum_{i=1}^{M_3}\zeta_i\right)^2\right] + o(1).$$

Similarly,

$$\begin{aligned} b\mathbb{E}\left[\left(\sum_{i=1}^{M_1-1}\xi_i\right)\eta\right] &= b\mathbb{E}\left[\left(\sum_{i=1}^{M_3}\zeta_i + \sum_{i=1}^{M_3}\gamma_i + \gamma^*\right)\eta\right] \\ &\leqslant b\mathbb{E}[\eta^2]^{1/2}\left\{\mathbb{E}\left[\left(\sum_{i=1}^{M_3}\zeta_i\right)^2\right]^{1/2} + \mathbb{E}\left[\left(\sum_{i=1}^{M_3}\gamma_i + \gamma^*\right)^2\right]^{1/2}\right\} \to 0. \end{aligned}$$

This means that

$$b\mathbb{E}\left[\left(\sum_{i=1}^{M_3}\zeta_i\right)^2\right] = b^{-1}\mathrm{Var}\left(\int_b^{1-b}\left|b^{-1/2}\Gamma_n^{(1)}(t) + g_{(n)}(t)\right|^p d\mu(t)\right) + o(1).$$

Define

$$X_{n,t} = b^{-1/2}\int_{t-b}^{t+b} k\left(\frac{t-y}{b}\right) dW(L(y)) + g_{(n)}(t). \tag{B.1.6}$$

Then, from Lemma B.1.3, it follows that

$$
\begin{aligned}
&b^{-1}\mathrm{Var}\left(\int_b^{1-b}\left|b^{-1/2}\Gamma_n^{(1)}(t)+g_{(n)}(t)\right|^p\,d\mu(t)\right)\\
&=\frac{1}{b}\int_b^{1-b}\int_b^{1-b}\left\{\mathbb{E}\left[|X_{n,t}X_{n,u}|^p\right]-\int_{\mathbb{R}}\left|\sigma_n(t)x+g_{(n)}(t)\right|^p\phi(x)\,dx\right.\\
&\qquad\qquad\left.\cdot\int_{-\infty}^{+\infty}\left|\sigma_n(u)y+g_{(n)}(u)\right|^p\phi(y)\,dy\right\}w(t)w(u)\,dt\,du\\
&=\frac{1}{b}\int_b^{1-b}\int_b^{1-b}\int_{\mathbb{R}^2}\left\{|A_n(t,u,x,y)|^p-\left|\sigma_n(t)x+g_{(n)}(t)\right|^p\right\}\\
&\qquad\qquad\cdot\left|\sigma_n(u)y+g_{(n)}(u)\right|^p w(t)w(u)\phi(x)\phi(y)\,dx\,dy\,dt\,du\\
&=\frac{1}{b}\int_b^{1-b}\int_b^{1-b}\int_{\mathbb{R}^2}\left\{|A_n(t,u,x,y)|^p-\left|\sqrt{L'(t)}Dx+g_{(n)}(t)\right|^p\right\}\\
&\qquad\cdot\left|\sqrt{L'(u)}Dy+g_{(n)}(u)\right|^p w(t)w(u)\phi(x)\phi(y)\,dx\,dy\,dt\,du+o(1)
\end{aligned}
$$

where

$$A_n(t,u,x,y)=g_{(n)}(t)+\sigma_n(t)\rho_n(t,u)y+\sqrt{1-\rho_n^2(t,u)}\sigma_n(t)x$$

with $\rho_n(t,u)$ and $\sigma_n(t)$ as defined in (B.1.8) and (B.1.7).

First we consider the case $nb^5\to 0$ and show that we can remove the $g_{(n)}$ functions from the previous integral. Indeed, since

$$
\begin{aligned}
&\left|\left|\sqrt{L'(u)}Dy+g_{(n)}(u)\right|^p-|\sqrt{L'(u)}Dy|^p\right|\\
&\leqslant p2^{p-1}|g_{(n)}(u)|^p+p2^{p-1}|\sqrt{L'(u)}Dy|^{p-1}|g_{(n)}(u)|
\end{aligned}
$$

we obtain

$$
\begin{aligned}
&\left|V_n-\frac{1}{b}\int_{I_n^2}\int_{\mathbb{R}^2}\left|\sqrt{l(u)}Dy\right|^p B_n(t,u,x,y)w(t)w(u)\phi(x)\phi(y)\,dx\,dy\,dt\,du\right|\\
&\leqslant\frac{c}{b}\int_{I_n^2}\int_{\mathbb{R}^2}\left\{\left|g_{(n)}(u)\right|^p+\left|\sqrt{l(u)}Dy\right|^{p-1}\left|g_{(n)}(u)\right|\right\}\\
&\qquad\qquad\cdot|B_n(t,u,x,y)|w(t)w(u)\phi(x)\phi(y)\,dx\,dy\,dt\,du,
\end{aligned}
$$

where $I_n=[b,1-b]$,

$$V_n=b^{-1}\mathrm{Var}\left(\int_b^{1-b}\left|b^{-1/2}\Gamma_n^{(1)}(t)+g_{(n)}(t)\right|^p\,d\mu(t)\right)$$

and

$$B_n(t,u,x,y)=|A_n(t,u,x,y)|^p-\left|\sqrt{l(t)}Dx+g_{(n)}(t)\right|^p.$$

Note that, if $|t-u| \geqslant 2b$, then $\rho_n(t,u) = 0$ and the previous integrand is equal to zero. Hence, a sufficient condition for the left hand side of the previous inequality to converge to zero is to have

$$b^{-1}\int_b^{1-b}\int_b^{1-b} \mathbb{1}_{\{|t-u|<2b\}} \left|g_{(n)}(u)\right|^p \left|g_{(n)}(t)\right|^p \, du\, dt \to 0.$$

and

$$b^{-1}\int_b^{1-b}\int_b^{1-b} \mathbb{1}_{\{|t-u|<2b\}} \left|g_{(n)}(u)\right|^p \, du\, dt \to 0.$$

This is indeed the case because $g_n(u) = O\left((nb)^{1/2}b^2\right)$ uniformly w.r.t. u and $(nb)^{1/2}b^2 \to 0$. In the same way we can remove also the other $g_{(n)}$ functions from the integrand, i.e.

$$V_n = \frac{1}{b}\int_{I_n^2}\int_{\mathbb{R}^2} \left|\sqrt{l(u)}Dy\right|^p B_n'(t,u,x,y)w(t)w(u)\phi(x)\phi(y)\, dx\, dy\, dt\, du + o(1)$$

where

$$B_n'(t,u,x,y) = \left|\sigma_n(t)\rho_n(t,u)y + \sqrt{1-\rho_n^2(t,u)}\sigma_n(t)x\right|^p - \left|\sqrt{l(t)}Dx\right|^p$$

With the change of variable $t = u + sb$, we get

$$V_n = \int_b^{1-b} \int\limits_{\substack{1-u/b \\ |s|\leqslant 2}}^{(1-b-u)/b} \int_{\mathbb{R}}\int_{\mathbb{R}} \left\{\left|yr(s) + \sqrt{1-r^2(s)}x\right|^p - |x|^p\right\}$$
$$\left|\sqrt{l(u)}\sqrt{l(u+sb)}D^2y\right|^p w(u)w(u+sb)\phi(x)\phi(y)\, dx\, dy\, ds\, du + o(1),$$

where $r(s)$ is defined in (6.1.7). The continuity of the functions l and w and the dominated convergence theorem yield

$$V_n = \int_b^{1-b} \int\limits_{|s|\leqslant 2}\int_{\mathbb{R}^2} \left|\sqrt{l(u)}\right|^{2p} D^{2p}|y|^p \left\{\left|yr(s) + \sqrt{1-r^2(s)}x\right|^p - |x|^p\right\}$$
$$\cdot\, w(u)^2\phi(x)\phi(y)\, dx\, dy\, ds\, du + o(1).$$

Then, with the change of variable $yr(s) + \sqrt{1-r^2(s)}x = z$ we can write equivalently

$$V_n = \frac{1}{2\pi}\int_{\mathbb{R}^3} |y|^p \left\{|z|^p - \left|\frac{z - r(s)y}{\sqrt{1-r^2(s)}}\right|^p\right\} e^{-\frac{z^2+y^2-2rzy}{2(1-r^2(s))}} \frac{1}{\sqrt{1-r^2(s)}}\, dz\, dy\, ds$$
$$\cdot\, D^{2p}\int_b^{1-b} \left|\sqrt{L'(u)}\right|^{2p} w(u)^2\, du + o(1)$$
$$= \sigma_1 D^{2p}\int_0^1 \left|\sqrt{L'(u)}\right|^{2p} w(u)^2\, du + o(1)$$

where σ^1 is defined in (6.1.10).

Let us now consider the case $nb^5 \to c_0^2 > 0$. First we show that the $g_{(n)}(u)$ functions can be replaced by $g(u)$ defined in (6.1.5). Indeed, $g_{(n)}(u) = g(u) + o((nb)^{1/2}b^2)$, where the big O term is uniform w.r.t. u and similar calculations to those of the previous case allow us to conclude that

$$V_n = \frac{1}{b}\int_{I_n^2}\int_{\mathbb{R}^2}\left|\sqrt{l(u)}Dy + g(u)\right|^p B_n'(t,u,x,y) \cdot w(t)w(u)\phi(x)\phi(y)\,dx\,dy\,dt\,du + o(1)$$

where

$$B_n'(t,u,x,y) = \left|g(t) + \sqrt{L'(t)}D\left[\rho_n(t,u)y + \sqrt{1-\rho_n^2(t,u)}x\right]\right|^p - \left|\sqrt{L'(t)}Dx + g(t)\right|^p.$$

With the change of variable $t = u + sb$, we get

$$V_n = \int_{I_n}\int_{\substack{(b-u)/b \\ |s|\leqslant 2}}^{(1-b-u)/b}\int_{\mathbb{R}^2}\left\{\left|g(u+sb) + \sqrt{l(u+sb)}D[yr(s) + \sqrt{1-r^2(s)}x]\right|^p - \left|g(u+sb) + \sqrt{l(u+sb)}Dx\right|^p\right\} \cdot \left|g(u) + \sqrt{L'(u)}Dy\right|^p w(u)w(u+sb)\phi(x)\phi(y)\,dx\,dy\,ds\,du + o(1).$$

Again, by the continuity of the functions l, w and g and the dominated convergence theorem we obtain that A_n converges to

$$\int_0^1\int_{\mathbb{R}^3}\left|g(u) + \sqrt{L'(u)}Dy\right|^p\left\{\left|g(u) + \sqrt{L'(u)}D[yr(s) + \sqrt{1-r^2(s)}x]\right|^p - \left|g(u) + \sqrt{L'(u)}Dx\right|^p\right\} w(u)^2\phi(x)\phi(y)\,dx\,dy\,ds\,du,$$

which is exactly $\theta^2(p)$ defined in (6.1.11). □

Lemma B.1.3. *Let* $l(t)$ *be a differentiable function on* $[0,1]$ *such that* $\inf_{[0,1]} l(t) > 0$ *and* $\sup_{[0,1]}|l'(t)| < \infty$. *Define* $L(t) = \int_0^t l(u)\,du$ *and, for* $t \in [0,1]$, *let* $X_{n,t}$ *be as in* (B.1.6). *It holds*

$$\mathbb{E}\left[|X_{n,t}X_{n,u}|^p\right] = \int_{\mathbb{R}}\int_{\mathbb{R}}\left|g_{(n)}(t) + \sigma_n(t)\rho_n(t,u)y + \sqrt{1-\rho_n^2(t,u)}\sigma_n(t)x\right|^p \cdot \left|\sigma_n(u)y + g_{(n)}(u)\right|^p \phi(x)\phi(y)\,dx\,dy,$$

where

$$\sigma_n^2(t) = l(t)D^2 + O(b^2), \qquad \sigma_n(t,u) = b^{-1}\int k(t-y)k(u-y)l(y)\,dy. \tag{B.1.7}$$

and

$$\rho_n(t,u) = \frac{\int k\left(\frac{t-y}{b}\right) k\left(\frac{u-y}{b}\right) l(y)\,dy}{b\sqrt{D^2 l(t) + O(b^2)}\sqrt{D^2 l(u) + O(b^2)}} \tag{B.1.8}$$

Proof. First, note that

$$(X_{n,t}, X_{n,u}) \sim N\left(\begin{bmatrix} g_{(n)}(t) \\ g_{(n)}(u) \end{bmatrix}, \begin{bmatrix} \sigma_n^2(t) & \sigma_n(t,u) \\ \sigma_n(t,u) & \sigma_n^2(u) \end{bmatrix}\right).$$

Hence, $X_{n,t}|X_{n,u} = x_2$ is distributed as

$$N\left(g_{(n)}(t) + \frac{\sigma_n(t)}{\sigma_n(u)}\rho_n(t,u)\left(x_2 - g_{(n)}(u)\right), (1-\rho_n^2(t,u))\sigma_n^2(t)\right)$$

and

$$\begin{aligned}\mathbb{E}\left[|X_{n,t}|^p|X_{n,u}\right] = \int_{\mathbb{R}} \Bigg| & g_{(n)}(t) + \frac{\sigma_n(t)}{\sigma_n(u)}\rho_n(t,u)\left(X_{n,u} - g_{(n)}(u)\right) \\ & + \sqrt{1-\rho_n^2(t,u)}\sigma_n(t)x\Bigg|^p \phi(x)\,dx\end{aligned}$$

Consequently, we obtain

$$\begin{aligned}&\mathbb{E}\left[|X_{n,t}X_{n,u}|^p\right] = \mathbb{E}\left[\mathbb{E}\left[|X_{n,t}X_{n,u}|^p|X_{n,u}\right]\right] \\ &= \int_{\mathbb{R}}\int_{\mathbb{R}} \left|g_{(n)}(t) + \sigma_n(t)\rho_n(t,u)y + \sqrt{1-\rho_n^2(t,u)}\sigma_n(t)x\right|^p \phi(x)\,dx \\ &\qquad \cdot \left|\sigma_n(u)y + g_{(n)}(u)\right|^p \phi(y)\,dy \\ &= \int_{\mathbb{R}}\int_{\mathbb{R}} \left|g_{(n)}(t) + \sigma_n(t)\rho_n(t,u)y + \sqrt{1-\rho_n^2(t,u)}\sigma_n(t)x\right|^p \\ &\qquad \cdot \left|\sigma_n(u)y + g_{(n)}(u)\right|^p \phi(x)\phi(y)\,dx\,dy.\end{aligned}$$

□

Proof of Proposition 6.2.3. We first prove (i). For each $t \in [0, b)$, we have

$$\begin{aligned}
\tilde{\lambda}_n^s(t) - \lambda(t) &= \int_0^{t+b} k_b(t-u)\, d\Lambda_n(u) - \lambda(t) \\
&= \int_0^{t+b} k_b(t-u)\, d(\Lambda_n - \Lambda)(u) + \int_0^{t+b} k_b(t-u)\, d\Lambda(u) - \lambda(t) \\
&= \int_0^{t+b} k_b(t-u)\, d(\Lambda_n - \Lambda)(u) + \int_{-1}^{t/b} k(y)[\lambda(t-by) - \lambda(t)]\, dy \\
&\quad - \lambda(t) \int_{t/b}^{1} k(y)\, dy.
\end{aligned}$$

Note that

$$\begin{aligned}
&\left| \int_0^{t+b} k_b(t-u)\, d(\Lambda_n - \Lambda)(u) \right| \\
&\quad = \frac{1}{b^2} \left| \int_0^{t+b} (\Lambda_n - \Lambda)(u) k'\left(\frac{t-u}{b}\right) du \right| \\
&\quad \leqslant c b^{-1} \sup_{u \leqslant 2b} |M_n(u) - M_n(0)| \\
&\quad \leqslant c b^{-1} \left\{ \sup_{u \leqslant 2b} \left| M_n(u) - n^{-\frac{1}{2}} B_n \circ L(u) \right| + n^{-\frac{1}{2}} \left| B_n \circ L(u) - B_n \circ L(0) \right| \right\} \\
&\quad = O_P\left(b^{-1} n^{-1+1/q}\right) + n^{-1/2} b^{-1} \sup_{y \leqslant cb} |B_n(y)| = O_P\left((nb)^{-1/2}\right),
\end{aligned} \tag{B.1.9}$$

uniformly in $t \in [0, b]$, and that according to (6.2.7),

$$\left| \int_{-1}^{t/b} k(y)[\lambda(t-by) - \lambda(t)]\, dy \right| = O(b),$$

Moreover, for $t \leqslant b/2$,

$$\lambda(t) \int_{t/b}^{1} k(y)\, dy \geqslant \inf_{t \in [0,1]} \lambda(t) \int_{1/2}^{1} k(y)\, dy = C > 0.$$

Now, define the event

$$\begin{aligned}
\mathcal{A}_n = \Bigg\{ \sup_{t \in [0,b]} \Bigg(&\left| \int_0^{t+b} k_b(t-u)\, d(\Lambda_n - \Lambda)(u) \right| \\
&+ \left| \int_{-1}^{t/b} k(y)[\lambda(t-by) - \lambda(t)]\, dy \right| \Bigg) \leqslant C/2 \Bigg\}.
\end{aligned}$$

Then, $\mathbb{P}(\mathcal{A}_n) \to 1$ and on the event $\mathcal{A}_n$, $|\tilde{\lambda}_n^s(t) - \lambda(t)| \geqslant C/2$. Consequently we obtain

$$\begin{aligned}\mathbb{E}\left[\int_0^b \left|\tilde{\lambda}_n^s(t) - \lambda(t)\right|^p \, d\mu(t)\right] &\geqslant \mathbb{E}\left[\int_0^{b/2} \left|\tilde{\lambda}_n^s(t) - \lambda(t)\right|^p \, d\mu(t)\right] \\ &\geqslant \mathbb{E}\left[\mathbb{1}_{\mathcal{A}_n} \int_0^{b/2} \left|\hat{\lambda}_n^s(t) - \lambda(t)\right|^p \, d\mu(t)\right] \\ &\geqslant c\mathbb{P}(\mathcal{A}_n) b,\end{aligned} \tag{B.1.10}$$

for some $c > 0$. Hence

$$(nb)^{p/2}\mathbb{E}\left[\int_0^b \left|\tilde{\lambda}_n^s(t) - \lambda(t)\right|^p \, d\mu(t)\right] \geqslant cb(nb)^{p/2}\mathbb{P}(\mathcal{A}_n) \to \infty,$$

because $b(nb)^{p/2} \geqslant b(nb)^{1/2} = (nb^3)^{1/2} \to \infty$.

In order to prove (ii), due to (6.1.14), we can bound

$$b^{-1/2}\left|\int_0^b (nb)^{p/2} \left|\tilde{\lambda}_n^s(t) - \lambda(t)\right|^p \, d\mu(t) - \int_0^b \left|g_{(n)}(t)\right|^p \, d\mu(t)\right|$$

by

$$\begin{aligned}&p2^{p-1}b^{-1/2}(nb)^{p/2}\int_0^b \left|\int_0^{t+b} k_b(t-u)\, d(\Lambda_n - \Lambda)(u)\right|^p \, d\mu(t) \\ &\quad + p2^{p-1}b^{-1/2}\left(\int_0^b (nb)^{p/2}\left|\int_0^{t+b} k_b(t-u)\, d(\Lambda_n - \Lambda)(u)\right|^p \, d\mu(t)\right)^{1/p} \\ &\qquad\qquad \cdot \left(\int_0^b \left|g_{(n)}(t)\right|^p \, d\mu(t)\right)^{1-1/p}.\end{aligned}$$

According to (B.1.9)

$$\left|\int_0^{t+b} k_b(t-u)\, d(\Lambda_n - \Lambda)(u)\right| = O_P\left((nb)^{-1/2}\right),$$

uniformly in $t \in [0, b]$. Furthermore, using (6.2.6), (6.2.7), and (6.2.8), we have

$$g_{(n)}(t) = O\left((nb)^{1/2}\right), \tag{B.1.11}$$

uniformly for $t \in [0, b]$. Hence, we obtain

$$\begin{aligned} &b^{-1/2} \left| \int_0^b (nb)^{p/2} \left|\tilde{\lambda}_n^s(t) - \lambda(t)\right|^p \, d\mu(t) - \int_0^b \left|g_{(n)}(t)\right|^p \, d\mu(t) \right| \\ &\quad \leqslant O_P\left(b^{1/2}\right) + O_P\left(n^{(p-1)/2} b^{p/2}\right) \to 0, \end{aligned}$$

because $n^{(p-1)/2} b^{p/2} = (bn^{1-1/p})^{p/2} \to 0$.

Next we deal with (iii). Again by means of (6.1.14), we can bound

$$b^{-1/2} \left| \int_0^b (nb)^{p/2} \left|\tilde{\lambda}_n^s(t) - \lambda(t)\right|^p \, d\mu(t) - \int_0^b \left|Y_n(t) + g_{(n)}(t)\right|^p \, d\mu(t) \right|$$

by

$$\begin{aligned} &p2^{p-1} b^{-\frac{1}{2}} \int_0^b \left| (nb)^{1/2} \int_0^{t+b} k_b(t-u) \, d(\Lambda_n - \Lambda - n^{-\frac{1}{2}} B_n \circ L)(u) \right|^p \, d\mu(t) \\ &\quad + p2^{p-1} b^{-\frac{1}{2}} \left(\int_0^b \left|Y_n(t) + g_{(n)}(t)\right|^p \, d\mu(t) \right)^{1-1/p} \\ &\quad \cdot \left(\int_0^b \left| (nb)^{\frac{1}{2}} \int_0^{t+b} k_b(t-u) \, d(\Lambda_n - \Lambda - n^{-\frac{1}{2}} B_n \circ L)(u) \right|^p \, d\mu(t) \right)^{1/p} \end{aligned}$$

Note that

$$\sup_{t \in [0,b]} |Y_n(t)| = \sup_{t \in [0,b]} \left| b^{1/2} \int_0^{t+b} k_b(t-u) \, dB_n(L(u)) \right| = O_P(1),$$

and, as in (B.1.9),

$$\begin{aligned} &\left| \int_0^{t+b} k_b(t-u) \, d(\Lambda_n - \Lambda - n^{-1/2} B_n \circ L)(u) \right| \\ &\quad \leqslant \frac{1}{b} \sup_{u \leqslant 2b} \left| (\Lambda_n - \Lambda - n^{-1/2} B_n \circ L)(u) \right| = O_P\left(b^{-1} n^{-1+1/q}\right), \end{aligned}$$

uniformly for $t \in [0, b]$. Together with (B.1.11), we obtain

$$\begin{aligned} &b^{-1/2} \left| \int_0^b (nb)^{p/2} \left|\tilde{\lambda}_n^s(t) - \lambda(t)\right|^p \, d\mu(t) - \int_0^b \left|Y_n(t) + g_{(n)}(t)\right|^p \, d\mu(t) \right| \\ &\leqslant O_P\left(b^{-1/2} (nb)^{p/2} b n^{-p+p/q} b^{-p}\right) + O_P\left(b^{-1/2} b (nb)^{p/2} n^{-1+1/q} b^{-1}\right) \\ &= b^{-1/2} (nb)^{p/2} n^{-1+1/q} \left\{ O_P\left((n^{-1+1/q} b^{-1})^{p-1}\right) + O_P(1) \right\}. \end{aligned}$$

Because $n^{-1+1/q}b^{-1} = O(1)$, the term within the brackets is of order $O_P(1)$, and since $b^{p-1}n^{p-2+2/q} \to 0$, the right hand side tends to zero. This proves (6.2.11).

Then, by Jensen's inequality, we get

$$\begin{aligned}
&b^{-1}\mathrm{Var}\left(\int_0^{cb} |Y_n(t) + g_{(n)}(t)|^p \, d\mu(t)\right) \\
&= \frac{1}{b}\mathbb{E}\left[\left(\int_0^{cb} |Y_n(t) + g_{(n)}(t)|^p \, d\mu(t) - \int_0^{cb} \mathbb{E}\left[|Y_n(t) + g_{(n)}(t)|^p\right] d\mu(t)\right)^2\right] \\
&\geqslant \frac{1}{b}\mathbb{E}\left[\left|\int_0^{cb} |Y_n(t) + g_{(n)}(t)|^p \, d\mu(t) - \int_0^{cb} \mathbb{E}\left[|Y_n(t) + g_{(n)}(t)|^p\right] d\mu(t)\right|\right]^2.
\end{aligned} \tag{B.1.12}$$

Note that $Y_n(t) \sim N(0, \sigma_n^2(t))$, where,

$$\sigma_n^2(t) = b^{-1}\int_0^{t+b} k^2\left(\frac{t-u}{b}\right) L'(u)\, du = \int_{-1}^{t/b} k^2(y) L'(t-by)\, dy,$$

if B_n is a Brownian motion, and

$$\sigma_n^2(t) = \int_{-1}^{t/b} k^2(y) L'(t-by)\, dy + O(b),$$

if B_n is a Brownian bridge. Now, choose $\epsilon > 0$. Then

$$\liminf_{n\to\infty} \mathbb{P}\left(\epsilon \leqslant Y_n(0) \leqslant 2\epsilon\right) > 0 \quad \text{and} \quad \liminf_{n\to\infty} \mathbb{P}\left(-2\epsilon \leqslant Y_n(0) \leqslant -\epsilon\right) > 0.$$

For $c > 0$, define the events

$$\begin{aligned}
\mathcal{A}_{n1} &= \{\epsilon/2 \leqslant Y_n(t) \leqslant 3\epsilon, \text{ for all } t \in [0, cb]\}, \\
\mathcal{A}_{n2} &= \{-3\epsilon \leqslant Y_n(t) \leqslant -\epsilon/2, \text{ for all } t \in [0, cb]\},
\end{aligned}$$

and let

$$\mathcal{B}_n = \left\{\int_0^{cb} \mathbb{E}\left[|Y_n(t) + g_{(n)}(t)|^p\right] d\mu(t) > \int_0^{cb} |g_{(n)}(t)|^p \, d\mu(t)\right\}.$$

Then, since Y_n has continuous paths, we have

$$\liminf_{n\to\infty} \mathbb{P}(\mathcal{A}_{n1}) > 0 \qquad \text{and} \qquad \liminf_{n\to\infty} \mathbb{P}(\mathcal{A}_{n2}) > 0.$$

Moreover, $Y_n(t) > 0$ on the event $\mathcal{A}_{n1}$, and from (6.2.9), it follows that $Y_n(t) + g_{(n)}(t) < 0$, for n sufficiently large. Therefore, for n sufficiently large, we have on $\mathcal{A}_{n1}$,

$$\int_0^{cb} |Y_n(t) + g_{(n)}(t)|^p \, d\mu(t) \leqslant \int_0^{cb} |\epsilon/2 + g_{(n)}(t)|^p \, d\mu(t). \tag{B.1.13}$$

Similarly, $Y_n(t) < 0$ on the event $\mathcal{A}_{n2}$ and $Y_n(t) + g_{(n)}(t) < 0$, for large n, so that on $\mathcal{A}_{n2}$,

$$\int_0^{cb} |Y_n(t) + g_{(n)}(t)|^p \, d\mu(t) \geqslant \int_0^{cb} |-\epsilon/2 + g_{(n)}(t)|^p \, d\mu(t). \tag{B.1.14}$$

Next, let

$$X_n = \int_0^{cb} |Y_n(t) + g_{(n)}(t)|^p \, d\mu(t) - \int_0^{cb} \mathbb{E}\left[|Y_n(t) + g_{(n)}(t)|^p\right] d\mu(t)$$

and write

$$\mathbb{E}\left[|X_n|\right] \geqslant \mathbb{E}\left[|X_n| \, \mathbb{1}_{\mathcal{A}_{n1}}\right] \mathbb{1}_{\mathcal{B}_n} + \mathbb{E}\left[|X_n| \, \mathbb{1}_{\mathcal{A}_{n2}}\right] \mathbb{1}_{\mathcal{B}_n^c}. \tag{B.1.15}$$

Consider the first term on the right hand side. Because for n large, $Y_n(t) + g_{(n)}(t) < 0$ on the event $\mathcal{A}_{n1}$, we have $|Y_n(t) + g_{(n)}(t)| \leqslant |g_{(n)}(t)|$. It follows that on the event $\mathcal{A}_{n1} \cap \mathcal{B}_n$:

$$\begin{aligned}\int_0^{cb} |Y_n(t) + g_{(n)}(t)|^p \, d\mu(t) &\leqslant \int_0^{cb} |g_{(n)}(t)|^p \, d\mu(t) \\ &< \int_0^{cb} \mathbb{E}\left[|Y_n(t) + g_{(n)}(t)|^p\right] d\mu(t).\end{aligned}$$

This means that we can remove the absolute value signs in the first term on the right hand side of (B.1.15). Similarly, $Y_n(t) + g_{(n)}(t) < 0$, for n sufficiently large on the event $\mathcal{A}_{n2}$, so that on the event $\mathcal{A}_{n2} \cap \mathcal{B}_n^c$:

$$\begin{aligned}\int_0^{cb} |Y_n(t) + g_{(n)}(t)|^p \, d\mu(t) &\geqslant \int_0^{cb} |g_{(n)}(t)|^p \, d\mu(t) \\ &\geqslant \int_0^{cb} \mathbb{E}\left[|Y_n(t) + g_{(n)}(t)|^p\right] d\mu(t),\end{aligned}$$

so that we can also remove the absolute value signs in the second term on the right hand side of (B.1.15). It follows that the right hand of (B.1.15) is equal to

$$\begin{aligned}&\mathbb{E}\left[-X_n \mathbb{1}_{\mathcal{A}_{n1}}\right] \mathbb{1}_{\mathcal{B}_n} + \mathbb{E}\left[X_n \mathbb{1}_{\mathcal{A}_{n2}}\right] \mathbb{1}_{\mathcal{B}_n^c} \\ &\geqslant \left(\int_0^{cb} |g_{(n)}(t)|^p \, d\mu(t) - \int_0^{cb} \left|\frac{\epsilon}{2} + g_{(n)}(t)\right|^p d\mu(t)\right) \mathbb{P}(\mathcal{A}_{n1}) \mathbb{1}_{\mathcal{B}_n} \\ &\quad + \left(\int_0^{cb} \left|-\frac{\epsilon}{2} + g_{(n)}(t)\right|^p d\mu(t) - \int_0^{cb} |g_{(n)}(t)|^p \, d\mu(t)\right) \mathbb{P}(\mathcal{A}_{n2}) \mathbb{1}_{\mathcal{B}_n^c},\end{aligned}$$

by using (6.2.7) and (6.2.8). Furthermore, for the first term on the right hand side

$$|g_{(n)}(t)|^p - |\epsilon/2 + g_{(n)}(t)|^p = |g_{(n)}(t)|^p \left(1 - |\epsilon_n(t) + 1|^p\right),$$

where $\epsilon_n(t) = \epsilon/(2g_{(n)}(t)) = O((nb)^{-1/2}) \to 0$, due to (6.2.6), (6.2.7) and (6.2.8), where the big-O term is uniformly for $t \in [0, b]$. This means that, for n large, $1 + \epsilon_n(t) > 0$, and by a Taylor expansion

$$|1 + \epsilon_n(t)|^p = 1 + p\epsilon_n(t) + O((nb)^{-1}).$$

It follows that

$$\begin{aligned}
&\int_0^{cb} |g_{(n)}(t)|^p \, d\mu(t) - \int_0^{cb} |\epsilon/2 + g_{(n)}(t)|^p \, d\mu(t) \\
&= \int_0^{cb} |g_{(n)}(t)|^p \left\{1 - |\epsilon_n(t) + 1|^p\right\} d\mu(t) \\
&= -p \int_0^{cb} |g_{(n)}(t)|^p \epsilon_n(t) \, d\mu(t) + cb \sup_{t \in [0,cb]} |g_{(n)}(t)|^p O((nb)^{-1}) \\
&= p(\epsilon/2) \int_0^{cb} |g_{(n)}(t)|^{p-1} \, d\mu(t) + O\left(b(nb)^{(p-1)/2}\right) \\
&= O\left(b(nb)^{(p-1)/2}\right)
\end{aligned}$$

due to (B.1.11). Similarly

$$\int_0^{cb} |-\epsilon/2 + g_{(n)}(t)|^p \, d\mu(t) - \int_0^{cb} |g_{(n)}(t)|^p \, d\mu(t) = O\left(b(nb)^{(p-1)/2}\right).$$

Going back to (B.1.12), since $\mathbb{P}(\mathcal{A}_{n1}) \to 1$ and $\mathbb{P}(\mathcal{A}_{n2}) \to 1$, we conclude that

$$b^{-1}\mathrm{Var}\left(\int_0^{cb} |Y_n(t) + g_{(n)}(t)|^p \, d\mu(t)\right) \geqslant b^{-1} O\left(b(nb)^{(p-1)/2}\right)^2.$$

The statement follows from the fact that $b^{-1}(nb)^{p-1}b^2 = n^{p-1}b^p \to \infty$.

Finally, one can deal in the same way with the L_p-error on the interval $(1-b, 1]$. □

Proof of Proposition 6.2.6. By definition we have

$$\begin{aligned}
&(nb)^{p/2} \int_0^b \left|\hat{\lambda}_n^s(t) - \lambda(t)\right|^p d\mu(t) \\
&\quad = \int_0^b \left|(nb)^{1/2} \int_0^{t+b} k_b^{(t)}(t-u) \, d(\Lambda_n - \Lambda)(u) + \bar{g}_{(n)}(t)\right|^p d\mu(t),
\end{aligned}$$

where

$$\bar{g}_{(n)}(t) = (nb)^{1/2} \left(\int k_b^{(t)}(t-u)\lambda(u) \, du - \lambda(t)\right). \tag{B.1.16}$$

When B_n in assumption (A2) is a Brownian motion, we can argue as in the proof of Theorem 6.2.1. By means of (6.1.14) we can bound

$$b^{-1/2}\Bigg|(nb)^{p/2}\int_0^b \left|\hat{\lambda}_n^s(t)-\lambda(t)\right|^p d\mu(t) \\ -\int_0^b \left|b^{-1/2}\int_0^{t+b} k^{(t)}\left(\frac{t-u}{b}\right) dB_n(L(u))+\bar{g}_{(n)}(t)\right|^p d\mu(t)\Bigg|,$$

from above by

$$p2^{p-1}b^{-1/2}b^{-p/2}\int_0^b \left|\int_0^{t+b} k^{(t)}\left(\frac{t-u}{b}\right) d(B_n\circ L-n^{1/2}M_n)(u)\right|^p d\mu(t) \\ +p2^{p-1}b^{-\frac{1}{2}}\left(b^{-\frac{p}{2}}\int_0^b \left|\int_0^{t+b} k^{(t)}\left(\frac{t-u}{b}\right) d(B_n\circ L-n^{\frac{1}{2}}M_n)(u)\right|^p d\mu(t)\right)^{\frac{1}{p}} \\ \cdot\left(\int_0^b \left|b^{-\frac{1}{2}}\int_0^{t+b} k^{(t)}\left(\frac{t-u}{b}\right) dB_n(L(u))+\bar{g}_{(n)}(t)\right|^p d\mu(t)\right)^{1-\frac{1}{p}}. \tag{B.1.17}$$

Similar to (6.2.4),

$$\begin{aligned}&\sup_{t\in[0,b]}\left|\int_0^{t+b} k^{(t)}\left(\frac{t-u}{b}\right) d(B_n\circ L-n^{1/2}M_n)(u)\right| \\ &\leqslant\left|\int_{-1}^{t/b}\left\{\psi_1\left(\frac{t}{b}\right)k(y)+\psi_2\left(\frac{t}{b}\right)yk(y)\right\} d(B_n\circ L-n^{1/2}M_n)(t-by)\right| \\ &\leqslant C\sup_{t\in[0,1]}\left|B_n\circ L(t)-n^{1/2}M_n(t)\right| \\ &=O_P(n^{-1/2+1/q}).\end{aligned} \tag{B.1.18}$$

Note that here we used the boundedness of the coefficients ψ_1 and ψ_2. Similar to the proof of Theorem 6.2.1, the idea is to show that

$$b^{-1/2}\int_0^b \left|b^{-1/2}\int_0^{t+b} k^{(t)}\left(\frac{t-u}{b}\right) dB_n(L(u))+\bar{g}_{(n)}(t)\right|^p d\mu(t)\to 0, \tag{B.1.19}$$

in probability. We first bound the left hand side of (B.1.19) by

$$Cb^{-1/2}\int_0^b \left\{|\bar{g}_{(n)}(t)|^p+b^{-p/2}\left|\int_0^{t+b} k^{(t)}\left(\frac{t-u}{b}\right) dB_n(L(u))\right|^p\right\} d\mu(t).$$

According to (1.2.13), a Taylor expansion gives

$$\begin{aligned}
\sup_{t\in[0,b]} |\bar{g}_{(n)}(t)| &= (nb)^{1/2} \sup_{t\in[0,b]} \left| \int_0^{t+b} k_b^{(t)}(t-u)\lambda(u)\,du - \lambda(t) \right| \\
&= (nb)^{1/2} \sup_{t\in[0,b]} \left| \int_{-1}^{t/b} k^{(t)}(y)\,[\lambda(t-by)-\lambda(t)]\,dy \right| \\
&= (nb)^{1/2} b^2 \sup_{t\in[0,b]} \left| \frac{1}{2} \int_{-1}^{t/b} k^{(t)}(y) y^2 \lambda''(\xi_{t,y})\,dy \right| \\
&= O_P\left((nb^5)^{1/2}\right) = O_P(1).
\end{aligned}$$

Furthermore,

$$\begin{aligned}
&\mathbb{E}\left[\left| \int_0^{t+b} k^{(t)}\left(\frac{t-u}{b}\right) dB_n(L(u)) \right|^p \right] \\
&= \int_{\mathbb{R}} \left(\int_0^{t+b} \left(k^{(t)}\left(\frac{t-u}{b}\right)\right)^2 L'(u)\,du \right)^{p/2} |x|^p \phi(x)\,dx \\
&= b^{p/2} \int_{\mathbb{R}} \left(\int_0^{t+b} \left(k^{(t)}\left(\frac{t-u}{b}\right)\right)^2 L'(u)\,du \right)^{p/2} |x|^p \phi(x)\,dx \\
&= O(b^{p/2}),
\end{aligned}$$

where ϕ denotes the standard normal density. This proves (B.1.19) for the case that B_n is a Brownian motion.

When B_n in (A2) is a Brownian bridge, then we use the representation $B_n(u) = W_n(u) - uW_n(L(1))/L(1)$, for some Brownian motion W_n. In this case, by means of (6.1.14), we can bound

$$\begin{aligned}
b^{-1/2} \Bigg| \int_0^b \left| b^{-1/2} \int_0^{t+b} k^{(t)}\left(\frac{t-u}{b}\right) dB_n(L(u)) + \bar{g}_{(n)}(t) \right|^p d\mu(t) \\
- \int_0^b \left| b^{-1/2} \int_0^{t+b} k^{(t)}\left(\frac{t-u}{b}\right) dW_n(L(u)) + \bar{g}_{(n)}(t) \right|^p d\mu(t) \Bigg|
\end{aligned}$$

by

$$p2^{p-1}b^{-\frac{1}{2}}\int_0^b \left| b^{-\frac{1}{2}}\frac{W_n(L(1)}{L(1)}\int_0^{t+b} k^{(t)}\left(\frac{t-u}{b}\right)L'(u)\,du+\bar{g}_{(n)}(t)\right|^p d\mu(t)$$

$$+\frac{p2^{p-1}}{\sqrt{b}}\left(\int_0^b \left|\frac{W_n(L(1))}{\sqrt{b}L(1)}\int_0^{t+b} k^{(t)}\left(\frac{t-u}{b}\right)L'(u)du+\bar{g}_{(n)}(t)\right|^p d\mu(t)\right)^{\frac{1}{p}}$$

$$\cdot\left(\int_0^b \left|b^{-\frac{1}{2}}\int_0^{t+b} k^{(t)}\left(\frac{t-u}{b}\right)dW_n(L(u))+\bar{g}_{(n)}(t)\right|^p d\mu(t)\right)^{1-\frac{1}{p}},$$

which tends to zero in probability, due to (B.1.19). □

B.2 ISOTONIZED KERNEL ESTIMATOR

Lemma B.2.1. *Assume (A1)-(A2) and let* $\tilde{\lambda}_n^s$ *be defined in* (6.1.2). *Let* k *satisfy* (6.1.1) *and let* $p \geqslant 1$. *If* $b \to 0$, $nb \to \infty$, *and* $1/b = o(n^{1/4})$, *then*

$$\mathbb{P}\left(\tilde{\lambda}_n^s \text{ is decreasing on } [b, 1-b]\right) \to 1.$$

Proof. The proof is completely similar to that of Lemma 3.6.14. Note that (3.1.7) follows from our Assumption (A2) and that here λ is a decreasing function.

We use the fact that on $[b, 1-b]$, $\tilde{\lambda}_n^s$ is the standard kernel estimator of λ given by (6.4.2) and we get

$$\frac{d}{dt}\tilde{\lambda}_n^s(t) = \int_{t-b}^{t+b}\frac{1}{b^2}k'\left(\frac{t-u}{b}\right)d\left(\Lambda_n-\Lambda\right)(u)+\int_{t-b}^{t+b}\frac{1}{b^2}k'\left(\frac{t-u}{b}\right)\lambda(u)\,du. \tag{B.2.1}$$

The first term on the right hand side of (B.2.1) converges to zero because in absolute value it is bounded from above by

$$\frac{1}{b^2}\sup_{x\in[0,1]}|\Lambda_n(x)-\Lambda(x)|\sup_{y\in[-1,1]}|k''(y)| = O_p(b^{-2}n^{-1/2}) = o_p(1),$$

according to Assumption (A2) and the fact that $1/b = o(n^{-1/4})$. Moreover, integration by parts gives

$$\int\frac{1}{b^2}k'\left(\frac{t-u}{b}\right)\lambda(u)\,du = \int_{-1}^{1}k(y)\lambda'(t-by)\,dy.$$

Hence, the second term on the right hand side of (B.2.1) is bounded from above by a strictly negative constant because of Assumption (A1). We conclude that $\tilde{\lambda}_n^s$ is decreasing on $[b, 1-b]$ with probability tending to one. □

Corollary B.2.2. *Assume (A1)-(A2) and let* $\tilde{\lambda}_n^s$ *and* $\tilde{\lambda}_n^{GS}$ *be defined in* (6.1.2) *and Section 6.4, respectively. Let* k *satisfy* (6.1.1). *Let* $0 < \gamma < 1$ *and* $p \geqslant 1$. *If* $b \to 0$, $nb \to \infty$, *and* $1/b = o(n^{1/4})$, *then*

$$\mathbb{P}\left(\tilde{\lambda}_n^s(t) = \tilde{\lambda}_n^{GS}(t) \textit{ for all } t \in [b^\gamma, 1-b^\gamma]\right) \to 1.$$

Proof. The proof is completely similar to that of Lemma 3.3.2, but now we want to extend the interval to $[b^\gamma, 1-b^\gamma]$, which is not fixed but approaches the boundaries as $n \to \infty$. In this case we define the linearly extended version of Λ_n^s by

$$\hat{\Lambda}_n^*(t) = \begin{cases} \Lambda_n^s(b^\gamma) + \left(t - b^\gamma\right)\tilde{\lambda}_n^s(b^\gamma), & \text{for } t \in [0, b^\gamma), \\ \Lambda_n^s(t), & \text{for } t \in [b^\gamma, 1-b^\gamma], \\ \Lambda_n^s(1-b^\gamma) + \left(t - 1 + b^\gamma\right)\tilde{\lambda}_n^s(1-b^\gamma), & \text{for } t \in (1-b^\gamma, 1]. \end{cases}$$

Choose $0 < \delta < 2$. It suffices to prove that, for sufficiently large n,

$$\mathbb{P}\left(\hat{\Lambda}_n^* \text{ is concave on } [0,1]\right) \geqslant 1 - \delta/2, \tag{B.2.2}$$

and

$$\mathbb{P}\left(\hat{\Lambda}_n^*(t) \geqslant \Lambda_n^s(t), \text{ for all } t \in [0,1]\right) \geqslant 1 - \delta/2. \tag{B.2.3}$$

To prove (B.2.2), define the event

$$A_n = \left\{\tilde{\lambda}_n^s \text{ is decreasing on } [b, 1-b]\right\}.$$

On the event A_n the curve $\hat{\Lambda}_n^*$ is concave on $[0,1]$, so

$$\mathbb{P}\left(\hat{\Lambda}_n^* \text{ is concave on } [0,1]\right) \geqslant \mathbb{P}(A_n),$$

and the result follows from Lemma B.2.1. To prove (B.2.3), we split the interval $[0,1]$ in five intervals $I_1 = [0, b)$, $I_2 = [b, b^\gamma)$, $I_3 = [b^\gamma, 1-b^\gamma]$, $I_4 = (1-b^\gamma, 1-b]$ and $I_5 = (1-b, 1]$. Then, as in Lemma 3.3.2, we show that

$$\mathbb{P}\left(\hat{\Lambda}_n^*(t) \geqslant \Lambda_n^s(t), \text{ for all } t \in I_i\right) \geqslant 1 - \delta/10, \quad i = 1, \ldots, 5. \tag{B.2.4}$$

For $t \in I_3$, $\hat{\Lambda}_n^*(t) = \Lambda_n^s(t)$, so (B.2.4) is trivial. For $t \in I_2$, by the mean value theorem,

$$\begin{aligned} \hat{\Lambda}_n^*(t) - \Lambda_n^s(t) &= \Lambda_n^s(b^\gamma) + \left(t - b^\gamma\right)\tilde{\lambda}_n^s(b^\gamma) - \Lambda_n^s(t) \\ &= (b^\gamma - t)\left[\tilde{\lambda}_n^s(\xi_t) - \tilde{\lambda}_n^s(b^\gamma)\right], \end{aligned}$$

for some $\xi_t \in (t, b^\gamma) \subset (b, b^\gamma)$. Thus,

$$\mathbb{P}\left(\hat{\Lambda}_n^*(t) \geqslant \Lambda_n^s(t), \text{ for all } t \in I_2\right) \geqslant \mathbb{P}(A_n) \geqslant 1 - \delta/10,$$

for n sufficiently large, according to Lemma B.2.1. The argument for I_4 is exactly the same.

Next, we consider $t \in I_1$. We have

$$\begin{aligned}
&\hat{\Lambda}_n^*(t) - \Lambda_n^s(t) \\
&= \Lambda_n^s(b^\gamma) + \left(t - b^\gamma\right)\tilde{\lambda}_n^s(b^\gamma) - \Lambda_n^s(t) \\
&= [\Lambda_n^s(b^\gamma) - \Lambda^s(b^\gamma)] + [\Lambda^s(t) - \Lambda_n^s(t)] + \Lambda^s(b^\gamma) - \Lambda^s(t) - \left(b^\gamma - t\right)\tilde{\lambda}_n^s(b^\gamma) \\
&\geqslant -2 \sup_{t\in[0,1]} |\Lambda_n^s(t) - \Lambda^s(t)| + \Lambda^s(b^\gamma) - \Lambda^s(t) - (b^\gamma - t)\lambda(b^\gamma) \\
&\qquad\qquad + \left(b^\gamma - t\right)\left[\lambda(b^\gamma) - \tilde{\lambda}_n^s(b^\gamma)\right],
\end{aligned} \tag{B.2.5}$$

where Λ^s is the deterministic version of Λ_n^s,

$$\Lambda^s(t) = \int_{(t-b)\vee 0}^{(t+b)\wedge 1} k_b^{(t)}(t-u)\Lambda(u)\,du.$$

For the first term on right hand side of (B.2.5), note that

$$\begin{aligned}
\sup_{t\in[0,1]} |\Lambda_n^s(t) - \Lambda^s(t)| &= \sup_{t\in[0,1]} \left| \int_{(t-b)\vee 0}^{(t+b)\wedge 1} k_b^{(t)}(t-u)\left[\Lambda_n(u) - \Lambda(u)\right] du \right| \\
&= \sup_{t\in[0,1]} \left| \int k^{(t)}(y)\left[\Lambda_n(t-by) - \Lambda(t-by)\right] dy \right| \\
&\leqslant \sup_{t\in[0,1]} |\Lambda_n(t-by) - \Lambda(t-by)| \int \sup_{t\in[0,1]} \left|k^{(t)}(y)\right| dy \\
&= O_P\left(n^{-1/2}\right),
\end{aligned} \tag{B.2.6}$$

due to Assumption (A2). Moreover, for the third term on right hand side of (B.2.5), for $t \in (b, 1-b)$, we have

$$\begin{aligned}
\left|\lambda(t) - \tilde{\lambda}_n^s(t)\right| &\leqslant \left|\lambda(t) - \int k_b(t-u)\lambda(u)\,du\right| + \left|\int k_b(t-u)\,d(\Lambda - \Lambda_n)(u)\right| \\
&= \left|\int k(y)[\lambda(t) - \lambda(t-by)]\,dy\right| + \frac{1}{b}\left|\int k'(y)(\Lambda - \Lambda_n)(t-by)\,dy\right| \\
&= O(b^2) + O_P(b^{-1}n^{-1/2}).
\end{aligned} \tag{B.2.7}$$

For the second term on right hand side of (B.2.5), for $t \in [0, b)$, we write

$$\begin{aligned}
&\Lambda^s(b^\gamma) - \Lambda^s(t) - (b^\gamma - t)\lambda(b^\gamma) \\
&= \int_{b^\gamma - b}^{b^\gamma + b} k_b(b^\gamma - u)\Lambda(u)\,du - \int_0^{t+b} k_b^{(t)}(t-u)\Lambda(u)\,du - (b^\gamma - t)\lambda(b^\gamma) \\
&= \int_{b^\gamma - b}^{b^\gamma + b} k_b(b^\gamma - u)[\Lambda(u) - \Lambda(b^\gamma)]\,du - \int_0^{t+b} k_b^{(t)}(t-u)[\Lambda(u) - \Lambda(t)]\,du \\
&\qquad\qquad + [\Lambda(b^\gamma) - \Lambda(t) - (b^\gamma - t)\lambda(b^\gamma)] \\
&= \int_{-1}^{1} k(y)[\Lambda(b^\gamma - by) - \Lambda(b^\gamma)]\,dy - \int_{-1}^{t/b} k^{(t)}(y)[\Lambda(t - by) - \Lambda(t)]\,dy \\
&\qquad\qquad - \frac{1}{2}(b^\gamma - t)^2\lambda'(\xi_t) \\
&\geqslant \int_{-1}^{1} k(y)[\Lambda(b^\gamma - by) - \Lambda(b^\gamma)]\,dy - \int_{-1}^{t/b} k^{(t)}(y)[\Lambda(t - by) - \Lambda(t)]\,dy \\
&\qquad\qquad - \inf_{t\in[0,1]} |\lambda'(t)| b^{1+\gamma} + \frac{1}{2}\inf_{t\in[0,1]} |\lambda'(t)| b^{2\gamma}
\end{aligned} \tag{B.2.8}$$

where $\xi_t \in (t, b^\gamma)$. Furthermore, the first two integrals on the right hand side can be written as

$$\begin{aligned}
&\frac{b^2}{2}\int_{-1}^{1} k(y)y^2\lambda'(\xi_{1,y})\,dy - \frac{b^2}{2}\int_{-1}^{t/b} k^{(t)}(y)y^2\lambda'(\xi_{2,y})\,dy \\
&\quad\geqslant -\frac{b^2}{2}\left|\int_{-1}^{1} k(y)y^2\lambda'(\xi_{1,y})\,dy - \int_{-1}^{t/b} k^{(t)}(y)y^2\lambda'(\xi_{2,y})\,dy\right| \\
&\quad\geqslant -\frac{b^2}{2}\left|\int_{-1}^{1} k(y)y^2\lambda'(\xi_{1,y})\,dy - \int_{-1}^{t/b} k^{(t)}(y)y^2\lambda'(\xi_{2,y})\,dy\right| = O(b^2),
\end{aligned}$$

with $\xi_t \in (t, b^\gamma)$, $|\xi_{1,y} - b^\gamma| \leqslant by$ and $|\xi_{2,y} - t| \leqslant by$. This means that

$$\mathbb{P}\left(\hat{\Lambda}_n^*(t) - \Lambda_n^s(t) \geqslant 0, \text{ for all } x \in I_1\right) \geqslant \mathbb{P}\left(Y_n \leqslant \frac{1}{2}\inf_{t\in[0,1]} |\lambda'(t)| b^{2\gamma}\right),$$

where

$$\begin{aligned}
Y_n &= O_P(n^{-1/2}) + O(b^\gamma)\left\{O(b^2) + O_P(b^{-1}n^{-1/2})\right\} \\
&\qquad + O(b^2) - \inf_{t\in[0,1]} |\lambda'(t)| b^{1+\gamma} \\
&= O_P(b^{1+\gamma}).
\end{aligned}$$

Hence, for n large enough, this probability is greater than $1 - \delta/10$, because $\gamma < 1$. □

B.3 CLT FOR THE HELLINGER LOSS

Lemma B.3.1. *Assume (A1)-(A3) hold. If λ is strictly positive, we have*

$$\int_0^1 \left(\sqrt{\hat{\lambda}_n^s(t)} - \sqrt{\lambda(t)}\right)^2 d\mu(t) = \int_0^1 \left(\hat{\lambda}_n^s(t) - \lambda(t)\right)^2 (4\lambda(t))^{-1} d\mu(t) + O_P\left((nb)^{-3/2}\right).$$

The previous results holds also if we replace $\hat{\lambda}_n^s$ with the smoothed Grenander-type estimator $\tilde{\lambda}_n^{SG}$.

Proof. As in the proof of Lemma 4.1.1, we get

$$\int_0^1 \left(\sqrt{\hat{\lambda}_n^s(t)} - \sqrt{\lambda(t)}\right)^2 d\mu(t) = \int_0^1 \left(\hat{\lambda}_n^s(t) - \lambda(t)\right)^2 (4\lambda(t))^{-1} d\mu(t) + R_n,$$

where

$$|R_n| \leqslant C \int_0^1 \left|\hat{\lambda}_n^s(t) - \lambda(t)\right|^3 d\mu(t)$$

for some positive constant C only depending on $\lambda(0)$ and $\lambda(1)$. Then, from Corollary 6.2.7, it follows that $R_n = O_P\left((nb)^{-3/2}\right)$. When dealing with the smoothed Grenander-type estimator, the result follows from Theorem 6.3.4. □

Theorem B.3.2. *Assume (A1)-(A3) hold and that λ is strictly positive.*

i) If $nb^5 \to 0$, then it holds

$$(b\sigma^{2,*}(2))^{-1/2}\left\{2nbH(\hat{\lambda}_n^s, \lambda)^2 - m_n^*(2)\right\} \xrightarrow{d} N(0,1).$$

ii) If $nb^5 \to C_0^2 > 0$ and B_n in Assumption (A2) is a Brownian motion, then it holds

$$(b\theta^{2,*}(2))^{-1/2}\left\{2nbH(\hat{\lambda}_n^s, \lambda)^2 - m_n^*(2)\right\} \xrightarrow{d} N(0,1),$$

iii) If $nb^5 \to C_0^2 > 0$ and B_n in Assumption (A2) is a Brownian bridge, then it holds

$$(b\tilde{\theta}^{2,*}(2))^{-1/2}\left\{2nbH(\hat{\lambda}_n^s, \lambda)^2 - m_n^*(2)\right\} \xrightarrow{d} N(0,1),$$

where $\sigma^{2,*}$, $\theta^{2,*}$, $\tilde{\theta}^{2,*}$ *and* m_n^* *are defined, respectively, as in* (6.1.9), (6.1.11), (6.1.12) *and* (6.1.8) *by replacing* $w(t)$ *with* $w(t)(4\lambda(t))^{-1}$.

If $p < \min(q, 2q-7)$ *and*

$$\frac{1}{b} = o\left(n^{(1/3-1/q)\min(q/(2p),1)}\right),$$

the same results hold also when replacing $\hat{\lambda}_n^s$ *by the smoothed Grenander-type estimator* $\tilde{\lambda}_n^{SG}$.

Proof. According to Lemma B.3.1, it is sufficient to show that the results hold if we replace $2H(\hat{\lambda}_n^s, \lambda)^2$ by

$$\int_0^1 \left(\hat{\lambda}_n^s(t) - \lambda(t)\right)^2 (4\lambda(t))^{-1}\, d\mu(t) = \int_0^1 \left(\hat{\lambda}_n^s(t) - \lambda(t)\right)^2 d\tilde{\mu}(t),$$

where

$$d\tilde{\mu}(t) = \frac{1}{4\lambda(t)}\, d\mu(t) = \frac{w(t)}{4\lambda(t)}\, dt.$$

It suffices to apply Corollary 6.2.7 with a weight $\tilde{\mu}$ instead of μ.

For the smoothed Grenander estimator the result would follow from Theorem 6.3.4. □

BIBLIOGRAPHY

Akakpo, Nathalie, Fadoua Balabdaoui, and Cécile Durot (2014). "Testing monotonicity via local least concave majorants." In: *Bernoulli* 20.2, pp. 514–544.

Albers, M.G. (2012). "Boundary Estimation of Densities with Bounded Support." Master's thesis. ETH Zürich.

Alizadeh Noughabi, Hadi and Naser Reza Arghami (2011). "Testing exponentiality based on characterizations of the exponential distribution." In: *J. Stat. Comput. Simul.* 81.11, pp. 1641–1651.

Anevski, D. and O. Hössjer (2006). "A general asymptotic scheme for inference under order restrictions." In: *Ann. Statist.* 34.4, pp. 1874–1930.

Anevski, Dragi and Ola Hössjer (2002). "Monotone regression and density function estimation at a point of discontinuity." In: *Journal of Nonparametric Statistics* 14.3, pp. 279–294.

Balabdaoui, Fadoua, Hanna Jankowski, Marios Pavlides, Arseni Seregin, and Jon A. Wellner (2011). "On the Grenander estimator at zero." In: *Stat. Sinica* 21.21, pp. 873–899.

Banerjee, Moulinath (2008). "Estimating monotone, unimodal and U-shaped failure rates using asymptotic pivots." In: *Statist. Sinica* 18.2, pp. 467–492.

Baraud, Yannick, Sylvie Huet, and Béatrice Laurent (2005). "Testing convex hypotheses on the mean of a Gaussian vector. Application to testing qualitative hypotheses on a regression function." In: *Ann. Statist.* 33.1, pp. 214–257.

Birgé, Lucien (1986). "On estimating a density using Hellinger distance and some other strange facts." In: *Probability Theory and Related Fields* 71.2, pp. 271–291.

Birgé, Lucien and Pascal Massart (1993). "Rates of convergence for minimum contrast estimators." In: *Probability Theory and Related Fields* 97.1-2, pp. 113–150.

Birke, M. and H. Dette (2007). "Testing strict monotonicity in nonparametric regression." In: *Math. Methods Statist.* 16.2, pp. 110–123.

Birke, Melanie and Natalie Neumeyer (2013). "Testing monotonicity of regression functions—an empirical process approach." In: *Scand. J. Stat.* 40.3, pp. 438–454.

Breslow, N. and J. Crowley (1974). "A large sample study of the life table and product limit estimates under random censorship." In: *Ann. Statist.* 2. Collection of articles dedicated to Jerzy Neyman on his 80th birthday, pp. 437–453.

Brunk, H. D. (1958). "On the estimation of parameters restricted by inequalities." In: *Ann. Math. Statist.* 29, pp. 437–454.

Burr, Deborah (1994). "A comparison of certain bootstrap confidence intervals in the Cox model." In: *J. Amer. Statist. Assoc.* 89.428, pp. 1290–1302.

Cheng, Kuang Fu and Pi Erh Lin (1981). "Nonparametric estimation of a regression function." In: *Z. Wahrsch. Verw. Gebiete* 57.2, pp. 223–233.

Cheng, Ming-Yen, Peter Hall, and Dongsheng Tu (2006). "Confidence bands for hazard rates under random censorship." In: *Biometrika* 93.2, pp. 357–366.

Chernoff, Herman (1964). "Estimation of the mode." In: *Ann. Inst. Statist. Math.* 16, pp. 31–41.

Chung, Daehyun and Myron N. Chang (1994). "An isotonic estimator of the baseline hazard function in Cox's regression model under order restriction." In: *Statist. Probab. Lett.* 21.3, pp. 223–228.

Cox, D. R. (1972). "Regression models and life-tables." In: *J. Roy. Statist. Soc. Ser. B* 34. With discussion by F. Downton, Richard Peto, D. J. Bartholomew, D. V. Lindley, P. W. Glassborow, D. E. Barton, Susannah Howard, B. Benjamin, John J. Gart, L. D. Meshalkin, A. R. Kagan, M. Zelen, R. E. Barlow, Jack Kalbfleisch, R. L. Prentice and Norman Breslow, and a reply by D. R. Cox, pp. 187–220.

Cox, D. R. and D. Oakes (1984). *Analysis of survival data*. Monographs on Statistics and Applied Probability. Chapman & Hall, London, pp. ix+201. ISBN: 0-412-24490-X.

Csörgő, Miklós, Edit Gombay, and Lajos Horváth (1991). "Central limit theorems for L_p distances of kernel estimators of densities under random censorship." In: *Ann. Statist.* 19.4, pp. 1813–1831.

Csörgő, Miklós and Lajos Horváth (1988). "Central limit theorems for L_p-norms of density estimators." In: *Probab. Theory Related Fields* 80.2, pp. 269–291.

Doss, Charles R. and Jon A. Wellner (2016). "Global rates of convergence of the MLEs of log-concave and s-concave densities." In: *The Annals of Statistics* 44.3, pp. 954–981.

Durot, Cécile (2002). "Sharp asymptotics for isotonic regression." In: *Probab. Theory Related Fields* 122.2, pp. 222–240.

Durot, Cécile (2007). "On the $\mathbb{L}_p$-error of monotonicity constrained estimators." In: *Ann. Statist.* 35.3, pp. 1080–1104.

Durot, Cécile, Piet Groeneboom, and Hendrik P. Lopuhaä (2013). "Testing equality of functions under monotonicity constraints." In: *J. Nonparametr. Stat.* 25.4, pp. 939–970.

Durot, Cécile, Vladimir N. Kulikov, and Hendrik P. Lopuhaä (2012). "The limit distribution of the L_∞-error of Grenander-type estimators." In: *Ann. Statist.* 40.3, pp. 1578–1608.

Durot, Cécile and Hendrik P. Lopuhaä (2014). "A Kiefer-Wolfowitz type of result in a general setting, with an application to smooth monotone estimation." In: *Electron. J. Stat.* 8.2, pp. 2479–2513.

Durot, Cécile and Hendrik P. Lopuhaä (2018). "Limit Theory in Monotone Function Estimation." In: *Statist. Sci.* 33.4, pp. 547–567.

Durot, Cécile and Anne-Sophie Tocquet (2003). "On the distance between the empirical process and its concave majorant in a monotone regression framework." In: *Ann. Inst. H. Poincaré Probab. Statist.* 39.2, pp. 217–240. ISSN: 0246-0203.

Eggermont, P. P. B. and V. N. LaRiccia (2000). "Maximum likelihood estimation of smooth monotone and unimodal densities." In: *Ann. Statist.* 28.3, pp. 922–947.

Friedman, J. and R. Tibshirani (1984). "The monotone smoothing of scatter plots." In: *Technometrics* 26, pp. 243–350.

Geer, Sara A (2000). *Empirical Processes in M-estimation*. Vol. 6. Cambridge university press, Cambridge.

Gijbels, Irène, Peter Hall, M. C. Jones, and Inge Koch (2000). "Tests for monotonicity of a regression mean with guaranteed level." In: *Biometrika* 87.3, pp. 663–673.

Gill, Richard D. (1984). "Understanding Cox's regression model: a martingale approach." In: *J. Amer. Statist. Assoc.* 79.386, pp. 441–447.

González-Manteiga, W., R. Cao, and J. S. Marron (1996). "Bootstrap selection of the smoothing parameter in nonparametric hazard rate estimation." In: *J. Amer. Statist. Assoc.* 91.435, pp. 1130–1140.

Grenander, Ulf (1956). "On the theory of mortality measurement. II." In: *Skand. Aktuarietidskr.* 39, 125–153 (1957).

Groeneboom, P. (1983). *Estimating a monotone density*. Vol. II. Proceedings of the Berkeley conference in honor of Jerzy Neyman and Jack Kiefer, Berkeley, Calif.: Wadsworth Statist./Probab. Ser., pp. 539–555.

Groeneboom, Piet (1989). "Brownian motion with a parabolic drift and Airy functions." In: *Probab. Theory Related Fields* 81.1, pp. 79–109.

Groeneboom, Piet, Gerard Hooghiemstra, and Hendrik P. Lopuhaä (1999). "Asymptotic normality of the L_1 error of the Grenander estimator." In: *Ann. Statist.* 27.4, pp. 1316–1347.

Groeneboom, Piet and Geurt Jongbloed (2010). "Generalized continuous isotonic regression." In: *Statist. Probab. Lett.* 80.3-4, pp. 248–253.

Groeneboom, Piet and Geurt Jongbloed (2012). "Isotonic L_2-projection test for local monotonicity of a hazard." In: *J. Statist. Plann. Inference* 142.7, pp. 1644–1658.

Groeneboom, Piet and Geurt Jongbloed (2013). "Smooth and non-smooth estimates of a monotone hazard." In: *From probability to statistics and back: high-dimensional models and processes*. Vol. 9. Inst. Math. Stat. (IMS) Collect. Inst. Math. Statist., Beachwood, OH, pp. 174–196.

Groeneboom, Piet and Geurt Jongbloed (2014). *Nonparametric estimation under shape constraints*. Vol. 38. Cambridge University Press.

Groeneboom, Piet and Geurt Jongbloed (2015). "Nonparametric confidence intervals for monotone functions." In: *Ann. Statist.* 43.5, pp. 2019–2054.

Groeneboom, Piet, Geurt Jongbloed, and Birgit I. Witte (2010). "Maximum smoothed likelihood estimation and smoothed maximum likelihood estimation in the current status model." In: *Ann. Statist.* 38.1, pp. 352–387.

Groeneboom, Piet and Jon A. Wellner (1992). *Information bounds and nonparametric maximum likelihood estimation*. Vol. 19. DMV Seminar. Basel: Birkhäuser Verlag, pp. viii+126. ISBN: 3-7643-2794-4.

Groeneboom, Piet and Jon A. Wellner (2001). "Computing Chernoff's distribution." In: *J. Comput. Graph. Statist.* 10.2, pp. 388–400.

Hall, Peter (1992). "Effect of bias estimation on coverage accuracy of bootstrap confidence intervals for a probability density." In: *Ann. Statist.* 20.2, pp. 675–694.

Hall, Peter and Ingrid Van Keilegom (2005). "Testing for monotone increasing hazard rate." In: *Ann. Statist.* 33.3, pp. 1109–1137.

Haywood, John and Estate Khmaladze (2008). "On distribution-free goodness-of-fit testing of exponentiality." In: *J. Econometrics* 143.1, pp. 5–18.

Hess, Kenneth R, Dan M Serachitopol, and Barry W Brown (1999). "Hazard function estimators: a simulation study." In: *Statistics in medicine* 18.22, pp. 3075–3088.

Huang, Jian and Jon A. Wellner (1995). "Estimation of a monotone density or monotone hazard under random censoring." In: *Scand. J. Statist.* 22.1, pp. 3–33.

Huang, Youping and Cun-Hui Zhang (1994). "Estimating a monotone density from censored observations." In: *Ann. Statist.* 22.3, pp. 1256–1274.

Jammalamadaka, S. Rao and Emanuele Taufer (2003). "Testing exponentiality by comparing the empirical distribution function of the normalized spacings with that of the original data." In: *J. Nonparametr. Stat.* 15.6, pp. 719–729.

Kalbfleisch, John D. and Ross L. Prentice (2002). *The statistical analysis of failure time data*. Second. Wiley Series in Probability and Statistics. Wiley-Interscience [John Wiley & Sons], Hoboken, NJ, pp. xiv+439.

Kiefer, J. and J. Wolfowitz (1976). "Asymptotically minimax estimation of concave and convex distribution functions." In: *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* 34.1, pp. 73–85.

Kim, Arlene K. H., A. Guntuboyina, and Richard J. Samworth (2016). "Adaptation in log-concave density estimation." In: *arXiv:1609.00861*.

Kim, Arlene K. H. and Richard J. Samworth (2016). "Global rates of convergence in log-concave density estimation." In: *The Annals of Statistics* 44.6, pp. 2756–2779.

Kim, JeanKyung and David Pollard (1990). "Cube root asymptotics." In: *Ann. Statist.* 18.1, pp. 191–219.

Komlós, János, Péter Major, and Gábor Tusnády (1975). "An approximation of partial sums of independent RV'-s, and the sample DF. I." In: *Probability theory and related fields* 32.1, pp. 111–131.

Kosorok, Michael R. (2008). "Bootstrapping the Grenander estimator." In: *Beyond parametrics in interdisciplinary research: Festschrift in honor of Professor Pranab K. Sen*. Vol. 1. Inst. Math. Stat. Collect. Beachwood, OH: Inst. Math. Statist., pp. 282–292.

Kulikov, Vladimir N. and Hendrik P. Lopuhaä (2004). "Testing for a monotone density using L_k-distances between the empirical distribution function and its concave majorant." In: *Report-028, Eurandom.*

Kulikov, Vladimir N. and Hendrik P. Lopuhaä (2005). "Asymptotic normality of the L_k-error of the Grenander estimator." In: *Ann. Statist.* 33.5, pp. 2228–2255.

Kulikov, Vladimir N. and Hendrik P. Lopuhaä (2006). "The limit process of the difference between the empirical distribution function and its concave majorant." In: *Statist. Probab. Lett.* 76.16, pp. 1781–1786. ISSN: 0167-7152.

Kulikov, Vladimir N. and Hendrik P. Lopuhaä (2008). "Distribution of global measures of deviation between the empirical distribution function and its concave majorant." In: *J. Theoret. Probab.* 21.2, pp. 356–377.

Le Cam, L. (1973). "Convergence of estimates under dimensionality restrictions." In: *The Annals of Statistics* 1, pp. 38–53.

Le Cam, Lucien (1970). "On the weak convergence of probability measures." In: *Annals of Mathematical Statistics* 41, pp. 621–625.

Lopuhaä, Hendrik P. and Gabriela F. Nane (2013). "Shape constrained nonparametric estimators of the baseline distribution in Cox proportional hazards model." In: *Scand. J. Stat.* 40.3, pp. 619–646.

Major, P. and L. Rejtő (1988). "Strong embedding of the estimator of the distribution function under random censorship." In: *Ann. Statist.* 16.3, pp. 1113–1132.

Mammen, Enno (1991). "Estimating a smooth monotone regression function." In: *Ann. Statist.* 19.2, pp. 724–740.

Marron, J. S. and W. J. Padgett (1987). "Asymptotically optimal bandwidth selection for kernel density estimators from randomly right-censored samples." In: *Ann. Statist.* 15.4, pp. 1520–1535.

Meintanis, Simos G. (2007). "Test for exponentiality against Weibull and gamma decreasing hazard rate alternatives." In: *Kybernetika (Prague)* 43.3, pp. 307–314. ISSN: 0023-5954.

Mukerjee, Hari (1988). "Monotone nonparameteric regression." In: *Ann. Statist.* 16.2, pp. 741–750.

Müller, Hans-Georg and Jane-Ling Wang (1990). "Locally adaptive hazard smoothing." In: *Probab. Theory Related Fields* 85.4, pp. 523–538.

Nane, Gabriela F (2013). *Shape Constrained Nonparametric Estimation in the Cox Model*. Delft University of Technology, Delft.

Nane, Gabriela F (2015). "A likelihood ratio test for monotone baseline hazard functions in the Cox model." In: *Statistica Sinica* 25, pp. 1163–1184.

Pal, Jayanta Kumar, Michael Woodroofe, and Mary Meyer (2007). "Estimating a Polya frequency function$_2$." In: *Complex datasets and inverse problems*. Vol. 54. IMS Lecture Notes Monogr. Ser. Inst. Math. Statist., Beachwood, OH, pp. 239–249.

Prakasa Rao, B. L. S. (1969). "Estimation of a unimodal density." In: *Sankhyā Ser. A* 31, pp. 23–36. ISSN: 0581-572X.

Prakasa Rao, B. L. S. (1970). "Estimation for distributions with monotone failure rate." In: *Ann. Math. Statist.* 41, pp. 507–519. ISSN: 0003-4851.

Ramlau-Hansen, Henrik (1983). "Smoothing counting process intensities by means of kernel functions." In: *Ann. Statist.* 11.2, pp. 453–466.

Ramsay, J. O. (1998). "Estimating smooth monotone functions." In: *J. R. Stat. Soc. Ser. B Stat. Methodol.* 60.2, pp. 365–375.

Ren, Jian-Jian and Mai Zhou (2011). "Full likelihood inferences in the Cox model: an empirical likelihood approach." In: *Ann. Inst. Statist. Math.* 63.5, pp. 1005–1018.

Sen, Bodhisattva, Moulinath Banerjee, and Michael Woodroofe (2010). "Inconsistency of bootstrap: the Grenander estimator." In: *Ann. Statist.* 38.4, pp. 1953–1977.

Seregin, Arseni and Jon A. Wellner (2010). "Nonparametric estimation of multivariate convex-transformed densities." In: *The Annals of Statistics* 38.6. With supplementary material available online, pp. 3751–3781.

Shorack, Galen R. and Jon A. Wellner (1986). *Empirical processes with applications to statistics*. Wiley Series in Probability and Mathematical Statistics:

Probability and Mathematical Statistics. John Wiley & Sons, Inc., New York, pp. xxxviii+938.

Tanner, Martin A. and Wing Hung Wong (1983). "The estimation of the hazard function from randomly censored data by the kernel method." In: *Ann. Statist.* 11.3, pp. 989–993.

Tsiatis, Anastasios A. (1981). "A large sample study of Cox's regression model." In: *Ann. Statist.* 9.1, pp. 93–108.

Van de Geer, Sara A. (2000). *Applications of empirical process theory*. Vol. 6. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge, pp. xii+286.

Van der Vaart, Aad W. and Mark J. van der Laan (2003). "Smooth estimation of a monotone density." In: *Statistics* 37.3, pp. 189–203.

Van der Vaart, Aad W. and Jon A. Wellner (1996). *Weak convergence and empirical processes*. Springer Series in Statistics. With applications to statistics. Springer-Verlag, New York, pp. xvi+508.

Van Geloven, N., I. Martin, P. Damman, R. J. de Winter, J. G. Tijssen, and H. P. Lopuhaä (2013). "Estimation of a decreasing hazard of patients with acute coronary syndrome." In: *Stat. Med.* 32.7, pp. 1223–1238.

Wang, Xiao and Michael Woodroofe (2007). "A Kiefer-Wolfowitz comparison theorem for Wicksell's problem." In: *Ann. Statist.* 35.4, pp. 1559–1575.

Wang, Yazhen (1994). "The limit distribution of the concave majorant of an empirical distribution function." In: *Statist. Probab. Lett.* 20.1, pp. 81–84.

Wellner, Jon A. (2015). "Musings about shape constrained estimation and inference: some problems." In: *Presentation at workshop Shape Constrained Inference: Open Problems and New Directions*, Lorentz Centre, Leiden, The Netherlands, 2015.

Wells, Martin T. (1994). "Nonparametric kernel estimation in counting processes with explanatory variables." In: *Biometrika* 81.4, pp. 795–801.

Westling, Ted and Marco Carone (2018). "A unified study of nonparametric inference for monotone functions." In: *arXiv preprint arXiv:1806.01928*.

Wong, Wing Hung and Xiaotong Shen (1995). "Probability inequalities for likelihood ratios and convergence rates of sieve MLEs." In: *Ann. Statist.* 23.2, pp. 339–362.

Woodroofe, Michael and Jiayang Sun (1993). "A penalized maximum likelihood estimate of $f(0+)$ when f is nonincreasing." In: *Statist. Sinica* 3.2, pp. 501–515. ISSN: 1017-0405.

Wright, F. T. (1982). "Monotone regression estimates for grouped observations." In: *Ann. Statist.* 10.1, pp. 278–286.

Xu, Gongjun, Bodhisattva Sen, and Zhiliang Ying (2014). "Bootstrapping a change-point Cox model for survival data." In: *Electron. J. Stat.* 8.1, pp. 1345–1379.

Zhang, Shunpu and Rohana J Karunamuni (1998). "On kernel density estimation near endpoints." In: *Journal of Statistical Planning and Inference* 70.2, pp. 301–316.

SUMMARY

In this thesis we address the problem of estimating a curve of interest (which might be a probability density, a failure rate or a regression function) under monotonicity constraints. This problem arises in different fields of application. For example, in survival analysis, a monotone hazard rate reflects the property of aging, meaning that components become less reliable as the survival time increases. Monotonic regression relationships are reasonable in econometrics (demand-price), biometrics (age-height) etc. In such situations, incorporating monotonicity constraints in the estimation procedure leads to more accurate results. There is a large body of literature on isotonic estimation which focuses on two methods: constrained nonparametric maximum likelihood estimation and a Grenander-type procedure. These estimators are piecewise constant and converge at rate cube-root n. On the other hand, smooth estimators are usually preferred because, among other reasons, they have a faster rate of convergence and a nicer graphical representation.

The main concern of this thesis is investigating large sample distributional properties of smooth isotonic estimators. In the first part we focus on the pointwise behavior of estimators for the hazard rate in survival analysis while the second part is dedicated to global errors of estimators in a general setup, which includes estimation of a probability density, a failure rate, or a regression function.

In Chapter 2 we consider kernel smoothed Grenander-type estimators for a monotone hazard or a monotone density in the right censoring model. Weak convergence of the estimators at a fixed point is established through a Kiefer-Wolfowitz type of result and the use of a boundary corrected kernel is proposed to avoid inconsistency at the boundary regions. Once having derived the asymptotic distribution of the estimators, a practical application is constructing pointwise confidence intervals. Simulations show that smoothing leads to more accurate results and undersmoothing is preferred with respect to bias estimation.

Chapter 3 focuses on smooth isotonic estimation of the baseline hazard in the Cox regression model, which is a generalization of the right censoring model that takes into account the presence of covariates. Four different estimators are obtained by combining an isotonization step with a smoothing step and alternating the order of smoothing and isotonization. We show that

three of the estimators are asymptotically equivalent with a Gaussian limit distribution at rate $n^{2/5}$ while the fourth one has a different asymptotic bias. Hence, there is no preference between the four methods on the basis of large samples theoretical properties. Moreover, we assess the finite sample performance of the estimators by means of simulation studies for constructing pointwise confidence intervals. We investigate both asymptotic confidence intervals using undersmoothing to avoid bias estimation and bootstrap confidence intervals. We observe that the second method performs better and the estimators have comparable behavior (with smoothed maximum likelihood estimator and maximum smoothed likelihood estimator being slightly more accurate). However, it is presently not clear how to "optimally" choose the smoothing parameter which in practice might be an issue.

Chapter 4 provides a central limit theorem for the Hellinger error of Grenander-type estimators of a monotone function obtained by approximating the squared Hellinger distance with a weighted L_2-distance. A goodness of fit test under the assumption of a non-increasing density is proposed based on the Hellinger distance between a parametric estimator and the Grenander-type estimator. Its performance is investigated through a simulation study on testing exponentiality.

In Chapter 5 we proceed by considering the process $\widehat{\Lambda}_n - \Lambda_n$, where Λ_n is a cadlag step estimator for the primitive Λ of a nonincreasing function λ on $[0, 1]$, and $\widehat{\Lambda}_n$ is the least concave majorant of Λ_n. We extend the results in Kulikov and Lopuhaä, 2006, 2008 to the general setting considered in Durot, 2007. Under this setting we prove that a suitably scaled version of $\widehat{\Lambda}_n - \Lambda_n$ converges in distribution to the corresponding process for two-sided Brownian motion with parabolic drift and we establish a central limit theorem for the L_p-distance between $\widehat{\Lambda}_n$ and Λ_n. Such result is then used in the Chapter 6 for dealing with smoothed Grenander-type estimators.

Finally, in Chapter 6 we provide central limit theorems for the L_p-error of smooth isotonic estimators obtained by kernel smoothing the Grenander-type estimator or isotonizing the kernel estimator. Both of them are connected to the behavior of the L_p-error of the ordinary kernel estimator, for which we extend the results from Csörgő and Horváth, 1988. However, differently from Csörgő and Horváth, 1988 we find that the limit variance changes depending on whether the approximating process of $\Lambda_n - \Lambda$ is a Brownian motion or a Brownian bridge. As an application we consider testing monotonicity of a regression function on the basis of the L_2-distance between the ordinary kernel estimator and the smoothed Grenander estimator. The limit theorem cannot be directly used for determining the critical region of the test due to the presence of unknown parameters in the limit

distribution. Instead, we use a bootstrap procedure and compare the performance of the test with other tests available in the literature.

SAMENVATTING

In dit proefschrift behandelen we het probleem van het schatten van een curve (zoals een kansdichtheid, een hazard functie, of een regressiefunctie) onder de aanname van monotonie. Dit probleem doet zich voor in verschillende toepassingsgebieden. Bijvoorbeeld in de analyse van levensduren wordt veroudering weerspiegeld door een monotone hazard functie, evenals het minder betrouwbaar worden van onderdelen naarmate de overlevingstijd toeneemt. Monotone regressierelaties komen vaak voor in de Economie (bijvoorbeeld tussen de omvang van de vraag en de prijs), in de Biologie (bijvoorbeeld tussen leeftijd en lengte), enz. In dergelijke situaties leidt het opnemen van monotonie in de schattingsprocedure tot nauwkeuriger resultaten. Er is een uitgebreide literatuur over isotone schatters, die zich richt op twee methoden: de niet-parametrische maximum likelihood schatter en een soort van Grenander-schatter. Deze schatters zijn stuksgewijs constant en convergeren met snelheid $n^{1/3}$. Aan de andere kant, hebben gladde schatters meestal de voorkeur omdat ze, naast andere redenen, een hogere convergentiesnelheid hebben en een mooiere grafische weergave.

Het belangrijkste doel van dit proefschrift is het onderzoeken van asymptotische verdelingseigenschappen van gladde isotone schatters. In het eerste deel concentreren we ons op het puntsgewijze gedrag van schatters voor de hazard functie bij de analyse van levensduren, terwijl het tweede deel gewijd is aan globale afwijkingen in een algemene opzet die het schatten van een kansdichtheid, een hazard functie, en een regressiefunctie bevat.

In Hoofdstuk 2 beschouwen we glad gemaakte Grenander-schatters voor een monotone hazard en een monotone kansdichtheid in het rechts censurerings model. Puntsgewijze zwakke convergentie van de schatters wordt aangetoond met behulp van een soort van Kiefer-Wolfowitz resultaat, gebruikmakend van van een rand-gecorrigeerde kernfunctie om inconsistentie aan de rand te voorkomen. Nadat de asymptotische verdeling van de schatters is afgeleid, is een praktische toepassing het construeren van puntsgewijze betrouwbaarheidsintervallen. Simulaties laten zien dat het glad maken leidt tot meer accurate resultaten en dat "undersmoothing" de voorkeur heeft boven het schatten van de "bias".

Hoofdstuk 3 richt zich op gladde isotone schatters voor de baseline hazard in het Cox model, een generalisatie van het rechts-censurerings model

dat rekening houdt met de aanwezigheid van covariaten. Vier verschillende schatters worden verkregen door twee soorten van isotonisatie te combineren met een "smoothing"-stap in twee verschillende volgorden. We laten zien dat de schatters asymptotisch normaal verdeeld zijn en consistent zijn met snelheid $n^{2/5}$. Drie van de schatters zijn asymptotisch equivalent terwijl de vierde schatter een andere asymptotische bias heeft. Derhalve is er geen voorkeur voor een van de vier methoden op basis van theoretische eigenschappen voor grote steekproeven. Verder onderzoeken we de prestaties van de schatters bij eindige steekproeven bij het construeren van puntsgewijze betrouwbaarheidsintervallen door middel van simulaties. We onderzoeken zowel asymptotische betrouwbaarheidsintervallen, met behulp van "undersmoothing" om het schatten van de bias te vermijden, alsmede bootstrap-betrouwbaarheidsintervallen. We zien dat de tweede methode beter presteert en dat de schatters een vergelijkbaar gedrag vertonen (waarbij de glad gemaakte maximum likelihood schatter en de maximum smoothed likelihood schatter iets nauwkeuriger zijn). Het is echter momenteel niet duidelijk hoe de gladheidsparameter "optimaal" kan gekozen worden, hetgeen in de praktijk een probleem kan zijn.

Hoofdstuk 4 bevat een centrale limietstelling voor de afwijking in Hellinger-afstand voor Grenander-schatters voor een monotone functie, verkregen door de gekwadrateerde Hellinger-afstand te benaderen met een gewogen L_2-afstand. Onder de aanname van een niet-stijgende kansdichtheid stellen we een goodness-of-fit toets voor op basis van de Hellinger-afstand tussen een parametrische schatter en de Grenander-schatter. De prestaties hiervan worden onderzocht door middel van simulaties voor het toetsen van exponentialiteit.

In Hoofdstuk 5 gaan we verder met het proces $\widehat{\Lambda}_n - \Lambda_n$, waarbij Λ_n een stuksgewijs constante schatter is voor de primitieve Λ van een niet-stijgende functie λ op $[0, 1]$, en $\widehat{\Lambda}_n$ de kleinste concave majorant is van Λ_n. We breiden de resultaten in Kulikov and Lopuhaä, 2006, 2008 uit naar de algemene opzet die in Durot, 2007 wordt gebruikt. Voor deze opzet bewijzen we dat een geschikt geschaalde versie van $\widehat{\Lambda}_n - \Lambda_n$ in verdeling convergeert naar het overeenkomstige proces voor een tweezijdige Brownse beweging met parabolische drift. Bovendien leiden we een centrale limietstelling af voor de L_p-afstand tussen $\widehat{\Lambda}_n$ en Λ_n. Dit resultaat wordt vervolgens gebruikt in Hoofdstuk 6 voor het behandelen van glad gemaakte Grenander-schatters.

Tenslotte behandelen we in Hoofdstuk 6 centrale limietstellingen voor de afwijking in L_p-afstand voor gladde isotone schatters die verkregen zijn door glad maken van een Grenander-schatters of door het isotoniseren van een kernschatter. In beide gevallen is het gedrag gerelateerd aan dat van

de L_p-afstand voor een reguliere kernschatter, waarvoor we de resultaten van Csörgő and Horváth, 1988 uitbreiden. Echter, anders dan in Csörgő and Horváth, 1988, zien we dat de limietvariantie verschillend is, afhankelijk van of het benaderende proces voor $\Lambda_n - \Lambda$ een Brownse beweging is of een Brownse brug. Als een toepassing beschouwen we het toetsen van monotonie van een regressiefunctie op basis van de L_2-afstand tussen de kernschatter en de glad gemaakte Grenander-schatter. De limietstelling kan niet rechtstreeks worden gebruikt voor het bepalen van het kritieke gebied van de toets vanwege de aanwezigheid van onbekende parameters in de limietverdeling. In plaats daarvan gebruiken we een bootstrap procedure en vergelijken we de prestaties van de toets met andere toetsen die in de literatuur beschikbaar zijn.

ACKNOWLEDGEMENTS

At the end of a four year life-changing experience, I would like to take a moment to reflect on how I got here. It has been a period of learning, at both scientific and the personal level. However, I did not achieve everything by myself. Throughout this period, I have received help and support from numerous people and it is a pleasure to pen down this thank you note to all of them.

First, I would like express my sincere gratitude to my supervisor for providing me the opportunity to join TU Delft and work on this PhD project. Thank you Rik for helping me evolve as an independent researcher through your guidance and supervision, for allowing me to knock on your door whenever I needed help, for answering my emails even when on holiday and for always trusting my work (more than I did myself). Your suggestions and feedbacks have been essential to the success of this thesis. I learned a lot from your expertise. Next, I would like to thank Geurt, in particular for his inspiring enthusiasm. Thank you Geurt for always talking to me in Dutch (and Italian) and for your efforts in learning Albanian. I appreciate that you have both been very supportive of my work and have continuously encouraged me to grow as a researcher. Your advice and suggestions on my research and career have been valuable. On a lighter note, I would like to thank you both for changing your seats on our way back from Banff to keep me company during the flight.

Sincere thanks go to my Italian professor, Maurizio Pratelli, for being my first academic guide and for instigating, through his lectures, my passion for probability and statistics. Not only he accompanied me during the bachelor and master studies but also introduced me to this PhD project. He encouraged me to apply and made sure that everything was going well. Thank you professor for helping, supporting and motivating me in all these years.

Apart from the work presented in this thesis, in the last four years I have had the opportunity to work on other projects as well. Thank you Dario for guiding me through my first journal article which resulted from my master thesis. It was immensely helpful. Also, I am glad to have worked with Juan in a different field of statistics. This has been possible thanks to her belief in my capabilities and her introduction to the topic. I am grateful to Cecile who (together with Rik) made my two-month research visit in Paris possible. I greatly appreciate her support and I enjoyed working with her. Moreover, I would like to thank all the members of the Statistics and Applied Probability group at TU Delft, in particular the PhD students, for

creating a pleasant working environment. Thank you also to Carl for always solving my computer problems and for his daily lunch call. I am grateful also to my dissertation committee for devoting their time to read my thesis.

I would like to take this opportunity to express my thanks to those friends that made my time in Delft enjoyable and unforgettable. Starting with Gaia who, with her honest opinions about life and work in the Netherlands, gave me the courage to undertake this journey. Thank you Gaia for being so helpful when I first arrived in Delft, for being a friend to whom I could always turn for advice and for your company outside the office hours. Francesca is the second person I met in Delft when I came for the interview in November 2014. I remember having breakfast together at Juliana hotel and then the long straight walk towards the highest building of the campus. Your company made that day easier for me but I had no idea that, later on, you would become such an important person in my life. Thank you Fra for your kind support throughout the last three years, for your valuable advices, for always finding a solution to my problems. The triad would not have been complete without Martina, the spirit of all the fun activities we have done. Thank you Marti for being the best office mate and the best personal guide in Paris but, most of all, for being always there when I needed. I will never forget the wonderful time the three of us have spend together and our daily discussions sharing each-others troubles and happiness. I would also like to thank Federico, not only because he asked me to do so, but also for his company in the Dutch classes and for the long random walks in Leiden and Delft. Special mention goes also to Andrea, Lixue, Dan, Gioia and Clarissa for being very good friends and for the time we have spent together.

Words cannot express how grateful I am to my parents for their efforts to offer me a better life. I cannot thank them enough for always encouraging me to move forward and search the best for me (even when it is not the best for them), for supporting my choices and being every day beside me during the happy and hard moments. I owe everything to them. I would like to thank my best (and only) brother for always cheering me up and for trying to make me realize that there are more important things in life than study and work. I thank my two special aunts who are like sisters to me and, of course, also Leon for loving me as much as he loves Santa Claus. Last, but not least, thank you Birbal for being so supportive, for patently listening my complaints about things that do not go as I want, for helping me feel more self-confident and optimistic. You are the only person that manages to make me stop thinking and worrying too much, but instead enjoy the present. Thank you for being in my life.

THANK YOU!

CURRICULUM VITAE

Eni Musta was born on September 15, 1991 in Elbasan, Albania. She completed her pre-university education at Dhaskal Todri high school in Elbasan, in 2010 and in the same year she began her studies in Mathematics at the University of Pisa in Italy. She obtained her B.Sc. degree *cum laude* in 2013 with a thesis on infinitely divisible distributions and applications to Lévy processes. From 2013 to 2015 she continued her Master studies at the University of Pisa and wrote a thesis on Malliavin Calculus with applications to statistical inference under the supervision of Prof. M. Pratelli and Dr. D. Trevisan. She obtained her M.Sc. degree *cum laude* in March 2015 and in April of the same year she started her PhD research in Statistics at Delft University of Technology under the supervision of Dr. H. P. Lopuhaä and Prof. dr. ir. G. Jongbloed.

PUBLICATIONS

with H. P. Lopuhaä. (2017) "Smooth estimation of a monotone hazard and a monotone density under random censoring". *Statictica Neerlandica 71.1*, pp. 58-82.

with H. P. Lopuhaä. (2017) "Isotonized smooth estimators of a monotone baseline hazard in the Cox model". *Journal of Statistical Planning and Inference 191*, pp. 43-67.

with H. P. Lopuhaä. (2018) "Smoothed isotonic estimators of a monotone baseline hazard in the Cox model". *Scandinavian Journal of Statistics 45.3*, pp. 753-791.

with H. P. Lopuhaä. (2018) "A central limit theorem for the Hellinger loss of Grenander-type estimators". *To appear in Statistica Neerlandica*.

with H. P. Lopuhaä. (2018) "The distance between a naive cumulative estimator and its least concave majorant". *Statistics and Probability Letters 139*, pp. 119-128.

with H. P. Lopuhaä. (2018) "Central limit theorems for global errors of smooth isotonic estimators". *Major revision requested by Electronic Journal of Statistics*.